Broad-scale flood modelling in the cloud: validation and sensitivities from hazard to impact

Fergus McClean

A thesis submitted to Newcastle University in partial fulfilment of the requirements for the degree of Doctor of Philosophy within the Faculty of Science, Agriculture and Engineering

JUNE 2019

Acknowledgements

I would like to sincerely thank both Prof. Richard Dawson and Prof. Chris Kilsby for their excellent support, advice and motivation throughout my work on this thesis.

Thanks to Vassilis Glenis for help with modelling, Miles Clement for providing Sentinel 1 data, Microsoft for the cloud computing time, and NERC for funding the research through the Data, Risk and Environmental Analytical Methods (DREAM) training centre.

I would like to thank my friends in the Cassie building for the entertaining lunchtimes and everyone who I have worked with throughout my studies.

Finally, thank you to my family for their encouragement and Kay for her patience.

Abstract

Broad-scale flood modelling is a growing research area with applications in insurance, adaption and response. This has been fuelled by increasing availability of continental-global datasets providing inputs to a mounting array of models. However, outputs vary greatly and validation is challenging.

This research developed a novel, consistent methodology for assigning performance scores to models using a range of gridded datasets and an accurate numerical 2D hydrodynamic modelling system. Validation using both extent and discharge was conducted for Storm Desmond in Northern England and the global applicability of the methodology demonstrated across Europe and in Indonesia. To meet computational demands, a cloud computing framework was implemented using a PostgreSQL database. Visualisation of results was achieved using a newly designed web interface. Finally OpenStreetMap data was overlaid to demonstrate the sensitivity of impacts to flood model inputs.

The main findings are that relative importance of precipitation and topographic data changes depending on the metrics used for validation. More variability in peak discharge error was found between models using different rainfall inputs (22-70%) than different DEMs (9-37%). Conversely, flood extent critical success index (CSI) was more sensitive to the choice of topography (25-32%) than rainfall (27-30%), though overall variability in CSI was low. This was echoed in the impacts analysis with higher sensitivity of feature inundation to topography than rainfall. Importantly, there was far more overall variability in discharge accuracy than extent which indicates that reproduction of peak discharge is a more powerful measure for assessing model performance. Models driven by global-continental precipitation products underestimated peaks more than those using Met Office rain gauge data, though better performance was demonstrated by replacing ERA-Interim with the updated ERA5 dataset.

The research highlights a growing need for more robust validation of broad scale flood simulations, and the difficulties this presents. Strong influence of dataset choice on infrastructure inundation has consequences for insurance premiums, development planning and adaptation to climate change risks which should not be ignored.

Acronyms

1DM: 1 Dimensional Model
20CRv2: 20th Century Reanalysis Version 2
2DM: 2 Dimensional Model
AVHRR: Advanced Very High Resolution Radiometer
AW3D: Advanced Land Observing Satellite World in 3D
BAG: Basisregistratie Adressen en Gebouwen
CaMa-Flood: Catchment-based Macro-scale Floodplain
CFSR: Climate Forecast System Re-analysis
CityCAT: City Catchment Analysis Tool
CORINE: Coordination of Information on the Environment
CRU: Climate Research Unit
DEM: Digital Elevation Model
DFO: Dartmouth Flood Observatory
DIVAST: Depth Integrated Velocity and Solute Transport
DLR: National Aeronautics and Space Research Centre of the Federal Republic of Germany
ECMWF: European Centre for Medium Range Weather Forecasting
ERA: European Re-Analysis
ERS: European Remote Sensing
FLOPROS: Global Database of FLOod PROtection Standards
GFM: Global Flood Model
GHM: Global Hydrological Model
GRASS: Geographic Resources Analysis Support System
GRD: Global Rivers Dataset (Andreadis et al., 2013)
GRDC: Global Runoff Data Centre
HBV: Hydrologiska Byråns Vattenbalansavdelning
HDM: Hydro-dynamic Model
HEC-RAS: Hydrologic Engineering Center's River Analysis System
HOWAS: HOchWAsserSchäden (Flood Damage) Database
HPC: High Performance Computing
HYDE: History Database of the Global Environment
IGN: Institut Géographique National
IMAGE: Integrated Model to Assess the Greenhouse Effect
ISCCP: International Satellite Cloud Climatology Project
JRA: Japanese Meteorological Agency (JMA) Re-Analysis
JRC: Joint Research Council
KGE: Kling Gupta Efficiency
Mac-PDM: Macro-scale—Probability-Distributed Moisture model
MAF: Mean Annual Flood
MERIT: Multi-Error-Removed Improved-Terrain
MODIS: Moderate Resolution Imaging Spectroradiometer
NCEP: National Centers for Environmental Prediction
NSE: Nash Sutcliffe Efficiency
OSM: Open Street Map
PCRGLOBWB: PCRaster GLOBal Water Balance
postgres: PostgreSQL Database
RDMS: Relational Database Management System
RFFA: Regional Flood Frequency Analysis
RMSE: Root Mean Square Error
SAR: Synthetic Aperture Radar
SIPSON: Simulation of Interaction between Pipe flow and Surface Overland flow in Networks

SRTM: Shuttle Radar Topography Mission
SSM/I: Special Sensor Microwave Imager
SWEs: Shallow Water Equations
UIM: Urban Inundation Model
UNET: Unsteady NETwork
USGS: United States Geological Survey
VGI: Volunteered Geographic Information
VU: Vrije Universiteit Amsterdam

Contents

Figures

Tables

Chapter 1. Introduction

1.1 Preface

Flood risk management requires understanding of where flooding is likely to occur, to enable targeted mitigation (Plate, 2002). However, there are currently fundamental difficulties with producing useful broad-scale information about floods (Ward et al., 2015). A range of user groups depend on this information to make decisions. For example, insurers and reinsurers rely on risk maps to set policy premiums (Michel-Kerjan & Kunreuther, 2011). City and transport planners require knowledge of flood hazard to enable resilient and sustainable development (Jabareen, 2013). Designing effective adaption measures including flood defences and mitigation measures like green infrastructure, depends on reliable depth estimates (White, 2008). Depth is also important when deciding if a car or train will be able to traverse a road or railway (Pregnolato et al., 2017). Velocity becomes relevant when deciding on evacuation routes for people during an event (Bernardini et al., 2017). Regardless of the application, the main issues surrounding flood modelling are concerned with data availability, resolution and uncertainty (Di Baldassarre & Uhlenbrook, 2012). Data is required to simulate floods, validate these simulations and estimate what is inundated, by what depth, at what velocity, for how long.

Flood risk modelling is often carried out at a local or city scale with simulations calibrated to a unique area (Pappenberger et al., 2007). This is useful for understanding impacts in that area, however flood events often impact large regions and even entire countries (Pant et al., 2018). The potentially widespread nature of flooding has consequences for reinsurance companies, who may incur significant costs during such large events (Insurance Information Institute, 2018). The reinsurance industry are therefore one example of a group with strong interests in the probability and magnitude of broad-scale flooding. Connected infrastructure and supply

chains mean impacts can extend far beyond the hazard zone itself (Haraguchi & Lall, 2014). Therefore, it makes sense to assess flood risk not only locally but also at the regional to national scale. However, such modelling poses many challenges, including data quality, validation and computational expense. Previous studies have already demonstrated the potential for producing flood hazard maps with near-global coverage (Ward et al., 2013; Winsemius et al., 2013; Sampson et al., 2015; Dottori et al., 2016), though validation issues have caused scepticism as to their reliability (Trigg et al., 2016; Ward et al., 2015). Furthermore, the ability of currently available datasets to create well validated simulations of historical widespread flood events, let alone forecast into the future, is still not fully understood.

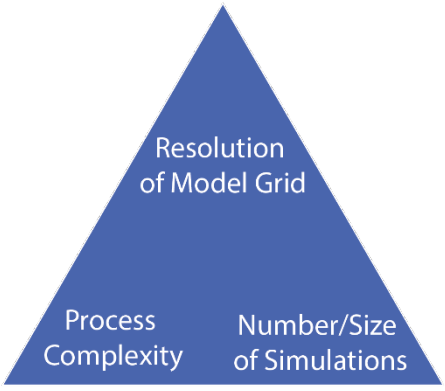

Figure 1-1 – Computational trade-off between model resolution, process complexity and number or size of simulations.

In terms of computational requirements, there are three primary dimensions which affect the capability of a broad-scale flood modelling system. These are summarised in Figure 1-1. To increase either resolution or the number and size of simulations, process complexity must be compromised to maintain computational expense at a constant amount. Of course, increasing computing power or allowing more time to run simulations would allow the process representation to remain fully realistic, however these are often not acceptable options. Therefore many studies have understandably simplified the hydrodynamics to account for the

size and resolution of the domain (Bates et al., 2010). While this allows faster simulations, it can compromise accuracy and makes validation even more vital.

While broad-scale flood modelling studies do generally carry out validation, the approaches used can vary greatly in how comprehensively and robustly they assess performance (Wing et al., 2017; Kutija et al., 2014; Zischg et al., 2017; Tyrna et al., 2016). Therefore, more work is needed to standardise and improve the available methods. One of the reasons why flood model outputs can be difficult to validate at broad-scale is that they often represent a return period event or set of events and not a historical flood (Paprotny et al., 2017; Pappenberger et al., 2012; Sampson et al., 2015). This tends to be because the methodology is specifically designed for simulating return period floods. For example, if a method involves estimating extreme river flows based on the size of the catchment, this does not directly relate to a past event. Return period event analysis is necessary for decision-making about development of both infrastructure and flood mitigation measures (Volpi et al., 2015). However, as return periods are idealisations of reality, comparing them to any given real world event or set of events does not make sense.

Nevertheless, validation of model outputs for return period events remains common in the literature. This leads to benchmarking approaches that involve comparing results to other model outputs which use similar data (Wing et al., 2017) or asking local experts to assess them based on their knowledge of the area (Tyrna et al., 2016). These “verified” outputs are then frequently used to assess damage or impacts using exposure data, which, itself, is subject to uncertainty (Wing et al., 2018). However, without standardised and comprehensive validation procedures, it is difficult to make conclusions about the relative accuracy of risk estimates from different studies.

There can be motivation to over-sell a particular model or dataset by using certain metrics and not others or presenting results from certain areas and not others. This is another attraction of model inter-comparison approaches, as two model outputs driven by the same or similar datasets will almost always be more similar than a model output and a set of observations. The reliance on inter-model comparison is understandable given limited availability of high quality validation data and added work requirements to model actual events for validation as well as return period events for planning purposes. However, just because validation using real events and observed data is difficult does not mean it should not be attempted. Without robust validation, decision makers will be unaware of the reliability of the information they are being provided. Robust validation is therefore a critical aspect of any flood modelling study and particularly broad-scale studies

1.2 Aims and Objectives

The main aim of this thesis is to better understand the potential and limitations of modelling floods at broad scales using non-localised data and parameters. Along with this primary aim, the thesis seeks to provide an transparent and transferable approach to validation which may be utilised in future studies and demonstrate the potential provided by on demand cloud computing for flood simulation. The major objectives of the research have been designed in relation to these aims and are outlined below.

1. **Create an automated system to deploy a hydrodynamic model on the cloud (Ch. 3)**

Often, simplifications are made to modelling systems to enable them to simulate large areas. However, this thesis uses a fully hydrodynamic 2D modelling system to remove the uncertainty which would be associated with these simplifications. This uncompromising approach to hydrodynamics means each simulation is highly computationally intensive. This thesis therefore presents an approach to resolve the problem of increased computational

demand by distributing and automating the workload using cloud computing, provided by Microsoft Azure. A highly supported and feature-rich Relational Database Management System (RDMS), PostgreSQL (postgres) is used for storage and some processing of the vast amounts of geospatial data involved.

2. **Validate the modelling system in an area with reliable observations for a real event (Ch. 4)**

This thesis presents a robust validation methodology for quantification of broad-scale flood model performance. To assess the performance of the automated modelling system, a real-world event across Northern England is simulated and metrics calculated based on both observed extent and discharge measurements. Often, either discharge or extent is assessed in isolation which does not provide a complete picture of performance, therefore using both in a consistent way is important to prevent an unrealistic view of accuracy. Using a historical event simulation is also important as return periods are not good at capturing spatial variations in precipitation fields.

3. **Assess the sensitivity of flood flows and extents to inputs and parameters (Ch. 5)**

The automated approach to model setup allows direct comparison of model outputs using a range of datasets and parameters. This thesis presents a comprehensive representation of model performance by focussing not only on a single combination of input datasets and parameters, but looking across datasets to create a better picture of how dataset and parameter choices can influence performance. Comparing multiple models using the same inputs, as seen in previous studies, does allow better understanding of the effects of different methodologies. However, this does not provide a full picture as uncertainty is often primarily derived from the underlying data rather than the set of equations used for simulations.

4. **Present applications and use cases for flood model outputs (Ch. 6)**

Using a fully hydrodynamic 2D modelling system at broad-scale means widespread inundation of assets can be estimated over time and not just for the maximum flood extent. OpenStreetMap data is used to demonstrate the potential of this approach by calculating the magnitude of road and building inundation over time. To understand how the variability in model outputs caused by dataset selection can propagate to the analysis of impacts, the effect of changing precipitation and DEM inputs on estimated infrastructure inundation over time is assessed. The international transferability of the methodology is demonstrated across Europe and in Indonesia.

1.3 Thesis Structure

The thesis has been structured in accordance with the objectives outlined above, with objectives 1-4 corresponding to Chapters 3-6 respectively. Following this introduction, the literature on large scale flood modelling is reviewed with a specific focus on validation methods and risk calculation. Then, Chapter 3 describes the modelling framework for the following analysis in Chapter 4, Chapter 5, and Chapter 6. The framework includes novel software for deploying simulations to the Microsoft Azure Cloud. Cloud computing is required to meet the computational demands of the modelling system as it does not simplify any of the fluid dynamics involved in flood propagation. This approach is different to many broad-scale studies which aim to simplify the shallow water equations to improve simulation speeds, using approximations such as inertial or diffusive formulations. Chapter 4 explores validation of the modelling system in Northern England using a real storm event which caused widespread damage to property and disruption of infrastructure networks. This initial validation uses a single set of default parameters and datasets to characterise performance and develop the methodology. Chapter 5 progresses to compare results from models using a range of common input datasets and assess the sensitivity of the modelling framework to data sources and pre-

processing parameters. This approach is highly novel as broad-scale flood risk studies usually use a single combination of datasets and parameters to produce outputs. Chapter 6 provides some sample applications of the modelling framework for estimating inundation of roads and buildings over time. The findings and implications of the research are discussed in Chapter 7.

Chapter 2. Literature Review

The literature surrounding the implementation and use of broad-scale flood modelling systems to estimate flood risk has been analysed and is summarised here in three sections. Firstly, the main aspects of hydrological and hydrodynamic modelling at scale are discussed. This section focuses on methodological approaches and sensitivities to parameters and data inputs. Following this is a summary of the various approaches to model validation that have been applied in the literature. Validation is critical when modelling floods and inconsistency between approaches used in different studies has led to difficulties in comparing their performance. Finally, the use of flood model outputs in quantifying risk to infrastructure both now and into the future is addressed. Uncertainty in flood model outputs can be magnified when used to estimate risk across large areas, therefore clarity in methodology is of great importance here. This summary provides context for the research in the following chapters and highlights the key issues that are addressed later.

2.1 Large Scale Flood Models

2.1.1 Introduction

“All models are wrong but some are useful” (Box, 1979).

In the 21st Century, as computing power has expanded rapidly, global flood models (GFMs) have started to become a serious option for assessing risk and potential consequences of future events around the world. Their use ranges from insurance to preparedness and emergency response and they can also be combined with climate models to simulate future scenarios (Pappenberger et al., 2012). A number of approaches have been taken to modelling floods at regional to continental scales, with varying levels of success. These models have looked predominantly at fluvial flooding but some also consider pluvial and coastal risk. As the insurance industry has such a high demand for model outputs, environmental consultancies

have also begun to produce GFMs for market, though their methodology is often unpublished and it is not generally possible to verify commercial outputs due to the costs involved in purchasing them.

2.1.2 Reproducibility

A fundamental problem with many complex GFMs which include long workflows is that the study is often not repeatable by another researcher using only the information contained within the published paper. Even if all the information is present, differing results may be found due to mistakes in code (Hutton et al., 2016). Exacerbating these problems, some data used is also private and only provided on a commercial basis. These challenges could be overcome by including all of the input data along with all of the code or processing steps. However this can be impractical if the code and/or data is not open source. There can also be an element of code shyness whereby modellers do not want to display their messy and potentially imperfect code to the world. Even if all data and code is provided, setup can be time consuming and require expensive computational resources and there is no incentive for other researchers to test them.

2.1.3 Extreme Discharge Estimation Methods

To create floods, an input of water is required onto a 2D surface, either indirectly as rainfall or directly as runoff. To create this input of water, most major studies estimate extreme return period river discharge. The discharge estimation process can vary greatly. For example, research groups at Bristol University and VU Amsterdam use contrasting approaches. Ward et al. (2013) of VU use the global hydrological model PCR-GLOBWB which has more recently been adopted by Muis et al. (2015). Climate re-analysis data is used to force PCR-GLOBWB and output flood volumes. These volumes are then down-scaled and input to a hydrodynamic inundation model (HDM). In contrast, Sampson et al. (2015) use Regional Flood Frequency

Analysis (RFFA), applied to gauged data from the Global Run-off Data Centre (GRDC), to estimate return period flood magnitudes before passing these to a similar HDM. There are both benefits and limitations of each method.

RFFA is fast and based on real data but stricken by a lack of densely and evenly distributed gauging stations. Smith et al (2015) found the method to perform better in wetter, more temperate regions. Catchment area was found to be the most successful predictor of Mean Annual Flood (MAF), which may present problems when looking at smaller and wetter catchments. Hydrological models such as PCR-GLOBWEB are more physically based and therefore able to simulate future scenarios using rainfall from climate models rather than requiring discharge forecasts. They can also be validated against observed discharge records and provide time series outputs rather than only peak discharge. However, hydrological models are more computationally expensive and require rainfall data which in itself introduces further uncertainty. Both approaches will now be discussed in more detail.

2.1.4 Global Hydrological Models

Global hydrological models (GHMs) attempt to incorporate physical processes including infiltration, interception and evapotranspiration to convert rainfall into runoff. LISFLOOD, developed by the European Joint Research Centre, is a commonly used example, created for large and trans-national catchments. The schematic in Figure 2-1 summarises how rainfall is routed into river channels via both the surface and subsurface. Along with elevation and rainfall, GHMs require data about soil, land use and temperature which may not always be available. Due to the range of processes being simulated, depending on the time step and number of cells, GHMs can be computationally expensive. When rainfall volumes are low, during dry seasons, these processes have a large influence on channel discharge. However, during large precipitation events, the land becomes saturated which limits infiltration,

temperatures are often low and rapid input of water to the system also makes evaporation a minimal factor. Therefore, during large floods, many processes may not need to be simulated, as nearly all rainfall instantly becomes surface runoff. However, many studies still simulate runoff in this way (Alfieri et al., 2013; Timonina et al., 2015; Dottori et al., 2017).

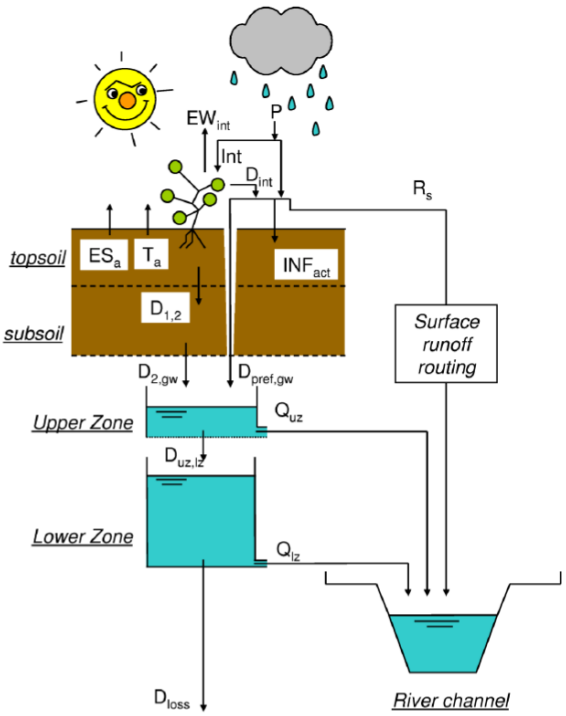

Figure 2-1 - LISFLOOD Schematic (Burek et al., 2013).

Depending on the land use and antecedent conditions, infiltration may influence flood impact but this requires additional data which is not always available. Including an infiltration component may also increase computational demand unnecessarily. GHMs generally require calibration to achieve reasonable performance which is not possible for the majority of earth's surface due to a lack of recorded flow data. Efforts have been made to estimate calibration parameters from gauged donor catchments with similar characteristics using the Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Beck et al., 2016). However, for the selected catchments, an improvement of only 0.01 median daily NSE from -0.02 to -0.03 was observed from spatially uniform to regionalized parameters. As both values are below zero, the mean discharge would have been more accurate than the simulation, and the regionalisation

process appears not to have significantly improved performance. More recently, Seibert et al (2018) demonstrated some better improvements on NSE by using regionalised parameters, from 0.67 to 0.72 using the SHETRAN hydrological model. However, this study used high quality local data and still required nearby gauged donor catchments to be present for regionalisation.

2.1.5 Regionalisation

In contrast to the physically based GHM method, regionalisation is entirely statistical and attempts to estimate extreme discharge for ungauged rivers based on their catchment characteristics and proximity to existing gauges. This is the basis for flood risk estimation in the UK, using the Flood Estimation Handbook statistical method (Robson & Reed, 1999). Estimating extreme return period discharge in this way is a well-established method. The relative contribution of catchment characteristics and proximity to other gauges can vary though. Merz and Blöschl (2005) found spatial proximity to be a better predictor than other attributes in 575 Austrian catchments. At large scales, this spatial proximity to good data can be limited which may negatively impact the performance of regionalisation.

A common element of regionalisation is estimation of the index flood or average annual maximum discharge. Smith et al (2015) tested index flood estimation on a global database of discharge and found median errors for the 100 year flood of 56%, with far larger errors in many areas. They found that adding more variables than catchment area caused minimal improvements. Using only area, the errors in temperate and continental regions were 107% and 164% respectively. The errors in median discharge were lower, though the mean and not median was then used to scale regional growth curves.

As discharge is calculated using a purely statistical method, it is often difficult to understand the source of such large error margins. Padi et al (2011) developed regional envelope curves

for the African continent based solely on catchment area with calibrated parameters for different biomes, however no overall performance measures were calculated. Lehner et al (2006) tested their global integrated water model in 21 European basins using regionalisation and found that performance decreased at longer return periods and in smaller catchments. Aziz et al (2014) applied a regionalisation method involving artificial neural networks in Australia and found the best predictor variables to be area and design rainfall intensity. Mean relative error for the 100 year flood across 90 stations was 44%. This high error margin was found in a relatively data rich area, demonstrating the inherent limitations of regionalisation.

2.1.6 Hydrodynamic Models

Hydrodynamic models (HDMs) differ from GHMs in that they ignore infiltration and evaporation processes in favour of a momentum based approach, often treating the land surface as impermeable. In the context of large floods, this is a reasonable assumption to make as storm rainfall rate usually exceeds the rate of infiltration, meaning that during extreme events, this effect is not necessarily significant. The Shallow Water Equations (SWEs), derived from the equations for conservations of mass and linear momentum, effectively represent the movement of water over a surface under gravity and can be used to route surface runoff during floods. Across hydraulic jumps, where elevation changes abruptly, the SWEs no longer hold and a separate method must be applied. In HDMs, there is always a trade-off between resolution, accuracy and computational expense.

There are various approximations of the SWEs representing different levels of complexity. In general, the more complex the formulation, the more computationally expensive the simulation is to run but the more accurate the results. However, a benchmarking study by the Environment Agency found that less complex approximations were not always faster (Neelz &

Pender, 2010). Accuracy also depends on data quality and often quality of data is low and a more complex scheme would not improve results.

Some examples of simplified schemes used in HDMs are the diffusive wave approximation of the SWEs (Hunter et al., 2005), the inertial formulation of the SWEs (Bates et al., 2010) and the kinematic wave model (Hoch et al., 2016). Hoch et al (2016) found the kinematic wave equation to significantly underperform, compared to solving the full SWEs, suggesting a lack of backwater effects as a key cause. This trade-off in accuracy is normal when making simplifications to improve computational performance.

The big advantage of HDMs is that they often outperform GHMs in terms of timing accuracy, however this accuracy can be compromised by simplifications of the physics involved. Fewtrell et al (2011) compared the inertial and diffusive formulations of the SWEs and found depth estimates were similar, however velocities were more varied at 1m resolution. These errors in velocity can cause increasing timing errors at larger scales. Tyrna et al (2016) suggest that very high resolution elevation data is more important than full representation of hydrodynamics, though they did not calculate or compare any performance metrics with a full SWE simulation.

2.1.7 Discretization

The SWEs can be solved using finite difference (FD), volume (FV) or elements (FE) methods. FE allows modelling of any geometric shape but is more complex and has less physical significance. FV and FD allow fluxes with more physical significance. Horritt et al (2007) demonstrated better performance from a FV method but greater sensitivity to Manning's n when compared to finite elements against a remotely sensed extent derived from SAR imagery. CityCAT (City Catchment Analysis Tool) is an example of a FV model and LISFLOOD-FP of finite-difference. FD does not conserve mass over the entire domain and FE does not

conserve mass over each element or node (Bradford & Sanders, 2002). Additionally neither can deal effectively with flow discontinuities. FV solves these issues by using a Riemann solver at flux boundaries to conserve mass and momentum. Variable grids have been used in HDMs to reduce computational demand by creating larger cells in areas of lower topographical variability or in regions where accuracy is not required. However, this introduces a time step choice problem as water takes longer to travel across some cells than others. If this is not accounted for properly, it can lead to breaching of the Courant–Friedrichs–Lewy (CFL) constraint which essentially states that water cannot travel more than one cell in one time step.

2.1.8 1D vs 2D

Another choice to be made in hydrodynamic modelling is whether to use 1D, 2D or a combination of approaches. 1D methods are less computationally expensive but are sometimes less accurate. 2D formulations can be highly computationally demanding but usually simulate flood wave propagation more effectively. Horritt and Bates (2002) tested the ability of 1D models HEC-RAS and LISFLOOD-FP against TELEMAC-2D to simulate inundation area. The models were calibrated against inundation area and discharge. LISFLOOD was found to perform worse in all tests, with HEC-RAS and TELEMAC producing similar performance statistics. Leandro et al (2009) compared SIPSON (1D) with UIM (2D) and found the 2D model to better represent flow over terrain, however the 1D formulation did produce good results within channels. Lin et al (2006) integrated a 2D model, DIVAST with the 1D ISIS and found the 1D only simulation to significantly underestimate flood depths compared to the 2D model.

In attempts to overcome the computational expense of fully 2D models, they have been used to calibrate faster 1D versions (Leandro et al., 2011). This is a reasonable solution to the lack of calibration data, however performance outside of the calibrated area and time period

cannot be guaranteed. For unseen areas, a 2D model will always be required to calibrate the 1D version for different rainfall events which means the reduction in computational expense will be supressed. Lhomme et al (2006) find a 1D model to perform comparably with a 2D model in the steeper streets of Nimes, France, however in flatter areas, particularly cross-roads, the outputs were significantly different. Vojinovic and Tutulic (2009) compare 1D only and 1D-2D versions of MIKE in St Maarten, The Netherlands. They highlight the increased computational, software licencing and expert time expenses of using 2D models and increased complications and stability issues when the models are coupled explicitly. These concerns can be a problem but depend on which modelling software is used.

2.1.9 Digital Elevation Models

Digital Elevation Models (DEM) are required by all flood models. However, a vertically accurate and high resolution DEM is not available in many areas, limiting model performance. Therefore various studies have attempted to improve DEM quality by applying correction algorithms which seek to remove effects such as noise from radar back scatter.

The Shuttle Radar Topography Mission (SRTM) dataset (Farr et al., 2007) is currently one of the best openly available DEMs to offer near-global coverage and is widely used in broad-scale flood modelling (Sampson et al., 2015, p.4; Domeneghetti, 2016, p.2904). The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM also provides near-global coverage, however has been shown to contain more surface artefacts and noise than SRTM (Rexer & Hirt, 2014; Jarihani et al., 2015). These surface artefacts cause routing problems for flood simulation and therefore SRTM is more widely used for this purpose (Dottori et al., 2016; Winsemius et al., 2013; Pappenberger et al., 2012; Yamazaki et al., 2011; Ward et al., 2013; Sampson et al., 2015). Limitations of SRTM have been assessed in studies

discussed in the following paragraphs and improvements made using a range of post-processing algorithms.

Falorni et al (2005) highlight the strong effect of relief on SRTM vertical accuracy and tested a wavelet filter to remove speckling. The filter was found to be effective at removing vertical noise, however did not alter the mean elevation in the Little Washita Basin which remained 3.4m below the USGS benchmark. This study demonstrates the need for pre-processing of raw SRTM data before it can be used for flood modelling but was limited to two small basins.

Baugh et al (2013) found an improved RMSE of water elevations and high and low water extent after applying a vegetation removal procedure to SRTM and using the corrected DEM in a hydrodynamic model. Vegetation removal is clearly important, however the study focussed on an area with widespread dense and tall vegetation where effects on results will be greatest. These densely vegetated areas are often not the regions of most interest when analysing flood risk as they are usually not highly populated.

The HydroSHEDS project (Lehner et al., 2013) applied a number of corrections to the original SRTM data specifically for hydrological modelling. HydroSHEDS is described in more detail in Section 5.2.1.2. Yamazaki et al (2017) present the Multi-Error Removed Improved Terrain (MERIT) DEM, a combination of SRTM and Advanced Land Observing Satellite World in 3D (AW3D), with absolute bias, stripe and speckle noise and tree height bias removed. MERIT is described in more detail in Section 5.2.1.3.

Despite all of these attempts to improve the original SRTM data, its quality remains low and landscape changes are rendering it outdated. Sampson et al (2016), along with Schumann and Bates (2018), have underlined the need for better open source DEMs as this data gap is currently limiting the estimation of flood risk. They point out that quality can be lowest in urbanised areas where vulnerability is highest, due to interaction with structures and

vegetation. While freely available global DEMs remain limited, a larger number of commercial higher resolution products are available (Table 1). If made available to the research community, these commercial DEMs may provide significant improvements in flood model performance.

Table 1 - Commercial and freely available global DEM products, adapted from Proietti etl al (2017). Only original and not derived products are included, lower resolution versions have been excluded.

	Source	Spatial Res. (m)	Vertical Accuracy (m)	Horizontal Accuracy (m)
Commercial				
Reference 3D	Airbus	30	10	10
SPOT DEM	Airbus	30	10	10
WorldDEM	Airbus	30	4	10
ALOS World 3D 5m	JAXA	5	5 (RMSE)	5 (RMSE)
NextMap 10	Intermap	10	10 (LE95)	-
VRICON 0.5	Vricon	0,5	3	3
Euro-Maps 3D	GAF AG	5	10	10
Freely Available				
SRTM 1arcsec (30m)	NASA	30	6-9	7-12
ASTER GDEM	NASA	30	15-20	15-20
ALOS WORLD 3D 30m	JAXA	30	7	7
TANDEM-x 12m	DLR	12	10	10

TanDEM-X is a relatively new product from the German Aerospace Centre, DLR (2018). The 90m version is fully open, while 30m and 12m products are available on request. As this is a new product, it has not been fully processed to remove artefacts, outliers, noisy areas or voids (DLR, 2019). Initial efforts are being made to correct the original data (Archer et al., 2018), however producing a fully hydrologically conditioned product will require considerable work. Due to its relative nascence and lack of corrected products, TanDEM-X is yet to be widely used within the flood modelling community. AW3D (JAXA, 2019) is another openly available global DEM but is also not widely used to model floods for the same reasons.

A number of commercial global DEM products are also available for purchase, as shown in Table 1. These products provide additional options to organisations with enough funding to acquire them. However, as they are not openly accessible, their use in research is very limited.

2.1.10 Model Resolution

The effect of DEM resolution on HDMs has been assessed in previous studies. This is due to the major incentive of reduced computational expense with decreasing resolution. The findings tend to highlight $100m^2$ as the minimum cell size to ensure water pathways are consistent. Savage et al (2016) conducted a sensitivity analysis of the LISFLOOD-FP model by varying friction, inflow hydrographs and DEM resolution. DEM resolution was found to affect water depth predictions more than timing and peak discharge. Mateo et al (2017) found a very minor increase in NSE of discharge outputs when increasing resolution from 11-1 km in a multiple downstream connectivity (MDC) version of CaMa-Flood.

Horritt and Bates (2001) tested LISFLOOD-FP on DEM resolutions ranging from 1000-10m and found that below 100m, there was no improvement on inundation area accuracy, when compared with satellite observations. However, flood wave travel times were more dependent on resolution. Longer travel times were found with increasing resolution until around 20m, after which they were slightly reduced. These results may have been influenced by calibration at different scales with the effects potentially representing only varying friction parameters and not just DEM cell size. Altenau et al (2017) found an improved NSE of 0.844 at 10m compared with 0.742 and 0.633 at 25 and 100m. CSI of extent also improved at higher resolutions. However, again different values of Manning's n were used in each case.

Yu and Lane (2006) found small differences in resolution to have a significant effect on both inundation extent and flood wave timing. 4, 8, 16 and 32m resolutions were used and smaller cell sizes resulted in reduced inundation area. The authors attributed this to the smoother and less complex topography at lower resolutions. In urban areas, Fewtrell et al (2008) find that lower resolution grids are also limited by their inability to resolve structures such as buildings.

Falter et al (2013) tested resolutions from 25-500m and found 100m to be a good compromise between run time and accuracy.

2.1.11 Friction Parameters

Manning's coefficient of friction is the most consequential variable parameter in any flood model. By tuning Manning's n up and down, a better match with observed discharge can be achieved. This would be acceptable if tuning resulted in a more physically accurate value, however more often the calibration accounts for other inaccuracies in the data or modelling process. Therefore a best fit Manning's n may only work well for a specific time period or area and cannot be readily transferred to ungauged areas or periods in time. Despite these issues, calibration using Manning's n is common practice as it provides an easy way to superficially improve model performance. As n has a large influence on model outputs, it is important to communicate along with any results or analysis, however its value is often not stated (Sampson et al., 2015).

Problems with equifinality in Manning's n have been encountered by Pappenberger et al (2005), who also found differences between effective and 'physically real' values. A sensitivity analysis on the River Alzette in Luxembourg using HEC-RAS found Manning's n to be responsible for 63.75% of model uncertainty split between the floodplain, channel and outflow (Pappenberger et al., 2008). Friction in the channel was also found to be by far the most consequential parameter, as illustrated in Figure 2-2. NSE is not necessarily the best measure of accuracy of extremes, however this does show how significant friction is, particularly within channels. Often, friction parameters are assigned based on land use, however this does not always improve results. Werner et al (2005) found that using different friction parameters for different land use types had little effect on model performance in the River Meuse as the dominant land use type had an overriding effect. As a consequence, having

a variable friction parameter on the flood plain would appear not to be beneficial. However this does depend on the proportions and properties of the land uses in any given area.

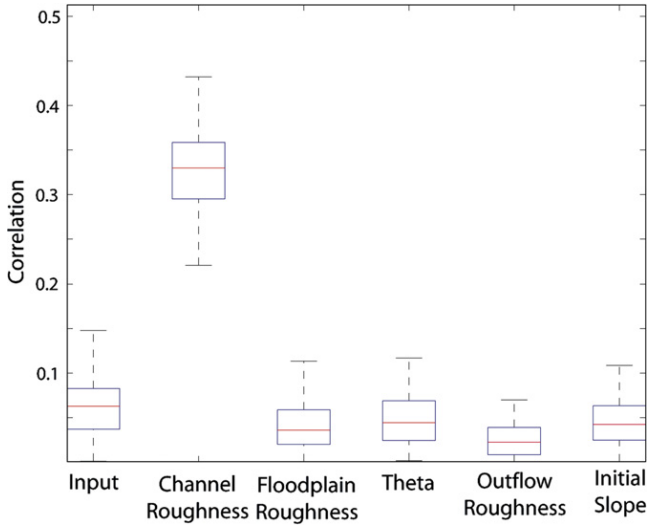

Figure 2-2 - Correlation to Manning's n of model parameters for the River Alzette in Luxemburg (Pappenberger et al., 2008).

Fewtrell et al (2008) optimised Manning's n for simulations at varying resolutions to the extent of a benchmark 2m simulation and found lower resolutions to be optimised at higher values of Manning's n. In this way, an unrealistic friction coefficient can correct for artificial smoothness at lower resolutions. For example, both Rollason et al (2018) and Wood et al (2016) calibrate their outputs using friction parameters. However, using a calibrated Manning's n value is often not practical and requires an observed flow series. Over-fitting may also result from optimising to a particular gauge and time period if techniques such as cross-validation are not applied properly. This can mean that the friction parameter may seem to produce good results but does not perform as well with unseen data. Although calibration using friction is common in local studies, the method is limited by data availability and therefore it is recommended that this should be avoided in broad-scale models if possible.

2.1.12 Rainfall

Rainfall inputs are required to drive HDMs. In data rich regions, this may be drawn from high resolution good quality gauges with long records. However, in many cases such data is not available. To create globally available gridded rainfall data sets, a number of groups have produced re-analysis products. These are generally a combination of interpolation and forecasting, based on observed data. The quality is always spatially variable and can be influenced by the density of gauges, steepness of terrain and algorithms used to grid the data. As with flood model outputs, validation is challenging. Nonetheless, re-analysis rainfall data has been widely used in large scale flood modelling due to its often continental to global scale availability and internal consistency. However, differences between re-analysis products, satellite data and gauge observations can be large. Sampson et al (2014) tested ERA Interim against surface gauge, radar and microwave observations, using a loss ratio model to highlight differences in final outputs (Figure 2-3). Radar rainfall was similar to the gauge data, however the re-analysis rainfall produced far shallower loss curves, indicating an under-representation of the hazard.

No two re-analysis products are the same. Donat et al (2014) found systematic differences in annual extremes between ECMWF, NCEP and JRA products caused partly by different approaches to gridding. However, extreme precipitation patterns were still generally correlated. Sperna Weiland et al (2015) tested the NCEP Climate Forecast System Re-analysis (CFSR), ERA-Interim and a combined CRU and ERA-40 dataset, which were found to have generally comparable climatology's. The datasets were most similar over the Rhine and Murray watersheds where measurement density and data quality were high. Unsurprisingly, simulated discharge was also the most similar in these areas. The main outcome of the study was that no parameterisation of the model performed consistently well, in terms of discharge, across all basins and therefore the authors state that rainfall data quality must be increased

before model complexity. Andreadis et al (2017) tested the ECMWF 20CRv2 dataset in Australia using LISFLOOD-FP and found relatively good correlation of 0.82 for the ensemble mean flooded area with a benchmark validated model driven by upstream observed flows. This similarity in extent is partly explained by both benchmark and test runs using the same elevation model. To properly assess the performance of rainfall data, discharge validation is required. However, discharge performance metrics, such as those discussed in Section 2.2.2, were not included in this study despite an acknowledgement that observed peaks were underestimated (Andreadis et al., 2017, p.1032).

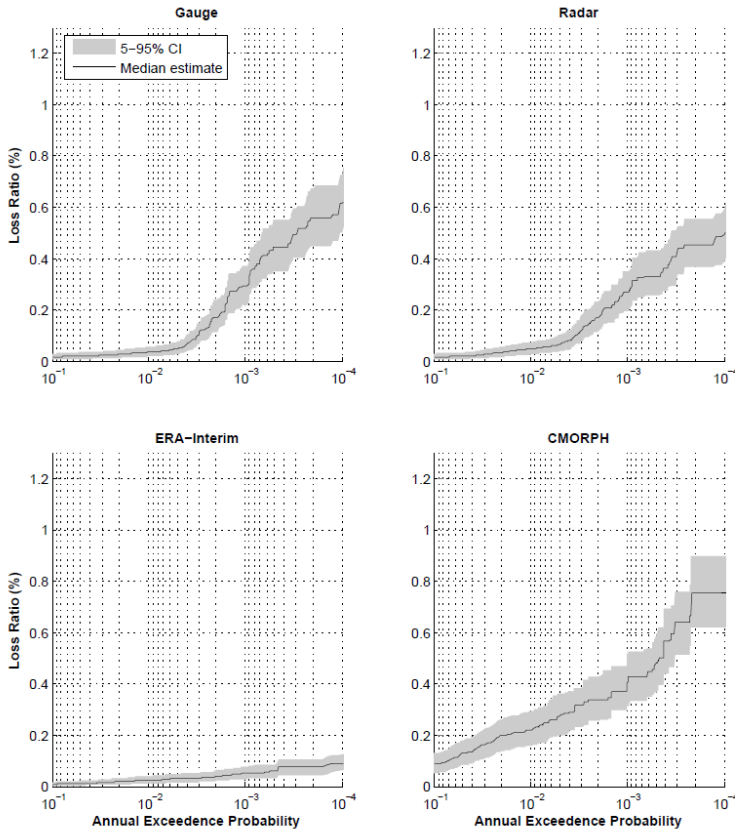

Figure 2-3 - Exceedance probabilities of losses as a proportion of total insured portfolio for Dublin, Ireland, using a range of precipitation data sources (Sampson et al., 2014).

2.2 Validation Methods

2.2.1 Introduction

Understanding how and where a model is underperforming is critical to understanding its usefulness and reliability. To estimate this error, validation is required. However, 2D hydrodynamic model results are inherently difficult to validate as measuring water depth at every grid cell for every time step of an event is not possible. The volume of calibration and validation data is always small compared to that produced by the model (Hunter et al., 2007). An example of these inherent limitations is Tyrna et al (2016) who compared locations of fire brigade alarms with the modelled extent and asked local engineers to confirm the plausibility of their results. Fire brigade alarms give no indication of depth or timings and the opinions of engineers are subjective and open to bias. This semi-quantitative approach is common amongst many studies which attempt to validate 2D flood model results and is understandable due to the challenges involved in collecting good validation data. Kutija et al (2014) attempted to validate their results for Newcastle upon Tyne using crowd-sourced photographs, however the exact timings of the images were unknown and the depths were only approximated by eye. Therefore interpretation of the results was very open to subjectivity and does not critically assess performance over time.

Rollason et al (2018) suggested a flexible approach to validation, incorporating volunteered geographic information (VGI) such as interviews, community mapping, photos and videos. They proposed three categories of validation: pathways, impacts, and timeline. However they found collected data to be messy and difficult to quantitatively quality control, leading to questions about its usefulness. Spatial and temporal variability in VGI is also inevitably present and comparison of performance between catchments would be highly influenced by the quality of VGI in each area. Zhou et al (2012) used a 2D inundation model to develop a risk

framework, however the validation was based on a single year of data collection which did not include any extreme events, again illustrating how challenging collecting validation data can be. This highlights that not only are many high quality measurements required, they need to be made over very long time periods in order to capture relevant events.

Yu et al (2016) used public reports of flooding corresponding to point locations, but with no measurement of depth, and found 96% of points falling within their modelled extent. Although this appears to represent good agreement, the lack of depth measurements means the accuracy is highly dependent on what depth threshold is chosen for defining the flooded extent from the modelled output. As observations don't tend to be recorded at locations which are not flooded, the over-estimation of extent or false alarm rate is not measured. Zischg et al (2017) used data from insurance claims to validate 2D flood model extent with metrics such as success index. The insurance data was found to produce similar validation metrics to documented extent outlines, however this approach does not give an accurate indication of timing. Clearly there is a need for a more robust approach to validating results from 2D flood models.

There are two broad categories of approaches to validation. The first involves continuous observations over time, limited to a single location and the second uses snapshots of observations, continuous over space but limited to discrete and often infrequent time steps. Time-continuous observations generally correspond to river flow gauges but may also be security cameras or any other continuous recording device at a single location. Space-continuous observations are generally captured from either an aircraft or satellite with sensors attached. Depths are often more difficult to calculate from these images but representations of maximum water extent may be produced and used for validating flood

simulations. Ultimately, a combination of both space-continuous and time-continuous observations are required to do the best validation possible for a given set of results.

Table 2 summarises a range of studies which have used either satellites, flow gauges or a combination to validate their outputs. Extent validation from satellites is most common and MODIS data is frequently used. GRDC is the most common source of streamflow data. While the data is widely available, GRDC records are limited to daily resolution, which may not capture flood peaks accurately. Distributed level observations have also been used but this is less common due to the relative difficulty in obtaining these observations (Neal et al., 2009). Generally, when streamflow data is used, it is to validate the hydrological or fluvial component of the model and therefore does not validate the 2D flood plain element. For the hydrodynamic element to be tested against flow gauges, a significant duration must be simulated across the entire domain to capture the event at each gauge. This is usually not carried out due to the large computational effort required. Both streamflow and extent validation will now be addressed in more detail.

Table 2 - Summary of modelling approaches and validation methods (E=extent, D=discharge, A=quantitative, B=qualitative) for major large scale flood studies.

Study	Modelling Approach	Output Resolution	Validation Time period	Method	Datasets used	Extent Dataset	Discharge Dataset
(Dottori et al., 2016)	Finite Volume 2D	30" (1km)	2000-2013	A	E	MODIS accumulated	GRDC (shape comparison)
(Falter et al., 2013)	Diffusive wave or volume spreading	25-500m	August 2002	A	E	Regional model outputs	-
(Paprotny et al., 2017)	1D model & planar extrapolation	100m	Return periods	A	E	JRC & regional model outputs	-
(Pappenberger et al., 2012)	GHM routed to rivers	1km	Return periods	A	E	DFO & Global Assessment Report maps	-
(Sampson et al., 2015)	RFFA routed to inertial SWEs	90m	Return periods	A	E	Regional model outputs	-
(Schumann et al., 2013)	GHM routed to 2D	1km	17 Feb 2007	A	E	Landsat	-
(Ward et al., 2013)	GHM routed to 2D	30" (1km)	Return periods	A	D	-	GRDC
(Ward et al., 2017)	GHM routed to 2D	30" (1km)	Return periods	A	E	Regional model outputs, Landsat, DFO MODIS	
(Wing et al., 2017)	RFFA routed to inertial SWEs	30m	Return periods	A	E	FEMA Hazard	
(Yamazaki et al., 2011)	GHM routed to channels	25km	1993-2001	A	E&D	SSM/I and ISCCP, ERS scatterometer, and AVHRR	GRDC
(Falter et al., 2016)	1D channels routed to 2D	100m	1990-2003	A – discharge B - extent	E&D	DLR	Unspecified source
(Hoch et al., 2016)	GHM routed to 2D with GRDC data	2.5-10km	1 July 1989 extent 1985-1991 discharge	A – extent B – discharge	E&D	Landsat	GRDC (single station)
(Alfieri et al., 2013)	GHM routed to 1D	0.1° (10km)	Jan 2009 – Jan 2011	B	E	MODIS	-
(Castellarin et al., 2011)	Quasi-2D finite difference	2-90m	RP (500yr)	B	D	-	Stream gauge for similar event
(Ghimire et al., 2013)	Cellular automata	2m	RP (100yr)	B	D	-	Finite Difference Model
(Rudari et al., 2015)	RFFA	90m	Return periods	B	E	UNOSAT	-
(Tyrna et al., 2016)	Cellular automata & Gauckler-Manning-Strickler equation	1m	"2010 event"	B	E	Photographs/ fire alarms	-
(Winsemius et al., 2013)	GHM routed to 2D	1km	10 July – 10 August 2004	B	E	DFO MODIS	-

2.2.2 Streamflow Metrics

Most metrics for comparing discharge model results to observations are designed to capture the characteristics of the entire series. In flood inundation modelling, though, this is not necessarily the best option and accuracy at the upper end of the hydrograph is far more relevant than for the whole series. Therefore, measures which look specifically at the highest values in a series are more useful. The obvious choice is the peak residual, measuring the difference between the largest values in the two series. Falter et al (2016) use a variation of this which sums all observed values over a bank-full threshold and compares this to the sum of the corresponding values in the simulated series. This peaks over threshold approach may be sensitive to the estimated bank-full value. Bank-full discharge was assumed to be the 2-year return period flow here but this may not be the case as channel capacity can be affected by geomorphological factors and may not be stationary over time (Slater et al., 2015).

Nash Sutcliffe Efficiency (NSE) has been used to evaluate short durations lasting only a few days (Rollason et al., 2018), however the metric is really designed for longer time series, usually encompassing at least one hydrological year. This is because the observed mean is used as a baseline to compare simulated outputs against. If the period is short, the mean and therefore measured performance will be highly sensitive to the start and end times. Kling Gupta Efficiency (KGE) is a similar alternative to NSE which is also not very appropriate for assessing flood simulations for the same reasons. Timing or phase errors can also be calculated by taking the difference between the observed and modelled peak times.

2.2.3 Streamflow Validation

River gauge data is the standard benchmark for assessing performance of fluvial models. Outputs are often then routed to 2D inundation models, so validation of channel stage and discharge is critical in getting the inundation correct. A good example of this is Falter et al

(2016), who use 14 gauges on the Elbe and tributaries to validate a 1D channel network model which is then sent to a 2D inertia model when stage exceeds dyke height. Both NSE and peak errors were calculated for the simulated discharge series, along with RMSE and mean bias for water levels. Alternatively, a 2D simulation can be directly tested against gauge data. Hoch et al (2016) tested coupled and 2D only simulations, forced with upstream observations, against gauge data at Obidos. R^2, RMSE and KGE were used as indicators here. Using only a single gauge limits the study though, as uncertainty in the gauge observations could have a large influence. Nguyen et al (2016) tested a coupled 2D hydrological-hydraulic model on the Baron Fork River basin on the Oklahoma-Arkansas border using two USGS stream gauges. Peak error, phase (time) error, RMSE, bias, r^2 and NSE were used here. In three out of four test events, the hydrological model performed better when calibrated without a hydraulic routing scheme. Water stage was also compared with observations, showing a consistent over-estimation from the coupled model.

Yamazaki et al (2011) validated CaMa-Flood in 30 major basins around the world using discharge from the most downstream GRDC station for various two year periods. R^2 and NSE metrics were used on the daily series. Large variability in NSE was found across the basins. Castellarin et al (2011) validated a quasi-2D model (UNET) using seven stream gauges along the River Po. Stage series and peak flow were visually compared, however no performance metrics were calculated. Dottori et al (2016) compare modelled and observed return period maximum discharge for 17 globally distributed gauges. As return period events were used, the modelled series could not be compared with corresponding observations, so the authors found similar peak discharge events and overlaid the hydrographs for a visual comparison (Figure 2-4). This method is highly subjective and sensitive to which observed event is chosen for comparison. The graphs presented are quite misleading as they plot the simulated hydrograph onto an axis with real dates, which do not exist in the hypothetical return period

event. Ward et al (2013) calculate percentage discharge peak errors, using corresponding return period discharge in observed GRDC series. This approach is more robust as the corresponding observed peaks are selected objectively.

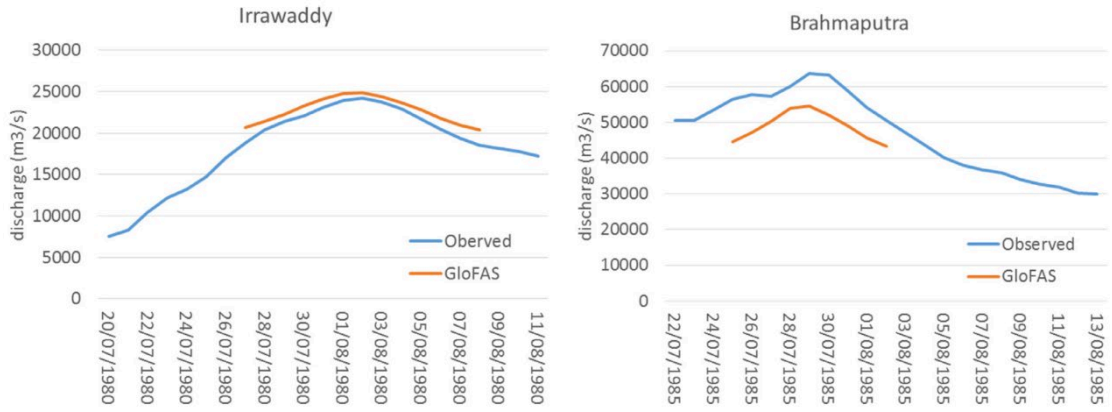

Figure 2-4 - Comparing return period model outputs with similar real world observation (Dottori et al., 2016).

2.2.4 Extent Metrics

Performance measures such as Hit Rate (HR), False Alarm Ratio (FAR) and Critical Success Index (CSI) (Donaldson et al., 1975) are commonly used to compare results with extents derived from remote sensing imagery or other models (Ward et al., 2017; Wing et al., 2017; Dottori et al., 2016). HR measures the percentage of observed flooded cells that are also flooded by the model. FAR is the percentage of cells that are incorrectly flooded by the model out of the total number of model flooded cells. CSI is the number of hits divided by the sum of hits, misses and false alarms. All of these measures intentionally ignore cells that are correctly identified as non-flooded, which is useful because they are not important in this context.

Differences in resolution between observed and modelled grids can complicate the evaluation process. One approach, taken by Wing et al (2017) is to resample the higher resolution grid to match the lower resolution grid and assign cells as wet if the larger cell is covered by more than 50%. Reporting HR in isolation can also be highly misleading as covering the entire

domain with flood water would produce a HR of 100%, however typically there should be some areas which remain unflooded. An example of this is seen in the abstract of Wing et al (2017) which quotes HR without FAR. CSI fixes this problem by taking into account the proportion of false alarms as well as correctly flooded cells (Schaefer, 1990). It has also been referred to as overall accuracy (Renschler & Wang, 2017) and is the best and most widely used measure of agreement between gridded flood extents.

In remote sensing, performance scores such as overall, user's and producer's accuracy, kappa, tau and conditional kappa are used to determine the performance of pixel classification algorithms (Stehman, 1997). User's accuracy corresponds to FAR, while producer's accuracy is HR in this context. Overall accuracy is the HR of multiple classes combined, which is not useful in flood mapping as only the flooded classes is of interest. Renschler and Wang (2017) modified overall accuracy by removing the correctly predicted non-flooded cells, in effect converting it to CSI. Kappa and tau deal with cases where classes may be assigned correctly by chance (Zhenkui & Redmond, 1995), which is not a concern in this context as hydrodynamic flood models are entirely deterministic.

2.2.5 Extent Validation

The most popular quantifiable and comparable way to validate the extent of model outputs is through the use of satellite remote sensing imagery (Hoch et al., 2016; Mateo et al., 2017; Andreadis et al., 2017). However, there is always a limited number of snapshots available for any location as satellite orbits only pass once every number of days. This temporal resolution problem severely limits their use in 2D model validation. Often, remote sensing products have not been validated themselves against any ground observations and the process of converting images to flood extent can vary greatly. Depending on the type of sensor used, satellite observations may also require a clear sky, which is inherently less likely during a flood event.

Despite these issues, there is no viable alternative so this remains the best method of authenticating flood extents using observations.

The Dartmouth Flood Observatory (DFO) database is a widely used source of flood extents derived from MODIS imagery (Dottori et al., 2016; Pappenberger et al., 2012; Ward et al., 2017). The method used to derive extent from the optical imagery is described by Brakenridge & Anderson (2006). Pixels are classified based on MODIS bands 1 and 2 by a non-supervised algorithm such as the Iterative Self-Organizing Data Analysis Technique (ISODATA). Manual intervention is then required to designate the resulting classes as water or land and remove cloud shadows that were classified incorrectly. This method is effective but, as with all methods using optical imagery, is limited when clouds are present.

Winsemius et al (2013) compared their 30 year return period results with extents accumulated between 1985-2010 in the DFO database, however no metrics were used and the comparison was highly qualitative (Figure 2-5). Mateo et al (2014) provide maps of simulated and observed extent from MODIS and AMSR-E, but again no quantifiable metrics were used. Another qualitative comparison was conducted by Yamazaki et al (2011) using SAR-derived datasets, however the authors did this time provide absolute values for total inundated area from both the model and satellite observations. Comparing total inundated area can be misleading though as it does not describe the ratio of hits, misses, false alarms and correct negatives. Schumann et al (2013) validate their results for the Zambezi 2007 flood using a single 30m ICESat image, using only a ratio of correctly predicted cells in low lying areas. The image itself was also obstructed by cloud cover so is likely to under-represent extent. Pappenberger et al (2012) compared results for the 20 globally largest basins with a combination of DFO extents and the flood hazard map produced as part of the Global Assessment Report on disaster risk

reduction. Hit rates, misses, false alarms and correct negatives were plotted for return periods of 2-75 years, however the precise methodology for comparison was not fully presented.

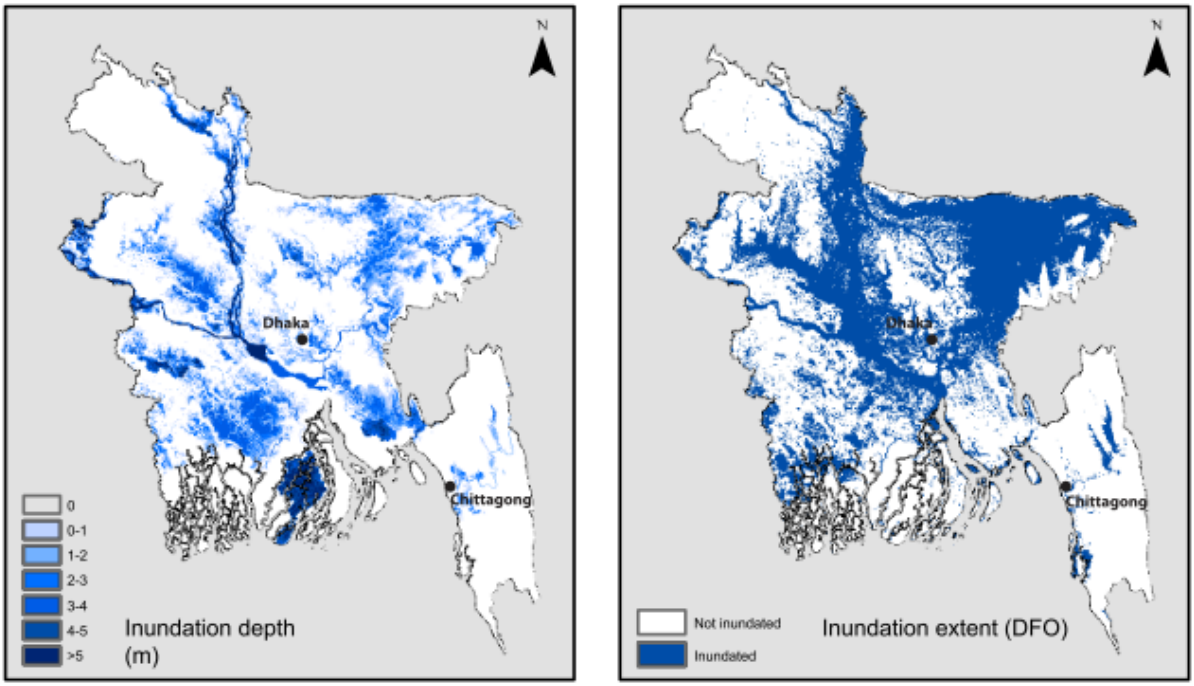

Figure 2-5 - A qualitative visual comparison over Bangladesh of a return period model output with remotely sensed extents accumulated over the return period (Winsemius et al., 2013).

Dottori et al (2016) of the JRC in Italy used MODIS observations compiled by the Dartmouth Flood Observatory (DFO), accumulated over 2000-2013 and combined them with maps from the UN Operational Satellites Programme (UNOSAT) to evaluate their Global Flood Awareness System (GloFAS). On comparison of Critical Success Index (CSI), GloFAS performs better than Alfieri et al (2013) and Winsemius et al (2013), but not as well as Sampson et al. (2015), though comparison is hindered by different model resolutions and channel network extents. Tarekegn et al (2010) use 250m MODIS flood observations to validate a 2D hydrodynamic model on an ASTER-generated DEM. Renschler and Wang (2017) validate HEC-RAS using both airborne LiDAR and flow gauge recordings from the 2011 floods across New York State. Perks et al (2016) have demonstrated the applicability of georeferenced UAV data recorded during an event in validation of flood models. This is quite expensive though and challenging to implement across large areas.

New developments are taking place in earth observation technology, such as the European Space Agency's Sentinel-1 and Sentinel-2 missions for Copernicus (Clement et al., 2018), SmallSats (Ruf & Warnock, 2018) and High Altitude Platforms (HAPs) (Di Vito et al., 2018). These developments have the potential to provide datasets with better temporal and spatial resolution than have been previously available. This will be highly beneficial for improving the robustness of flood model extent validation. Copernicus' fully open data policy (European Commision Parliament, 2013) will accelerate the use of Sentinel data in research and ensure reliable accessibility into the future.

2.2.6 Volunteered Geographic Information

VGI data can provide readily available and widespread coverage of a flood event, however is significantly limited by difficulties in ascertaining its validity. Despite this, many studies have utilized VGI to supplement conventional validation methods. Smith et al (2017) use crowd-sourced textual and photographic data from Twitter to validate a 2D simulation of Newcastle Upon Tyne at 2m resolution. They found that depth information was often lacking within tweets and the timestamp may not correspond to the time the observation was made if a photo or statement is posted later in the day. In the context of real time hazard management, they found twitter data about inundation extent can often come much later than the inundation actually occurs. Twitter was again used by McDougall (2011) to map the 2010/11 Queensland floods. They identified good data at the beginning of the event but as it progressed, spurious postings and incorrect information became more prevalent.

The photo sharing platform Flickr has also been integrated into a flood monitoring framework and tested on UK flooding in 2007-2009. Longueville et al (2010) attempted to improve the credibility of VGI from Flickr by using reasoning techniques and past observations. However, only the text attached to images was processed which significantly limits the information

which can be retrieved. The results indicated VGI was successful at detecting major floods in the DFO database. They also found significant noise in the data causing false positives which is a limiting factor of VGI. VGI has been used to look at impacts on road infrastructure by Schnebele et al (2014). They estimated flood damage in New York City caused by Hurricane Sandy using data from YouTube, Twitter and Google Crisis Map. They found the combination of data sources useful as missing areas from one are more likely to be covered by another. Flood identification from multiple platforms also lends more authority to their validity.

2.2.7 Model Inter-comparison

Previous studies have validated model results based on results of other models using similar methods and data for the same return period (Wing et al., 2017; Sampson et al., 2015). This is of some value as large differences in outputs will indicate a problem, however if the results match, this does not necessarily show that they are correct, only that the two models have found the same results based on similar input data. Falter et al (2013) compare two simplified models with outputs from the more complex InfoWorks RS 2D. This does allow benchmarking of the more simple modelling methods against a more complex approach, however it does not say anything about the accuracy of either. The main issue is that a good match with a more complex model can easily be interpreted as an indication of accuracy, which is not necessarily the case. Both complex and less complex versions may be highly inaccurate. Of course, if the results are effectively caveated and explained, this problem can be negated, but it is often not in the interest of the authors to insist that their results say nothing about accuracy.

Elshorbagy et al (2017) use a local hydraulic model produced by the city of Calgary, Canada to test their data based approach to flood hazard mapping. Although it may be useful in identifying similarities or differences in approach, this does not indicate which is more correct. Despite these issues, model inter-comparison studies have been conducted without the use

of any observed data. For example, Trigg et al (2016) compiled a flooded area vs return period graph for a number of modelling chains (Figure 2-6). Inter-comparison of large scale models in this way provides a useful view of how disparate their results are and which can be used interchangeably, however still does not present accuracy. Other examples include Dimitriadis et al (2016) and Hunter et al (2008) who compared stage between various simulations but did not include any observed data. Without attempting to simulate real events, with recorded observations at gauges and from satellites, it is not possible to calculate accuracy.

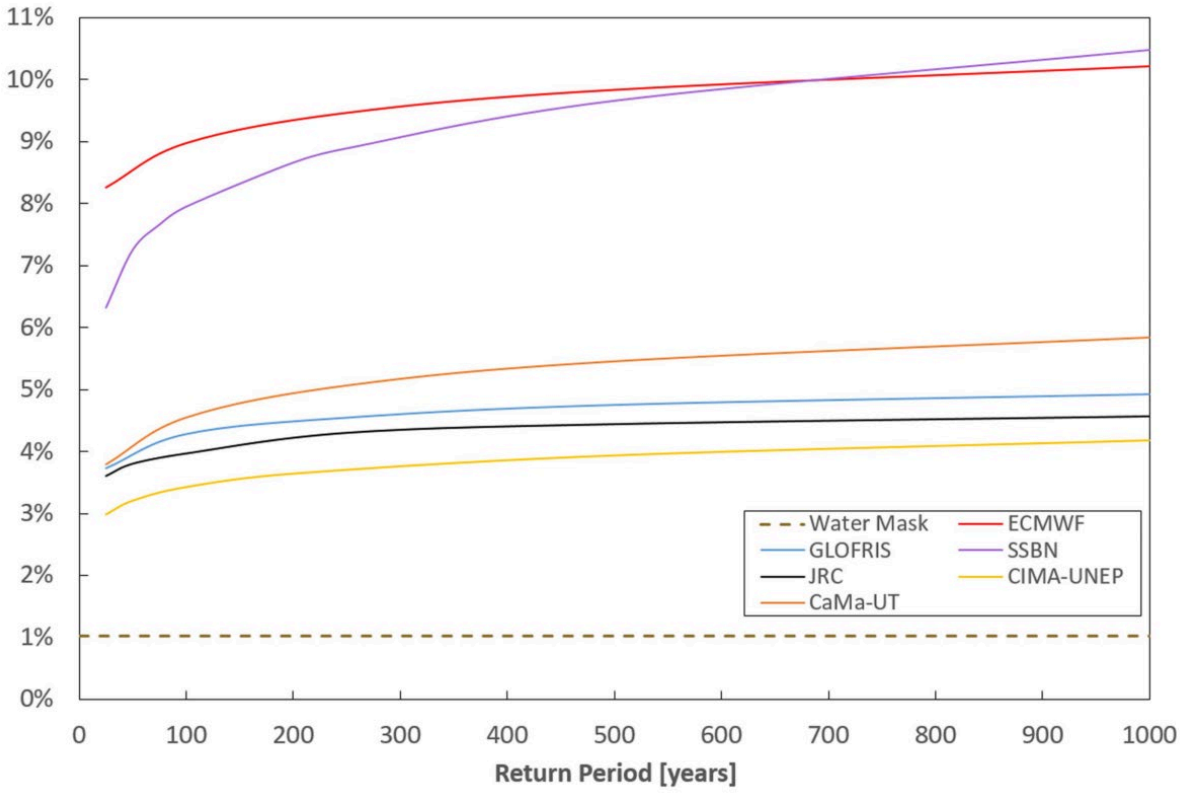

Figure 2-6 – An example of model inter-comparison without any observed data (Trigg et al., 2016).

2.3 Impacts Analysis

2.3.1 Introduction

Outputs from large scale inundation models have been used to calculate impacts on affected infrastructure and population. The modelling chain can often be long and complex with uncertainty introduced at each stage, an example of which is shown in Figure 2-7. The consequence of these often long modelling chains is that it can be difficult to interpret

whether a signal is genuine or simply an artefact of the methodology. Calculations are often made without any data to validate against and therefore their accuracy cannot be assessed. This is a problem because different studies produce widely different damage estimates and there is no way to compare validity. However, results can still be useful as a best approximation. The temporal aspect of flood impact is generally not considered as it presents a more computationally demanding task and the results can be difficult to interpret.

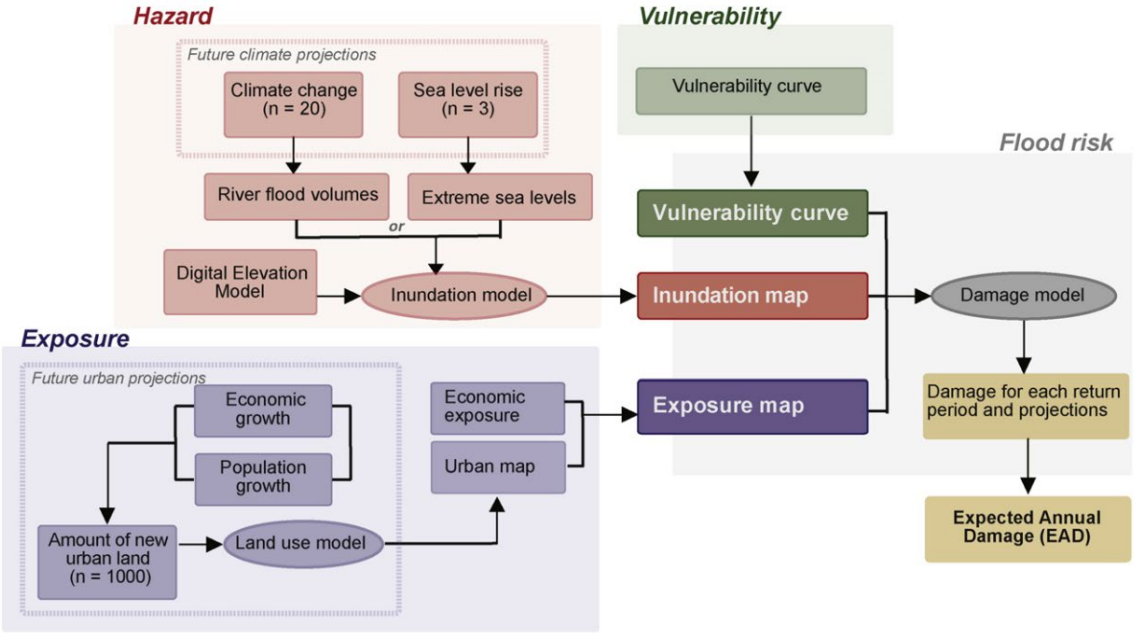

Figure 2-7 - An example of a long risk modelling chain from Muis et al (2015).

2.3.2 Quantifying Current and Future Risk

Impact analysis is useful for both general policy making and local investment. Wing et al (2018) used outputs from LISFLOOD-FP to derive estimates of population numbers exposed to flood return periods. For the conterminous US, 2.6-3.1 times more people were found to be at risk than previous estimates. This difference is largely a consequence of the improved coverage of flood model results used in the study compared with the baseline patchy FEMA local studies. Alfieri et al (2017) use hazard maps from Dottori et al (2016) in combination with population and land use data to generate both current and future impacts globally at 1km resolution. Future changes were not found to be significant in most of Africa and Oceania. The largest

increases in risk were found across Asia, the US and Europe. However, Hirabayashi et al (2013) found large increases in flood frequency in eastern Africa. They used CaMa-Flood with an inundation component and an ensemble of climate forecasts with a limited validation in only 32 basins.

Validation of modelled impacts is rarely carried out but Ward et al (2013) manage to demonstrate a correlation over time between observed fatalities and exposed population, shown in Figure 2-8. They present annual expected impacts at 30" resolution, equivalent to around 1 km at the equator. The study highlighted sensitivities to both flood protection standards and climate forcing data used. Feyen et al (2012) used LISFLOOD and the CORINE Land Cover 2000 dataset to derive a current and future damage assessment for Europe. Depth damage functions were again used and rescaled to fit the country's GDP. The final outputs were at 100m resolution for return periods between 2 and 500 years. Flood protection was accounted for by discarding floods at lower return periods. Population exposure was assessed using a 100m density map. A further study by Mokrech et al (2014) uses the same fluvial model but also incorporates coastal flooding from the DIVA database at 10' spatial resolution. They conclude that 28.6 million people, 6% of Europe, are at risk from the 100 year flood event, with most risk concentrated on the western coasts. They found that net future flood risk will not increase but a higher proportion of inundation will occur in the north-east.

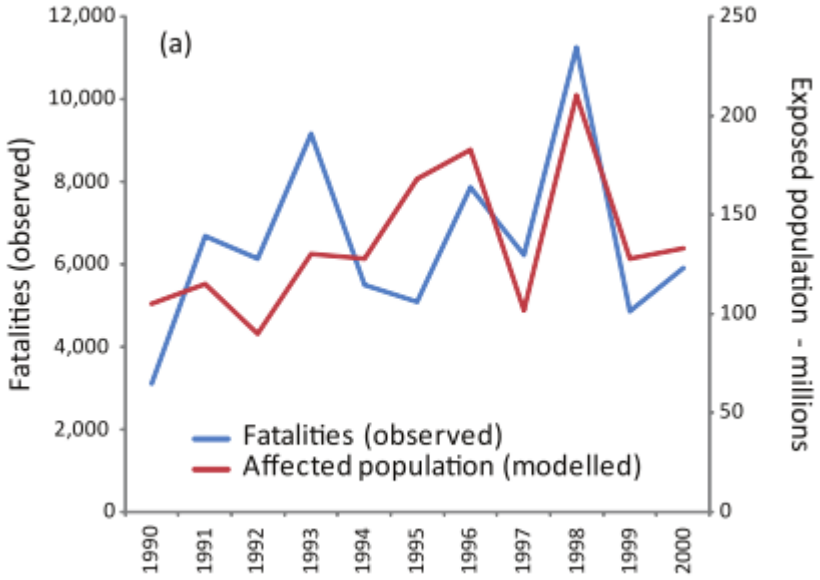

Figure 2-8 - Reported and modelled impacts for the period 1990–2000, whereby reported impacts are taken from the NatCatSERVICE database of Munich Re (Ward et al., 2013).

The CORINE dataset has also been used by Thieken et al (2015) to highlight the limitations of considering return period events as occurring homogenously across a catchment. Arnell and Gosling (2014) calculate present and future fluvial flood risk globally at 0.5' resolution using a daily water balance model, Mac-PDM.09. The climate forcing dataset is disaggregated statistically from monthly to daily for a 30 year period. Routing was not simulated so catchments were effectively limited to around 2,500 km^2. Population and GDP data from the IMAGE 2.3 projections, also at 0.5' resolution was used to project impacts into the future. They found considerable uncertainty in the magnitude of changes in impact between climate model projections.

2.3.3 Hazard vs Exposure

The predominant cause of increasing exposure to floods is development in flood prone areas, rather than the flood prone areas increasing in size. Ceola et al (2014) used night time light data to illustrate the increasing density of population along large rivers. An up-scaled version of the Global Human Settlement Layer Global Population Grid was used along with the FLOPROS database, GlobCover 2009 land use data and damage functions from Moel et al

(2016). Elshorbagy et al (2017) also use nightlight data, this time over Canada, to assess flood exposure at the national scale by classifying luminosity into five exposure classes from very low to severe. These classes were combined with a simple hazard layer derived from topographic descriptors to produce risk maps, excluding vulnerability.

One problem with night lights is that they cannot be used as a proxy in areas without consistent electricity supply. In areas that do though, this provides a useful, frequently updated measure and can easily be used to map changes in exposure over time. There is clearly uncertainty associated with the degree of development expansion at continental to global scales. Muis et al (2015) deal with this by producing 1000 projections for urban expansion between 2000-2030. They find that urban extent may increase between 215-357% by 2030. They cite this as the main driver for increasing risks rather than the highly uncertain changes in fluvial or pluvial flood risk. Sea level rise was found to further elevate the exposure-induced increase in flood risk by 19-37%. The baseline 2000 urban extent was derived from the MODIS v5 land cover product with a 0.5 km resolution, resampled by Muis et al to 1 km.

The HYDE population database, at 1 km resolution, has been used to calculate global population at risk from 100 year floods as 805 million people in 2010 (Jongman, Ward, et al., 2012). 73% of this value was in Asia with the largest relative increase found in Sub-Saharan Africa between 1970 and 2010. However, there is a clear over-estimation issue with such low resolution exposure data. Jongman et al (2014) use property level information to assess exposure in the Netherlands between 1960 and 2012. The exposure dataset consists of 8.9 million addresses in the BAG database with information about use, surface area and year of construction. They found the total number of properties exposed to increase by 216%. Worryingly, development has been more rapid in flood-prone than non-flood-prone areas. It is possible to further subset populations into groups to look at patterns in who is most

exposed. One clear delineation is wealth and Winsemius et al (2018) have found that poor people are disproportionately exposed. They used household survey data from USAID and DHS with 1 km fluvial risk outlines from GLOFRIS. Interestingly, a threshold of 0m depth (i.e. any water was considered as flooding) was used, which may not correspond to substantial impacts. Africa was shown to have the largest disparity in exposure between income levels.

2.3.4 Vulnerability and the Depth-Damage Function (DDF)

Given depths of flooding cause varying levels of damage depending on where they occur. For example, in highly developed and adapted areas, building guidelines may recommend ground floors to be assigned non-essential uses such as garages, meaning flood depths must be far greater to have an impact than areas with primary facilities on the ground floor. This spatial variability in vulnerability is usually characterised by the depth/damage function, as suggested by Messner et al (2007). There is, again, a large amount of uncertainty associated with deriving depth-damage functions, as illustrated by Merz et al (2004). They used the 4000 record Bavarian HOWAS dataset describing damages from nine floods in Germany to illustrate the limitations of these functions. It was demonstrated that the uncertainty in damage estimates depends on the number of buildings and variability of use. It is suggested that various other parameters than water depth are required to ascertain the damage on buildings. Alternatives to traditional depth-damage curves, such as tree-based approaches (Merz et al., 2013), exist but are generally very data intensive. Jongman et al (2012) also identify large uncertainties depending on the vulnerability curve used. Damage curves are often developed in one region and applied to another, however Budiyono et al (2015) stress that care is needed to assess whether the damage relationship is transferable. They found damages several orders of magnitude larger than previously reported for floods in Jakarta during 2002 and 2007 by using more localised vulnerability curves. Freni et al (2010) point to depth damage curves as the

main limiting factor in urban flood risk estimation due to the limited advancements compared to the rapid development of inundation models.

2.3.5 Infrastructure Damage

Even less research exists on development of depth-damage functions specific to infrastructure such as roads and railways. Unlike buildings, road and rail networks may become impassable during an event without becoming damaged in the long term. Temporary impassability does not have a directly calculable monetary cost but can have major impacts during an event, preventing or reducing mobility. Flow velocity has been found to be more influential than depth on the magnitude of damage to roads (Kreibich et al., 2009). The size of road, quality of maintenance and surface material are all required to calculate the long-term damage caused by flooding. Temporary impassability however is much easier to quantify as it is largely independent of the size and quality of road or rail infrastructure. Pregnolato et al (2015) developed a literature based flood-depth to car-safe-speed curve in order to assess the impacts of extreme rainfall on transport networks. They found 30cm to be a maximum depth for roads to remain passable. Yin et al (2016) also use 30cm as a threshold, based on the height of standard car air inlets. They present lengths of roads inundated for different durations at each return period, using a 2m 2D inundation model, FloodMap-HydroInundation2D, for the Shanghai central business district. Inundation of transportation networks does not only affect the directly inundated regions. If no similar alternative routes are available, there are consequences for the rest of the network. Particular regions of networks have more of an impact than others. If critical links are inundated, they have a disproportional effect on the entire network (Sohn, 2006).

2.3.6 OpenStreetMap for Impacts Analysis

OpenStreetMap (OSM) (Haklay & Weber, 2008) is the world's largest source of free geographic data such as roads and buildings globally. The data is sourced from a combination of individual contributors and pre-existing databases, combined into one schematized format consisting of points, lines and polygons. The data has been used for flood risk analysis by a number of studies. Schelhorn et al (2014) use OSM data to assess elements at risk in Cologne, Germany. Eckle et al (2016) look specifically at flooding of critical infrastructure within the OSM database. Weiser and Zipf (2007) have developed an emergency routing tool using OSM which can adapt to developing disasters. Rahman et al (2012) created a mobile based application using OSM to provide early disaster warnings and evacuation guidelines in Bangladesh. The data has also been used for identification of locally perceived risk (Klonner et al., 2016). Rapid updateability has allowed time critical information to be added to OSM during disasters such as the 2015 Nepalese earthquake (Rieger & Byrne, 2015), 2010 Haiti Earthquake and 2013 Typhoon Yolanda (Palen et al., 2015). Schorlemmer et al (2017) have been working towards collecting vulnerability information for buildings globally, using the OSM framework.

Due to the volunteered nature of much of the data, there is some uncertainty as to the accuracy of the records. Haklay (2010) looked at this using Ordnance Survey data for London and found an average of 6m difference in position of points and 80% overlap of motorways. In 71% of the area in England where features are present, better coverage was found in the OS data, with 29.3% having comprehensive coverage in OSM. This study was following only four years of data collection though and completeness has been growing exponentially since then. In the same year, Girres and Touya (2010) assessed OSM quality in France and found an average of 2.19m error in road locations when compared with BD TOPO® data from IGN, the French national mapping agency. Brovelli et al (2017) present a grid based approach to evaluate the completeness of OSM roads, however do not conduct any real analysis.

Completeness of roads was found to be 45%. Cipeluch et al (2010) compared OSM with Google and Bing maps for five areas in Ireland and found no major differences in completeness.

2.4 Summary

A review of the literature concerning broad-scale flood hazard and risk modelling has been presented. The challenges presented include reproducibility, how to estimate extreme discharge, data availability and quality, computational requirements of accurate numerical schemes and wide variation in approaches to validation. Many modelling systems exists but fundamental data limitations mean they experience similar problems. These uncertainties propagate into impact analyses and cause large differences in damage estimates between studies. Improving elevation and precipitation data coverage and quality would perhaps make the most fundamental difference in reducing uncertainties in model outputs and is further discussed in Chapter 7.

The need for further research into validation approaches is clear. Along with this, uncertainty analysis in infrastructure impact modelling is almost non-existent. The reasons for these gaps are understandable. The underpinning cause is an overwhelming lack of high quality observations for validation of model outputs. This thesis does not attempt to solve the data shortage problem but aims at tackling how best to use the data that is available to create a transparent and informative outline of the performance of a simulation and therefore how much confidence can be placed in any resulting estimated potential impacts on infrastructure systems. The next chapter outlines the approach to modelling taken in this thesis which is then used in later chapters to tackle the validation challenges experienced by previous studies.

Chapter 3. Cloud-based Modelling Framework

Evaluating HDM performance using many combinations of inputs and parameters across large areas at high resolution presents challenges in data storage, processing and analysis. Computationally demanding processing needed to be distributable across many High Performance Computing (HPC) machines and the results centrally accessible. A range of options were initially considered before a final framework was designed which met the requirements. The Microsoft Azure cloud computing platform provides HPC resources for modelling, and a local PostgreSQL (postgres) database is relied upon for storing data inputs and outputs. A software stack, including Python, JavaScript and SQL was designed in three main components. Firstly, a python library, named Automatic Catchment Analysis Tool (autocat), handles data uploads, downloads and processing of data outside of hydrodynamic modelling. This data and processing workflow is shown in Figure 3-1. A second python library, azure_api, handles creation and deletion of virtual machines on Microsoft Azure. Finally, GeoServer is combined with a Python Tornado and Angular JavaScript web app to create a front-end user interface for viewing and analysing results, called flood-map.

The open source Anaconda Python 3.6 environment from Continuum Analytics was used as a base for code development. Most of the require d libraries are already packaged with this distribution but some additional geospatial and data processing packages were required. The Geospatial Data Abstraction Library (GDAL) was used for handing all raster and geometry data. The NetCDF4 library was required for extracting climatic data. Finally, the python-wrapped postgres driver psycopg2 was used for connecting to the database. To create an API which accepts requests to create and delete Azure virtual machines, the compute and network components of the Microsoft Azure Python module were installed using PIP. The PyCharm Interactive Development Environment (IDE) from JetBrains was used throughout the

development process, providing a built-in database manager and Python, JavaScript and SQL code suggestion and correction. The Quantum Geographic Information System (QGIS) was also used in development to quickly and easily visualise geographical datasets on top of OSM base maps. Hydrodynamic simulations were carried out using an innovative 2D modelling system, CityCAT (City Catchment Analysis Tool), developed at Newcastle University.

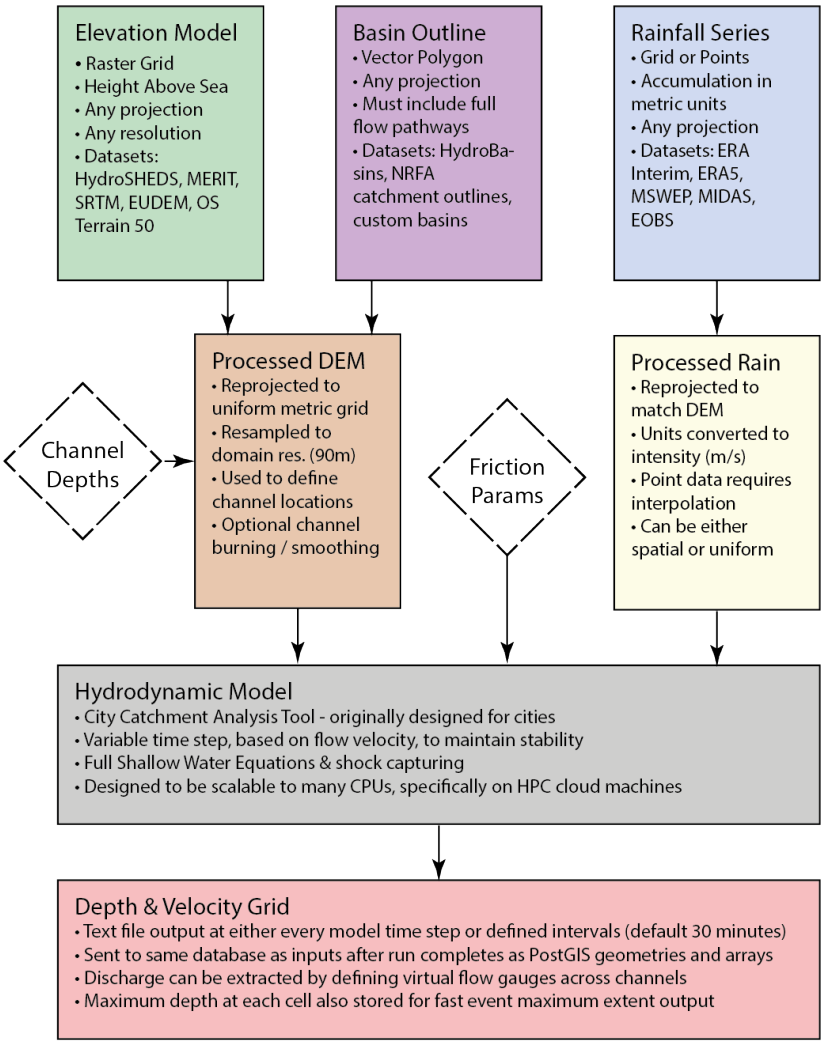

Figure 3-1 - Generalised modelling workflow.

3.1 PostgreSQL Database

PostgreSQL is a well-supported and documented RDMS which supports a wide range of data types. Python is able to connect and send or receive data conveniently through the Psycopg2

adapter. The main attraction of PostgreSQL is the spatial extension PostGIS, designed by the Open Source Geospatial Foundation. PostGIS adds support for geographic objects, including both rasters and geometries, and spatial queries. This means data can be conveniently extracted for a given domain from a large dataset covering continental-scale areas. Using PostGIS also means other software, such as GeoServer and QGIS can easily interact with the database from anywhere on the internet. A virtual machine running Ubuntu 16.04.4 LTS, PostgreSQL 9.6 and PostGIS 2.3 with 5TB of storage and 8GB of RAM was used to host the database. Postgres was configured to be accessible from the internet with username and password authentication.

3.1.1 Schema

To be able to be stored in the database, each dataset must first be converted into a tabular layout with specified data types for each column. This layout, called a schema, is sometimes already included within the input data. For example, Shapefiles are basically geospatial tables and can be automatically inserted to the database using a Windows GUI tool called PostGIS 2.0 Shapefile and DBF loader. This was the case for the Global Rivers Dataset (GRD) (Andreadis et al., 2013) and HydroSHEDS basin outlines (Lehner & Grill, 2014). Raster datasets, such as DEMs are also easy to upload using the raster2pgsql command line application. All raster file types were converted to GeoTIFF format before import and where data was provided in tiles, a virtual file (VRT) was used to combine them. To improve efficiency of access from the database, all rasters were stored in 50x50 pixel tiles. Rainfall and river flow data usually come in more disparate formats and need some conversion to fit into postgres tables. Python scripts were designed for each of these datasets and the schema is shown in Table 3. Additional tables were also created to store information about model runs, including the data sources and parameters. The process is now discussed in more detail.

Table 3 - Database schema for rainfall and discharge data.

Data Source	Schema
Rainfall (point)	gauge_id (integer) geom (geometry), timeseries (float array) start_date (timestamp)
Rainfall (grid)	geom (geometry) timeseries (float array)
Flow gauges	gauge_id (integer) geom (geometry) metadata (json)
Flow values	gauge_id (integer) date (timestamp) value (float)

Precipitation time series from both gridded products and point gauge data were used in the research. The gridded datasets were originally in NetCDF format. The time series for each grid point were converted into postgres numeric arrays and cell locations into PostGIS polygons. Each gridded dataset was stored in its own table, meaning that the start dates and time steps were the same for each row. This information was stored in the table comments and included in the Python code used to extract the data for running simulations. To reduce the time required to read each cell and write to the database, multiprocessing was used in Python to parallelise and distribute the cells between available processors. The point gauge data was in text format, with a file for locations and one for observations. All observations for every gauge were stored in a single file so each gauge had to be separated. The data also contained duplicate values, relating to different versions, and values for both hourly and daily accumulated series. The unique hourly data was extracted using the Pandas Python library and inputted to the database. Start dates were stored as a separate field in the same table to enable time series to be later rebuilt.

River flow gauge data was also required for validation of model results. As multiple time series were available for each gauge at different temporal resolutions, two separate tables were used this time, allowing each observation to be designated its own timestamp. Datasets from

the NRFA and GRDC were both in a textual format and easily readable by Pandas. The series were again stored in postgres numeric arrays and gauge locations as PostGIS points in a separate table. Metadata was generally available for gauges and was stored alongside their location in the flow gauges table as JSON key value pairs.

For running many simulations with various setup parameters, a table to store model locations, start and end times, data sources and pre-processing parameters was required. To facilitate adding of more parameters as the methodology develops, they are stored as a JSON group of key-value pairs. Other information included in the runs table includes the Git commit ID of the code that was used to run the model, an arbitrary descriptive name, the hostname of the machine that was used, the size of this machine if it was on Azure and a start and end time for the processing. To group multiple runs into sets, using the same parameters but for different areas, a run groups table was also added with a lookup of group ID to an array of run IDs and a designated group name. To simplify sensitivity analyses, these groups were further organised into group sets for comparing input data and parameters.

The schema for the results table is similar to that of the gridded rainfall datasets. Each row refers to one cell from an individual run. The rows contain polygon geometries both in the model spatial projection (European Lambert Azimuthal Equal Area or Universal Transverse Mercator) and in WGS84 for use in creating web maps. Further details about spatial projections are included in Section 3.3.1. Time series are stored in postgres numeric arrays. The date and time of the event that was simulated is already stored in the runs table, so no temporal information is required.

To provide data on infrastructure for impacts analysis, OpenStreetMap (OSM) data in compressed PBF format was downloaded of the required areas from Geofabrik (2018). As OSM is usually stored in PostgreSQL, many open source tools are available for inputting

downloaded data to the database. Two of the most popular are Osmosis, written in Java and osm2pgsql, written in C++. Both are available as cross-platform command line applications. The PBF data was converted into 3 tables containing points, lines, and polygons using osm2pgsql (Burgess et al., 2018). An additional table called roads is created containing data from the other 3 selected for low-zoom rendering. Attributes referring to the properties of geometries are stored in columns based on key-value pair tags associated with that object. Common tags and what they refer to are available on the OSM wiki (OpenStreetMap, 2019). Particular types of feature can easily be selected based on these tags.

3.1.2 Communication

Figure 3-2 shows the main channels of communication required between the database, model VMs and visualisation platform. Connection to the database was primarily through the Python library psycopg2 and a specific wrapper was designed to create new connections with a cancellable backend. PyCharm, QGIS and PgAdmin were also all used to connect to send and retrieve information from the database. GeoServer has built in PostGIS support and was able to connect directly to the database. Functions and methods were written in Python with nested SQL for requesting specific data based on a run, group or set ID. To optimise retrieval of required data based on spatial locations and ID numbers, appropriate indexes were added to all tables. Indexes require additional hard disk space but massively increase query performance. As new data is added, indexes are automatically updated.

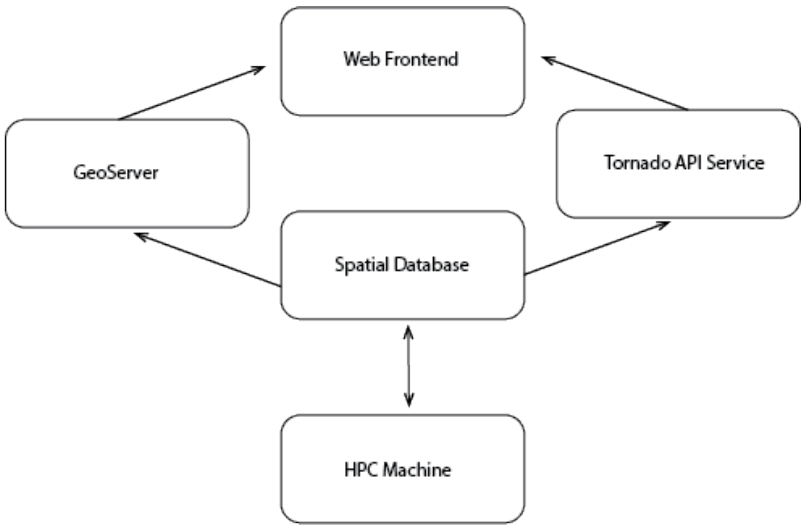

Figure 3-2 - Communication between database, remote machines and web-based frontend.

3.1.3 Maintenance

To keep the database running efficiently, occasional maintenance was required. For example, following large additions to the results table, ANALYSE was executed to collect updated statistics. These statistics can then be used by the query planner to speed up query execution. Occasionally large amounts of data were deleted, requiring a VACUUM to be run which reclaims storage on the hard disk. VACUUM and ANALYSE can be run in a single combined statement.

3.2 Microsoft Azure

Running broad-scale 2D flood simulations requires machines with a large number of fast CPUs. One option is to purchase many of these machines but that would be prohibitively expensive and a waste of resources as they would only be used for short periods. Cloud computing provides a solution to this issue by allowing expensive computational resources to be leased on demand at an hourly rate. There are a number of companies offering these services, such as Amazon, Google and IBM. However, Microsoft Azure was selected due to its generous research grant sponsorship scheme. A $20,000 allowance was received for modelling on Azure Infrastructure as a Service (IaaS). There are a range of high specification machines available

on Azure, categorised into series with a designated alphabet letter. The most suitable series for this research was the H series, designed for “high-performance computing workloads”. H series machines are available with either 8 or 16 3.2 GHz processors, 1 or 2TB hard drives and 56 or 112 GB of RAM. The machine used depends on the size of the domain to be modelled and is designated in the runs table. The Azure Python API has simplified and sped up the process of creating many Windows instances of specific size and power asynchronously, based on the same custom image. This means many basins and events can be simulated easily and quickly by splitting runs between machines, as illustrated in Figure 3-3. The Azure Web Portal and mobile app also allow easy monitoring of run progress and remote desktop connection to machines which require debugging.

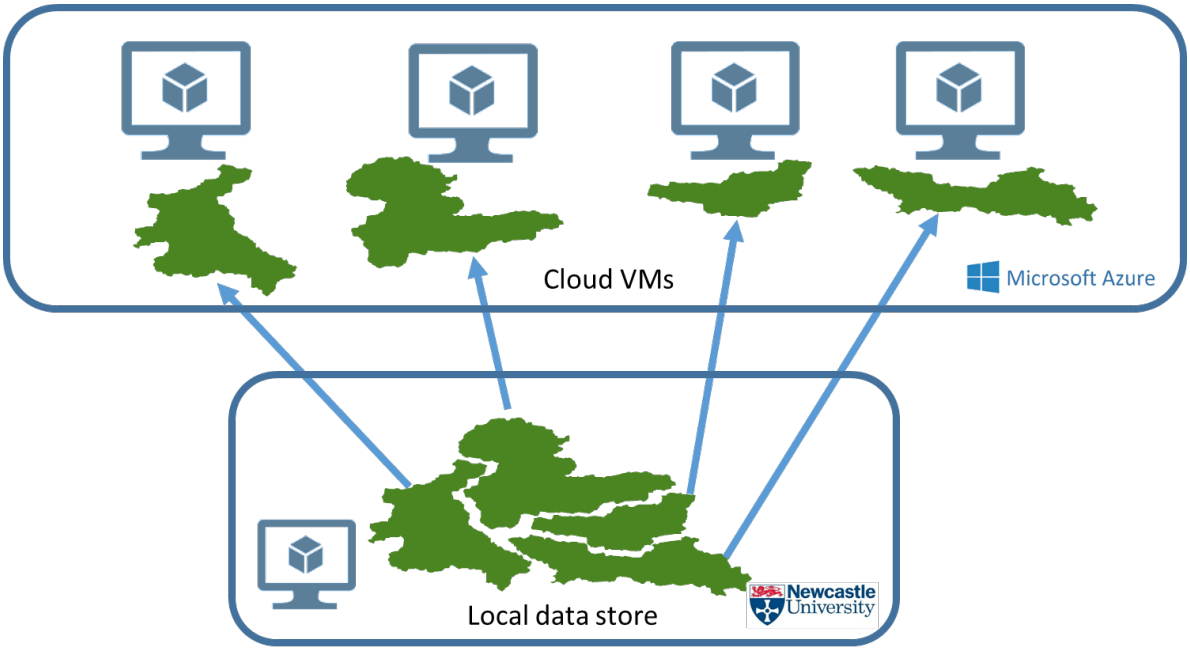

Figure 3-3 - Domain splitting to distribute modelling computation.

3.2.1 Custom Images

The first step in deploying to Microsoft Azure was to create an image including all of the required base software. Windows Server 2012 was the operating system used. On a small VM instance, Anaconda 3 with the addition of psycopg2 and GDAL, QGIS and Git were then installed. A Windows Batch script was included in the ProgramData folder. The script clones

the Git repository into the temporary storage disk, followed by a JSON file, also stored in ProgramData, which includes the credentials required to connect to the database and a number of other settings. The output of the batch script is recorded in a text file for easy debugging. The image was then generalised and stored online for use when creating new VMs. Occasionally, it was necessary to update the custom VM image with the latest security patches. This was achieved by spinning up a small VM instance, applying the updates, generalising to a new custom VM and then updating the name in the API which creates new machines.

3.2.2 RESTful VM Management

To manage the array of VMs required to run multiple models simultaneously, software was also designed for a central management server. Flask is a micro-framework for Python which allows simple request despatching from HTTP requests. An API was designed to handle requests to create and delete virtual machines, using the Azure Python module (Figure 3-4). This allowed VMs to be created programmatically using the custom image and to delete themselves after a simulation had finished. The API also handles logging at various steps in the workflow so the progress of all VMs can be tracked in one central location. When multiple simulations are starting, it is very helpful to watch their setup progress from an SSH session into the Azure hosted management server to check for errors. As the Linux server is on the same subnet as the Windows images, it makes it possible to restrict requests to only that network, ensuring a basic level of security. When a request to create new machines is sent, the runs table is checked for non-started runs and if found, machines are created with names corresponding to the ID numbers of available runs. This enabled their hostname to be used when fetching the relevant setup information. The size of machines and a name prefix may also be specified. When a delete call is made, a new Python thread is started which first deletes the instance itself, followed by its public IP address, network interface and finally virtual hard

disk. A Linux management server was created on Azure to run the flask python web app as a Representational State Transfer (REST) service.

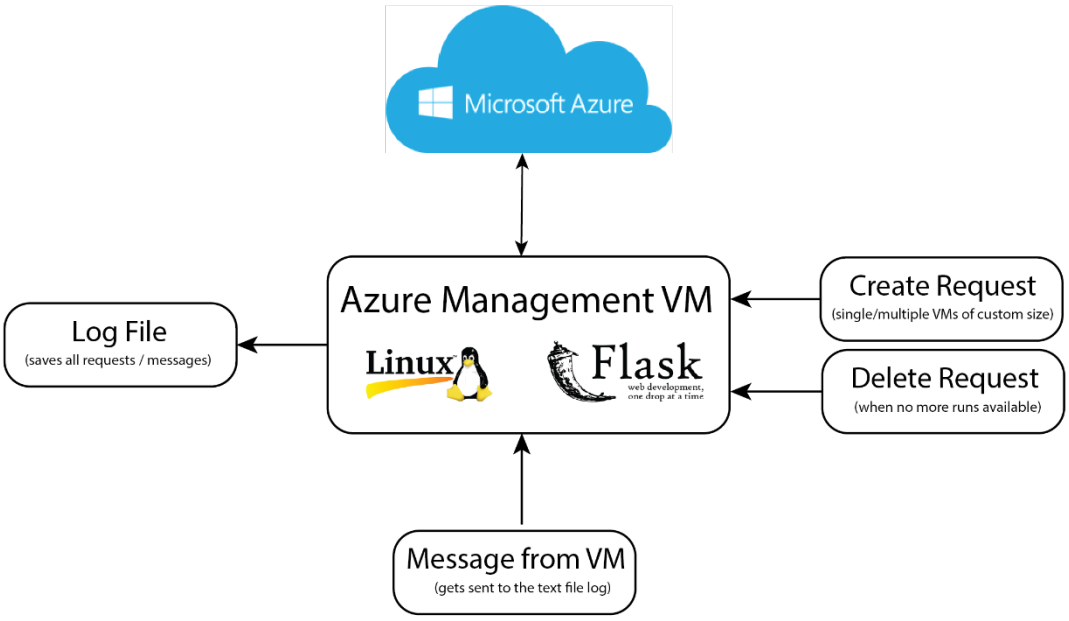

Figure 3-4 - RESTful VM Management with Flask and Microsoft Azure.

3.2.3 VM Workflow

Figure 3-5 illustrates the basic steps which happen when a new VM starts. The initiation of model setup on each new VM is done using Windows task scheduler, which runs a batch file at start up to pull code from a private GitHub repository and executes a main python script. The dotted line looping back to the beginning indicates that there is the option to run multiple simulations, one after the other, on a single machine. When VMs start in headless mode, the HOMEPATH environment variable is not set, which is required by GRASS (GIS software used in model setup), so this is also assigned in Python. If a new run is successfully retrieved, its record is updated with a run start time, commit ID (a reference to the version of code used to setup the model) and the name of the machine running the simulation. A folder is created on the VM to store the input and output files for the selected simulation. All of the required input data for the run is then downloaded from the database. Any pre-processing steps specified in

the setup parameters of the run are carried out and the hydrodynamic model starts as a sub-process. Once the simulation finishes (which can take anywhere from less than a day to multiple weeks), results are sent back to the same database for analysis and visualisation before the VMs delete themselves by sending a request to the Flask app. The approach of using an integrated Python module to carry out all modelling and pre/post-processing is highly adaptable and the GitHub repository can be updated without making any changes to anything on Azure. The generalised data and modelling workflow illustrated in Figure 3-1 will now be discussed in more detail.

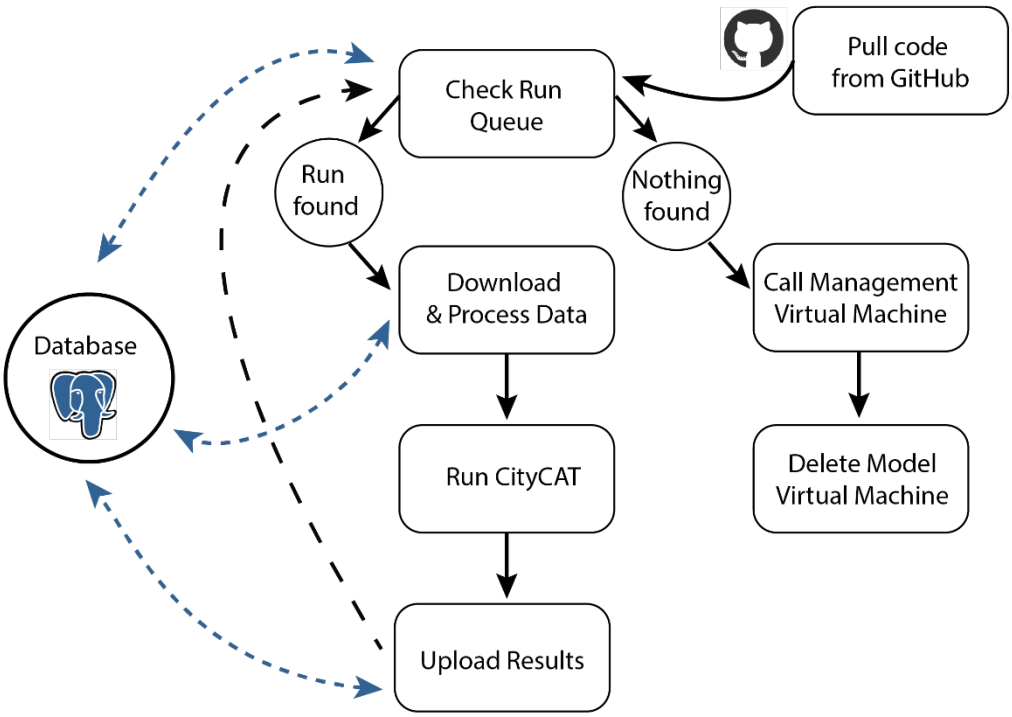

Figure 3-5 - Remote machine data and modelling workflow.

3.3 Pre-processing

Automatic selection, download and conversion of input data into the correct format is carried out by Python. Depending on the setup parameters stored in the runs table, additional processing stages can be implemented on the rainfall or DEM. The main processing pipeline is shown in Figure 3-6.

Figure 3-6 - Model input data pre-processing workflow.

3.3.1 DEM Download and Re-projection

After a run has been selected, the first stage is to download an original copy of the DEM as it's held in the database. The original coordinate reference system is maintained and a buffer is applied around the edge of the domain. This buffer allows filters to be applied to the DEM cells without incurring edge effects. DEM data is stored as 50x50 pixel tiles, allowing a spatial index to speed up extraction of the required area. Building of the entire DEM in memory on the database server was not possible for large domains, so the tiles instead get downloaded individually and combined on the client machine using a GDAL VRT. This ensures enough memory is available to build the entire DEM for large basins. To slice the DEMs for the domain, a basin outline is required. Basin outlines by default are provided by the HydroSHEDS project but custom basin outlines may also be specified or catchments upstream of a gauge can be used. The HydroSHEDS outlines were combined based on their main basin ID number to form catchment areas with an outlet at the sea. To specify which outline is used, a basin ID parameter is held in the runs table.

To allow a fixed cell size in metres for model runs, the original DEM is re-projected into either European Equal Area (3035) aligned with the projection datum or a specified coordinate reference system (CRS) for the area, e.g. UTM, held in the run parameters. This assumes the domain with terrain removed to be a flat surface, which is not the case in reality. As long as a suitable spatial projection with minimal area distortion is chosen, it is assumed that this does not significantly affect flood simulations. From this point onwards, all data is also re-projected into the specified CRS to match the DEM. The outline of the selected basin in the designated

CRS is then downloaded and saved to disk. This outline is used to crop the re-projected DEM with GDAL.

3.3.2 River Channels

A river network dataset, including approximate channel depths and widths, was required in order to enable testing of a further stream burning process for rivers which were not captured by the HydroSHEDS project. Andreadis et al (2013) provide a suitable Global River widths and depths Database (GRD), accessible in Shapefile format. Their method uses the relationship between channel profile and mean discharge to approximate the size of channels derived using a version of SRTM data. This channel data is downloaded for the selected basin to be used during model setup. Following is a more detailed explanation of the process.

3.3.3 Channel Burning and Smoothing

Although main channels are already burned in during the initial conditioning process of some DEM products, a further optional channel burning stage has been designed to also capture smaller channels. Firstly, the channels from the GRD which intersect the selected basin are converted from lines to polygons using their width estimate and saved to disk with a depth attribute attached to each polygon. As the channels from the GRD do not align precisely with the location where modelled channels appear in the re-projected DEM, a second channel network is created using a flow accumulation algorithm on the DEM. A least cost search algorithm within GRASS is used to calculate the number of upstream cells (Metz et al., 2011) and channels assigned based on an upstream area threshold of 1 km^2. The nearest depth values in the GRD channel network are then used to assign depths to the newly created DEM-based channel network. These values are subtracted from the DEM itself to burn in the channels. The process is summarised below:

a) Using GRASS, create a flow accumulation raster using the DEM
b) Rasterise the GRD rivers shape file using depth as values

c) Interpolate the GRD depth values across the entire domain
d) Create a DEM based rivers raster using an accumulation threshold
e) Save GRD depth values into DEM based rivers raster where channels exist
f) Subtract depths from DEM and output modified DEM raster

As another optional step, channels can be smoothed using a median filter with a specified window size which defaults to 3x3. This was designed to remove artificial blockages in channels such as bridges and any remaining noise, as shown in Figure 3-7. To select cells within channels, the flow accumulation raster was used. Although median filters don't ensure all blockages are completely removed, they can significantly improve results in areas where bridges cause major problems for flow routing (Jarihani et al., 2015).

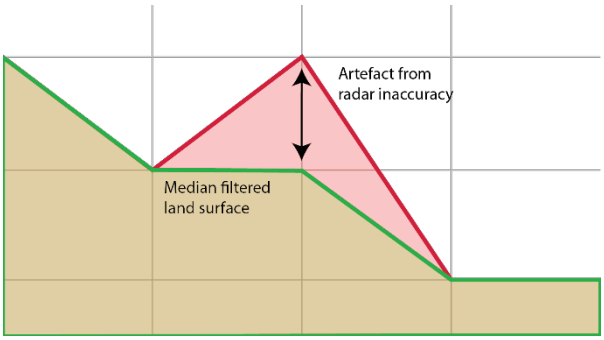

Figure 3-7 - Median filtering of elevations within channels.

3.3.4 Friction Coefficients

Friction coefficients are defined for cells within channels and on the flood plain. The two values are assigned within run parameters and fed to CityCAT in text format. The definition of a cell as within or outside of a channel is done using the flow accumulation raster with an upstream threshold of 1 km^2. When doing direct comparisons between DEMs, river channels were not included and channel friction was set to match flood plain friction to ensure the datasets were analysed in their original form before any further modifications or parameterisations were made. The default floodplain friction coefficient was set at 0.03 to correspond with the short grass pasture value from Chow (1959).

3.3.5 Rainfall Download and Interpolation

The rainfall table to use is defined within the run parameters and can be one of either ERA Interim, ERA5, MSWEP, E-OBS or MIDAS. In the case of MIDAS, it is interpolated to a 1 km grid using inverse distance weighting with a buffer of 50 km around the basin to select the relevant gauges. Both the perimeter for gauge selection and the resolution of the interpolation, along with an interpolation algorithm, are all designated within the run parameters. By default, inverse distance weighting is used. The time series and geographic locations for each grid cell are then written to disk in text format. Rainfall depths are converted to velocities using the period they were observed over.

3.4 Hydrodynamic Modelling (CityCAT)

The City Catchment Analysis Tool (CityCAT), developed by Glenis et al (2018) at Newcastle University, was designed primarily to simulate flooding in cities and solves the full shallow water equations (SWEs) numerically using a Generalised Osher-Solomon Riemann solver. The lack of compromise on accuracy is the main reason CityCAT was chosen for this study as this puts the focus on data rather than the model itself. The numerical methods used in CityCAT have been validated using analytical solutions (Glenis et al., 2018). This means the largest source of uncertainty is unquestionably from the input data and not any approximation in the numerical methods, as can be the case in modelling approaches which have been simplified to increase simulation speed. For this research, the interval between results outputs was set at 30 minutes, however the actual model time step varies depending on the speed of water and grid resolution to maintain stability. Cell wetting and drying is incorporated along with a shock capturing scheme to deal with hydraulic jumps.

3.4.1 Land Surface

CityCAT provides a number of other capabilities, such as excluding buildings, adding subterranean pipes, green rooves and infiltration, which are not used here for the following reasons. The 90m resolution of simulations means that removing cells intersecting buildings from the domain would result in large missing areas, therefore buildings are not removed. Pipes and infiltration are both excluded on the basis that reliable enough datasets are not readily available at the scales required. Incorporating infiltration would also add another parameter to calibrate and this does not align well with the minimal calibration philosophy of this thesis. This means that the land surface is effectively represented within CityCAT as an impermeable surface with the only way for water to leave the domain being by exiting via the open boundary at the edge of the basin. As a result of the exclusion of pipes and infiltration, it would be expected that all modelled discharge peaks would be overestimated. In general, however, this was not found to be the case, as is discussed in the following chapters.

3.4.2 Pluvial Modelling

CityCAT can be categorized as a pluvial flood model. All flood models either incorporate physically based simulations, statistical methods or a combination of both for estimating flows in ungauged catchments. One feature both GHMs and regionalisation have in common is they concentrate on river channels and route flow out onto the flood plain from there. Floods therefore cannot form in areas not directly connected to the main channel. When flooding occurs not connected to a channel, directly from rainfall without traversing a river system, it is often referred to as pluvial. Pluvial flooding is often ignored in broad-scale studies which usually focus on fluvial processes. Using a purely pluvial simulation not only cuts down the amount of data and parameterisation required, it also incorporates flooding which occurs away from channels.

3.4.3 Computational Demand

To meet the resulting computational demands of the complex solver, the software architecture of CityCAT has been designed to distribute calculations across an unlimited number of CPU cores. However, reading and writing to memory is still a bottleneck. This means that after a certain number of cores, there is no speed benefit to be gained by adding more as the previous and next time steps cannot be read and written fast enough. Therefore, an approach to split domains into multiple upstream catchments feeding into separate simulations downstream was tested. However, challenges with getting time steps to match up and the requirement for flow between domains at the outlets meant this could not be implemented effectively. It was therefore concluded that entire basins had to be simulated as one domain to ensure no extraneous effects were caused at artificial boundaries.

3.4.4 Spin-up period

To ensure water is flowing in a realistic way throughout the basin before the main rainfall event occurs, the modelled period is extended to start before the beginning of the main event. The period of time simulated before the main event is known as the spin-up period. The required spin-up period depends on how much rainfall occurs leading up the main peak and on the size of the basin itself. The spin-up period dictates the total model period. An acceptable total model period for the size of catchments and amount of rainfall was found to be 1 month. To arrive at this duration, simulations of 2 weeks, 3 weeks and 2 months were assessed. At the gauge located on the Tyne at Bywell, the 2 and 3 week runs produced lower peaks than the 1 month simulation, while the 2 month run produced an equivalent peak to the 1 month simulation. This led to the conclusion that 1 month is a suitable model period for catchments with similar characteristics to the Tyne and with a similar rainfall series. For larger catchments, or where there is longer duration lower, intensity rainfall leading up to the main peak, a longer model period may be required to produce a representative flood peak.

Initial conditions were tested in an attempt to reduce the required spin-up period, however this would lead to calibration based on initial conditions which was avoided throughout the study on the principle that local calibration is not possible at large scales. When looking at return period events, a spin-up period is also not available as no previous data exists for an artificial event. Therefore, a big advantage of simulating past events is that a realistic spin-up period can be simulated.

3.5 Post-processing

Once simulations are complete, the results are converted into SQL for input to the database. The main steps involved are shown in Figure 3-8. To translate the text file outputs into postgres rows, geometries are first created from the first output file, then depth and velocities are read from all files. The time series for each geometry is stacked and stored into a Pandas data frame containing all cells. To ensure correct ordering of time steps, a natural sort was applied to the results files in Python. The time series are then sent as postgres numeric arrays and the cell geometries as PostGIS polygons into unique rows, along with their run number. As the number of cells is often large for a single query, they are first chunked into sets of 100. The advantage of using PostGIS geometries and arrays is that GeoServer can readily read this data to produce web maps with time and run number filters. In situations where the results dataset is too large to fit into memory, cells are split into groups and one group at a time is read from disk and sent to the database. The number of cells in each group can be defined by the amount of memory available on disk.

In order to automate the validation process, objects are included in the main python library which can be used to compare model results with both gauge data and satellite extent observations. When a run completes, performance indicators relating to discharge and extent are calculated and stored within the metrics table of the database. Metrics are then able to

be read by the web application when viewing results. Further details about the validation methods used are provided in Chapter 4.

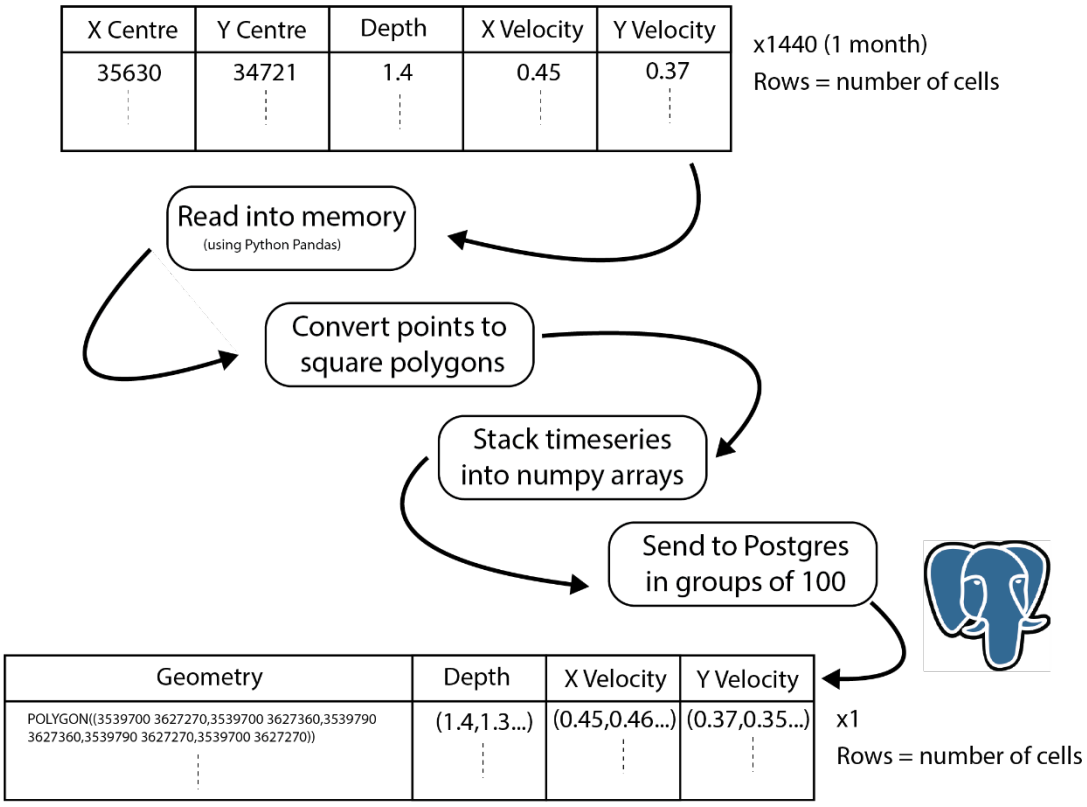

Figure 3-8 - Converting CityCAT results to a suitable database schema.

3.6 Visualisation

No currently available software was able to display model outputs in a way that was easily navigable and efficient. Therefore, a web application was designed to display and analyse model results, as pictured in Figure 3-9. The app uses the Python Tornado web framework to connect directly into the main database and retrieve spatial data. The Leaflet JavaScript library is used to display an interactive web map. The Angular library was used to connect JavaScript components with the html page. A sidebar, designed with Angular Material layout, provides sliders and selection options to navigate between runs and time steps

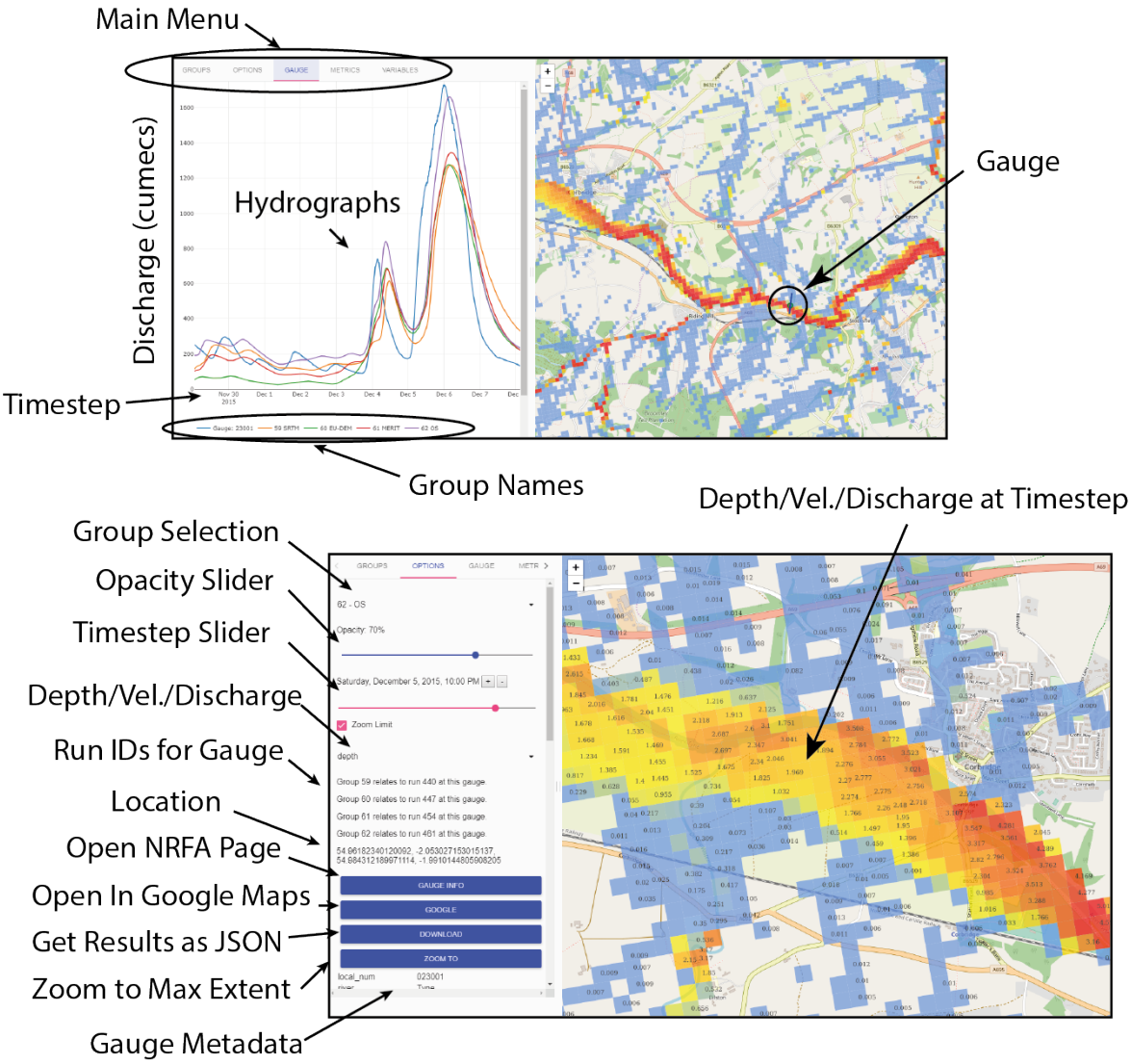

Figure 3-9 - Web application designed in AngularJS, using GeoServer, Leaflet and a Tornado backend.

To visualise individual cells, sending all of the spatial information to a web browser would be too much for it to handle, therefore GeoServer is used as an intermediary. GeoServer renders tiles of cells as images which are then returned to the browser as a Leaflet WMS layer. Parameterised SQL views are used to access run groups in the results table at given time steps. Style parameters may also be passed to GeoServer to render the cells with different colour scales. Performance metrics at gauges can be displayed with a proportional symbol radius. The web application is served from a Linux virtual machine hosted at Newcastle University running Ubuntu 16.04.3.

3.7 Summary

This chapter has laid out the automated computational framework for the rest of the thesis. It makes use of on demand high performance computing from Microsoft Azure and a highly supported and feature-rich geospatial relational database. Visualisation has been made possible using an interactive web-based application, connected to the database through an API designed specifically for this research. The framework has been designed here specifically for flood modelling applications, however with some generalisation could be applied to any environmental modelling. The distributed nature of the framework allows numbers of simultaneous computationally demanding simulations limited only by the monetary budget of the project. The following chapters use this framework to simulate floods across a range of areas and analyse the efficacy of various input datasets and parameters.

Chapter 4. Validation using Discharge and Extent Observations

Validation is vital to understand if a model is providing realistic outputs or if unanticipated factors are causing large errors. As seen in Chapter 2, the term validation has been used in the literature with very loose specifications. Models are often compared to results from other models using similar or the same input data and similarity is seen as an indication of good performance. Certain metrics, such as hit rate, can also provide a highly misleading view of model skill. This chapter aims to develop a transparent and representative approach to validation, using a range of real observations and metrics to describe the performance of the workflow outlined in Chapter 3. Both channel flows and extents are assessed as neither provide the entire picture on their own and could result in major flaws being missed if not combined with the other. To ensure calibration does not give an over-optimistic view of performance, model parameters also remain identical throughout this chapter, with a Manning's friction coefficient of 0.03 for the entire domain (Section 3.3.4).

4.1 Model Inputs

To produce simulation results for workflow validation, input data was required. At this stage, it is primarily the 2D model being tested as it has not been used at 90m resolution before, having been designed for 1-5m grids. Therefore the choice of input data is not highly consequential here and a more in depth look at a wider range of possible datasets is provided in Chapter 5.

4.1.1 Met Office Gauges (MIDAS)

As it was considered important to get the rainfall element as accurate as possible for useful initial validation results, it was decided that the best available local data should be used. The UK holds some of the longest and most reliable rainfall records globally, stored in the Met Office Integrated Data Archive System (MIDAS), many reaching as far back as 1959 at daily

resolution, with the longest record starting in 1853. Initially, readings were made manually, but since 1970, automation has become increasingly prevalent and now almost every aspect of the process is automated. Manual checks by observers are still made though, to ensure gauges remain functional and assess any maintenance requirements. All instruments are calibrated to international standards.

There are daily, hourly and sub-hourly datasets available. As temporal resolution is important for capturing flood peak magnitude and timing, the daily product was excluded, despite its much longer records and more extensive availability. The sub-hourly dataset, was too limited in availability for the study, therefore, with an average spacing of less than 20 km as shown in Figure 4-1, the hourly dataset was selected. MIDAS Hourly consists of records describing rainfall accumulation to the nearest 0.1 mm, predominantly from the Basic Synoptic Network. Also included is information about event duration, version number and a quality control flag. A gridded product already exists from the Centre for Ecology and Hydrology (CEH), called Gridded Estimates of Areal Rainfall (GEAR), however the temporal resolution is only daily. Therefore, interpolation is carried out on the hourly MIDAS data using Inverse Distance Weighting (IDW) to create a gridded product at the same spatial resolution as CEH-GEAR (1 km). The effect of rainfall resolution is discussed in Section 5.1.3 with results from simulations using MIDAS data interpolated at different resolutions. To select gauges to include in each basin's interpolation, a buffer of 0.5° (~55 km) was used. Records with a version number not equal to 1 and/or accumulation period not equal to 1 hour were excluded.

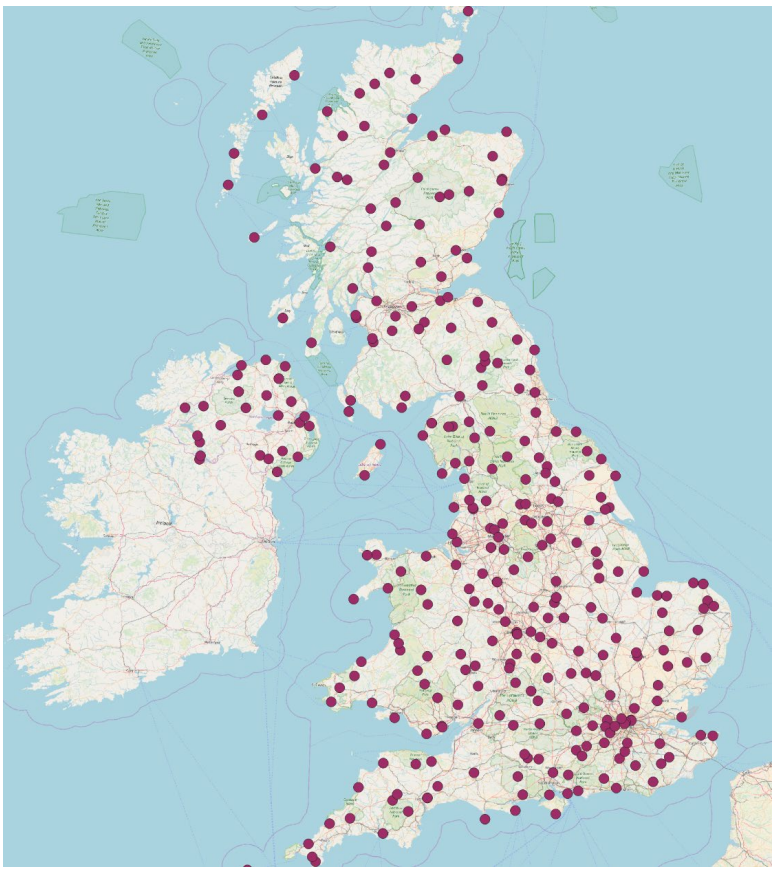

Figure 4-1 - Location of MIDAS hourly rain gauges around the UK (Met Office, OpenStreetMap).

4.1.2 HydroSHEDS

The choice of DEM dataset for validation was based on its global availability, common use amongst global flood studies and included hydrological conditioning. HydroSHEDS is based on SRTM with a number of processes applied to make the original data more suitable for hydrological modelling (Figure 4-2) (Lehner et al., 2013). A major processing component was to burn in large rivers and lakes contained in ESRI's ArcWorld dataset and the Global Lakes and Wetlands Database (GWLD) by Lehner and Döll (2004). All rivers were burned with a depth of 12m along the centre line, reducing to 2m at the edge of a 0.005 degree (~500m) buffer area. Only large rivers are included in ArcWorld, therefore many smaller channels remain unaffected. All lake polygons were reduced in elevation by 14m with a 0.0025 degree buffer. Rivers and lakes contained in the SRTM Water Body Data (SWBD) were also lowered by 10m. Other filters were applied to prevent coastal mangroves blocking flow, remove spikes and

wells, model valleys, fill sinks and carve through barriers in main river channels. In the carving process, channels with more than 1000 upstream cells were selected and all rising reaches along their courses flattened. Some areas were manually corrected using a generated river network to highlight spurious flow pathways.

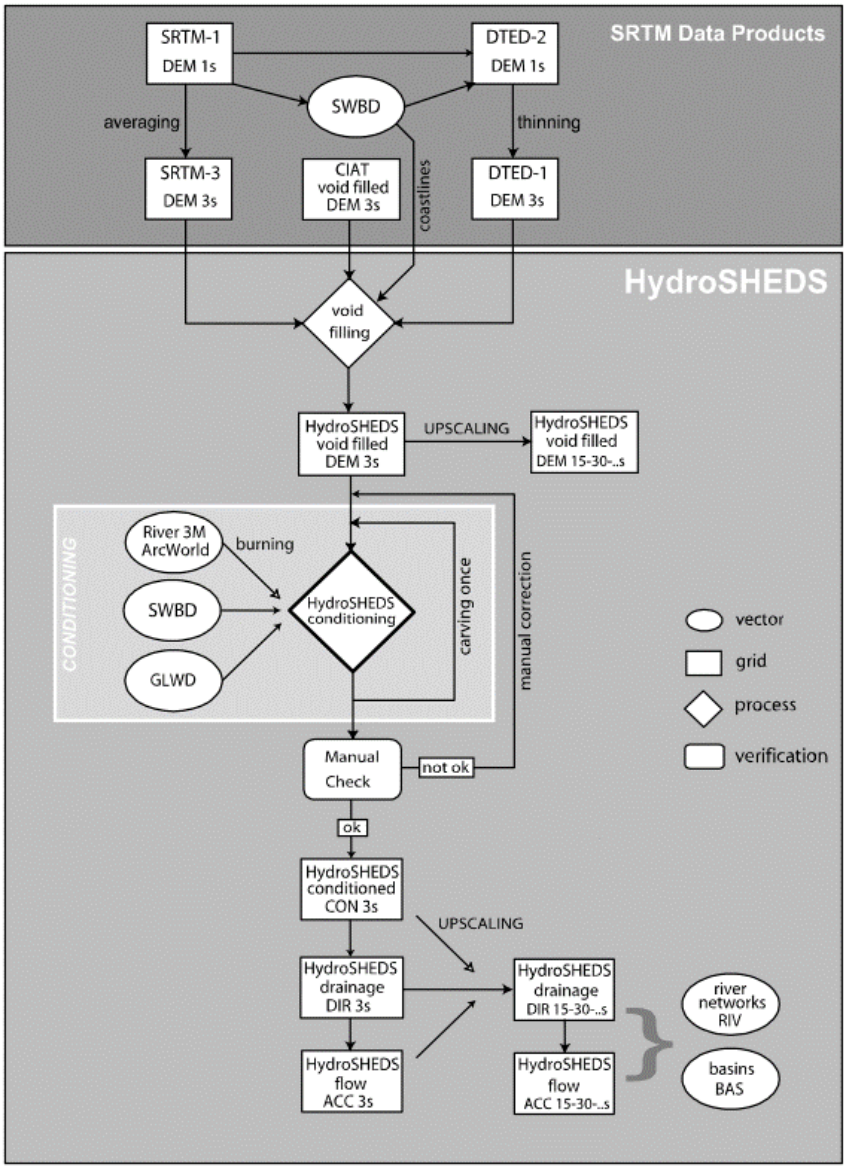

Figure 4-2 - HydroSHEDS workflow (Lehner et al., 2013).

4.2 Study Area

To validate the model and later test different combinations of datasets and setup parameters, a set of basins was required. The choice of region was based both on availability and quality of data, the size of the constituent basins and occurrence of recent broad-scale flooding.

Northern England has a high density of good quality river and rainfall gauges which are readily available at high temporal resolution and the basins are of a manageable size to model hydro-dynamically. Therefore a group of basins with sizes between 1-10 thousand km^2 was selected in this area, shown in Table 4. These basins cover the largest rivers in Northern England and have seen significant flooding in recent years. The basin boundaries are defined within the global HydroBASINS dataset (Lehner & Grill, 2014), with ID numbers listed in Table 4.

Table 4 – Selected basins in Northern England.

BASIN NAME	HYDROSHEDS BASIN ID	AREA (KM^2)
Ouse	2120053240	9350
Tyne	2120053050	2892
Eden	2120050770	2343
Ribble	2120050460	1908
Tees	2120053100	1827
Wear	2120053070	1181
Lune	2120050510	1067

4.3 Storm Desmond

The standard approach to precipitation when modelling floods is to generate uniform return period events based on statistical analysis of historical data. This aligns well with building design criteria, however does not recreate a realistic broad-scale event as rainfall never occurs uniformly across large areas. Therefore, applying a uniform rainfall will inevitably result in a misrepresentation of the distribution of widespread impacts. However, complex localised convective processes mean weather patterns are different in each event and extremely difficult to predict long term. This makes generating a return period spatial rainfall event impossible, leaving two other options available. Either an ensemble of realistic events can be generated stochastically or historical data can be used to recreate past events. The latter option provides the advantage of allowing real independent observations to be used to validate the outputs of simulations using spatial rainfall, which would not be possible when creating artificial fields. Therefore, this is the approach that has been taken here to identify

and explore the issues surrounding broad-scale fluvial and pluvial flood modelling and validation.

The selection of a rainfall event was limited by the period of hourly MIDAS data available, with many records only beginning in 2010. As this hourly dataset was required as a benchmark, any events occurring before MIDAS observations were widely available had to be discounted. This left a clear event choice and Storm Desmond (named by the UK Met Office) was selected for both its magnitude and large spatial area of influence. A combination of a previously wet period and an intense burst of rainfall across much of Northern England in November-December 2015 caused widespread flooding from both surface water and fluvial sources. The recent timing also makes Desmond relatable for many people who can remember the news stories and may have been directly or indirectly impacted both during and after the event. The recovery process was still ongoing 2 years after the storm (Figure 4-3). At many river flow gauges, the event caused some of the largest flows on record, with the highest ever mean weekly flow observed at the Tyne at Bywell, as shown in Figure 4-4.

The MIDAS dataset provides high spatial and temporal resolution rainfall data for the event and is subject to rigorous quality control checks. However, despite its high quality, MIDAS hourly is clearly not equivalent to the truth as factors such as under-catch, interpolation methodology and gauge density all affect the accuracy of values. This became clear in the initial simulation results presented later in this chapter. However, MIDAS remains the best openly available UK dataset with recordings of the selected event and therefore it was used to analyse the storm itself.

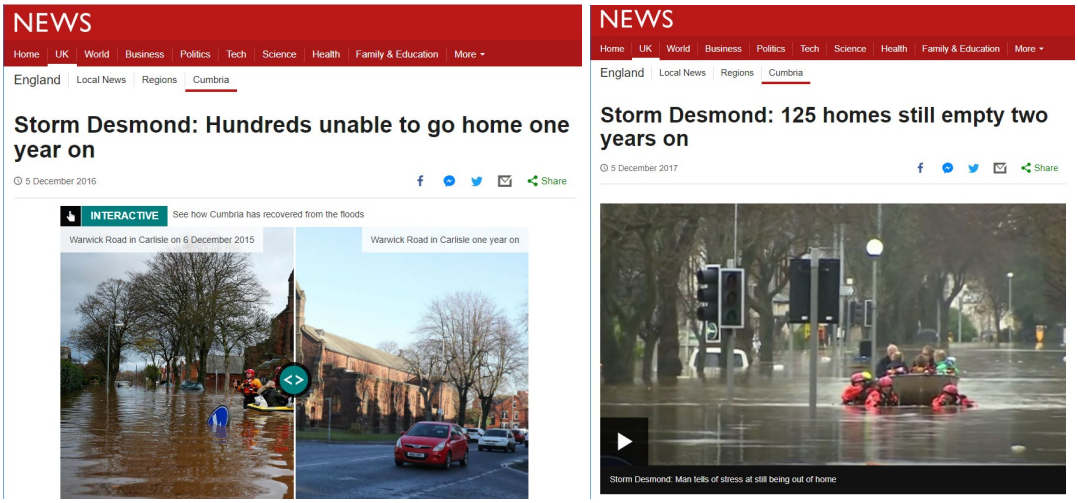

Figure 4-3 – Aftermath of Storm Desmond (BBC, 2015, 2017).

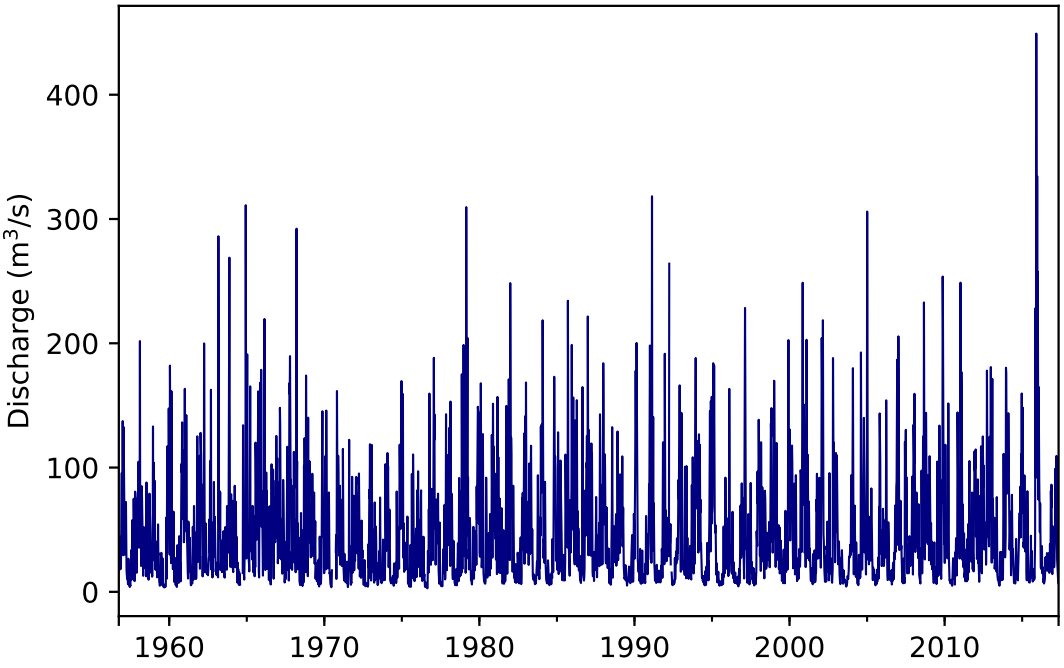

Figure 4-4 – Weekly mean flow at the Tyne @ Bywell.

A spatially averaged time series of the interpolated rainfall for each basin within the study area is shown in Figure 4-5. The Lune and Eden basins saw the largest amount of rainfall, reaching nearly 6 mm/hr. The Tyne received less rainfall during the main event on 5-6 Dec, however is subject to the second largest burst of rainfall in the preceding shorter downpour

on 3 Dec. The Ouse basin had the lowest rainfall intensity during the period, however its much larger catchment area means this equates to a greater total volume of water entering the system. Spatially, the event was mainly concentrated in the west of the study area, as shown in Figure 4-6. Figure 4-6 also demonstrates that interpolation causes unrealistic rainfall patterns, with intensities gradually increasing away from low points in the Ouse basin which would not generally be the case in reality. Spatial patterns could feasibly be improved with the use of radar imagery, however this is outside the scope of the current study.

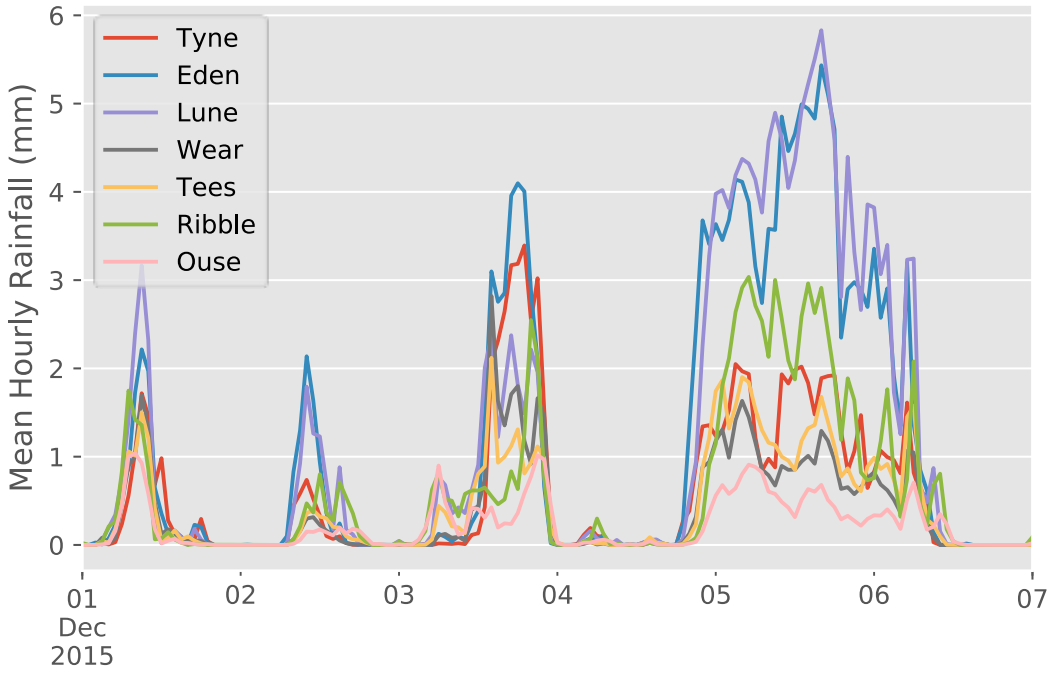

Figure 4-5 - Mean MIDAS rainfall values for each basin during the main storm period (mm/hr).

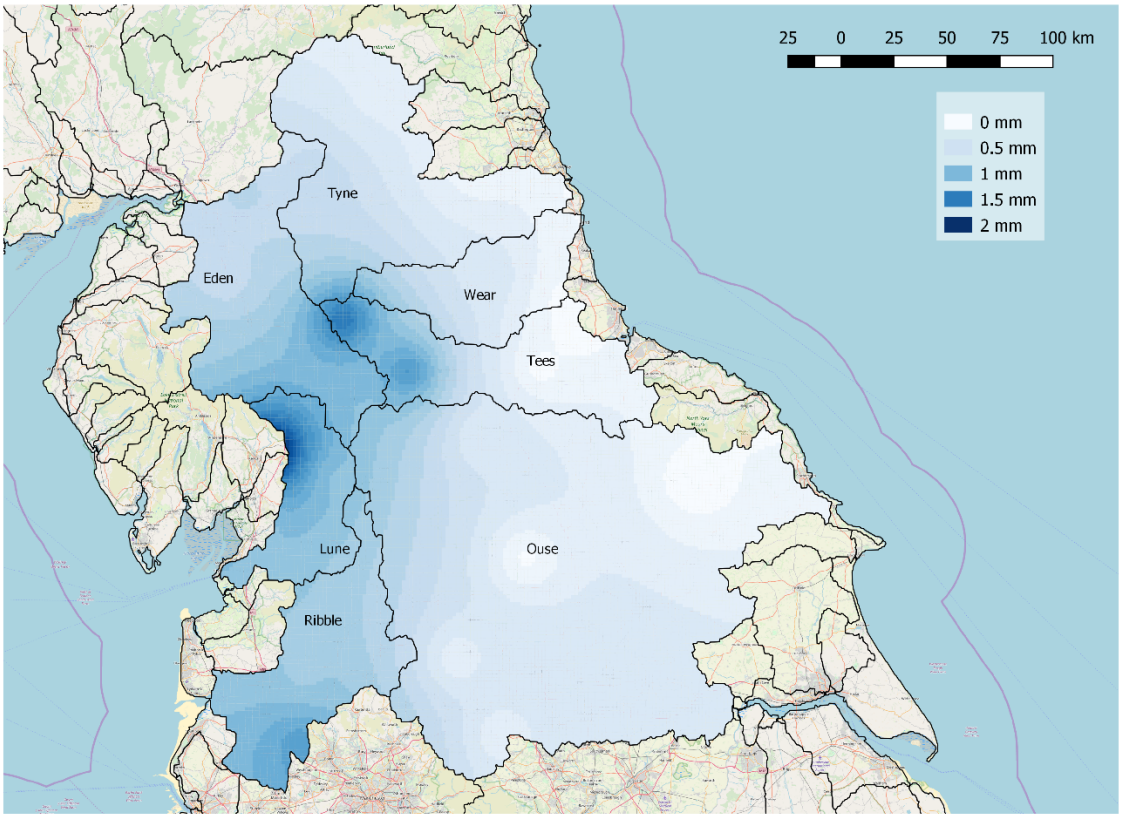

Figure 4-6 - Spatial distribution of mean hourly rainfall around the study area, interpolated to 1 km spatial resolution, for the period 1-7 December 2015.

4.3.1 Water Balance

As another assessment of the MIDAS data and the event itself, water balance ratios were calculated based on observed discharge and rainfall. The period from 12 November – 12 December was used to calculate ratios between observed rainfall and discharge in mm. To convert the discharge from m^3/s to mm, values were divided by the appropriate catchment area in m^2 and multiplied by the time period between observations in seconds. Catchment area values were obtained from the National River Flow Archive (NRFA). Rainfall cells which intersected the upstream area of each gauge were used to calculate total inputs and outputs of water for the selected period.

Figure 4-7 compares the total values for each flow gauge, coloured by their main basin. The Lune basin saw the greatest intensity of rainfall, while the River Eden produced the largest

total output of water for its size. The Ouse basin has some of the lowest input and output volumes while discharge tends to exceed rainfall to a greater degree than other catchments.

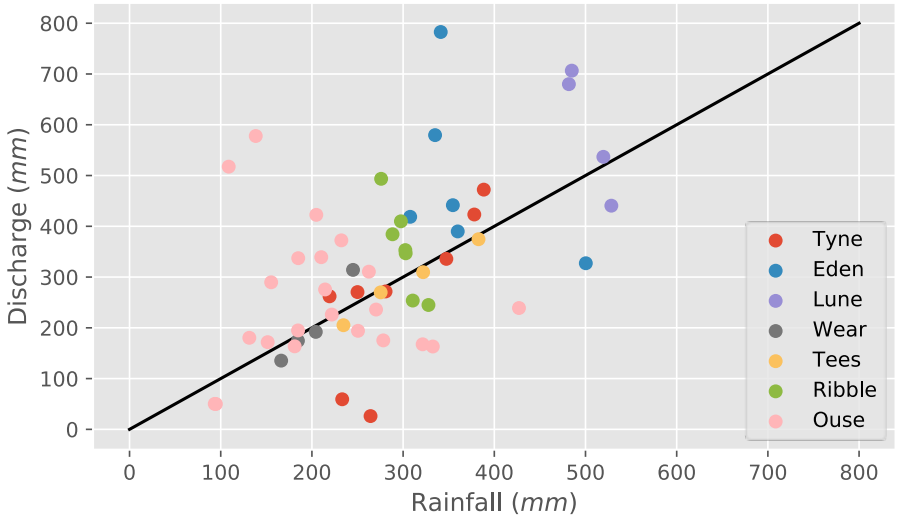

Figure 4-7 - Observed mass balance plot for all gauges in the catchment using EA flow data and interpolated MIDAS rainfall.

Table 5 includes the calculated ratios for each gauge, ordered by catchment area. Many ratios are over 1, indicating more water left the catchment than entered it. This is most likely to be caused by both rainfall and flow gauging errors. The interpolation is also limited to the number of gauges available in the area. Figure 4-8 and Table 6 show the number of rainfall gauges available in each basin per area, including the ~55 km buffer. The densest gauge network occurs over the Lune basin, coinciding with the highest intensity of rainfall. The Ouse has the most gauges overall but the least dense network due to its greater area. This also aligns with the lowest amount of rainfall input. The relationship between number of gauges present and intensity of rainfall could be linked to small patches of intense precipitation having a greater chance of falling in areas without gauges when network density is lower, therefore causing an under-representation of the true amount of rainfall.

Table 5 - Mass balance of largest 20 catchments within the study area, sorted by water balance ratio. Area is in km² and both rainfall and discharge are in mm.

Basin	River	Location	Area	Rainfall	Discharge	Ratio
Ouse	Ure	Westwick Lock	914.6	186.6	337.2	1.81
Ouse	Swale	Crakehill	1363	132.5	180.3	1.36
Ribble	Ribble	Samlesbury	1145	304.4	353.1	1.16
Ribble	Ribble	New Jumbles Rock	1053	304.9	346.8	1.14
Ouse	Ouse	Skelton	3315	152.6	171.8	1.13
Eden	Eden	Great Corby	1373	356.6	389.9	1.09
Tyne	North Tyne	Reaverhill	1007.5	250	270.4	1.08
Lune	Lune	Caton	983	522.7	536.8	1.03
Ouse	Wharfe	Tadcaster	818	222.1	226.1	1.02
Tees	Tees	Broken Scar	818.4	272.1	269.9	0.99
Tyne	Tyne	Bywell	2175.6	281.6	271.6	0.96
Tyne	South Tyne	Haydon Bridge	751.1	348.8	335.8	0.96
Tees	Tees	Low Moor	1264	232	205.1	0.88
Ouse	Aire	Armley	691.5	273.3	236	0.86
Wear	Wear	Chester le Street	1008.3	165.8	135.4	0.82
Ouse	Aire	Lemonroyd	865	252.9	194.1	0.77
Ouse	Aire	Beal Weir	1932.1	279	175.2	0.63
Ouse	Derwent	Malton A64 Road Bridge	1360.1	93.9	50.2	0.53
Ouse	Derwent	Buttercrambe	1586	95.1	49.9	0.52
Ouse	Calder	Methley	930	320.7	167.5	0.52

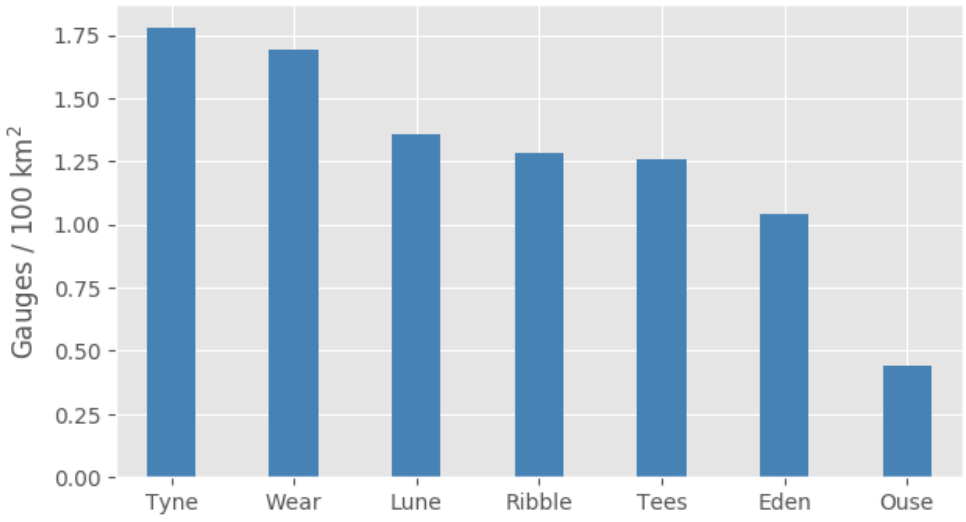

Figure 4-8 - Density of available gauges in each basin.

Table 6 - Number of MIDAS rainfall gauges used in creating grids for each basin.

Basin	Tyne	Wear	Lune	Ribble	Tees	Eden	Ouse
Area (km^2)	1181	1067	1908	2343	1827	2892	9350
MIDAS Gauges Available	21	18	26	30	23	30	41
Gauges available per 100 km^2	1.78	1.69	1.36	1.28	1.26	1.04	0.44

9 out of 20 stations in Table 5 present a larger amount of discharge than rainfall. Two possible explanations for this are that either rainfall is not being fully captured by the gauge network and interpolation has caused an under-estimation, or discharge is being over-estimated by the flow gauges. As incorrect gauge readings tend to under- rather than over-estimate values of either flow or rainfall, the first option is more convincing. The period selected for calculating a water balance can also affect the ratio values as previous rainfall may not reach the gauge until after the selected time period. Meanwhile, not all rainfall falling on the catchment will reach the gauge before the end of the period. If these two volumes are not balanced, then the ratio will be influenced by processes outside of the period, causing misleading results. Ideally, a period would be selected which begins after a considerable duration of very low rainfall and with very low rainfall towards the end to allow all inputs to pass the flow gauge. However, this is almost always not practical as large rainfall events generally occur during wet seasons. Therefore, the effect should be considered when interpreting results.

4.4 Validation Methodology

4.4.1 Discharge

Comparison of modelled and observed discharge at flow gauges was the primary method used for validation. This required a discharge time series to be extracted from the 2D model results. As channels are not explicitly represented in CityCAT, a process was designed for selecting appropriate cells and extracting a combined discharge series. The main steps required are shown in Figure 4-9.

Step 1 involves creating a line, termed a virtual gauge, perpendicular to the channel and reaching across its width. Virtual gauges are positioned close to existing flow gauges so series can be directly compared. Initially an automated method of virtual gauge creation was tested, however due to the variety of complications in gauge locations, this was not always effective. Therefore a tool to create them manually has been integrated into the web interface. At each gauge, they were placed across the simulated channel, close to the given location. Virtual gauges are stored in their own table within the database. To prevent discharge being included in opposing directions, the sum of velocities in the x and y directions is first calculated and its sign used to determine the dominant flow direction for each axis. Each cell is then divided into its x and y components and components which do not match the dominant direction are filtered out. To filter out components which are flowing into other selected cells, a point is created in the direction of flow one width from its centroid. This point is checked against all of the selected cells to make sure it does not intersect. This leaves only cells bordering the downstream flux boundary, as shown in Figure 4-10. After filtering, the remaining cell depths are multiplied by the grid resolution and the velocities in the given direction to produce discharge time series.

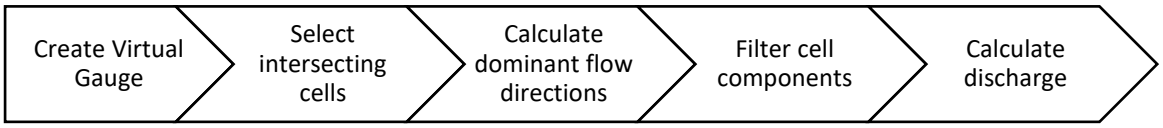

Figure 4-9 - Discharge extraction workflow.

Placement of virtual gauges has an effect on the final discharge output and there is always some leeway in where they should reasonably be placed. This means there is some degree of uncertainty in the modelled discharge series related to the placement of the virtual gauge, however the approach used was the only viable option for extracting a time series.

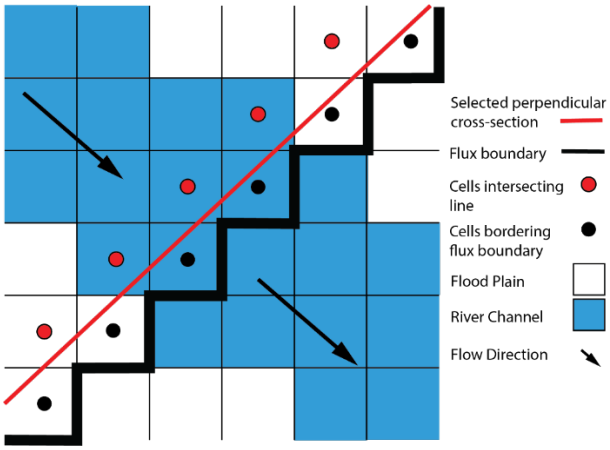

Figure 4-10 - Diagram showing a hypothetical case of the discharge extraction method.

Only daily river flows are available from the National River Flow Archives (NRFA) which was not frequent enough to validate 30 minute model results. Therefore a freedom of information request was submitted to the Environment Agency (EA) for around 100 UK gauges with 15 minute resolution. Extracted discharge time series were compared with this data to assess model accuracy. To select the gauges with available data, the minimum and maximum available gauge time step is compared with the run start and end times.

It is not practical to examine similar plots for all gauges and parameter sets and it would be difficult to identify systematic trends. Therefore, performance measures are required. Performance measures, or metrics, such as the Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) and Kling-Gupta Efficiency (KGE) (Gupta et al., 2009) are widely used across all areas of hydrology, though are not necessarily suited to looking at large flood events (Schaefl & Gupta, 2007). As a more suitable alternative, flow peak and timing errors are calculated to provide a simulation performance summary directly related to the largest flood wave. The main limitation in comparing modelled and observed peaks is that if a greater peak occurs in the modelled sequence, far from the maximum observed value, the timing error will be over-estimated and the wrong peak values will be compared. This is not a common problem, but

can occur when there is not a single major peak. The effect is discussed further when interpreting results.

4.4.2 Extent

Flood extent is difficult to measure. Satellites make only intermittent passes and produce outlines which in themselves are uncertain and difficult to verify, particularly in urban areas. However, they are the only coherent source of available data for observed extent. Synthetic Aperture Radar (SAR) data from the Sentinel 1 satellite constellation, processed using change-detection to produce flood extents, was used for validation of the modelled extents (Clement et al., 2018). This approach is an ongoing area of research and therefore is subject to uncertainty, however is able to produce observations with a high frequency compared with other methods as it is not hindered by cloud presence.

The method used for creating extents from imagery involves taking the difference in backscatter between a baseline image with normal water extent and an image with flooding present. Any areas in the flooded image with a negative change in backscatter above a certain threshold are classified as flooded based on the premise that water will produce a more specular reflection and therefore a large reduction in backscatter. Any regions more than 20 m above nearest drainage or with more than a 3° slope were assumed not to be flooded to account for radar shadow. River channels are not shown as flooded in the results as they exhibit no difference between images from flooded and non-flooded periods. However, in model results, all river channels and lakes are represented as flooded. If the change-detection data was used by itself for validation, performance scores would be underestimated and lakes and river channels would be designated as false alarms. Therefore, water body locations from OS VectorMapLocal were used to also define water bodies as having been flooded.

There is a difference in resolution between model results and satellite observations, with Sentinel 1 producing outputs on a 10m grid. To reconcile this difference, a model cell is counted as being observed as flooded if at least 50% of its area is covered by satellite-observed flooded cells. As Sentinel 1 is unlikely to pass over at the correct moment to capture the maximum flood extent, the satellite observations are also likely to be an under-estimation. Due to the way a surface is defined as water by the Sentinel 1 processing method, floods on flat surfaces such as airports and roads may never be detected as they are flat in both flooded and non-flooded cases. This is another factor which may cause under-estimation in the observed extents.

To extract maximum extent from the satellite observations, all polygons assigned as flooded across every time step within the modelled period are compared with cells containing maximum simulated depth. The threshold for model cells being designated as inundated was 30cm, as the nature of directly adding rainfall to the domain surface means that using zero would lead to massive overestimation of flood extent. This threshold was also chosen based on the literature review in Section 2.3.5, which found 30cm to be a maximum depth for roads to remain passable (Pregnolato et al., 2017). A range of hazard layers can be produced based on different depth thresholds to suit a given application. As one of the proposed applications for hazard layers produced here is road inundation (Section 6.1.3), 30cm was also used to generate the layers used in validation to ensure that the performance scores were relevant.

Due to inherent limitations in the Sentinel 1 observations, polygons from the EA Flood Zone 2 dataset (Environment Agency, 2019), including a combination of high resolution local model outputs and observations, were also used for extent validation. This additional analysis provides an illustration of how much better performance may appear when results from another model are used instead of real observations. The Flood Zone 2, Areas at Risk from

Rivers and Sea, dataset comprises a number of categories of polygons, including Tidal Models, Tidal Events, Fluvial Models, Fluvial Events, undefined and various combinations of these. As coastal flooding is not taken account of in this thesis, only polygons with a category containing the word Fluvial were extracted for validation purposes. The same methods as those used for Sentinel 1 data are implemented to deal with resolution inconsistency. As use of a local benchmark model is common practice in the literature, using EA Flood Zone 2 provides further context for the performance metrics calculated using Sentinel 1 data.

The extent performance metrics used are summarised in Table 7 and include false alarm ratio (FAR), hit rate (HR) and critical success index (CSI) (Donaldson et al., 1975). The CSI is the most appropriate indicator for the purpose as it assesses both the ability to flood flooded cells and keep non-flooded cells clear. As discussed in Chapter 2, although HR may be high, it is a misleading measure as if the entire domain is flooded by the model, this would result in a HR of 100%. HR and FAR are included as they provide a better understanding of what is contributing to the CSI score, for example if both HR and FAR and low, the model is under-predicting extent or if they are both high, extent is being over-estimated. To achieve a good CSI score, a combination of high HR and low FAR are required. Cells correctly identified as non-flooded are ignored by the metrics used here as they are not important (Table 8).

Table 7 - Extent metric equations based on totals from Table 8 (Donaldson et al., 1975).

METRIC	EQUATION
Hit Rate (HR)	hits / (hits + misses)
False Alarm Ratio (FAR)	false alarms / (hits + false alarms)
Critical Success Index (CSI)	hits / (hits + misses + false alarms)

Table 8 - Error/confusion matrix for calculating CSI, HR and FAR (Stehman, 1997).

		Model	
		Flooded	**Non-flooded**
Benchmark	**Flooded**	Hits	Misses
	Non-flooded	False alarms	Ignored

4.5 Results

4.5.1 Gauge Validation

To assess the performance of the modelling workflow from Chapter 3 across all available gauges for Storm Desmond in the study area, peak flow (Q_{max}), extracted from gauge observations was plotted against modelled Q_{max} in Figure 4-11. The largest observed Q_{max} was on the Lune at Caton, upstream of Lancaster, with a peak of 1740 m^3/s, followed closely by the Tyne at Bywell (1730 m^3/s) and Eden at Great Corby (1490 m^3/s).

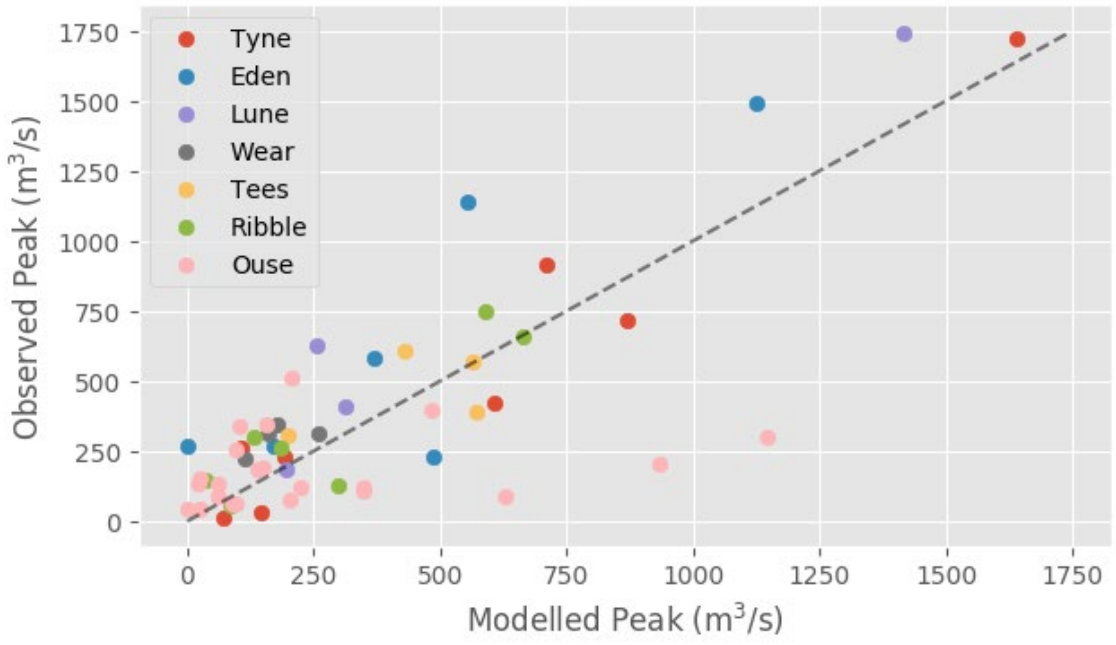

Figure 4-11 - Observed vs modelled peaks from runs with interpolated MIDAS rainfall using the HydroSHEDS DEM.

Despite its size, the Ouse basin displays the lowest observed peaks, which are grossly over-estimated by the simulations. Possible explanations for this may include the flow gauges in this area not being well designed to cope with extreme flows or the relatively low density of rainfall gauges, as shown in Figure 4-8. Lower flows are also associated with smaller catchment areas, which means that fewer rain grid cells influence the results, increasing sensitivity to spurious variations in precipitation inputs. Furthermore, the Ouse basin is more urbanised than the other regions, which could contribute to external factors such as water abstraction which was not included in the model. Urbanisation can also affect the DEM, causing artefacts which may slow down or impede runoff. Other causes of incorrect peaks include the location of reservoirs in the catchment which result in processes not captured by the model, particularly at reservoirs with controlled outlets and scheduled releases. The model does not include infiltration either, which may be a major factor in the Ouse basin, leading to lower observed flows as more water is stored beneath the surface.

Generally, model performance is shown to improve with Q_{max}. This was expected, due to the reducing proportion of rainfall and DEM grid cell size to catchment area in larger catchments. Larger catchment areas also naturally produce higher peaks and at these high flows, hydrodynamics become increasingly influential relative to other ongoing processes such as infiltration and evaporation.

To ensure the results are not excessively influenced by the erroneous effects of DEM grid and rainfall resolution and focus on locations with higher flows, gauges with Q_{obs} of less than 550 m^3/s were excluded from the analysis. This flow threshold was decided upon based on an optimisation between mean peak error and number of gauges, as shown in Figure 4-12. Other threshold variables were considered but area was not well correlated with the observed peaks and the gauge with the fourth largest peak had a disproportionately low QMED, as shown in

Figure 4-13. Using a 550 m^3/s peak threshold left 12 flow gauges, listed in Table 9, for comparison with observed flow data in further analysis.

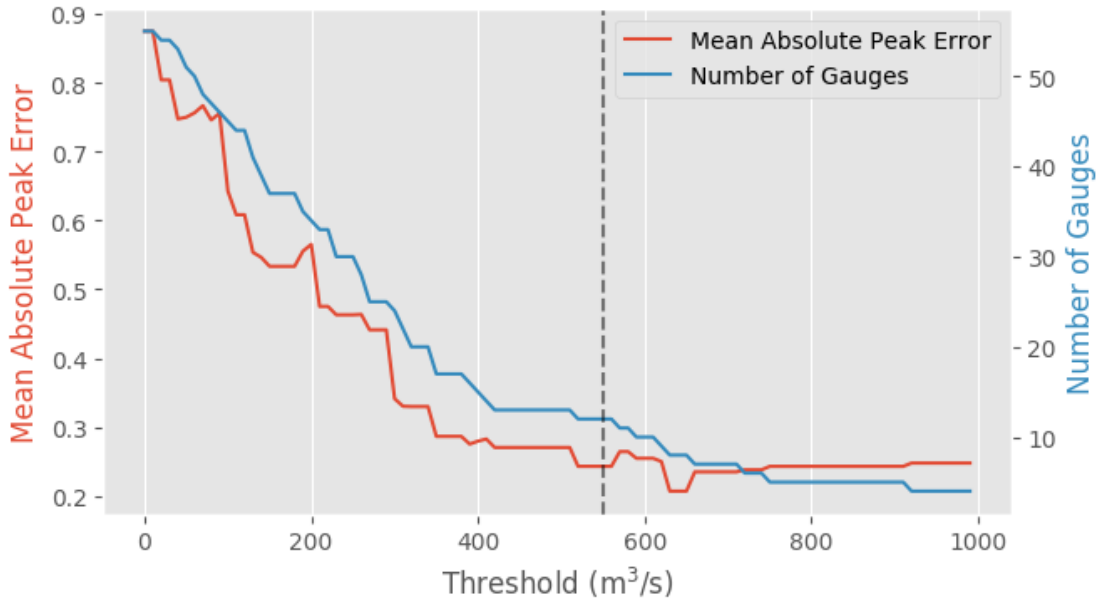

Figure 4-12 - Effect of observed peak threshold (m^3/s) on number of gauges selected and average peak error. The vertical line marks 550 m^3/s.

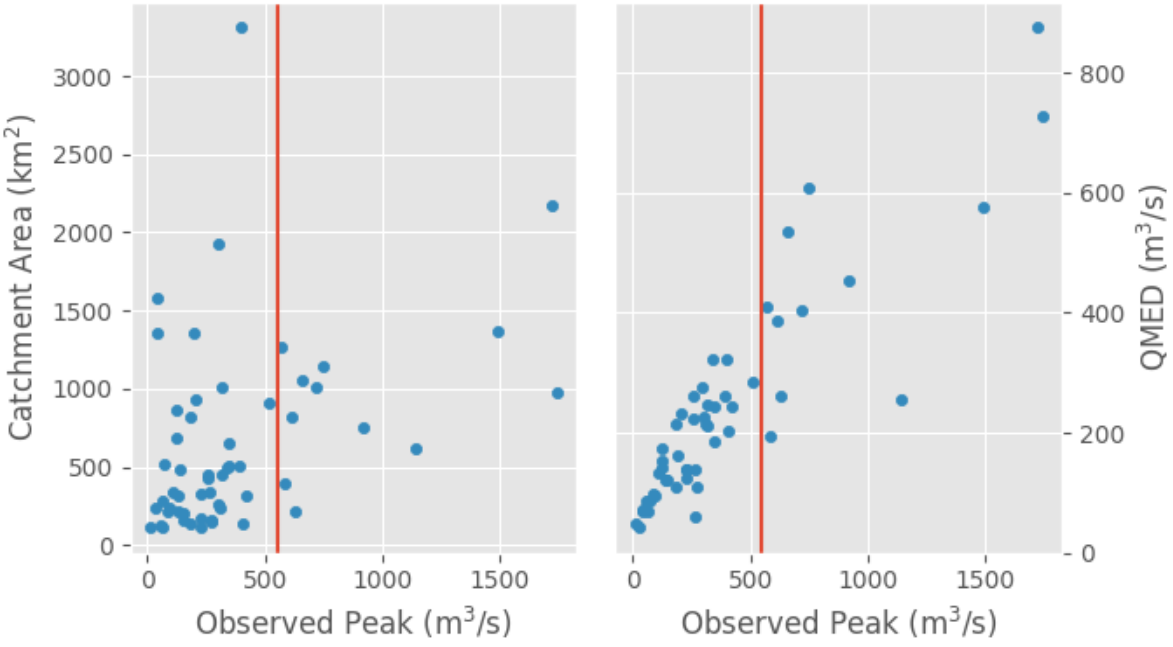

Figure 4-13 - Catchment areas and QMED relative to maximum discharge during Storm Desmond.

Hydrographs showing modelled vs observed discharge at each of these remaining gauges are shown in Figure 4-14. In general, the flood wave shape is well represented across all gauges, with a discernible smaller peak preceding the main event. At all locations apart from the North Tyne at Reaverhill, flood peaks are underestimated by the model, due either to unrealistically low precipitation data or numerical dispersion caused by the simulation. The largest underestimate is at the Eden at Temple Sowerby with best matches seen at Low Moor, New Jumbles Rock and Bywell, on the Tees, Ribble and Tyne.

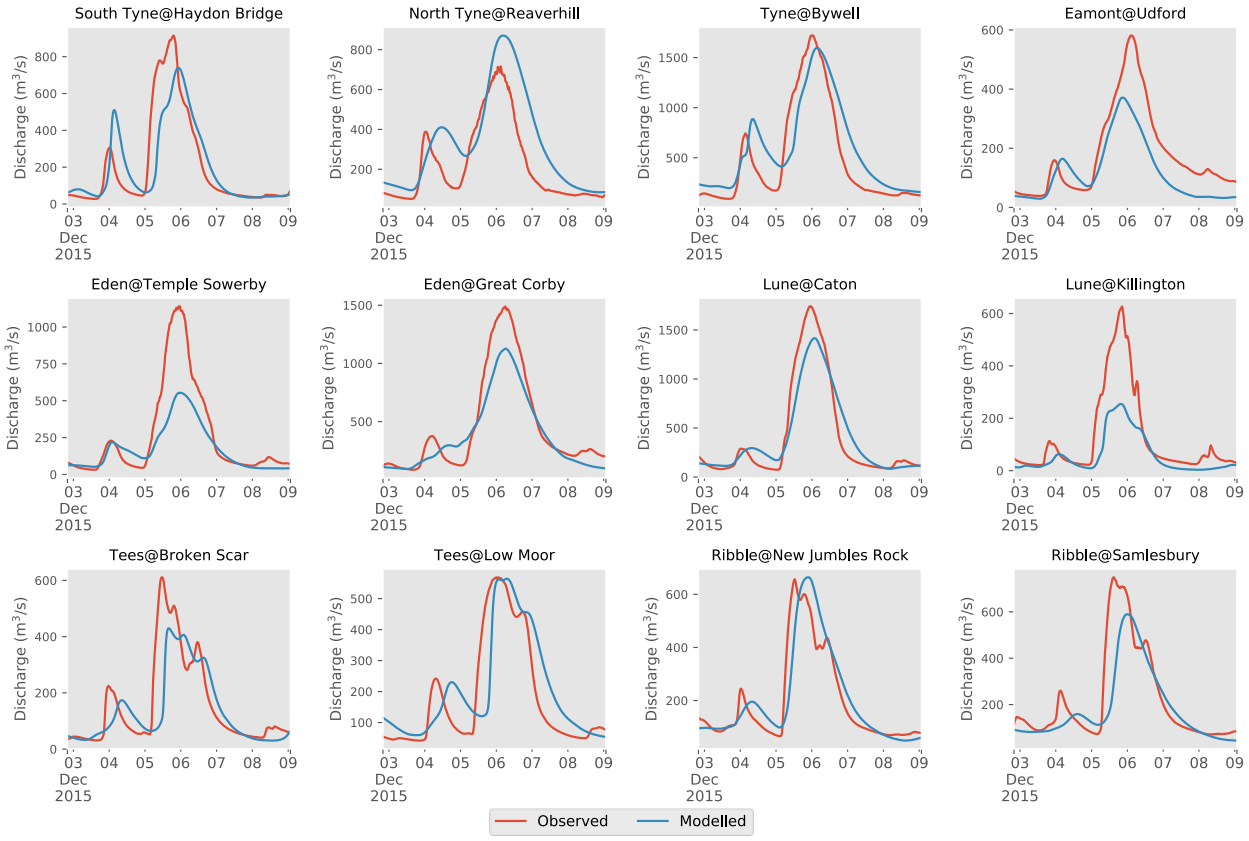

Figure 4-14 – Hydrographs for gauges with observed peaks over 550 m³/s within the study area, showing modelled discharge (m³/s) in blue against observed in red.

A plot of observed vs modelled peaks is also shown in Figure 4-15, clearly demonstrating this negative bias. In terms of timing, modelled peaks almost always reached the gauges later than the observed peaks, as illustrated in Figure 4-16. A possible explanation for the delayed peaks is the numerical dispersion induced by the numerical scheme. This dispersion means the

modelled flood wave travels slightly slower than the observed and arrives at the gauge later. The dispersion effect could be accounted for by adjusting the friction coefficient but this is not carried out here as it would introduce an element of calibration. Avoiding calibration ensures that the methodology is applicable in areas with no gauge data.

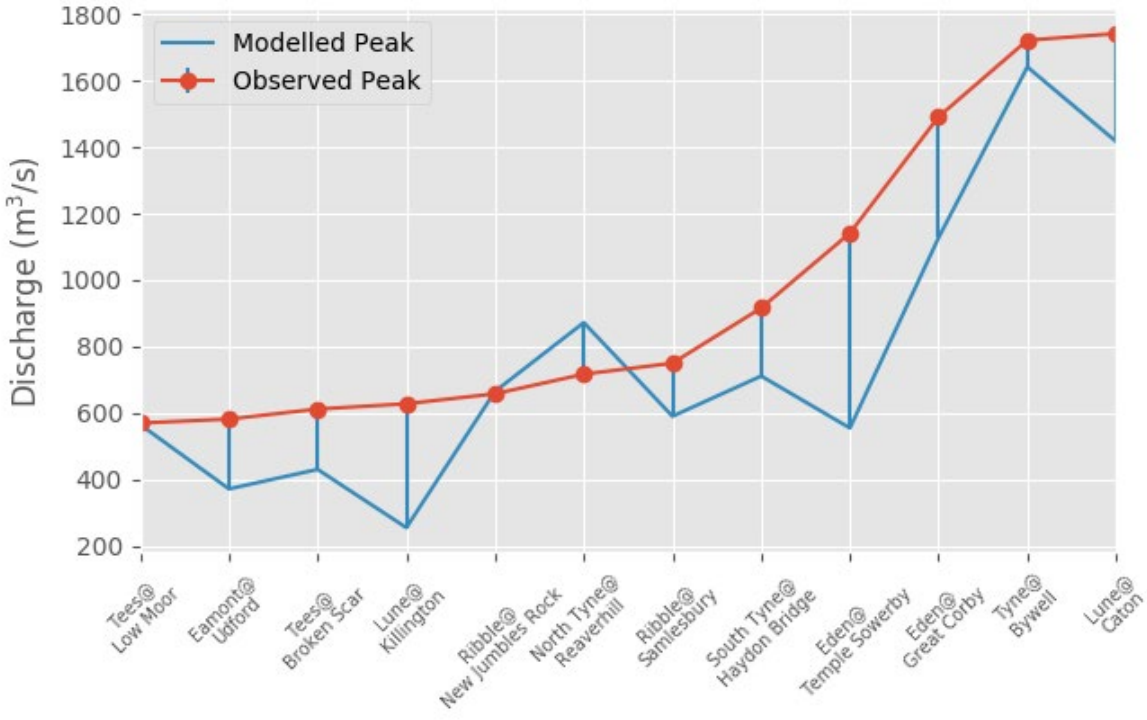

Figure 4-15 - Discharge error at gauges with an observed peak over 550 m³/s.

Table 9 - Selected catchments: Area (km²), QMED (m³/s) and observed peaks (m³/s), sorted by size, to nearest integer.

Gauge ID	River	Location	Area	QMED	Peak
23001	Tyne	Bywell	2176	876	1720
76017	Eden	Great Corby	1373	576	1490
25009	Tees	Low Moor	1264	410	569
71001	Ribble	Samlesbury	1145	607	749
71009	Ribble	New Jumbles Rock	1053	534	657
23003	North Tyne	Reaverhill	1008	404	716
72004	Lune	Caton	983	727	1740
25001	Tees	Broken Scar	818	387	611
23004	South Tyne	Haydon Bridge	751	454	915
76005	Eden	Temple Sowerby	616	256	1140
76003	Eamont	Udford	396	196	581
72005	Lune	Killington	219	262	627

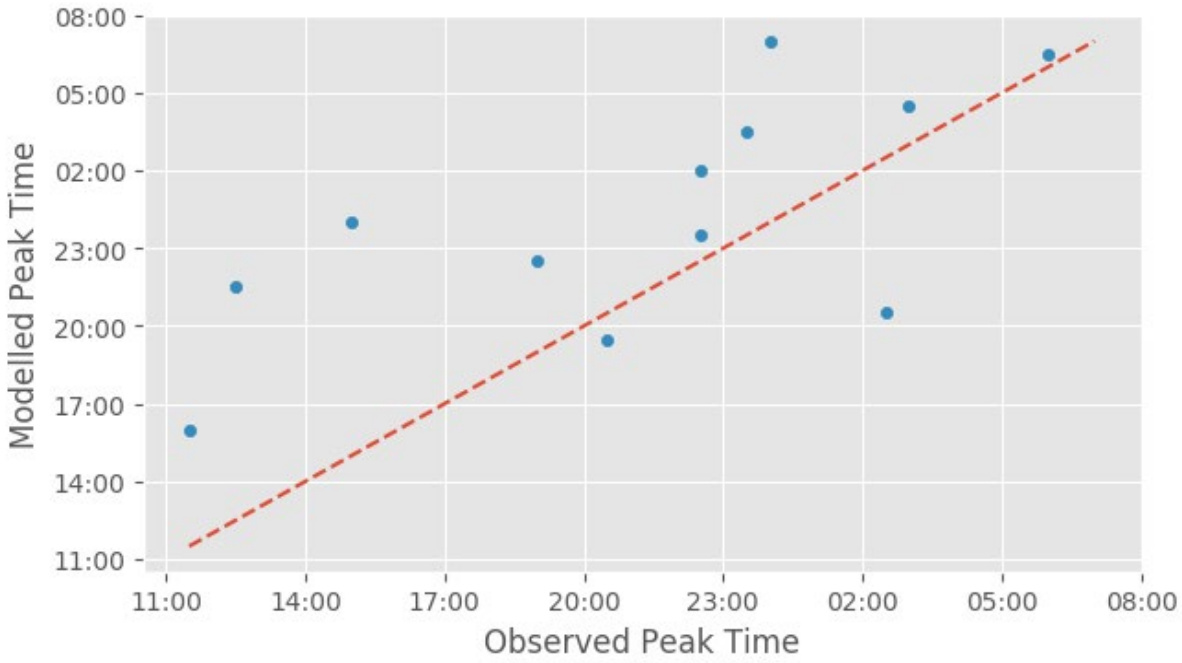

Figure 4-16 - Peak timing at gauges with observed peaks over 550 m³/s.

4.5.2 Extent Validation

To complement the gauge-based validation, extents were also compared with both outputs from a change-detection study using Sentinel 1 data and the Environment Agency's Flood Zone 2, which incorporates both model outputs and observed events. Figure 4-17 shows a visual comparison of the Sentinel 1 output and modelled maximum depths for an area north-east of Carlisle. The change detection methods applied are sensitive to flat surfaces, such as roads, and often define them as water bodies during initial calibration as they are flat in both flooded and non-flooded states. Therefore these flat areas do not get classified as flooded during an event, as seen in Figure 4-17. This means roads which are correctly inundated by the model will count as a false alarm when compared with Sentinel 1 output. There is also patchiness evident in the Sentinel 1 results, possibly caused by vegetation or structures blocking the satellite view of the ground surface. Again, this will result in an under-estimation

of model performance as in reality water would be present under vegetation or within structures, at least in some cases.

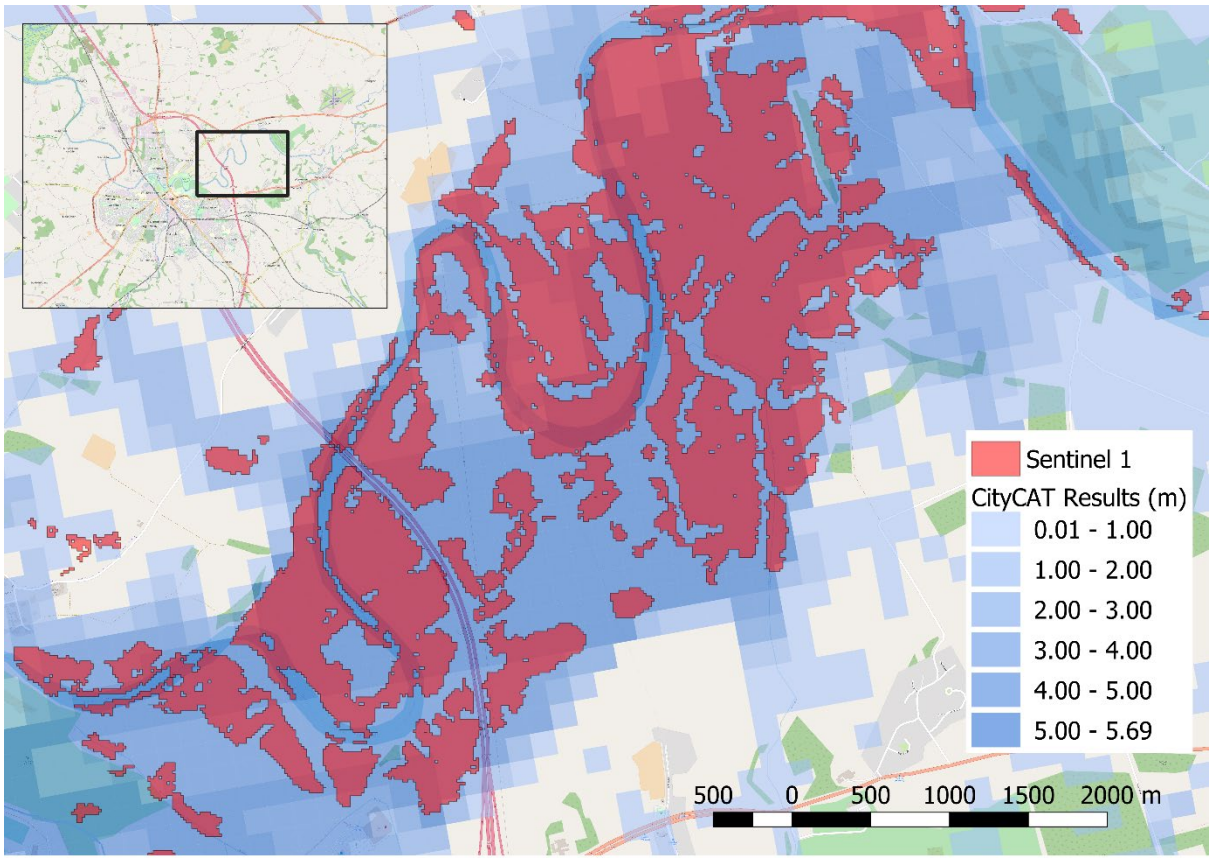

Figure 4-17 – Extent comparison NE of Carlisle between maximum CityCAT output depths (m) using MIDAS rainfall and the HydroSHEDS DEM with accumulated change detection results using SAR from Sentinel 1.

As a result of these factors, the performance metrics resulting from the comparison with Sentinel 1 combined with water bodies from VectorMapLocal are disappointingly poor as shown in Table 10. However, this does not necessarily reflect a lack of model skill and could be caused by unrepresentative Sentinel 1 results. The Eden basin provides the best results with a CSI of 0.16 and Wear the worst with only 0.04. The exact cause of this variation remains unknown but could be at least partially a result of different proportions of land-uses causing variation in the number of surface features. It is not difficult to recognise how model skill could be over-exaggerated by quoting only hit rate as a score without combining this with false

alarms to get the CSI. The Lune, for example, has a hit rate of 0.8 but this is combined with a FAR of 0.84 to create a low CSI score of only 0.15.

Table 10 - Extent performance metrics by basin using Sentinel 1 data and EA Flood Zone 2 for validation.

	EA			SENTINEL		
	CSI	**HR**	**FAR**	**CSI**	**HR**	**FAR**
Lune	0.34	0.63	0.48	0.15	0.80	0.84
Ribble	0.28	0.44	0.47	0.12	0.50	0.86
Eden	0.37	0.67	0.45	0.16	0.74	0.83
Tees	0.29	0.55	0.51	0.09	0.44	0.90
Tyne	0.30	0.66	0.55	0.14	0.56	0.85
Ouse	0.17	0.22	0.44	0.12	0.34	0.84
Wear	0.29	0.61	0.54	0.04	0.22	0.95

Looking at EA metrics, there is improved performance across the board, as illustrated in Figure 4-18. The best performance is again seen in the Eden basin at 0.37 CSI but in contrast to Sentinel results, the Ouse has the worst match, with only 0.17 CSI, still higher than the best Sentinel score. The fact that CityCAT results are more similar to the EA layer than Sentinel 1 is no surprise, as they are both model outputs using relatively similar data. Figure 4-20 shows the clearly better match with the EA outputs than Sentinel observations. Despite this, there are also evident patches where there is a good match between CityCAT output and Sentinel observations which is not replicated in the Flood Zone 2 layer. This may indicate improved capability of CityCAT vs EA models in some areas away from channels which is not manifested in the performance metrics.

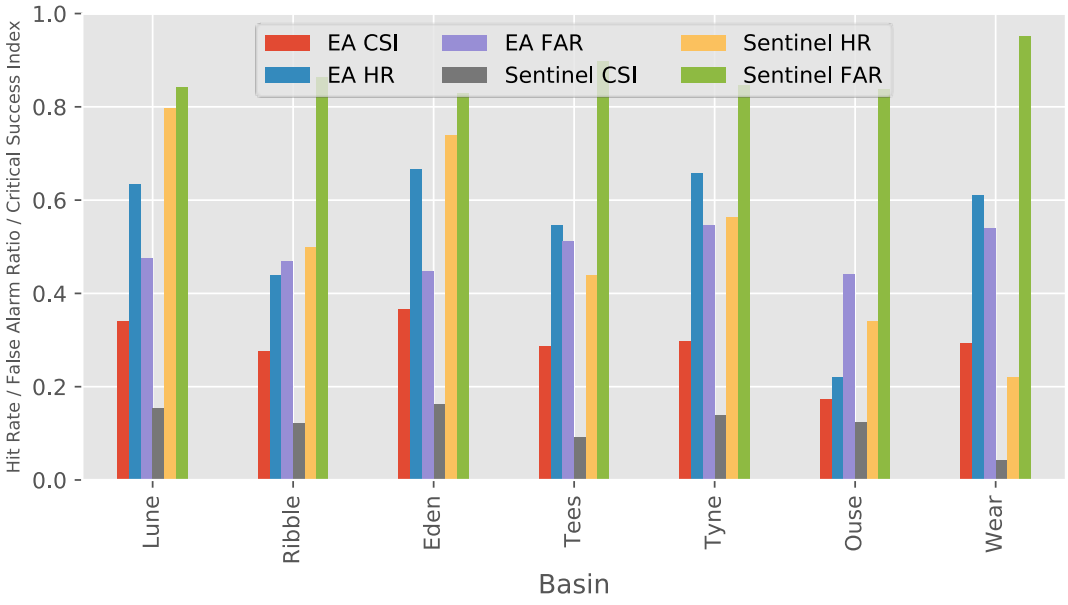

Figure 4-18 - Extent metrics for all basins for Storm Desmond using MIDAS rainfall and HydroSHEDS DEM.

The problem is deciding which layer is closest to the truth. The Sentinel 1 output itself, although technically observed data, is subject to various limitations and therefore is not necessarily a better representation of reality. Another particular issue with remotely sensed data is the intervals at which it is recorded. As satellites tend to make passes at most every 48 hours, it is unlikely that a snapshot will be taken at the peak of a flood extent. This is demonstrated in Figure 4-19 which shows that there were passes on either side of the main flood peak as recorded at a number of river gauges, but none during the main event on December 5-6, where flooding is most likely. This further exacerbates the underestimation of extents in the observed data, therefore reducing performance scores. On the other hand, the EA model results are at high resolution and use good quality data but are not real observations so while they may actually be closer to reality than the Sentinel observations, this is difficult to prove.

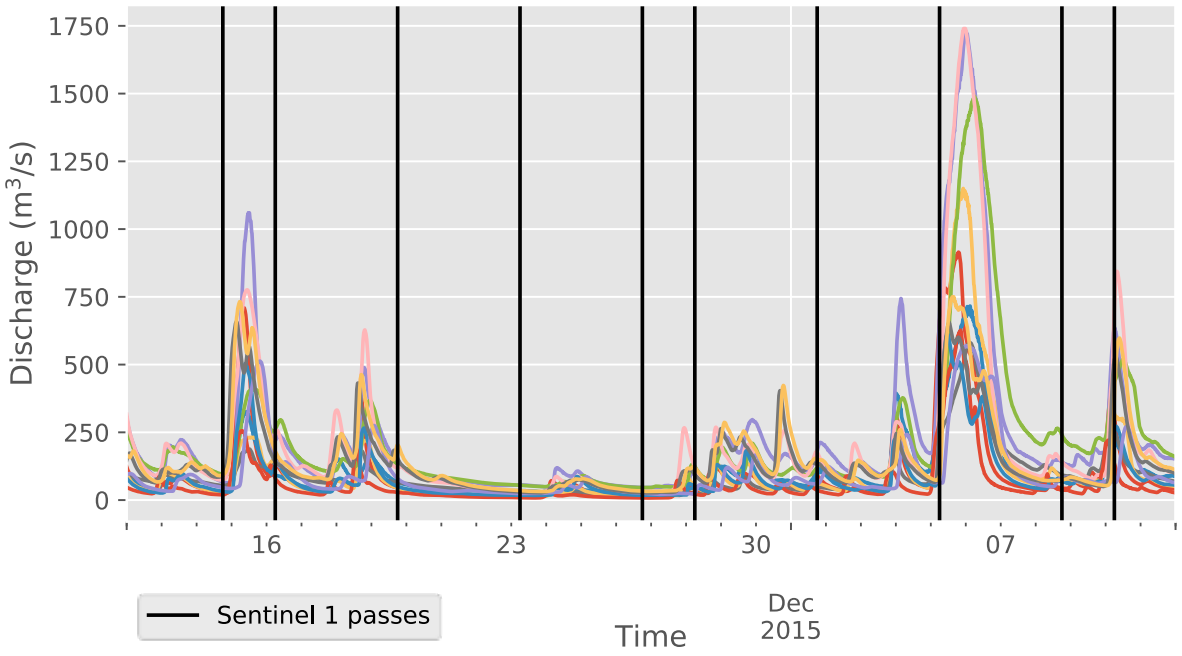

Figure 4-19 - Times at which sentinel 1 snapshots were recorded, relative to hydrographs from the 12 flow gauges in the study area with the highest peaks.

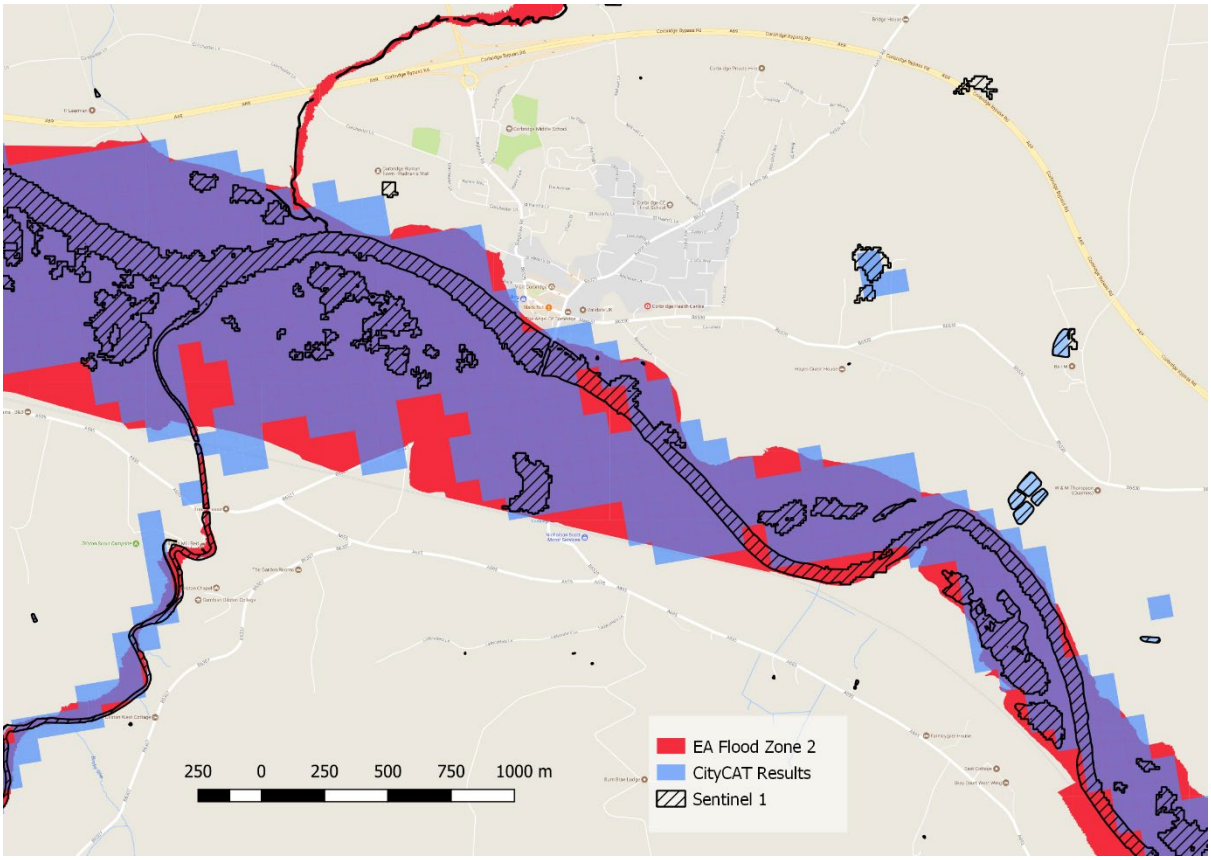

Figure 4-20 – Comparison of flood extent at Corbridge from EA Flood Zone 2, Sentinel 1 change detection and CityCAT results using MIDAS and HydroSHEDS.

4.5.3 Catchment Characteristics

To assess whether there were any physical explanations for the spatial variation in model performance, peak and timing errors were correlated against a range of catchment characteristics held by the NRFA, the coefficients of determination for the top ten correlated descriptors are shown in Figure 4-21. Figure 4-22 illustrates the strong negative correlation between peak error and PROPWET. It is clear that there is a significant effect on flood peaks caused by a lack of infiltration and any knowledge of soil wetness within the model.

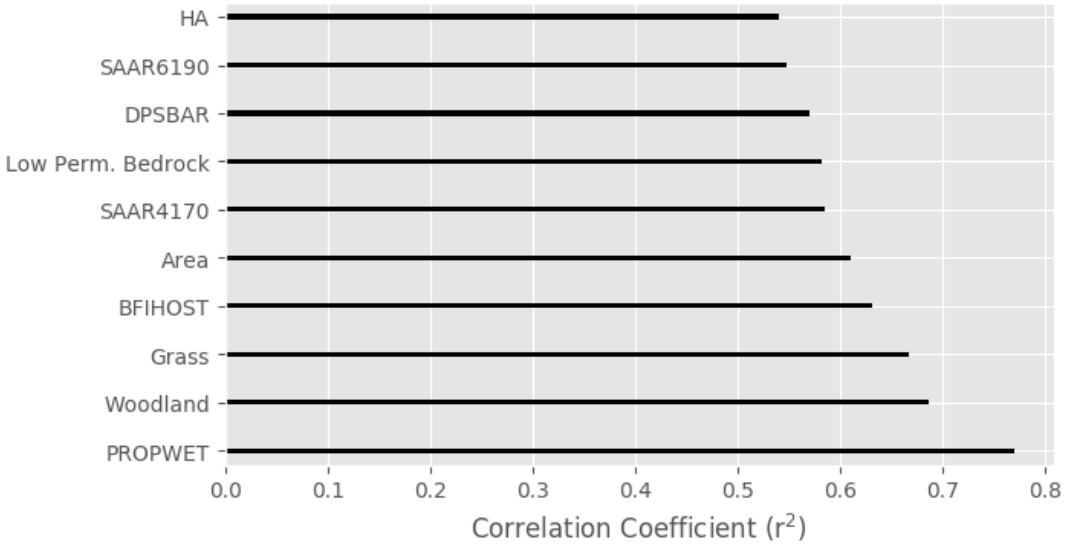

Figure 4-21 – Absolute correlation of catchment characteristics with absolute peak error percentage.

In areas with wetter soil, real flows are likely to be higher as infiltration is low, which is not captured in the model results, causing under-estimation of peaks. The opposite effect takes places when soil is dry, with increased infiltration causing real flows to be reduced which, again, is not captured by the model. It might be expected that due to no infiltration being included in the model, all peaks should be over-estimated, however it is anticipated that the deficit is caused by under-estimation of rainfall, as a result of the limited number of gauges in the network.

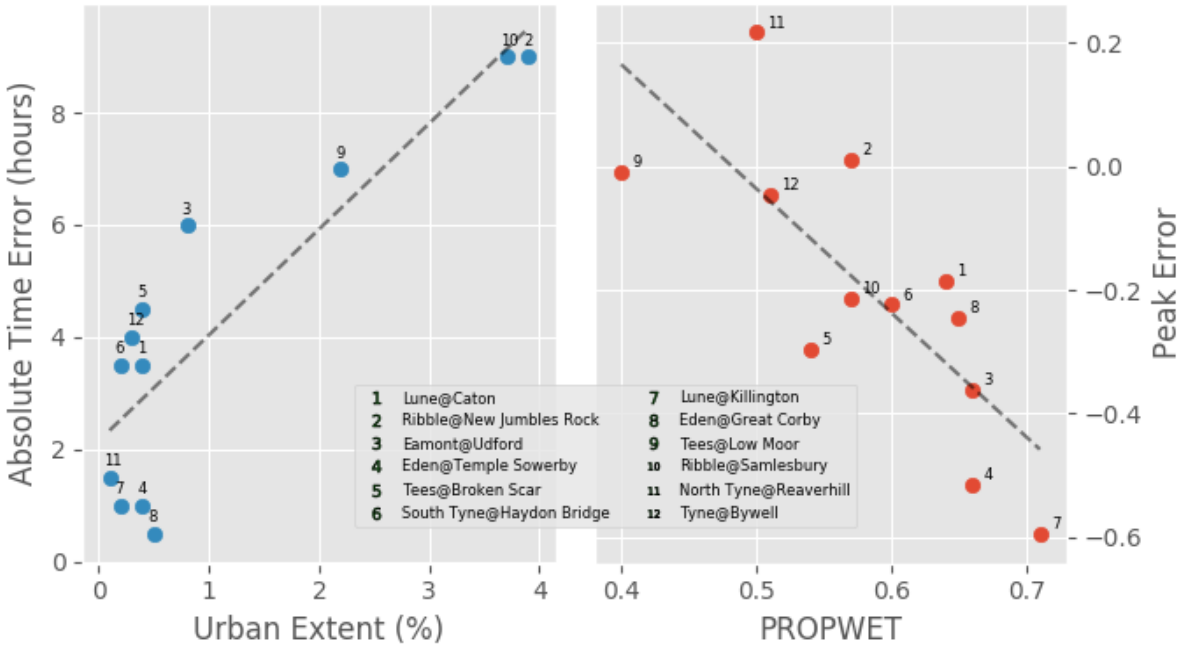

Figure 4-22 - Correlation between PROPWET and Peak error (r^2=-0.77, p=0.003) and Urban Extent and timing error (r^2=0.86, p=0.0003).

Although outside the scope of the present study, the PROPWET of the catchment could be accounted for in the model by allowing spatially variable infiltration. However, soil moisture data is generally harder to obtain than precipitation and DEMs, even more so at larger scales. An increasing proportion of wooded area was also negatively correlated with peak error, due either to increased DEM error in these areas or increased water storage and evapotranspiration not represented within the model. Looking at timing, the highest correlation (0.69) was found with urban extent, plotted in Figure 4-22 (left). The trend is not as significant as between PROPWET and peak error, nonetheless is it a potential cause of the largest delays in timing. One explanation for this is reduced friction in urban areas causing the real flood wave to propagate faster than its modelled counterpart, as this region of lower friction is not incorporated within the model. It is possible to account for this by including lower friction coefficients in urban areas, however that is outside the scope of the thesis.

4.6 European Basins

To test the transferability of the data and methods implemented in the UK and assess the performance of the workflow using only globally available data for larger catchments, floods in five European basins were simulated. The choice of basins was based largely on catchment area and coverage of major cities across France, Germany and Italy. The selected rivers are listed in Table 11. Due to their size, run times for these catchments are approximately 30 times slower than the previous UK based simulations. Therefore, only one simulation was carried out in each case.

As MIDAS gauges only cover the UK, a global reanalysis dataset from the European Centre for Medium-range Weather Forecasting (ECMWF), European Re-Analysis Interim (ERA Interim), was used instead. ERA Interim records rainfall on an 80 km grid at 3 hour intervals and is discussed in further detail in the next chapter. Instead of simulating a single period such as Storm Desmond, a different event was selected for each of the 5 basins based on the presence of a large discharge peak in the observed record. The search period was limited to the duration of the ERA Interim record which spans from 1979-present. Although the quality and availability of validation data is generally lower in the rest of Europe than the UK, some daily series from the GRDC were available and these were used to analyse how effectively flood peaks were being captured in these larger basins. No extent observations were readily available for the selected events and therefore the analysis is limited to peak performance only.

Table 11 - European flood events by basin.

BASIN ID	BASIN NAME	AREA (KM^2)	EVENT START	EVENT END
2120024170	Elbe	138,891	2002-07-20	2002-08-20
2120021030	Loire	117,034	1979-01-22	1979-02-22
2120016510	Rhone	96,868	1979-10-01	1979-10-21
2120022150	Seine	73,474	1980-02-20	1980-03-20
2120013010	Po	73,458	1981-09-08	1981-10-04

Four hydrographs are shown in Figure 4-23. There is considerable discrepancy with the location of gauges on the Seine and so they were excluded from this section but the results are used in Chapter 6. In the four hydrographs shown, spin up periods are more noticeable than in UK basins due to the larger areas requiring more time for water to accumulate and reach the gauge's location. Despite the lag caused by this long spin-up period, the shape and peak timing of these selected gauges is promising and demonstrates that the workflow can effectively be applied across much larger domains, maintaining the lack of calibration.

Notably, flood peaks tended to be overestimated by the model in contrast to the underestimation observed in the UK and the degree of error was much greater. This could be due to the shorter time-step of the model than flow gauges, causing the peak to be under-represented in the observed record. The gauges themselves could also be less effective at capturing high flows which is particularly difficult in these larger channels. Another problem is the over-burning of channels in the HydroSHEDS DEM which results in unrepresentatively high conveyance. To improve model performance in these larger basins, a longer spin-up period is required, however this was not practical given the amount of time currently required to run the modelling system. By improving computational performance, this may be possible in the future. Using the updated ERA-5 dataset in place of ERA-Interim or replacing the DEM with a dataset which does not have unrepresentatively deep channels burned in would also be likely to improve results.

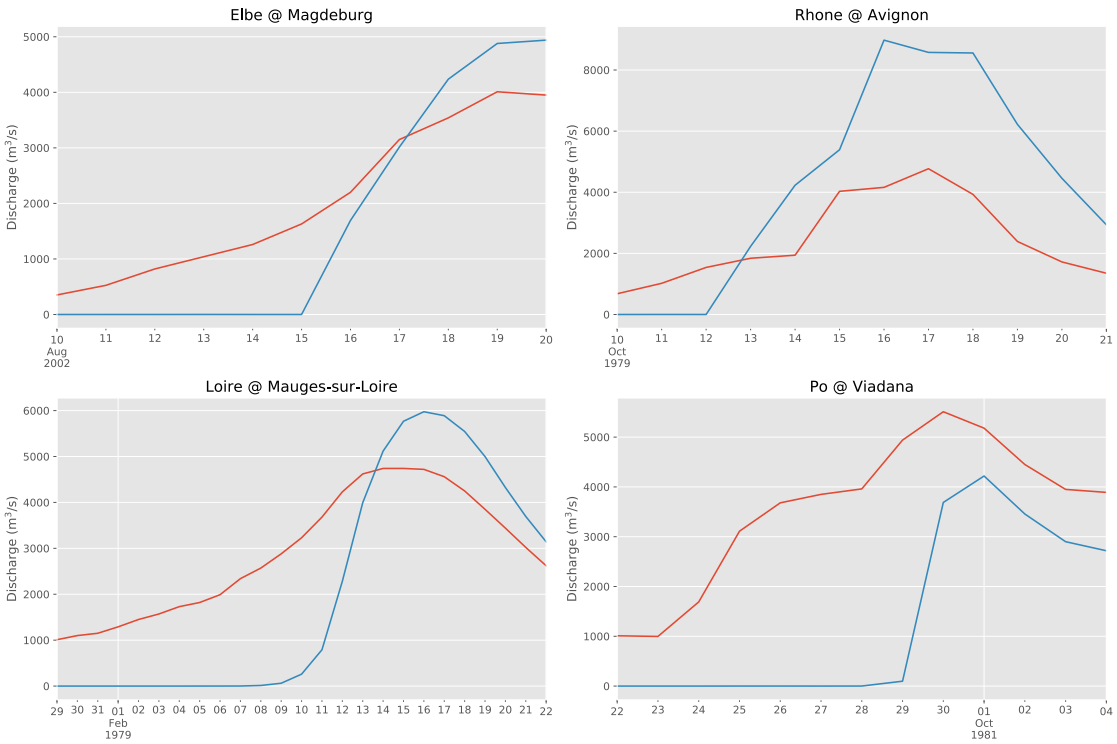

Figure 4-23 - Modelled (blue) vs observed (red) discharge (m^3/s) at four gauges from European basins. Modelled discharge is resampled from 30 minute to daily resolution to match observations.

4.7 Summary

The automated and un-calibrated modelling workflow outlined in Chapter 3 has been validated for both Storm Desmond in the UK and a range of events across Europe using methods orientated towards capturing the accuracy of maximum flood peaks and extents. Results indicate that good performance is possible at gauges with high flow observations using the 90m HydroSHEDS DEM with MIDAS rainfall in the UK and ERA-Interim re-analysis in Europe.

The difficulty in assessing extent performance has been discussed, with the main limitation being the highly challenging nature of collecting reliable distributed and frequent observations during an event. Ground surveys would be prohibitively expensive and satellites can only pass once an orbit, often missing the main flood peak. Even when a satellite happens to pass over at the correct time, buildings, vegetation and flat surfaces can inhibit the ability of processing methods such as change-detection to accurately determine whether or not an area is flooded.

Comparison with other models using theoretically more accurate data, such as the comparison with EA Flood Zone 2 here, is a possible way around these problems. The results suggest improved performance when compared with the EA model extent rather than Sentinel 1. However, as this does not involve any real observations, the usefulness of the approach is questionable. River gauges are in general more reliable, however during large precipitation events, discharge may reach outside of reliable rating curve bounds meaning observed flow values may not be accurate either, particularly if the river is out of bank. These validation difficulties permeate all flood simulation studies and are an inherent problem with trying to recreate such extreme events.

For Storm Desmond, flood peak performance has been shown to correspond to the proportion of time soils are wet, with wetter soils resulting in under- and drier soils over-estimation of maximum discharge. This could be accounted for in future studies by incorporating infiltration, however as this thesis is concerned with assessing the feasibility of modelling floods with limited data, this has not been addressed here. Nonetheless, usability of 90m global terrain data for 2D hydrodynamic modelling has been established. The automated cloud-based workflow means this can readily be applied globally, though accuracy can be difficult to characterise in many areas where validation data is limited and large basins present computational limitations using the CityCAT modelling system. The following chapter will progress to use a wider range of input datasets and parameters to assess the sensitivity of flood peaks and extents to these variables.

Chapter 5. Sensitivity to Data & Parameters

So far, only a single combination of input datasets and parameters has been tested. In this chapter, the rapid setup and performance analysis capabilities of the workflow are utilised to test models using a range of datasets and parameters. Usually, such sensitivity tests are not possible for computationally demanding hydrodynamic models like CityCAT, however the automated and distributed approach of this research makes them possible.

Simulation outputs were compared based on peak error, timing error and extent similarity. Flow gauge observations, Sentinel 1 remotely sensed extents and Environment Agency hazard maps were used to produce the required metrics. Gauges with peak flows below 550 m^3/s were excluded from the analysis as they produce unreasonably poor results, due to the resolution of both the DEM surface and rainfall, as outlined in Chapter 4. Both Chapter 3 and Chapter 4 also contain more technical methodological descriptions and may be referred to for further details.

5.1 Precipitation

Excluding coastal flooding, the primary driver of any source of flood water is the initial input of precipitation. The duration, intensity and spatial extent of a precipitation event, determines the reach and magnitude of its impacts. Therefore, the choice of precipitation data (*P*) when simulating floods is critical. Inaccurate *P* will undoubtedly lead to a spurious and potentially misleading understanding of the risk posed by a given event. This effect is further exacerbated when low quality *P* is used to project risk into the future, with planning decisions being made based on the results. Unfortunately, understanding which source of *P* is the best representation of reality is challenging. There is also spatial variation in the availability and quality of *P*, largely dependent on the number of rain gauges available in a region and the length and quality of their records. Therefore, there is no single best source for all time

periods, resolutions and areas, necessitating a prioritisation of criteria when selecting the most suitable dataset for an application. More accurate data is often available locally but inaccessible or difficult to attain, therefore continental or global *P* is a popular option for broad-scale flood modelling despite its generally lower resolution and accuracy. However, given the wide range of products available, it can be difficult to choose which is likely to give the best results for a certain use case.

To decide which *P* is the most accurate, more reliable local data is required as a baseline to validate against. However, the quantity and quality of local data is limited across much of the globe, particularly in developing regions. Commonly, local data is available only at specific locations and not as a gridded product, therefore gridded products can only be tested at these locations, unless a further interpolation step is carried out. Also, local data may in fact be of lower accuracy than the large scale products if the rain gauges on the ground are of such poor quality or have been influenced by enough human error that climate models do a better job of estimating *P*. There is no way to check which is more correct by looking at precipitation alone. To cross-validate *P* accuracy, an independent source of data is required. River flow data (*Q*) has been used for this purpose in the past (Beck, Vergopolan, et al., 2017) and presents a viable option for assessing *P* performance in the context of flood events. Not only can individual products be tested against each other, but the effects of temporal and spatial resolution can also be assessed.

5.1.1 Datasets

The modelling framework outlined in Chapter 3 was used to assess a range of *P* datasets, including ground-based gauges, climate reanalysis and remote sensing products. Each has its own set of spatiotemporal resolution and extent characteristics, ranging from 1 to 80 km cell sizes and hourly to daily time steps. The availability of each source is also highly variable, with

reanalysis data having continental to global reach in an uninterrupted series and hourly gauge data being limited to point locations with relatively short and sometimes discontinuous records. For all rainfall product test simulations, the Digital Elevation Model (DEM) was taken directly from the 90m HydroSHEDS product. Further details can be found on HydroSHEDS in Section 4.1.2. Simulation results were compared with both flow gauge data and satellite observations for the period 12 November - 12 December 2015 which encompasses Storm Desmond (3-8 December 2018) across Northern England. All precipitation data has been downloaded directly from the internet and is freely available for use in research. Some intermediary steps were required to translate the original formats into postgres tables which are explained in Section 3.1.1. The datasets themselves, summarised in Table 12 will now be discussed in more detail.

Table 12 - Precipitation data time intervals and spatial resolutions

DATASET	TIME INTERVAL	RESOLUTION
MIDAS	1 hour	1 km
ERA Interim	3 hour	~80km
ERA 5	1 hour	~30km
EOBS	24 hour	~30km
MSWEP	3 hour	~30km

5.1.1.1 MIDAS

The MIDAS dataset, outlined in Section 4.1.1, was used as a benchmark local precipitation dataset. In a best case scenario, local, high quality rainfall data would always be used. However, long enough usable records at a suitable density are not always available, particularly in developing areas. Therefore lower resolution datasets, as outlined in the following sections, often provide the only options.

5.1.1.2 ECMWF Re-analysis

The European Centre for Medium-Range Weather Forecasts (ECMWF), based in Reading, is the largest provider of re-analysis products in Europe. Each describes a unique time period

with varying spatial and temporal resolutions using a collection of climate observations. Data from satellites, ground-based instruments and observers is used to drive computational models of the atmosphere and ocean and generate estimates for a range of variables at every specified grid cell and time step. ERA40 (Uppala et al., 2005), the initial attempt to produce a coherent re-analysis product, was the largest experiment of its kind and was highly influential at the time, however has since been superseded by the vastly improved ERA-Interim (Dee et al., 2011). The most recent release from ECMWF was ERA5 in July 2017, which is still under continued development. Compared with ERA Interim, ERA5 supports an increased spatial and temporal resolution and eventually will replace the earlier dataset. Both provide estimates of a large number of climatic variables up until the present day. The major differences are shown in Table 13. From both datasets, the variable used was total precipitation.

5.1.1.3 E-OBS Station Interpolation

E-OBS is a 0.25 degree resolution gridded daily dataset providing temperature, pressure and precipitation information across Europe (Haylock et al., 2008). The spatial grid was designed to match the monthly Climate Reanalysis Unit (CRU) dataset. Monthly totals are interpolated first, then indicator and universal kriging is used to interpolate the daily values before monthly and daily estimates are combined. The number of gauges included in interpolation depends on the time step and can range between 500 and 2000.

To assess how well extremes are captured in the final product, a cross-validation dataset was created by selectively removing each station, using its neighbours to interpolate the missing records and comparing the interpolated series to the observed data. The median reduction factor for precipitation using the cross-validation data was found to be 0.66, demonstrating the large impact of interpolation on extremes. Hofstra et al (2009) carried out further testing of E-OBS using local datasets with denser gauge networks and highlighted the need for

understanding of the limitations of accuracy in the interpolated product. In the UK, E-OBS was compared to a 5 km resolution daily dataset provided by the Met Office, interpolated using 4400 stations. The RMSE of E-OBS when compared to the Met Office data was 2.17 mm daily, which is more consequential for water supply as a resource than flood volumes.

Table 13 - Differences between ERA-Interim and ERA5 (ECMWF, 2019)

	ERA-INTERIM	ERA5
Temporal Resolution	3-hourly	1-hourly
Spatial resolution	79 km	31 km
Period covered	1979 - present	1950 - present
Input observations	As in ERA-40 and from Global Telecommunication System	In addition, various newly reprocessed datasets and recent instruments that could not be ingested in ERA-Interim
Model input	As in operations (inconsistent SST)	Appropriate for climate (e.g. CMIP5 greenhouse gases, volcanic eruptions, SST and sea-ice cover)
Production Period	August 2006 – end 2018	Jan 2016 – end 2017, then continued in near real-time
Assimilation system	IFS Cycle 31r2 4D-Var	IFS Cycle 41r2 4D-Var
Satellite data	RTTOV-7, clear-sky, 1D-VAR rainy radiances	RTTOV-11, all-sky for various components
Variational bias scheme	Satellite radiances	Also ozone, aircraft and surface pressure data

5.1.1.4 MSWEP Ensemble

The Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset from Princeton University offers a weighted-ensemble of re-analysis products, gauge interpolations and satellite observations at 0.1 degrees (Beck, Van Dijk, et al., 2017). The initial version provided 3 hourly rainfall on a 0.25 degree grid before being upgraded to the higher resolution second and current version (Beck, 2017). MSWEP has undergone quality control to ensure erroneous gauges are fully or partially excluded. This process included checking the cumulative distribution function, daily totals, and correlation with gridded datasets and number of unique values, among other steps. The satellite and reanalysis data were tested against ground-based

gauges using 3-day averages. Bias correction was then carried out using the number of wet days in the gauges and reanalysis or satellite datasets.

When tested using hydrological modelling (HBV) against a range of other similar precipitation products, including ERA Interim, MSWEP performed the best on NSE, R^2, mean absolute error (MAE) and bias among other estimates. In terms of 99.9th percentile error, which is the most relevant for flooding applications, ERA-Interim outperformed version 1 of MSWEP with 27.82 to 29.30 mm/d, until the latest version was released with only 14.90 mm/d. MSWEP is freely available online at gloh2o.org.

5.1.2 Inter-comparison

Model performance when using MSWEP, E-OBS, ERA-Interim and ERA5 rainfall was measured against the performance of the models driven using MIDAS data from Chapter 4. Performance was assessed in terms of flood peak, timing and extent metrics. All models were found to generally under-estimate flood peaks at the 12 selected gauges, as seen in Figure 5-1. Overall, models using MIDAS data tended to outperform those using the more widely available products. One explanation is the higher spatial and temporal resolution of MIDAS compared to the other datasets. Comparing performance of models using the two ECMWF re-analysis datasets, those driven by ERA 5 outperformed others using its predecessor ERA Interim. Again, a large contributor to this improvement is the increased spatial and temporal resolution of ERA5 compared with ERA Interim.

Model performance was most variable across gauges when using MSWEP rainfall, which could be explained by the wide range of sources which are drawn on to create the dataset. Although EOBS is derived directly from precipitation gauge observations, models using E-OBS data performed to a similar standard as those using ERA-5. This may have been caused by the low temporal resolution of EOBS, which is limited to daily time steps. E-OBS also only interpolates

observations with no representation of physical processes, as opposed to the more dynamic data assimilation and modelling in ERA5. However, as the information describing what observations were used in creating ERA5 is currently not easily accessible, it is difficult to suggest what may be causing this.

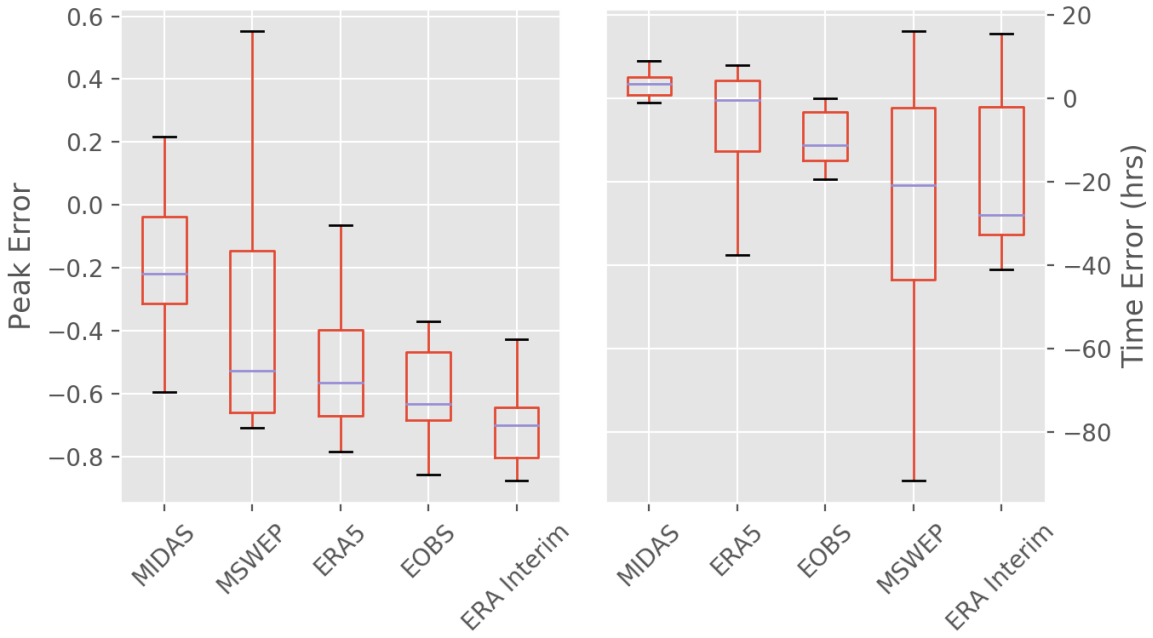

Figure 5-1 – Peak and timing error using HydroSHEDS 90m DEM and a range of rainfall products.

Notably, the performance of models using ERA Interim data is lower here than when the same data was used to simulate large European basins in Section 4.6. This is likely to be a consequence of either the increased number of 80 km ERA-Interim cells falling within the larger basins or a better representation of atmospheric dynamics over the more continental European areas. The storm events causing large peaks in rivers such as the Po and Rhone are also undoubtedly larger than Storm Desmond and therefore more likely to be realistically captured by the re-analysis data.

Timing error is shown in Figure 5-1 (right) between observed and simulated peaks at the selected gauges for models using each *P* dataset. There is greater variation in performance here than in the peak error results. Again, models using MIDAS data demonstrate the highest

level of accuracy, closely followed by those driven with ERA5. This makes sense as MIDAS and ERA5 are joint first in terms of temporal resolution, with hourly time steps.

Unfortunately, spurious error measurements have been caused in some cases, as an earlier peak exceeds the main peak in the observed series, causing an over-representation of timing error, as seen in Figure 5-2. This false peak problem is clearly affecting simulations using MSWEP rainfall, as seen in the timing errors. It is difficult to draw any meaningful conclusions about model timing performance when the wrong peak is used to generate a timing error. The problem is caused by an earlier burst of rainfall being over-estimated and the more intense, main period being under-estimated.

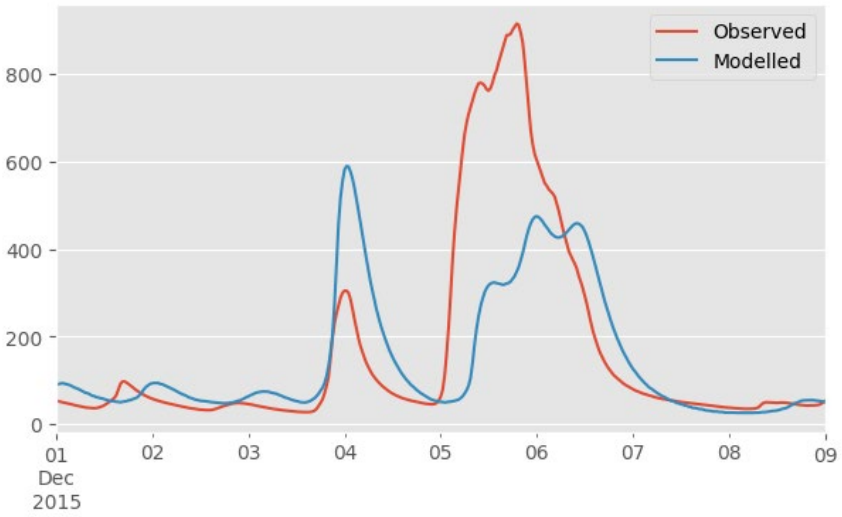

Figure 5-2 - Example of a spurious early peak from a simulation using MSWEP rainfall causing incorrect timing error.

In the case of MSWEP, it is difficult to speculate as to the direct reason for a spurious early peak as it is not clear what proportion of gauged, radar and re-analysis data has been used in a given area. Some attempts were made at using peak identification algorithms to find the nearest high point to the observed and use this to calculate both peak and timing error, however this is also subject to misinterpretation. If the main peak is delayed and preceded by

a smaller peak, timing error will be underestimated and peak error overestimated. Therefore, the standard approach of comparing the highest peaks in the series had to be used.

Flood peaks from models using E-OBS data are approximately 12 hours behind the observed peaks, on average, which could be explained by the daily values being positioned at 12am rather than 12pm, in the middle of the accumulation period. Some difficulty was found in ascertaining the original observing period of the data used when creating E-OBS, which may explain this offset (Haylock et al., 2008). The dramatic improvement in model performance seen when upgrading from ERA Interim rainfall to ERA5 is partly due to an increase in temporal resolution from 3 hourly to hourly, but a better representation of extreme values has also contributed to ensuring that the correct peak is designated as the largest in the series.

Hit Rate (HR), False Alarm Ratio (FAR) and Critical Success Index (CSI) were used to compare the extent similarity of models using each *P* dataset to remotely sensed extents. Further detail about the metrics and methods used can be found in Section 4.4.2. Metrics were calculated for each basin and the results were averaged to produce over-all performance statistics. When compared with extents from Sentinel 1, there is minimal variation in the performance of models using different rainfall datasets. Figure 5-3 (right) shows some internal variation in hit rate and false alarm ratio, however this does not correspond to improved CSI due to a consistently high FAR, indicating that CityCAT is flooding more area than observed by Sentinel 1. For example, considering hit rate alone, models driven by MSWEP appear to perform best, but the same simulations also produced a high FAR, leading to a very similar CSI to the other datasets. Models using ERA-Interim produced the lowest HR results but their CSI was not significantly lower than those using MSWEP.

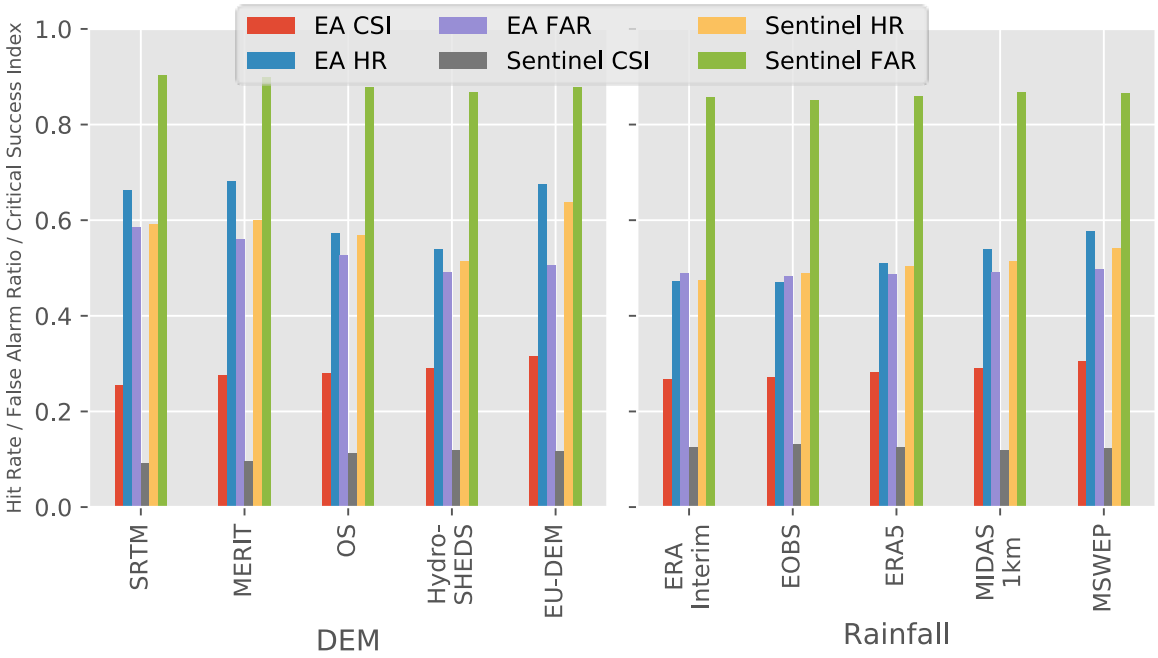

Figure 5-3 - Metrics describing the accuracy of maximum extent when compared with outputs from Sentinel 1 for the same period and polygons from EA Flood Zone 2.

Again, the EA data did not produce a large amount of variation in CSI. Models using ERA Interim and EOBS produced the lowest CSI values of 0.27. When ERA5, MIDAS and MSWEP were used, scores improved to 0.28, 0.29 and 0.30 respectively. Overall, the performance is still generally relatively low, however these scores ignore internal variations within basins which are shown in Figure 4-18. Despite the lack of significant variability in CSI, their ranking does approximately correspond to the spatial resolution of the datasets, as shown in Table 12. ERA Interim has a grid spacing of ~80 km which equates to cells more than seven times the size of ERA5, MSWEP and EOBS, each with ~30 km resolution. Although MIDAS has the highest resolution grid at 1 km, models using MIDAS data did not produce the highest CSI. This may have been caused by the sparsity of rain gauges or could indicate that the accuracy of rainfall input is no longer the limiting factor and the DEM quality is more significant.

The extent comparison method fails to take into account depths and a threshold of 30 cm has been used to assign cells as flooded or not flooded from the model (Section 4.4.2). As neither

the data from Sentinel 1 nor the EA distinguishes depths, there is no convenient way to verify if this threshold is suitable. Any reflective surface may also be classified as flooded in the change-detection algorithm, creating incorrectly flooded areas.

5.1.3 Spatial Resolution

The size of rainfall grid cells effects the precision with which they are able to represent the spatial distribution of rainfall during an event. To quantify this effect, models were run with MIDAS data at varying interpolation grid resolutions, maintaining all other parameters constant. The results of these runs are shown in Figure 5-4. Notably there is only a significant difference in peak error when resolution is reduced below 10 km. The resolution at which model performance deteriorates will depend highly on the size of the catchment, with smaller basins requiring higher resolution rainfall to capture variations more effectively. The density of the gauge network will also affect how influential grid resolution is as if gauges are generally more than 10 km apart, increasing the resolution below this level will not add any more useful detail. The difference in performance between models with 500m, 1 km and 5 km resolutions is not significant enough to draw any conclusions about their ranking as the variation between them is so minor. As other *P* datasets, such as ERA-Interim, have significantly lower resolution than MIDAS, any under-estimation of peak flows when using these datasets could simply be a result of reduced resolution.

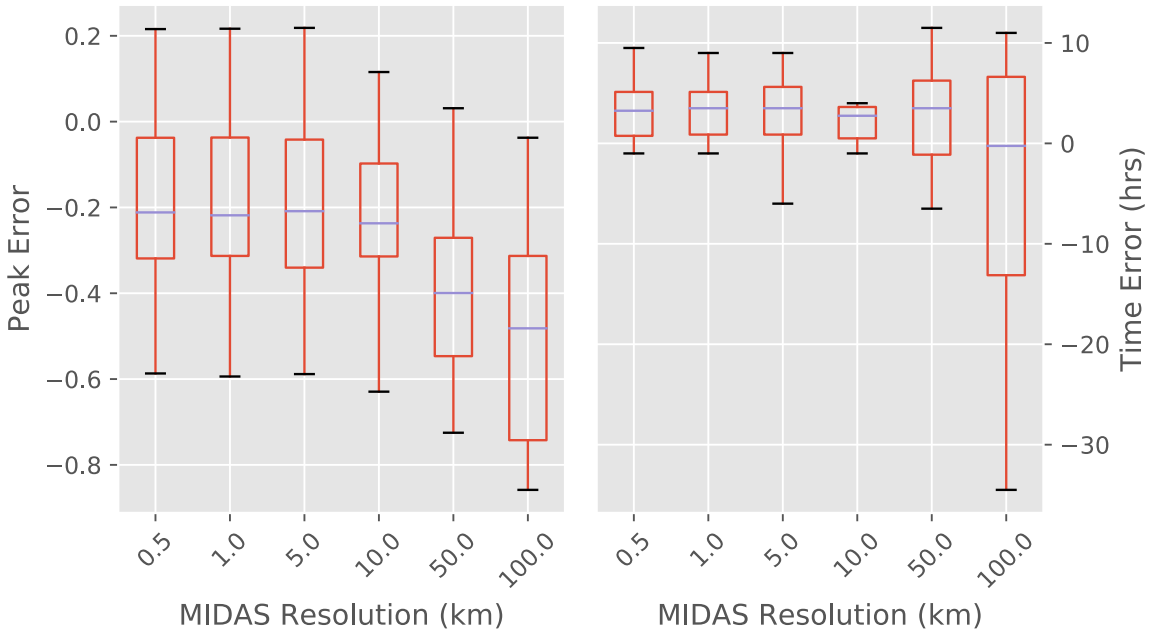

Figure 5-4 - Effects of interpolation resolution (km) of MIDAS rainfall on peak and timing error.

5.2 Digital Elevation Models

Land surface is the most influential factor in the distribution of flood water. The topography of a domain dictates the flow pathways and speed of movement of rainfall after it reaches the surface. Therefore using accurate information to describe this surface is critical in producing realistic flood simulations. Large scale elevation data is usually collected either from satellites or aircraft carrying sensors to measure distances to the earth's surface. In developed countries, such as the UK, LiDAR is used to produce high precision and resolution DEMs, however, due to the low altitude that these missions must be flown at, the coverage is generally patchy and can be of variable quality. Recent developments in airborne lidar and particularly miniaturisation of sensors have enabled the use of UAVs in data collection, greatly reducing costs. Single-photon-sensitive sensors have also entered the market which allow a wider area to be scanned while maintaining the same point density as conventional LiDAR (Mandlburger, 2019). These advances will allow more countries to match the high standard set by the Netherlands with their 0.5 m national-coverage dataset (Google, 2019).

Where LiDAR data is not currently available, the alternative is to use data from satellites, which provides more consistent coverage and performance, with the caveat of generally lower resolution and some systematic errors. Whether from satellites or aircraft, original datasets often contain issues affecting hydrological connectivity which need to be addressed before they can be used in modelling. This can be either an automated or manual process. Automation means large datasets can be processed quickly, however often is not able to correctly detect artefacts. Manual inspection and correction usually results in a better outcome, however is extremely time intensive. A number of projects have produced products which have been manually or automatically processed or both. Surprisingly there are relatively few published studies which directly compare different DEM datasets in the context of flood modelling.

5.2.1 Datasets

The choice of datasets to include in the study was based on availability, completeness, quality and popularity in the literature. Table 14 shows a summary of the resolution and coverage of each DEM. SRTM is the highest resolution but least processed product. Both HydroSHEDS and MERIT are directly derived from SRTM, while EU-DEM includes it as a source. The most localised dataset is OS Terrain, which is also the only product to be created using airborne rather than satellite-based sensors. A more technical outline of how rasters were processed to be used in the model is included in Chapter 3. Each DEM is now described in more detail.

Table 14 – Spatial resolution and coverage of selected elevation datasets.

Dataset	Spatial Res. (m)	Coverage	Source Dataset	Vertical Error (m)
SRTM (Farr et al., 2007)	30	~Global	SRTM	10
HydroSHEDS (Lehner et al., 2013)	90	~Global	SRTM	10
MERIT (Yamazaki et al., 2017)	90	~Global	SRTM & AW3D	10
EU-DEM (Bashfield & Keim, 2011)	50	Europe	SRTM & ASTER	Unknown
OS Terrain 50 (Ordnance Survey, 2015)	50	Great Britain	LiDAR	4

5.2.1.1 SRTM

NASA's Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007), flown in 2000, remains the highest quality set of freely and globally available land elevation observations available, covering almost the entire earth surface between 60 degrees north and south. A single satellite, shown in Figure 5-5, carrying two radar antenna connected by a mast, was used to measure distances to earth from two fixed locations on the instrument, allowing elevation to be calculated using interferometry.

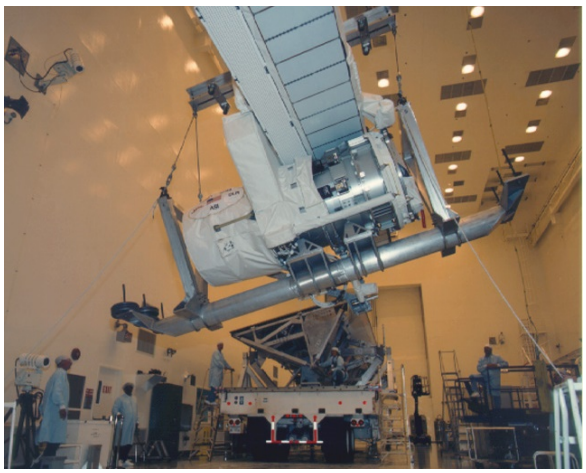

Figure 5-5 - SRTM in production (JPL, 1999)

The data can be accessed via NASA's Earth Explorer web interface or free batch download desktop software. SRTM was initially only available at a reduced 90m, but in late 2015, NASA

released the full 30m resolution version. However, inherent limitations in the way observations were made means that there are systematic inaccuracies (Figure 5-6) along with gaps in the data. Therefore efforts have been made by many groups to improve the overall accuracy of the dataset, particularly for hydrological and hydraulic modelling. NASA is currently working on its own version of a corrected SRTM product called NASADEM (JPL, 2019), however this remains in development. SRTM is widely accepted to be better than the similar alternative ASTER dataset (Rexer & Hirt, 2014; Jarihani et al., 2015).

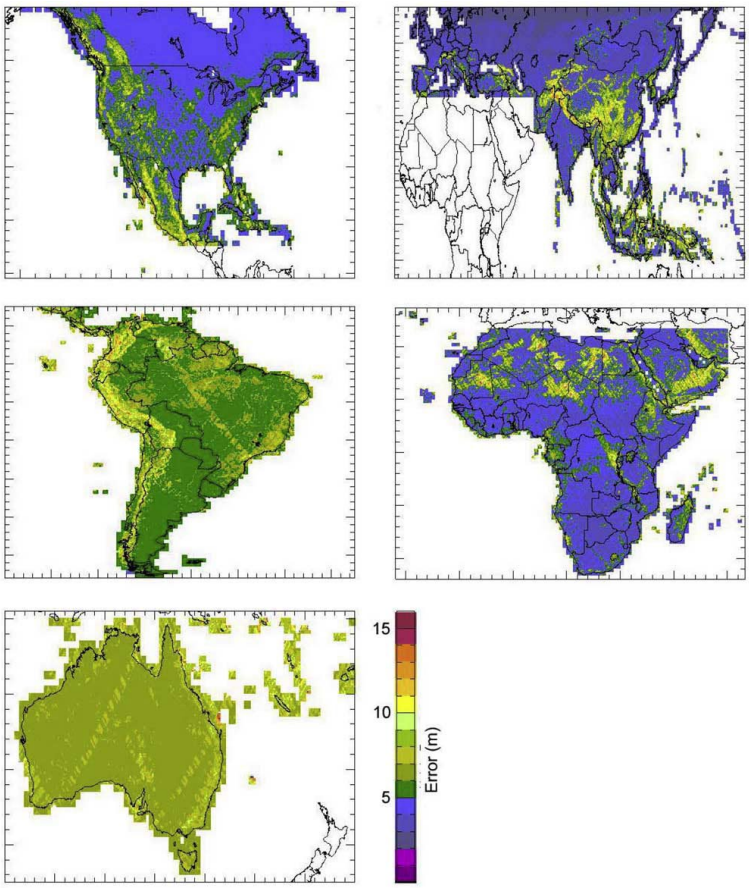

Figure 5-6 - Absolute vertical errors for the five SRTM continents (Rodriguez et al., 2005).

5.2.1.2 HydroSHEDS

The HydroSHEDS dataset, previously discussed in Section 4.1.2, is a processed version of SRTM, available at the lower 90m resolution. This dataset has already been used for validation and will serve as a baseline to compare other DEMs against.

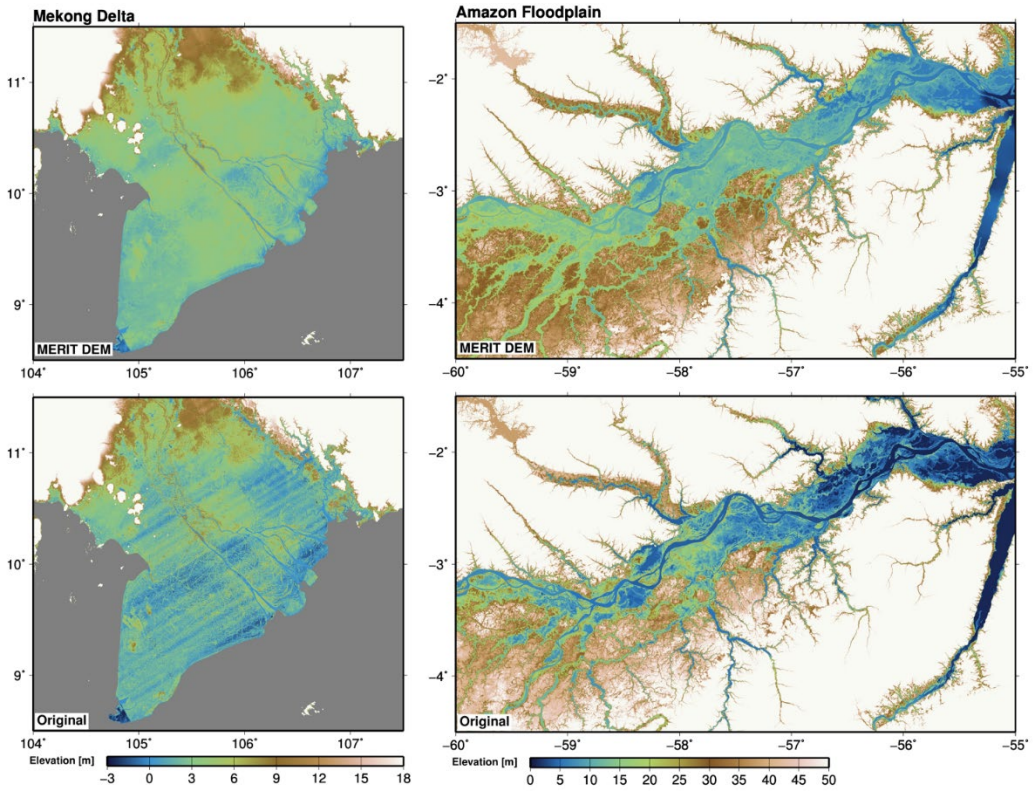

Figure 5-7 - Comparison of MERIT DEM and Original SRTM DEM (Yamazaki, 2018)

5.2.1.3 MERIT

The Multi-Error-Removed-Improved-Terrain (MERIT) DEM, produced by Yamazaki et al (2017) seeks to further improve the accuracy of SRTM, particularly on flood plains. The project aimed to remove speckle noise from surface reflectance, stripe noise from motion errors of the sensor, absolute bias caused by limited ground control points and tree height bias where the radar incorrectly classifies canopies as the land surface. An example of the final product is shown in Figure 5-7. The accuracy of MERIT was tested using ICEsat which is able to penetrate forest canopies. Positive bias in forested areas was found to be reduced compared with the original SRTM data however there was little difference in mountainous regions due to large topographic variability within pixels. 58% of corrected pixels were within 2m of the ICEsat elevations, compared with only 39% in SRTM (Yamazaki et al., 2017). Validation was also conducted across eastern England using the UK 1m LIDAR data where the processed DEM again showed improved agreement compared to the original SRTM data.

5.2.1.4 EUDEM

The European Union DEM (Bashfield & Keim, 2011) was created by Intermap Technologies for the European Environment Agency using a combination of SRTM and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. As part of its creation, inland water bodies and river elevations were flattened and channels were stepped to maintain a consistent downstream negative slope. Channel locations were derived primarily from the IMAGE2006 land cover dataset. Pits and bumps were then manually removed via proprietary user interface software. Using a combination of LIDAR and the privately licenced Intermap NEXTMap DEM, RMSE was found to be approximately 7m (Bashfield & Keim, 2011) but a full analysis of accuracy has not been published.

5.2.1.5 OS Terrain 50

The Ordnance Survey (OS) provides freely available elevation data for the entire of Great Britain at 50m resolution (Ordnance Survey, 2015). OS flies regular imaging missions in light aircraft to collect new data and updates their 'Terrain' products yearly. The final regular gridded dataset is derived from a triangular elevation model (TIN) which is used as it is able to capture edges of features more accurately. All vegetation, buildings and supported structures such as bridges are removed to create a bare earth surface. However, permanent surface structures such as dams, bridge revetments and earthworks are left unmodified. Water bodies are flattened to the height of the lowest surrounding elevation value. To assess accuracy, the DEM was compared with GPS points and found to have a vertical RMSE of 4m. This is the only DEM tested which does not include SRTM in some way.

5.2.2 Inter-comparison

The following findings are from simulations of the period leading up to and including Storm Desmond in 2015, using the five DEMs described above, with 0.03 Manning's n everywhere

and using MIDAS rainfall data interpolated to 1 km resolution. Figure 5-8 shows the spread of modelled peak error relative to the observed discharge peak at gauges with observed maxima over 550 m^3/s. Looking at the median errors only, models using OS Terrain 50 appear to achieve the best results, with a median error of only -0.09. However, the performance of models using OS Terrain is more variable across gauges than of those using HydroSHEDS which have the next lowest median error.

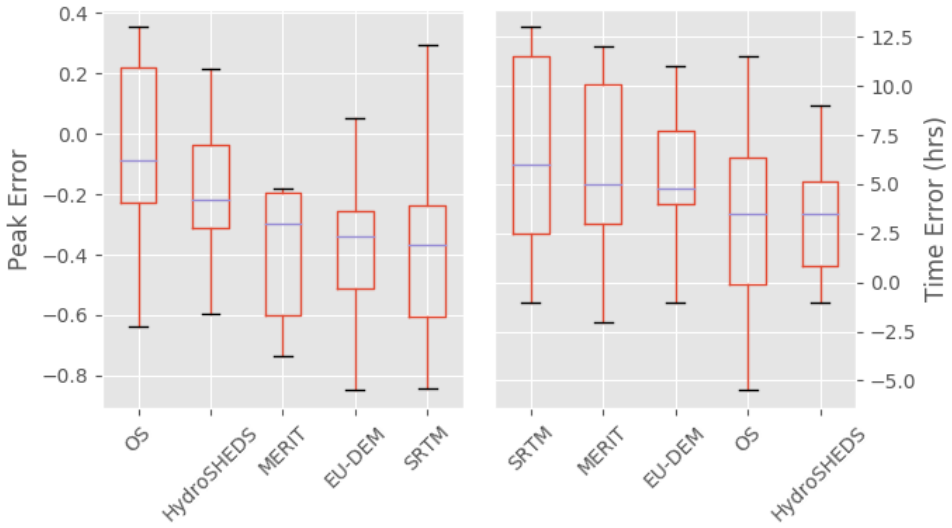

Figure 5-8 - Peak and timing error for selected DEMs using MIDAS 1km rainfall during Storm Desmond.

Models using HydroSHEDS data produced the lowest peak error at the highest number of gauges, as shown in Table 15. In terms of mean absolute percentage error, models using HydroSHEDS outperformed others by a considerable margin, as shown in Figure 5-9, with an average error of only 27%. When the more recently released MERIT DEM was used instead of HydroSHEDS, models surprisingly did not perform as well. HydroSHEDS has been around for longer and it was expected that improvements in model performance would be seen when using the more recently released MERIT dataset. This result could be down to the more manual correction approach taken by the HydroSHEDS team. Models using raw SRTM data produced a peak of zero at a number of gauges, including the Ure at Westwick Lock. This was

due to flow not being conveyed properly in the uncorrected original surface. The median peak errors of models using MERIT, EU-DEM and SRTM data were similar, with MERIT showing the least variance and median error closest to zero. Compared with choice of rainfall dataset, the choice of DEM dataset had less effect on flood peak errors but there was still clear variations. Models using HydroSHEDS and OS Terrain were slightly more accurate in terms of timing, but there was generally less variation in timing than peak accuracy.

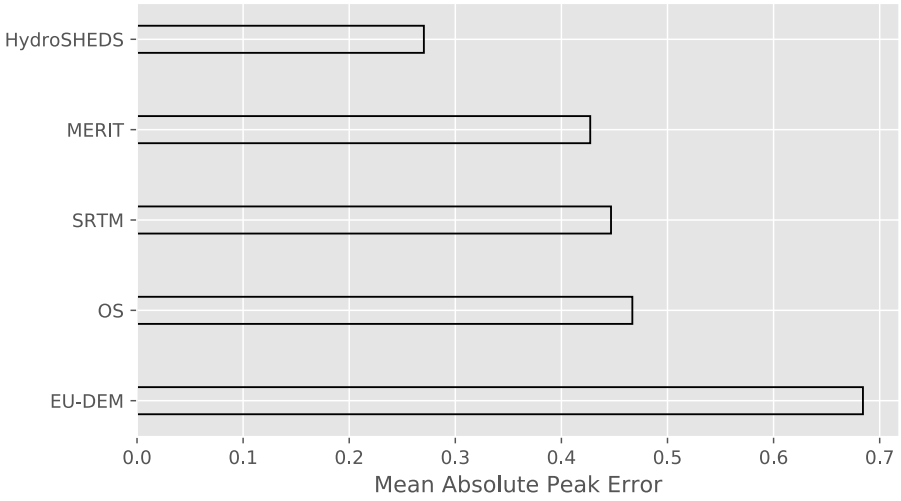

Figure 5-9 - Mean absolute peak error of selected DEMs.

Figure 5-3 indicates that models using EU-DEM data produced better extent performance across all basins, when EA maps were used for validation. This could be down to the fact that EU-DEM is already re-projected into European Lambert Azimuthal Equal Area meaning that there was no pre-processing required to adjust the grid for modelling. There is also a resolution advantage as EU-DEM is originally at 50m but OS Terrain has the same benefit and does not perform as well on extent as it did on median peak error. Among the results from models using the other DEM datasets, there is inconsistent ranking depending on if Sentinel 1 imagery or EA maps are used to calculate metrics. If Sentinel data is used, models using HydroSHEDS narrowly outperform those using EU-DEM. Models using MERIT and SRTM show a significant decline in performance compared to those using other DEMs, when validated

against Sentinel 1. In contrast, when using EA extents for validation, there was a more gradual increase in CSI between models using different datasets. Overall there was more variation in CSI between models using different DEMs than those using different rainfall datasets. This was expected as the shape of the land surface has a more dominant effect on flow pathways, nonetheless is still an important finding.

Table 15 - Peak error for each DEM at each gauge with an observed peak above 500 m^3/s. Lowest absolute error is highlighted in bold.

GAUGE	EU-DEM	MERIT	HYDROSHEDS	OS	SRTM	PEAK
Lune @ Caton	-0.36	-0.35	**-0.19**	-0.27	-0.21	1740
Tyne @ Bywell	-0.30	-0.21	-0.05	**-0.03**	-0.29	1720
Eden @ Great Corby	-0.46	-0.63	**-0.24**	-0.48	-0.59	1490
Eden @ Temple Sowerby	-0.57	-0.73	**-0.51**	-0.64	-0.74	1140
South Tyne @ Haydon Bridge	-0.25	-0.20	-0.22	**-0.19**	-0.25	915
Ribble @ Samlesbury	-0.31	-0.24	**-0.21**	0.22	-0.32	749
North Tyne @ Reaverhill	3.30	**-0.19**	0.22	0.35	0.30	716
Ribble @ New Jumbles Rock	-0.49	-0.53	**0.01**	-0.21	-0.64	657
Lune @ Killington	-0.85	-0.59	-0.59	**0.23**	-0.55	627
Tees @ Broken Scar	-0.26	-0.18	-0.30	**-0.14**	-0.41	611
Eamont @ Udford	-0.80	-0.66	-0.36	**0.07**	-0.84	581
Tees @ Low Moor	0.05	-0.19	**-0.01**	2.40	0.22	569
Ure @ Westwick Lock	-0.90	-0.84	**-0.60**	-0.84	na	513

5.2.3 Resolution

To analyse performance sensitivity to grid resolution, models were run using the same DEM, resampled to between 50 and 150m, at 10m intervals, using nearest neighbour interpolation. Out of the selected products, only EU-DEM and OS Terrain reach 50m resolution, therefore HydroSHEDS, SRTM and MERIT were excluded from this analysis. Only a single catchment was used to assess the effects of resolution, as reducing cell sizes below 90m greatly increases simulation run times. The South Tyne at Haydon Bridge was selected for this purpose due to its relatively small size, high observed peak during Storm Desmond and good performance during initial validation. Simulated peaks are influenced by the number of cells used to calculate them. This means interpretation of any trends in peak performance metrics is

difficult as the number of cells used can differ between resolutions. Figure 5-10 illustrates this with a lack of any significant trends when using peak error as a metric.

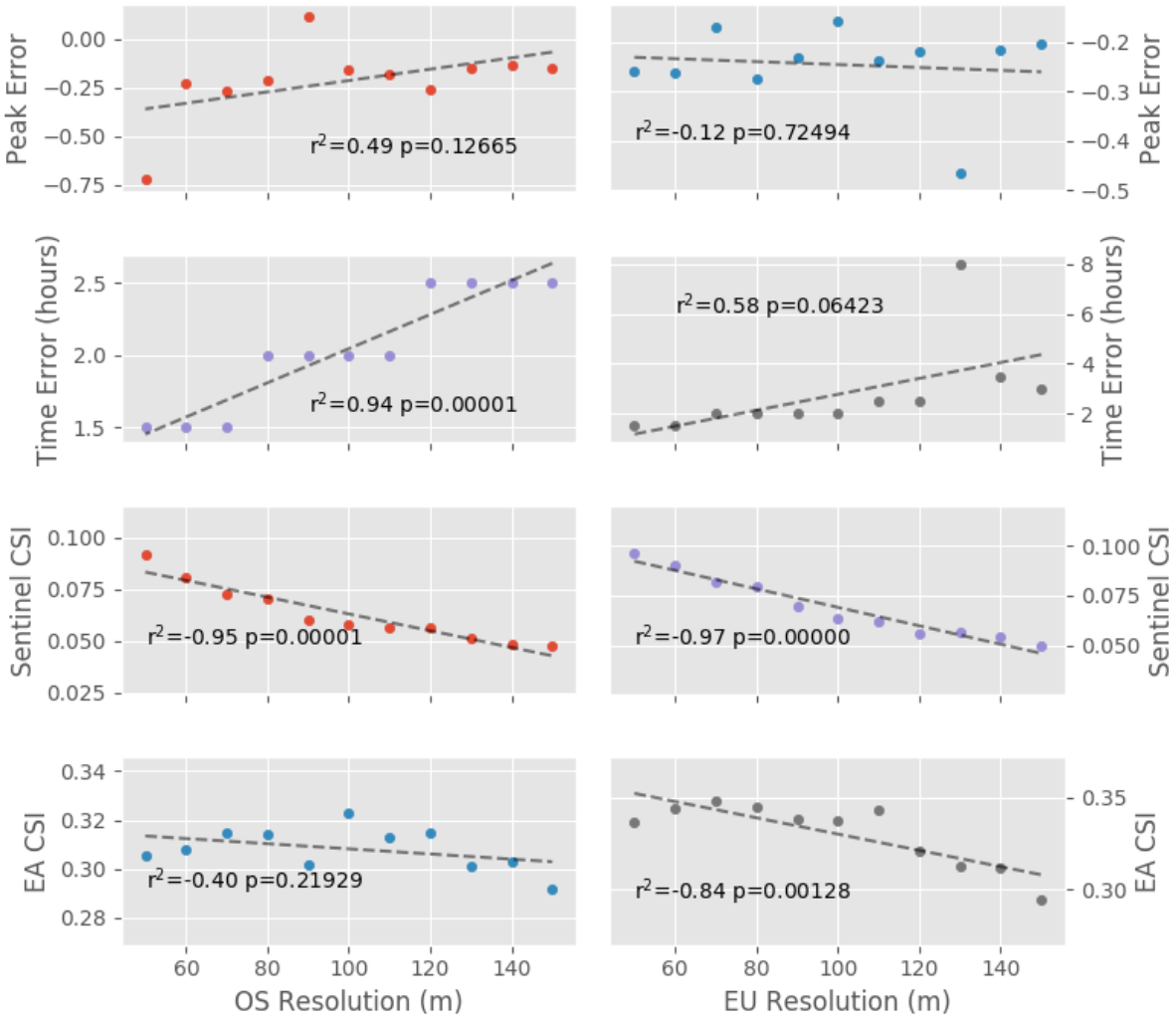

Figure 5-10 - Effect of DEM resolution on flood peaks and extents for the South Tyne at Haydon Bridge.

The outlier at 130m using EUDEM is caused by extra cells downstream of the virtual gauge which are not present at the other resolutions as a consequence of the relative position to the origin of the coordinate system. Another outlier at 50m resolution using the OS Terrain data is possibly caused by spurious storage in artefacts that are removed by resampling the grid to a lower resolution.

One explanation for the general lack of peak accuracy improvement with increased resolution, aside from the discharge calculation methodology, could be related to friction. As the number of cells in the domain increases, the precision with which flow channels are represented also increases, meaning they are less straight. If channels are less straight they experience more friction, slowing the flow and reducing the peak. Water also takes longer to travel to the gauge if the channel is less straight as the travel distance increases, increasing the time to peak and therefore diffusing the maximum flow. This effect is counteracted by a reduced relative influence of friction on each channel cell at the higher resolutions, as water depths within the channels become greater. Lowering the resolution also has a smoothing effect on the DEM itself which would act to increase flood peaks.

These effects combine to result in limited differences in flood peaks with changing resolution. In terms of timing however, the results show a clear slowdown of the flood wave at lower DEM resolutions, particularly when using the OS Terrain product. This is not consistent with the idea that larger cell sizes reduce roughness and cause a speedup in flood wave propagation as the peaks are clearly arriving later at lower resolution. It seems that increased distance to the gauge or surface flatness are not in fact contributing significantly and faster conveyance due to greater depths within channels at higher resolution is more important. Looking at extent metrics, in all cases apart from when using EA data to validation results from models using OS data, a significant negative trend is apparent with CSI. Therefore, increasing resolution has a clear positive effect on the similarity of model results to observations and the local EA model results. As the extent from Sentinel 1 is at 10m resolution, this can be expected as the higher resolutions tested were more similar to the reference data. In the context of extent accuracy, DEM resolution is perhaps the most important factor, however as the flood peak results have shown, this does not necessarily correlate with accurate depths and velocities.

5.2.4 Resampling Method

Each DEM had to be resampled to ensure each grid cell remained in the same position across datasets and did not vary in size. However, there is more than one way to resample DEMs and the choice of resampling method inevitably influences model performance. Therefore a range of methods were tested, including nearest neighbour, average and bilinear, using the command line utility gdalwarp (GDAL, 2018). Resampling is carried out during the re-projection process so cell locations are altered only once. The default nearest neighbour (NN) approach takes the closest point on the original DEM to each new cell and transposes it directly, keeping the value the same. A second method takes the average of a window around each new cell from the original values. Finally bilinear is the most sophisticated method tested and weights the average of the cells within the window using their distance to the target cell in the resampled DEM. NN can result in some spurious artefacts but prevents any of the smoothing effects which are caused by the other methods. Both average and bilinear resampling cause the DEM to become more smooth, speeding up the transport of flood water and leading to higher peak discharge and a shorter time to peak, as shown in Figure 5-11.

Looking at mean absolute peak error (Table 16), models which used nearest neighbour resampling achieved increased performance. However this is largely contributed to by the North Tyne at Reaverhill, where the flood peak is grossly over-estimated when using average and bilinear resampling. These outliers are likely to be an effect of the discharge calculation methodology and virtual gauge placement, therefore the mean values are misleadingly skewed. Moreover, NN has the highest median absolute peak and timing errors, with bilinear slightly outperforming average resampling. The consequences of the lack of smoothing in NN are apparent in the significantly delayed timing, however in many cases, such as the Tyne @ Bywell (Figure 5-12), this has not affected the peak value. In terms of extent, models using average and bilinear resampling produce almost identical results to two decimal places (Table

16). Models using NN have higher HR but also higher FAR, due to a larger maximum extent causing a slight decrease in CSI compared to those using the other two methods. Thanks to its largely increased processing speeds and reduced smoothing effect, NN resampling has been used throughout this thesis, however bilinear resampling may provide improved model performance.

Table 16 - Extent performance metrics and mean absolute peak error (MAPE) for resampling methods.

		EA			SENTINEL		
Resampling method	**MAPE**	**CSI**	**HR**	**FAR**	**CSI**	**HR**	**FAR**
average	0.29	0.31	0.53	0.45	0.13	0.51	0.86
bilinear	0.29	0.32	0.52	0.43	0.13	0.50	0.85
near	0.24	0.29	0.54	0.49	0.12	0.51	0.87

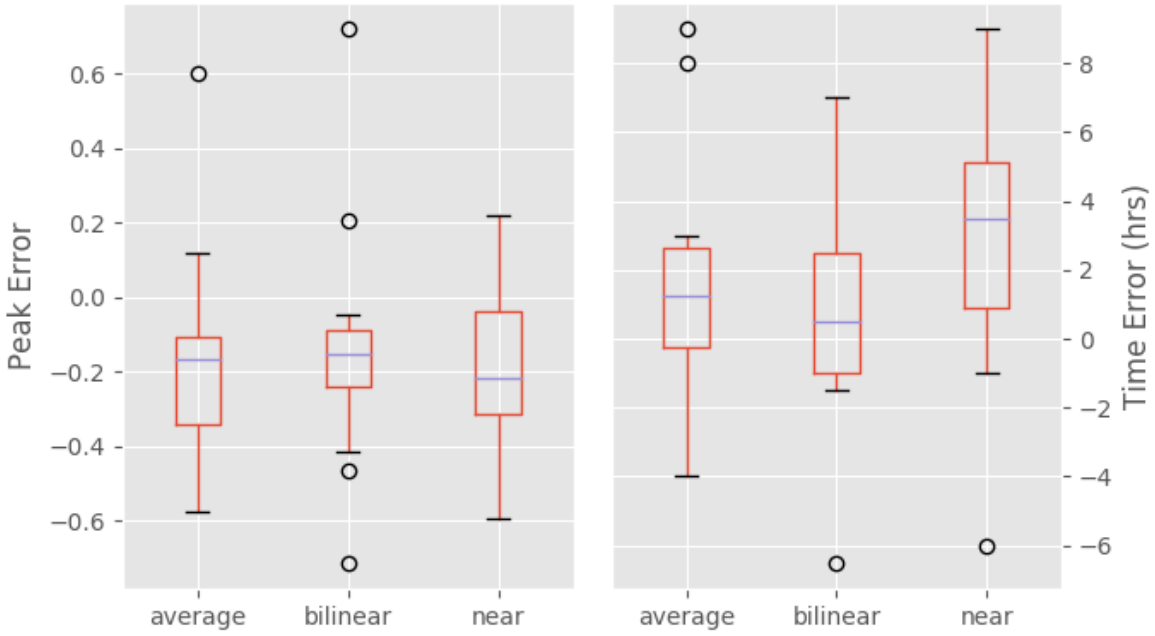

Figure 5-11 - Effect of resampling method on peak error and timing.

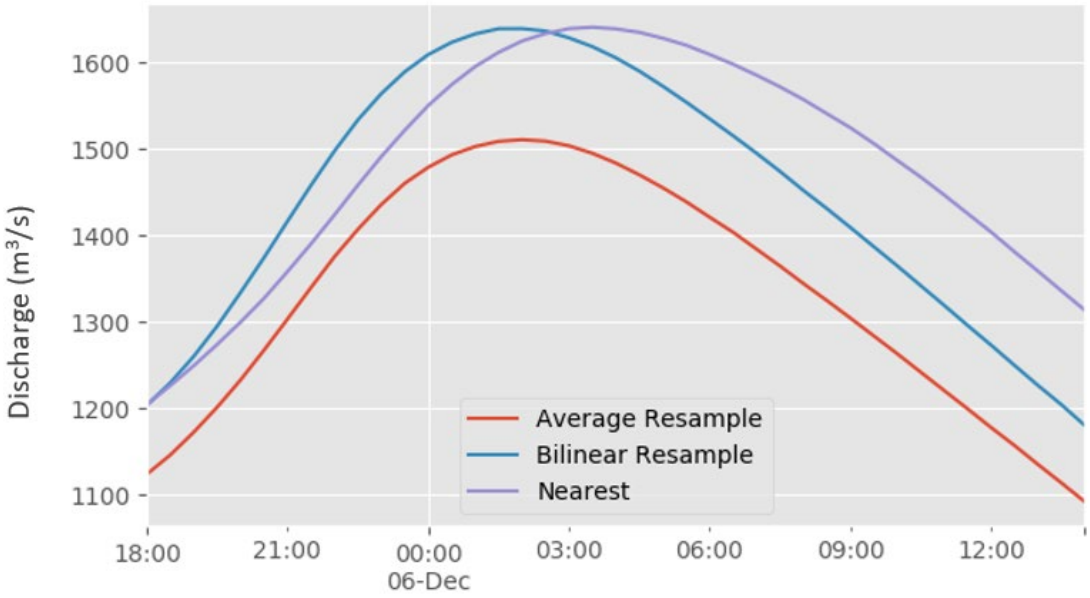

Figure 5-12 - Example hydrographs at the Tyne at Bywell showing the effect of resampling method on peaks (m³/s).

5.3 Friction

The Gauckler-Manning coefficient of friction (n) is a key parameter in any hydraulic or hydrological simulation. Its value is usually either calibrated or determined using the relationship between flow velocity, hydraulic radius and channel slope. The value of n on flood

plains has been fixed here at 0.03, which is the standard value for pasture with short grass and cultivated areas with no crop from Chow (1959), shown in Table 17. To assess the effects of n within channels, simulations have been run using a range of values, keeping all other factors constant. In all cases, the original HydroSHEDS DEM has been used as it was demonstrated in Section 5.2.2 to provide the best performance within the study area.

Table 17 - Manning's n values for flood plains (Chow, 1959).

DESCRIPTION	MINIMUM	NORMAL	MAXIMUM
a. Pasture, no brush			
1.short grass	0.025	0.030	0.035
2. high grass	0.030	0.035	0.050
b. Cultivated areas			
1. no crop	0.020	0.030	0.040
2. mature row crops	0.025	0.035	0.045
3. mature field crops	0.030	0.040	0.050

Figure 5-13 shows a reducing flood peak relative to the observed as channel friction increases. This was expected as more friction leads to slower flows, which both directly affects discharge and increases the time to peak. Slower flows mean the flood wave will be more dissipated, causing a lower maximum discharge. Looking at the effects on peak timing, also shown in Figure 5-13, the trend is even clearer. In both cases, 0.01 would appear to be an optimum value, however this could be due to its compensating effects on other factors that may be slowing down the flow or decreasing the peak, such as inaccurate precipitation or elevation data, or numerical dispersion. For this reason, calibration using Manning's n has not been carried out. Friction is also affected by sediment on channel beds. If flows become sufficiently high, this sediment will become entrained, reducing friction levels. However, the model is currently not capable of representing a time variable value for n in this way, therefore n is likely to be over-estimated during high flows, meaning that peaks will be under-estimated.

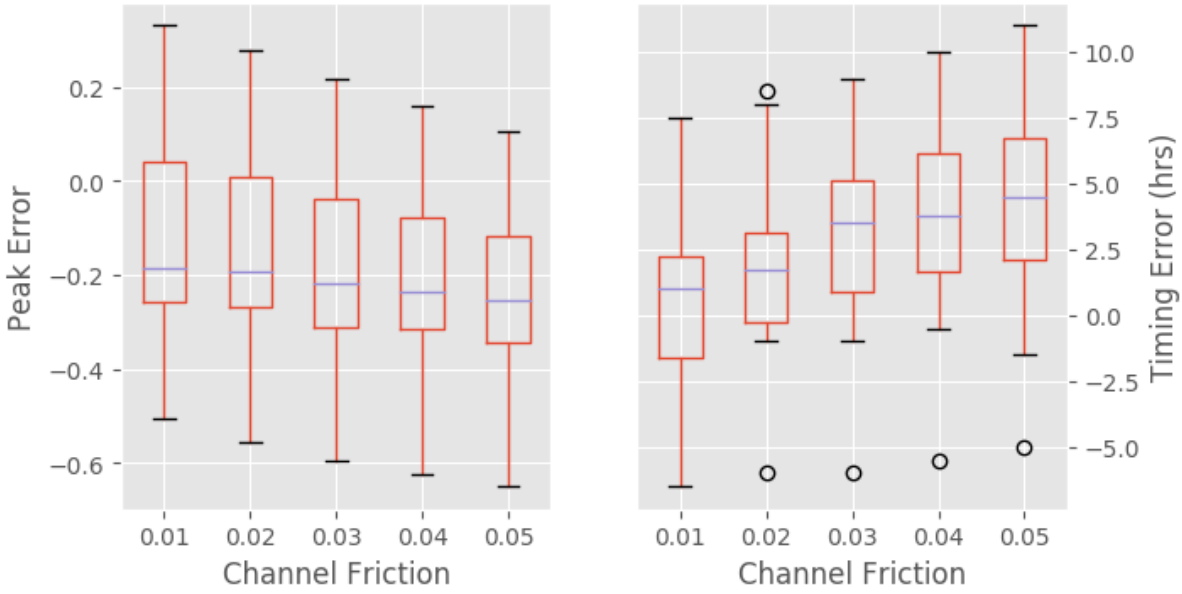

Figure 5-13 - Effect of channel friction coefficient on flood peak and timing error.

The effect of varying channel friction on extent metrics was negligible, with a variation in CSI of less than 0.005. There was slightly more of an effect on FAR and HR, however as they both increased by a small amount, with FAR increasing more, this led to a marginally increased CSI. The increased HR is caused by more water leaving the channels as friction prevents the flow moving fast enough and causes a backwater effect meaning the flooded extent becomes larger. For example, the maximum flooded area increased by 3% for the Tyne basin by only changing the channel friction parameter. However these additional flooded cells frequently are not aligned with cells detected as wet by Sentinel 1 and therefore increase the FAR.

5.4 Channels

To assess the effects of channel characteristics, the event was simulated first without any channels present in the DEM, then with rivers burned in and finally using a median filter with a 3x3 window on cells categorised as being within channels to smooth the bed. The methods used are described in more detail in Chapter 3. The original SRTM dataset contains noise characteristic of any uncorrected radar-derived elevation data. This noise causes artificial

roughness in channel profiles, slowing down flood wave propagation. The HydroSHEDS project did attempt to remove this effect by flattening rising reaches to prevent a downstream uphill slope but only processed channels with an upstream area over 8 km. Therefore, the effect of an additional smoothing step on models using the HydroSHEDS dataset was tested by applying a median filter to the DEM within channel boundaries.

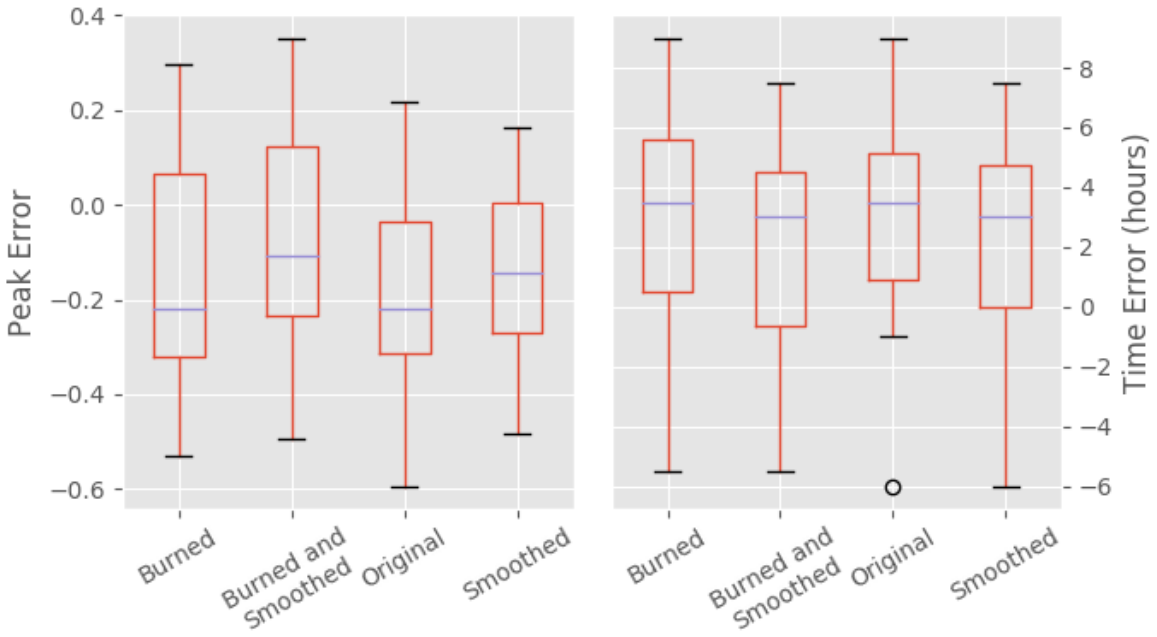

Figure 5-14 - Effect of channel burning and smoothing on peak and timing error

Figure 5-14 shows the consequences of these modifications for peak error and timing, followed by an example of the effects on hydrograph shapes in Figure 5-16. Burning a channel network led to generally increased peaks, as expected, but the median peak error value was almost identical to when the original DEM was used. Models with smoothed channel profiles did produce lower absolute peak errors though, and the variance in accuracy was also slightly reduced. Meanwhile, smoothing of river channels also reduced time delays more than burning. The median timing error was the same for models using a burned and non-burned

DEM. Burning and smoothing of channels had more effect on median flood peak error than median timing error.

In terms of extent, there was a marginal difference between performance scores from the three groups of runs, as shown in Figure 5-15. Models with a burned DEM performed the worst, followed by those using the original HydroSHEDS data and finally, the runs with smoothed channels showed a small increase in extent performance. Models with channels that were both burned and smoothed also performed worse than those using the original unmodified DEM. Although the margins are very small, these results indicate that smoothing channels could improve performance while burning may reduce it. However, this finding is limited to the HydroSHEDS DEM and datasets with no pre-processing may benefit from burning.

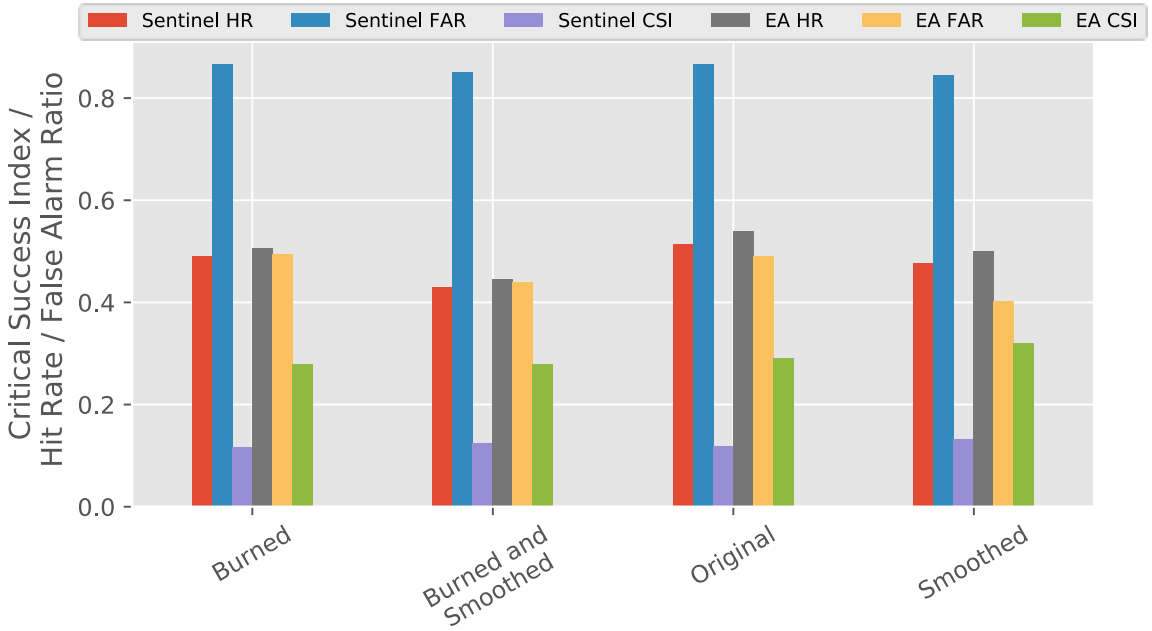

Figure 5-15 - Effect of channel burning and smoothing on mean flood extent performance.

Hit rate was similar between models with burned channels and those with smoothed channels, with the difference in CSI being caused by over-estimation of extent in the burned

models. This is counter-intuitive as deeper channels would be expected to reduce inundation. The effect could be caused by increased floodplain storage in the upper reaches of the basins when channels are not burned in. Combining smoothing and burning did not sum their effects and produced similar results as either in isolation, in terms of both peak timing and extent metrics. Although maximum discharge at some gauges was increased by combining channel smoothing and burning, as shown in Figure 5-16.

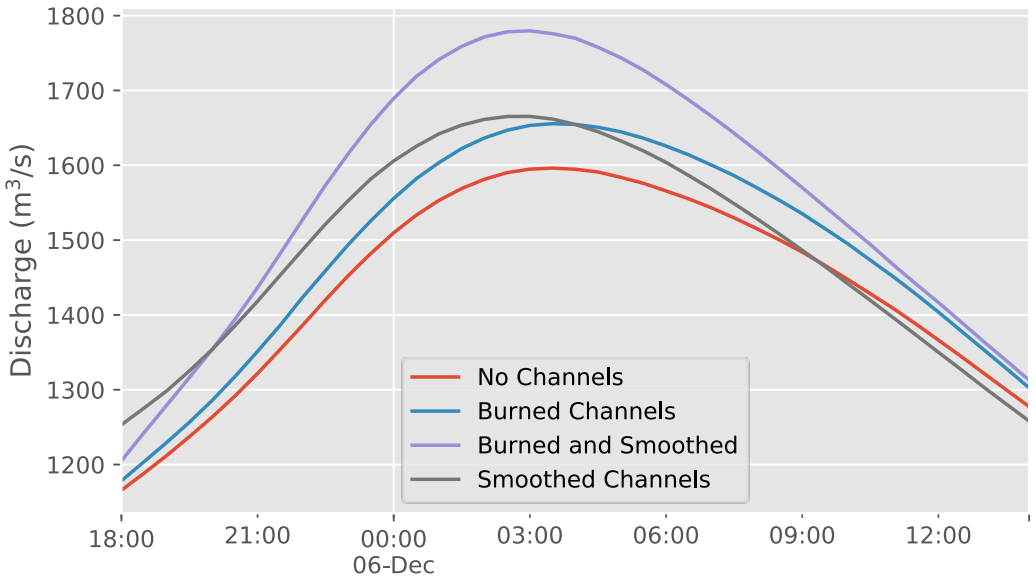

Figure 5-16 - Hydrograph examples for the Tyne at Bywell, demonstrating the effects of channel burning and smoothing. The y axis shows discharge in m^3/s.

5.5 Limitations

All DEMs contain inherent limitations as they attempt to capture a 3D surface in 2D. This means complex features such as bridges and built up areas are not represented perfectly. Sensors are also usually unable to penetrate water surfaces, meaning lakes and rivers need to be accounted for after the primary data collection process. When more complex structures are present, such as dams, weirs or locks, this becomes challenging and if the resolution is not fine enough, they are completely missing in the DEM surface. Therefore these structures are

very difficult to represent in broad-scale flood models which use standard 2D DEMs. 3D surfaces, such as point clouds, would enable better representation of these features, however they require detailed surveying which is expensive and not feasible across such large areas. River channel geometry is also not represented accurately within the model due both to data limitations and the grid resolution not being commensurate with the width of the channels.

5.6 Summary

Models using five different freely available DEMs and precipitation products have been tested using observations of discharge peaks and flood extents during Storm Desmond across Northern England. Out of the global elevation datasets used, HydroSHEDS has been demonstrated to be capable of producing the best performance scores, in terms of both flood peaks and extent and consistently outperformed models using MERIT and SRTM. Models using the OS Terrain product were similar in peak performance to those using HydroSHEDS, however this data is only available within the UK. Models using EU-DEM data excelled on comparison of maximum extents, however on average produced the least accurate discharge peaks.

A number of other factors related to the land surface have been found to influence performance. Reducing friction and burning and smoothing channels have been shown to increase and speed up flood peaks, as expected. The resampling method used when re-projecting DEMs also affects performance. Both average and bilinear resampling resulted in higher and faster peaks than nearest neighbour, due to the smoothing effect they have on the surface, though the effect on CSI was marginal. Man-made features such as bridges, dams and embankments are not properly captured by broad-scale DEMs, therefore more localised studies may be required if these features have a significant impact on the area of interest.

The main conclusion from precipitation analysis was that models driven by MIDAS data significantly outperformed all others using more widely available gridded datasets. Driving

models with ERA Interim data consistently produced the worst performance scores. ERA5, the ECMWF's update to ERA Interim, did produce improved model performance compared to its predecessor, which is promising for the future of re-analysis products. Models driven using MSWEP occasionally created spurious early peaks which caused the timing performance of these models to be much lower than when using other datasets. In general, model extent performance metrics were even less varied when using different precipitation datasets than DEMs.

The variation between possible flood model outputs, demonstrated in this chapter, shows that choice of input datasets and pre-processing steps should always be properly considered when conducting broad-scale modelling experiments or assessing the implications of results. Looking to the future, as new DEMs and precipitation products are released, it is important to test them on the same benchmarks as their predecessors and consider their accuracy specifically for the intended use. Examining a single event allows a clear and concise comparison of the available data, however in other events, the comparative performance between each may vary. Therefore, a useful direction for further research would be an analysis across many more events to create a more event independent summary of which are better. The next chapter advances to look at some examples of impact metrics which can be calculated using flood model outputs from this thesis, and their sensitivity to the choice of datasets compared above.

Chapter 6. Applications

Global Flood Risk Models (GFRMs) are increasingly used by the insurance industry to set premiums and make investment decisions (Ambiential, 2019; JBA, 2019; Swiss Re, 2019), by development planners at the local, regional and national level (de Moel et al., 2015) and by emergency responders during broad-scale events to get assistance to affected populations (Revilla-Romero et al., 2018). The flood hazard map products used by insurance organisations are usually proprietary and uncertainty is therefore generally under-communicated due to its negative effect on how the product is perceived. This is tempered by the hedging approach commonly taken in purchasing a range of flood hazard products from independent sources and combining them to reduce reliance upon a single potentially flawed methodology. In the context of insurance pay-outs, minor differences in the size of the inundation area will not greatly influence the cost to the company and therefore they are not necessarily very interested in extent precision. The more pressing question in this application is how frequently these events occur, a problem which has not been addressed directly here and which is affected by a more intractable form of uncertainty, subject to the lengths of observed records and (in)ability to forecast into the future.

However, in contrast to insurance industry applications, planning of new infrastructure and defences of existing infrastructure is much more sensitive to the exact location and depth of potential flood extents than their likelihood. Response during an event is also a suitable application, however as an event is taking place, there is often not enough time to provide actionable results which can affect decision making and collection of live data is usually more appropriate. Planning, though, is something which allows plenty of time for consideration of options and deciding upon the best methods and data to use when considering flood risk.

Particularly in developing nations where networks are still emerging and accurate information about flood hazard would allow for smarter planning for more resilient infrastructure.

However, there are a number of problems with using outputs from broad-scale models to achieve an understanding of risks to assets and populations. The primary concern is compounded uncertainty when uncertain flood model outputs are combined with uncertain data describing people and infrastructure. The inaccuracy of flood models is not only added to but multiplied when receptor data is incorrect. To exacerbate this problem, validation data is almost always unavailable at a level of detail commensurate with that produced by any impact models. Furthermore, studies are frequently interested in impacts into the future, where validation is not possible and an array of climate projections mean an array of possible impact scenarios, each with its own associated uncertainty. Nevertheless, assessments of flood risk are required to enable informed planning, therefore uncertainty needs to be highlighted while providing precise enough findings to enable decisions based on data.

When looking at impacts to infrastructure, there is frequently an added layer of complexity as dependencies exist both within and between networks such as transport, energy and communication. This reliance means that the impacts of flooding in a specific area can extend far beyond the region of inundation. For example, a collapsed bridge will put increased traffic pressure on alternative routes or an important node in a power network, if flooded, may remove power from an entire neighbourhood. The essence of this problem is that not all elements in an infrastructure network have the same importance and if critical nodes are at risk of flooding, this may be far more important to avoid or remedy than protecting an entire group of less important nodes. To address this, hazard models can be combined with network analysis to get a holistic picture of flood risk. When used for this purpose though, any errors

in flooded extent will be magnified by the interconnectedness of a network model and therefore it is even more vital to understand and characterise hazard uncertainty.

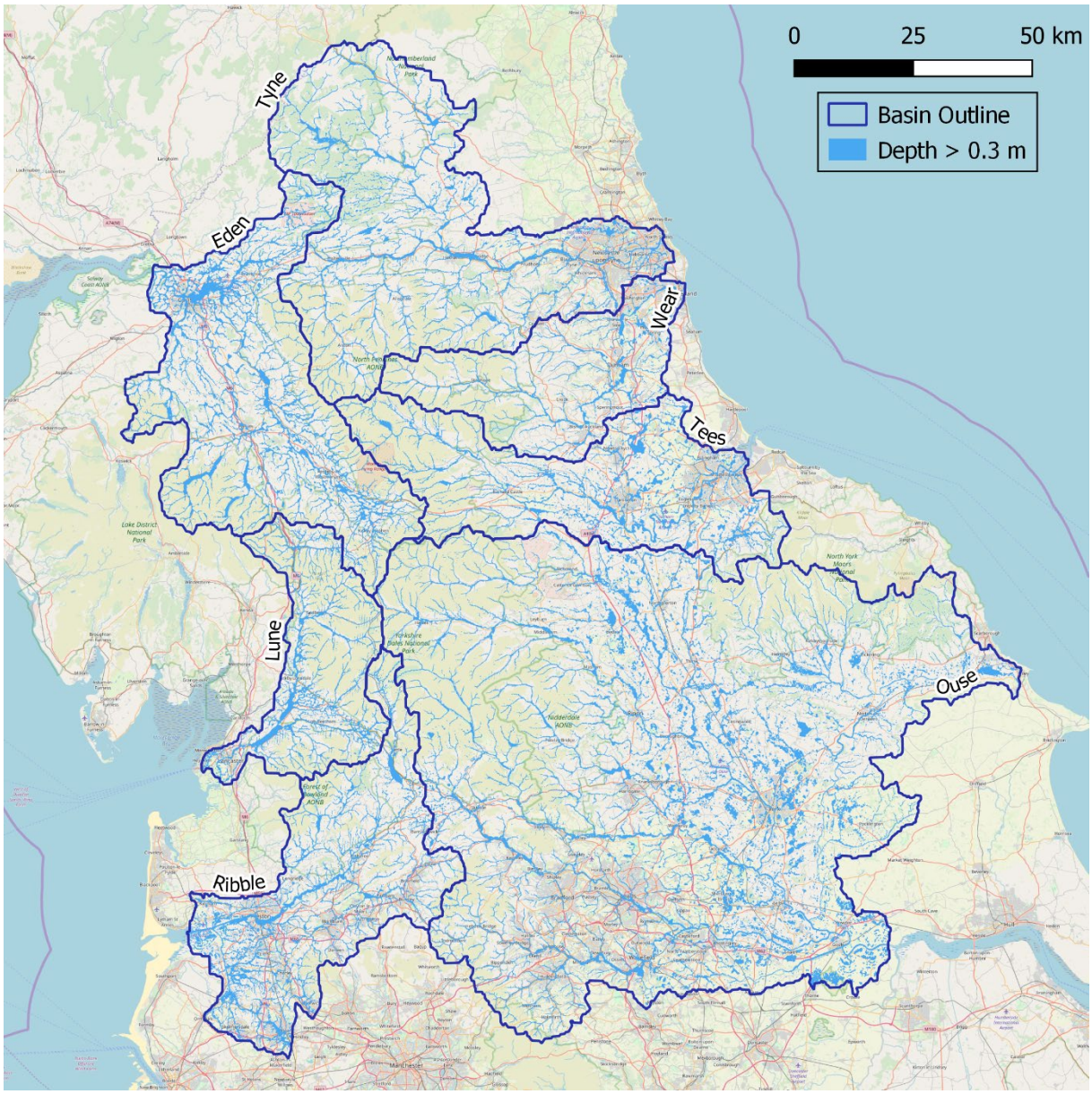

Figure 6-1 - Flood hazard layers produced Chapter 4, showing regions with maximum water depth greater than 0.3m.

6.1 Impacts on Infrastructure

As a first step towards a complete analysis of impacts, static infrastructure, including roads and buildings, was intersected with flood hazard layers, produced in Chapter 4 and Chapter 5. Figure 6-1 shows the general extent of flooding throughout the study area in Northern England. Roads and buildings were chosen due to their prominent availability in freely available datasets. Power and communication data, for example, tends to be more secure and

therefore not readily available for download. However the methodology applied to roads and buildings here is easily transferable to any other network data where it is available.

6.1.1 Data

OpenStreetMap (OSM) is a globally available source of geographical data describing features on the surface of the earth. Each feature exists as either a node, or a group of nodes, connected by a "way". Groups of ways, such as bus routes, are represented as relations. Fundamentally, OSM is a volunteered dataset in that individual users contributed information for free, in much the same way as Wikipedia. In many areas, this data is supplemented by governmental or freely available data. In the UK, the main source of non-volunteered information is the Ordnance Survey. This means the uncertainty present in the dataset for many of the less developed regions of the world is not as much of a factor in the UK and the dataset is seen to be reliable enough for testing the modelling framework developed in this research. The resolution of flood model results is also low enough that exact road locations will not have a significant influence on the results relative to the uncertainty in flood extent. The entire dataset can be freely downloaded from various online platforms and either used directly as XML or imported to a postgres database with a pre-defined schema. This process has been automated by a number of freely available command line tools.

6.1.2 Methods

To define features as either flooded or not, a threshold depth is required. The default depth used here for roads was 30cm, based on a consensus from the literature that this is the depth at which the average car can no longer make progress (Pregnolato et al., 2017). Undoubtedly depths lower than this value will delay traffic and may entirely prevent some cars from passing, however this effect is not included here. The roads data in OSM includes an array of

road sizes, all of which are not necessary for the analysis. Therefore primary roads were selected based on the metadata included in OpenStreetMap.

Buildings are more complicated to ascertain a uniform depth as it is depended on uses of individual floors within the building and further how vulnerable the building is to flood water both in terms of defences and materials used for walls, flooring and furniture. For example, if the ground floor is tiled and enough warning is given, damage to property may be minimal. Basements also prove difficult to represent effectively in any analysis, as water from the ground floor may move to the basement which continues to fill. To simplify interpretation, these effects were excluded here and the threshold depth for buildings was also set to 30cm as standard. There is also large variation in building value, which is generally not fully accounted for due to a lack of data. Therefore analysis of building impacts has been constrained to a count of affected properties as a first step. An important factor in flood impacts is the duration of inundation, which can only be accounted for if the analysis comprises a time dimension. Therefore analysis of impacts is undertaken at regular time steps throughout the flood simulation.

6.1.3 Results

Figure 6-2 shows the location of flooded areas relative to major roads and settlements within the Tyne basin. The largest regions of flooding are located near Hexham, Haltwhistle and Prudhoe and intersect parts of the A69. As can be seen in Figure 6-3, a large distance of roads and number of buildings became inundated before the main event even took place. This is largely due to the 90m resolution of the hydrodynamic model outputs. The length of roads is more sensitive to changes in inundated extent as once a large building is touched by a smaller extent, a larger extent will not affect its contribution to the total, whereas increased inundation of a single road will add to the total inundated distance. The road length inundation

series therefore follows the hydrograph shape more clearly than the building counts. The peaks in both inundated buildings and road distances occur slightly before the discharge peak at Bywell. This could be caused by the relatively downstream location of the gauge.

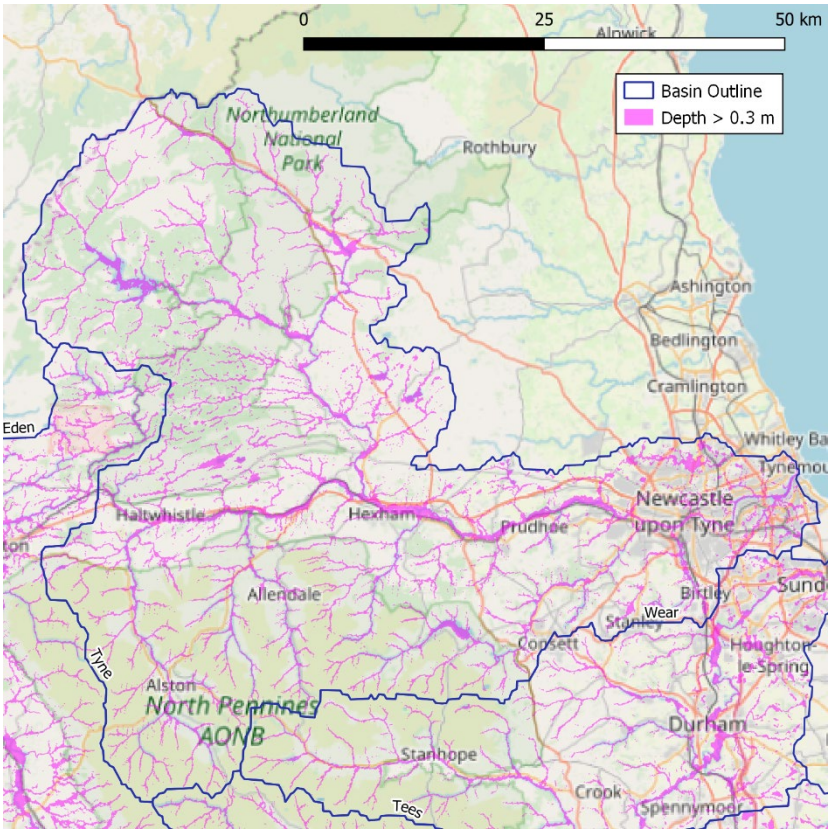

Figure 6-2 - OpenStreetMap data for Tyne basin overlaid with modelled flood extent for Storm Desmond.

Unfortunately, solving the initial over-inundation problem by masking out locations where rivers exists would not work. This is because the cell size of the model grid means that some buildings and roads exist within the gridded representation of the river channels. If these areas were excluded, the buildings and roads located on the river banks would never become flooded and this would result in an under-representation of the impact. The only feasible solution to this is to increase the flood model resolution, however there are no higher resolution globally available, corrected DEMs available for public use at the time of writing. The model used is also limited by the speed at which it can read from and write to memory and therefore is not able to scale up to very high resolutions across such large areas. Therefore a simplified version would have to be adopted in at least the majority of areas.

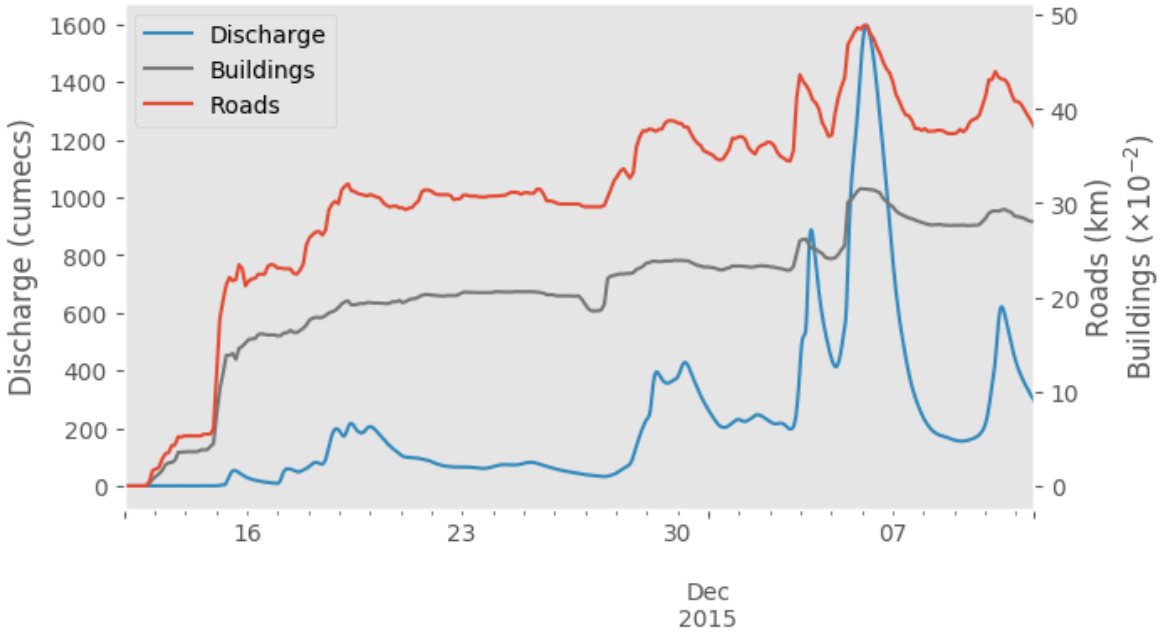

Figure 6-3 - OSM primary, trunk and motorway roads and OSM buildings inundated by more than 0.3 m of flood water in the Tyne basin. The hydrograph shown is from the corresponding model results at the Tyne at Bywell using HydroSHEDS elevation and MIDAS rainfall data.

The effect of depth thresholds on the number of buildings and distance of roads inundated was also assessed. Intersections were made from 10-50cm at 10cm intervals and the results can be seen in Figure 6-4. The shapes of both curves indicate that there is not a critical threshold within the bounds tested. There is a similar proportional difference in impact on the inundation levels between roads and buildings for the different inundation thresholds. This proportional difference is considerable, approximately 30% for both roads and buildings between 10cm and 50cm. The threshold of 30cm has been chosen as a default setting as this corresponds to the generally accepted depth at which cars cannot make progress, however it should be considered that the final results are highly sensitive to this value and 0.3m may be inappropriate for certain applications.

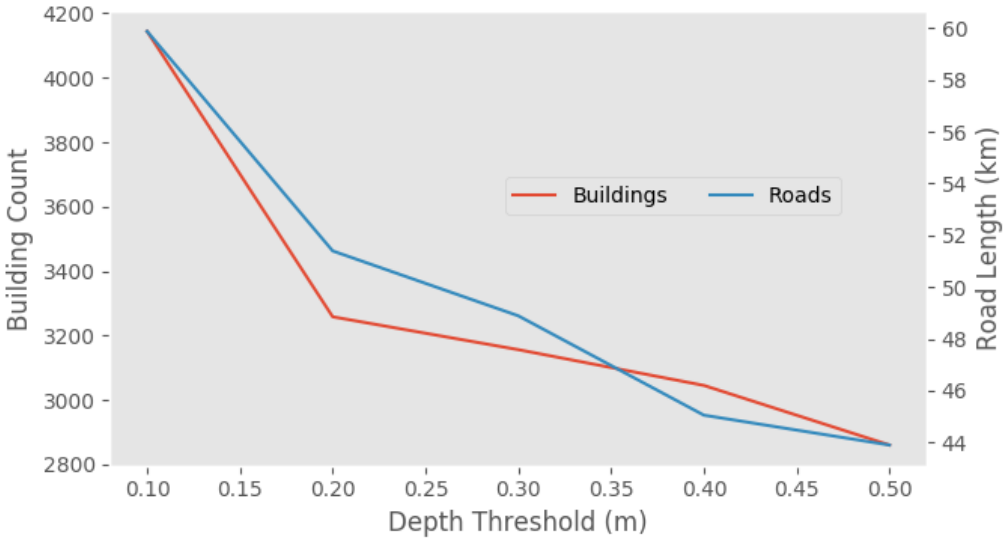

Figure 6-4 - The effect of depth threshold on building and road inundation in the Tyne basin.

To assess the distribution of impacts between regions, peak inundation is compared across the studied basins in Figure 6-5. As the amount of inundation is highly proportional to the size of the basin, these results have been standardized by dividing the road distances and building counts by the area of each basin in km^2. The Lune has by far the highest density of flooded roads with 33 m/km^2. Other basins were generally between 10 and 20 m/km^2. The lowest proportional road inundation was found in the Eden basin. There is a larger amount of variation in the roads results between basins, ranging from 0.3 $/km^2$ in the Tees to over 1.75 $/km^2$ in the Ribble basin. There is undoubtedly a high degree of uncertainty associated with any comparison, as discussed in previous chapters, however the ranking of basins gives a functional picture of which regions were most affected by the event.

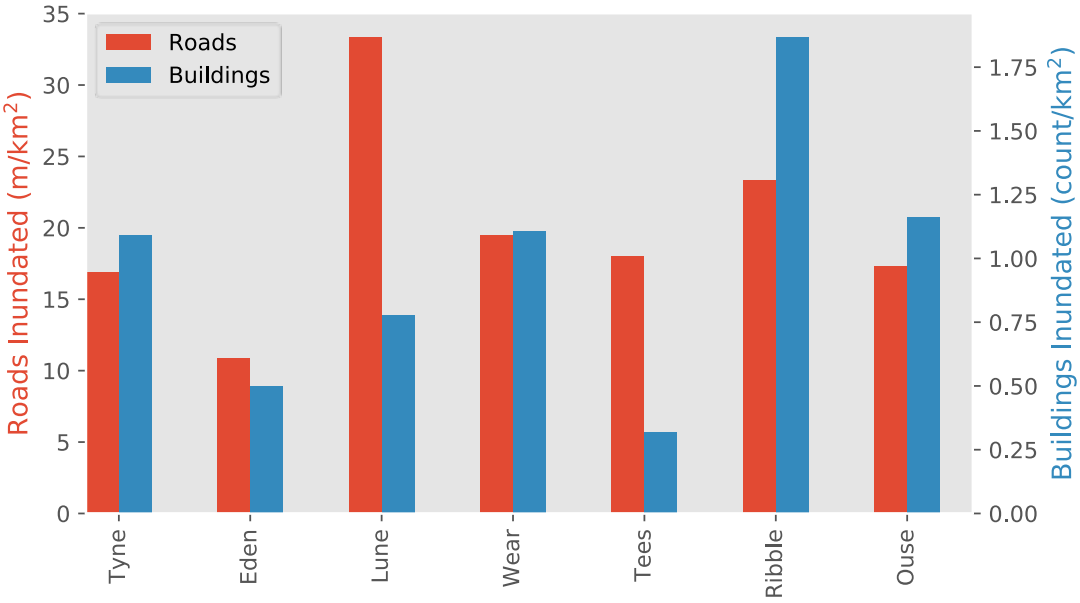

Figure 6-5 - Peak roads and buildings inundated across all modelled basins.

6.2 Sensitivity to Flood Model Inputs

Using a single set of flood model results provides a useful preliminary assessment of impacts, however does not allow for uncertainty associated with the choice of input datasets. Therefore, following the initial calculation of impacts with a single set of flood model results, the same process was repeated using outputs from flood models using each rainfall and DEM dataset discussed in Chapter 5, for the Tyne basin. Additional uncertainty from model parameters also exists, however is not discussed here due to the lack of calibration data describing infrastructure inundation and the relatively far less constrained bounds of possible variables. Throughout all simulations, friction remains 0.03 everywhere, at a grid resolution of 90m re-projected using NN resampling and with no additional modifications to the HydroSHEDS DEM. The depth threshold also remains at 0.3 m. As the process of extracting flood depths from the database takes a non-trivial amount of time for each run, depths were extracted at 6 hourly intervals rather than at every time step.

6.2.1 Digital Elevation Models

Figure 6-6 shows a comparison of inundated road distances and building counts using models based on each of the five DEMs tested in Chapter 5, driven using MIDAS rainfall data. The most notable result is that the model using HydroSHEDs data produced a considerably lower peak road inundation distance than models using other datasets. This could be a result of the included channel pre-processing to ensure downstream connectivity, reducing the backwater effect and therefore inundation of the flood plain. Models using EUDEM, SRTM and MERIT created the largest amount of road inundation, all close to 80 km. The most pronounced peak in road flooding, relative to the pre-storm period, was produced when EUDEM was used, to such a degree that the preliminary smaller peak, clearly visible in the other series, is almost indistinguishable. In the initial spin-up period, the model using OS-Terrain data causes the least impact, but then overtakes the HydroSHEDS model to have the fourth highest peak in both road distances and building counts.

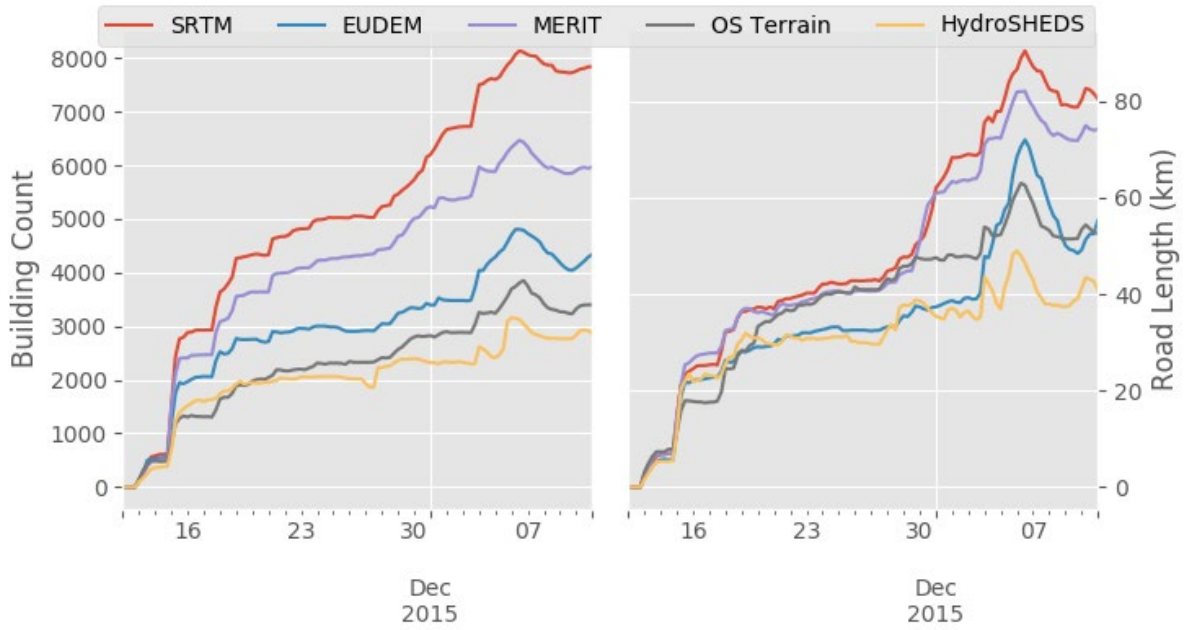

Figure 6-6 - Time series' of inundated road distances and building counts using a range of DEMs for the Tyne basin.

One explanation for the variance in peaks and temporal patterns is the differences in channel geometries between datasets. Deeper channels allow more water to leave the flood plain causing initially lower inundation. However if there are deeper channels, downstream conveyance will be increased leading to greater potential for flooding when banks are exceeded. Looking at buildings, peaks are less defined across all models and there is a larger disparity between maximum values than in the road distance results. However, the ranking of highest to lowest inundation magnitudes from models using each dataset remains consistent between roads and buildings, with SRTM and HydroSHEDS resulting in the most and least inundation respectively.

The magnitude of difference between datasets is much more evident when looking at infrastructure impacts rather than the extent performance measures discussed in Chapter 5. This suggests that while basin-wide CSI may be very similar between two model outputs, they can still produce very different impact estimates. One way to incorporate this into performance assessment could be to calculate CSI only at cells with infrastructure features present. This would provide a more representative assessment of performance in areas which affect impact estimates. Such an approach may become problematic when using a benchmark dataset which does not capture flooding well in urban areas, such as SAR, however future developments in both remote sensing and CCTV image processing may provide solutions here (Section 7.3.3).

6.2.2 Precipitation

Along with sensitivity to DEMs, the effect of precipitation dataset choice on infrastructure inundation was also tested. In a similar approach to Chapter 5, the DEM is held constant here as the HydroSHEDS dataset while the precipitation input is varied between models. The results of this analysis are shown in Figure 6-7. There is a clear reduction in the variability of

inundation compared with the results from models using different DEMs. This suggests that, at least at this resolution, the topography of the land surface has a greater influence on the final estimation of impacts. Therefore, development of better digital elevation models would seem to be more of a priority than increasing the quality of precipitation data, if the accuracy of impact estimates is more important. However, this does not take into account the severity of inundation, including the depth and velocity of water, which is more influenced by the nature of the precipitation input, as seen in Chapter 5.

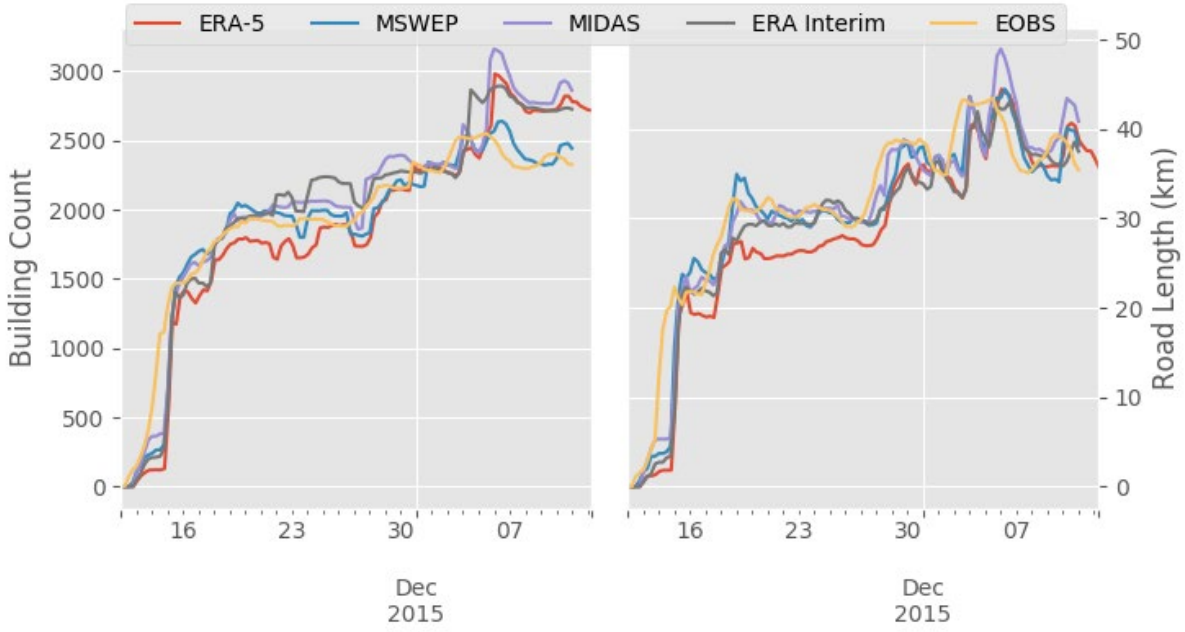

Figure 6-7 - Time series' of inundated road distances and building counts using a range of rainfall datasets for the Tyne basin.

The model using MIDAS data produced a higher peak inundation magnitude across both buildings and roads, and exceeded the other models for almost the entire series. This result is coherent with the higher discharge peaks produced in Chapter 5 by models driven using MIDAS data compared to those using other precipitation datasets. Looking at buildings, the EOBS model produced the lowest inundation peak, closely followed by MSWEP. The model

driven with MIDAS data stands out as the highest peak in the road inundation results with models using other datasets flooding approximately 5 km less distance.

6.3 Case Studies

The modelling framework, outlined in Chapter 3, can be applied anywhere in the world and has been tested in a number of locations. Outside of the UK, however, validation data is more sparsely available and difficult to trust. As the hydrodynamic model being used does not simplify any of the required calculations, there is also a limit to the size of the basin that can be simulated within a reasonable real-world time period. In theory, adding more CPUs to a machine should speed up the simulation linearly, however this does not happen in practice, as the speed of each time step is limited by the speed at which random access memory can be read and written. This limitation could be circumvented by simplifying the governing equations or creating a non-uniform grid, however that was not within the scope of this study. Therefore, the areas which could be tested within a reasonable time period were limited by the speed of the modelling system. Another consideration when modelling different regions of the world was the projection to use outside of Europe. In Jakarta, for example, the model has been applied using the Universal Transverse Mercator (UTM) projection which is available for any location on earth, however each version is constrained to a certain area, meaning that it may not be suitable for modelling basins larger than a single UTM region.

6.3.1 Europe

Results from the European simulations described in Chapter 4 of the Elbe, Loire, Rhone, Seine and Po were used to assess the potential spatiotemporal impacts to infrastructure in these regions of the modelled events. Maximum flood extents are shown for 6 selected cities in Figure 6-8. The effects of the channel burning process within the HydroSHEDS DEM can be clearly observed in these snapshots, particularly in Nantes and Avignon. As a result of this and

other topographic factors, there are some distinct variations in the extent of inundation between cities. Milan, for example, has the widest spread of flooding across the city, whereas in Paris, Nantes and Avignon, the majority of water reaches the main channels.

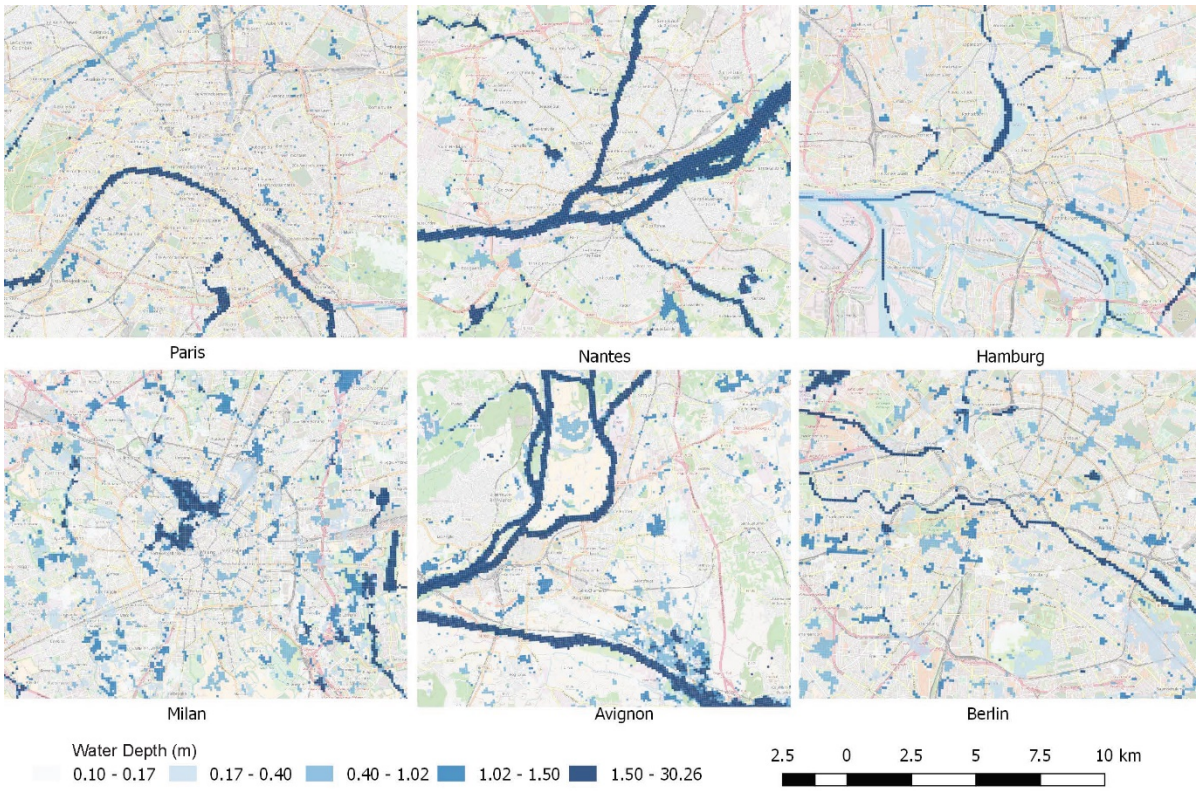

Figure 6-8 - Modelled maximum flood depth (m) maps from events in selected European cities.

To calculate infrastructure inundation in these areas, boundaries were required for the urban regions. The European Commission's Nomenclature of Territorial Units for Statistics (NUTS) system was used for this purpose. The results from the analysis are shown in Figure 6-9, with the corresponding NUTS IDs shown in brackets. The lack of a declining inundation period in most of the areas is a result of the position of the flood peak near to the end of the modelled period which could be rectified by moving the simulation end date later, however due to the run times involved, this was not practical. Avignon and Milan exhibit the largest impacts in terms of road distances, while Berlin experiences the highest number of inundated buildings.

The most variation between the shapes of inundation series between roads and buildings is seen in Avignon, while in other areas, the temporal distribution of inundation between infrastructure types is very similar. This could be a result of fewer roads being located close to channels in Avignon than other areas. However, a lot of this variation could be due to varying accuracy in the OpenStreetMap data and buildings or roads being positioned close enough to channels that they become falsely inundated. The lack of any infrastructural validation data is a major limitation and the reason why only a limited number of areas have been included for demonstration of capability purposes and it is not suggested that these results should be assumed to be accurate. Collection of infrastructure inundation validation data is perhaps an even larger challenge than flood extent and remains an open problem.

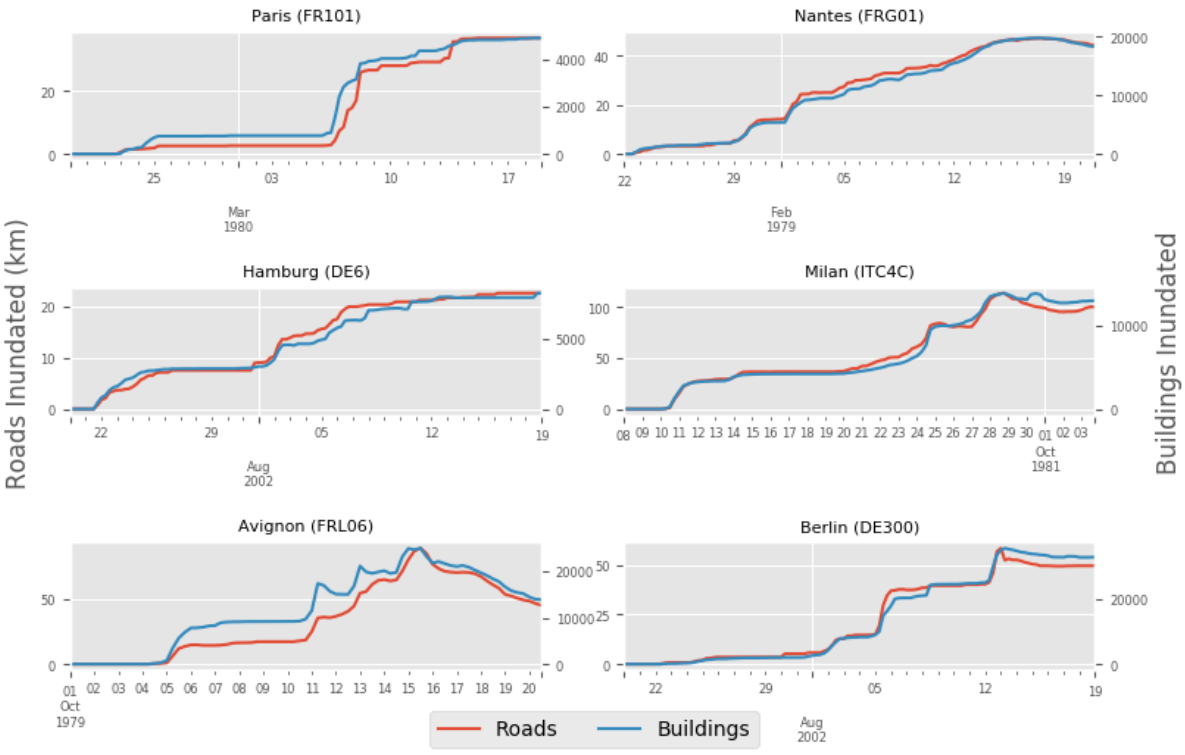

Figure 6-9 - Road and building inundation time series' for selected European cities.

6.3.2 Jakarta

Figure 6-10 illustrates the results from a more distant case study in Indonesia. The area was chosen due to interest from the insurance industry, thanks to the high density of vulnerable

communities and assets at risk of flooding and a lack of good quality modelling in the region. The topography of the area also means the region of interest was split into four basins, meaning simulations were extremely fast to execute. The same workflow as before is utilised, using ERA-Interim rainfall data and the HydroSHEDS DEM to model the widespread flooding that occurred in Jakarta in 2007.

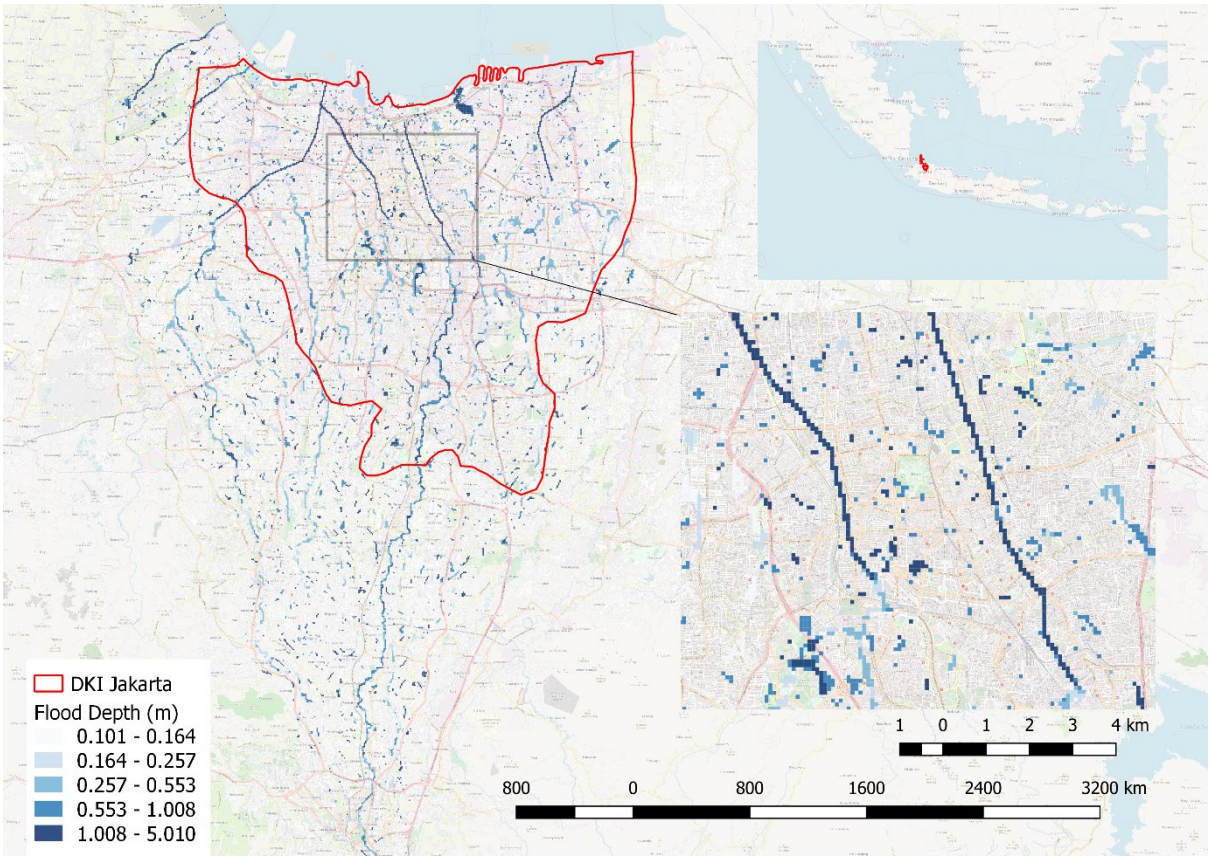

Figure 6-10 - Results from simulation of 2007 floods in Jakarta using ERA Interim rainfall data and HydroSHEDS DEM.

There are a number of additional factors which cause modelling difficulties in this region. Primarily, the large number of raised motorways perpendicular to flow pathways mean they are often interrupted, resulting in slowing or complete blocks in flood wave conveyance. The second major factor is the effect of channel burning within the HydroSHEDS DEM. As all channels are burned to 12m by default, the effect of channels is over-represented in the mode output. This means there is likely to be more flood plain inundation in reality than is

represented in the results. A lack of accessible and usable flow gauge observations in the region meant hazard validation was not possible here. Remote sensing techniques are also hindered in such built up areas as radar signals are interrupted by the dense array of buildings and other development. Therefore hydrodynamic modelling using globally available data provides one of the best solutions to characterize flood risk in the area.

To analyse infrastructure impacts, urban boundaries were again required. This time the limits of the DKI Jakarta province were used and the results of the inundation analysis are shown in Figure 6-11. Flooded area is also shown and corresponds closely to the building and road inundation pattern, indicative of the high density of development in the area. There are a much larger proportion of buildings to roads inundated compared with the European regions tested. Again, the lack of any information to use in validation of infrastructure inundation limits any conclusions which may be drawn from the analysis, however it does provide a demonstration of the global transferability of the processing workflow. If data describing asset inundation over time does become available in the future, this could readily be used to assess the performance of the framework.

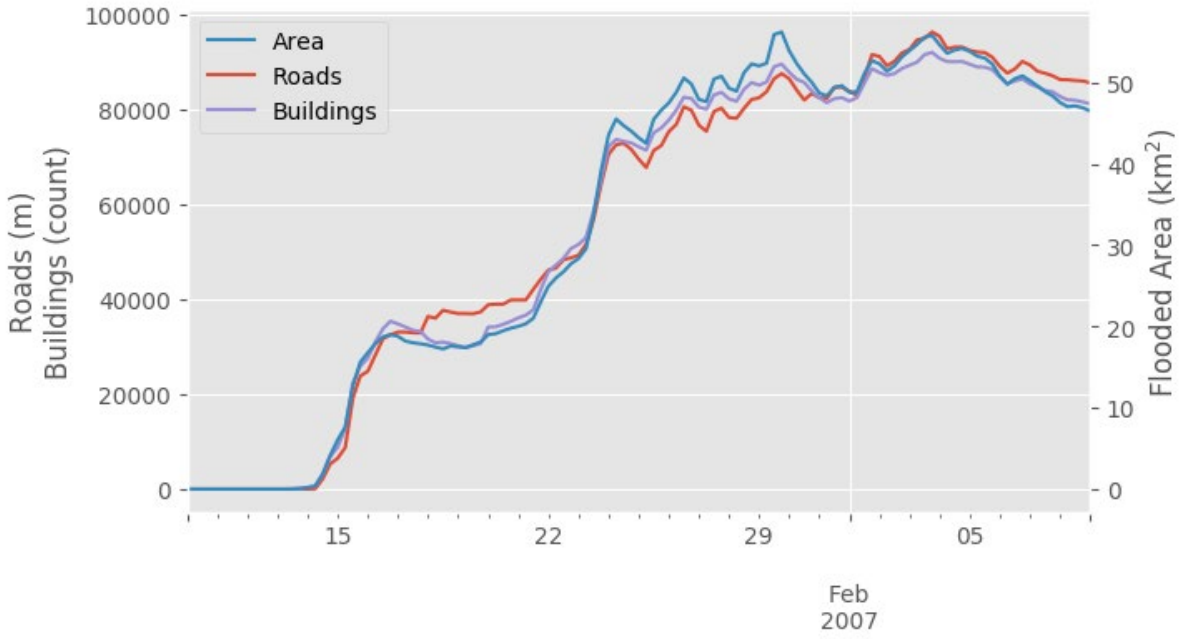

Figure 6-11 - Inundated roads and buildings in DKI Jakarta caused by flooding in 2007.

6.4 Summary

The globally available OpenStreetMap database has been used to calculate the degree of infrastructure inundation over time during Storm Desmond in Northern England. Sensitivity to flood model inputs has also been tested and the choice of DEM dataset was found to be more influential than the precipitation dataset used, with implications for future development of flood risk models. A general lack of infrastructure inundation validation data availability has hindered any analysis of how well the designed workflow is performing at calculating inundation. There is a clear need for this data, however cost of the number of sensors involved and maintenance requirements would be difficult to justify. Satellite observations may provide a solution here, however the Sentinel 1 results discussed in previous chapters often detect roads as being flooded when they are not due to the resemblance of the flat tarmac surface to water. To test the modelling framework in a wider range of basin sizes and topographies, simulations were also completed in a number of European and Indonesian basins. Validation was limited in Europe and not carried out in Indonesia, however the results demonstrate the global applicability of the automated processing workflow.

Chapter 7. Conclusions

This research aimed to better understand the potential and limitations of modelling floods at broad scales using non-localised data and parameters. To accomplish this, the research was divided into four main objectives and these have been met as described below.

A fully integrated and automated framework for deploying 2D hydrodynamic flood models of entire basins on the Microsoft Azure cloud has been developed, which meets the first research objective. To accomplish this, a relational spatial database system has been used to handle storage and retrieval of the required datasets and model outputs. This combination of a local database with on demand cloud computing provides a novel and powerful approach to running many large simulations and can be readily adapted for other modelling applications.

The framework incorporates a quantitative and transparent validation methodology, fulfilling objective two. The combination of extent and flow validation using observed data provides information about both the performance of depth and velocity estimates and spatial scale of floodplain inundation. Events in Europe and Indonesia have been used to demonstrate the global capabilities of the methodology.

The effects of dataset choice and model parameters on performance have been assessed, which addresses the third objective. Global and continental precipitation and elevation datasets have been tested against higher resolution and better quality national products. The effects of parameters including friction, channel burning, and resolution have been analysed.

OpenStreetMap (OSM) data has been used to demonstrate an application of the framework for impacts analysis and meet the fourth and final objective. The approach used enables the spatiotemporal and dynamic nature of flooding to be represented rather than using only a static inundation layer. Sensitivity of road and building inundation to flood model inputs was

analysed and the global availability of OSM utilised to apply the methodology in Jakarta and a range of European locations. The key findings of the research are discussed below, followed by their implications and possible areas for future research.

7.1 Key Findings

Cloud computing has been found to provide a viable and efficient platform for running computationally intensive flood simulations. Microsoft's grant scheme for research enables access to the required resources and their open-source Python API means automation of any processing workflow is possible. Using a spatial database made handling of the wide range and vast quantity of required data highly efficient. Automation and parallelisation of the model setup and analysis process meant more combinations of datasets and parameters could be tested as less time was spent on calculating performance metrics. More detailed validation was possible as the process was automated which meant model outputs could be tested against a range of datasets including observed flow records, remotely sensed imagery and benchmark Environment Agency models to calculate a variety of metrics.

The power of cloud computing enabled a fully hydrodynamic modelling system to be tested at spatial scales which it had never been applied to previously. This was critical in enabling discharge validation as depths and velocities could therefore be calculated over time. This meant the dynamic nature of the flood events could be represented rather than only providing maximum inundation extent. The fact that results are provided at frequent time intervals also opens up new avenues for research, particularly in terms of flood duration. Often, discharge cannot be used in validation as models output only maximum flood extent and it is argued that this severely limits the degree of validation possible. Duration and velocities of floods can be critical for determining the impact they have on an area and therefore there needs to be a way to check the accuracy of these variables. Validation using discharge gauges provides one

way to do this and is a means of assessing the timing accuracy of flood simulations rather than just their maximum extent.

Two categories of metrics were used for validation, relating to point flows and spatial extents. The findings are discussed within the context of how appropriate each metric is for a given application. Although extent might be seen as more relevant most of the time, a lack of depth data within flood extent observations often means the variation in severity of flooding within the extent is not checked. It may be possible to generate depths from extent observations where a high quality DEM is available, by combining these datasets (Cian et al., 2018). However, this still does not verify flow velocities. Extent observations tend to be limited in frequency, therefore timing cannot be assessed properly and there is unlikely to be an extent observation at the time of peak flood inundation. In cases where depths and velocities are important, such as in the design of adaption and mitigation measures, validation using flow gauges is vital to augment the extent-based approaches.

Looking at discharge gauges, better performance was demonstrated in areas with greater observed flows. The average antecedent level of soil moisture in a catchment was shown to affect the model's ability to capture flood peaks during Storm Desmond. Surprisingly, almost all peaks were underestimated. This was not anticipated as the effect of infiltration is not included within the model and therefore one would expect the model to overestimate flood peaks. The underestimation may be a result of either rainfall underrepresentation or numerical dispersion within the models. The most important and novel findings of the research are related to dataset choice for modelling in data-scarce regions as flood peaks and extents have not been comprehensively examined and compared to such a degree by previous studies. The variability in discharge peak accuracy between models driven by different precipitation inputs was more evident than between those using different DEMs. A

particularly significant reduction in performance was evident in models driven by global-continental gridded rainfall when compared to models using UK Met Office rainfall. There was promising improvement in model performance obtained by replacing ERA Interim with its updated version, ERA5. Models driven using the MSWEP dataset were shown to create erroneous early flood peaks. Other factors such as DEM resolution and friction were also shown to affect both the timing and magnitude of flood peaks.

The difficulties with trying to validate flood extents have been discussed and demonstrated. The main limitation of extent validation is the lack of reliable ground-based observations to validate satellite imagery. This means it becomes impossible to determine which source of flood extent is more accurate. Improved CSI scores were produced when CityCAT results were compared with the results of high resolution EA models rather than real observations from Sentinel 1. The range of extent performance was demonstrated to be marginally larger between models using different DEMs than different rainfall datasets, however overall, there was minimal difference between results. A lack of confidence in the reliability of either benchmark dataset hinders any ability to draw strong conclusions from these results. The significantly lower temporal resolution of the Sentinel observations compared with CityCAT outputs was identified as a key limitation. The disparity in spatial resolution was also a problem and significant improvements in CSI were seen with increased resolution of OS Terrain and EUDEM, using Sentinel observations as a benchmark.

To illustrate the importance of validation, the sensitivity of infrastructure impacts to flood model input dataset choices was assessed. Outputs from models using different DEMs were found to produce more varied road and building inundation results than models using different rainfall datasets. This indicates that the choice of DEM input has more effect than the choice of rainfall input on estimates of infrastructure inundation. This highlights the

importance of DEM quality when using flood model results for impacts analysis. While rainfall remains important for depths and velocities, it has been found to be less important for estimating the maximum extent of flooding. However, as no data describing the observed magnitude of infrastructure inundation during the modelled events was readily available, the inundation results have not been validated and therefore their credibility is limited.

7.2 Context and Implications

7.2.1 Research

This work is novel and builds on the existing literature concerning flood modelling using broad-scale datasets. Rather than using return periods, the focus has been on characterising historical events to enable real observations to be used in validation. Some previous studies included simulations of past events but did not combine the use of discharge gauges and satellite observations in such a comprehensive way. A direct comparison has also not commonly been drawn in the literature between using remotely sensed extent and localised benchmark flood models as a validation source. Dottori et al (2016), for example, validate against local models in the Po, Severn, Thames and Elbe whereas they use satellite imagery in the Niger, Ganges, Mekong, Irrawaddy and Tocantins basins, preventing a direct comparison. The finding from this thesis that performance appears better in the model-model comparison than when using observations from satellites has far-reaching implications. It reinforces the message that while inter-model comparisons are useful for benchmarking purposes, they should always be augmented with real observations either from the ground, air or space. Specifically, validation using observations over time and not just of maximum extent can provide a more holistic picture of model performance. The only way that progress can be in made in model performance is if historical events are simulated and results from these simulations assessed against independent observations.

For data-scarce regions, the methodology developed provides a parsimonious and effective way to produce precise estimates of flood depths and velocities across the domain and duration of an event and not just for the maximum extent. With the only requirements being elevation and rainfall, this is one of the least data-demanding setups of any existing study but is still able to produce reasonable results. The approach used differs from studies such as Sampson et al (2015), who regionalise extreme discharge using climate class, upstream annual rainfall, and catchment area. This means that rather than being dependent on often limited flow data, only rainfall is required to generate extreme discharge directly on the model grid. Of course, the results are therefore sensitive to the accuracy of rainfall inputs, however it is anticipated that in developing regions, gridded precipitation products and not discharge gauges will provide the more reliable source of data as discharge records are often not automated or quality controlled. The direct rain-on-grid approach to the entire domain also means flooding is not limited to the areas surrounding channels and pluvial flooding can be characterised.

The comprehensive sensitivity analysis to DEMs and rainfall datasets shows that broad-scale flood models should never be run using a single source of data. Instead, ensembles of results should be created using the range of inputs available, in much the same way that climate model results are currently delivered. This would ensure that uncertainty from dataset choice is included in the final outputs. If multiple sets of results are not acceptable for an application then the best performing datasets from the ensemble could be used. An over-arching theme of the work has been the difficulties in attempting to validate flood extents. This problem is pervasive in the literature and it is suggested that more open and comprehensive approaches to validation using all data sources available would lead to improved understanding and characterisation of how challenging modelling with limited data can be.

The integration with a PostgreSQL database and JavaScript frontend using WMS layers from GeoServer is a novel use of existing technology to swiftly visualise and interpret flood model results from anywhere in the world. Using a spatial database simplified the process of running parallel simulations with a range of datasets and parameters on the Microsoft Azure cloud and is a novel approach to automated sensitivity analysis for flood modelling. The use of database technology is pervasive in industry and is beginning to be adopted by researchers, however there are many barriers to this including the required technical skills and limited access to suitable hardware with the required level of permissions for hosting web applications and services. Publicly available cloud computing services, such as Microsoft Azure as used here, go some way to solving the latter problem. This research has demonstrated the usability and benefits of a database back-end for environmental modelling and this example provides motivation for other studies to adopt a similar approach. Another advantage to using a hosted database is that both inputs and results can easily be shared between researchers working in different institutions.

7.2.2 Insurance Industry

The main implication for the insurance industry is to ensure that when collecting ensembles of risk reports, there is not a reliance on a single source of data between studies using different modelling systems. The concept of using multiple hazard maps is useful as it builds a better picture of the agreement between modellers, however, if the underlying data is similar across all results, the usefulness of the approach is undermined. Of course validation is also important and flood model outputs should be validated using multiple data sources concerning both extent and discharge. A new standard of validation is required which can assign scores to each study to allow rapid comparison of reliability. A consistent approach needs to be applied to validate all model outputs, using the same set of observations. As velocity can be particularly important for flood damages, validation using discharge is

particularly important and is still not common among broad-scale studies. Using an approach which characterises flood hazard over time, and not just at the maximum extent, allows the duration of the event to be characterised, which is critical for assessing damages.

The approach used in this thesis characterises pluvial flood risk at the same time as fluvial, which can be important depending on the region of interest. The rainfall-runoff method does not depend on the availability of flow gauge records in the way that regionalisation approaches do and rainfall can usually be measured, interpolated or simulated more easily than discharge. The presence of flood defences has not been addressed but presents a major obstacle to both validation and impacts analysis in data scarce areas. Particularly if the density and distribution of defences changes over time. Sensitivity analyses are also critical to understand if a model is highly affected by a certain parameter or dataset choice. A higher resolution and more accurate global DEM has already been called for by researchers (Schumann & Bates, 2018; Sampson et al., 2016) and this thesis reinforces that message. The returns for a wide range of fields and applications would far offset the costs but the question remains as to who should pay for it.

7.2.3 Policy Makers

Planning for flood risk is increasingly difficult with the demand for development growing at an alarming rate. Climate change is also increasingly in the forefront of political discourse. But planning for future flood risk is an even more uncertain research area. As highlighted by this thesis, uncertainty is a major factor in any broad-scale flood risk analysis where data is limited and forecasting into the future can only exacerbate this. Though various modelling studies have predicted an increase in the frequency and magnitude of floods into the future, it is difficult to find definitive evidence for this without longer records than we currently have. Therefore, adaptability is critical to counter the evolving flood risk under uncertainty. Efforts

are already underway to accelerate work in this area, such as the Global Commission on Adaption (WRI, 2019). A particular focus of the commission is ecosystem-based adaption. This can involve measures such as green infrastructure which provide a degree of flood protection while also improving the well-being of people living in cities. Measures such as increased greenspace are sustainable and account for uncertainty as their financial case is not based only on the premise of increasing flood risk but also the 'liveability' of cities.

7.3 Future Work

There are many possible areas for future research, grouped here into data, model development and validation. These are outlined below.

7.3.1 Input Data

A central theme of the thesis has been the quality and availability of data for modelling and impact analysis. Data is required about both extreme precipitation events and the land surfaces that these events affect. Collection of higher resolution, longer duration, more accurate and readily available data would significantly improve the usefulness of broad-scale models. However, collecting this data can be expensive and time consuming or simply not possible given the resources available. The main priority is an updated globally available digital elevation model, which has already been widely requested by the scientific community. This would either require intergovernmental cooperation though projects such as the World Water Data Initiative or a charitable private company or group of companies with humanitarian objectives. There are currently a number of high resolution products on the market, however they are prohibitively expensive and also as a result, lacking in quality control. The organisations marketing these products would clearly be opposed to a better openly available version, however this is what is required to drive the field of broad-scale flood modelling forward. The extensibility of the framework developed in this thesis means that new sources

of data such as a better DEM or new rainfall products, could be easily incorporated and benchmarked against the previously existing datasets.

The UN's World Water Data Initiative is one of the leaders in the area of data collection, specifically for building resilience to water related hazards including flooding. A recommendation for the initiative in terms of broad-scale flood modelling is to focus on improving DEM and rain gauge data availability and quality. Citizen science projects, such as the Trans-African HydroMeteorological Observatory (TAHMO) and the UK Met Office's Weather Observations Website (WOW), provide promising sustainable approaches to more distributed and community supported rainfall data collection in the future.

7.3.2 Model Development

The modelling system used here is based on CPU architecture and although the code has been designed for distributed processing across multiple cores, there is still a limitation on reading from memory. Speed-up of CityCAT may be possible by moving to Graphical Processing Units (GPUs) which has not been possible until recently due to memory limitations meaning that the domain had to be divided. By using GPUs, the speed of calculations may increase, however optimisation of the code for this purpose is not trivial and research would be required to find out if simulation run times would actually be reduced.

Infiltration or the effect of land use on friction has also not been accounted for here in order to simplify the sensitivity analysis and recreate a scenario where only very limited data is available. However, both infiltration and variable friction across the floodplain could be incorporated in future work. Although a single modelling system has been used here, other hydrodynamic models, such as LISFLOOD or CaMa-Flood could be incorporated to allow for an assessment of the effect of changing the hydrodynamic engine and review the trade-off between performance and accuracy.

Additional processing stages such as infrastructure network impact analysis can readily be incorporated into the distributed processing workflow to enable automation of more complex analysis. This has not been carried out here due to the prohibitively coarse scale of the hazard maps for detailed network analysis. The inundation analysis was limited to only include flood depth but a velocity threshold could also be implemented to allow more detailed damage calculations. As flood defences and natural flood management measures such as reservoirs and green infrastructure have been ignored here, incorporating their influence is another area for future research. Though, DEM resolution will need to be improved before the scale of these defences will be commensurate with the model.

7.3.3 Validation

A key aspect of this thesis has been the ability to validate results using observed flood flows. In the UK, the temporal resolution and quality control of flow data is generally good. However, in other areas around the world, this is not the case and records are either difficult to access, incomplete or low quality. One solution to this could be using remotely sensed observations of river widths to derive flows, which is a developing area of research (Gleason & Smith, 2014). This would not solve the temporal resolution problem, however may provide a useful peak flow estimate for large channels during very extreme events. More open sharing of any privately held high resolution gauge data would also be highly beneficial to future studies.

Other sources of benchmark extent data can be included as new satellites are launched and new processing methods developed. A range of new commercial remote sensing products for flood monitoring such as ICEYE (2019) are becoming increasingly available and could be incorporated if their results became open for research purposes. However, the capability of satellites is very limited in urban areas where their sensors are often blocked from reaching ground level. Ultimately CCTV coupled with computer vision algorithms could provide the best

means of validating building inundation, however even then, building specific defences may provide protection from floodwater which would require even more detailed local knowledge.

A lack of validation data describing flood impacts to buildings and infrastructure has been a key limitation of the impacts analysis in this thesis. This data would provide a key missing source of accuracy information. Insurance claims may provide one avenue for this which has already been tested in a number of studies (Zischg et al., 2017), however exact timings and depths are often not included in the claims and the data itself can be difficult or impossible to access, so this is not a panacea.

7.4 Summary

This thesis has found a high level of variability in flood model outputs when using different sources of broad-scale input data and pre-processing methods, in terms of both extents and peak flows. The importance of using observed discharge in combination with extents for validation has been demonstrated. Data is critical to improve flood hazard modelling. Specifically DEMs and flood event observations must be improved in quality and availability. The reproducibility of modelling workflows needs to be improved and validation methods more standardised to enable clearer interpretation by researchers, the insurance industry and policy makers. Underlying data and sensitivities must be questioned and validation based on historical events included in any broad-scale flood risk assessment. The potential of cloud computing to enable these recommendations has been established. As more data becomes available and computing power improves, while costs reduce, this potential is only going to increase, providing a useful tool to facilitate a better understanding of broad-scale flood risk and uncertainty.

References

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., Roo, A. De, Salamon, P., Wyser, K., Feyen, L., Meteorological, S. & Centre, R. (2017) Global projections of river flood risk in a warmer world. *Earth's Future*. 5171–182.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J. & Pappenberger, F. (2013) GloFAS–global ensemble streamflow forecasting and flood early warning. *Hydrol. Earth Syst. Sci*. 17 (3), 1161–1175.

Altenau, E.H., Pavelsky, T.M., Bates, P.D. & Neal, J.C. (2017) The effects of spatial resolution and dimensionality on modeling regional-scale hydraulics in a multichannel river. *Water Resources Research*. 53 (2), 1683–1701.

Ambiential (2019) *Global FloodMap*^{TM}. [Online] [online]. Available from: https://www.ambientalrisk.com/global-floodmap/ (Accessed 31 May 2019).

Andreadis, K.M., Schumann, G.J.-P., Stampoulis, D., Bates, P.D., Brakenridge, G.R. & Kettner, A.J. (2017) Can atmospheric reanalysis datasets be used to reproduce flooding over large scales? *Geophysical Research Letters*.

Andreadis, K.M., Schumann, G.J.P. & Pavelsky, T. (2013) A simple global river bankfull width and depth database. *Water Resources Research*. 49 (10), 7164–7168.

Archer, L., Neal, J.C., Bates, P.D. & House, J.I. (2018) Comparing TanDEM-X Data With Frequently Used DEMs for Flood Inundation Modeling. *Water Resources Research*. 54 (12), 10,205-210,222.

Arnell, N.W. & Gosling, S.N. (2014) The impacts of climate change on river flood risk at the global scale. *Climatic Change*. 1–15.

Aziz, K., Rahman, A., Fang, G. & Shrestha, S. (2014) Application of artificial neural networks in regional flood frequency analysis: A case study for Australia. *Stochastic Environmental Research and Risk Assessment*. 28 (3), 541–554.

Di Baldassarre, G. & Uhlenbrook, S. (2012) Is the current flood of data enough? A treatise on research needs for the improvement of flood modelling. *Hydrological Processes*. 26 (1), 153–158.

Bashfield, A. & Keim, A. (2011) 'Continent-wide DEM Creation for the European Union', in *34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring*. [Online]. 2011 pp. 10–15.

Bates, P.D., Horritt, M.S. & Fewtrell, T.J. (2010) A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*. 387 (1–2), 33–45.

Baugh, C.A., Bates, P.D., Schumann, G. & Trigg, M.A. (2013) SRTM vegetation removal and hydrodynamic modeling accuracy. *Water Resources Research*. 49 (9), .

BBC (2017) *Storm Desmond: 125 homes still empty two years on*. [Online] [online]. Available from: https://www.bbc.co.uk/news/uk-england-cumbria-42236304 (Accessed 8 May 2019).

BBC (2015) *Storm Desmond: Hundreds unable to go home one year on*. [Online] [online]. Available from: https://www.bbc.co.uk/news/uk-england-cumbria-38194698 (Accessed 8 May 2019).

Beck, H. (2017) *MSWEP Version 2.1 documentation*. 1–18.

Beck, H.E., Van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. & De Roo, A. (2017) MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*. 21 (1), 589–615.

Beck, H.E., Dijk, A.I.J.M. van, Roo, A. de, Miralles, D.G., McVicar, T.R., Schellekens, J. & Bruijnzeel, L.A. (2016) Global-scale regionalization of hydrologic model parameters. *Water Resources Research*. 523599–3622.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J. & Wood, E.F. (2017) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*.

Bernardini, G., Postacchini, M., Quagliarini, E., Brocchini, M., Cianca, C. & D'Orazio, M. (2017) A preliminary combined simulation tool for the risk assessment of pedestrians' flood-induced evacuation. *Environmental Modelling and Software*. 9614–29.

Box, G.E.P. (1979) Robustness in the strategy of scientific model building. *Robustness in statistics*. 1201–236.

Bradford, S.F. & Sanders, B.F. (2002) Finite-Volume Model for Shallow-Water Flooding of Arbitrary Topography. *Journal of Hydraulic Engineering*. 128 (3), 289–298.

Brakenridge, R. & Anderson, E. (2006) 'MODIS-Based Flood Detection, Mapping and Measurement: the Potential for Operational Hydrological Applications', in Jiri Marsalek, Gheorghe Stancalie, & Gabor Balint (eds.) *Transboundary Floods: Reducing Risks Through Flood Management*. [Online]. Dordrecht: Springer Netherlands. pp. 1–12.

Brovelli, M.A., Minghini, M., Molinari, M. & Mooney, P. (2017) Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. *Transactions in GIS*. 21 (2), 191–206.

Budiyono, Y., Aerts, J., Brinkman, J.J., Marfai, M.A. & Ward, P. (2015) Flood risk assessment for delta mega-cities: a case study of Jakarta. *Natural Hazards*. 75 (1), 389–413.

Burek, P., van der Knijff, J. & de Roo, A. (2013) LISFLOOD DistributedWater Balance and Flood Simulation Model e Revised User Manual 2013,. JRC Technical Reports. Joint Research Centre of the European Commission. *Publications Office of the European Union, Luxembourg.*

Burgess, J., Pavlenko, A., Oosterhout, M. van, Hoffmann, S., Krueger, K., Ramm, F., Quinion, B., Amos, M., Kreiser, K. & Norman, P. (2018) *osm2pgsql*.

Castellarin, A., Domeneghetti, A. & Brath, A. (2011) Identifying robust large-scale flood risk mitigation strategies: A quasi-2D hydraulic model as a tool for the Po river. *Physics and Chemistry of the Earth*. 36 (7–8), 299–308.

Ceola, S., Laio, F. & Montanari, A. (2014) Satellite nighttime lights reveal increasing human exposure to floods worldwide. *Geophysical Research Letters*. 41 (20), 7184–7190.

Chow, V. Te (1959) *Open-channel hydraulics*. Vol. 1. McGraw-Hill New York.

Cian, F., Marconcini, M., Ceccato, P. & Giupponi, C. (2018) *Flood depth estimation by means of high-resolution SAR images and lidar data*. 3063–3084.

Cipeluch, B., Jacob, R., Winstanley, A., Mooney, P., Ciepluch, B., Jacob, R., Winstanley, A. &

Mooney, P. (2010) Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences*. 337.

Clement, M.A., Kilsby, C.G. & Moore, P. (2018) Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*. 11152–168.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G. & Bauer, P. (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. 137 (656), 553–597.

Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., Koutsoyiannis, D. & Efstratiadis, A. (2016) Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. *Journal of Hydrology*. 534478–492.

DLR (2018) *Global 3D elevation model from the TanDEM-X mission now freely available*. [Online] [online]. Available from: https://www.dlr.de/dlr/en/desktopdefault.aspx/tabid-10081/151_read-30139/#/gallery/32238 (Accessed 21 May 2019).

DLR (2019) *Overview of the TanDEM-X DEM Product Specification*. [Online] [online]. Available from: https://geoservice.dlr.de/web/dataguide/tdm90/ (Accessed 21 May 2019).

Domeneghetti, A. (2016) On the use of SRTM and altimetry data for flood modeling in data-sparse regions. *Water Resources Research*. 522901–2918.

Donaldson, R.J., Dyer, R.M. & Kraus, M.J. (1975) 'An objective evaluation of techniques for predicting severe weather events', in *9th Conference on Severe Local Storms, Norman, American Meteorological Society*. [Online]. 1975 pp. 321–326.

Donat, M.G., Sillmann, J., Wild, S., Alexander, L. V, Lippmann, T. & Zwiers, F.W. (2014) Consistency of Temperature and Precipitation Extremes across Various Global Gridded In Situ and Reanalysis Datasets. *Journal of Climate*. 27 (13), 5019–5035.

Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L. & Feyen, L. (2017) An operational procedure for rapid flood risk assessment in Europe. *Hazards Earth Syst. Sci*. 1751941111–1126.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A. & Feyen, L. (2016) Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*. 9487–102.

Eckle, M., de Albuquerque, J.P., Herfort, B., Leiner, R., Jacobs, C. & Zipf, A. (2016) Leveraging OpenStreetMap to support flood risk management in municipalities : A prototype decision support system. *Proceedings of the ISCRAM 2016 Conference – Rio de Janeiro, Brazil, May 2016*. (May), .

ECMWF (2019) *What are the changes from ERA-Interim to ERA5?* [Online] [online]. Available from: https://confluence.ecmwf.int//pages/viewpage.action?pageId=74764925 (Accessed 8 May 2019).

Elshorbagy, A., Bharath, R., Lakhanpal, A., Ceola, S., Montanari, A. & Lindenschmidt, K.-E. (2017) Topography-and nightlight-based national flood risk assessment in Canada. *Hydrol. Earth Syst. Sci*. 212219–2232.

Environment Agency (2019) *Flood Map for Planning (Rivers and Sea) - Flood Zone 2*. [Online]

[online]. Available from: https://data.gov.uk/dataset/cf494c44-05cd-4060-a029-35937970c9c6/flood-map-for-planning-rivers-and-sea-flood-zone-2 (Accessed 23 May 2019).

European Commision Parliament (2013) Commission Delegated Regulation (EU) No 1159/2013 of 12 July 2013. *Official Journal of the European Union*. 3091–6.

Falorni, G., Teles, V., Vivoni, E.R., Bras, R.L. & Amaratunga, K.S. (2005) Analysis and characterization of the vertical accuracy of digital elevation models from the Shuttle Radar Topography Mission. *Journal of Geophysical Research: Earth Surface*. 110 (F2), n/a-n/a.

Falter, D., Dung, N. V., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F. & Merz, B. (2016) Continuous, large-scale simulation model for flood risk assessments: Proof-of-concept. *Journal of Flood Risk Management*.

Falter, D., Vorogushyn, S., Lhomme, J., Apel, H., Gouldby, B. & Merz, B. (2013) Hydraulic model evaluation for large-scale flood risk assessments. *Hydrological Processes*. 27 (9), 1331–1340.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D. (2007) The Shuttle Radar Topography Mission. *Reviews of Geophysics*. 45 (2), .

Fewtrell, T.J., Bates, P.D., Horritt, M. & Hunter, N.M. (2008) Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes*. 22 (26), 5107–5118.

Fewtrell, T.J., Duncan, A., Sampson, C.C., Neal, J.C. & Bates, P.D. (2011) Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data. *Physics and Chemistry of the Earth*. 36 (7–8), 281–291.

Feyen, L., Dankers, R., Bodis, K., Salamon, P. & Barredo, J. (2012) Flood hazard in Europe in an ensemble of regional climate scenarios. *Climate Change*. 112 (1), 47–62.

Freni, G., Loggia, G. La & Notaro, V. (2010) *Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation*. (2005), 2979–2993.

GDAL (2018) *The gdalwarp utility*. [Online] [online]. Available from: https://trac.osgeo.org/gdal/wiki/UserDocs/GdalWarp (Accessed 9 May 2019).

Geofabrik (2018) *OpenStreetMap Data Extracts*. [Online] [online]. Available from: http://download.geofabrik.de/ (Accessed 9 May 2019).

Ghimire, B., Chen, A.S., Guidolin, M., Keedwell, E.C., Djordjević, S. & Savić, D.A. (2013) Formulation of a fast 2D urban pluvial flood model using a cellular automata approach. *Journal of Hydroinformatics*. 15 (3), 676–686.

Girres, J.F. & Touya, G. (2010) Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*. 14 (4), 435–459.

Gleason, C.J. & Smith, L.C. (2014) *Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry*. 1–4.

Glenis, V., Kutija, V. & Kilsby, C.G. (2018) A fully hydrodynamic urban flood modelling system representing buildings, green space and interventions. *Environmental Modelling and Software*. 109 (August), 272–292.

Google (2019) *AHN Netherlands 0.5m DEM, Raw Samples*. [Online] [online]. Available from: https://developers.google.com/earth-engine/datasets/catalog/AHN_AHN2_05M_RUW (Accessed 30 May 2019).

Gupta, H. V., Kling, H., Yilmaz, K.K. & Martinez, G.F. (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*. 377 (1–2), 80–91.

Haklay, M. (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment and Planning B: Planning and Design*. 37 (4), 682–703.

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

Haraguchi, M. & Lall, U. (2014) Flood risks and impacts: A case study of Thailand's floods in 2011 and research questions for supply chain decision making. *International Journal of Disaster Risk Reduction*.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. & New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research Atmospheres*. 113 (20), .

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H. & Kanae, S. (2013) Global flood risk under climate change. *Nature Clim. Change*. 3 (9), 816–821.

Hoch, J.M., Haag, A. V., van Dam, A., Winsemius, H.C., van Beek, L.P.H. & Bierkens, M.F.P. (2016) Assessing the impact of hydrodynamics on large-scale flood wave propagation - a case study for the Amazon Basin. *Hydrology and Earth System Sciences Discussions*. 1–25.

Hofstra, N., Haylock, M., New, M. & Jones, P.D. (2009) Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research Atmospheres*.

Horritt, M.S., Baldassarre, G. Di, Bates, P.D. & Brath, A. (2007) Comparing the performance of a 2-D finite element and a 2-D finite volume model of floodplain inundation using airborne SAR imagery. *Hydrological Processes*. 212745–2759.

Horritt, M.S. & Bates, P.D. (2001) Effects of spatial resolution on a raster based model of flood flow. *Journal of Hydrology*. 253 (1–4), 239–249.

Horritt, M.S. & Bates, P.D. (2002) Evaluation of 1D and 2D numerical models for predicting river flood inundation. *Journal of Hydrology*. 268 (1–4), 87–99.

Hunter, N.M., Bates, P.D., Horritt, M.S. & Wilson, M.D. (2007) Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology*. 90 (3–4), 208–225.

Hunter, N.M., Bates, P.D., Neelz, S., Pender, G., Villanueva, I., Wright, N.G., Liang, D., Falconer, R.A., Lin, B., Waller, S., Crossley, A.J. & Mason, D.C. (2008) Benchmarking 2D hydraulic models for urban flooding. *Proceedings of the Institution of Civil Engineers - Water Management*. 161 (1), 13–30.

Hunter, N.M., Horritt, M.S., Bates, P.D., Wilson, M.D. & Werner, M.G.F. (2005) An adaptive time step solution for raster-based storage cell modelling of floodplain inundation. *Advances in Water Resources*. 28 (9), 975–991.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C. & Arheimer, B. (2016) Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*. 527548–7555.

ICEYE (2019) *Flood Mapping*. [Online] [online]. Available from: https://www.iceye.com/applications/land/flood-mapping (Accessed 9 May 2019).

Insurance Information Institute (2018) *Facts + Statistics: Flood insurance*. [Online] [online]. Available from: https://www.iii.org/fact-statistic/facts-statistics-flood-insurance (Accessed 20 May 2019).

Jabareen, Y. (2013) Planning the resilient city: Concepts and strategies for coping with climate change and environmental risk. *Cities*. 31220–229.

Jarihani, A.A., Callow, J.N., McVicar, T.R., Van Niel, T.G. & Larsen, J.R. (2015) Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology*. 524489–506.

JAXA (2019) *ALOS Global Digital Surface Model 'ALOS World 3D - 30m (AW3D30)'*. [Online] [online]. Available from: https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm (Accessed 21 May 2019).

JBA (2019) *Global Flood Maps*. [Online] [online]. Available from: https://www.jbarisk.com/flood-services/maps-and-analytics/global-flood-maps/ (Accessed 31 May 2019).

Jongman, B., Koks, E.E., Husby, T.G. & Ward, P.J. (2014) Increasing flood exposure in the Netherlands: Implications for risk financing. *Natural Hazards and Earth System Sciences*. 14 (5), 1245–1255.

Jongman, B., Kreibich, H., Apel, H., Barredo, J.I., Bates, P.D., Feyen, L., Gericke, A. & Neal, J. (2012) *Comparative flood damage model assessment : towards a European approach*. 3733–3752.

Jongman, B., Ward, P.J. & Aerts, J.C.J.H. (2012) Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*. 22 (4), 823–835.

JPL (2019) *NASADEM: Creating a New NASA Digital Elevation Model and Associated Products*. [Online] [online]. Available from: https://earthdata.nasa.gov/earth-science-data-systems-program/competitive-programs/measures/nasadem (Accessed 30 May 2019).

JPL (1999) *Raised canister w/ main antenna in background at KSC*. [Online] [online]. Available from: https://www2.jpl.nasa.gov/srtm/images/photos/srtm_27_hi.jpg (Accessed 8 May 2019).

Klonner, C., Marx, S., Usón, T. & Höfle, B. (2016) Risk Awareness Maps of Urban Flooding via OSM Field Papers - Case Study Santiago de Chile. *Iscram 2016*. (May 2016), .

Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B. & Thieken, A.H. (2009) Is flow velocity a significant parameter in flood damage modelling ? *Natural Hazards and Earth System Sciences Discussions*. 91679–1692.

Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C.L., Robinson, J. & Kilsby, C. (2014) Model Validation Using Crowd-Sourced Data From A Large Pluvial Flood. 11th International Conference on Hydroinformatics

Leandro, J., Chen, A.S., Djordjević, S. & Savić, D.A. (2009) Comparison of 1D/1D and 1D/2D

Coupled (Sewer/Surface) Hydraulic Models for Urban Flood Simulation. *Journal of Hydraulic Engineering*. 135 (6), 495–504.

Leandro, J., Djordjević, S., Chen, A.S., Savić, D.A. & Stanić, M. (2011) Calibration of a 1D/1D urban flood model using 1D/2D model results in the absence of field data. *Water Science and Technology*. 64 (5), 1016–1024.

Lehner, B. & Döll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*. 296 (1–4), 1–22.

Lehner, B., Döll, P., Alcamo, J., Henrichs, T. & Kaspar, F. (2006) Estimating the impact of global change on flood and drought risks in Europe: A continental, integrated analysis. *Climatic Change*. 75 (3), 273–299.

Lehner, B. & Grill, G. (2014) *HydroBASINS*. 27 (2013).

Lehner, B., Verdin, K. & Jarvis, A. (2013) HydroSHEDS Technical Documentation Version 1.2. *EOS Transactions*. 89 (10), 26.

Lhomme, J., Bouvier, C., Mignot, E. & Paquier, A. (2006) One-dimensional GIS-based model compared with a two-dimensional model in urban floods simulation. *Water Science and Technology*. 54 (6–7), 83–91.

Lin, B., Wicks, J.M., Falconer, R.A. & Adams, K. (2006) Integrating 1D and 2D hydrodynamic models for flood simulation. *Water Management*. 159 (1), 19–25.

Longueville, B. De, Luraschi, G., Smits, P., Peedell, S., Groeve, T. De & Commission, E. (2010) Citizens As Sensors for Natural Hazards. *Geomatica*. 64 (1), 41–59.

Mandlburger, G. (2019) Recent Developments in Airborne Lidar. *GIM International*. 19 April.

Mateo, C.M., Hanasaki, N., Komori, D., Tanaka, K., Kiguchi, M., Champathong, A., Sukhapunnaphan, T., Yamazaki, D. & Oki, T. (2014) Assessing the impacts of reservoir operation to floodplain inundation by combining hydrological, reservoir management, and hydrodynamic models. *Water Resources Research*. 50 (9), 7245–7266.

Mateo, C.M.R., Yamazaki, D., Kim, H., Champathong, A., Vaze, J. & Oki, T. (2017) Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods. *Hydrology and Earth System Sciences Discussions*. 1–32.

McDougall, K. (2011) Using volunteered information to map the Queensland floods. *Proceedings of the 2011 Surveying and Spatial Sciences Conference*. (November), 13–23.

Merz, B., Kreibich, H. & Lall, U. (2013) *Multi-variate flood damage assessment : a tree-based data-mining approach*. 53–64.

Merz, B., Kreibich, H., Thieken, A., Schmidtke, R. & Hydrology, S.E. (2004) *Estimation uncertainty of direct monetary flood damage to buildings*. 153–163.

Merz, R. & Blöschl, G. (2005) Flood frequency regionalisation - Spatial proximity vs. catchment attributes. *Journal of Hydrology*. 302 (1–4), 283–306.

Messner, F., Penning-rowsell, E., Green, C., Tunstall, S., Veen, A. Van Der, Tapsell, S., Wilson, T., Krywkow, J., Logtmeijer, C., Fernández-bilbao, A., Geurts, P. & Haase, D. (2007) *Evaluating flood damages : guidance and recommendations on principles and methods principles and methods*.

Metz, M., Mitasova, H. & Harmon, R.S. (2011) Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrology and Earth*

System Sciences. 15 (2), 667–678.

Michel-Kerjan, E. & Kunreuther, H. (2011) Redesigning Flood Insurance. *Science*. 333 (6041), 408–409.

Moel, H. De, Huizinga, J. & Szewczyk, W. (2016) 'Flood damage curves for consistent global risk assessments', in *EGU General Assembly 2016*. [Online]. 2016 p. 14699.

de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E. & Ward, P.J. (2015) Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change*.

Mokrech, M., Kebede, A.S., Nicholls, R.J., Wimmer, F. & Feyen, L. (2014) An integrated approach for assessing flood impacts due to future climate and socio-economic conditions and the scope of adaptation in Europe. *Climatic Change*. 128 (3–4), 245–260.

Muis, S., Guneralp, B., Jongman, B., Aerts, J.C. & Ward, P.J. (2015) Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. *Sci Total Environ*. 538445–457.

Nash, E. & Sutcliffe, V. (1970) River flow forecasting through conceptual models: Part 1 - a discussion of priciples. *Journal of Hydrology*. 10282–290.

Neal, J.C., Bates, P.D., Fewtrell, T.J., Hunter, N.M., Wilson, M.D. & Horritt, M.S. (2009) Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology*. 368 (1–4), 42–55.

Neelz, S. & Pender, G. (2010) Benchmarking of 2D Hydraulic Modelling Packages protecting and improving the environment in England and. Environment Agency

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., AghaKouchak, A., Sanders, B., Koren, V., Cui, Z. & Smith, M. (2016) A high resolution coupled hydrologic–hydraulic model (HiResFlood-UCI) for flash flood modeling. *Journal of Hydrology*. 541401–420.

OpenStreetMap (2019) *Map Features*. [Online] [online]. Available from: https://wiki.openstreetmap.org/wiki/Map_Features (Accessed 9 May 2019).

Ordnance Survey (2015) *OS Terrain 50 User Guide*.

Padi, P.T., Baldassarre, G. Di & Castellarin, A. (2011) Floodplain management in Africa: Large scale analysis of flood data. *Physics and Chemistry of the Earth*. 36 (7–8), 292–298.

Palen, L., Soden, R., Anderson, T.J. & Barrenechea, M. (2015) Success & Scale in a Data-Producing Organization. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. 4113–4122.

Pant, R., Thacker, S., Hall, J.W., Alderson, D. & Barr, S. (2018) Critical infrastructure impact assessment due to flood exposure. *Journal of Flood Risk Management*. 11 (1), 22–33.

Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R. & Matgen, P. (2007) Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach. *Journal of Hydrology*. 333 (2–4), 275–287.

Pappenberger, F., Beven, K., Horritt, M. & Blazkova, S. (2005) Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *Journal of Hydrology*. 302 (1–4), 46–69.

Pappenberger, F., Beven, K.J., Ratto, M. & Matgen, P. (2008) Multi-method global sensitivity

analysis of flood inundation models. *Advances in Water Resources*. 31 (1), 1–14.

Pappenberger, F., Dutra, E., Wetterhall, F. & Cloke, H.L. (2012) Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrol. Earth Syst. Sci.* 16 (11), 4143–4156.

Paprotny, D., Morales-Nápoles, O. & Jonkman, S.N. (2017) Efficient pan-European river flood hazard modelling through a combination of statistical and physical models. *Natural Hazards and Earth System Sciences Discussions*. (January), 1–29.

Perks, M.T., Russell, A.J. & Large, A.R.G. (2016) Technical Note: Advances in flash flood monitoring using unmanned aerial vehicles (UAVs). *Hydrology and Earth System Sciences*. 20 (10), 4005–4015.

Plate, E.J. (2002) Flood risk and flood management. *Journal of Hydrology*. 267 (1–2), 2–11.

Pregnolato, M., Ford, A. & Dawson, R. (2015) 'Analysis of the risk of transport infrastructure disruption from extreme rainfall', in *12th International Conference on Applications of Statistics and Probability in Civil Engineering*. [Online]. 2015 p.

Pregnolato, M., Ford, A., Wilkinson, S.M. & Dawson, R.J. (2017) The impact of flooding on road transport: A depth-disruption function. *Transportation Research Part D: Transport and Environment*. 5567–81.

Proietti, S., Lorenzon, F., Uttenthaler, A., Klaus, A. & Probeck, M. (2017) *Overview of Global DEM: Assessment of the current global DEMs and requirements for an updated global DEM*.

Rahman, K.M., Alam, T. & Chowdhury, M. (2012) Location based early disaster warning and evacuation system on mobile phones using OpenStreetMap. *2012 IEEE Conference on Open Systems, ICOS 2012*.

Renschler, C.S. & Wang, Z. (2017) Multi-source data fusion and modeling to assess and communicate complex flood dynamics to support decision-making for downstream areas of dams: The 2011 hurricane irene and schoharie creek floods, NY. *International Journal of Applied Earth Observation and Geoinformation*. 62 (May), 157–173.

Revilla-Romero, B., Bevington, J., Allatson, C., Dempsey, S. & Wood, E. (2018) 'Flood Foresight: A pilot project in the Brahmaputra basin provides flood inundation forecasts in response to the summer 2017 floods.', in *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. [Online]. April 2018 p. 16093.

Rexer, M. & Hirt, C. (2014) Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2.1/v4.1) and validation against accurate heights from the Australian National Gravity Database. *Australian Journal of Earth Sciences*. 61 (2), 213–226.

Rieger, C. & Byrne, J. ~M. (2015) The Role of Citizen Science in Risk Mitigation and Disaster Response: A Case Study of 2015 Nepalese Earthquake Using OpenStreetMap. *AGU Fall Meeting Abstracts*. GC33A-1252.

Robson, A. & Reed, D. (1999) Flood Estimation Handbook. *Institute of Hydrology, Wallingford*.

Rodriguez, E., Morris, C.S., Belz, J.E., Chapin, E.C., Martin, J.M., Daffer, W. & Hensley, S. (2005) *An assessment of the SRTM topographic products*.

Rollason, E., Bracken, L.J., Hardy, R.J. & Large, A.R.G. (2018) The importance of volunteered geographic information for the validation of flood inundation models. *Journal of*

Hydrology.

Rudari, R., Silvestro, F., Campo, L., Rebora, N., Boni, G. & Herold, C. (2015) *Improvement of the global flood model for the GAR 2015*. 69.

Ruf, C.S. & Warnock, A.M. (2018) Remote Sensing of the Terrestrial Water Cycle with the NASA CYGNSS SmallSat Constellation. *AGU Fall Meeting Abstracts*.

Sampson, C.C., Fewtrell, T.J., O'Loughlin, F., Pappenberger, F., Bates, P.B., Freer, J.E. & Cloke, H.L. (2014) The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrology and Earth System Sciences*. 18 (6), 2305–2324.

Sampson, C.C., Smith, A.M., Bates, P.B., Neal, J.C., Alfieri, L. & Freer, J.E. (2015) A high-resolution global flood hazard model. *Water Resources Research*. n/a-n/a.

Sampson, C.C., Smith, A.M., Bates, P.D., Neal, J.C. & Trigg, M.A. (2016) Perspectives on Open Access High Resolution Digital Elevation Models to Produce Global Flood Hazard Layers. *Frontiers in Earth Science*. 3.

Savage, J.T.S., Pianosi, F., Bates, P., Freer, J. & Wagener, T. (2016) Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*. 529146–9163.

Schaefer, J.T. (1990) The Critical Success Index as an Indicator of Warning Skill. Weather and Forecasting 5 (4) p.570–575.

Schaefl, B. & Gupta, H. V. (2007) Do Nash values have value? *Hydrological Processes*. 21 (June), 2075–2080.

Schelhorn, S.J., Herfort, B., Leiner, R., Zipf, A. & Albuquerque, J.P. (2014) 'Identifying Elements at Risk from OpenStreetMap: The Case of Flooding', in S R Hiltz, M S Pfaff, L Plotnick, & P C Shih (eds.) *11th International ISCRAM Conference*. [Online]. 2014 University Park, Pennsylvania: . p.

Schnebele, E., Cervone, G. & Waters, N. (2014) Road assessment after flood events using non-authoritative data. *Natural Hazards and Earth System Sciences*. 14 (4), 1007–1015.

Schorlemmer, D., Beutin, T., Hirata, N., Hao, K., Wyss, M., Cotton, F. & Prehn, K. (2017) 'Global Dynamic Exposure and the OpenBuildingMap - Communicating Risk and Involving Communities', in *EGU General Assembly Conference Abstracts*. EGU General Assembly Conference Abstracts. [Online]. April 2017 p. 7060.

Schumann, G.J.-P. & Bates, P.D. (2018) The Need for a High-Accuracy, Open-Access Global DEM. *Frontiers in Earth Science*. 6 (December), 225.

Schumann, G.J.P., Neal, J.C., Voisin, N., Andreadis, K.M., Pappenberger, F., Phanthuwongpakdee, N., Hall, A.C. & Bates, P.D. (2013) A first large-scale flood inundation forecasting model. *Water Resources Research*. 49 (10), 6248–6257.

Seibert, J., Vis, M.J.P., Lewis, E. & van Meerveld, H.J. (2018) Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*. (February), 1120–1125.

Slater, L.J., Singer, M.B. & Kirchner, J.W. (2015) Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophysical Research Letters*. 42 (2), 370–376.

Smith, A., Sampson, C. & Bates, P. (2015) Regional flood frequency analysis at the global scale. *Water Resources Research*. 51 (1), 539–553.

Smith, L., Liang, Q., James, P. & Lin, W. (2017) Assessing the utility of social media as a data

source for flood risk management using a real-time modelling framework. *Journal of Flood Risk Management*. 10 (3), 370–380.

Sohn, J. (2006) Evaluating the significance of highway network links under the flood damage: An accessibility approach. *Transportation Research Part A: Policy and Practice*. 40 (6), 491–506.

Sperna Weiland, F.C., Vrugt, J.A., van Beek, R.P.H., Weerts, A.H. & Bierkens, M.F.P. (2015) Significant uncertainty in global scale hydrological modeling from precipitation data errors. *Journal of Hydrology*. 529, Part1095–1115.

Stehman, S. V (1997) Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*. 62 (1), 77–89.

Swiss Re (2019) *Enhanced risk models with Global Flood Zones*. [Online] [online]. Available from: https://www.swissre.com/institute/research/topics-and-risk-dialogues/natcat-and-climate/enhanced-risk-models-with-global-flood-zones-tool.html (Accessed 31 May 2019).

Tarekegn, T.H., Haile, A.T., Rientjes, T., Reggiani, P. & Alkema, D. (2010) Assessment of an ASTER-generated DEM for 2D hydrodynamic flood modeling. *International Journal of Applied Earth Observation and Geoinformation*. 12 (6), 457–465.

Thieken, A.H., Apel, H. & Merz, B. (2015) Assessing the probability of large-scale flood loss events: A case study for the river Rhine, Germany. *Journal of Flood Risk Management*. 8 (3), 247–262.

Timonina, A., Hochrainer-Stigler, S., Pflug, G., Jongman, B. & Rojas, R. (2015) Structured Coupling of Probability Loss Distributions: Assessing Joint Flood Risk in Multiple River Basins. *Risk Analysis*. n/a-n/a.

Trigg, M.A., Birch, C., Neal, J.C., Bates, P.D., Smith, A., Sampson, C.C., Yamazaki, D., Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P.J., Winsemius, H.C., Salamon, P., Dottori, F., Rudari, R., Kappes, M., Simpson, A., Hadzilacos, G. & Fewtrell, T.J. (2016) The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*. 11 (9), 1–10.

Tyrna, B., Assmann, A., Fritsch, K. & Johann, G. (2016) Large-scale high-resolution pluvial flood hazard mapping using the raster-based hydrodynamic two-dimensional model FloodAreaHPC. *Flood Risk Management*. 1–14.

Uppala, S.M., Kallberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., et al. (2005) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*. 131 (612), 2961–3012.

Di Vito, P., Fischer, D., Spada, M., Rinaldo, R. & Duquerroy, L. (2018) 'HAPs Operations and Service provision in Critical Scenarios', in *SpaceOps Conference*. [Online]. 2018 pp. 1–8.

Vojinovic, Z. & Tutulic, D. (2009) On the use of 1d and coupled 1d-2d modelling approaches for assessment of flood damage in urban areas. *Urban Water Journal*. 6 (3), 183–199.

Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F. & Koutsoyiannis, D. (2015) One hundred years of return period: Strengths and limitations. *Water Resources Research*. 51 (10), 8570–8585.

Ward, P.J., Jongman, B., Aerts, J.C.J.H., Bates, P.D., Botzen, W.J.W., DIaz Loaiza, A., Hallegatte, S., Kind, J.M., Kwadijk, J., Scussolini, P. & Winsemius, H.C. (2017) A global framework for

future costs and benefits of river-flood protection in urban areas. *Nature Climate Change*. 7 (9), 642–646.

Ward, P.J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., Muis, S., de Perez, E.C., Rudari, R., Trigg, M.A. & Winsemius, H.C. (2015) Usefulness and limitations of global flood risk models. *Nature Clim. Change*. 5 (8), 712–715.

Ward, P.J., Jongman, B., Weiland, F.S., Bouwman, A., Beek, R. van, Bierkens, M.F.P., Ligtvoet, W. & Winsemius, H.C. (2013) Assessing flood risk at the global scale: model setup, results, and sensitivity. *Environmental Research Letters*. 8 (4), 44019.

Weiser, A. & Zipf, A. (2007) 'Web Service Orchestration of OGC Web Services for Disaster Management Free and open GIS for disaster management -the project OK-GIS', in *Geomatics Solutions for Disaster Management*. [Online]. pp. 239–254.

Werner, M.G.F., Hunter, N.M. & Bates, P.D. (2005) Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*. 314 (1–4), 139–157.

White, I. (2008) The absorbent city: urban form and flood risk management. *Proceedings of the Institution of Civil Engineers - Urban Design and Planning*. 161 (4), 151–161.

Wing, O.E.J., Bates, P.D., Sampson, C.C., Smith, A.M., Johnson, K.A. & Erickson, T.A. (2017) Validation of a 30m resolution flood hazard model of the conterminous United States. *Water Resources Research*.

Wing, O.E.J., Bates, P.D., Smith, A.M., Sampson, C.C., Johnson, K.A. & Fargione, J.P. (2018) Estimates of present and future flood risk in the conterminous United States. *Environmental Research Letters*. 13 (3), 34023.

Winsemius, H.C., Van Beek, L.P.H., Jongman, B., Ward, P.J. & Bouwman, A. (2013) A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.* 17 (5), 1871–1892.

Winsemius, H.C., Jongman, B., Veldkamp, T.I.E., Hallegatte, S., Bangalore, M. & Ward, P.J. (2018) Disaster risk, climate change, and poverty: assessing the global exposure of poor people to floods and droughts. *Environment and Development Economics*. (May), 1–21.

Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., Corato, G., Matgen, P. & Bates, P. (2016) Calibration of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques. *Hydrology and Earth System Sciences*. 20 (12), 4983–4997.

WRI (2019) *Global Commission on Adaption*. [Online] [online]. Available from: https://gca.org/global-commission-on-adaptation (Accessed 24 May 2019).

Yamazaki, D. (2018) *MERIT DEM: Multi-Error-Removed Improved-Terrain DEM*. [Online] [online]. Available from: http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/ (Accessed 8 May 2019).

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S. & Bates, P.D. (2017) A high-accuracy map of global terrain elevations. *Geophysical Research Letters*.

Yamazaki, D., Kanae, S., Kim, H. & Oki, T. (2011) A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resources Research*. 47 (4), n/a-n/a.

Yin, J., Yu, D., Yin, Z., Liu, M. & He, Q. (2016) Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China.

Journal of Hydrology. 537138–145.

Yu, D. & Lane, S.N. (2006) Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: Mesh resolution effects. *Hydrological Processes*. 20 (7), 1541–1565.

Yu, D., Yin, J. & Liu, M. (2016) Validating city-scale surface water flood modelling using crowd-sourced data. *Environmental Research Letters*. 111748–9326.

Zhenkui, M. & Redmond, R.L. (1995) Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data. *Photogrammetric Engineering & Remote Sensing*. 61 (4), 435–439.

Zhou, Q., Mikkelsen, P.S., Halsn??s, K. & Arnbjerg-Nielsen, K. (2012) Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *Journal of Hydrology*. 414–415539–549.

Zischg, A.P., Mosimann, M., Bernet, D.B. & Röthlisberger, V. (2017) Validation of 2D flood models with insurance claims. *Journal of Hydrology*. 557250–361.