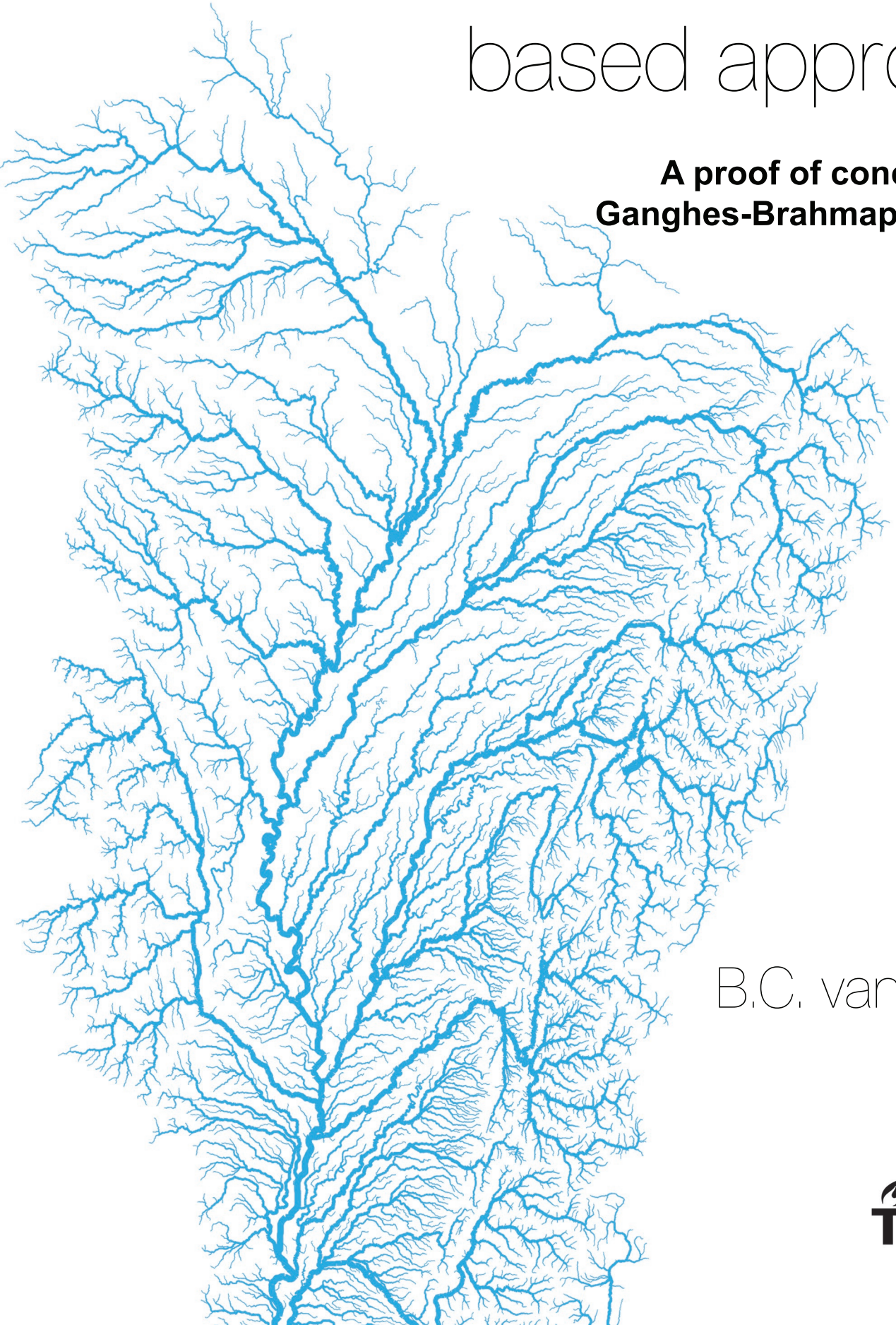
Riverine flood risk screening with a simple network-based approach

A proof of concept in the Ganghes-Brahmaputra basin



B.C. van Meurs

TU Delft

Riverine flood risk screening with a simple network-based approach

A proof of concept in the Ganghes-Brahmaputra basin

by

Bram Christiaan van Meurs

to obtain the degree of **Master of Science** in Engineering and Policy Analysis,
at the Faculty of Technology, Policy and Management,
Delft University of Technology,
to be defended publicly on Monday, December 9th, 2019 at 15:00.

Student number: 4177614

Committee:	Dr. M. E. Warnier,	Chair,	Systems Engineering section
	Dr. ir. T. Verma,	First Supervisor,	Policy Analysis section
	Dr. ir. J. H. Kwakkel,	Second Supervisor,	Policy Analysis section

Documentation can be found at https://rna.readthedocs.io/
and source code at https://github.com/bcvanmeurs/rna

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Executive summary

Floods cause major problems around the world. Over 35 million people were affected by floods in 2018. They have a growing worldwide impact on life and property. Changes in climate conditions lead to unanticipated variations in glacial runoffs, snowmelt and precipitation, which all significantly changing river flows. An imbalance in river network equilibrium leads to flooding and often ends up causing tremendous damage to society and the environment. Regions that are perceived to be downstream from the source of flooding may end up taking the brunt of the river force due to flood cascades. Floods account for about a third of all natural catastrophes worldwide, they cause more than half of all fatalities and are responsible for a third of the overall economic loss.

Modelling approaches are often used to determine flood consequences. Two types of flood models are commonly used: statistical models and flow simulation models. Statistical methods are easy to use but provide limited insight into flood problems. Flow simulation models' results can be very accurate, especially for hydraulic simulation models. However, these models are expensive to use and develop, and they require a lot of data. These requirements make them unsuitable for application in developing countries and analysing large watersheds. Flood risk screening models try to solve these problems. They are suitable for use in data-sparse regions and are efficient in terms of computational costs.

However, there is a lack of knowledge between river structure and cascading flood effects, and there is a lack of models that are efficient, easy to understand, use topological data and have the purpose of risk screening. In this research, we show a flood model based on complex network theory to efficiently study the cascading effects of floods in riverine systems. Cascading effects are defined as floods that occur as a result of water waves through the system that originate from upstream sources. The developed model uses the hydrological Muskingum routing method.

We found that it was possible, notwithstanding many assumptions and a lack of data, to reproduce system behaviour during an extreme flood event in the Ganges-Brahmaputra Basin. Satellite elevation data were used to construct the river network, and satellite precipitation data was used to feed the model. The model can indicate high risk reaches based on the simulated overflow, the flow exceeding a predefined capacity. No existing models are known that can do this, on a laptop, within seven minutes per simulated day, with limited data for a watershed that exceeds the size of one million square kilometres. The network structure of the model makes it possible to achieve a better understanding between river typology and cascading flood effects.

The model is not without its limitations. It cannot pinpoint when and where floods will occur, because it only calculates overflow. Moreover, flood failure mechanisms are not yet included in this model. Failure mechanisms will change model behaviour: when a flood occurs water temporarily leaves the system, which reduces downstream risk. Overflow cascades, therefore, would be shorter in reality than in this model.

The model is a proof of concept that shows the potential of a network theory-based risk screening method in flood simulation context. Its properties make it suitable for analysing the effects of changing precipitation patterns, which, for example, could originate from climate change studies. Another use case is real-time forecasting of discharge levels if the mode is combined with real-time discharge levels and precipitations forecasts. The model can be used as an early warning system: alerting when and where high discharge levels are expected.

We anticipate our model to be a starting point for policy screening and scenario analysis. Suggestions are made to include policy options within the model. Policy analysts can then use the model to compare different policy interventions for all kinds of (future) scenarios. The model should not be seen as a replacement of the advanced hydraulic simulation models, but as a complementary tool useful at an earlier moment in a design process with the purpose of screening options. Ultimately it can become a framework with the aim to support informed decision-making.

Preface

I was looking for a topic for my second thesis, after graduating in the Econometrics master Operations Research and Quantitative Logistics at Erasmus University. During an internship, I developed an optimisation algorithm for KLM Cargo The algorithm optimised the transportation of Cargo between the warehouse and the aircraft. While very interesting to develop and research, it was also a very practical topic and an application of existing knowledge. Therefore, when I got to opportunity to work on a more theoretical and new subject, I thought it would add nicely to my (academic) development.

This is why I was (and I still am) excited when Trivik explained to me an idea to develop a method to analyse rivers from a network analysis perspective. This was new and not really done before. It could lead to new insights and understanding of river structures. If successful, it could also help to study policy options to reduce river flood risk in developing countries. Ambitions were sky-high. I was going to solve everything.

But after the literature review and some discussions, I realised I had to take a step back. I wanted my model to be on par with existing models that have been in development for years. Something Martijn pointed out, is unrealistic. Other researchers have spent their whole career in forecasting floods. After all, this is a master thesis project of only six months. I am not going to solve the worldwide flood problem in this short period or develop a method that can do everything.

So I learnt to adjust my expectations: The development of a new method begins with baby steps and leads to exciting new questions. Because it is a new approach, I had the freedom to steer my own research, which I also found challenging sometimes. Challenging because there is uncertainty. There is uncertainty in which way to go, and if that direction would lead to any results. Nonetheless, I developed a working prototype that shows cascading effects and reproduces system behaviour. I now understand the different types of research, application versus exploration. Both come with their own challenges and solution strategies. Because of these challenges in explorative research, I am even more proud of the results.

Acknowledgements

Firstly, I would like to thank Trivik Verma, for guiding me through this process. We had many very constructive meetings, and I am proud to be your first graduate student in Delft. It was nice to discuss on a high level how to tell the story. Secondly, many thanks to Martijn Warnier for being very involved, from a distance, and helping me when challenges seemed to become burdens.

I would also like to thank all my friends for the fun time I have had during my whole study career. But special thanks to Laurane, with whom I studied the entire duration of the thesis and made the process less lonely. Of course, special thanks to my family, especially my parents, for always supporting me in all circumstances. Lastly, I would like to thank my partner, Tim, for his (everlasting) support and feedback.

I wish you pleasant reading,

Bram Christiaan van Meurs MSc
Delft, November 2019

Contents

List of Figures

1

Introduction

June 2013: early monsoon rainfall hits India. In a couple of days, more than 300 millimetres of rain has fallen in Northern India and Nepal. On June 21st Associated Press reports a death toll nearing 600. Landslides devastated large regions and complete villages. Over 30,000 people have been evacuated, and over 50,000 people were stranded. Floods washed away roads and bridges. Repairing the consequences of this disaster took months (Maqbool, 2013).

This year, 2019, the heaviest monsoon rains in 25 years have hit India. More than 1,600 people were killed by the floods since June, according to government data. The monsoon has already delivered 10% more rain than the 50-year average. The monsoon in this region typically lasts from June to September, but this year it is expected to withdraw more than a month later than usual. The floods brought down thousands of homes and large parts of farmland were submerged. Experts argue that the Indian government faces problems with flood prevention and forecasting (Reuters, 2019).

These are two Indian examples, but floods cause many problems around the world. Even more so, an increasing number of floods is linked to climate change. There is a growing worldwide impact on life and property. A study from 2000 shows that: floods account for about a third of all natural catastrophes, they cause more than half of all fatalities and they are responsible for a third of the overall economic loss. Floods cause the largest number of deaths in poor and densely populated countries of the world, particularly those in Asia (Knight & Shamseldin, 2005).

More recent numbers from the World Meteorological Organization (WMO) show that over 35 million people were affected by floods in 2018. Over two million people were displaced due to weather and climate-linked disasters. An estimated 30% of these displacements is linked to floods (World Meteorological Organization, 2019). Damages related to floods are expected to increase further.

On the one hand, climate change will cause more extreme precipitation events, even if though total precipitation decreases (Trenberth, 2011). These more intense rain patterns cause an increase in flood risk. On the other hand, population growth and pressure on land use increase the flood consequences or societal exposure. Most major cities are located on main river systems. Migration towards these cities is likely to increase. Sometimes there is no choice but to settle in exposed areas, such as in Bangladesh. This effect of increased societal exposure is called the expanding bull's eye effect (Ashley, Strader, Rosencrants, & Krmenec, 2014). It has great value to understand how and when floods happen because floods often have a huge impact on both economic and personal level. Thus, it would be great if it is possible to determine the high-risk regions in a river basin.

Flood basics There are different types of floods: coastal, pluvial and fluvial floods. Coastal floods occur in areas close to large bodies of open water. They are mainly the result of extreme tidal conditions caused by severe weather. Low-lying land is extra vulnerable to these types of floods. Pluvial floods are surface floods and occur when heavy rainfall creates a flood event independent of a water body. There are two common types: urban floods and hillside floods. Urban floods can occur when oversaturated drainage systems cause water flowing into streets. Hillside floods happen when hillsides are unable to absorb enough water. This research focuses on fluvial floods. Fluvial floods or riverine floods occur when excessive rainfall or snowmelt causes a river to exceed its capacity. A natural river can have overbank flooding, which occurs when water overflows the edges of a river. If a river has levees, overtopping or breaches can occur.

1.1. Background

Modelling approaches are often used to determine high-risk flood regions. Two types of models are mainly used to study river flows and flood risk, as will be explored in-depth in the literature review (chapter 2). The first type of models are empirical or frequentist models. The second type are full 2D or 3D simulation models. While the first type of models are simple to use and easy to understand, they provide limited insight into the flood problems. The second type of models use advanced hydraulics to give very accurate results. These models are problematic because they are expensive to use and develop, and they require a lot of data. These requirements make them difficult to use for developing countries and large scale river systems. Most of these countries are lacking knowledge, data and funds to use these complicated models. Therefore both model types are unsuitable to do a quick flood scenario analysis.

In between these models, there exist another group of models. These models use more simplifications than the detailed simulation models, but more data than the simple empirical models. These can add much value for policy analysts and decision-makers because of three reasons. These models can be used for rapid assessment of flood flows (and indirectly risk) in data-sparse regions. Policymakers can compare high-level policy options under different scenarios. These scenarios can also include effects as a result of climate change. This model type is especially suitable for developing countries with large river systems. Figure 1.1 shows an overview of the different model types.

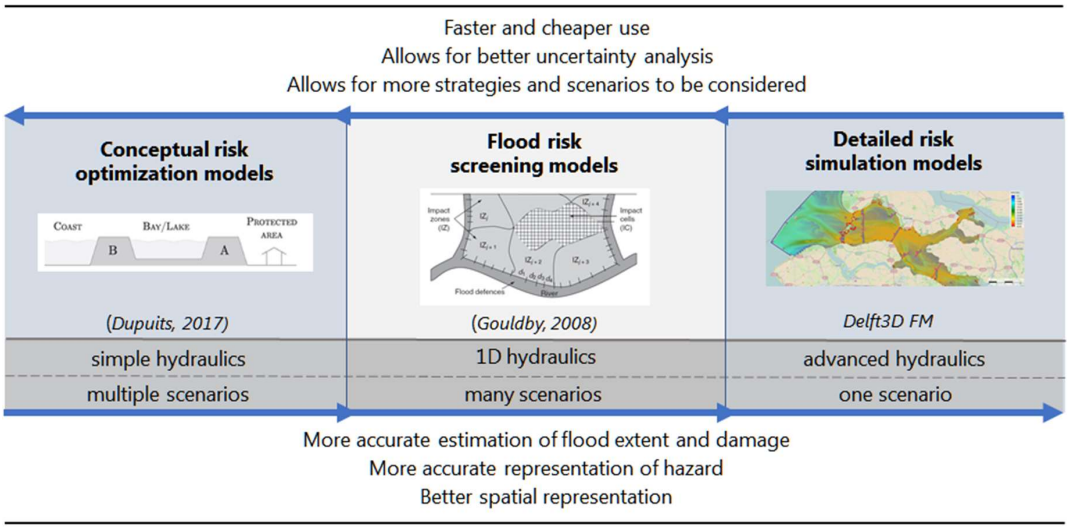


Figure 1.1: Different flood model types (van Berchum, 2019). The figure focuses on coastal flooding but holds for fluvial flooding as well. On the left, there are the simple conceptual models and simple hydrological models. These methods include the (unit) hydrograph methods and other statistical methods. Although simple to use, they do not give the dynamics in a large watershed. On the right, there are advanced hydraulic simulations. These hydrodynamic simulations include full 2D and 3D simulations of rivers in for example SOBEK/D-HYDRO Suite or 3Di. These models are accurate but are labour intensive to develop and use and require a lot of data. These models are typically used to analyse a single scenario. The goal is to develop a model in between these two types. It should be efficient enough that it allows for easy calculation and scenario analysis but should include enough dynamics to get a high-level system understanding.

1.2. Research objectives

From the previous section, it is clear that there is value in efficient methods to analyse flood flows and risks. This research focuses on analysing flood flows. Determining flood flows in river systems is called *flood routing*. Analysing flood risk is outside the scope of this research, although flood routing can be a first step to do so. The first research objective of this thesis is defined as:

> I) To develop a flood routing model that is efficient, uses less information, is easy to understand and is useful for policy analysts and decision-makers.

In this thesis, we choose to use complex network theory together with flood routing methods to develop a network model. The advantage of a network model is that it composed of simple elements and therefore, easy to understand. This idea leads to the second research objective:

> II) To represent river systems using complex network theory for understanding flood routing in riverine systems.

1.3. Thesis outline

This thesis is divided into four parts: I) introduction to the research problem, II) model development, III) a case study and IV synthesis. An overview of this thesis' outline is given in Table 1.1.

Table 1.1: Thesis outline

Part		Chapter
I	Introduction to the research problem	1. Introduction 2. Literature review 3. Research design
II	Model development	4. Modelling 5. Verification
III	Case study	6. Data 7. Results
IV	Synthesis	8. Discussion 9. Conclusion

Part I introduces the research problem. Chapter 2 presents a literature review of both flood modelling and complex systems. The purpose is to find out what models exist and if a complex network approach has been tried to model floods. The literature review leads to a knowledge gap and the research questions, presented in Chapter 3. This chapter also includes the research strategy on how to answer these questions.

Part II presents the model development. Chapter 4 starts with model conceptualisation: network theory and flood routing are combined in a single model. This chapter ends with model implementation. In this chapter, some practical issues are solved before being able to use and run this model. The developed model is verified and suitability is discussed in Chapter 5.

The developed model is applied in a case study in part III. The developed model makes use of open-source data sets. This data needs to be processed before it can be effectively used in the model. Chapter 6 describes the requirements of the data and data processing. This chapter also explains which assumptions were made because of data availability issues. Chapter 7 shows the results of the model applied to the case in the Ganges Brahmaputra basin.

Part IV ends this thesis by discussing the obtained results and model suitability. It presents recommendations for further research and the conclusions of this thesis.

2

Literature Review

Understanding, modelling and, predicting floods is a broad field of study with many different modelling approaches. Before explaining different types of models, the definition of flood risk is given. This definition provides context for the modelling techniques.

2.1. Flood risk analysis

According to the European Commission, flood risk is defined as follows:

> *"Flood risk means the combination of the probability of a flood event and of the potential adverse consequences for human health, the environment, cultural heritage and economic activity associated with a flood event."* (European Commission, 2007)

In the context of flood risk management, flood risk is often considered as an expected value: it is the probability of a flood event multiplied by its consequences. Flood event probability is generally expressed as the probability per unit time, for example per year. Consequences can be defined in multiple ways and can consist of: material damages, ecological damages, injuries and fatalities. For example: in the Netherlands, three types of risk consequences are considered: economic risk, group risk and individual risk (Expertisenetwerk Waterveiligheid, 2016).

Two factors that are involved in flood event probabilities: hydraulic loads and the corresponding failure rates. The hydraulic load is a collective term for the different forces on the flood defences. Given certain hydraulic loads, flood defences can fail with a certain probability. This probability is the failure rate. Together the hydraulic load and failure rate give the flood event probability, also called the flood hazard. Dependent on the inland situation, this will result in different flood scenarios with corresponding consequences. The set of all (n) flood event probabilities with (monetised) consequences will give an expected value: $E = \sum_{s_i}^{n} p_i d_i$, in which s_i is a scenario with corresponding probability p_i and consequence or damage d_i.

Another frequently used risk metric is the risk curve. A risk curve shows the probability of exceedance of a certain magnitude of consequence. In flood risk, this is often presented as an FN curve, which displays the probability of exceedance of N fatalities. A frequently used method to determine next actions is the use of a cost-benefit analysis. In such an analysis, an optimum total cost is found. The total cost includes mitigation cost and flood risk and is a function of a safety measure, for example, dyke height.

2.2. Flood models

There are two types of models: frequentist models (or statistical models) and simulation models.[1]

- Statistical models, frequency analysis of flood and flow data
  - QdF
  - Hydrograph methods
- Simulation-based models
  - Hydraulic routing
  - Hydrologic routing

The principles behind these methods will be explained briefly in the following subsections. This overview will give a better understanding of the current methods used, with their advantages and disadvantages.

2.2.1. Frequency analysis

In the frequency analysis, flood and flow data is linked to parameters in a distribution. A statistic such as the maximum flood record or the peaks-over-threshold data is selected for analysis. Fitting the parameters to a flood frequency distribution will give an estimation of flood risk (Cunderlik & Ouarda, 2006). In these methods, only observed data is used. No catchment geography or river network structure is included in these types of analysis. The major downside of this method is that it only relies on historical data. It cannot tell something about flood risk under changing conditions, for example, changing weather patterns as a result of climate change.

QdF (quantity, duration, probability) models are more advanced flood frequency methods. These methods summarize the frequency of flood events for different probabilities and durations by one compound formula (Cunderlik & Ouarda, 2006). In such a formula, the flood quantiles, $Q(d,F)$, are a function of probability F and duration d. In this function, several catchment or basin specific parameters are obtained by fitting to real-world discharge data. Another option is to fit these parameters to simulated data (see Subsection 2.2.2). Although this method does not use any geographical data as an input, specific catchment parameters can be obtained by fitting to real-world discharge data. However, no clear relation between catchment structure and these parameters can be found, and the primary frequency analysis downsides still apply.

A hydrograph displays the relation between rainfall and discharge in a river. Both rainfall and discharge data are used to construct observed flood hydrographs. These observed hydrographs can be used to analyse flood risk directly, or these data can be used to construct unit-hydrographs. The unit-hydrograph shows the flow response of one unit rainfall, e.g. 1 cm of rain uniformly distributed over a catchment. The unit hydrograph is related to soil and other area catchment specific parameters and topology. Methods exist to determine the unit hydrograph based on the catchment shape. A superposition of hydrographs can be used to predict flow based on arbitrary rainfall events (Dooge, 1959). Because hydrographs are based on observed data, they do not give the relation between the form of the catchment specific parameters and topology. This makes them less suitable for large regions without rain and flow data.

2.2.2. Continuous flow simulation

Nowadays, cheap computer power has made continuous flow simulation a viable option for the estimation of flood flows. These simulations make use of flood routing methods to estimate peak floods. Flood routing is a method to predict the change in the shape of a hydrograph as water moves through a river. There are two different technique classes for flood routing, hydraulic (distributed) routing or hydrologic (lumped) routing. Hydraulic routing is based on the partial differential equations for open water, called the Saint-Venant equations (Choudhury Parthasarathi, Shrivastava Rakesh Kumar, & Narulkar Sandeep M., 2002). Although more accurate, their practical application is restricted due to the high demand for computation technology and the reliability on quality and quantity of data (Karahan, 2012). Opposed to hydraulic routing methods, hydrologic routing methods are based on storage-continuity relationships and uses computational simplifications by reducing the number of routing parameters. The parameters are chosen so they can reproduce an outflow hydrograph. These simplified approaches

[1] In relation with Figure 1.1: the frequentist methods are the simple methods on the left in the figure. Hydraulic routing methods applied in continuous flow simulations are considered the methods on the right.

can solve many complicated problems efficiently (Choudhury Parthasarathi, 2007) and with reasonable accuracy (Karahan, 2012).

The Muskingum method is one such example method and is one of the most popular routing methods. This model simplifies a river reach by solving directly for the outflow discharges as a function of the inflow discharges. All geomorphological characteristics and hydraulic properties are lumped into a number of model parameters. There exist various numbers of Muskingum methods, each with a different number of parameters and their own strengths and weaknesses (Todini, 2007).

Some advantages of simulation models are: improved flood estimation in ungauged catchments, the possibility to estimate the effect of changes in urbanisation and land-use and to develop flood scenarios from future climate scenarios (Samuels, 2000). Many of these models exist; an overview study compares and categorises over seventy different watershed hydrology models (Singh & Woolhiser, 2002). The downside of these models is that they, especially hydraulic models, need a lot of information and parameters about the watershed or catchment. These models are therefore time-consuming to develop, implement and use (Choudhury Parthasarathi et al., 2002). Another downside is that these models do not directly give a measure for flood risk. Other statistical methods used on top of these simulations are necessary to give flood risk estimations.

2.3. Complex systems

One of the solutions might be to look at riverine floods from a complex system perspective. However, to fully understand a complex system, it is essential to understand the context in which it is developed. Although there is no official definition, a complex system is often defined as “a large group of relatively simple components with no central control and where organization and emergent non-trivial behaviour are exhibited” (Koç, 2015). There are three main concepts in this definition: the complex system has many components that have relatively simple individual behaviour, these components can interact with each other on different levels, and as a result, non-trivial system behaviour emerges.

These typical emergent features include cascading failures and phase transitions (Koç, 2015). Cascading failures are successive failures in a system. Cascades usually start with a single failure and spread to the system due to interactions between the components in the system. For example, an overload in a power grid may result in failure of a particular node. Because this node is deactivated, other nodes in the network must compensate, energy is rerouted. The nodes that compensate can, in turn, also overload and fail. This repetitive process can cause mayor blackouts in wrongly designed power grids. Cascades can be observed in many systems including, computer networks (Wang, Jiang, & Qian, 2014), supply chain networks (Tang, Jing, He, & Stanley, 2016), financial systems (Haldane & May, 2011), water distribution systems (Shuang, Zhang, & Yuan, 2014) and thus power grids (Dobson, Carreras, Lynch, & Newman, 2007; Hines, Balasubramaniam, & Sanchez, 2009).

There are several ways to model and analyse complex systems and their emergent features. These methods include, differential equations, cellular automata, agent-based models and complex networks. Differential equations are particularly suitable to simulate dynamic systems behaviours of identical elements. Cellular automata and agent-based models are suitable for systems that consist of elements that mainly interact locally. All these three methods face difficulties “when systems are composed of many non-identical elements that have diverse and multi-level interactions, local and non-local” (Rosas-Casals, n.d.). By considering complex systems as networks, these emergent behaviours of large, diverse systems can still be analysed.

2.3.1. Complex network theory and analysis

Networks are omnipresent. Humans form a network of social relationships and are also the result of a network of biochemical reactions. These are networks defined in an abstract space. Networks can also have a physical structure, such as electric power grids, telecommunications networks, the Internet, highways or subway systems (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006).

The mathematical study of networks started with the “Bridges of Königsberg” problem in the eighteenth century. Euler introduced graph theory with his solution, which has become a large field of research. A graph is a mathematical structure modelling relations between components in a system. Graphs consist of nodes that are connected by edges (Newman, 2003). Translating a real-world complex system into a graph makes such a graph a complex network.

Complex networks have been studied in social sciences. Patterns of connection between people are

considered essential to understand the functioning of human society. One can reconstruct a network in which nodes represent individuals and edges the interactions between them. Typical social network studies try to address issues of centrality (which individuals have the most influence) and connectivity (how are individuals connected). Nowadays, research has shifted from the analysis of small networks towards the analysis of large-scale scale properties of graphs.

There are several reasons why complex network analysis is an upcoming field of research. First of all, the availability of large datasets makes it possible to construct networks. Second, the availability of powerful computers has made it feasible to analyse their structure. Third, many of the parts of a complex system are understood very well, but their behaviour as a whole is not yet fully understood. Still, networks are challenging to understand; their structure can be complex, evolve, and connections or nodes can be diverse (Strogatz, 2001). Recent research has made a lot of progress in studying: food-webs (Cohen, Briand, & Newman, 1990; Williams & Martinez, 2000), cellular and metabolic networks (Jeong, Tombor, Albert, Oltvai, & Barabási, 2000; Kohn, 1999), electrical power grids (Koç, 2015), the World-Wide Web (Broder et al., 2000) and much more (Strogatz, 2001).

Traditionally networks were analysed as random networks, but recent research has shown that real-world networks are not random (Newman, 2003). Other methods and generalisations were developed to study complex networks; these include generalized random graphs, small-world networks, scale-free networks and evolving networks (Albert & Barabási, 2002).

2.3.2. Analysing cascades

Complex network analysis is frequently used to analyse cascades. As described in the previous section, cascades can be observed in many different systems. Examples of cascades are overload cascades, information cascades and avalanche cascades. Overload cascades can be modelled by removing edges or nodes in a network. After removing these nodes, loads within the network will redistribute. The resulting network can have other overloaded edges or nodes in it. These overloaded edges or nodes are then also removed, and these steps are repeated until a stable system is reached. Thus removal of a single node can trigger a cascade of overload failures (Motter & Lai, 2002). These types of cascades happen and are studied within power grids (Koç, 2015), banking systems (Haldane & May, 2011) and the internet. The effect of attack strategies on different types of networks topologies can be generalised. It can be concluded that the effect of cascades on complex networks is very large due to their non-random structure. This poses security threats in systems like the internet and power grids (Motter & Lai, 2002). Complex network analysis is, therefore, a valuable tool to find weak spots within these systems in order to make them more resilient to attacks. A common problem in analysing these cascades is that the capacity is not always known. A common assumption to solve this, for human-made networks, is to include a tolerance parameter to relate the capacity to the initial load (Motter & Lai, 2002).

In social networks, information cascades can occur. An information cascade occurs “when it is optimal for an individual, having observed the action of those ahead of him, to follow the behaviour of the preceding individual without regard to his own information” (Bikhchandani, Hirshleifer, & Welch, 1992). This means that, depending on the decision of the first few persons, the other individuals follow, regardless of their own information. The problem with these cascades is that they prevent aggregation of knowledge of multiple individuals. Ideally if information is aggregated, later individuals should converge to the correct action. However, once a cascade has started, actions do not convey private information or knowledge; thus, one action does not improve later decisions (Bikhchandani et al., 1992).

Avalanche cascades and dynamics are studied in sandpile dynamics. An example is the Bak-Tang-Wiesenfeld sandpile model (Bak, Tang, & Wiesenfeld, 1988). This model can be applied to scale-free networks to obtain the avalanche size and duration distribution of sand toppling (Lee, Goh, Kahng, & Kim, 2004). Every time a sand grain is added to a node in the network. When the height of the node exceeds a certain level, it becomes unstable, and the grains topple to adjacent nodes. If the adjacent nodes also become unstable, an avalanche can occur. Lee et al. (2004) show a different approach to studying cascading effects. This approach does not alter the network structure, and flows of sand can be studied. These examples above show that cascades can be defined in multiple ways.

2.3.3. State of the art

Some approaches to analysing river floods come close to complex network analysis. An overview of articles found is given in Table 2.1. Most articles use a kind of network representation in their method.

Table 2.1: State of the art of flood modelling using network representation.

Authors	Network representation	Method	Flow modelling
Naden (1992)	yes	Width function, unit hydrograph	no
Harley (1971)	yes	Kinematic wave	yes
Lehner and Grill (2013)	yes	Plug-flow, diffusion	no
Boyd et al. (1979)	yes	Connected storage elements, hydrographs	no
Bérod et al. (1995)	no	Hydrological cascades	yes
Mamede et al. (2012)	yes	Overspill avalanching	no

Bérod, Singh, Devred, and Musy (1995) briefly touch upon the idea that networks can be used, but there is no analysis.

The methods used within these networks vary greatly. Boyd, Pilgrim, and Cordery (1979) and Bérod et al. (1995) use a form of connected storage elements or hydrological cascades (not to be confused with cascades in the sense of cascading failures). In a hydrological cascade, flow is modelled as a linear series of storage reservoirs with outflow equal to the inflow of the next reservoir. Hydrological cascades have been used extensively in the past also to predict precipitation runoff.

In three cases, the goal of the network model itself is not used to analyse the resulting flood flows. Naden (1992) and Boyd et al. (1979) use their methods to improve (unit) hydrographs. With their approaches, spatial variability information is used to improve existing methods. Adjustments to the method of Boyd et al. (1979) however, could make flood modelling possible. Lehner and Grill (2013), propose an advanced vector-based routing method. However, it is not used to model discharge routing. Discharge is an input variable and is based on external runoff calculations. So this method is not suitable for flow modelling.

The model that comes closest to a network flow model is that of Harley (1971). He proposes a network representation of rivers and applies a kinematic wave model to determine flows within the network. Although represented as a network, most work is done on developing the kinematic model. They can model river flows quite accurately. The downside of his approach is that the kinematic wave model still needs a lot of data.

The model of Mamede, Araújo, Schneider, de Araújo, and Herrmann (2012) is a network flow routing model of reservoirs in Brazil. They show that the network is a scale-free network. Their model takes into account rainfall, evaporation an in and outflow from and to connected reservoirs. The result is flow data and avalanches of flow: they define an avalanche as a cascade of overflows of reservoirs. A single reservoir overflow can then cause trigger such a cascade. Unfortunately, their method is not suitable for routing of water in riverine systems because transmission losses through river transport are neglected.

3

Research design

3.1. Knowledge gap

In the previous chapter, several method types are discussed. In the empirical or frequency analysis methods, risk is based on observed flood flows, not on catchment geography or topological structure. In these methods, no information about the topology or river structure is used or understood. The same holds for catchment design methods such as QdF. QdF, on the other hand, does include catchment specific parameters that are fitted based on observed flows. In the case of flow simulations, advanced models are used that take a lot of time to develop and use. This is especially true for full three-dimensional simulations of river basins. Even though these simulations make use of river structure and geographical data, no real understanding between flood risk and river topology is developed. On the other hand, there are a few methods that try to incorporate topological data in a river network, as was discussed in Subsection 2.3.3. However, the goal of these models is neither related to flood flows nor cascades, or river structure is used in extensions of more traditional methods. Therefore the first knowledge gap identified is:

> I) There is a lack of knowledge between river structure and cascading flood effects.

Based on the articles and literature overview, the methods can be scored on several properties. The different models and their properties are shown in Table 3.1. Four properties are defined: fitting to observed data, use of topological data, parameter complexity and development effort. Fitting to observed data is primarily done in frequentist methods in which discharge levels are analysed statistically. Those methods do not use any topological data such as network structure. They are simple but provide limited insight. Fitting to observed data alone is also problematic in the event of a flood. If no flood has ever occurred, it will not happen in statistical methods. Moreover, if no topological data is used the consequences of a flood event are unknown. Flood risk maps are necessary for policy benefit.

Flow simulation models, on the other hand, use a lot of topological data. So much in fact that their parameter complexity is very high. These methods require many parameters to be known in advance (Karahan, 2012). Even if some parameters are available, others might not. This makes development expensive and time-consuming. These methods are therefore, less suitable for policy screening. The problem with the previously described methods is that either only empirical data is used or the relation between river network structure and risk is not well understood. Therefore a second gap, in between these models, is identified:

> II) There is a lack of models that are efficient, easy to understand, use topological data and have the purpose of risk screening.

To the best of our knowledge, no complex network analysis approach to analyse cascading river floods has been developed. Complex network analysis is a method to describe certain problems as graphical networks. As described in the previous chapter, this method has been used in various research fields to get a better understanding of the networks. In this research, we propose the use of complex network analysis to get a better understanding between river network structure and cascading floods. Ideally, a

method is developed that combines both observed data with topological data and has less development effort compared to continuous flow simulation methods. In the next section, the research questions and research design are proposed to develop such a method.

Table 3.1: Qualitative analysis of flood estimation and risk methods, showing data usage and development effort. Mixed methods are methods considered in Subsection 2.3.3. A very limited number of studies were found for the mixed method type.

Method	Fitting observed data	Topological data	Parameter complexity	Development effort
Empirical / frequentist	++	–	–	-
Catchment design: QdF	++	-	–	+
Continuous flow simulation	+	++	++	+++
Mixed methods	+	+-	+-	+

3.2. Research question

The literature review in the next chapter reveals that there are two knowledge gaps: There is a lack of knowledge between river structure and cascading flood effects. Moreover, there is a need for efficient and easy to understand models that have the purpose of risk screening and are useful for policy analysts and decision-makers. This gap leads to the following research question and sub-questions:

How to use complex network theory to analyse the cascading effects in riverine systems?

1. How to model a river as a complex network?
2. How to define cascading effects in riverine systems?
3. How to use a complex network model to analyse the cascading effects of floods?
4. To what extent can the river network model be used with data from the Ganges-Brahmaputra river basin?

The sub-questions divide this research into two parts: a conceptual part and a case study. In the first part, cascading effects are defined, and a theoretical model is developed. To answer the last two sub-questions the model is implemented and used in a case study. The main goal of this case study is to show its potential value and determine its suitability. For the initial research approach, see Appendix A.

4

Modelling

In this chapter, a model is developed that satisfies both research objectives. Model development is done in five steps. This chapter starts with the definition of cascades in riverine systems. Then the existing Muskingum flood routing method is explained in the second section. With this routing method, a network routing method is developed in section three. The routing method only tells us how water flows within a network. In order to use this method to model overflow or flood risk, the model is extended with inflows and capacities in section four. Model implementation and associated practical issues are described in the last section of this chapter. Blue blocks within this chapter summarise answers to the research sub-questions.

4.1. Cascades in riverine systems

Cascades from a complex network point of view are often seen as successive overload failures. This holds for different kinds of networks such as the Internet and power grids. The assumption for cascades to happen in such networks is that a failure, or a removal of nodes, leads to a new redistribution of loads. Cascades occur when the load distribution is highly heterogeneous, and the removed nodes are among those with higher loads. Otherwise, cascades are not expected (Motter & Lai, 2002).

No such thing as a river cascade exists in literature. Cascades in riverine systems have to be defined in order to continue. Riverine systems behave differently. A river network mostly has a tree-like structure, which is less heterogeneous and consists mostly of confluences. Removal of a node does not lead to a redistribution of flows within the network: water just flows downwards. On the other hand, once a river is flooded, water leaves the system and the downstream risk decreases. When flows levels are normalised, this water can flow back in the river. This is an essential difference from cascading failures in a traditional sense: a cascade results in an amplifying effect. In a riverine system, the flood leads to an attenuating effect. The cascading effects in a riverine system are defined as floods that occur as a result of water waves that originate from upstream sources.

Thus cascades are defined as propagation of (water) waves through a network. There are many ways to calculate the propagation of water through river sections. Several options and the difference between hydraulic and hydrologic models have been discussed in 2.2. One of the most popular and accessible methods is the Muskingum flood routing model. Because of its simplicity, this method will be used as one of the building blocks of the network. Another advantage is that the Muskingum parameters are linked to physical properties of a river reach. With these concepts, it is possible to summarise the answer to research sub-question 2:

How to define cascading effects in riverine systems?

The cascading effects in a riverine system are defined as floods that occur as a result of water waves that originate from upstream sources.

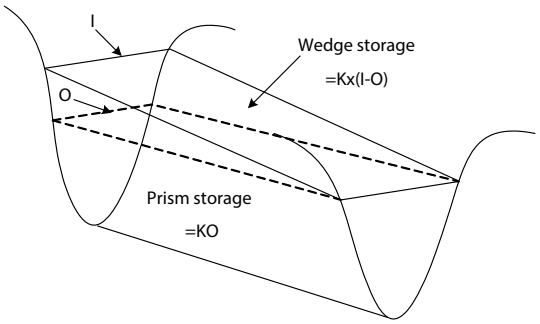


Figure 4.1: Prism and wedge storages in a channel reach defined as a function of inflow and outflow. These storages are used in the Muskingum flood-routing model (Yoo Chulsang, Lee Jinwook, & Lee Myungseob, 2017).

4.2. Muskingum flood routing

The Muskingum routing model simplifies a river reach by solving directly for the outflow discharges as a function of the inflow discharges. All geomorphological characteristics and hydraulic properties are lumped into several model parameters. In the most basic form, the model has two parameters, x and k and is based on two equations. The first equation is the continuity equation or conservation of mass equation. This equation states that the change in storage S is equal to the difference in inflow I and outflow O:

$$\frac{dS}{dt} = I - O. \tag{4.1}$$

In the Muskingum flood routing model, the finite approximation for dS/dt is used

$$\frac{dS}{dt} \approx \frac{S(t+\Delta t) - S(t)}{\Delta t}. \tag{4.2a}$$

The inflow and outflow are determined by a finite approximation as well, i.e.

$$I \approx \frac{I(t+\Delta t) + I(t)}{2}, \qquad O \approx \frac{O(t+\Delta t) + O(t)}{2}. \tag{4.2b}$$

By substituting equations 4.2b into equation 4.2a the difference of storage S between t and Δt is approximated by

$$S(t+\Delta t) - S(t) = \frac{I(t+\Delta t) + I(t)}{2}\Delta t - \frac{O(t+\Delta t) + O(t)}{2}\Delta t. \tag{4.3}$$

The Muskingum routing method is based on the assumption that water in a reach consists of a wedge and prisms storage component. This is shown in Figure 4.1 and is expressed in equation 4.4:

$$S = kO + kx(I - O) = k\left(xI + (1-x)O\right). \tag{4.4}$$

Thus the volume stored in the reach is expressed as a linear combination of the inflow discharge and outflow discharge (Todini, 2007). In this equation, x is a dimensionless parameter and k is a time parameter. An approximation for k is the travel time of water through the reach.

A value of $x = 0$ gives maximum attenuation, or maximum storage effect, and $x = 0.5$ minimum attenuation. For $x = 0.5$ the shape of the outflow becomes identical to the inflow hydrograph, this is the case of a linear channel without storage effect. For smaller values of x, the peak outflow becomes higher than the peak inflow, which is not physically reasonable (Yoo Chulsang, Lee Jinwook, & Lee Myungseob, 2017). Therefore the value of x has the following range: $0 \leq x \leq 0.5$. There are several ways to estimate both parameters x and k. k can easily be determined by dividing the length of the reach by the average flow velocity. x is often more challenging to estimate if inflow and outflow data is sparse x is assumed to have a value of $x = 0.2$ on average (Mays, 2010). When inflow and outflow hydrographs are available for a reach, there are several techniques to estimate both x and k (Karahan, 2012).

4.2.1. Single reach routing equations

In this subsection, the so-called routing equations for a single reach are derived. The routing equations specify the outflow of a reach based on inflow and its parameters x and k. Once we know the routing equation for a single reach, these can be combined in a network structure. By substituting equation 4.4 in 4.1 we derive the following:

$$\frac{d\,(kxI)}{dt} + \frac{d\,(k\,(1-x))\,O}{dt} = I - O. \tag{4.5a}$$

The Muskingum method assumes that k and x are constant in time:

$$kx\frac{dI}{dt} + k\,(1-x)\,\frac{dO}{dt} = I - O. \tag{4.5b}$$

This equation can be solved by using a finite difference approach, introduced in equation 4.2a:

$$\frac{dI}{dt} \approx \frac{I(t+\Delta t) - I(t)}{\Delta t}, \qquad \frac{dO}{dt} \approx \frac{O(t+\Delta t) - O(t)}{\Delta t}. \tag{4.5c}$$

Substitution of equations 4.2b and 4.5c in equation 4.5b leads to:

$$kx\frac{I(t+\Delta t) - I(t)}{\Delta t} + k\,(1-x)\,\frac{O(t+\Delta t) - O(t)}{\Delta t} = \frac{I(t+\Delta t) + I(t)}{2} - \frac{O(t+\Delta t) + O(t)}{2}.$$

Multiplied by $2\Delta t$ gives:

$$\begin{aligned} &2kx\,(I(t+\Delta t) - I(t)) + 2k\,(1-x)\,(O(t+\Delta t) - O(t)) = \\ &\Delta t\,(I(t+\Delta t) + I(t)) - \Delta t\,(O(t+\Delta t) + O(t))\,. \end{aligned}$$

Factoring out $O(t+\Delta t)$ gives us:

$$\begin{aligned} &(2k\,(1-x) + \Delta t)\,O(t+\Delta t) = \\ &(-2kx + \Delta t)\,I(t+\Delta t) + (2kx + \Delta t)\,I(t) + (2k\,(1-x) - \Delta t)\,O(t). \end{aligned}$$

Isolating $O(t+\Delta t)$ leads to:

$$O(t+\Delta t) = \frac{-2kx + \Delta t}{2k\,(1-x) + \Delta t}I(t+\Delta t) + \frac{2kx + \Delta t}{2k\,(1-x) + \Delta t}I(t) + \frac{2k\,(1-x) - \Delta t}{2k\,(1-x) + \Delta t}O(t).$$

Which can be simplified to

$$O(t+\Delta t) = C_1 I(t+\Delta t) + C_2 I(t) + C_3 O(t), \tag{4.6a}$$

where the three routing coefficients C_i are given by

$$C_1 = \frac{-2kx + \Delta t}{2k\,(1-x) + \Delta t}, \quad C_2 = \frac{2kx + \Delta t}{2k\,(1-x) + \Delta t} \quad \text{and} \quad C_3 = \frac{2k\,(1-x) - \Delta t}{2k\,(1-x) + \Delta t}. \tag{4.6b}$$

C_1, C_2 and C_3 are the routing coefficients and sum op to 1. All coefficients should be positive; therefore the following constraint holds:

$$2kx \le \Delta t \le 2k(1-x) \tag{4.7}$$

When x and k are known, and Δt satisfies 4.7, equations 4.6 can be used to determine every reach outflow $O(t)$ for any arbitrary inflow $I(t)$.

4.2.2. Advantages and limitations

The Muskingum method has some advantages and limitations. It is efficient because all hydraulic effects are lumped into two parameters for a reach. This makes the calculation very easy. The method is proven and verified: it is widely used in studies and literature (Ponce, Lohani, & Scheyhing, 1996).

Because all effects are lumped into two parameters, only those two have to be estimated. All together, it is also easy to understand.

Estimating these parameters can be done with known discharges. If those are not known, it is also possible to use the Muskingum-Cunge method. The Muskingum-Cunge makes the parameters dependent on hydraulic properties of the reach. This can result in better parameter estimations. It is also possible to use a time variable Muskingum-Cunge routing scheme or a non-linear Muskingum model. In such a Muskingum-Cunge routing scheme, the parameters become dependent on discharge. In non-linear Muskingum models, the relation consists of more parameters with a non-linear relation between storage and the inflow and outflow (equation 4.4). Both methods provide more accurate representations (Bozorg-Haddad, Abdi-Dehkordi, Hamedi, Pazoki, & Loáiciga, 2019; Todini, 2007).

Because the Muskingum routing method is a lumped method, some detail is lost. It is less detailed than methods that solve the Saint-Venant equations. Two issues arise when the Muskingum routing method is used for flood modelling. The first issue is that no flood mechanism is included in the method. For example, when an overbank flooding occurs, the water on the river bank will behave differently from water in the river. This water slows down and is temporarily stored on the river banks. This effect is not directly included in the Muskingum method.

The second issue is that no downstream effects are taken into account. The outflow from a river is independent of downstream conditions. Thus the method cannot model backwater effects or reverse flows. This means that it is not suitable for languid rivers and deltas. Tidal influences are significant in deltas. These tidal influences can cause reverse flows. A tidal flood in combination with rain flood can cause major damage within deltas. These effects cannot be studied with Muskingum routing or any other hydrological routing method (Todini, 2007).

4.3. Muskingum routing in a network

In this section, a routing network model is developed. In order to model floods within a network, we have to define network parts. Any network consists of nodes and edges. Edges are defined as river reaches. A river reach is a segment of a river that has similar properties. The nodes in the network are defined as the junctions between reaches. In real data reaches are often defined as river segments between junctions. There are two types of junction: confluences and bifurcations. A confluence is a junction where two reaches meet and continue as one. A bifurcation is a junction where a single reach forks into two reaches. Nodes, edges, confluences, bifurcations and regular connections are the building blocks of a riverine system. These elements have to be defined in order to build a network.

First, we define the reach or edge. A reach j has an outflow which is defined as Q_j^{out} and an inflow defined as Q_j^{in}. This inflow is a linear combination of outflows from other reaches and an external inflow defined as Q_j^{ext}, i.e.

$$Q_j^{\text{in}} = Q_j^{\text{ext}} + \sum_{i \neq j} w_{j,i} Q_i^{\text{out}}. \tag{4.8a}$$

The external inflow can be any other source of water other than from the previous reaches. The flow from one reach to another has to be positive: $w_{j,i} \geq 0$. Moreover, a reach i can have a maximum outflow weight of 1 due to the conservation of mass:

$$\sum_i w_{j,i} = 1. \tag{4.8b}$$

Thus follows that a single weight

$$0 \geq w_{j,i} \leq 1. \tag{4.8c}$$

Figure 4.2 shows an isolated edge that represents a reach.

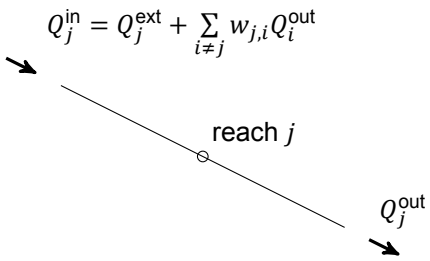


Figure 4.2: Schematic displaying a single reach represented as an edge with its inflow outflow. For weighting factors $0 \geq w_{j,i} \leq 1$ holds. These are typically zero except for the connected reaches. Note that a $\sum_j w_{j,i} = 1$ for all reaches i.

Reach j is fully characterised with the parameters x and k, respectively x_j and k_j. With the x_j and k_j, Equations 4.6b give the routing coefficients, $C_{j,1}$, $C_{j,2}$ and $C_{j,3}$ belonging to this reach j. These are used in Equation 4.6 to relate outflow to the inflow in the following way:

$$\begin{aligned} Q_j^{\text{out}}(t+\Delta t) = C_{j,1} &\left(Q_j^{\text{ext}}(t+\Delta t) + \sum_{i \neq j} w_{j,i} Q_i^{\text{out}}(t+\Delta t) \right) \\ &+ C_{j,2} \left(Q_j^{\text{ext}}(t) + \sum_{i \neq j} w_{j,i} Q_i^{\text{out}}(t) \right) \\ &+ C_{j,3} Q_j^{\text{out}}(t) \end{aligned} \tag{4.9a}$$

with

$$C_{j,1} = \frac{-2k_j x_j + \Delta t}{2k_j\left(1-x_j\right) + \Delta t}; \quad C_{j,2} = \frac{2k_j x_j + \Delta t}{2k_j\left(1-x_j\right) + \Delta t}; \quad C_{j,3} = \frac{2k_j\left(1-x_j\right) - \Delta t}{2k_j\left(1-x_j\right) + \Delta t}. \tag{4.9b}$$

In Figure 4.3 a single reach $j = 0$ is displayed with a single inflow Q_0^{ext} and outflow Q_0^{out}. This reach 0 has no previous inflow, only an external inflow. And thus for this reach Equation 4.9 becomes:

$$Q_0^{\text{out}}(t+\Delta t) = C_{0,1} Q_0^{\text{ext}}(t+\Delta t) + C_{0,2} Q_0^{\text{ext}}(t) + C_{0,3} Q_0^{\text{out}}(t). \tag{4.10}$$

This is the standard Muskingum routing equation as is used extensively in research. In the next subsections this reach representation is used to construct a full river network. The different connection types: normal connection, confluence and bifurcation are defined.

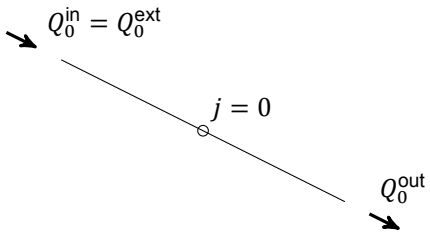


Figure 4.3: Schematic displaying a single reach represented as an edge with a single inflow and its outflow.

4.3.1. Connecting two reaches

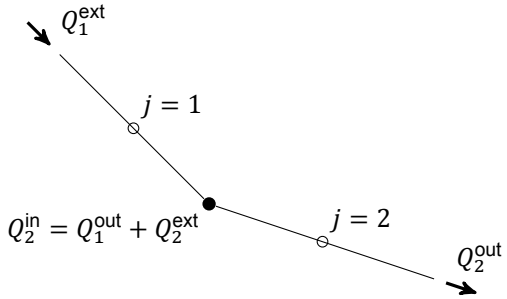


Figure 4.4: Two reaches represented as two connected edges in a network: the inflow of reach 2 is defined as the outflow of reach 1 plus an external inflow Q_2^{ext}. The external inflow can be any other source of water flowing in or out of this reach. Examples are: precipitation, ground water flow or surface water run-off.

Figure 4.4 shows a network representation of two reaches as edges 1 and 2 connected by a single node. Nodes have no functions or properties other than connecting reaches. Both reaches have their own outflow, respectively Q_1^{out} and Q_2^{out}. In this case the sum of the outflow of the previous reach Q_1^{out} and the external inflow Q_2^{ext} becomes the net inflow of reach 2, i.e.

$$Q_2^{\text{in}} = Q_1^{\text{out}} + Q_2^{\text{ext}}.$$

By substituting this into equation 4.9 it is possible to model the outflow of reach 2 based on the outflow of reach 1. The final expression becomes

$$\begin{aligned} Q_2^{\text{out}}(t+\Delta t) &= C_{2,1} Q_2^{\text{in}}(t+\Delta t) + C_{2,2} Q_2^{\text{in}}(t) + C_{2,3} Q_2^{\text{out}}(t) \\ &= C_{2,1}\left[Q_1^{\text{out}}(t+\Delta t) + Q_2^{\text{ext}}(t+\Delta t)\right] + C_{2,2}\left[Q_1^{\text{out}}(t) + Q_2^{\text{ext}}(t)\right] + C_{2,3} Q_2^{\text{out}}(t) \end{aligned} \tag{4.11}$$

In these equations, the routing coefficients $C_{2,i}$ are properties of reach 2 using x_2 and k_2.

4.3.2. Confluence of reaches

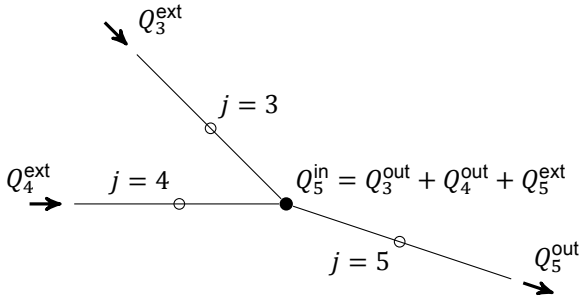


Figure 4.5: Confluence of reaches represented in a network: two reaches join and continue as a single reach. The inflow of reach 5 is the sum of the outflows of reach 3 and 4 plus an external inflow.

The same strategy can be applied to a confluence of reaches. Here reaches 3 and 4 join and flow into a single reach, number 5. The inflow of reach 5 is the sum of the outflows Q_3^{out}, Q_4^{out} and the external inflow Q_5^{ext}

$$\begin{aligned} Q_5^{\text{out}}(t+\Delta t) &= C_{5,1} Q_5^{\text{in}}(t+\Delta t) + C_{5,2} Q_5^{\text{in}}(t) + C_{5,3} Q_5^{\text{out}}(t) \\ &= C_{5,1}\left[Q_3^{\text{out}}(t+\Delta t) + Q_4^{\text{out}}(t+\Delta t) + Q_5^{\text{ext}}(t+\Delta t)\right] \\ &\quad + C_{5,2}\left[Q_3^{\text{out}}(t) + Q_4^{\text{out}}(t) + Q_5^{\text{ext}}(t)\right] + C_{5,3} Q_5^{\text{out}}(t) \end{aligned} \tag{4.12}$$

4.3.3. Bifurcation of reaches

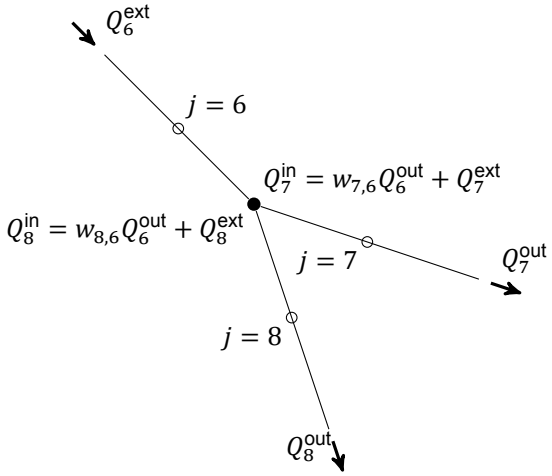


Figure 4.6: Bifurcation of reaches represented in a network: the outflow of a single reach is divided over two new reaches. Other parameters are needed to define the division of water: a fraction $w_{7,6}$ flows into reach 7 and a fraction $w_{8,6}$ in reach 8. Because of mass balance the following must hold: $w_{7,6} + w_{8,6} = 1$ and thus $w_{8,6} = 1 - w_{7,6}$.

In bifurcations the situation is different and extra parameters are necessary to define the division of water. The routing of water into each stream depends on many variables and can be dependent on the flow. A static division of water, independent of flow is assumed. In such a case the flow is independent of time and other variables. The weight factor $w_{7,6}$ is determined by the fraction of the water that flows into reach 7 from reach 6. Another fraction of the water flows into reach 8 and is denotes as the weigth factor $w_{8,6}$. In this bifurcation all the water is divided between the two reaches such that $w_{7,6} + w_{8,6} = 1$, satisfying equation 4.8b, and thus $w_{8,6} = 1 - w_{7,6}$. With the help of this fraction it is again possible to construct the routing equations for these two nodes:

$$\begin{aligned} Q_7^{out}(t+\Delta t) &= C_{7,1}Q_7^{in}(t+\Delta t) + C_{7,2}Q_7^{in}(t) + C_{7,3}Q_7^{out}(t) && \text{(4.13a)} \\ &= C_{7,1}\left[w_{7,6}Q_6^{out}(t+\Delta t) + Q_7^{ext}(t+\Delta t)\right] + C_{7,2}\left[w_{7,6}Q_6^{out}(t) + Q_7^{ext}(t)\right] + C_{7,3}Q_7^{out}(t) \\ Q_8^{out}(t+\Delta t) &= C_{8,1}Q_8^{in}(t+\Delta t) + C_{8,2}Q_8^{in}(t) + C_{8,3}Q_8^{out}(t) && \text{(4.13b)} \\ &= C_{8,1}\left[(1-w_{7,6})Q_6^{out}(t+\Delta t) + Q_8^{ext}(t+\Delta t)\right] \\ &+ C_{8,2}\left[(1-w_{7,6})Q_6^{out}(t) + Q_8^{ext}(t)\right] + C_{8,3}Q_8^{out}(t). \end{aligned}$$

4.3.4. Synthesis

In the last three sections, various essential network elements have been introduced and explained. Any river structure can be represented as a network with these three essential elements: connection, confluence and bifurcation. The adjusted Muskingum routing equations can be used to model flow through the network in a simplified way. To be able to model overflow the model needs to be extended, some relations with river surroundings have to be made. These relations are described in the next section. Nevertheless, with this information, we can answer the first research sub-question.

How to model a river as a complex network?

The river can be represented as a network by subdividing the river into reaches. Reaches form connections, confluences and bifurcations. Flow through this network is defined by applying the hydrological Muskingum routing model. A river system is fully characterised by network structure, external inflows, and the lag and attenuation parameters.

4.4. Modelling overflow

The model, as described in the previous section, can route flows through a network. However, with this information, we do not know how much water to route and when which network parts are overloaded. Two other crucial parameters need to be defined to model floods and overflows in this method. Those are the external inflow and the reach capacity.

4.4.1. External inflow

The inflow in each reach can consist of various components: groundwater flows, glacier discharge, precipitation directly falling onto the river surface or surface water flows as a result of precipitation. It is case dependent which inflows are considered in creating a model. Often inflow as a result of precipitation is modelled because it is usually a major source of water. There exist several runoff models to estimate flows resulting from rain:

- Unit hydrograph methods (Dooge, 1959)
- Peak discharge methods (Dalrymple & Benson, 1968), such as
  - Dickens' formula (Chandramohan, Jose, Purandara, & Venkatesh, 2018)
  - Ryve's formula (Ghosh, 1997)
  - Rational formula (Alexander, 1972)
- NRCS runoff curve number method (Williams J.R., Kannan N., Wang X., Santhi C., & Arnold J. G., 2012)

Most of these methods come down to the same basic formula:

$$Q = ciA^{\alpha}.$$

In which Q is the discharge, c is a runoff coefficient, i is the rainfall intensity, A the drainage area and α another scaling factor. All methods use different parameters for c and α, but the relationship between the discharge and the drainage area is proportional in all methods. Any inflow relation can be used within the developed model.

4.4.2. Capacity

There is extensive literature on flood risk for certain water levels and systems. These methods, however, are too detailed and require too much information to be applicable for a simple network model. In network analysis, a failure is often simplified to the exceedance of a capacity C: a flood occurs when the flow Q exceeds this capacity C. Determining the capacity is often difficult or not possible. In such cases, capacity is based on a tolerance parameter β and the initial load or flow Q^{initial}:

$$C = (1 + \beta)Q^{\text{initial}}.$$

Dependent on the available data, it is possible to set fixed capacities or use tolerance parameters. In the model the flow above the capacity is defined as the overflow, which is an indirect proxy for risk. So when external inflows and capacities are known, we can use the model to study the cascading effects.

4.5. Model implementation

The developed network routing model is implemented in Python. A single class has been created that handles the creation of the networks and calculation of flows through the networks. This class is called `RiverNetwork`. The package `NetworkX` is used for graph construction. There are three main steps in the model life cycle: model initialisation, setting the inflows of the reaches and calculating flow propagation through these reaches. These three steps are explained in the following three sections. The full source code can be found on Github at https://github.com/bcvanmeurs/rna.

4.5.1. Model initialisation

Model initialisation consists of five steps:

1. Graph creation,
2. Validating graph structure,
3. Node positioning,
4. Determine calculation order,
5. Calculate base load.

Graph creation

Initialisation requires an excel file to define the structure of the network. The file has two tabs: one with nodes and another with the edges. The nodes have the following parameters: `id`, `name`, `source`, `sink`, `avg_flow`, `draw_y`, `draw_x`. Both `source` and `sink` are boolean parameters telling if the corresponding node is either a source, sink or neither. `avg_flow` is the long term average flow through a node, also referred to as the base load. The draw parameters are coordinates that do not affect the model behaviour but are used for drawing networks.

The edges have the following parameters: `xpnode`, `node`, `fraction`, `x`, `k`. The parameter `pnode` denotes the starting node of the edge. `node` is the ending node of the edge. `fraction` is 1 for normal nodes and confluences, but a decimal number for bifurcations which tell what fraction is going into this edge. Fractions from the same bifurcation should sum to 1, i.e. satisfying equation 4.8b. `x` and `k` are the Muskingum parameters x and k as described in the previous section. A script validates whether all columns are available, correctly filled, and that only source nodes can contain an average flow.

Validating graph structure

The created graph must satisfy several conditions before it is accepted. These checks are performed after the graph is created and include the following:

- All nodes are connected.
- The graph contains no loops.
- The outgoing fractions on each node sum up to one (equation 4.8b).
- Every edge has an x and a k.

Node positioning

The node positioning makes sure that nodes are drawn according to the given coordinates. Also, node and edge labels for drawing are created. These labels are dynamically generated according to the Muskingum parameters and fractions.

Determine calculation order

When the graphs have been validated, a calculation order for the reaches has to be determined. This is the order in which the edges in the graph are traversed to calculate the flows. Relevant here is that all upstream flows should be calculated before downstream flows can be calculated. Determining order is done during initialisation because there is no need to determine it every single iteration. The order is determined based on a breadth-first search (BFS) algorithm. A breadth-first search algorithm starts at a root node and explores all of the neighbouring nodes at current depth before moving to nodes at the next depth level.

First, a temporary sink is added to the graph, all other sinks are connected to the temporary sink with temporary edges. Secondly, a reversed breadth-first search is performed starting from this temporary sink. The result is an edge order. This order is then reversed, and the temporary edges and node

are removed from the graph. The final order is the calculation order of the graph. An example of this procedure is given in Figure 4.7.

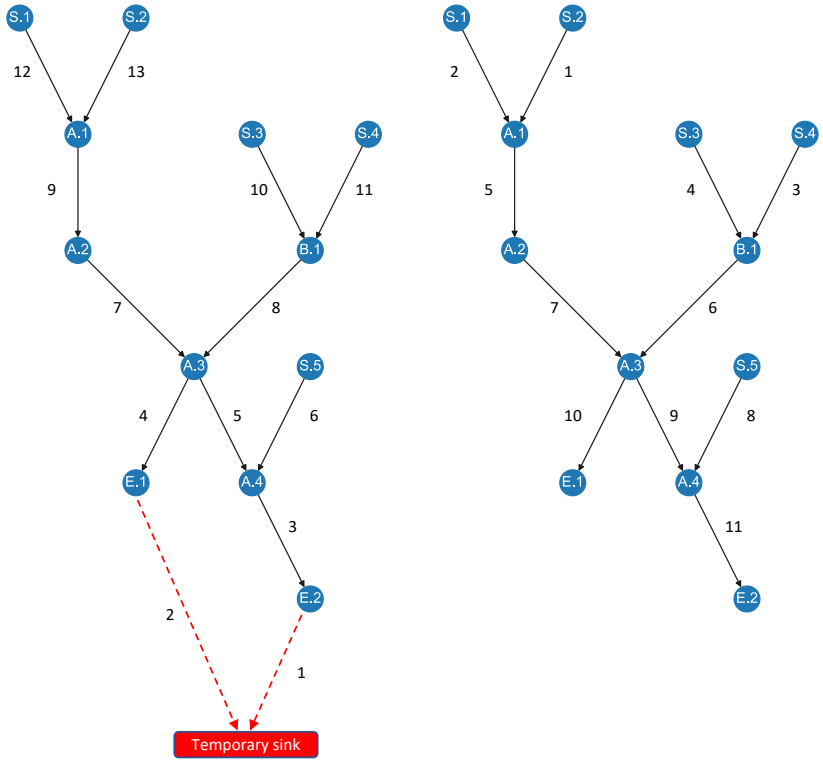


Figure 4.7: Example of the calculation order determining. In the left graph, a temporary node and edges are added in red. From this node, the reversed breadth-first search algorithm finds edges in a particular order. The order is indicated with numbers next to the edges. In the right graph, the order of these edges is reversed, and the temporary edges and node are removed. The order now corresponds to the calculation order that guarantees that all upstream parts are calculated before traversing downstream. This can be seen best at node A.3: all upstream edges are traversed before going downstream.

Calculate base load

The last step during initialisation is the calculation of the base load or average flow. The graph is traversed according to the calculation order to determine this base load. The average flow of the nodes is added to all neighbours of the nodes.

4.5.2. Setting inflows

The class RiverNetwork has three functions to set the inflows. It is possible to set a constant inflow, to select a predefined inflow from a dataset or to set a custom wave shape. These corresponding functions are:

- `set_constant_flow(node,timesteps)`
- `set_wave(node,shape_number,strength)`
- `set_shape(node,timesteps,shape)`

In all functions, the `node` is the string of the node to which the inflow will be set. `set_constant_flow()` will set a constant flow equal to the average flow rate. The number of time steps specifies how long the flow will be. `set_wave()` makes it possible to assign preset wave forms from a dataset. The location of this dataset should be added during initialisation. `shape_number` is the shape number in the dataset to add. `set_shape()` allows setting a custom wave shape, which should be an array. All inflows arrays should be of equal length (time steps), otherwise, the shortest array limits the maximum time step.

4.5.3. Calculating flow propagation

After model initialisation and setting inflows for every node, the resulting flows are calculated by calling the function `calc_flow_propagation()`. For every time step, in the predetermined calculation order, equations 4.10, 4.11, 4.12 and 4.13 are used to calculate the flow at the next time step. This procedure is repeated until all time steps are processed. The resulting flow rates can be plotted in graphs.

Time step determination

In this research, we aim to use one time step ΔT for the whole system. This is convenient when using data that contains external inflow. These datasets typically have a fixed time step. This is the time step that we ideally want to use: a smaller time step does not yield extra information but takes extra computation time, but resolution and thus information is lost when a larger time step is used. However, reaches vary in many ways. They have different lengths, widths, depths, steepness etc. These characteristics affect both the attenuation parameter x and the lag time k of a reach.

Let us restate the time step constraint for a single reach j from 4.7:

$$2k_j x_j \leq \Delta t_j \leq 2k_j(1 - x_j). \tag{4.14}$$

The general time step ΔT will have to meet this constraint for all reaches j. This will not be possible for all reaches. Therefore we choose to use the time step ΔT for all reaches j where $2k_j x_j \leq \Delta T \leq 2k_j(1-x_j)$ holds. There are two scenarios left. The first is that $\Delta T < 2k_j x_j$ for some reach j, the other is that $\Delta T > 2k_j(1 - x_j)$ for some reach j.

In the first case, $\Delta T < 2k_j x_j$, the reaches often very long. It is possible to split such a long reach j into a number n of smaller sub-reaches with equal length. For each newly created sub-reach j_m, $m \in 1, \ldots, n$, the lag time k_{j_m} then becomes k_j/n. The attenuation parameters will be considered equal for all sub-reaches m: $x_{j_m} = x_j$. The number n of newly created reaches is such that equation 4.14 is satisfied:

$$\frac{2k_j x_j}{n} = 2k_{j_m} x_{j_m} \leq \Delta T \leq 2k_{j_m}(1 - x_{j_m}) = \frac{2k_j(1 - x_j)}{n}.$$

Note that n should be integer; thus the smallest number of reaches needed is the most efficient, which is $n = \lceil 2k_j x_j / \Delta T \rceil$. Reach j in the graph will be replaced by sub-reaches $j_m, \in, \ldots, n$.

Reaches from the second case, $\Delta T > 2k_j(1-x_j)$ are often very short. These shorter reaches could be merged with other reaches to achieve a time step interval that includes ΔT. However, merging is not always trivial or not possible at all, for instance when a reach is the only reach in between two junctions. When merging is not possible, there is no possibility other than using a smaller time step for this reach. For these reaches the model uses sub time steps. This sub time step is Δt_j differs for the reaches j where $\Delta T > 2k_j(1 - x_j)$ holds. We choose the sub time step Δt_j such that constraint 4.14 holds and that $\Delta T / \Delta t_j$ is integer. This second constraint is added to use the output of the sub time step reaches as input for the main time step reaches and vice versa.

With the information presented in this chapter it is possible to answer research sub-question 3.

How to use a complex network model to analyse cascading effects of floods?

The Muskingum network routing model is implemented in Python. Combined with data that specifies the external inflows and capacity, it is possible to study the cascading effects of flood waves in a riverine system. If the cascading effects in the system exceed edge capacity, it is considered a flood.

5

Verification

This chapter consists of two sections. In the first section the implemented routing equations from the previous chapter are verified. The Muskingum routing method is a commonly used tool in hydrology. Therefore many online examples of the application the method on a single reach exist. In the second section, the network model is verified by inspecting model behaviour. No examples exist with which we can compare the network model.

5.1. Muskingum routing verification

Verification is done by comparing model output, with online examples. The same parameters and inputs are used. Three comparisons have been made. The first two examples consist of single reaches with different parameters and time steps. The third example consists of five reaches connected in a row. The full comparisons can be seen at https://rna.readthedocs.io/en/latest/model-verification.html. In this section, only the results of two examples are presented.

5.1.1. Single reach example

The first example is from Karahan (2012). Karahan uses a $x = 0.221$, a $k = 29.165$, and a Δt of 6 hours. Two effects are expected: a peak shift to the right, because of the lag time, and an attenuation effect, because of an x smaller than 0.5. Both effects can be seen in Figure 5.1. The model output is similar to that of Karahan. However, there is also a strange effect during the first couple of timesteps: the discharge decreases, which is unexpected behaviour. This behaviour can be explained by Karahans choice of $\Delta t = 6$ hours. According to equations 4.7: $2kx \leq \Delta t$, we find $12.9 \leq \Delta t$ when the used values for x and k are filled in. The chosen Δt is thus insufficient.

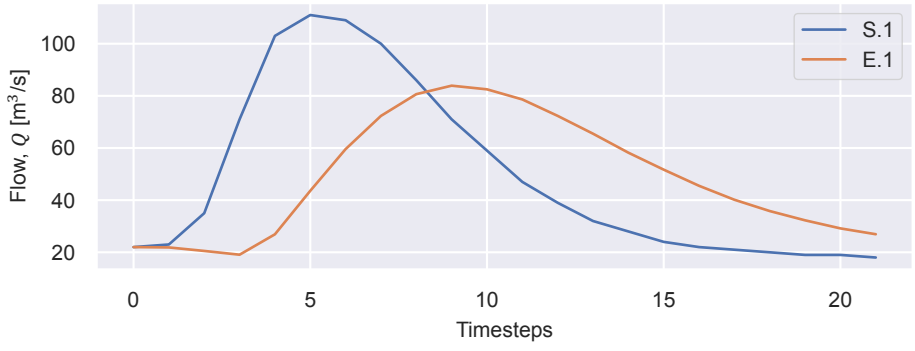


Figure 5.1: showing in and outflow discharge of the model. Both the lag effect and the attenuating can be seen. The lag can be seen because the discharge peak has shifted to the right. The attenuating effect can be seen because the maximum discharge has decreased. The output is the same to that of Karahan (2012). The full comparison can be seen at https://rna.readthedocs.io/en/latest/model-verification.html. The outflow seams to decrease in the first couple of timesteps, this is the result of a wrongly chosen Δt.

5.1.2. Multiple reaches: IJssel data

The second example is from Ciullo, de Bruijn, Kwakkel, and Klijn (2019). They apply Muskingum routing in 5 subsequent reaches. The results of the developed network model are exactly similar to the results of the code from Ciullo et al. (2019). In Figure 5.2a we can see the resulting flow for the five locations in Figure 5.2b. Their behaviour is as expected: each peak is shifted to the right and down as a result of lag and attenuation.

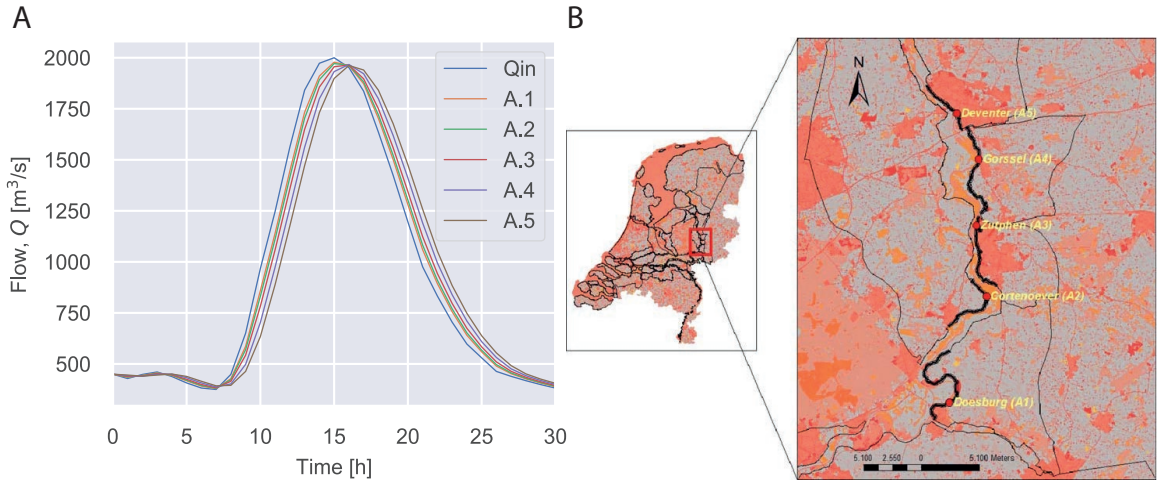


Figure 5.2: Muskingum applied to a five node network corresponding to five consecutive locations in the IJssel river. x and k for each reach are IJssel parameters from Ciullo, de Bruijn, Kwakkel, and Klijn (2019). a) graph showing the flows in the network. A shift in time and maximum between each node can be seen. b) The five points in the IJssel, showing where the flow has been calculated (Ciullo, de Bruijn, Kwakkel, & Klijn, 2019). The resulting flows from both the developed model as the reference code are the same. For all results see: https://rna.readthedocs.io/en/latest/model-verification.html.

5.2. Model behaviour

The model will be verified further with two steps: Firstly, we will take a look at the effect of the different parameters and verify their behaviour. Then we are going to take a look at the behaviour of the two other building blocks of the model: confluences and bifurcations.

5.2.1. Understanding parameters

First, we want to understand the effect of x. In figure 5.3 x will take the values 0, 0.25 and 0.5. Both k and Δt are kept constant at 1; these values ensure that equations 4.7 are satisfied. As a result, we can see that the x behaves as the attenuation parameter. All graphs have a peak at the same time step, so no shift in time has occurred. What differs is the height of each peak. For x = 0.5 no attenuation occurs, while for x = 0 maximum attenuation occurs. This behaviour is as expected from theory. Less clear is the (physical) relation between the value of x and the achieved attenuation.

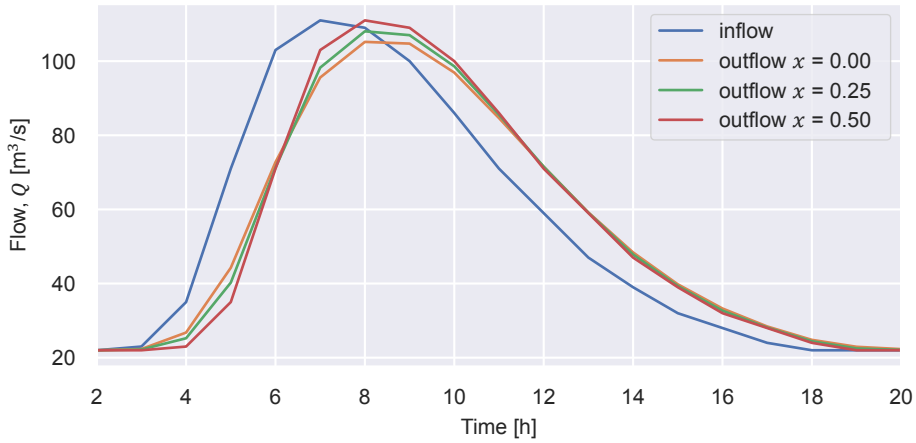


Figure 5.3: Inflow and outflow for a single river reach with $k = 1$, $\Delta t = 1$ and x varying between 0 and 0.5. x behaves as a attenuation parameter with decreasing x an increasing attenuation. There is no clear relation between the value of x and the achieved attenuation.

To understand the effect of k, we do a similar analysis. k is varied from 1 to 50, while x is fixed at 0.01 and Δt at 1. In figure 5.4 we can see that k is related to the delay or lag of the peak: the peaks shift to the right with increasing k. While the peaks shift to the right, also the attenuation increases. This makes sense because for constant x but increased reach length, attenuation is expected to increase. Meanwhile, the total volume, i.e. the area under the graph, remains the same.

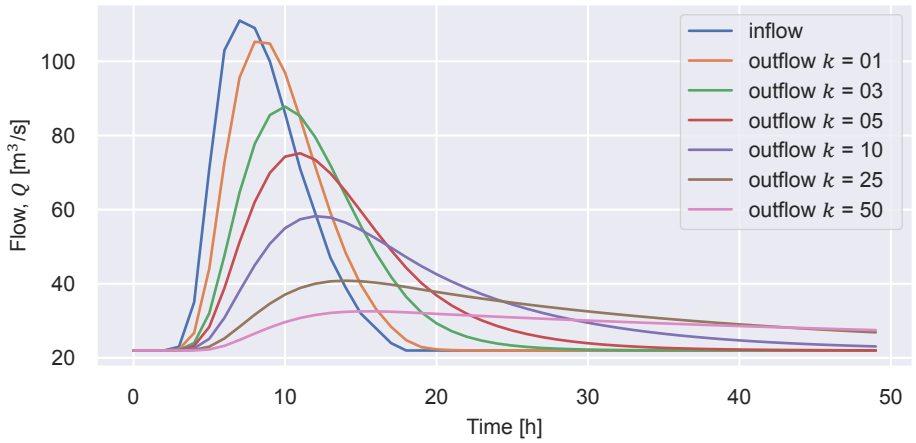


Figure 5.4: Inflow and outflow for a single river reach with $x = 0.01$, $\Delta t = 1$ and k varying between 1 and 50. It is clear that k behaves as a delay or lag parameter. With increasing k attenuation also increases, which is expected, because the longer the reach the larger the attenuation.

5.2.2. Confluences and bifurcations

In this section, we are looking at the behaviour of the main building blocks of the network model: confluences and bifurcations. In figure 5.5 an example behaviour of a small network such as shown in figure 4.5. The network consists of two reaches, 3 and 4 that join and continue in reach 5. The inflows of reaches 3 and 4 are omitted from the graph for clarity; only their outflows are plotted. These outflows are summed to form the inflow for the third reach: $Q_5^{\text{in}} = Q_3^{\text{out}} + Q_4^{\text{out}}$, no external inflows are present. The resulting intermediate inflow is plotted as the striped green line in the graph. Muskingum routing is then applied, which results in the red outflow. In this outflow, we see the attenuating and delaying effect. The two peaks are dampened and almost merged into one.

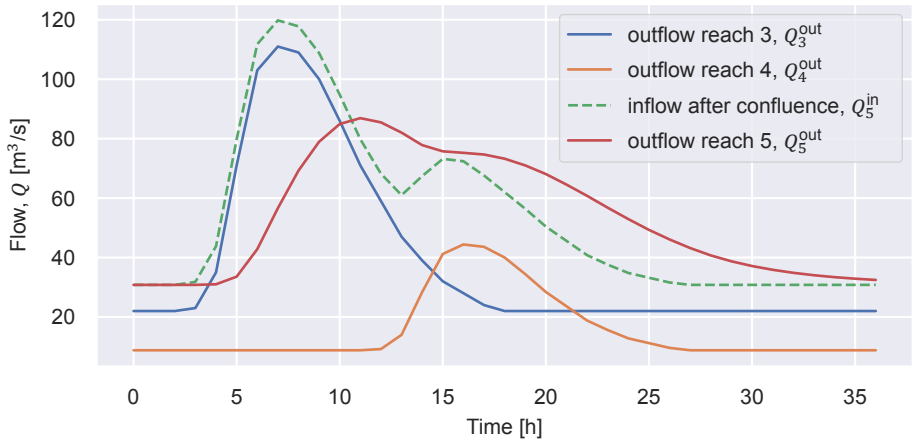


Figure 5.5: showing the outflow of two reaches that come together in a single reach as described in Figure 4.5. The two inflows are summed together to become the new inflow for the following reach, (shown in green). This inflow flows into the next reach and the resulting output is shown in red. $x = 0.1$, $k = 5$ and $\Delta t = 1$.

Example behaviour of a small network such as in figure 4.6 is shown in figure 5.6. This network consists of a single reach that divides into two reaches. The inflow of the first reach is omitted from the graph, and only its outflow is plotted. This outflow is first divided according to the fractional factor $w_{7,6} = 0.4$. The resulting (striped) intermediate flows are shown in the colours orange and green. These flows are scaled graphs compared to the outflow before bifurcation. These flows are then routed through their own reach, with different values for x and k. The resulting flows are shown in red and purple.

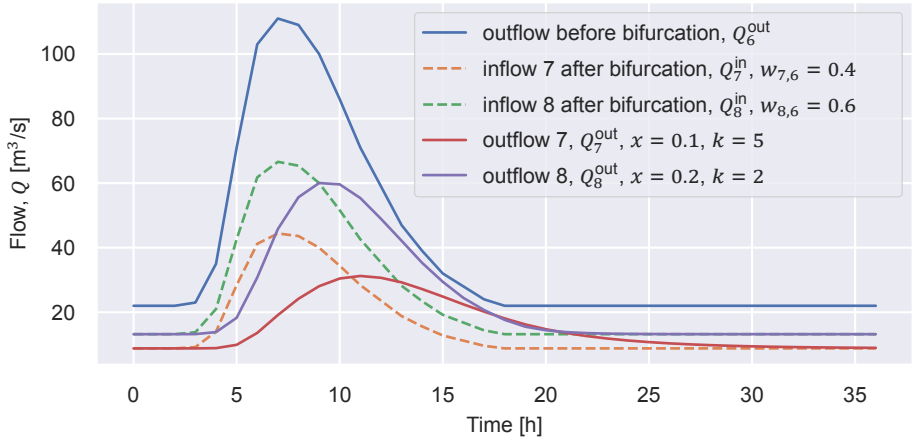


Figure 5.6: showing the outflow of a reach before bifurcation, the intermediate inflows and the outflows after bifurcation. The schematic of a bifurcation is shown in Figure 4.6. The green and orange intermediate flows are scaled version of the the blue line. The outflows are then the routed versions of these inflows. The red flow has a larger k and is therefore shifted further in time.

5.3. Synthesis

In this chapter two verification steps have been presented. Firstly the routing equations are tested and verified against three examples. These equations are implemented correctly within the model. Secondly, the basic network parts, confluences and bifurcations, are successfully tested to see if their behaviour is as defined. Also, the effect of the parameters x and y is checked. All together, we can conclude that the implemented model behaves as was designed in chapter 4.

6 Data

In the previous two chapters, we have developed and verified a flood routing model based on network theory. The next step is to apply this model to a case. The necessary steps related to data are explained in this chapter. Data is case dependent; therefore, this chapter starts with an introduction to the case. Afterwards, four topics related to data are covered. In the second section, we define the data requirements. These requirements, originate from model design choices, case description and data availability. Based on the requirements, several data sources are selected and combined. In the third section, data processing is explained. The necessary steps are described to extract and combine data. The fourth section shows the relevant data for the case. Based on the available data, the model, as described in the previous chapters, is slightly adjusted for the case. These adjustments are the topic of the fifth and last section.

6.1. Case introduction

The chosen region is the Ganges Brahmaputra watershed, with focus on the Ganges river basin, which is the largest in India. In this region, floods happen regularly and have huge impacts. The Ganges river rises in the Himalaya, travels across fertile plains and flows into the Bay of Bengal. From north to south, the landscape plunges from an area with peaks, as high as the Mount Everest, to the flat Ganges plains less than 100 meters above sea level. The eastern part is a flat delta characterised by the Sundarbans mangrove systems. The total area of the basin is estimated at one million square kilometres and includes Nepal, a quarter of Bangladesh and nearly one-third of India (World Bank, 2014). An elevation map of the Ganges river basin is shown in figure 6.1.

The origin of the Ganges is the Gangotri Glacier in the Indian state Uttarakhand. Its major tributaries include the Yamuna, Gomti, Ghaghara, Gandaki and Kosi (see figure 6.2). The basin is part of the larger Ganges Brahmaputra Meghna basin. The river system is a complex interplay of monsoonal runoff, glacier melt and groundwater resources. The South Asian monsoon system is the most significant factor in defining the climate and hydrology of the Ganges river system. During an average year, 1,200 billion cubic meters of precipitations falls in the basin. The monsoon brings about 80% of the annual rainfall in just three months (i.e. June, July, and September) (World Bank, 2014).

The Ganges basin is home to more than 655 million people in its area. Poverty is widespread, with poverty rates around 30% and an average GDP per capita under 2 dollars per day. Because of the huge population, combined with poverty and extreme population density, the Ganges Basin can be considered a unique global challenge. Climate has always been both beneficial and a challenge. The natural cycle of inundation provides soil moisture that increases production. However, peak flows cause more extensive and devastating floods, which disrupt social life and economic activity. Traditionally, flood plains were only used for agriculture without building infrastructure and housing. However, the extreme population density has led to permanent habitation in flood plains. The World Bank considers the Ganges Basin as one of the most climate-vulnerable areas in the world and as a ‘hot spot’ for climate-induced conflicts (World Bank, 2014). Because of its size and the aforementioned problems the Ganges basin was selected for the case. Specifically, the 2013 floods are selected, in which the death toll neared 600 people and landslides devastated large regions and complete villages.

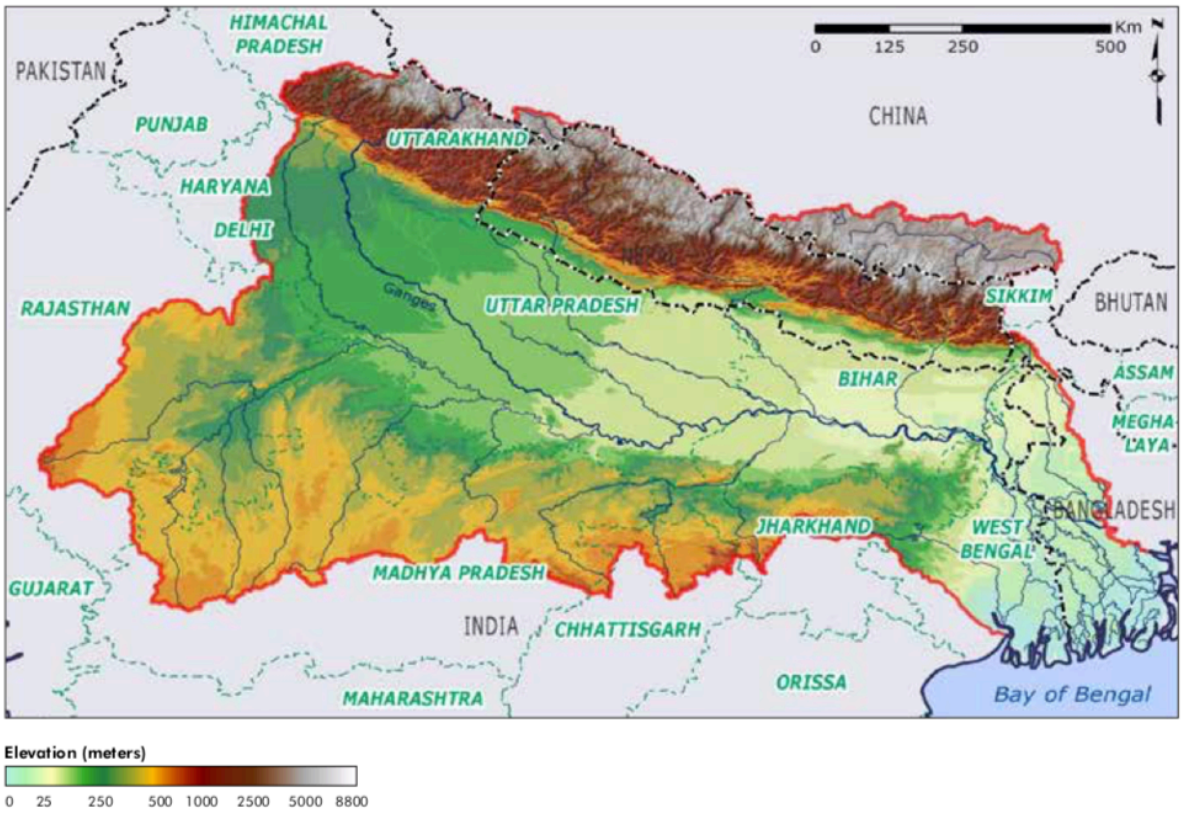


Figure 6.1: The elevation map of the Ganges basin. The Ganges rises in the Himalaya and within 200 kilometres the landscape changes from mountainous to the flat Ganges plains. The river ends in a flat delta and flows through the earth's largest mangrove ecosystem. (World Bank, 2014)

Figure 6.2: The full Ganges Brahmaputra Maghna basin showing the different rivers and basins. The coloured areas are the different river basins: the Ganges in orange, the Brahmaputra in puple and the Meghna basin in green. The main tributaries of the Ganges are Chambol, Yamuna, Gomti, Ghaghara, Gandaki and Kosi. The main outlet of the Ganges is called the Padma in Bangladesh. The secondary outlet is called the Hooghly river, which flows out to the Bay of Bengal near Kolkata. (Image source: https://upload.wikimedia.org/wikipedia/commons/3/34/Ganges-Brahmaputra-Meghna_basins.jpg)

6.2. Requirements and data sources

We have to define the data requirements before we can use the model. These requirements follow directly from the model and case. First of all, we need a network representation of a watershed, in which river reaches are represented as edges. This network cannot have cycles because the water in natural riverine systems can only flow downstream. Each reach needs two model-specific parameters x and k to perform routing, as we have seen in chapter 4. An edge capacity is needed if we want to obtain an indication of flood risk or overflow. The last requirement is inflow data, the information on how much water enters the system. Without inflow data, there is nothing to simulate. All requirements, including examples of inflow data, are listed below in four types:

1. **Network representation of a watershed in which each edge is defined as a river reach.**
    - The network cannot have cycles
2. **Capacity and average flow of the river reaches**
    - Used to analyse when a river reach overflows
3. **Inflow data**
    - Precipitation with a runoff coefficient
    - Glacier discharge data
    - Infiltration
    - Groundwater flow and storage
    - Evaporation
    - Evapotranspiration
4. **Model specific requirements**
    - x, the attenuation parameter
    - k, the lag parameter
    - Δt, the preferred time step

We need types 1), 2) and 4) to construct the network model, but it suffices to start with one type of inflow. Precipitation data was chosen, because floods in northern India are mainly caused by rainfall, as was described in the case introduction. In the next four subsections, sources for each data type are discussed.

6.2.1. Network representation

Several data sources were considered for each category. The HydroSHEDS (Lehner, Verdin, & Jarvis, 2008) (**Hydro**logical data and maps based on **SH**uttle **E**levation **D**erivatives at multiple **S**cales) dataset is selected to create a network representation. This dataset is produced by computers and based on digital elevation data. Several procedures are used to create a hydrological conditioned digital elevation model (DEM) and a flow directions model in a raster format. Based on these flow directions, a vectorised river network is created. The river network thus corresponds to local minima in the digital elevation model. For a complete description of the process, see appendix B.1. Figure 6.3 shows an artificially created example of river reaches in the HydroSHEDS database. A river reach is defined as a stretch between two tributaries, or between the respectively start or end of the network and a tributary.

The advantages are that it is well-structured data, and the network does not contain cycles or double connections. Moreover, it has a very high spatial resolution of 15 arc-sec (which corresponds to 450 m at equator). There are two main disadvantages. The first is that there are no bifurcations in this dataset (see appendix B.1). The second is that the method does not work well in areas with little relief, such as flood plains and deltas.

Other datasets that were considered are, for example, open street maps data. These sets, however, do include bifurcations, but the main problem is that they are inconsistent because they are human-made. Example problems are, river reaches occur double, cycles can exist and reaches are not precisely connected. Because of these drawbacks, HydroSHEDS data was preferred.

6.2.2. Capacity and average flow of reaches

Another advantage of using HydroSHEDS data is that it is recently extended with another dataset: GloRiC (Dallaire, Lehner, Sayre, & Thieme, 2019) (**Glo**bal **Ri**ver **C**lassification). In this dataset, river reach

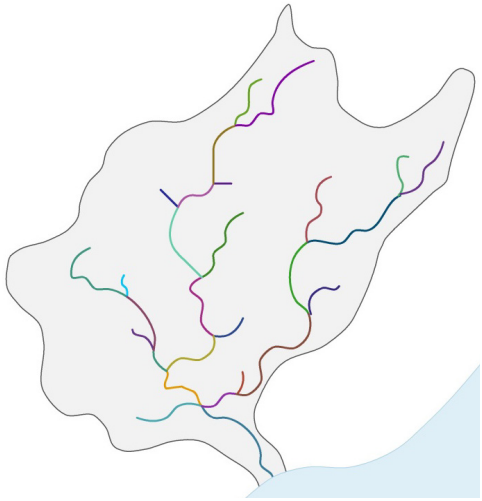


Figure 6.3: Artificially created example of river reaches in the HydroSHEDS/GloRiC dataset. A river reach, indicated by its own colour, is defined as a stretch between two tributaries, or between the start/end of the network and a tributary. The grey contour indicates the watershed. The river ends at the sink, a lake or ocean, indicated with the blue area. We can see that no bifurcations exist in this dataset.

properties are added from a numerous amount of other sources. One of these sources is WaterGAP (Water – Global Assessment and Prognosis) (Alcamo et al., 2003), which is a global water use and hydrology model. WaterGAP adds long term average discharge [m^3/s] and flow regime variability to the dataset. Flow regime variability is defined as 'maximum long-term average monthly discharge / long-term average discharge'. Both values are rescaled from 0.5 degree pixel resolution (corresponding to 50 km at the equator). The average discharge can be used to set a static flow on the reaches. Furthermore, although we do not know the capacity of each river reach, the flow regime variability can be a good proxy.

6.2.3. Inflow data, precipitation

As described above, precipitation is selected as inflow data because it is the leading cause of floods in northern India. This was the most difficult dataset to find because it has specific requirements: we need both a reasonable spatial resolution and a high temporal resolution. A lot of precipitation data is aggregated to daily statistics. This is not useful in flood modelling because the intensity is crucial in flood modelling. It matters whether three millimetres of rain fall in 15 minutes or in 24 hours. Another problem is that rain gauges can have a high temporal resolution, but the spatial resolution is meagre. Therefore the measurements they make are not representative of a whole area. For example, they could easily miss local rainfall events.

To tackle these problems, satellite precipitation data is used. NASA has the Precipitation Measurement Missions, PMM (NASA, 2019c). PMM consist of two missions: the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM). TRMM lasted from 1997 till 2015 and GPM is its successor. GPM, as the successor, has a higher spatial resolution and better sensitivity for light rainfall. The principle is the same: rainfall is estimated by combining measurements from several satellites. The satellites make use of both active radars and passive radiometers to measure reflected and natural thermal radiation. These radar reflectivities and natural radiation are processed by several algorithms to obtain precipitation information. The resulting data is calibrated with ground measurements. For more detail on the measurements, see appendix B.2.

The final results are calibrated rainfall intensities with a temporal resolution of 30 minutes and a spatial resolution of 0.1 degrees (this corresponds to roughly 11km at the equator) (Huffman, Stocker, Bolvin, Nelkin, & Tan, 2019). There are some downsides of satellite precipitation measurement. The main downside is that the GPM can miss pockets of heavy rain in smaller areas, or short periods of (more or less) intense rain, because data is collected over large areas and is based on when satellites pass overhead. Therefore local rainfall totals measured by satellite can differ from ground measurements. The processed, calibrated, and combined data is called IMERG. IMERG stands for Integrated Multi-satellitE Retrievals for GPM.

The rainfall data combined with a runoff coefficient as described in Subsection 4.4.1 is used as the external inflow for each reach. The runoff coefficient differs between 0.2 and 0.5 per month (Chowdhury & Ward, 2004). To be able to use this data, the rainfall data in grid format has to be linked to the reaches in vector format. This is explained in the next section, data processing.

6.2.4. Model-specific parameters

So far, we have discussed the external datasets. Ideally, we want to have datasets that supply the model parameters x, the attenuation, and k, the lag time, for all reaches. Unfortunately, such dataset does not exist, and these parameters are very challenging to estimate. Both parameters require detailed flow information, for each reach, to fit them (Karahan, 2012). Alternatively, they require complicated optimisation procedures for ungauged reaches (David et al., 2011; Yoo Chulsang et al., 2017).

Because of this lack of information, the values for these parameters are estimated based on proxies. Many sources in literature state a value of x =0.2 can be used if the real value is unknown. Therefore we will use this estimation.

The parameter k is based on reach length and flow velocity in the Ganges river. The relation then is k = length/velocity. Reach length is already specified in the HydroSHEDS dataset, so we need flow velocities. One study found a long term (11 years) average of 2 m/s (Bhutiani, Khanna, Kulkarni, & Ruhela, 2016). This value of 2 m/s has been used in the model, whereas the reach length is supplied by the HydroSHEDS dataset. There are several issues with selecting a fixed long term average: firstly, in cases of rainfall flow discharge will increase, and with increasing discharge flow velocity will also increase. For example, a maximum of 54 m/s was measured in the same time frame. Secondly, in reality, flow velocity will differ from region to region. For example, it is expected that mountainous regions have higher flow velocities than flatlands or deltas due to a more substantial height difference.

The third model parameter Δt, can be chosen freely within constraint 4.7. To allow for efficient calculation, the time step is based on the temporal resolution of the precipitation data. The GPM has a spatial resolution of thirty minutes. Therefore, taking a smaller time step does not add any resolution and will only increase the run time. While taking a larger time step will result in a loss of information. Therefore $\Delta t = 30$m equal to the temporal resolution of the GPM data.

An overview of all data requirements, the corresponding data sources and assumptions are given in table 6.1.

Table 6.1: Summary of data requirements, corresponding sources and assumptions. One data requirement is not yet discussed, which are the micro watersheds. The need for micro watersheds will be explained in Section 6.3. [1] Chowdhury and Ward (2004) [2] Bhutiani, Khanna, Kulkarni, and Ruhela (2016) [3] Elbashir (2011), Al-Humoud and Esen (2006), Karahan (2012), Mays (2010)

Data requirement	Source	Type	Useful properties	Spatial resolution	Temporal resolution	Comment
Network representation	HydroSHEDS - flow	Vector	Length	15 arc-s / 450m	-	No bifurcations
Capacity and avg flow	GloRiC	Vector	Avg flow & variability	15 arc-s / 450m	-	From WaterGAP
Micro watersheds	HydroSHEDS - dir	Grid	Flow directions	3 arc-s / 90m	-	-
Rain data	PMM	Grid	Precipitation in mm/h	0.1 deg / 11km	30 min	High temporal resolution
Runoff coefficient	[1]	-	0.2 - 0.5	-	-	-
Stream velocity	[2]	-	1 - 54 m/s (avg: 2 m/s)	-	-	
Model parameter x	[3]	-	0.2 if unknown	-	-	-
Model parameter k	-	-	reach length / velocity	-	-	Based on flow velocity
Model parameter Δt	Equal PMM resolution	-	-	-	30 min	-

6.3. Data processing

In the previous section, the different datasets were described. Processing has to be done in order to connect different datasets. There are two main issues: one, we want to extract the watershed from the dataset that contains all river reaches worldwide. Two, we want to know how much rain flows into which reach for every time step. The second issue is divided into two processing steps: first, we create micro watersheds belonging to each reach. Second, we use these micro watersheds to determine the inflow for each reach. A short description of these processing steps is given below. The full diagram showing all steps is given in appendix B.3.

6.3.1. Extracting river network

The HydroSHEDS and GloRic dataset contain all river reaches in the world. Each river reach entry consist of properties and a georeferenced vector. The Gloric dataset is too big to load into memory at once; therefore, the first step is to filter out only river reaches in Asia. To reduce the number of reaches, reaches with an average flow less than 1 m^3/s are filtered out. Most of these reaches are just perpendicular to a main stream, and they do not add relevant detail. The last step is to only select the reaches of the Ganges-Brahmaputra-Meghna basin.

This is done by manually selecting the reach that flows into the ocean in the Bay of Bengal. A script is used to find all reaches that flow into this reach in a recursive manner. These extracted reaches are saved into a separate file for later processing. The result can be seen in figure 6.4 in which the plotted thickness is scaled with the average flow. The red reaches are reaches with an average flow exceeding 200 cubic meters per second. When we compare this figure with figure 6.2 we can see that the main tributaries correspond well with the red coloured reaches. A notable difference is the missing of bifurcations in the dataset: the Hooghly river is missing, the river flowing through Kolkata into the Bay of Bengal.

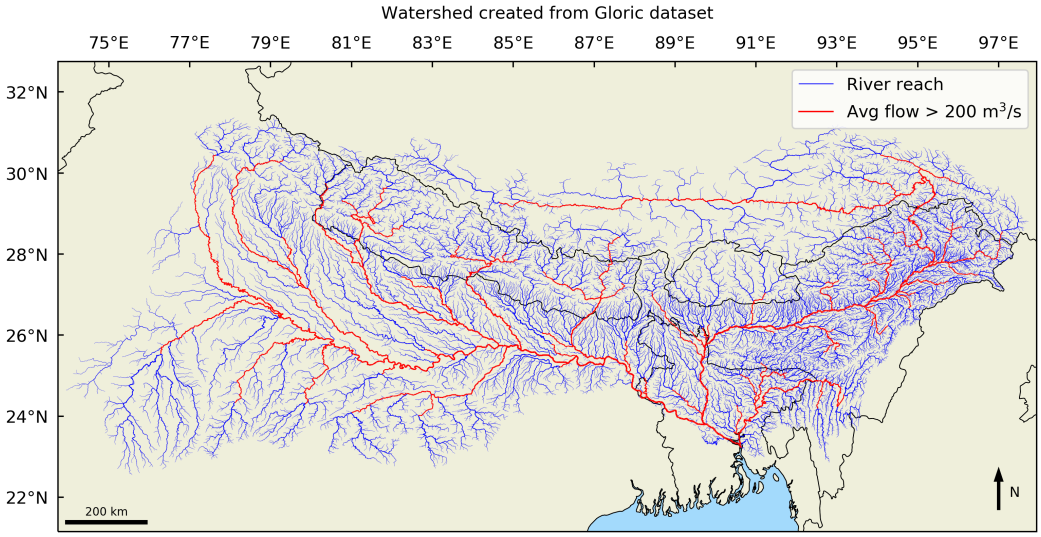


Figure 6.4: Figure showing the reaches extracted from the GlorRiC dataset. When compared to figure 6.2 the main tributaries are in close agreement. One notable difference is that the HydroSHEDS dataset does not contain bifurcations. This can be seen, for example, in the delta where the Hoogly river is missing that flows through Kolkata.

6.3.2. Translation layer: micro watersheds for each river reach

As explained before, we want to know how much rain flows into which reach at every time step. The challenge here is that the reaches of the river network are defined in a vector format, whereas the rain data is specified in a grid format. In order to know how much rain flows into each river reach, we need a physical area associated with each river reach. The rain that falls in this area will eventually flow in the corresponding reach. For a complete river system, this area is called a watershed. Therefore, this area is defined as the micro watershed, a watershed for each river reach.

The procedure to create these micro watersheds uses a combination of flow direction data in a grid format and the reach vectors as obtained in the previous step. The algorithm works as follows: for each reach vector, the algorithm determines the first and last pixel from the flow data that correspond with the vector. This can be seen in figure 6.5a. All pixels that flow into this pixel are added recursively. For this last pixel, the algorithm continues but does not add pixels in the directions of the inflowing vector(s). When the vector corresponds to a source reach, all inflowing pixels are added. In the last step, the areas of the selected pixels are combined into a polygon corresponding to the micro watershed of a reach. As can be seen in figure 6.5b. This procedure is repeated for each vector, and the result is a map with micro watersheds that cover the complete watershed. Multiple of these micro watersheds are shown in figure 6.3. We have now created a transition layer that can map rainfall data to reaches.

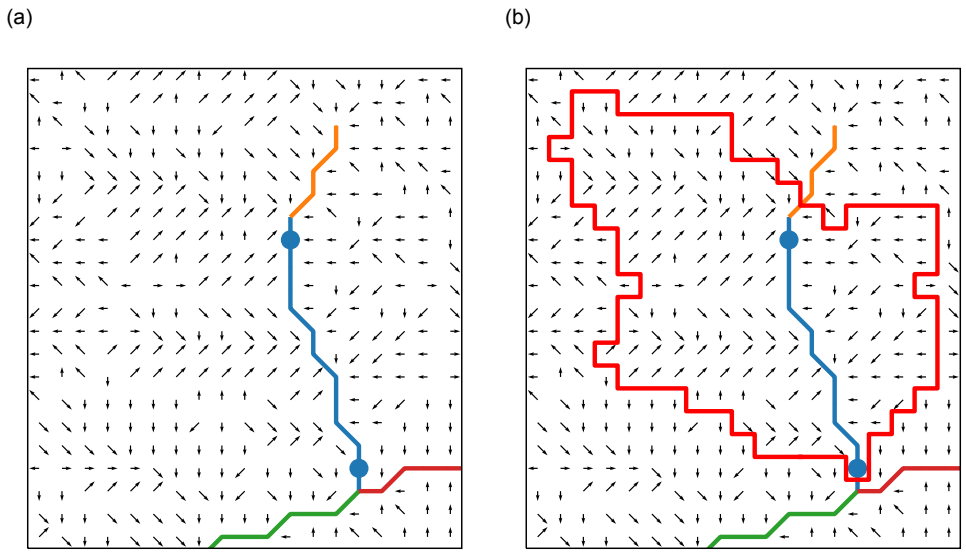


Figure 6.5: Figures showing the procedure to select micro watersheds. Figure a) shows the reaches in different colours, with flow directions for the underlying pixels. We want to determine the micro watershed of the blue reach. The start and ending pixels of the reach are selected from the flow direction data, these are the blue dots. All pixels are selected that have an inflow that ends in the ending pixel. Figure b) shows the polygon that consists of the selected pixels, this is the micro watershed. The micro watershed tells that rain falling in this region will end up in this reach.

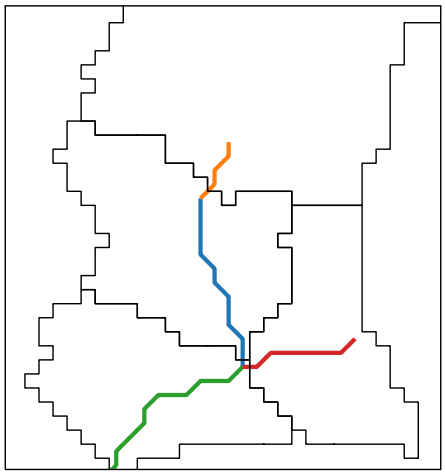


Figure 6.6: Showing multiple micro watersheds in the network. All micro watersheds should cover the complete watershed without gaps.

6.3.3. Precipitation: from grid to micro watershed

The GPM data of the PMM is in a grid format and has to be converted to rain rates for each reach. The data is provided in an HDF5 data format and has to be transformed into georeferenced data, e.g. a GeoTIFF. These GeoTIFFs are cropped because only rain data in the region is necessary. Then, the data is resampled to the same spatial resolution as the micro watersheds. In the final step, the polygons of the micro watersheds are overlaid on the rain data. An average is taken for all pixels inside the polygon. This average is the rainfall intensity for this watershed in this half hour. The intensities for all half hours are combined into a single data file. To obtain the inflow, the intensity is multiplied by the runoff coefficient and the area of the micro watershed. The area for each polygon had to be calculated in an area-preserving coordinate reference system.

6.4. Case precipitation data

In June 2013 heavy monsoon rainfall caused extreme floods in northern India. This case was chosen for a first analysis because the consequences were enormous and many lives were lost. Another reason for selecting this case is that NASA has published rainfall data and flood maps. These data are useful for qualitative comparison with model results. Figure 6.7 shows the total rainfall fallen in Northern India and Nepal over the course of a week published by NASA in 2013. We reproduced the rainfall data from the raw data sets in figure 6.8. We can see quite some similarities, but the main difference is that there seems to be a patch of rain missing in China, just across the border with India.

Calibration effects and different data sources can explain this difference in pattern. The first map, created by NASA, is based on data from NASA's MPA: Multisatellite Precipitation Analysis. NASA estimates rainfall by combining measurements from several satellites. These estimations are calibrated with measurements from the Tropical Rainfall Measuring Mission, TRMM. The figure was published on the 22nd of June, and this means that possibly no ground calibration has taken place. It can take weeks to months to finalise ground calibration.

Figure 6.8 is based on data from NASA's GPM, Global Precipitation Mission. GPM is the successor of TRMM and contains more detailed and light rainfall. The successor of MPA is IMERG, and MPA data from the past is translated into the IMERG datasets. The data shows considerable similarities to NASA's rainfall map. Note that the data used has a higher resolution (0.1 and 0.25 degrees, respectively). Differences can be explained by the transformation from MPA to IMERG datasets or because of difference in calibration.

For the analysis itself, the two days in this week with the most rainfall were selected. Reducing the number of days reduces the run time of the model. These days are June 16th and 17th. Figure 6.8b shows the rainfall in these days. The areas where they received heavy rainfalls were the same over these days.

6.5. Model case adjustments

The model developed in chapter 4 is slightly adjusted for the case. Some simplifications could be made because of data formats. First of all, the river network dataset, HydroSHEDS, does not contain any bifurcations. Because of this, the network contains an equal amount of nodes and edges. Node properties can be considered edge properties and vice-versa. Computational load is reduced by not having to define these edge properties.

An issue with the data is that in and outflow of nodes are not always balanced: the summed flow of two reaches before a confluence is not always equal to the average flow after the confluence. Water is either added or subtracted from the reaches. This can have multiple reasons: there can be errors in the average flow values. It can also be the case that water comes from another source than rainfall, for instance, groundwater flows. Water can also flow out of the reaches, because the model behind these values, WaterGAP, includes water use. Examples are water use by farmers, large industries or other processes such as evaporation. An extra static inflow is defined to match the inflow and outflows of reaches. A negative inflow accounts for outflow. We assumed that these static flows are independent of river discharge. After determining these static inflows, water from the source reaches flows into the subsequent reaches as expected.

After all data preparations and model adjustments, we are ready to start the simulation. In the next chapter the results for the case will be presented.

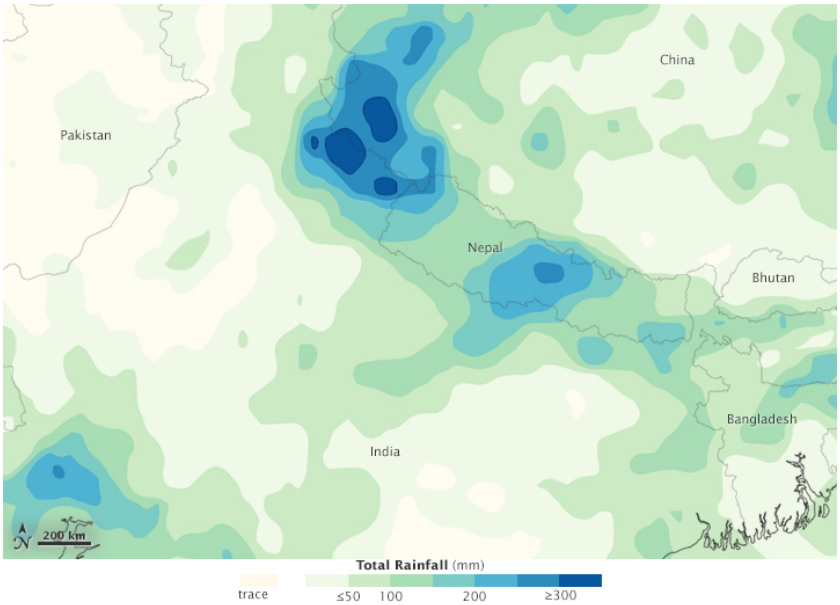


Figure 6.7: Figure from NASA for validation of rainfall data. June 14 - 20, 2013. The map is based on data from NASA's MPA, Multisatellite Precipitation Analysis. NASA estimates rainfall by combining measurements from several satellites. These estimations are calibrated with measurements from the Tropical Rainfall Measuring Mission, TRMM. Because data collected over large areas and is based on when satellites pass overhead, the MPA can miss pockets of heavy rain in smaller areas or short periods of (more or less) intense rain. Local rainfall totals measured by satellite can differ from ground measurements. (NASA, 2013)

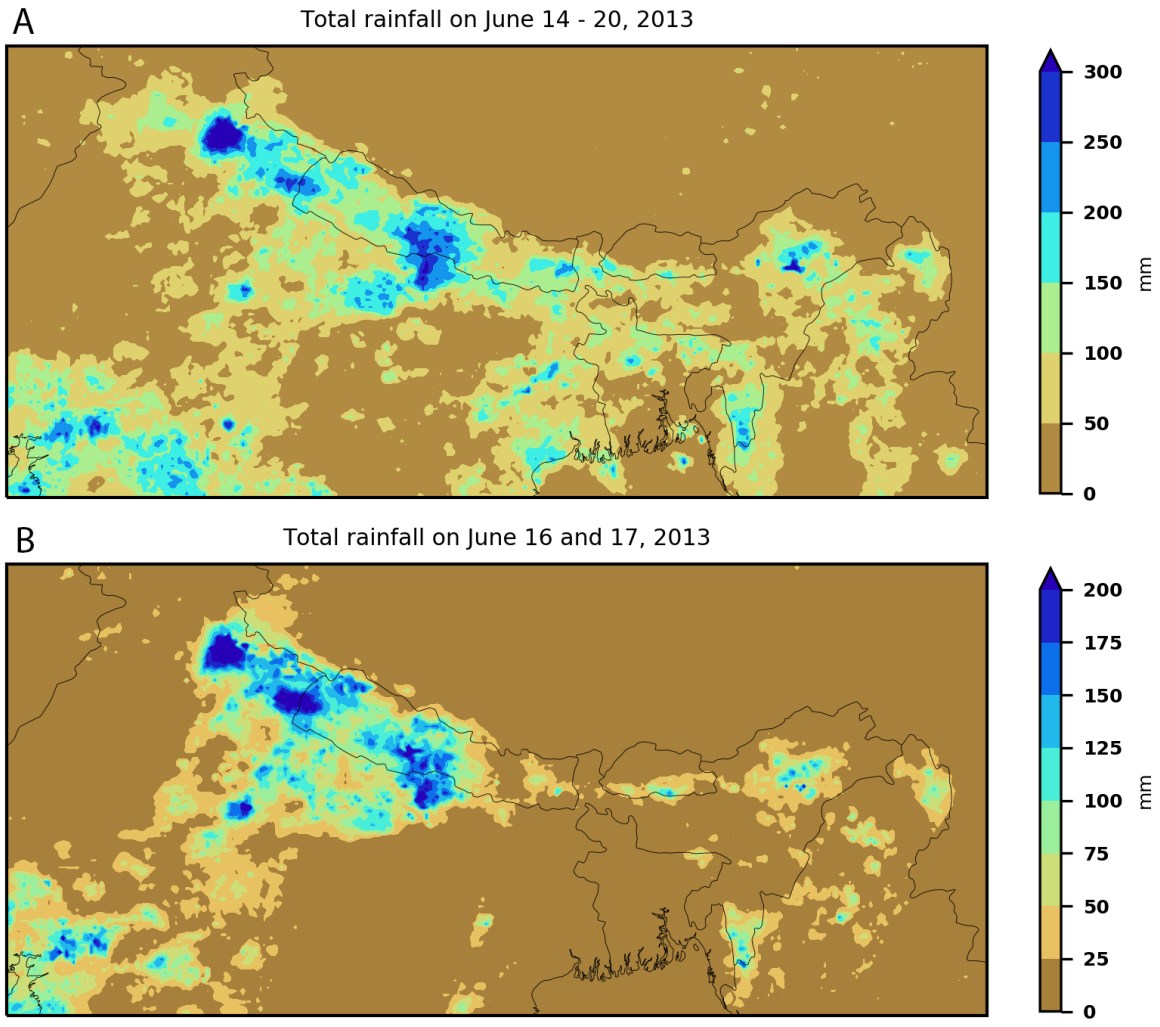


Figure 6.8: Figure a) showing total precipitation in North India on the days of the flooding. The map is based on data from NASA's GPM, Global Precipitation Mission. GPM is the successor of TRMM and contains more detailed and light rainfall. The successor of MPA is IMERG and MPA data from the past is translated into the IMERG datasets. IMERG stands for Integrated Multi-satellitE Retrievals for GPM. The data shows considerable similarities to figure 6.7, But the data used has a higher resolution (0.1 vs 0.25 degrees). Differences can be explained by the transformation from MPA to IMERG datasets. Moreover, datasets are calibrated with ground measurements, but this can take a couple of months. This figure uses calibrated data, while it is unknown for figure 6.7. b) Showing the same data but only for the 16th and 17th of June. Scaling is adjusted to less rainfall. This data is used as input data for the model.

7

Results

All collected data is merged and added to the model as described in Chapter 4. For this simulation, the following parameters are used: $x = 0.2$, a flow velocity of 2 m/s and a runoff coefficient of 0.5. The model was run on a laptop computer and the run time is 7 minutes per simulated day. Before showing the results in the whole river basin, some concepts have to be clarified.

The plot in figure 7.1 shows a discharge graph for a single reach. Inflow and outflow of the node are plotted in blue and orange, respectively. The red and purple lines indicate the average and maximum discharge (long term monthly average, based on GloRiC). Green is the resulting runoff in this river reach. Overflow is defined as the discharge rate above long term monthly average discharge. The overflow is indicated with the red surface in the graph. In this example, we can see the cascading effect of floods. Rain in this reach was not intense enough to cause the excess flow. Thus rain from upstream parts caused the overflow in this reach.

Figure 7.2 shows the maximum overflow of each reach for the full basin. This is the maximum for the two days, so compared to figure 7.1, it would be the overflow value at t = 13 – the larger this overflow, the darker red the colours. Moreover, the scaling is plotted logarithmically because the differences in flow are very large. To understand where these patterns come from, we combined rain data and flow data in a single figure (7.3). Here we clearly see the rain patterns in the mountains in the north. The resulting overflows occur downstream, which is a clear example of flood cascades happening. At the locations of the rain, there do not have to be any issues, nevertheless downstream there can be.

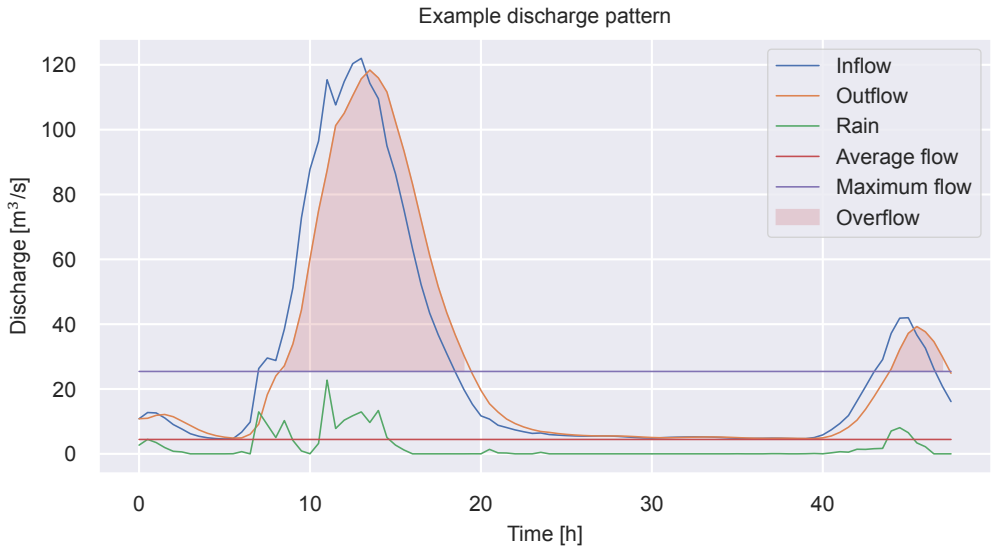


Figure 7.1: Discharge graph for a single reach showing the cascading effect. Discharge from rain in this reach is not large enough to cause overflow; the green line stays below the purple line. Thus the overflow is caused by rain upstream in the riverine system.

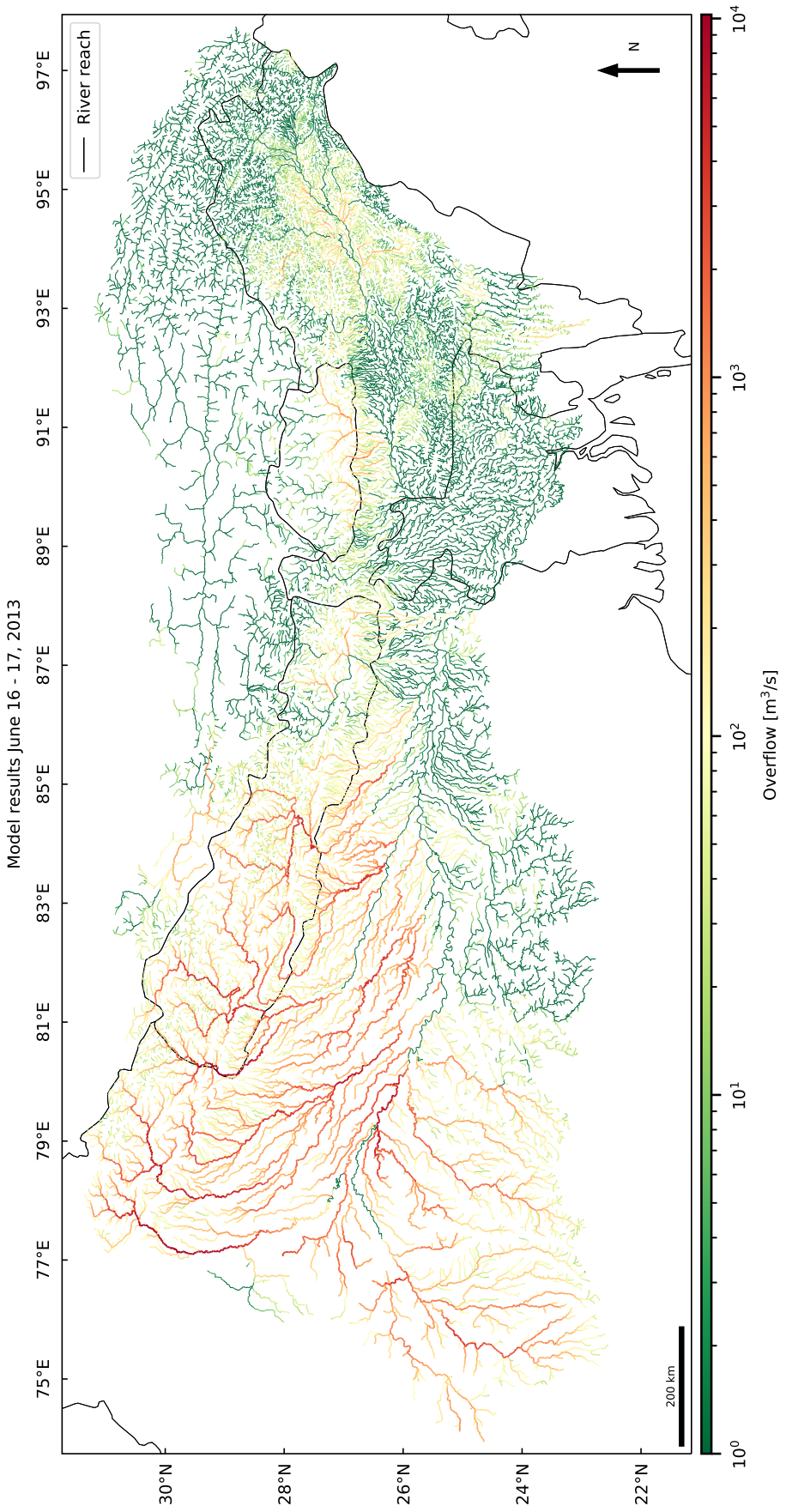


Figure 7.2: Figure showing model results for the whole basin for 16th and 17th of June. The colour indicates the amount of flow above the monthly long term maximum. The maximum value over both days is taken and plotted on a logarithmic scale. The main overflows are in the Ganges, Yamuna and Ghaghara River.

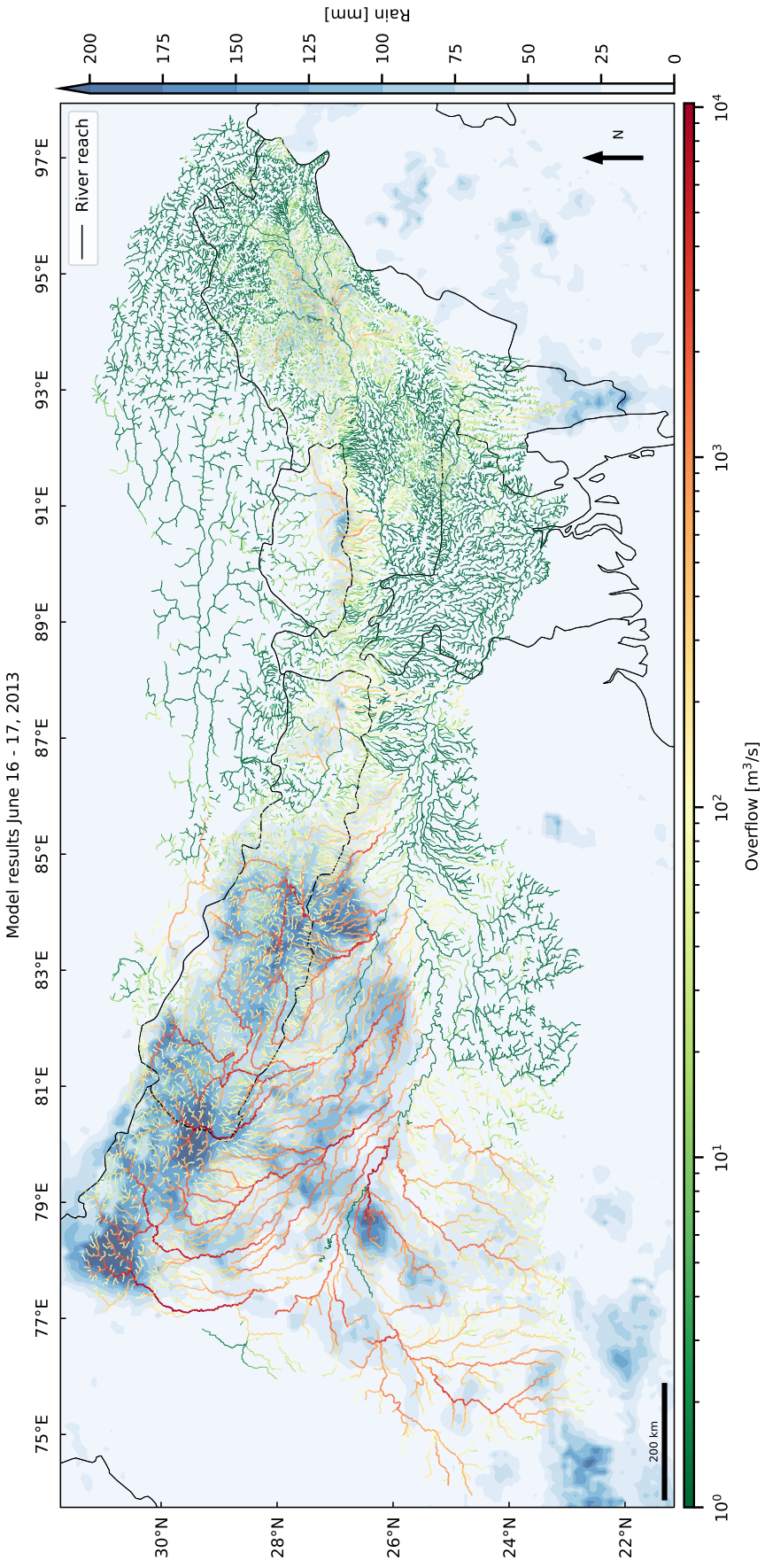


Figure 7.3: Figure similar to 7.2 showing model results for the whole basin for 16th and 17th of June. The green to red colour indicates the amount of flow above the monthly long term maximum. The blue colour gradient indicates the total amount of rain fallen in the two days at these locations. Rainfall in the north causes a cascading effect: downstream rivers show a large overflow.

7.1. Rainfall - overflow correlation

We can obtain a measure for the locality of effect by plotting rainfall versus overflow for each reach. This is presented in figure 7.4. If rainfall causes a local overflow, we would expect a correlation between increasing rainfall and overflow in that reach. If that is not the case, rainfall causes overflows downstream than there is no clear correlation between the two. Figure 7.4 shows that there is no clear correlation between rainfall and overflow within the same reach. This figure shows that the effect of rainfall is not local, but downstream, which is also an indication of the cascading effect.

Another observation from the plot is that the dots that form horizontal patterns, for example, just above 8000 m^3/s. These patterns exist because of subsequently connected reaches. If a reach is flooded, the neighbouring reaches are likely to be flooded by the same amount. These dots, at the same overflow level, with slightly different rainfall, appear as horizontal 'lines' in the graph.

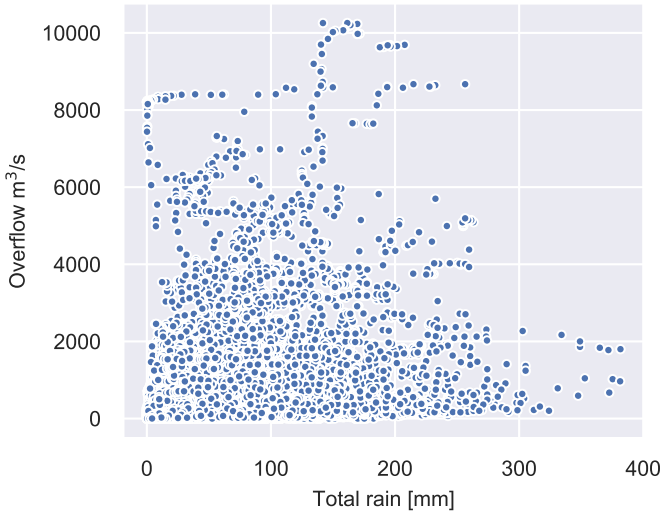


Figure 7.4: Showing correlation between rainfall and overflow. No clear correlation can be seen. A positive correlation would indicate local effects of rain: the larger the rainfall in the area, the larger the overflow. The lack of correlation can indicate the cascading effect of water waves: overflows happen downstream of the direct source.

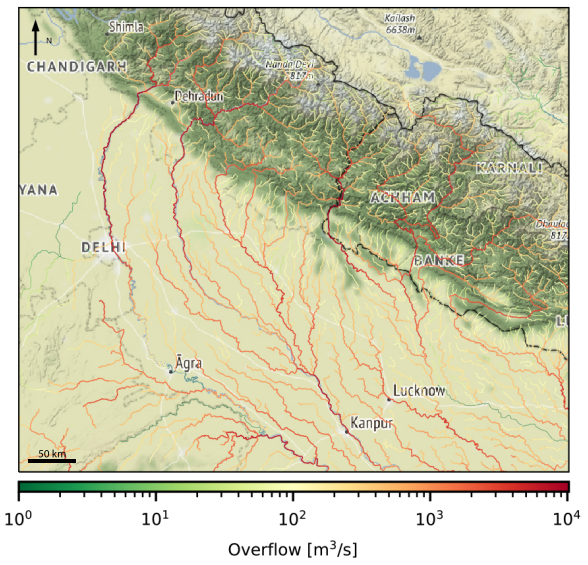
7.2. Comparison with satellite images

We compare the model results with satellite imagery from NASA to check model behaviour for this case. The intended use of this model is to discover high-risk regions. By comparing with the satellite images, we can see whether the overflow areas match. Figure 7.5 shows four sub-figures. The upper figures show model results. Figure 7.5a shows an overlay of the river network on a terrain map. The colours again indicate the amount of overflow. Its neighbour, figure 7.5b shows the same network combined with a rain map.

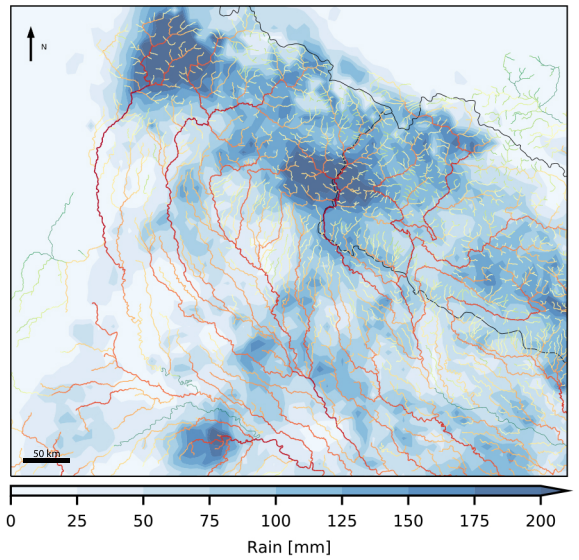
The lower figures show MODIS satellite images. MODIS stands for MODerate resolution Imaging Spectroradiometer. As the name suggests, it acquires data from multiple spectral bands: 36 in total. The two MODIS satellites cover the earth every 1 to 2 days. Their goal is to provide measurements in large scale global dynamics. These multiple wavelengths make it possible to identify, for example, water, vegetation and ice.

The lower MODIS images figures in 7.5 show a combination of visible and infrared light. They are falsely coloured to help see the difference between water and land. Water appears blue, vegetation in bright green, clouds are pale blue-green and cast shadows. Glacier ice and snow in the Himalayas are coloured pale blue to cyan.

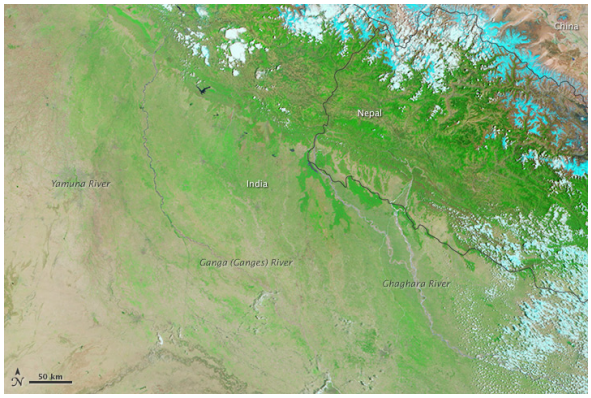
The left MODIS image (7.5c) shows the area before the floods, on May 30th. The right image (7.5d) shows the area after the floods, June 21st. The resulting water in the rivers and river banks is clearly visible in the after image. The rivers that are full of water are, from left to right, Yamuna, Ganges and Ghaghara. In the upper figures, we can see the same three rivers that are dark red. These are the rivers with the simulated high overflows.



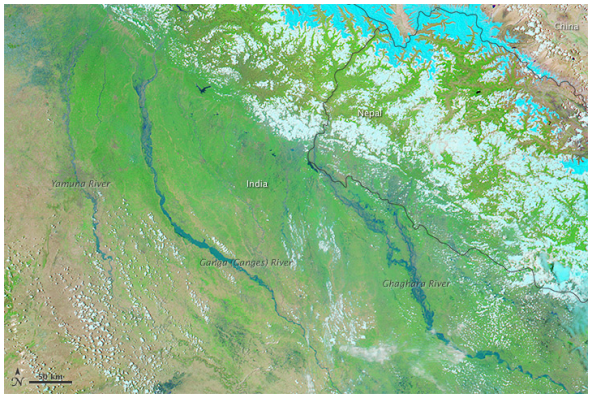
(a) Model results plotted on terrain map



(b) Model results overlaid on rain map



(c) MODIS image before floods, May 30, 2013



(d) MODIS image after floods, June 21, 2013

Figure 7.5: Figures showing model results and satellite imagery of northern India. In figures a) and b) we can see the results. In figure a) the model results are overlaid on a terrain map. This terrain map shows the region and some large cities. In Figure b) the model results are overlaid on the rain map of June 16th and 17th. In Figures c) and d) show MODIS images of the same region, before and after the floods. These MODIS images use a combination of infrared and visible light to make a better distinction between water and land. Water appears blue, vegetation in bright green, clouds are pale blue-green and cast shadows. Glacier ice and snow in the Himalayas are coloured pale blue to cyan. In image c) Yamuna, Ganges and Ghaghara clearly flooded over: they show a lot of water on the river banks. When we compare figure d) with a) we can see that the same three rivers are highlighted dark red. In the model, these rivers thus show a very high overflow.

8

Discussion

8.1. Interpretation of results

The results show the simulated overflow: the discharge above the capacity. Capacity is defined as the maximum long-term average monthly discharge. An overflow thus does not indicate an actual flood, as we do not know when an actual flood will happen. It can be seen as an indication of flood risk: the larger the overflow, the more risk there is for an actual flood.

In the results, we can clearly see the cascading effects as defined in section 4.1. Water from upstream sources: in this case, rain in upstream regions, causes large overflows in downstream reaches. The simple network model, that consist of small pieces with relatively simple behaviour, is thus able to show complex emerging behaviour. One could argue that this is emergent non-trivial behaviour because overflow occurs in places where it is not necessarily expected. As a result, the second research objective is fulfilled: a river system is represented using complex network theory for understanding flood routing in riverine systems.

The first research objective was to develop a model that is efficient, uses less information, is easy to understand and useful for policy analysts and decision-makers. Here we reflect on efficiency, information use and understandability. In the next section, usefulness is discussed. Because of the simple network representation, the model is easy to understand even for people without a mathematical or computer science background. The network representation also makes the model very efficient. It can simulate the whole Ganges-Brahmaputra watershed in seven minutes for a full day[1]. Furthermore, information usage is reduced by using a hydrologic routing model. But because of the network structure, spatial information is still included.

To compare the developed model with the existing models, the developed model is added to table 3.1. The resulting table, table (8.1), is shown here. Compared to statistical methods, the model uses more data and is more development effort is required. The use of more data is necessary to include topological data. More data automatically leads to a much higher parameter complexity. Still much less data is used compared to continuous flow simulations, especially hydraulic routing models.

Table 8.1: Qualitative analysis of flood estimation and risk methods, showing data usage and development effort. Adjusted version of table 3.1 in which the developed model is added.

Method	Fitting observed data	Topological data	Parameter complexity	Development effort
Empirical / frequentist	++	–	–	-
Catchment design: QdF	++	-	–	+
Continuous flow simulation	+	++	++	+++
Mixed methods	+	+-	+-	+
Developed method	+	+	+	+

[1] Laptop with Intel Core i7-3720Q, 2.6 GHz, 16GB ram (2012)

8.2. Modelling for policy analysis

This research started with the research aim:

> To develop a flood routing model that is efficient, uses less information, is easy to understand and is useful for policy analysts and decision-makers.

The goal was to develop a model that is useful and suitable for policy analysts and decision-makers. More specifically the model should be useful for risk screening. This leads to the two other requirements 1) the model should be efficient and 2) it should be understandable.

The model should be efficient in order to be able to calculate and compare different scenarios. Scenario analysis is a common tool in risk screening and modelling for policy analysis. As we have seen, the model is efficient: it runs a single day in seven minutes. This runtime is achieved without optimising the code, which leaves room for improvement.

Model simplicity is important because policy analysts and especially decision-makers should have confidence in a model. A black box model or model that is hard to understand will lead to reduced confidence in its outcomes, and this undermines usability. The concepts behind the developed model are based on network theory. It is this decomposition that leads to small and understandable parts. This simplicity is believed to increase willingness to comprehend and use this model. Moreover, the use of publicly available data sources also reduces the development effort and investments necessary for data gathering.

So how can this model be used by policy analysts for risk screening? Three use cases have been identified. The first use case is the identification of vulnerable reaches. The overflow criterium is a measure or proxy for flood risk, as mentioned in the previous section. The larger the overflow, the more risk there is for an actual flood. A screening procedure could be as follows: The availability of rain datasets make it possible to simulate over long time periods. Overflow, as the risk proxy, can then be aggregated to get indications of vulnerable reaches within the river system.

More information can be extracted by combining the results of such analysis with geographical information. There are two example options: policy analysts could overlay population density maps to identify high risk reaches, i.e. these are flowing through densely populated areas. Another option is to overlay critical infrastructures on the risk map. This could lead to the determination of crucial bridges, roads and other infrastructures. Based on the results, vulnerable infrastructure can be reinforced, or alternative evacuation routes can be designed.

A second use case is to simulate the effect of climate change. It is expected that climate change results in more intense precipitation events, even if though total precipitation decreases (Trenberth, 2011). With this model, it is possible to simulate the effect of any arbitrary rainfall event. These events could be amplified past events or generated. The generated events can be based on forecasted rainfall distribution, for instance forecasts resulting from climate change studies.

A third use case is to forecast the discharge levels within the river system. To enable forecasting the model should be coupled to real-time water levels and flows in the system. It is possible to simulate the resulting downstream effects with this flow of information. Also, the expected rainfall within the next hours or days can be included in this system. The model can then be used as an early warning system: alerting when and where high levels are expected.

8.2.1. Policy options

The three use cases, discussed in the previous section, can provide insight into the riverine system and identify weaknesses. The next step is to simulate policy interventions. There are two ways to adjust the model for policy interventions: parameters adjustments and model changes. An overview of policy options and corresponding model adjustments is given in Table 8.2. Two of the interventions have been selected to clarify further.

Lowering embankments This option can be implemented in two ways. The first suggestion is to use a parallel network in which the river consists of two edges. One edge for the main channel and a second edge for the embankment. If a flood occurs, the main channel will reach its capacity and water flows into the embankment. The embankment has its own Muskingum parameters, notably a longer lag time. As long as river discharge is above the main channel capacity, the water levels in the main channel and embankment are equal. The parallel edge then acts as a storage system for the main with its own

downstream flow to the next embankment section. Another option is to use a variable time Muskingum scheme. An explanation of this method is described in Subsection 4.2.2. The Muskingum parameters change, as soon as discharge exceeds the main channel capacity. In this way, it is possible to account for the extra storage effect and longer lag time in the embankment. There is a mass conservation issue when using the variable time Muskingum method; those should be taken into account (Todini, 2007).

Reservoirs and hydro dams Implementing reservoirs and hydro dams is straight forward. Because the model is time step based one can add another type of node. This node can have a storage capacity and a policy to determine out and inflow. For hydro dams, this inflow cannot be regulated. The policy for the outflow can determine on the water level, inflow and outflow discharges. These policies can also be adaptive to precipitation forecasts or upstream flow measurements. For example: if a lot of rain or inflow is expected, outflow discharge can be increased to make room in the reservoir.

8.3. Limitations

In this section, the limitations of the model are discussed.

8.3.1. Muskingum routing

In subsection 4.2.2 the limitations of the Muskingum method itself were discussed. The first limitation is that no flood or failure mechanisms are included in the model. The effects of for example overbank flow, or overtopping are not included. Suggestions on how to implement these are described in 8.2.1 and C.

Another problem has to do with the chosen time step and rainfall data. Because the time step is set to 30 minutes, no flash floods can be studied. Flash floods can occur unexpectedly within several minutes. The developed model is not suitable for studying effects that happen such a short time span.

8.3.2. Failure mechanisms

Failure mechanisms are not yet included in the model. First, a proof of concept is developed before implementing these mechanisms. Instead, the assumption was made that water continues flowing even when an overflow happens. However, in reality, overbank flows have a considerable effect on river dynamics. Water that flows out of the river slows down and stays on the banks. This storage and buffering effect was not taken into account in this model. Because this water is temporarily taken out of the river, downstream flood risk is reduced. This would mean that the overflow cascades are shorter in reality than in this model.

8.3.3. Parameter estimation

The model uses five parameters that can influence the outcome. A full list given in Table 8.3. It is possible to divide the consequences into two categories: Spatial consequences and impact consequences. Spatial consequences are defined as the change in location as a result of changing parameters. Im-

Table 8.2: Different policy options and how they adjust the model to implement.

Policy option	Possible in model	Parameter or model adjustments	Comment
Smoothening Deepening river Creating dikes	yes	adjust capacity parameter	options result in more capacity
Lowering embankments	yes	use parallel network use variable time Muskingum	see 8.2.1 see 4.2.2
Secondary channels	yes	add edges to network	
Reservoirs / hydro dams	yes	extend model	see 8.2.1
Land-use change	yes	adjust runoff parameter	
Widening small sections (e.g. removing bridges)	no		too detailed to implement

Table 8.3: Overview of parameters assumptions, consequences and possible improvements.
[1] Estimation techniques for ungauged reaches: David et al. (2011), Yoo Chulsang, Lee Jinwook, and Lee Myungseob (2017). Another option is to classify reaches into several classes: e.g. mountainous, midland and lowland and use different sets of x and k. [2] Usage of runoff maps based on remote sensing data: Shi et al. (2007), de Smedt, Yongbo, and Gebremeskel (2000), Zade, Ray, Dutta, and Panigrahy (2005)

Parameter	Assumption	Consequence	Improvement
x, attenuation	0.2, because unknown	• Equal attenuation in whole watershed	Better estimation [1]
k, lag	Based on static flow velocity of 2 m/s	• Equal velocity in whole watershed • Flow differs in reality (e.g. mountainous regions) • Flow independent of discharge	Better estimation [1]
Runoff coefficient	0.5, because unknown	• Equal for whole watershed • Differs from region to region in reality	Use runoff coefficient maps [2]
Average flow	long term average discharge	• No effect of seasonality included • Long term effect is disregarded	Use real-time data
Capacity	maximum long term average monthly discharge	• Assumed to be bank full discharge • Possibly incorrect proxy	Investigate locally

pact consequences are defined as the change in overflow at the same place as a result of a change in parameters. With these two definitions, x and k both have a spatial (and impact) consequence: a lower x decreases the flood wave sooner, leading to a consequence that has a shorter reach. A longer lag time, k, keeps the flood wave longer in the reach, which changes the downstream consequences.

The average flow, or base load, and capacity both only have a local effect, and can be defined as impact consequences. There is an exception, when a failure mechanism is triggered, in that case, downstream risk reduces. The same holds for the runoff coefficient. This parameter only affects the discharge quantity, not its wave shape. It is thus an impact consequence, although rain from multiple sources combined can be problematic downstream. Thus it can be argued that there is an indirect spatial consequence.

With these two effects in mind, we may conclude that impact consequences are essential when we want to know at what discharge level a failure occurs. A precise capacity is only necessary when studying specific reaches in, for example, densely populated areas. On the other hand, the spatial effects are important throughout the basin, because it is unknown beforehand what the origin of the water is.

8.3.4. Data sources

Limited data availability has resulted in several assumptions to estimate parameters, but data structure also poses some limitations to the model. Firstly: the HydroSHEDS dataset does not include bifurcations. Real riverine systems have bifurcations. This also holds for the case: in Bangladesh, the Ganges bifurcates into several tributaries. This makes the model not suitable for analysis in this region. However, bifurcations often occur when the flow velocity is low, and backwater effects or tidal influences become significant. As mentioned before the Muskingum routing method is not suitable for modelling in this type of regions.

8.4. Improvements

There are two approaches to improve model results: improve data fed into the model or improve the model itself. Suggestions for parameter improvements are included in Table 8.3. Apart from using complex estimation techniques, it is also possible to classify reaches and use a different set of parameters for each class. Example classes are: mountainous, midland and lowland reaches. Runoff coefficient maps are used in other simulations and could easily be used in this model. These maps can be generated by various remote sensing techniques. Improving the capacity parameter is least important and only necessary where floods will have large impacts. Therefore it is recommended to study capacity locally at for example bottlenecks or close to cities. To verify improvements, the results can be compared with actual flows.

Then there are model improvements; some have already been discussed in the policy options subsection (8.2.1). The failure mechanisms subsection (8.3.2) points out the most important limitation:

there is no failure mechanism included in this model. Some suggestions on how to include this are presented here. The first is using the variable time Muskingum-Cunge routing method, as described in subsection 4.2.2. Second, a parallel network can be used to model bank overflow (8.2.1). Alternatively, using storages that can collect water that overtops or when levees breach. These storages return water to the adjacent reaches when discharge levels are below the capacity.

One should keep in mind that, both with improving data as with model improvements, that the purpose of this model is high-level screening. Thus improving and adding data should only be done when this results in different model behaviour. A few per cent improvement on discharge levels does not add value if it takes much effort. For the Ganges-Brahmaputra case, it is suggested to add glacier discharge data and maybe groundwater flows and storage. Glacier discharge is a significant source of water, especially in summer, which will have a substantial effect on flood levels. For model improvement, it is suggested to implement a failure mechanism because this will change model behaviour. A complete overview and description of all improvements is given in appendix C

8.4.1. Damage estimation

Important in risk screening is the potential damage a flood can have. This damage or consequence is dependent on several factors. Most notably, the population density around and how far a flood will extend beyond the river banks. The first, as suggested earlier, is easy to include with population density maps or infrastructure maps. The latter is more challenging to estimate. Traditionally hydraulic simulation methods are used to model overbank flows. But by using such a method, all advantages are lost.

An alternative to estimate flood outside of the river is proposed. The idea is to include a cross-section profile to each river reach. Such a cross-section profile can be based on a digital elevation model, as provided in the HydroSHEDS dataset or other more accurate sources. If a cross-section is known, together with the assumed flow velocity and discharge volume, it is possible to estimate water levels. If the water level exceeds the reach capacity, the area cross-section profile can tell how far the water will flow outside of the river banks. A risk map can then be constructed that does not only include an overflow colour but also include a meaningful width. The width of the lines on the map can show the expected flood impact beyond the river boundaries. Combining these with population density and infrastructure maps can give a better risk indication.

Considering all the information so far, we will answer the fourth and last research sub-question.

To what extent can the river network model be used with data from the Ganges-Brahmaputra river basin?

The developed river network model is successfully applied to the Ganges-Brahmaputra river basin as a proof of concept. The model can, notwithstanding all assumptions and limitations, show similar flood patters as was shown in a qualitative comparison. It can indicate a high risk reaches with an overflow measure and can be useful for policy screening. There are also suggestions for improvements: the model should be improved by implementing failure mechanisms, and parameter estimation will lead to more accurate results.

9 Conclusion

This research started with the main research question: “How to use complex network theory to analyse the cascading effects in riverine systems?” A conclusion to this study can be derived by revisiting the sub-research questions. To discuss the societal and scientific relevance, we restate the research goals and reflect upon successfulness.

9.1. Revisiting research questions

How to model a river as a complex network? In Chapter 4 is shown that we can model a river as a complex network by using the existing Muskingum routing model. The Muskingum routing model is a hydrological routing model. In such a model, the routing parameters are reduced to two parameters: a lag and an attenuation parameter. These two parameters fully characterise a river reach. This is also called a lumped model because all geomorphological and hydraulic properties are lumped into two model parameters.

The Muskingum routing model is translated into a network model. A river system is divided into segments, the reaches and those reaches are linked to form any arbitrary river network structure. Each reach has the two Muskingum routing parameters, inflow from the previous reach and an external inflow. When bifurcations occur, a fraction parameter determines the division of flow discharge. The external inflow, parameters and network structure fully characterise a river system.

How to define cascading effects in riverine systems? In traditional complex networks, cascading effects are defined as successive overload failures. This can occur when a node or link in such network is removed, and loads within the network will redistribute. The resulting network can have other overloaded nodes or links in the network that will break down. This process repeats itself until a stable situation is reached. Eventually, it can cause large blackouts in, for example, electricity grids or communication systems.

Cascades in riverine systems are different. A river network mostly has a tree-like structure, which is less heterogeneous and consists mostly of confluences. An overload or overflow does not lead to a redistribution of flows within the network: water flows downwards. So in this research, the cascading effect in a riverine system is defined as floods that occur as a result of water waves that originate from upstream sources.

A simple example of such an effect is the following. Consider a Y like river structure in which there are three reaches. This river structure can be a part of a larger structure. In this system, it rains in the two upstream regions simultaneously. The rain causes high water levels in the two upper reaches, but they do not overflow. So in the two arms of the Y there is no problem. But downstream, where the two arms meet, the waves come together and result in a large flood. In this small example, we see the cascading effect of upstream rainfall occurring in a downstream reach, which results in a flood.

How to use a complex network model to analyse cascading effects of floods? The Muskingum network routing model is implemented as a simulation model in Python to study this cascading effect of flood waves in riverine systems. This model can be used to study any riverine structure. It can simulate wave patterns through the systems. If these wave patterns exceed link capacity, it is considered a flood. Because the model is graph-based and not cell-based, it is very efficient. This efficiency leads to new possibilities, such as scenario analysis and policy comparisons.

To what extent can the river network model be used with data from the Ganges-Brahmaputra river basin? In this research, the developed river network model is successfully applied to the Ganges-Brahmaputra river basin. The hardest part of development was finding suitable data. The model is used with rainfall data and can show the flow above the long term monthly maximum, defined as the overflow. It cannot show the precise tipping point when flooding will occur, but it can indicate reaches with high overflows. This overflow measure shows which reaches have a high risk of flooding. A qualitative comparison with a flooding case in 2013 shows that the model can, notwithstanding all assumptions and simplifications, show similar flood patterns.

The model also has its limitations. The main limitation is the inability to model failure mechanisms. In reality, water flows out of the river, causes damage, and is collected in outflow areas. This buffering effect reduces downstream flow and flood risk. On the other hand, flow is spread out over a long period, causing high water levels for an extended period of time. Another limitation is the inability to give a meaningful quantitative risk number, such as expected damage.

But the model has potential: It is efficient, it can simulate the whole Ganges-Brahmaputra basin within seven minutes per day without highly optimised code. Speed and simplicity make it interesting for scenario analysis. Many different parameters and settings can be used to study its effects. The results can be combined with population density maps and infrastructure maps to indicate high-risk areas and infrastructure. Moreover, because it is a simulation model and not a statistical model, the effects of for example, increased rain patterns can be taken into account. This makes it also useful in, for example, climate change studies.

How to use complex network theory to analyse the cascading effects in riverine systems? We can now answer the main research question. Key in this answer is to embed hydrologic flood routing within a network representation of a river basin. This decomposition of a river basin in a network of reaches with hydrologic properties allows for efficient calculation of water flows. By adding rainfall inflow data to the model, it is possible to simulate the resulting water waves through the system. The defined reach capacity enables to study when en where these cascading effects happen.

9.2. Societal and scientific relevance

The first research objective was to develop a flood routing model that is efficient, uses less information, is easy to understand and is useful for policy analysts and decision-makers. This objective is closely related to societal relevance. Useful was defined as suitable for risk screening. The model, as developed and applied in the case study, can show this risk. The overflow indicates risk, and it is not (yet) a quantitative risk measure such as an expected value. With the model, it is possible to reproduce behaviour in the Ganges-Brahmaputra basin, and it shows which reaches with high flow levels were most likely to flood. Moreover this result was achieved within a run time of 7 minutes per simulated day.

In its current form, the model is thus suitable to show vulnerable reaches on a map as a result of precipitation patterns. Although shown for a single case, it is easy to run the model for more historical precipitation events. By doing so, vulnerable rivers or reaches can be identified. This makes the model already useful for policy analysts and decision-makers. Usefulness can be extended further to include predicted (changing) precipitation patterns. The model can then be used as a flood forecasting tool. When using long term changing precipitation patterns, it can give future scenarios that include effects from climate change.

Although not yet tried, the model can also be used to compare policy options. Some policies require just a parameter change; others require model modifications. These policy options can be compared under several precipitation scenarios – the model than effectively becomes a tool for policy screening. However, the model should not be seen as a replacement for hydraulic modelling and simulation tools.

It lacks detail and accuracy to be used for detailed design. It should be seen as a complementary tool useful at an earlier moment in a design process and with the purpose of screening options.

This thesis is a clear example of an MSc thesis in the EPA program because of two factors. The first factor is that a model framework is developed with the aim to support informed decision-making. Advanced (geographical) data analytics and modelling simulation techniques are used to achieve this aim. The second factor is that the research in this thesis contributes to one of the EPA 'Grand Challenges': 'water'. The model and possibility for screening policy options are directly aimed at understanding and reducing flood risks in riverine systems.

The second research objective was to represent river systems using complex network theory for understanding flood routing in riverine systems. This is where a clear scientific relevance and contribution emerge. As we have shown in Section 2.3.3, there were ideas to use network theory in flood modelling. However, these models are either not used for flow modelling (Boyd et al., 1979; Lehner & Grill, 2013; Naden, 1992) or still use a lot of data (Harley, 1971). We have shown that combining flood routing theory with complex network theory is possible and yields useful insights. Firstly we have seen that in a riverine system a flood or failure is the result of cascading water waves through the system. Once such a failure occurs the downstream risk is reduced, which is an attenuating effect. This is opposite to the definition of a cascading failure in traditional context: such failure often leads to an amplifying effect. We have also shown that, with a relatively simple network model and a combination of different data sources, it is possible to reproduce system behaviour. Moreover, the model can be used to get insight in the relation between river structure and flood effects. This proves that looking from a network perspective to a riverine system can be useful.

9.3. Future work

There are two clear practical suggestions for future work, described in Section 8.4 and a more theoretical suggestion. The first suggestion is to implement failure mechanisms in the model. These failure mechanisms will change model behaviour and provide more insight into the system. The behaviour will then be closer to reality. The second suggestion for future work is to implement policy options within the model to make it ready for policy screening. A third suggestion is a more theoretical approach. Network science has indicators such as betweenness and centrality. It could be interesting to see if there is a correlation between these indicators and high-risk reaches.

Bibliography

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. doi:10.1103/RevModPhys.74.47

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., & Siebert, S. (2003). Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrological Sciences Journal*, *48*(3), 317–337. doi:10.1623/hysj.48.3.317.45290

Alexander, G. N. (1972). Effect of catchment area on flood magnitude. *Journal of Hydrology*, *16*(3), 225–240. doi:10.1016/0022-1694(72)90054-6

Ashley, W. S., Strader, S., Rosencrants, T., & Krmenec, A. J. (2014). Spatiotemporal Changes in Tornado Hazard Exposure: The Case of the Expanding Bull's-Eye Effect in Chicago, Illinois. *Weather, Climate, and Society*, *6*(2), 175–193. doi:10.1175/WCAS-D-13-00047.1

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, *38*(1), 364–374. doi:10.1103/PhysRevA.38.364

Berchum, E. C. van. (2019). *Rapid screening and evaluation of flood risk reduction strategies*.

Bérod, D. D., Singh, V. P., Devred, D., & Musy, A. (1995). A geomorphologic non-linear cascade (GNC) model for estimation of floods from small alpine watersheds. *Journal of Hydrology*, *166*(1-2), 147–170. doi:10.1016/0022-1694(94)02591-X

Bhutiani, R., Khanna, D. R., Kulkarni, D. B., & Ruhela, M. (2016). Assessment of Ganga river ecosystem at Haridwar, Uttarakhand, India with reference to water quality indices. *Applied Water Science*, *6*(2), 107–113. doi:10.1007/s13201-014-0206-6

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, *100*(5), 992–1026. doi:10.1086/261849

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175–308. doi:10.1016/j.physrep.2005.10.009

Boyd, M. J., Pilgrim, D. H., & Cordery, I. (1979). A storage routing model based on catchment geomorphology. *Journal of Hydrology*, *42*(3), 209–230. doi:10.1016/0022-1694(79)90048-9

Bozorg-Haddad, O., Abdi-Dehkordi, M., Hamedi, F., Pazoki, M., & Loáiciga, H. A. (2019). Generalized Storage Equations for Flood Routing with Nonlinear Muskingum Models. *Water Resources Management*, *33*(8), 2677–2691. doi:10.1007/s11269-019-02247-2

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., … Wiener, J. (2000). Graph structure in the Web. *Computer Networks*, *33*(1), 309–320. doi:10.1016/S1389-1286(00)00083-9

Chandramohan, T., Jose, M. K., Purandara, B. K., & Venkatesh, B. (2018). Revision of Empirical Coefficients of Commonly Used Flood Formulae Using Flow Data from Karnataka Rivers. In V. P. Singh, S. Yadav, & R. N. Yadava (Eds.), *Hydrologic Modeling* (pp. 555–563). Water Science and Technology Library. doi:10.1007/978-981-10-5801-1_39

Choudhury Parthasarathi. (2007). Multiple Inflows Muskingum Routing Model. *Journal of Hydrologic Engineering*, *12*(5), 473–481. doi:10.1061/(ASCE)1084-0699(2007)12:5(473)

Choudhury Parthasarathi, Shrivastava Rakesh Kumar, & Narulkar Sandeep M. (2002). Flood Routing in River Networks Using Equivalent Muskingum Inflow. *Journal of Hydrologic Engineering*, *7*(6), 413–419. doi:10.1061/(ASCE)1084-0699(2002)7:6(413)

Chowdhury, M. R., & Ward, N. (2004). Hydro-meteorological variability in the greater Ganges–Brahmaputra–Meghna basins. *International Journal of Climatology*, *24*(12), 1495–1508. doi:10.1002/joc.1076

Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., & Klijn, F. (2019). Accounting for the uncertain effects of hydraulic interactions in optimising embankments heights: Proof of principle for the IJssel River. *Journal of Flood Risk Management*, e12532. doi:10.1111/jfr3.12532

Cohen, J., Briand, F., & Newman, C. (1990). *Community Food Webs: Data and Theory*. Biomathematics. doi:10.1007/978-3-642-83784-5

Cunderlik, J. M., & Ouarda, T. B. M. J. (2006). Regional flood-duration–frequency modeling in the changing environment. *Journal of Hydrology*, *318*(1), 276–291. doi:10.1016/j.jhydrol.2005.06.020

Dallaire, C. O., Lehner, B., Sayre, R., & Thieme, M. (2019). A multidisciplinary framework to derive global river reach classifications at high spatial resolution. *Environmental Research Letters*, *14*(2), 024003. doi:10.1088/1748-9326/aad8e9

Dalrymple, T., & Benson, M. A. (1968). Measurement of peak discharge by the slope-area method. doi:10.3133/twri03a2

David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., & Eijkhout, V. (2011). River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology*, *12*(5), 913–934. doi:10.1175/2011JHM1345.1

Dobson, I., Carreras, B. A., Lynch, V. E., & Newman, D. E. (2007). Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *17*(2), 026103. doi:10.1063/1.2737822

Dooge, J. C. I. (1959). A general theory of the unit hydrograph. *Journal of Geophysical Research*, *64*(2), 241–256. doi:10.1029/JZ064i002p00241

Elbashir, S. T. (2011). *Flood Routing in Natural Channels Using Muskingum Methods* (Master's thesis, Dublin Insitute of Technology, Dublin).

European Commission. (2007). Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. Retrieved July 3, 2019, from http://data.europa.eu/eli/dir/2007/60/oj/eng

Expertisenetwerk Waterveiligheid. (2016). *Grondslagen voor hoogwaterbescherming.* OCLC: 1023024002. Den Haag: ministerie van Infrastructuur en Milieu.

Ghosh, S. N. (1997). *Flood Control and Drainage Engineering: Second Edition*. CRC Press.

Haldane, A. G., & May, R. M. (2011). Systemic risk in banking ecosystems. *Nature*, *469*(7330), 351–355. doi:10.1038/nature09659

Harley, B. M. (1971). *A modular distributed model of catchment dynamics.* (Doctoral dissertation, Massachusetts Institute of Technology, Cambridge).

Hines, P., Balasubramaniam, K., & Sanchez, E. C. (2009). Cascading failures in power grids. *IEEE Potentials*, *28*(5), 24–30. doi:10.1109/MPOT.2009.933498

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., & Tan, J. (2019). GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06. *Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: 18-10-2019*. doi:10.5067/GPM/IMERG/3B-HH/06

Al-Humoud, J. M., & Esen, I. I. (2006). Approximate Methods for the Estimation of Muskingum Flood Routing Parameters. *Water Resources Management*, *20*(6), 979–990. doi:10.1007/s11269-006-9018-2

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651. doi:10.1038/35036627

Karahan, H. (2012). Predicting Muskingum flood routing parameters using spreadsheets. *Computer Applications in Engineering Education*, *20*(2), 280–286. doi:10.1002/cae.20394

Knight, D., & Shamseldin, A. (2005). *River basin modelling for flood risk mitigation*. CRC Press. Retrieved from https://www.taylorfrancis.com/books/e/9780367803209

Koç, Y. (2015). *On Robustness of Power Grids* (Doctoral dissertation, TU Delft, Delft). doi:10.4233/uuid:d6ddd5fe-7280-4e10-877c-9fb90139e886

Kohn, K. W. (1999). Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. *Molecular Biology of the Cell*, *10*(8), 2703–2734. doi:10.1091/mbc.10.8.2703

Lee, D. .-.-S., Goh, K. .-.-I., Kahng, B., & Kim, D. (2004). Sandpile avalanche dynamics on scale-free networks. *Physica A: Statistical Mechanics and its Applications*. Proceedings of the Conference A Nonlinear World: The Real World, 2nd International Conference on Frontier Science, *338*(1), 84–91. doi:10.1016/j.physa.2004.02.028

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. doi:10.1002/hyp.9740

Lehner, B., Verdin, K., & Jarvis, A. (2008). New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union*, *89*(10), 93–94. doi:10.1029/2008EO100001

Mamede, G. L., Araújo, N. A. M., Schneider, C. M., Araújo, J. C. de, & Herrmann, H. J. (2012). Overspill avalanching in a dense reservoir network. *Proceedings of the National Academy of Sciences*, *109*(19), 7191–7195. doi:10.1073/pnas.1200398109

Maqbool, R. (2013). Death toll in Indian monsoon flooding nears 600. Retrieved October 21, 2019, from https://news.yahoo.com/death-toll-indian-monsoon-flooding-nears-600-142740840.html?guccounter=1

Mays, L. W. (2010). *Water resources engineering*. Tempe: John Wiley & Sons.

Motter, A. E., & Lai, Y.-C. (2002). Cascade-based attacks on complex networks. *Physical Review E*, *66*(6), 065102. doi:10.1103/PhysRevE.66.065102

Naden, P. S. (1992). Spatial variability in flood estimation for large catchments: The exploitation of channel network structure. *Hydrological Sciences Journal*, *37*(1), 53–71. doi:10.1080/02626669209492561

NASA. (2013). Severe Flooding in Northern India, Nepal. Retrieved October 21, 2019, from https://earthobservatory.nasa.gov/images/81450/severe-flooding-in-northern-india-nepal

NASA. (2019a). GPM Mission Concept. Retrieved November 21, 2019, from https://pmm.nasa.gov/GPM

NASA. (2019b). PMM Precipitation Algorithms. Retrieved November 21, 2019, from https://pmm.nasa.gov/science/precipitation-algorithms

NASA. (2019c). Precipitation Measurement Missions | An international partnership to understand precipitation and its impact on humankind. Retrieved November 21, 2019, from https://pmm.nasa.gov/precipitation-measurement-missions

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, *45*(2), 167–256. doi:10.1137/S003614450342480

Ponce, V., Lohani, A., & Scheyhing, C. (1996). Analytical verification of Muskingum-Cunge routing. *Journal of Hydrology*, *174*(3-4), 235–241. doi:10.1016/0022-1694(95)02765-3

Reuters. (2019). Over 1,600 Dead In India's Heaviest Monsoon In 25 Years: Report. Retrieved October 31, 2019, from https://www.ndtv.com/india-news/over-1-600-dead-in-indias-heaviest-monsoon-in-25-years-report-2110356

Rosas-Casals, M. (n.d.). *Topological Complexity of the Electricity Transmission Network. Implications in the Sustainability Paradigm* (Doctoral dissertation, Universitat Politècnica de Catalunya). Re-

trieved June 25, 2019, from https://www.tdx.cat/bitstream/handle/10803/5824/TMRC1de1.pdf?...1

Samuels, P. (2000). An Overview of Flood Estimation and Flood Prevention, Germany.

Shi, P.-J., Yuan, Y., Zheng, J., Wang, J.-A., Ge, Y., & Qiu, G.-Y. (2007). The effect of land use/cover change on surface runoff in Shenzhen region, China. *CATENA*. Influences of Rapid Urbanization and Industrialization on Soil Resource and Its Quality in China, *69*(1), 31–35. doi:10.1016/j.catena.2006.04.015

Shuang, Q., Zhang, M., & Yuan, Y. (2014). Node vulnerability of water distribution networks under cascading failures. *Reliability Engineering & System Safety*, *124*, 132–141. doi:10.1016/j.ress.2013.12.002

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical Modeling of Watershed Hydrology. *Journal of Hydrologic Engineering*, *7*(4), 270–292. doi:10.1061/(ASCE)1084-0699(2002)7:4(270)

Smedt, F. de, Yongbo, L., & Gebremeskel, S. (2000). Hydrologic modelling on a catchment scale using GIS and remote sensed land use information. *WIT Transactions on Ecology and the Environment*, *45*. doi:10.2495/RISK000281

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*(6825), 268. doi:10.1038/35065725

Tang, L., Jing, K., He, J., & Stanley, H. E. (2016). Complex interdependent supply chain networks: Cascading failure and robustness. *Physica A: Statistical Mechanics and its Applications*, *443*, 58–69. doi:10.1016/j.physa.2015.09.082

Todini, E. (2007). A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. *Hydrology and Earth System Sciences*, *11*(5), 1645–1659. doi:https://doi.org/10.5194/hess-11-1645-2007

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1-2), 123–138. doi:10.3354/cr00953

Wang, J., Jiang, C., & Qian, J. (2014). Robustness of Internet under targeted attack: A cascading failure perspective. *Journal of Network and Computer Applications*, *40*, 97–104. doi:10.1016/j.jnca.2013.08.007

Williams J.R., Kannan N., Wang X., Santhi C., & Arnold J. G. (2012). Evolution of the SCS Runoff Curve Number Method and Its Application to Continuous Runoff Simulation. *Journal of Hydrologic Engineering*, *17*(11), 1221–1229. doi:10.1061/(ASCE)HE.1943-5584.0000529

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180. doi:10.1038/35004572

World Bank. (2014). *Ganges Strategic Basin Assessment A Discussion of Regional Opportunities and Risks*. World Bank. Washington.

World Meteorological Organization. (2019). *WMO Statement on the state of the global climate in 2018*. WMO. World Meteorological Organization.

Yoo Chulsang, Lee Jinwook, & Lee Myungseob. (2017). Parameter Estimation of the Muskingum Channel Flood-Routing Model in Ungauged Channel Reaches. *Journal of Hydrologic Engineering*, *22*(7), 05017005. doi:10.1061/(ASCE)HE.1943-5584.0001507

Zade, M., Ray, S. S., Dutta, S., & Panigrahy, S. (2005). Analysis of runoff pattern for all major basins of India derived using remote sensing data. *Current Science*, *88*(8), 1301–1305. Retrieved November 10, 2019, from https://www.jstor.org/stable/24110301

Zinoviev, D. (2018). *Complex network analysis in Python: Recognize - construct - visualize - analyze - interpret*. The Pragmatic Programmers. OCLC: 1028629134. Raleigh, North Carolina: The Pragmatic Bookshelf.

A

Research approach

The whole research process is shown in a research flow diagram in Figure A.1. The process is divided into four phases: information gathering (1), model conceptualisation (2), model implementation (3) and synthesis (4). For each phase an approach is described in the following subsections.

A.1. Phase one: information gathering

In phase one all relevant data and literature are collected for further use in this research. Relevant literature was studied in the previous part and consists of: current ways of flood modelling, flood risk and complex network analysis. These theories and methods create the context and starting point for this research. During the course of this research there will be reflected upon these theories and methods. When necessary relevant parts will be added and irrelevant parts will be removed. Apart from these theories and methods modelling information is also necessary. This consist of practical help for performing complex network analysis as well as data for the case study. Practical help has been identified in the book “Complex Network Analysis in Python” (Zinoviev, 2018). Python is chosen as the programming language because of language experience and the availability of suitable complex network analysis packages and tools. Data necessary for the case study will be described in Phase 3.

A.2. Phase two: model conceptualisation

Phase two concerns model conceptualisation. In model conceptualisation a theoretical model is defined independent of programming language A conceptualisation is an abstract demarcation of the most important objects, processes and time-relations. It is is significant to to set boundaries for model conceptualisation. With too tight boundaries, the problem might not be solvable. With too vague boundaries the model will be come very complex. Since limited studies were found on using complex network analysis in river flood modelling were used, a step based approach is proposed.

In this step based approach we start with the most simple models that exclude time and are flow balanced, meaning the in and outflow of river reaches must be in equilibrium. In the first iterations concepts such as: precipitation, evaporation, transpiration, surface runoff and subsurface flows will be omitted. When these models are working gradually other concepts will be added. Ideas for concepts to add are, but not limited to: routing methods such as the Muskingum routing method and precipitation and surface runoff.

A.3. Phase three: model implementation

After model conceptualisation, model implementation follows. The implementation is divided into two parts: developing theoretical/small models and performing a case study on a river basin. The theoretical and small models are the building blocks or river reaches needed to construct a whole river. In the case study these building blocks are combined to form a river model.

A.3.1. Theoretical models

The development of the models consist of three steps: In the implementation step the conceptual models are translated into working computer models. As a programming language Python is chosen. After implementation the models are verified by looking if they behave as expected. These small models are then used and analysed to see if already a better understanding of river topology can be attained. Implementation and conceptualisation will not follow the waterfall approach but an incremental or evolutionary approach.

A.3.2. Case study

In the case study the small models will be combined to form a river basin. This model has to be parametrised with the data. This river basin has not yet been chosen. Section of a river basin is strongly dependent on the availability of data and data requirements. Data requirements are not yet known and will follow from model conceptualisation. After listing these requirements suitable river basins will be searched. There is a risk involved that not sufficient data can be found. In that case model parameters can either be simplified, estimated or uncertainty analysis can be performed. The ultimate goal of the case study is to validate the model. In order to reach this goal, ideally the results are compared to existing data or simulation results of other models. The availability of such data is thus also a requirement for the selection of a river basin.

A.4. Phase four: synthesis

After model implementation we take a step back and try to synthesise these results. We try to find generalised lessons in the discussion and conclusion. We will try to indicate weather a complex network analysis approach on river flood estimation is worth pursuing and extending. Recommendations for future work will therefore be given.

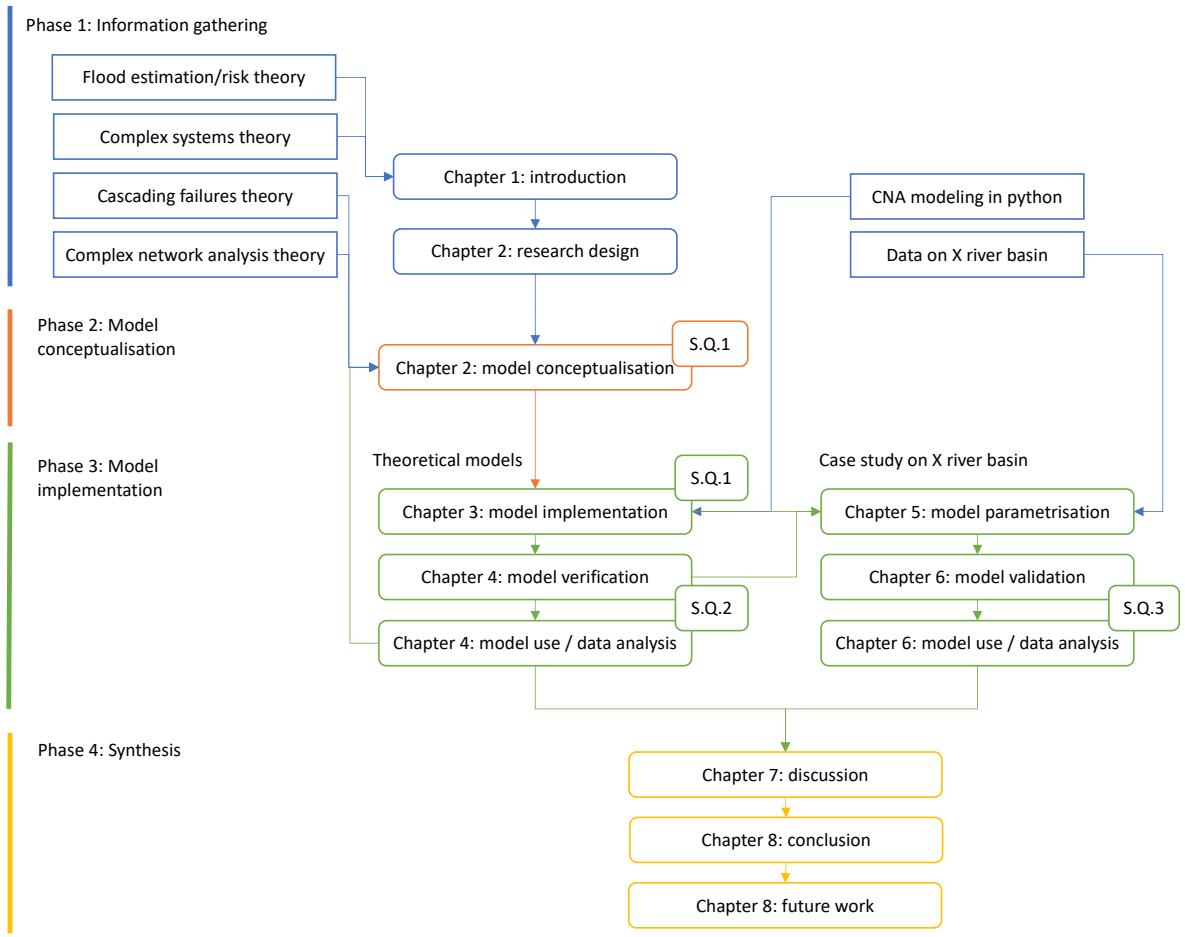


Figure A.1: Research flow diagram showing all relevant process steps.

B

Data sources and processing

B.1. HydroSHEDS creation

There are several steps involved in the creation of HydroSHEDS (Lehner et al., 2008). For full explanation see https://www.hydrosheds.org/images/inpages/HydroSHEDS_TechDoc_v1_2.pdf. Here the main steps are briefly explained, these are:

- Void filling
- Sink identification
- HydroSHEDS conditioning and burning
- Manual inspection

The basis is SRTM (Shuttle Radar Topography Mission) elevation data. Two types, SRTM-1 and DTED-2 are upscaled to 3s and combined. SRTM water body data (SWBD) is used to exclude water bodies from the digital elevation model. Because elevation alone is not a good indication for water bodies. Then void filling is applied, because SRTM data contains regions of no-data (voids). Voids occurred because of radar specific problems, such as shadow effects in mountainous areas. This causes problems for deriving hydrological products, which require continuous flow surfaces. Sinks, mainly inland natural sinks have been identified manually if deeper than 10 meters and larger than 10 km^2.

Hydrologic conditioning is applied to solve the hydrologic errors in the dataset. These steps consist of: deepening of open water surfaces, to ensure they become sinks. Weeding of coastal zones: because vegetation such as mangrove heightens the DEM, which can cause backwaters. Cells are randomly lowered to enforce breakthroughs through this vegetation. Stream burning: known existing rivers are “burned” into the DEM, which means creating gorges into the elevation surface. For HydroSHEDS only large rivers were burned into the elevation surface to prevent excessive alterations. These procedures in combination with other filtering steps and removal of barriers in rivers are repeated and manual corrections are applied. An overview of all steps can be seen in Figure B.1.

The process results in three raster file products: Hydrologically conditioned elevation, drainage directions and flow accumulations. Drainage directions maps define the direction of flow from a cell to its steepest down-slope neighbour. The river network in vector format is directly derived from the drainage direction layer and flow accumulation. The network corresponds thus to local minima in the DEM. There is only one drainage direction per cell allowed, therefore the single flow direction algorithm does not allow for river bifurcations. Other regions prone to errors are (Lehner et al., 2008):

- Areas of low or not well-defined relief
- Areas with varying vegetation cover and low-relief topography (e.g. large river floodplains)
- Low-relief coastal areas (in part due to the effect of mangroves)
- Braided rivers and deltas (the use of a single flow direction per pixel, no bifurcations)
- Large-scale roads or clearings in the vegetation cover of low-relief areas (causing artificial depressions in the elevation data)
- Elevated “barriers” in the elevation that in reality have no effect on flow connectivity (e.g. bridges)
- Areas of no-data voids in the original SRTM data

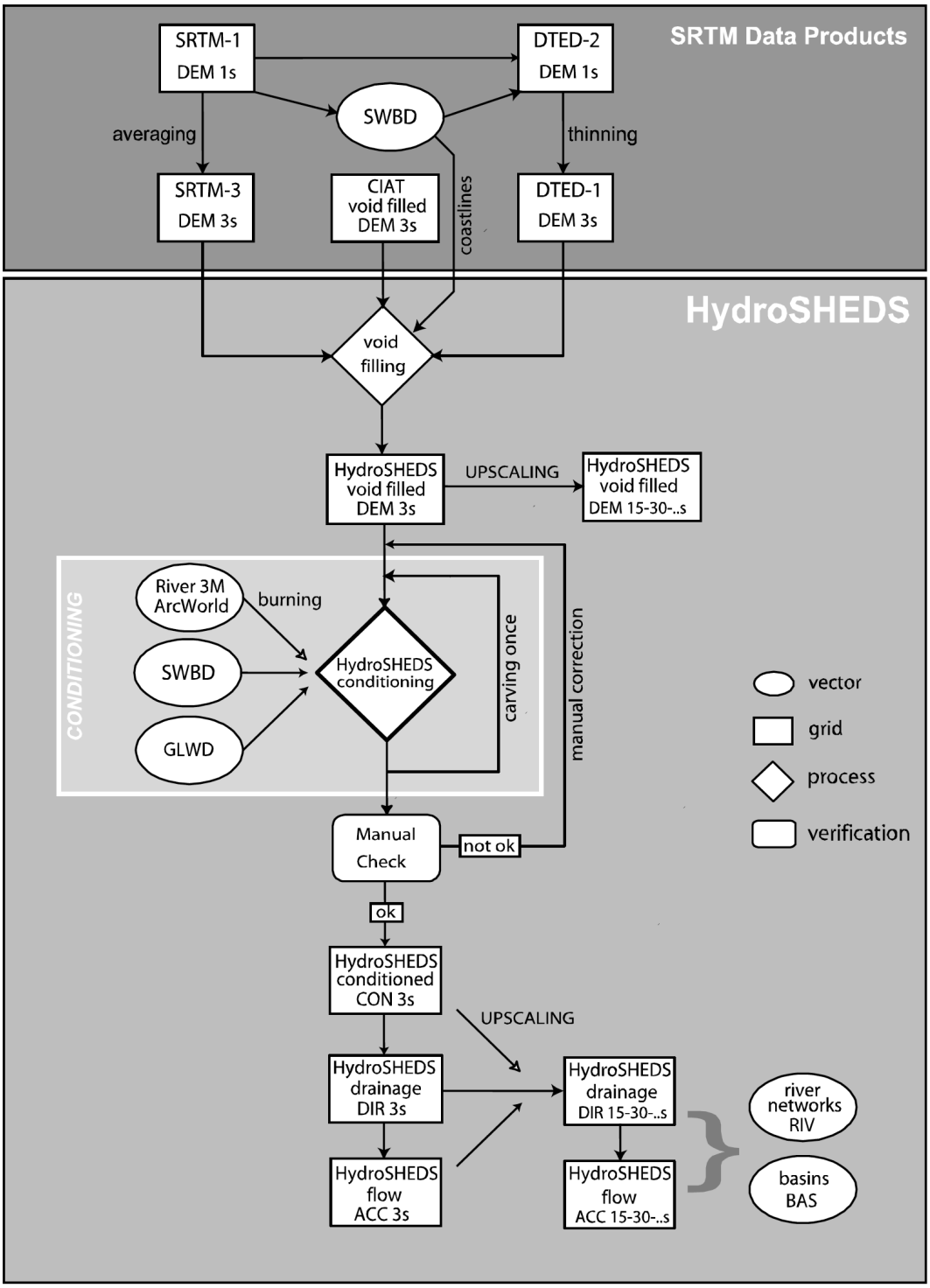


Figure B.1: Flowchart of the generation of HydroSHEDS. For full explanation see https://www.hydrosheds.org/images/inpages/HydroSHEDS_TechDoc_v1_2.pdf.

B.2. NASA precipitation measurement missions

Both TRMM and GPM use two types of sensors to measure rain: active radars and passive radiometers. Active radars such as the Dual-Frequency Precipitation Radar (DPR) transmit signals and measure the reflected signal back to the radar. The signal returned to the radar receiver (called radar reflectivity) provides a measure of the size and number of rain/snow drops at multiple vertical layers in the cloud (see the left figure in B.2).

Passive precipitation radiometers, such as the GPM Microwave Imager (GMI), measure natural thermal radiation (called brightness temperatures) from the complete observational scene including snow, rain, clouds, and the Earth's surface (see the right figure in B.2) (NASA, 2019a).

The data from these sensors and satellites is combined and processed by several algorithms to transform the radar reflectivities and brightness temperatures into precipitation information. The resulting data is verified and calibrated against Ground Validation data (NASA, 2019b). This data is called IMERG which stands for Integrated Multi-satellitE Retrievals for GPM. Older TRMM data is also converted into IMERG datasets. Validation can take weeks to several months and leads to a non-calibrated ('early') and a 'final' calibrated dataset.

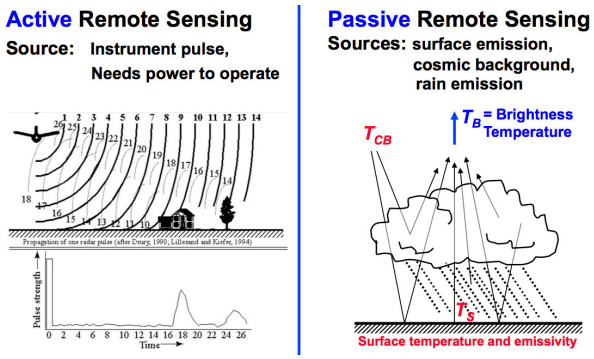


Figure B.2: Remote sensing fundamentals: showing the difference in active and passive remote sensing. Active sensors needs power to send a pulse and measure reflectivity, whereas passive sensors only receive surface emission. (NASA, 2019a). https://pmm.nasa.gov/image-gallery/active-and-passive-remote-sensing-diagram.

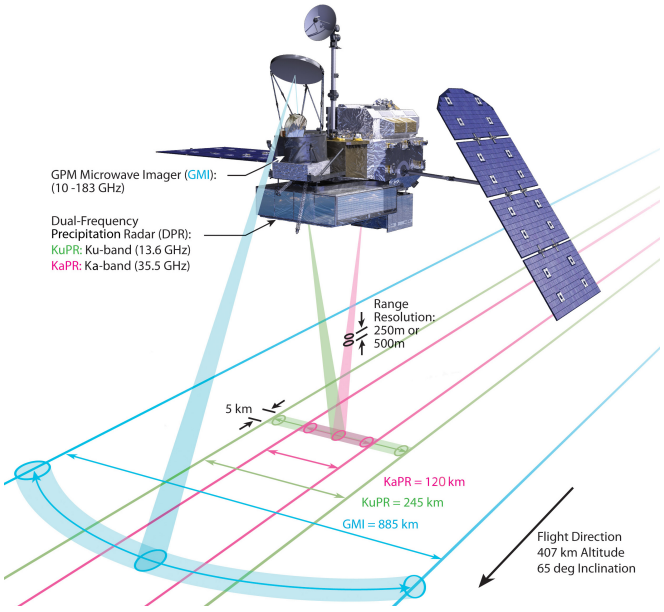


Figure B.3: Figure of the GPM core observatory showing the different measurement sensors and swath coverage. The Dual-Frequency Precipitation Radar (DPR, green and pink) is the active sensor and the GPM Microwave Imager (GMI, blue) is an passive radiometer. The DPR corresponds to the left figure and the GMI to the right figure in B.2. https://pmm.nasa.gov/image-gallery/diagram-swath-coverage-gpm-sensors. (NASA, 2019a).

B.3. Data processing

Figure B.4 shows all steps taken in data processing. The orange trapezoids are datasets and the blue rectangles are processing steps. The following enumerated list corresponds with the processing steps in Figure B.4. These steps are executed in the file `main_data.py`

1. The GloRiC / HydroSHEDS dataset is filtered for reaches only in Asia and with an average flow larger than 1 m^3/s. This reduces the size of the dataset, so that it can be loaded into memory and reaches that do not add any relevant detail are omitted.
2. A recursive searching procedure starts at a selected reach that flows into the ocean. The algorithm finds the reaches that flow into this reach. The resulting reaches are saved into a separate dataset. This dataset contains the network structure in vector format together with reach properties. Relevant properties are: reach length, average flow an flow regime variability. An example is shown in Figure 6.4
3. Before we can create the micro watersheds, the flow direction dataset is cropped. It is cropped to the boundaries of the river network. Cropping reduces file size and speeds up the operation to create micro watersheds.
4. The flow direction dataset and river network dataset are combined to create the micro watersheds. For each reach the corresponding pixels that flow into this reach are found. These pixels are added to a list. This list describes for each reach, which pixels flow into the reach.
5. To visualise the micro watersheds and to perform further calculations the pixel areas are combined into polygons. This is a spatial join operation, which is considered an expensive (slow) operation. Process 4 and 5 are shown in Figure 6.5
6. Before we continue we need rainfall data. The PMM data is a worldwide dataset in HDF5 format. This is not a standard data format that is easy to read with python GIS tools. The set is read with custom scripts and translated into default georeferenced tif files. Also this dataset is cropped with boundaries of the river network.
7. The PMM rain data has another coordinate reference system than the HydroSHEDS dataset. To ensure easy processing the dataset is resampled in the exact same coordinate reference system as the HydroSHEDS flow direction dataset. This creates temporarily large files on disk, but makes further processing much more easy.
8. Now there are two processing steps that result in the same dataset. The first option is to use the micro watershed polygons. The rasterstats packes is used to determine which pixels lie within each watershed polygon. The package returns statistics for each polygon, the average is used as the rain intensity in the micro watershed. The point in polygon problem is slow, therefore an alternative solution is proposed.
9. The alternative route is to use direct pixel matching. The coordinate reference systems of the rainfall data is exactly similar to that of the flow direction grid. This trick makes it possible to directly lookup all pixels for each watershed and compute the average. This procedure is much faster.

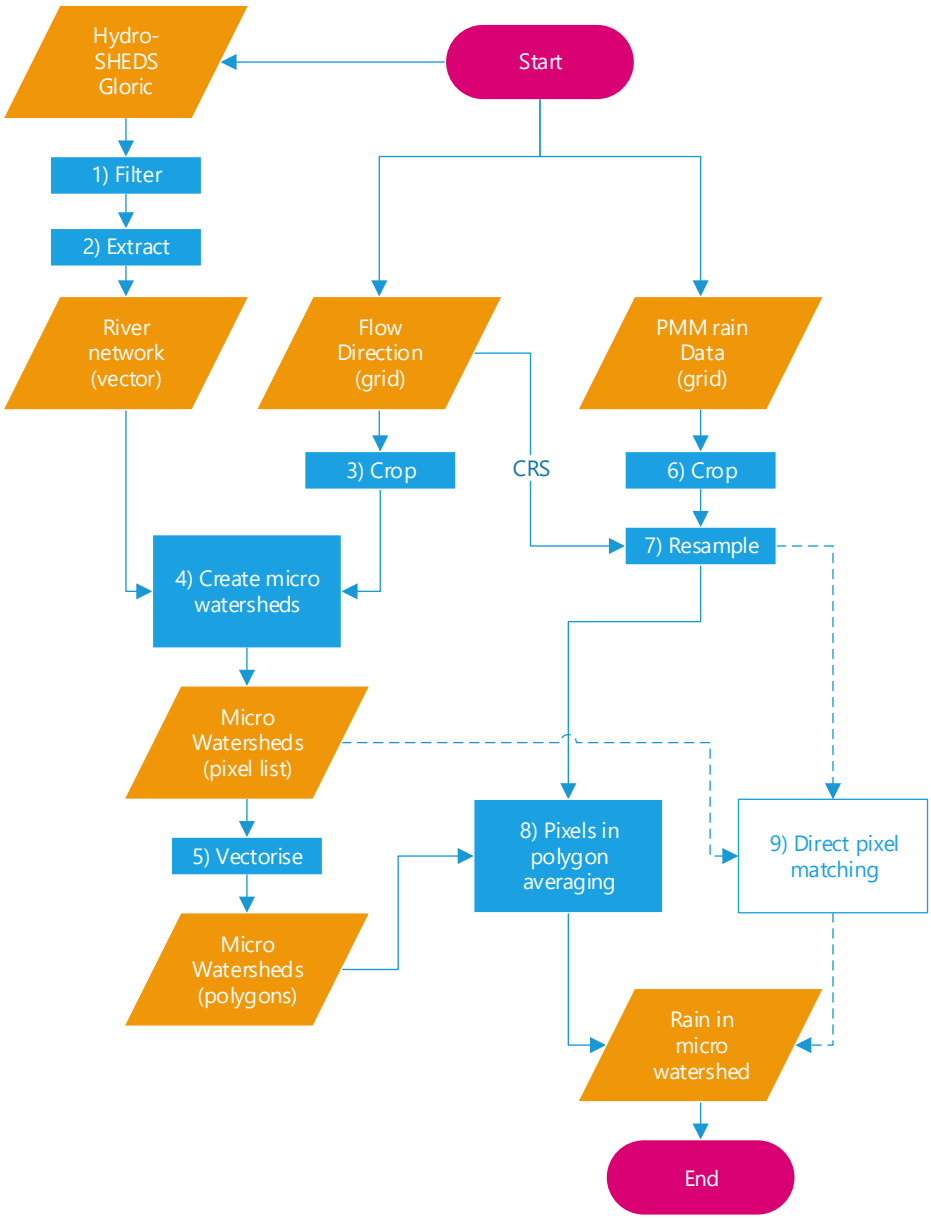


Figure B.4: Schematic showing all steps taken in data processing. Blue blocks are processing steps and orange blocks are datasets. The processing steps corresponds to the file main_data.py.

C

Reflection

The table on the following page shows suggestions for improvements and further research. The important topics have been discussed in Section 8.4

Table C.1: Full list of model improvements. The improvements are divided in categories.

Category	Improvement	Description
Data quality	Estimate x and k	Various estimation techniques can be used to get estimations for ungauged networks. These methods use optimisation techniques to minimise discharge difference at gauged reaches. See for example: David et al. (2011), Yoo Chulsang, Lee Jinwook, and Lee Myungseob (2017).
	Classify x and k	Another simple approach is to classify the reaches in a number of groups. Each class then can have the same (static) variables. This adds heterogeneity to the watershed and is important because mountainous reaches have completely different behaviour from lowland reaches.
	Runoff coefficient maps	Runoff coefficient maps are used in other runoff estimation models. These maps can easily be introduced in this model in the same way as rainfall data is mapped to reaches. See for example: Shi et al. (2007), de Smedt, Yongbo, and Gebremeskel (2000), Zade, Ray, Dutta, and Panigrahy (2005)
	Use actual and historic flow data	Actual flow or history flow data is necessary to improve the base load of the model. Currently the long term average is the level to which rainfall water is added. This can improve seasonality effects and long term effects.
	Capacity estimation	Detailed capacity estimation is only necessary at reaches where consequences can be high. For examples in reaches next to cities. There is no single method to determine capacity as this is highly dependent on river shape.
Data additions	Glacier discharge data, Ground water flows, Infiltration, Evaporation, Evotranspiration	Adding more data is possible but should only be done if the added data adds relevant behaviour to the case or scenario analysis. For example: in the Ganges-Brahmaputra case it is suggested to add glacier discharge information as this has a large effect on water levels in summer coincides with monsoon time.
Model improvements	Failure mechanism: parallel network	A parallel network can be used to model overflow on the river banks. The parallel network can have different Muskingum parameters to account for larger storage effect and longer lag time. Water levels in the main river channel and river banks should are equal, as long as the main river channel discharge exceeds its capacity. If water level lowers the parallel reach will feed back water to the main channel.
	Failure mechanism: storages	The concept of adding storages is similar to a parallel network but simpler. Instead of a complete parallel network storage functions are assigned to each reach. When the reach overflows, which can be overtopping or a levee breach, water from the reach is stored in this function. When river discharge decreases again water is returned back to the network.
	Muskingum-Cunge	Muskingum-Cunge can be used to more accurately determine the Muskingum parameters based on river geometries.
	Variable time Muskingum-Cunge	Variable time Muskingum methods can be used to take into account the effects of changing flow velocity on the lag time and attenuation. With some more adjustments this can also account for overflow situations. In such situation the flow velocity is reduced and storage effect increased.
	Model speed-up	The model has a runtime of 7 minutes without code optimisations. Runtime can be improved by rewriting core components with a package like Cython or Numba. An improvement of a factor 100 can be expected.