Aggregation and Disaggregation in Flood Risk Models

A Cumulative Dissertation by
Seth Bryant

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Engineering (Dr.-Ing.)
in Natural Hazards Research

Submitted to the Faculty of Science at the
UNIVERSITY OF POTSDAM

2024

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-65095
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-650952

DECLARATION OF ORIGINALITY

I declare that this work has not been submitted to any other institution of higher education, and that it was prepared independently and exclusively with the specified funds.

This dissertation was developed through independent research. Complete documentation of the scholarly literature and aids applied in the project are provided in the text.

Location and Date	Seth Bryant

SUMMARY

Floods continue to be the leading cause of economic damages and fatalities among natural disasters worldwide. As future climate and exposure changes are projected to intensify these damages, the need for more accurate and scalable flood risk models is rising. Over the past decade, macro-scale flood risk models have evolved from initial proof-of-concepts to indispensable tools for decision-making at global-, national-, and, increasingly, the local-level. This progress has been propelled by the advent of high-performance computing and the availability of global, space-based datasets. However, despite such advancements, these models are rarely validated and consistently fall short of the accuracy achieved by high-resolution local models. While capabilities have improved, significant gaps persist in understanding the behaviours of such macro-scale models, particularly their tendency to overestimate risk. This dissertation aims to address such gaps by examining the scale transfers inherent in the construction and application of coarse macro-scale models. To achieve this, four studies are presented that, collectively, address exposure, hazard, and vulnerability components of risk affected by upscaling or downscaling.

The first study focuses on a type of downscaling where coarse flood hazard inundation grids are enhanced to a finer resolution. While such inundation downscaling has been employed in numerous global model chains, ours is the first study to focus specifically on this component, providing an evaluation of the state of the art and a novel algorithm. Findings demonstrate that our novel algorithm is eight times faster than existing methods, offers a slight improvement in accuracy, and generates more physically coherent flood maps in hydraulically challenging regions. When applied to a case study, the algorithm generated a 4 m resolution inundation map from 30 m hydrodynamic model outputs in 33 s, a 60-fold improvement in runtime with a 25% increase in RMSE compared with direct hydrodynamic modelling. All evaluated downscaling algorithms yielded better accuracy than the coarse hydrodynamic model when compared to observations, demonstrating similar limits of coarse hydrodynamic models reported by others. The substitution of downscaling into flood risk model chains, in place of high-resolution modelling, can drastically improve the lead time of impact-based forecasts and the efficiency of hazard map production. With downscaling, local regions could obtain high resolution local inundation maps by post-processing a global model without the need for expensive modelling or expertise.

The second study focuses on hazard aggregation and its implications for exposure, investigating implicit aggregations commonly used to intersect hazard grids with coarse exposure models. This research introduces a novel spatial classification framework to understand the effects of rescaling flood hazard grids to a coarser resolution. The study derives closed-form analytical solutions for the location and direction of bias from flood grid aggregation, showing that bias will always be present in regions near the edge of inundation. For example, inundation area will be positively biased when water depth grids are aggregated, while volume will be negatively biased when water elevation grids are aggregated. Extending the analysis to effects of hazard aggregation on building exposure, this study shows that exposure in regions at the edge of inundation are an order of magnitude more sensitive to aggregation errors than hazard alone. Among the two aggregation routines considered, averaging water surface elevation grids better preserved flood depths at buildings than averaging of water depth grids. The study provides the first mathematical proof and generalizeable treatment of flood hazard grid aggregation, demonstrating important mechanisms to help flood risk modellers understand and control model behaviour.

The final two studies focus on the aggregation of vulnerability models or flood damage functions,

investigating the practice of applying per-asset functions to aggregate exposure models. Both studies extend *Jensen's inequality*, a well-known 1906 mathematical proof, to demonstrate how the aggregation of flood damage functions leads to bias. Applying Jensen's proof in this new context, results show that typically concave flood damage functions will introduce a positive bias (overestimation) when aggregated. This behaviour was further investigated with a simulation experiment including 2 million buildings in Germany, four global flood hazard simulations and three aggregation scenarios. The results show that positive aggregation bias is not distributed evenly in space, meaning some regions identified as "hot spots of risk" in assessments may in fact just be hot spots of aggregation bias. This study provides the first application of Jensen's inequality to explain the overestimates reported elsewhere and advice for modellers to minimize such artifacts.

In total, this dissertation investigates the complex ways aggregation and disaggregation influence the behaviour of risk models, focusing on the scale-transfers underpinning macro-scale flood risk assessments. Extending a key finding of the flood hazard literature to the broader context of flood risk, this dissertation concludes that *all else equal, coarse models overestimate risk*. This dissertation goes beyond previous studies by providing mathematical proofs for how and where such bias emerges in aggregation routines, offering a mechanistic explanation for coarse model overestimates. It shows that this bias is spatially heterogeneous, necessitating a deep understanding of *how* rescaling may bias models to effectively reduce or communicate uncertainties. Further, the dissertation offers specific recommendations to help modellers minimize scale transfers in problematic regions. In conclusion, I argue that such aggregation errors are epistemic, stemming from choices in model structure, and therefore hold greater potential and impetus for study and mitigation. This deeper understanding of uncertainties is essential for improving macro-scale flood risk models and their effectiveness in equitable, holistic, and sustainable flood management.

ZUSAMMENFASSUNG

Hochwasser sind weiterhin die größte Ursache für wirtschaftliche Schäden und Todesopfer durch Naturkatastrophen weltweit. Da prognostiziert wird, dass künftige Klima- und Expositionsänderungen diese Schäden noch verstärken werden, steigt der Bedarf an genaueren und skalierbaren Hochwasserrisikomodellen. In den letzten zehn Jahren haben sich makroskalige Hochwasserrisikomodelle von ersten Machbarkeitsversuchen zu unverzichtbaren Instrumenten für Entscheidungsträger auf globaler, nationaler und zunehmend auch lokaler Ebene entwickelt. Dieser Fortschritt wurde weitgehend durch das Aufkommen von Hochleistungsrechnern und die Verfügbarkeit globaler, weltraumgestützter Datensätze vorangetrieben. Trotz dieser Fortschritte werden diese Modelle jedoch nur selten validiert und bleiben durchweg hinter der Genauigkeit zurück, die von hochauflösenden lokalen Modellen erreicht wird. Obwohl sich die Möglichkeiten verbessert haben, gibt es immer noch erhebliche Lücken im Verständnis der Skalenübertragungen, die typischerweise in großskaligen Modellen verwendet werden, insbesondere zu deren Tendenz zur Überschätzung von Risiken. Diese Dissertation zielt darauf ab, diese Lücken zu schließen, indem sie diese Skalentransfers untersucht, die mit der Konstruktion und Anwendung solcher groben makroskaligen Modelle einhergehen. Zu diesem Zweck werden vier Studien vorgestellt, die gemeinsam die Expositions-, Gefährdungs- und Vulnerabilitätskomponenten des Risikos betrachten, welche durch Upscaling oder Downscaling beeinflusst werden.

Die erste Studie konzentriert sich auf eine Art von Downscaling, bei dem grobe Überschwemmungsraster auf eine feinere Auflösung umskaliert werden. Eine solche Herabskalierung von Überschwemmungsdaten wurde zwar bereits in zahlreichen globalen Modellketten eingesetzt, doch ist unsere Studie die erste, die sich speziell auf diese Komponente konzentriert und eine Bewertung des Standes der Technik sowie einen neuartigen Algorithmus liefert. Die Ergebnisse zeigen, dass unser neuartiger Algorithmus achtmal schneller ist als bestehende Methoden, eine leichte Verbesserung der Genauigkeit bietet und physikalisch kohärentere Überschwemmungskarten in hydraulisch anspruchsvollen Regionen erzeugt. Bei der Anwendung in einer Fallstudie erzeugte der Algorithmus eine Überschwemmungskarte mit einer Auflösung von 4 Metern aus den Ergebnissen eines 30-metrigen hydrodynamischen Modells in 33 Sekunden, was eine 60-fache Verbesserung der Laufzeit bei einem 25-prozentigen Anstieg des RMSE im Vergleich zur direkten hydrodynamischen Modellierung darstellt. Alle bewerteten Downscaling-Algorithmen erbrachten – im Vergleich zu Messergebnissen – eine bessere Genauigkeit als das grobe hydrodynamische Modell, was die Grenzen der groben hydrodynamischen Modelle aus früheren Studien bestätigt. Der Einsatz von Downscaling in Hochwasserrisikomodellketten anstelle von hochauflösender Modellierung kann die Vorlaufzeit von auswirkungsbasierten Vorhersagen und die Effizienz der Erstellung von Gefahrenkarten drastisch verbessern. Downscaling könnte es regional erlauben, hochauflösende lokale Überschwemmungskarten aus einem globalen Modell zu generieren, ohne dass teure Modellierung oder Fachwissen erforderlich sind.

Die zweite Studie befasst sich mit der Aggregation von Gefahren und deren Auswirkungen auf die Exposition, wobei implizite Aggregationen untersucht werden, die üblicherweise Verwendung finden, wenn Gefahrenraster mit groben Expositionsmodellen verknüpft werden. In dieser Studie wird ein neuartiges räumliches Klassifizierungsframework eingeführt, um die Auswirkungen der Umskalierung von Hochwassergefahrenrastern auf eine gröbere Auflösung zu verstehen. Die Studie leitet analytische Lösungen in geschlossener Form für die räumliche Ausprägung und die Richtung des systematischen

Fehlers durch die Aggregation von Überschwemmungsrastern ab und zeigt, dass dieser Fehler in Regionen nahe der Überschwemmungsgrenze immer auftritt. So wird beispielsweise die Überschwemmungsfläche positiv beeinflusst, wenn Einstautiefenraster aggregiert werden, während das Hochwasservolumen negativ beeinflusst wird, wenn Wasserhöhenraster aggregiert werden. Bei der Ausweitung der Analyse auf die Auswirkungen der Gefahrenaggregation auf die Gebäudeexposition zeigt diese Studie, dass die Exposition in Regionen am Rande der Überschwemmung um eine Größenordnung empfindlicher auf Aggregationsfehler reagiert als die Hochwassergefahr allein. Von den beiden betrachteten Aggregationsroutinen bewahrte die Mittelung von Wasseroberflächen-Höhenrastern die Überschwemmungstiefen an Gebäuden besser als die Mittelung von Einstautiefenrastern. Die Studie liefert den ersten mathematischen Beweis und eine verallgemeinerungsfähige Aggregationsmethode für Hochwassergefahrenraster und zeigt wichtige Mechanismen auf, die Hochwasserrisikomodellierern helfen können, das Modellverhalten zu verstehen und zu kontrollieren.

Die letzten beiden Studien konzentrieren sich auf die Aggregation von Vulnerabilitätsmodellen oder Hochwasserschadensfunktionen und untersuchen die Praxis der Anwendung von gebäudespezifischen Funktionen auf aggregierte Expositionsmodelle. Beide Studien erweitern die Jensen'sche Ungleichung, einen bekannten mathematischen Beweis, um zu zeigen, wie die Aggregation von Flutschadensfunktionen zu systematischen Fehlern führt. Die Anwendung von Jensens Beweis aus dem Jahr 1906 in diesem neuen Kontext zeigt, dass typischerweise konkave Hochwasserschadensfunktionen zu einer systematischen Überschätzung führen, wenn sie aggregiert werden. Dieses Verhalten wurde anhand eines Simulationsexperiments mit 2 Millionen Gebäuden in Deutschland, vier globalen Hochwasserrisikosimulationen und drei Aggregationsszenarien weiter untersucht. Die Ergebnisse zeigen, dass die positive Aggregationsverzerrung nicht gleichmäßig im Raum verteilt ist, was bedeutet, dass einige Regionen, die in den Bewertungen als „Hochrisikogebiete" identifiziert wurden, in Wirklichkeit bloß „Hochfehlergebiet" seien könnten. Dies ist die erste Anwendung der Jensen'schen Ungleichung zur Erklärung der Überschätzungen, die in anderen Studien für grobskalige Hochwasserrisikomodelle berichtet wurden.

Diese Dissertation untersucht die komplexe Art und Weise, wie Aggregation und Disaggregation das Verhalten von Risikomodellen beeinflussen, und konzentriert sich dabei auf die Skalentransfers, die den makroskaligen Hochwasserrisikobewertungen zugrunde liegen. Indem sie eine wichtige Erkenntnis aus der Literatur zur Hochwassergefahr auf den breiteren Kontext des Hochwasserrisikos ausweitet, zeigt diese Dissertation, dass grobe Modelle das Risiko überschätzen, wenn alle anderen Faktoren gleich sind. Diese Dissertation geht über frühere Studien hinaus, indem sie mathematische Beweise dafür liefert, wie und wo systematische Fehler in Aggregationsroutinen entstehen, und bietet eine mechanistische Erklärung für die Überschätzung in groben Modellen. Es wird gezeigt, dass diese Verzerrung räumlich heterogen ist, was ein tiefes Verständnis der Art und Weise erfordert, wie die Umskalierung Modelle verzerren kann, um Unsicherheiten effektiv zu reduzieren oder zu kommunizieren. Die Dissertation enthält spezifische Empfehlungen, die Modellierern helfen sollen, Skalenübertragungen in problematischen Regionen zu minimieren. Abschließend argumentiere ich, dass solche Aggregationsfehler epistemischer Natur sind, da sie aus der Wahl der Modellstruktur resultieren, und daher ein größeres Potenzial und einen stärkeren Anstoß zur Verbesserung bieten sollten. Dieses tiefere Verständnis von Unsicherheiten ist unerlässlich, wenn makroskalige Hochwasserrisikomodelle ihr volles Potenzial als Werkzeuge für ein gerechtes, ganzheitliches und nachhaltiges Zusammenleben mit extremen Überschwemmungen entfalten sollen.

TABLE OF CONTENTS

ABBREVIATIONS

CSI. Critical Success Index.

DD. Dry-Dry resample case.

DEM. Digital Elevation Model.

DP. Dry-Partial resample case.

EAD. Estimated (or Expected) Annual Damages.

FN. False Negative.

FP. False Positive.

GIS. Geographic Information Systems.

OSM. Open Street Map.

RL. Relative Loss.

RMSE. Root Mean Square Error.

SRTM. Shuttle Radar Topography Mission.

TN. True Negative.

TP. True Positive.

WP. Wet-Partial resample case.

WSE. Water Surface Elevation.

WSH. Water Surface Height.

WW. Wet-Wet resample case.

Chapter 1

INTRODUCTION

Water is essential and ubiquitous, comprising roughly 60% of the human body (Mitchell et al., 1945) and covering 70% of the earth's surface (Gleick, 1993). However, when water arrives unexpectedly and inundates the unprepared, the consequences can be devastating. For example, in June 2013 when 300 mm of rain fell on snow in Southern Alberta, Canada severing the country's main highway and killing 5 (Pomeroy et al., 2016); and May 2018 when rapid snowmelt led to record water levels in New Brunswick, Canada inundating 13,000 homes (Hrabluk, 2019); and when a cyclone stalled over central Europe in July 2021 dropping 157 mm of rain in 72 hours killing 184 in Germany (Dietze et al., 2022), and in June 2022 when an early monsoon coincided with a record-breaking heat wave to displace an estimated 33 million and kill 1,200 in Pakistan (Mallapaty, 2022). Events like these contribute to floods remaining the largest source of economic damages and fatalities from natural disasters globally (Douris and Kim, 2021).

1.1 Motivation and Background

Unexpected flooding can have vast and complex consequences for humans. For example, flooding can: displace populations (Vestby et al., 2024), lead to crop loss and food insecurity (Reed et al., 2022), disable power production (Jackson and Gunda, 2021), spread disease (Paterson et al., 2018), disrupt transportation (Pregnolato et al., 2017), and lead to loss of life (Petrucci, 2022). To classify these diverse consequences, three dimensions are commonly used in hazards research. The first dimension, tangibility, relates to the ease with which the consequence can be monetized. In this dimension, consequences are divided into tangible (easily priced or enumerated) and intangible (difficult to monetize or enumerate) (Merz et al., 2010). The second dimension, mechanistic, pertains to how the loss was realized. This includes direct consequences, which are damages caused by contact with flood waters, debris, or ice; and indirect consequences, which are those damages resulting from interruptions, or recovery, from flooding (Parker et al., 1987). The third dimension is value-based and divides consequences into themes that facilitate decision-making and risk communication. For example, the Sendai Framework uses 38 indicators (e.g., mortality, exposed population, economic loss), to track global progress in disaster risk reduction (UNISDR, 2017).

Despite the challenges and occasional devastation of living with floods, flooding plays an important role in landscapes, ecosystems, and societies. For example, 15,000-18,000 years ago, in what is thought to be the largest ever on earth, a flood with a maximum outburst discharge of $20 \times 10^9\ \mathrm{m^3\ s^{-1}}$ sculpted much of the Columbia Basin in the Northwestern US (O'Connor et al., 2020). Such megafloods are not unique to Earth, with recent data from Mars suggesting outburst flooding is responsible for carving roughly a quarter of Martian valleys (Goudge et al., 2021). On shorter timescales, periodic river flood pulses are essential for a healthy ecosystem, maintaining temporal and spatial habitat variability, high biodiversity, and biotic and abiotic interactions like the influx of nitrogen (Petsch et al., 2023). When rivers are left connected to their floodplains, neighbouring communities also benefit from soil formation, water quality improvements, and groundwater recharge (Talbot et al., 2018). These examples demonstrate the complexity and inter-dependence between flood-cycles, society, and the environment.

Numerous frameworks have been proposed to classify flood events, most describing the flood generation

or initiation process (Mohor, 2022). The simplest classification only divides coastal and inland flooding (UNISDR, 2009), but often the latter is further divided into fluvial, flash, ice-jam, and glacial lake outburst (IRDR, 2014). Rain-driven inundation (without major up-catchment contributions) is commonly called *pluvial* flooding. When this specifically affects urban areas and drainage networks, it is often called *urban* flooding (Zevenbergen et al., 2010). Importantly, other triggers or mechanisms exist that don't fit neatly into these categories, such as dam breaks, controlled releases, groundwater-driven floods, and landslide-triggered floods (Buttle et al., 2016). Often multiple mechanisms interact in complex ways, partially explaining why floods have posed consistent challenges for societies throughout history.

1.1.1 Managing Floods

Pre-modern societies coexisted with *regular* flood cycles through a mix of dependence (e.g., Nile, Yangtze, and Euphrates civilizations) and mobility (Hoeksema, 2006; McBain, 2012; Angelakis et al., 2023). On the other hand, *extreme* events have always been more challenging, driving many civilizations to attempt "flood control". This paradigm remained popular until the mid-20th century when implementation costs and frequent failures led many to question the treatment of extreme floods as an engineering problem to be solved with physical infrastructure (National Research Council, 2013). In response, modern decision-making around floods emphasizes management rather than control, acknowledging the futility and harm of the latter. Similarly, decades of practice and research have shown that effective flood management must be risk-based, where both costs and benefits are weighed in a probabilistic way (Merz et al., 2014; Sayers, 2012; Schanze et al., 2006; Simonovic, 2013). While there is no consensus on a precise definition of *flood risk*, with authors adapting it to different contexts and frameworks (Shi, 2019; Ranke, 2016; Scheuer et al., 2012; Kaplan and Garrick, 1981; Zevenbergen et al., 2010), a simple definition of *risk* is "the product of probability and consequence" widely attributed to Knight (1921). In a flood management context, the term is often employed to highlight the importance of considering not just hydraulic processes or *hazards* but also their consequences to objects, systems, or elements of concern — called *assets* — and the probability of those consequences occurring. To better conceptualize the relationship between hazards and consequences, flood risk is often deconstructed into the following components: hazard, exposure (assets subject to harm), and vulnerability (factors affecting an asset's susceptibility to impact) (UNISDR, 2009). Crichton's risk triangle graphically illustrates this concept (Figure 1.1), showing how changes in magnitude of any of these ingredients are positively related to a change in risk (Crichton, 1999). Modern conceptualizations of flood risk expand upon this foundation, viewing risk as the product of complex, dynamic interactions within interdependent systems (Birkmann, 2013). For example, the Bogardi, Birkman, and Cardona (BBC) conceptual framework asserts that risk is produced through the intersection of hazard and vulnerability, which in turn trigger management; a cycle that is dynamic in space and time (Birkmann, 2006). Barendrecht et al. (2017) provides a less-specific and more broad framework of six inter-connected systems: technology, environment, society, hydrology, politics, and the economy. While such system-based frameworks differ in their organization and connections, they emphasize some common features and challenges of flood risk management. For example, they highlight that flood risk generation and management are cyclical; risk can never be eliminated, only reduced; and the beneficial or damaging aspects of floods depend on context (Bedient, 2013; Sayers, 2012).

Despite broad acceptance of holistic flood risk management among experts, many communities struggle to implement such practices in the face of both technical and political obstacles (Tariq et al., 2020). For example, unprecedented or unexpected events remain a major challenge, as these floods often overtop or

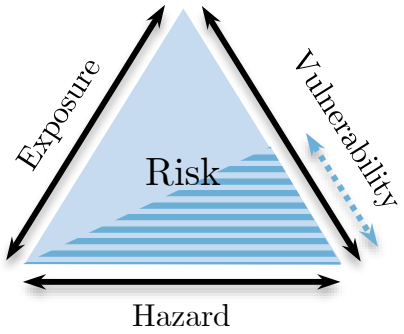


Figure 1.1: Crichton's conceptual risk triangle where *risk* is represented by the area of a triangle with leg dimensions corresponding to each of the three risk components. Dashed arrow shows an example where vulnerability is reduced leading to a reduction in risk (i.e., area) (Crichton, 1999)

exceed the expectations of management and infrastructure with devastating consequence (Kreibich et al., 2022; Bertola et al., 2023). However, a more comprehensive perspective evaluating similar cross-border catchments demonstrates that most such events, while surprising locally, are unsurprising at a continental scale (Bertola et al., 2023). While increased investment and improved management can mitigate impacts of even unprecedented events, Kreibich et al. (2022) found this occurred in only two out of 45 cases, highlighting a persistent challenge. This obstacle is exacerbated by the unequal distribution of resources, data, models, knowledge, and risk (Gerl et al., 2016; Antwi-Agyakwa et al., 2023; Poussard et al., 2021; Peduzzi et al., 2009; Ward et al., 2020; Yarveysi et al., 2023; Winsemius et al., 2016). For example, Wing et al. (2022)'s US-wide model revealed that current and future flood risk disproportionately burdens poor and Black communities. This underscores both the opportunity and the obligation for the flood modelling community to develop and implement more accessible and comprehensive models.

1.1.2 Past and Future Changes

Recent observations across all components of risk suggest floods and society are increasingly dynamic, further challenging managers and the models they depend on. For example, a global analysis from 2000-2018 shows that extreme precipitation has increased in two-thirds of meteorological stations, likely driven by rising temperatures (Sun et al., 2021). Because the relationship between extreme precipitation and river flooding is complex, this same consistent increase has not been observed in all regions, with both increases and decreases in discharge observed in the past decades (Blöschl et al., 2019; Tramblay et al., 2019; Kemter et al., 2020). Focusing on flood timing, Blöschl et al. (2017) analyzed a pan-European database and revealed that warmer temperatures have shifted flooding either earlier or later in the hydrologic year depending on the region. Similar findings were reported for ice-jam flood timing in Canada (Rokaya et al., 2018). These changes in flood timing can disrupt both ecosystems (Rytwinski et al., 2020) and emergency response (Yu et al., 2020). Corresponding dynamics in exposure are similarly alarming. Using daily 250 m multi-spectral global imagery, Tellman et al. (2021) identified a 20 to 24% increase in the proportion of the global population exposed to floods between 2000 and 2015; driven primarily by settlement trends. Leveraging a gridded reconstruction of flood exposure across 37 European countries and a novel historical flood database, Paprotny et al. (2018) analyzed population and land use data at local scales. Their findings reveal an increase in annually inundated area and impacted individuals since 1870, alongside a significant reduction in fatalities and a recent downward trend in financial losses. Overall, there is strong evidence of a substantial increase in global flood risk and economic losses in

recent decades (Kundzewicz et al., 2014, 2019; Duan et al., 2022). Attributing these observed changes to specific drivers (e.g., exposure, vulnerability, climate change) is more challenging (Metin et al., 2018); with most finding exposure (Tramblay et al., 2019; Kundzewicz et al., 2014, 2019), but some finding vulnerability (Kreibich et al., 2017), or hazard (Löwe et al., 2017) as the dominant driver of increased flood risk.

The Intergovernmental Panel on Climate Change has assessed that "flood risks and societal damages are projected to increase with every increment of global warming" (IPCC, 2023, p. 62), suggesting future changes will intensify existing flood management challenges. For example, fluvial flood risk in Europe is projected to increase seven- to tenfold by the end of the century, driven primarily by socioeconomic growth alongside climate change (Steinhausen et al., 2022). Employing a global model, Dottori et al. (2018) developed a river flood database by downscaling multi-model hydrological ensemble simulations to create gridded daily runoff projections at $0.5° \times 0.5°$ resolution. With this, they assessed impacts by overlaying flood inundation maps with global population data, estimating that under a 1.5°C warming scenario, flood-related fatalities could rise by 70-83%. However, uncertainties in these forecasts are large and comparisons difficult owing to different boundary conditions, forcings, and assumptions (Kundzewicz et al., 2019).

1.1.3 Flood Risk Modelling

To support risk-based decision-making and modern flood management, communities and governments frequently undertake flood risk assessments (Messner, 2007; Salman and Li, 2018; Scheuer et al., 2012; Merz and Thieken, 2004; IWR and USACE, 2017). For example, the Delta Risk Management Strategy Assessment employed coupled models for waterworks operations, hydrodynamics, and levee failure to guide risk reduction strategies and mitigation measures in the Sacramento-San Joaquin Delta in California (URS, 2008). Similarly, the VNK2 Project in the Netherlands estimated failure probabilities and consequences for each levee system, informing maintenance priorities, research efforts, and flood-related legislation (VNK, 2012). At the core of quantitative flood risk assessments like these is a *flood risk model* or collection of models used to understand floods and their societal consequences.

Flood risk models are a collection of procedures, algorithms, and mathematical functions used to simplify, study, and quantify the components of our world that relate to flood risk. In this way, each such model formalizes a framework or understanding of *how* flood risk manifests within society. While modern hazard science recognizes flood risk as arising from complex, interconnected systems (Barendrecht et al., 2017), most studies focus on acute, localized events where many broader systemic components can be reasonably simplified. Under these assumptions, contemporary flood risk models often use a static, sequential chain of submodels, each representing a specific stage in the transformation of rainfall into societal consequences. For example, an early model chain developed by Falter et al. (2016) implements five such submodels: hydrometeorology, rainfall-runoff, river routing, inundation, and damages; each representing a sub-discipline with rich literature, diverse methods, and complex uncertainties. Applying Crichton (1999)'s risk triangle framework (Figure 1.2), Falter et al. (2016)'s initial submodels (hydrometeorology to inundation) address the *hazard* component, while the damage submodel simulates both *exposure* and *vulnerability*. Figure 1.2 shows a simple schematic of this, where the hazard submodel(s) compute hazard indicators (e.g., flood depth, velocity) which are passed to the *damage model* to compute event-based consequences (e.g., replacement cost) for a simulated flood event. In this framework, the damage model itself is composed of an *exposure model* (a dataset of potentially flood-exposed elements, e.g., houses,

with relevant attributes), a *vulnerability model* (a functional relation for relative loss), and calculations to link these. During the *damage calculation* step (Figure 1.2), hazard indicators (e.g., flood depth) are spatially joined or intersected with the exposure model to provide the features needed by the vulnerability model to quantify consequences. The simplest vulnerability models are uni-variable functions of water depth and flood damage widely called *depth-damage functions* (White, 1945). While depth-damage functions are still common, research in the past two decades has advanced the understanding of flood damage processes and, along with new data and techniques, has led to more sophisticated and accurate multi-variate models (Schröter et al., 2014).

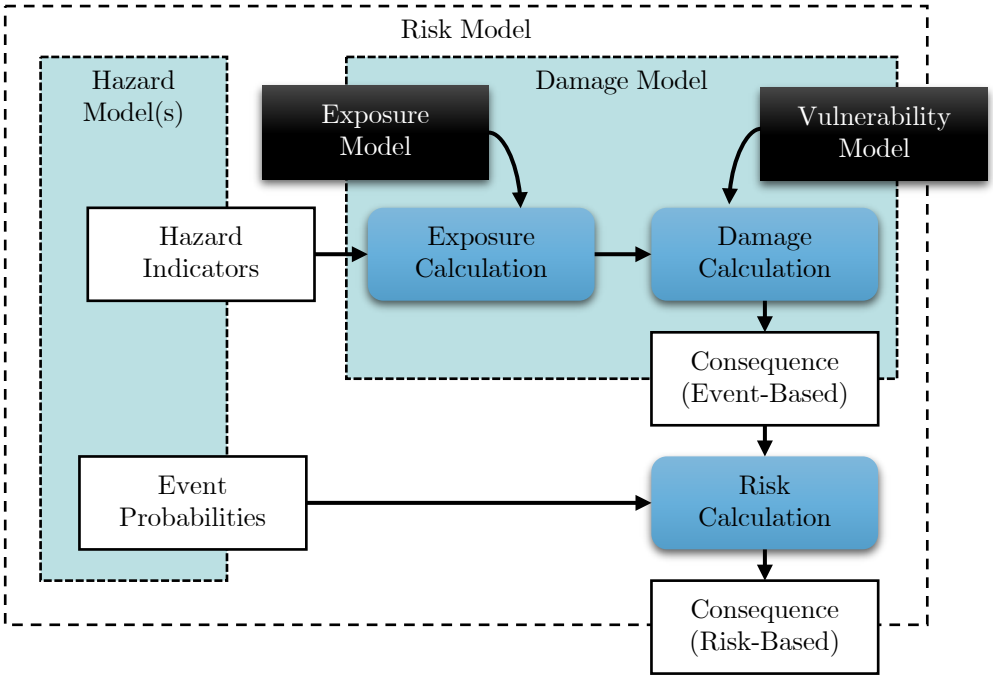


Figure 1.2: Schematic of a flood risk model.

In data scarce regions, developing such models can be challenging, especially in post-event analysis. To support disaster management in these contexts, hazard and exposure estimates derived directly from space-based observations have become increasingly popular (Cohen et al., 2018, 2022; Bryant et al., 2022a). For example, in Bryant et al. (2022a) I developed an algorithm for estimating flood depths from interferometric synthetic aperture radar derived flood footprints and a high-resolution terrain model. This algorithm implemented a series of filters and normalizations using height above nearest drainage (HAND) estimates. While the new algorithm improved upon the accuracy of existing methods, these gains came at the expense of algorithm complexity and higher computational cost. In Mester et al. (2023), I applied this same flood depth algorithm to support the estimation of displacement caused by the 2019 Tropical Cyclone Idai in Mozambique, with and without climate change. The study found a relatively minor climate change signal (2.7% to 3.2%); however, the methodology can be applied to similar events to understand the influence of drivers or help allocate relief.

A comprehensive flood risk assessment requires consideration of a wide range of potential flood events (hazards), alongside their corresponding exposures and vulnerabilities, under both current and future conditions. However, such comprehensive analysis is difficult and rare. Most studies employ a limited

set of scenarios (e.g., current vs. future conditions) and selected flood events with associated exceedance probabilities (e.g., 50- and 100-year floods) (URS, 2008; VNK, 2012). This event selection brings additional assumptions and complexities, particularly with respect to spatial dependence in large domains (see discussion below). State-of-the-art models address this through continuous simulations with large time spans and synthetic climate forcings. This allows for more accurate probability estimates through ordering events by consequence magnitude (Sairam et al., 2021). To facilitate communication and decision-making on floods and their probabilities, models often integrate estimated damages across exceedance probabilities to obtain the expected annual damages (EAD) (US Army Corps of Engineers, 1996; Messner, 2007; Merz et al., 2010; Penning-Rowsell et al., 2013). However, some have criticized the over-reliance on EAD, for example due to the implicit assumption that society is "risk neutral". This assumption overlooks the out-sized importance of truly rare and catastrophic events on a society and presumes that decision makers are comfortable with some degree of loss (Haimes, 2009; IEC and ISO, 2019; Merz et al., 2009).

1.1.4 Model Domain and Scale

To be useful, models simplify reality for easier understanding (Eisenberg et al., 2019). A key simplification is often the use of spatial and temporal boundaries, in contrast to reality which is effectively unbounded. In flood risk models, the spatial boundary is a modeller-defined region or geodetic polygon called the model *domain* or *extent*. While obvious in some studies (e.g., a continent), domain selection can be subjective, often influencing the analysis (Metin et al., 2020). Domains range from local to global, suiting different research questions (de Moel et al., 2015). This spectrum is often divided into micro-, meso-, and macro-scales, implying a trade-off between effort and accuracy (Messner and Meyer, 2006). To these three scales de Moel et al. (2015) adds a fourth "supra-national" scale for global and continental models; however, they observe that with recent advancements in data and computing the gap between global- and micro-scale models is shrinking.

Fluvial flood risk models at the local or micro-scale, which focus on a single catchment or river reach, are the most commonly used and have the longest history of application (FEMA, 2012; Messner, 2007; Merz et al., 2010). For example, Bryant et al. (2022b) developed a stochastic flood risk model, constructed from five years of remote-sensing data, property assessment records, and hazard scenarios, to simulate the influence of flood protections and urban development on flood risk in a mature urban neighbourhood along the Bow River in Alberta, Canada. These fit-for-purpose models often benefit from rich data and calibration; however, their cross-comparison is difficult as contexts and assumptions can vary widely between studies (Guo et al., 2021). Further, construction and computation costs for such models typically limit their application to wealthier regions (Gerl et al., 2016; Antwi-Agyakwa et al., 2023).

With the advent of space-based data collection and computational advancements, *macro-scale* national and global flood risk models have become increasingly popular. For example, Peduzzi et al. (2009) developed the Disaster Risk Index for a quantitative assessments of risk from four hazards modelled by intersecting intensity with global population density at $0.5° \times 0.5°$ resolution. Winsemius et al. (2016) built the first global flood risk model to project future changes under socioeconomic and climate scenarios, using the GLOFRIS model chain (Winsemius et al., 2013) to compute fluvial hazards at $0.5° \times 0.5°$ resolution, then downscaling to $30'' \times 30''$ with a hole-filled Shuttle Radar Topography Mission (SRTM) global digital elevation model (DEM). Their results indicated a 20% increase in flood risk across the 20 major global river basins. Sampson et al. (2015) developed a similar global hazard modelling chain with

LISFLOOD-FP as the hydraulic engine to simulate inundation at $30'' \times 30''$ resolution downscaled to $3'' \times 3''$ using a variant of the same DEM. Focusing on the US, Wing et al. (2022) employed the model chain of Bates et al. (2021) to simulate five return periods at $1'' \times 1''$ resolution, further downscaled to $\frac{1}{9}'' \times \frac{1}{9}''$ using a LiDAR-derived DEM mosaic. They coupled these simulations with US Army Corps depth-damage functions (FEMA, 2012) aggregated to census-tracts to compute EAD for 2020 and 2050 climate states, demonstrating correlations between flood risk and demographics.

1.1.5 Scale Transformations

Flood risk modellers, like all applied modellers, employ numerous transfers to translate information across scales. Bierkens et al. (2000) identifies three such scales: the *observation scale* (where measurements like water height or flood damage are made), the *model scale* (defined by the model's native output resolution, e.g., $30'' \times 30''$ flood depths), and the *policy scale* (where findings are reported, e.g., per-country flood risk projections). To this I add a fourth scale: the *consumer scale* where data products are transferred to other projects and parties (e.g., output datasets). When an operation transfers from a smaller to a larger scale, like when outputs are aggregated for reporting, this is called *upscaling* and the reverse *downscaling*. Relevant operations for upscaling and downscaling flood models are a *change of domain*, like upscaling the domain from a city to a country, and a *change of support*, where support is defined as the geographic size of the discretization (e.g., $10\,\mathrm{m}^2$ per-pixel). These transfers apply to both spatial and temporal scales; however, the remainder of this dissertation focuses on spatial transfers and changes of support, like resolution changes from $30'' \times 30''$ to $0.5^\circ \times 0.5^\circ$.

In the context of grid-based flood risk models, support is closely related to the more familiar term *resolution*, which describes the geographic dimensions of a rectangular grid cell (e.g., $10\,\mathrm{m} \times 10\,\mathrm{m}$) (Degbelo and Kuhn, 2018). For datasets like DEMs used within these models, terms like matrix, array, or the Geographic Information Systems (GIS)-specific term *raster* may be used interchangeably — the latter originally borrowed from the realm of digital photography (Bach et al., 1997). To avoid ambiguity with the terms *increasing* and *decreasing* resolution, this dissertation adopts *fine* to describe higher resolution and *coarse* for lower resolution. Therefore:

$$\text{fine} \rightarrow \text{coarse} \equiv \text{upscaling or aggregating}$$
$$\text{coarse} \rightarrow \text{fine} \equiv \text{downscaling or disaggregating}$$

Numerous algorithms exist for grid transformations like the above, commonly termed *rescaling* or *resampling* operations. Common downscaling algorithms include *Nearest Neighbour* (2D) which uses the nearest coarse grid centroid value, *Bilinear* which employs linear interpolation in two directions, and *Bicubic Convolution* which fits a curve through a 4×4 neighbourhood of cells (Keys, 1981; Parker et al., 1983). Upscaling algorithms can employ any aggregating statistic like the mean, mode, max, or min on a neighbourhood of cells. However, while data transformations like these are relatively simple to implement, their implications on model behaviour are rarely considered.

1.1.6 Issues of Scale

The following well-developed research areas offer relevant insights into issues of scale in flood risk modelling: 1) spatial dependence of flood generation processes; 2) disaggregation employed to construct exposure models; and 3) sensitivity to resolution. These are summarized in turn in the following paragraphs.

Traditional flood risk assessments calculate EAD at catchment scales, using return periods derived from individual discharge gauges, then aggregate these results for regional mapping or total EAD estimation (Ward et al., 2013; Hall et al., 2005; Sampson et al., 2015). While these maps are useful at the local-scale, this aggregation approach inherently assumes complete dependence between catchment flood processes, neglecting the spatial heterogeneity characteristic of real floods and their underlying drivers (Vorogushyn et al., 2018; Nguyen et al., 2020). To quantify the error introduced by this spatial dependence assumption, Metin et al. (2020) employed a continuous simulation model, demonstrating that the traditional approach tends to overestimate risk for larger, more extreme flood events. Similar findings have been reported for smaller domains (Lamb et al., 2010; Wyncoll and Gouldby, 2015).

Disaggregation is a common technique in exposure modelling, used to map coarse, non-flood-specific datasets (e.g., national census data) onto higher-resolution asset geometries (e.g., individual buildings) (Röthlisberger et al., 2018; Seifert et al., 2010b; Wünsch et al., 2009; Thieken et al., 2006; Paprotny et al., 2020b,a; Figueiredo and Martina, 2016; Nievas et al., 2022; Sieg and Thieken, 2022; Molinari and Scorzini, 2017). Generalizing findings from disaggregation studies is challenging, as accuracy often depends heavily on the quality of the source data rather than the disaggregation itself. Examples of both under- and over-estimation have been reported (Sieg and Thieken, 2022; Röthlisberger et al., 2018; Figueiredo and Martina, 2016; Molinari and Scorzini, 2017).

In general, computational cost scales with grid size, motivating numerous studies on the sensitivity of hydrodynamic inundation models to resolution changes. These studies generally demonstrate that resolution plays a critical role in flood hazard prediction accuracy, with coarser resolutions consistently overestimating inundation area and depth for fluvial floods (Fewtrell et al., 2008; Savage et al., 2016; Papaioannou et al., 2016; Alipour et al., 2022; Banks et al., 2015; Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021; Xafoulis et al., 2023). This overestimation at coarser resolutions is likely due to poorly defined river channels, resulting in reduced conveyance (Muthusamy et al., 2021).

Some studies have considered the sensitivity to resolution in vulnerability or exposure, with most also finding that coarse models tend to overestimate (Komolafe et al., 2015; Brussee et al., 2021). However, the sensitivity to resolution can vary based on the model type and the data used (Ghimire and Sharma, 2021). Furthermore, when both hazard variables and building attributes are aggregated, the bias can shift from negative to significantly positive, making it difficult to predict ex-ante (Pollack et al., 2022). Table 3.1 summarizes the few studies to have considered resolution sensitivity of a complete risk model, all showing coarse models overestimate damage and risk (Apel et al., 2009; Sieg et al., 2019; Seifert et al., 2010b; Komolafe et al., 2015; Brussee et al., 2021; Ghimire and Sharma, 2021; Pollack et al., 2022). These studies generally highlight that errors from missing data and uncertainties in vulnerability models are often more significant than those arising from typical aggregations (Pollack et al., 2022; Brussee et al., 2021); however, the magnitude of these uncertainties varies widely. For example, in Brussee et al. (2021)'s sensitivity analysis of a Dutch loss-of-life model, they find a 13% overestimate for the 100 m vs. 25 m resolution model compared with a 65% overestimate when the "capacity for escape" parameter is excluded from the vulnerability model. However, the magnitude of scale transfer considered by such uncertainty quantification studies is generally less than the magnitudes implemented in recent global models (Ward et al., 2020).

While research on various scaling aspects has improved flood risk modelling, our understanding of

scale transfers remains limited, specifically the systematic overestimation bias observed in coarse models. Computational limits often necessitate relatively coarse resolutions for macro-scale inundation modelling, motivating the downscaling of outputs using high-resolution DEMs (Sampson et al., 2015; Bates et al., 2021; Winsemius et al., 2016). Implicit rescaling frequently occurs during damage calculations (Figure 1.2), for instance, when spatially detailed hazard indicators must be aggregated to match the resolution of coarse exposure data (Seifert et al., 2010a; Sieg and Thieken, 2022; Thieken et al., 2016; Jongman et al., 2012). This mismatch between component resolutions creates further scaling issues, for example when applying asset-level vulnerability models to aggregated exposure data (Hall et al., 2005; Ward et al., 2020; Sairam et al., 2021). The effect of these three rescaling operations on model behaviour has never been explored systematically and is the focus of this dissertation.

1.2 Research objectives and overview

Two observations motivate this dissertation: 1) policy makers increasing reliance on coarse-scale flood risk models, and 2) the gap in understanding of the impacts and application of scale transformations within these models. In response, this dissertation seeks to illuminate and understand the often "hidden" scale transfers inherent to modern macro-scale flood risk modelling. To accomplish this, this dissertation is guided by the overarching question:

> How do the aggregations and disaggregations currently employed in flood risk models affect their behaviour?

To explore specific transformations and the model behaviours they incite, the four core chapters summarized in Figure 1.3 are structured around the following sub-questions:

The first question focuses on the disaggregation or downscaling of flood hazards from model to consumer scale:

Research Question: Chapter 2

How does enhancing the resolution of flood hazard data improve accuracy?

This question investigates an overlooked component integrated in some macro-scale flood risk model chains where, following the hydrodynamic model, inundation grids are downscaled to a fine resolution, typically to create a more realistic flood map for consumers.

The second question addresses the aggregation of hazards and the subsequent effects on exposure:

Research Question: Chapter 3

How does the aggregation of flood hazards affect model behaviour?

This question investigates the implicit aggregations common in risk calculations (see Figure 1.2) necessary to intersect hazard grids with coarse exposure models.

The third question focuses on vulnerability models or damage functions and their application to aggregated exposure models:

Research Question: Chapter 4 and 5

How does the aggregation of damage functions affect model behaviour?

This addresses risk models where per-asset damage functions are used in conjunction with aggregate exposure models, a common feature of macro-scale models.

Together, these sub-questions investigate potential model behaviour changes across the hazard, exposure, and vulnerability components of risk models (Figure 1.3). These sub-questions are addressed sequentially in the following four chapters. The sixth and final chapter integrates these findings, showing how collectively the findings address the overarching research question.

This dissertation examines the use of aggregation and disaggregation techniques employed in previous studies, many of which were conducted by the author or co-authors. Rather than solely relying on simulations, which can be difficult to generalize; Chapters 3, 4, and 5 offer mathematical explanations and proofs to explain the overestimation bias reported by others for coarse models. Chapters 4 and 5 both investigate damage function aggregation, but with different emphasis on breadth vs. specificity. Chapter 5 employs a more theory-focused approach to estimate aggregation bias on 344 damage functions; while Chapter 4 employs a Germany-wide flood hazard model in a simulation experiment to quantify bias from four damage functions.

The core chapters leverage three case studies to demonstrate the mathematical proofs and novel methods: the 2021 Ahr flood in Germany (Chapter 2), the 2018 Saint John River flood in Canada (Chapter 3), and a global flood hazard model provided by others for all of Germany (Chapter 4). A theme throughout this dissertation is GIS operations commonly employed in risk model studies but rarely discussed — let alone evaluated systematically. To accomplish this, I employ modern hydrodynamic models, high performance computing, data-rich case studies, and in Chapter 4 and 5 an extensive global library of common vulnerability functions. In this way, this dissertation lies at the intersection of GIS and flood risk modelling.

The findings of this dissertation are primarily intended to inform methodological advancements in flood risk modelling, rather than directly evaluate flood risk or guide policy decisions. My aim is to provide modellers with insights and potential performance improvements, ultimately contributing to the field's progress and societal resilience to extreme flooding.

1.3 Author contributions

Chapters 2, 3, 4, and 5 of this dissertation present research reproduced from articles published, or under review in international peer-reviewed journals. To promote knowledge dissemination and reuse, these articles, along with associated code and data, are available under open-access and open-source licenses. As of February 2024, Chapters 2 and 3 have been published, Chapter 4 has been accepted and is in the typesetting stage, while Chapter 5 is in the second round of review. All work within these articles and this dissertation is authored by Seth Bryant with contributions summarized in Table 1.1 while the *Zusammenfassung* was graciously translated by Matthias Kemter.

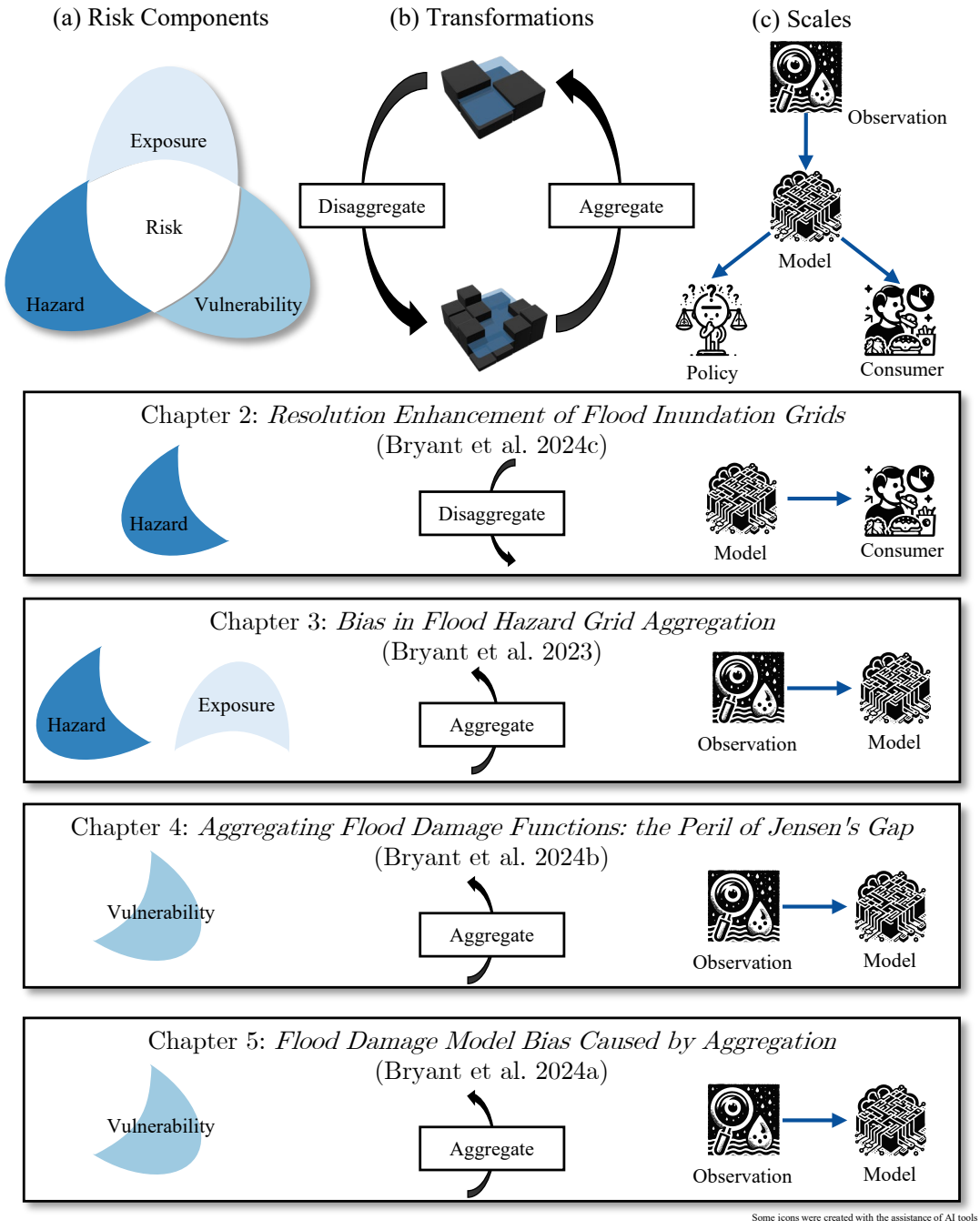


Figure 1.3: Graphical outline of dissertation showing the three primary dimensions of inquiry arranged as columns: a) components of risk (adapted from Field et al. (2014)); b) direction of transformation; and c) source and destination scales.

Table 1.1: Contribution summary for core chapters following CRediT (Contributor Roles Taxonomy) (NISO CRediT Working Group, 2022)

	Chapter 2	**Chapter 3**	**Chapter 4**	**Chapter 5**
Role	Bryant et al. (2024c)	Bryant et al. (2023)	Bryant et al. (2024b)	Bryant et al. (2024a)
Conceptualization	**SB**	**SB**	**SB**	**SB**
Data Curation	**SB**, HA	**SB**, HA	**SB**	**SB**
Formal Analysis	**SB**	**SB**	**SB**	**SB**
Investigation	n/a	n/a	n/a	n/a
Methodology	**SB**	**SB**	**SB**	**SB**
Software	**SB**, GS, HA	**SB**, HA	**SB**	**SB**
Validation	**SB**	**SB**	**SB**	**SB**
Funding Acquisition	BM, HK	BM, HK	BM, HK	BM, HK
Resources	n/a	n/a	n/a	n/a
Writing - Original Draft	**SB**	**SB**	**SB**	**SB**
Writing - Review & Editing	All	All	All	All
Visualization	**SB**	**SB**	**SB**	**SB**
Supervision	BM, HK	BM, HK	BM, HK	BM, HK
Project Administration	n/a	n/a	n/a	n/a

Key: **SB** - Seth Bryant, HA - Heiko Apel, GS - Guy Schumann, BM - Bruno Merz, HK - Heidi Kreibich

Chapter 2

RESOLUTION ENHANCEMENT OF FLOOD INUNDATION GRIDS

Bryant, S., Schumann, G., Apel, H., Kreibich, H., and Merz, B.: *Technical Note: Resolution enhancement of flood inundation grids*, Hydrology and Earth System Sciences, 28, 575–588, https://doi.org/10.5194/hess-28-575-2024, 2024.

Abstract

High-resolution flood maps are needed for more effective flood risk assessment and management. Producing these directly with hydrodynamic models is slow and computationally prohibitive at large scales. Here we demonstrate a new algorithm for post-processing low-resolution inundation layers by using high-resolution terrain models to disaggregate or downscale. The new algorithm is roughly eight times faster than state-of-the-art algorithms and shows a slight improvement in accuracy when evaluated against observations of a recent flood using standard performance metrics. Qualitatively, the algorithm generates more physically coherent flood maps in some hydraulically challenging regions compared to the state-of-the-art. The algorithm developed here is open source and can be applied in conjunction with a low-resolution hydrodynamic model and a high-resolution DEM to rapidly produce high-resolution inundation maps. For example, in our case study with a river reach of 20km, the proposed algorithm generated a 4m resolution inundation map from 32m hydrodynamic model outputs in 33seconds, compared to a 4m hydrodynamic model runtime of 34minutes. This 60-fold improvement in runtime is associated with a 25% increase in RMSE when compared against the 4m hydrodynamic model results and observations of a recent flood. Substituting downscaling into flood risk model chains for high-resolution modelling has the potential to drastically improve the efficiency of inundation map production and increase the lead time of impact-based forecasts, helping more at-risk communities prepare for and mitigate flood damages.

2.1 Introduction

Over the past decade, there has been significant progress in the development and implementation of flood models for large continental river basins and at the global scale. This is due to several factors including the rise in flood-related disaster damages, advancements in computing, and the availability and quality of global datasets (Ward et al., 2020; Nones and Caviedes-Voullième, 2020). As these models underpin risk management activities, ranging from early warning to land-use planning to disaster response, their accuracy and efficiency are important considerations for improving disaster resilience (de Moel et al., 2009).

Because of the computational demands of hydrodynamic models, resolution has been extensively studied and found to be one of the parameters of most importance for accuracy (Horritt and Bates, 2001; Fewtrell et al., 2008; Savage et al., 2016; Papaioannou et al., 2016; Alipour et al., 2022) with most finding inundation area and flood depth overestimated at coarser resolutions (Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021; Banks et al., 2015). In a study comparing fine and coarse models with resolution ranging from 1m to 50m and identical roughness, Muthusamy et al. (2021) used separate resolutions for the channel and floodplain. They found an overestimate in water depths and attributed it to the poorly defined coarse river channel (e.g., thalweg depth underestimated or

steep bank misrepresentation) and a subsequent reduction in conveyance (Muthusamy et al., 2021).

There are three primary hazard grids included in most flood risk models: Water Depth or Water Surface Height (WSH), Water Surface Elevation (WSE), and the Ground Elevations (DEM) which can be related by $WSE = DEM + WSH$. Often, WSE or WSH grids are produced from a hydraulic analysis or some model structured on the DEM, leading to a natural pairing of resolution, datum, and domain (i.e., the real-world region associated with the model). For large-scale studies, the resolution (s) of the DEM is relatively coarse (30-100m), resulting from the process and data used to construct the terrain model, or from some post-process upscaling introduced to obtain the resolution desired by the hydraulic analysis (i.e., coarsening model resolution to reduce complexity and runtime). For this latter case, or any case where supplementary fine-resolution ($s1$) DEM grids are available, applications like flood damage modelling or impact-based forecasting may benefit from enhancing WSE grids through downscaling or disaggregation to obtain a finer-resolution without the need for expensive or unstable hydrodynamic modelling. Unlike super-resolution techniques, which seek a high-resolution image from a single low-resolution image (Dong et al., 2015), flood hazard grid downscaling is a well-posed problem that uses the high-resolution DEM_{s1} and simple hydraulic assumptions to seek the WSE_{s1} grid that would be generated by an otherwise equivalent high-resolution hydrodynamic model.

While many flood risk model studies maintain a single resolution throughout the analysis (Hall et al., 2005; Sairam et al., 2021), examples of both upscaling and downscaling hazard grids are common. Upscaling, where hazard model output grids are post-processed to coarsen resolution, is generally undertaken to facilitate intersection with some exposure data, which is generally the most coarse data grid in flood risk model chains. This upscaling is achieved either through simple averaging (Seifert et al., 2010a; Sieg and Thieken, 2022) or some unspecified method (Thieken et al., 2016; Jongman et al., 2012). Examples of downscaling in the literature may employ it to reverse some earlier coarsening which was applied to improve hydrodynamic model stability or efficiency (Schumann et al., 2014; Sampson et al., 2015) or enhance some remote-sensing derived inundation product (Fluet-Chouinard et al., 2015; Aires et al., 2017).

In the first and only study (we are aware of) to investigate downscaling 2D calibrated hydrodynamic models, Schumann et al. (2014) developed a method using a nearest-neighbour search with $N_4(P)$ adjacency (querying values from only the N, S, E, and W adjacent or neighbouring cells, rather than $N_8(P)$ which queries all eight neighbouring cells) and a search radius of half the coarse-resolution. The researchers tested their algorithm using two models: a fine model with a 30m resolution and a coarse model with a 600m resolution, each calibrated separately. When comparing the downscaled grid to the results of the fine hydrodynamic model, they discovered an overestimate in water levels and negligible differences in volume. This method substantially improved computation times (compared to hydrodynamic modelling) and provided the basis for some large-scale flood models (Sampson et al., 2015; Bates et al., 2021). The CaMa-Flood project (Yamazaki et al., 2011) has developed a FORTRAN script with a similar algorithm to downscale results of their global river model; however, this script has not been described in any publication we are aware of. With the objective of operationalizing a 2D hydrodynamic model, Fraehr et al. (2023) developed a modelling framework that integrates a Guassian Process learning model with a low-fidelity hydrodynamic model to yield high-resolution depth and inundation estimates.

As part of their work to enhance the VIIRS (Visible Infrared Imaging Radiometer Suite) 375m resolution near real-time global flood inundation product, Li et al. (2022) developed a seven-stage downscaling

and correction pipeline. Leveraging global datasets for tree cover, land cover, permanent water bodies, and river networks, VIIRS water fractions were first converted to water levels then corrected using simple hydraulic assumptions. These 375m resolution water level grids were downscaled and converted to depths by intersecting with a global 30m DEM using a two stage algorithm. In the first stage, a nearest-neighbour search is employed with $N_4(P)$ adjacency and a search radius of one fine pixel starting with the lowest elevation pixel. A similar process is applied to the dry cells in the second stage. This work demonstrates a useful application of downscaling to generate finer-resolution flood-related earth observation data; however, the method is not directly applicable to enhancing coarse-resolution water grids produced through hydrodynamic modelling because it relies on coarse global data.

While downscaling flood grids is used by many global hazard models, to our knowledge only one study has addressed 2D downscaling of hydrodynamic model results (Schumann et al., 2014) and no studies have provided a methods comparison. Addressing this, our objectives are twofold: 1) present our newly developed downscaling approach; and 2) evaluate and compare our new approach to the state of the art downscaling approach of Schumann et al. (2014) and two simple algorithms using a data-rich case study.

2.2 Resample Case Framework

To better communicate and understand the challenges and solutions to rescaling flood hazard grids, we adapt the *Resample Case Framework* from Bryant et al. (2023) to classify each cell in the $s2$ domain into one of four cases with similar disaggregation behaviour. Each case is defined by comparing the local coarse water depth value ($WSH_{s2,j}$) to the corresponding fine values ($WSH_{s1,i}$) where cell j is composed of a block of i cells as shown graphically in Fig. 2.1 and defined explicitly as:

$$case_j = \begin{cases} \text{dry-dry (DD)} & \text{if } max(WSH_{s1,i}) = 0 \\ \text{dry-partial (DP)} & \text{if } WSH_{s2,j} = 0 \text{ and } max(WSH_{s1,i}) > 0 \\ \text{wet-partial (WP)} & \text{if } WSH_{s2,j} > 0 \text{ and } max(WSH_{s1,i}) > 0 \\ \text{wet-wet (WW)} & \text{if } min(WSH_{s1,i}) > 0 \end{cases} \tag{2.1}$$

where the first part of the $case_j$ label code is determined by the coarse cell (WSH_{s2}), and the second letter by the extremes of the fine cells (WSH_{s1}). The quadrants in Fig. 2.1 Panel (b) provide a simple example of four such groups whose corresponding case labels are shown on Panel (a). Because domain resample case classification is dependent on both input and output grids, classification is not directly used in any downscaling algorithms – instead, we use the framework to communicate the process and challenges of downscaling flood hazard grids.

Beginning with the simplest case, dry-dry (DD) zones are trivial and can be ignored for flood hazard rescaling operations as they remain unaltered: dry before and after rescaling (i.e., $WSH = 0$). Wet-wet (WW) zones are also relatively simple as the group of WSE_i cells should be roughly equivalent to their parent WSE_j cell; an easy task for classic grid resampling tools like bilinear resampling. Wet-partial (WP) zones can be similarly obtained, but with the additional step of removing dry $s1$ cells with an exceedance mask (i.e., cells where $WSE_{s1} < DEM_{s1}$ are set to null). Most difficult is the treatment of dry-partial (DP) zones which require some propagation or searching beyond the original coarse ($s2$) inundation footprint. In fact, this propagation challenge is a similar problem to that of a classic 2D hydrodynamic model; however, downscaling must employ simplifying assumptions to maintain advantageous computation times. This requires more sophisticated algorithms to propagate wet cells laterally like horizontal projection. Such horizontal projection may introduce artifacts like isolated flooding where disconnected groups of cells

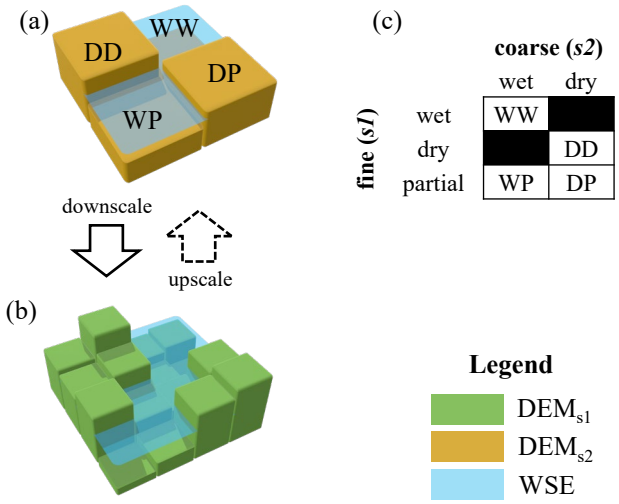


Figure 2.1: Framework for classification of flood hazard resample case. Panel (a) shows conceptual coarse grids and the corresponding resample case calculated from Eq. 2.2. Panel (b) shows the corresponding fine grids while Panel (c) shows the case label acronyms. D, W, and P stand for *dry*, *wet*, and *partial*, respectively.

in low-lying terrain are shown as flooded (e.g., behind levees). Because high-resolution hydrodynamic models generally employ cell-connected routines, this isolated flooding can be considered erroneous within the paradigm of hydrodynamic downscaling. However, this paradigm can be a poor representation of actual flood behaviour in areas with high groundwater connectivity or imperfect flood defences. In summary, the downscaling problem can be broken into three zones: the first two (WW and WP) are relatively simple and, while some alternate approaches are possible, we do not expect large differences in performance in these zones. The final dry-partial (DP) zone is more challenging and we therefore expect differences in treatment and performance between algorithms in this zone.

2.3 Methods

To validate and compare the novel resolution enhancement or downscaling algorithm, we first compute the requisite coarse-resolution ($s2 = 32m$) input grid (WSE_{s2}) using a calibrated hydrodynamic model. Using this input grid (WSE_{s2}) and a fine-resolution terrain layer (DEM_{s1}) we apply the novel downscaling algorithm to compute a fine-resolution enhanced grid ($WSE_{s1=4m}$). Using the same inputs, we then compute similar enhanced grids for the state of the art downscaling algorithm from Schumann et al. (2014) and two simple algorithms representing solutions of minimum complexity. These enhanced grids (WSE_{s1}), along with results from a calibrated fine-resolution hydrodynamic model ($s1 = 4m$), are then evaluated against the maximum inundation extents and high water marks observed during a 2021 flood in Germany to validate and demonstrate the performance improvements of the novel algorithm.

2.3.1 Novel CostGrow Algorithm

The novel *CostGrow* algorithm employs the four phases summarized in Fig. 2.2: 1) grid resampling; 2) least-cost mapping; 3) filtering high-and-dry sub-grid cells; and finally 4) an isolated-cell filter; all of which are parameterless. In the first grid resampling phase, various techniques have been developed by

others for applications in image analysis and spatial analysis (Bierkens et al., 2000) with bilinear being the most common for terrain manipulations in hydraulic applications (Heritage et al., 2009; Muthusamy et al., 2021) as it provides a smooth result while preserving centroid values. For downscaling, bilinear resampling computes the $s1$ value from the four adjacent $s2$ centroid values weighted by distance as seen in Fig. 2.2a (notice the $s2$ values are preserved by the center $s1$ cell). *CostGrow* implements bilinear resampling from the popular spatial analysis package GDAL (GDAL/OGR contributors, 2022). In the second phase, the resampled grid is extrapolated using a cost-distance analysis, a common GIS algorithm for computing the path of least cost, determined by weighting distance and some cost map to obtain the *effective distance* from source cells to sink cells Foltête et al. (2008). For this study *CostGrow* implements a cost-distance routine with $N_8(P)$ adjacency and a neutral cost surface (Lindsay, 2014, CostAllocation). This first maps the dry portion of the domain in terms of catchment areas for each boundary cell $WSE_{s1,i}$ from the previous phase, then maps the corresponding boundary $WSE_{s1,i}$ cell value to each of its catchments. In effect, this grows or horizontally projects each $WSE_{s1,i}$ boundary cell value outwards, filling the dry domain with the $WSE_{s1,i}$ values that are closest in distance. For the toy example shown in Fig. 2.2b, this is a simple extrapolation onto the dry right-hand side of the domain. Future implementations could employ a non-neutral cost surface to incorporate levees or some other flood obstructions into the analysis. In the third phase, high-and-dry cells are filtered from this cost-distance map by comparing cell-by-cell to the terrain values (where $DEM_{s1} > WSE_{s1,i}$ set $WSE_{s1,i} = NULL$) as shown by the blank cells in Fig. 2.2c. This often results in many isolated pockets of flooding in low-lying areas shown beyond the initial contiguous WSE_{s2} flood (see Fig. 2.2c red circle). In the final phase, these isolated or disconnected groups of flooded cells are filtered from the result such that only the largest or main flooded water body remains. To accomplish this, the filtered grid is converted to a binary inundation grid, from which each contiguous clump is identified and ranked according to size (Lindsay, 2014, Clump) (see Fig. 2.2d). From the largest clump, an inverted mask is generated and applied to the water level grid to remove isolated flooding cells from the result.

2.3.2 Validation and Comparison

Case Study, Data, and Hydrodynamic Modelling

To evaluate the aforementioned downscaling algorithms, data obtained from the July 2021 flooding of the Ahr River in Germany is used. This was the most extreme flood event to hit the region in living memory, with precipitation exceeding a 500-year return period (Dietze et al., 2022), a difficult to estimate peak discharge (Vorogushyn et al., 2022), and 134 casualties in the Ahr valley (Szönyi M. and Roezer V., 2022). The data used for this study is summarized in Table 2.1 and Fig. 2.3.

To construct the coarse water grid for use as an input in downscaling (WSE_{s2}) and a second grid for validation and comparison (WSE_{s1}), coarse ($s2 = 32m$) and fine ($s1 = 4m$) resolution twin hydrodynamic models are calibrated to the observed inundation extents using the Critical Success Index. The hydrodynamic models are constructed in the 2D raster-based *RIM2D* framework (Apel, 2023) and run on a Tesla P100 GPU. RIM2D implements a simplified version of the shallow water equations after Bates et al. (2010). A reconstructed hydrograph is used for the upstream boundary condition and other model parameters are described in Apel et al. (2022). The model terrain is generated through bilinear resampling of the bare earth DEM (Table 2.1) to the target resolution. For the treatment of urban areas, blocking-out buildings has been shown to be a more accurate way to represent buildings within a relatively high-resolution hydrodynamic model (Bellos and Tsakiris, 2015). In coarse models however,

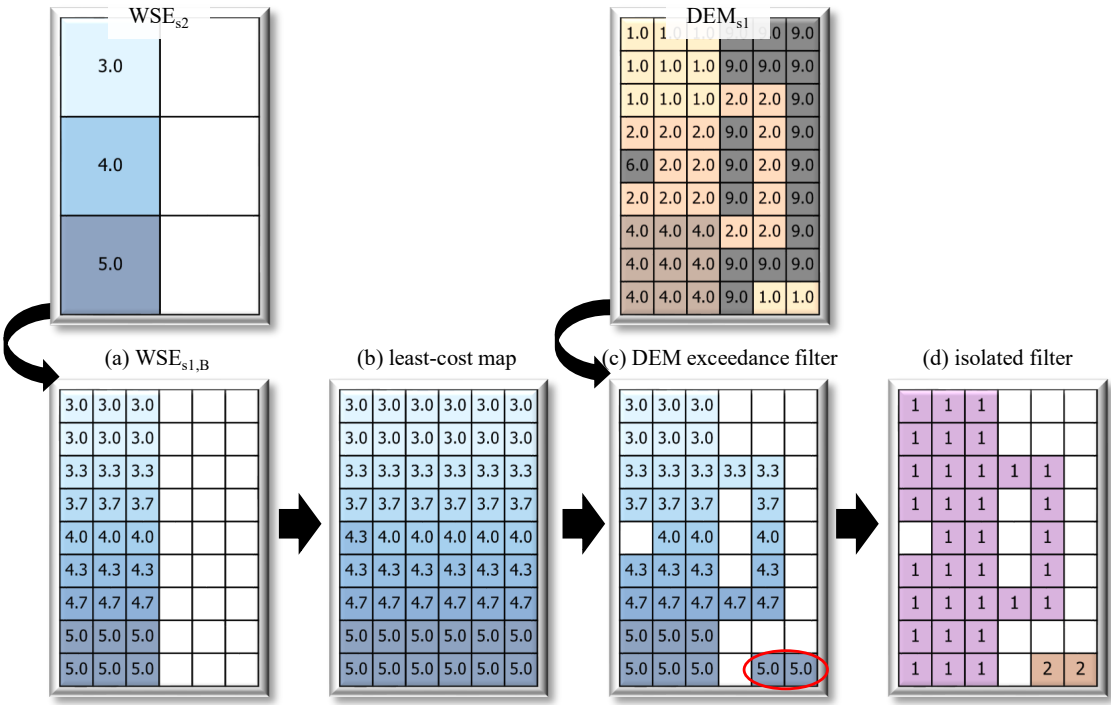


Figure 2.2: Toy example of the novel *CostGrow* downscaling algorithm showing WSE, DEM, and clump analysis grids for the four phases: a) bilinear resampling of coarse water level WSE_{s2} grid; b) extrapolation using least-cost mapping from $WSE_{s1,B}$ values; c) application of DEM exceedance mask to filter high-and-dry cells; and d) results of clump analysis, from which the isolated mask is generated to filter all but the largest group (1 in this case). See text for details.

especially where the building size (and space between) is smaller than a grid-cell, blocking-out reduces model performance, especially in and around buildings. Regardless, blocking-out could be included into a downscaling algorithm (and the high-resolution validation model); however, as this requires a more complicated algorithm and is not included in the current state of the art methods against which we compare, we opted to avoid blocking-out and instead apply a separate roughness coefficient to built-up areas in both models (and all downscaling algorithms) to capture the blocking effects of buildings (a.k.a. the "urban porosity approach").

To obtain accurate maximum WSE grids from the twin hydrodynamic models, a calibration routine is used to optimize model roughness using the Critical Success Index (CSI) (see Table A.2 for definition) of the maximum simulated inundation calculated against the observed inundation from Table 2.1. Two unique Manning's roughness values (built-up and channel/floodplain) are treated as free parameters for each model and optimized while a third roughness value for forested areas is held fixed ($n = 0.2\frac{s}{m^{1/3}}$) as Apel et al. (2022) showed this third region to have negligible influence. The three roughness values are spatially allocated according to land cover (Table 2.1) as described in Apel et al. (2022). Finally, the optimal roughness values are obtained using a mix of trial-and-error and the Newton-Conjugate Gradient algorithm (Nocedal and Wright, 2006; Virtanen et al., 2020) to optimize the Critical Success Index (CSI).

The best performing effective roughness values for the twin hydrodynamic models $s2 = 32m$ and $s1 = 4m$ is 0.867 and 0.175 $\frac{s}{m^{1/3}}$ for urban areas and 0.089 and 0.133 $\frac{s}{m^{1/3}}$ for channel areas with a CSI of 0.885

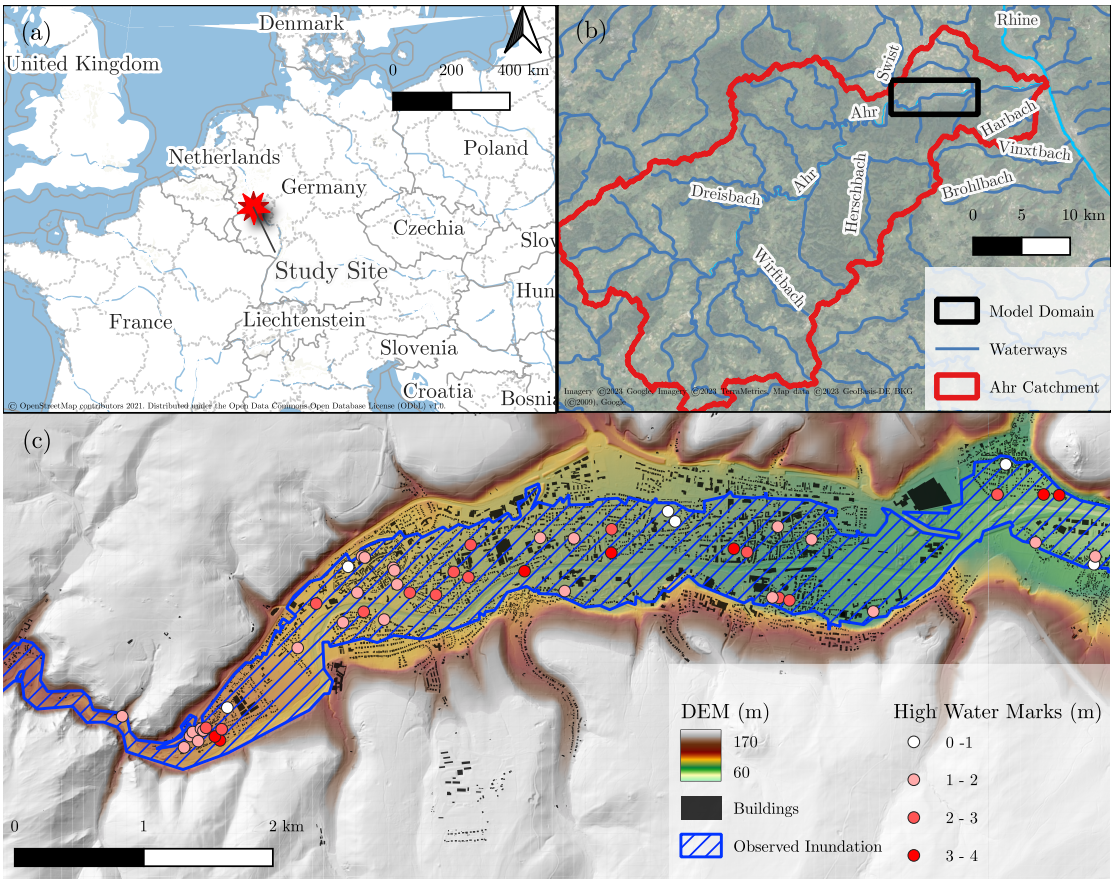


Figure 2.3: Study site maps showing: (a) location map; (b) Ahr catchment map; and (c) downscaling domain with main datasets (see Table 2.1 for descriptions).

and 0.914 respectively as shown on Fig. A.1 and A.2. These counterintuitive relative roughnesses are a result of differences in floodplain-channel dynamics necessary to match the observed inundation footprint between the two models. In the coarse hydrodynamic model, the river channel is poorly represented by the 32m resolution which is roughly 3 times larger than the channel. Thus, the flow in the channel, as well as the channel-floodplain interactions, show different dynamics compared to the more realistic fine-resolution model. The calibration routine compensates for these differences with the disproportionate roughness values reported above. However, as our focus is on downscaling performance, the less-accurate representation of channel dynamics and water levels (as opposed to inundation extents) provided by the coarse model are inconsequential considering we apply the fine and coarse hydrodynamic model results comparatively in all scenarios. Additional figures, performance measures, and discussion for the calibration are provided in Section A. To remove any boundary effects, the hydrodynamic model results are cropped to a smaller domain for the downscaling analysis (13.4 x 6.6 km to 8.9 x 3.5 km; see Fig. A.3).

Downscaling Algorithms for Comparison

To demonstrate the performance of the novel *CostGrow* algorithm relative to similar algorithms, two simple algorithms representing solutions of minimum complexity and the state of the art from Schumann

Table 2.1: Summary of data used in case study

Type	Metadata	Reference
DEM	0.5 m resolution bare earth DEM created from aerial LiDAR survey from September 22 to October 24, 2021 in twelve sessions with a RIEGL scanner LMS-VQ780i with 20 points/m^2 achieved	(Milan Geoservice GmbH, 2023)
High water marks	75 high water marks at buildings reported by residents	(Apel et al., 2022)
Inflow hydrograph	30 hour hydrograph at Altenahr gauge with maximum depth of 10.2 m reconstructed by Landesamt für Umwelt Rheinland-Pfalz	(Apel et al., 2022)
Building locations	Building footprint polygons downloaded from OSM on 2022-11-14	(OpenStreetMap contributors, 2022)
Observed inundation	Polygon of maximum flood extents compiled from an aerial survey on July 16th and 20th and a second survey on July 24th and 29th	(Landesamt für Umwelt Rheinland-Pfalz, 2022)
Land cover	Gridded land cover inventory reflecting 2017-2018 conditions and updated in 2020	(Copernicus Land Monitoring Service, 2018)

et al. (2014) are described below and included in the comparison.

The first simple algorithm considered here is a bilinear grid resampling (see *Resample* in Fig. 2.4c) which is identical to the above described first phase of the novel *CostGrow* algorithm. This algorithm is the only one considered that does not make use of the fine-resolution terrain values (DEM_{s1}) and therefore carries an obvious limitation in wet-partial (WP) regions, where sub-grid high-and-dry ground elevations may be present within a wet coarse cell (see red circle in Fig. 2.4a). To address this, the second simple algorithm we consider (see *TerrainFilter* in Fig. 2.4d) builds and applies a terrain exceedance mask ($WSE_{s2} < DEM_{s2}$) which removes those cells where depths are negative from the resulting WSE_{s1}. Neither of these simple algorithms treat cells outside the wet coarse domain (i.e., the dry-partial (DP) zone remains dry).

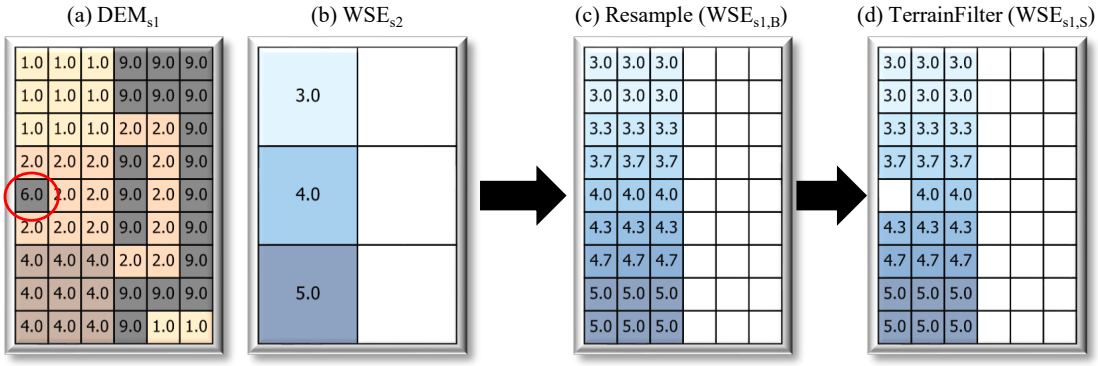


Figure 2.4: Toy example of simple flood downscaling inputs and algorithms showing: a) input fine-resolution terrain grid; b) input coarse-resolution water level grid; c) *Resample* downscaling result; and d) *TerrainFilter* downscaling result described in the text.

To the best of our knowledge, Schumann et al. (2014)'s method (*Schumann14*) is the state of the art in 2D flood grid downscaling algorithms. This algorithm was developed to downscale 1D/2D hybrid

inundation model results from 600m to 30m by employing a two tier approach: first, the 1D channel regions are downscaled assuming a water surface plane between sections; second, 2D floodplain regions are downscaled using a nearest-neighbour search. For our study, we focus on the floodplain portion of the algorithm for which the source code was provided to the study team and which has roughly the three steps shown in Fig. 2.5. First, a search zone is built using a buffer of width one half the coarse-resolution around all wet cells in the coarse domain (i.e., wet-partial (WP) and wet-wet (WW) regions). An alternate buffer distance parameter is possible, but here we select the same parameter value as Schumann et al. (2014). Second, within the search zone, each fine ($s1$) cell searches for the nearest coarse ($s2$) cell using a nearest neighbour *city-block* (also called *Manhattan*) search algorithm which replicates the same $N_4(P)$ adjacency used by their inundation model. Finally, the WSE_{s1} nearest-neighbour search result is combined with a simple grid resample (also using nearest-neighbour) and a terrain exceedance mask is applied. While this algorithm improves upon the simple approaches, the blocky WSE_{s2} values remain in the fine WSE_{s1} result, isolated flooding artifacts are introduced (red circle in Fig. 2.5d), and the per-cell nearest-neighbour search is computationally expensive. Finally, the algorithm was originally written in the *MATLAB* programming language and was not made public or widely shared.

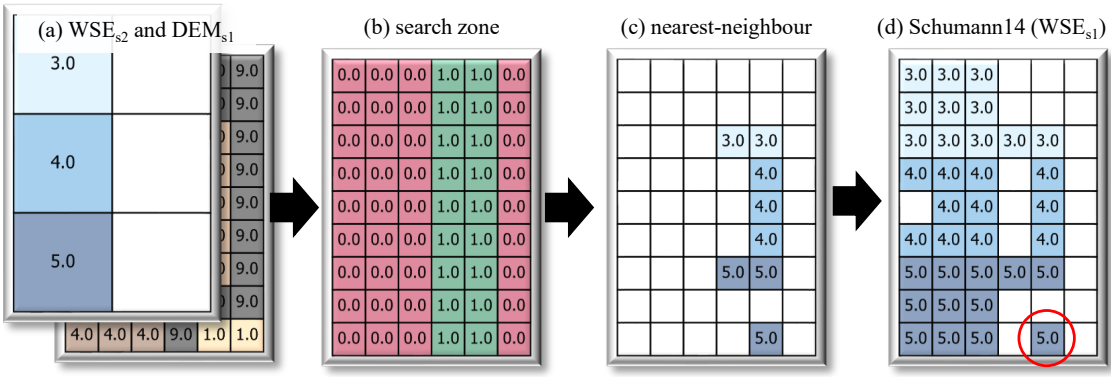


Figure 2.5: Toy example of Schumann et al. (2014)'s floodplain downscaling algorithm showing: a) input fine-resolution terrain grid and coarse-resolution water level grid (see Fig. 2.4a); b) search zone; c) nearest-neighbour search result; and d) WSE_{s1} downscaling result. See text for details.

Compared to the above *Schumann14* algorithm, we expect three main advantages from the novel *CostGrow*. First, *CostGrow* should be substantially more computationally efficient by avoiding the cell-by-cell nearest neighbour search. Second, *CostGrow* should be slightly more accurate in reproducing fine-resolution inundation extents by avoiding the fixed search radius and including an isolated flooding filter. Third, the accuracy of water levels should improve due to the incorporation of the aforementioned inundation extent mechanisms in dry-partial regions and the replacement of the initial nearest-neighbour resampling with bilinear resampling. These expectations are tested below using a case study. Table 2.2 provides a brief summary of all the downscaling algorithms considered by this study and the two hydrodynamic models used.

2.4 Analysis and Results

Water level grids obtained from the twin calibrated hydrodynamic models and the four downscaling algorithms and their corresponding inundation performance are shown in Fig. 2.6. For the models, Fig. 2.6a4 and b4 show the fine ($s1$) parameterization while the results for the *Resample* downscaling algorithm (Fig. 2.6a0 and b0) show the performance of the coarse ($s2$) parameterization (because this

Table 2.2: Downscaling algorithms and hydrodynamic models included in this evaluation showing resample $case_j$ applicability (see Fig. 2.1) and case study runtime to achieve output grids ($WSE_{s1=4m}$ for all except *Hydro. (s2)* which outputs $WSE_{s2=32m}$).

Name	$case_j$	Runtime (secs)	Method Description
Hydro. (s1)	n/a	1917.0	fine ($s1 = 4m$) hydrodynamic model
Hydro. (s2)	n/a	32.0	coarse ($s2 = 32m$) hydrodynamic model
CostGrow	WW, WP, DP	1.0	*TerrainFilter* plus cost-distance mapping and isolated-cell filtering
Resample	WW	0.1	simple bilinear grid resampling
TerrainFilter	WW,WP	0.2	*Resample* plus filtering of high-and-dry cells
Schumann14	WW, WP, DP	8.7	nearest-neighbour search and grid resampling from Schumann et al. (2014)

algorithm does not alter the coarse extents). This suggests both parameterizations reproduce the observed inundation well, with the fine ($s1$) obtaining a slightly better CSI as expected considering the complex topography. The coarse ($s2$) model converged on lower water levels and extents to obtain the optimal CSI, as shown by the lower Error Bias (0.205 vs. 1.029) and the calibration contour plots (Fig. A.1 and A.2).

Fig. 2.6 suggests there are some discrepancies between the observed inundation, which has a single contiguous inundation extents without holes, and the DEM which contains some micro-topography that likely would have remained dry during the flood (see Fig. 2.8 point A), suggesting the observed inundation is slightly conservative (i.e., overestimates the true flood extents). These discrepancies may be attributable to the methods used by Landesamt für Umwelt Rheinland-Pfalz (2022) to map the extents or the earthworks undertaken between the inundation mapping surveys (late July) and the LiDAR survey (October). Regardless, these discrepancies are relatively minor and we consider them negligible for our research objective of evaluating the *CostGrow* algorithm.

Comparing the inundation performance of the downscaling algorithms in general, the more complex algorithms performed better, with the novel *CostGrow* and *Schumann14* algorithms performing similarly. Specifically, the four common inundation performance metrics in Fig. 2.6 panels b0-b3 show that *CostGrow* and *Schumann14* have nearly identical performance, with all metrics (with one exception) out performing the simple algorithms *Resample* and *TerrainFilter*. The exception being that *TerrainFilter* has the lowest False Alarm rate (0.014), an artifact which can be explained by two concurrent hypothesis: 1) the *TerrainFilter* algorithm does not address dry-partial (DP) regions but does filter wet-partial (WP) regions, giving it the smallest inundation area and making it the least likely to overestimate inundation; and 2) the over-estimation bias in the observed inundation discussed earlier favours methods that underestimate.

The inundation metrics reported in Fig. 2.6 are sensitive to both the hydraulic character of the study region and the particular domain selected for analysis e.g., we expect a broad-flat floodplain or different boundary conditions to yield different metric values. For the results reported here, we selected the domain (8.9 x 3.5 km) by balancing hydraulic continuity, controlling for boundary effects, and computational cost; however, other similar domains were tested during study preparation and the relative ranking of performance between the downscaling algorithms was found to be consistent, with *CostGrow* outperforming *Schumann14* slightly for some domains. For example, the CSI of the detail area (Fig. 2.6 blue box) is 0.813 for *CostGrow* and 0.811 for *Schumann14*. This aligns with our expectation that *CostGrow*'s inundation results would slightly outperform that of *Schumann14* given the absence of a fixed search radius and use of an isolated filter in the *CostGrow* algorithm. Regardless, this evaluation suggests that, while gains in computation performance are substantial, gains in standard qualitative inundation performance

over *Schumann14* are negligible. Qualitatively however, *CostGrow* generates more physically coherent depth grids in some fringe areas as shown in Fig. 2.8 and A.4.

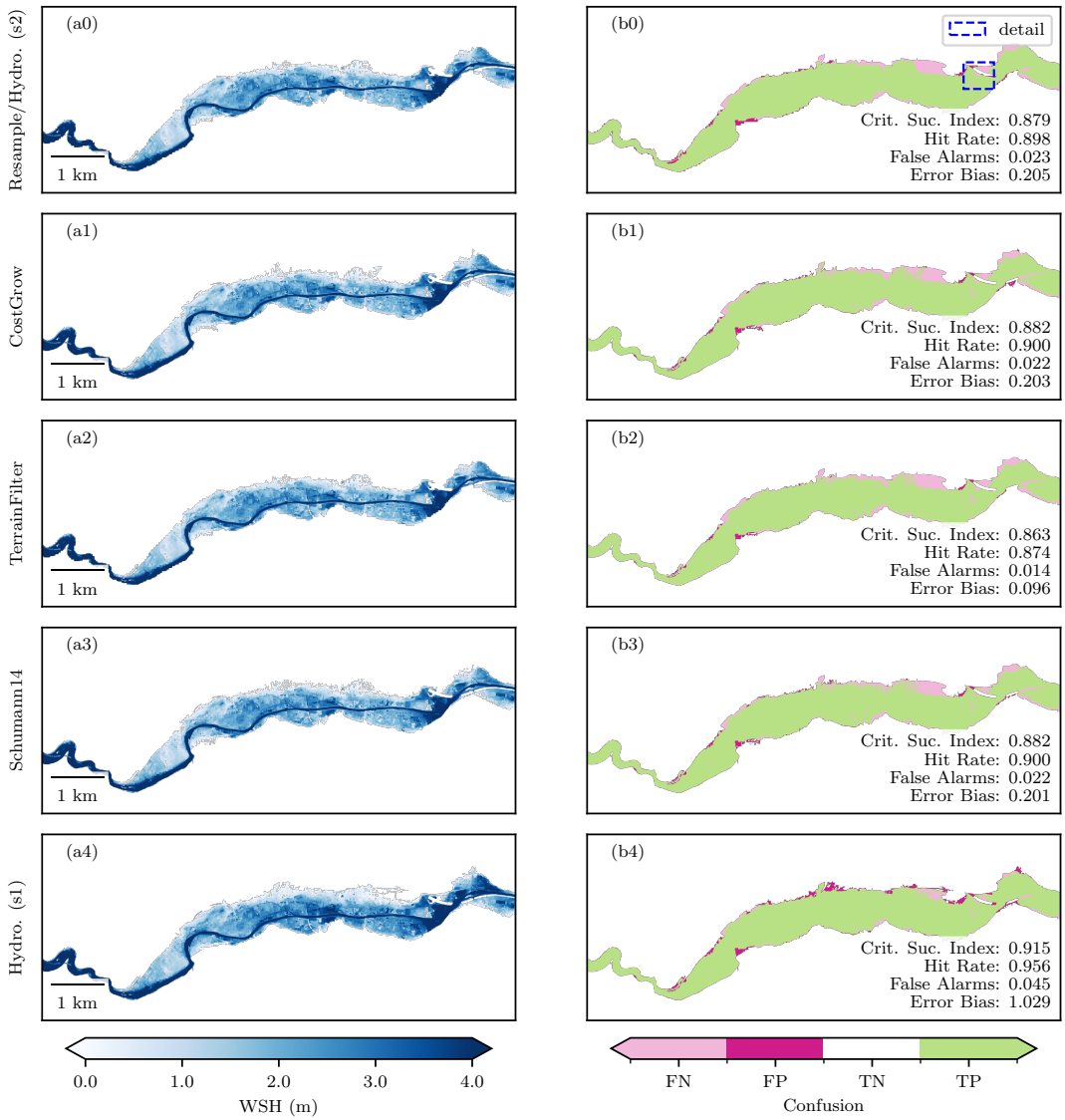


Figure 2.6: Downscaled and hydrodynamic model WSH_{s1} results and corresponding inundation performance. Left-side panels (a) show the water depth grids (WSH_{s1}) obtained from the coarse and fine hydrodynamic model (a0 and a4) and the four downscaling algorithms (a0-a3; see Table 2.2 and text for description). Similarly, the right-side panels (b) show the domain inundation classification confusion map (False Negative (FN), False Positive (FP), True Negative (TN), and True Positive (TP) cells – see Table A.1) and the common inundation performance metrics defined in Table A.2. The *Resample* algorithm and coarse hydrodynamic model results ($s2$) are identical and therefore shown as one panel. See Fig. 2.8 for detail area.

Examining the water level performance of the hydrodynamic models, Fig. 2.7 shows that the coarse ($s2 = 32m$; panel a0) and fine ($s1 = 4m$; panel c1) reproduce the observations well. This is remarkable considering the models were calibrated on inundation extents (using CSI), not water levels, and that the water levels are reported by residents. Similar to inundation performance, Fig. 2.7 also shows the fine model ($s1 = 4m$) performs best while the coarse ($s2 = 32m$) slightly underestimates (see Fig. A.3

for a map of WSE differences between $s2 = 32m$ and $s1 = 4m$). Fig. 2.7 also shows that, like for inundation performance, *CostGrow* and *Schumann14* have better performance than the simple algorithms and the coarse hydrodynamic model; however, the performance of *CostGrow* slightly surpasses that of *Schumann14*. Given the more comparable inundation performance, we conclude the advantage seen here emerges from *CostGrow's* application of bilinear resampling as opposed to *Schumann14's* nearest-neighbour resampling; however, owing to the relatively small scale ratios (4:32), this advantage is minor. Comparing the simple algorithms in Fig. 2.7 (panel c0 and a1) shows that the treatment of wet-partial regions provides no improvement in reproducing high water marks, unlike the advantages seen for inundation performance. We hypothesize this owes to the absence of any high water mark observations on dry cells in wet-partial regions. In other words, *TerrainFilter* only improves the filtering of False Positives when compared to the *Resample* result and False Positive regions can not be evaluated by high water mark observations (as these regions are dry in reality).

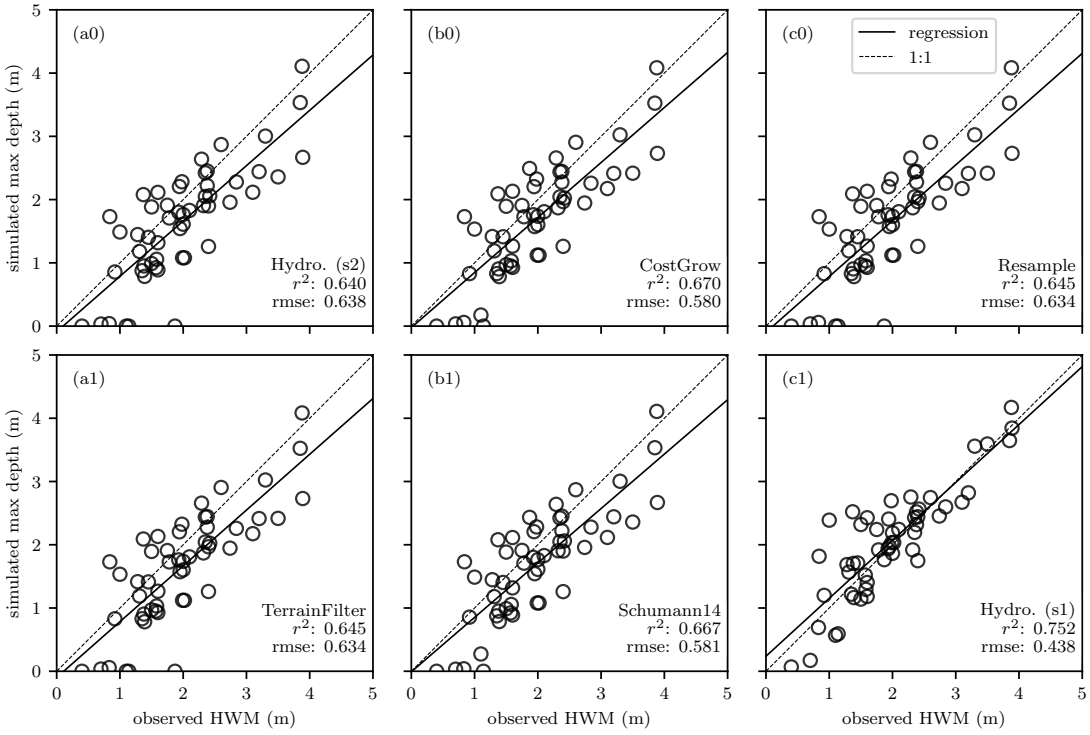


Figure 2.7: Linear correlation between July 2021 resident-reported high water marks (see Table 2.1) and maximum simulated depth of twin hydrodynamic models (a0, c1) and downscaling algorithms (b0, c0, a1, b1).

Focusing on a small region, Fig. 2.8 shows a portion of the domain where floodwaters likely flowed behind a highway embankment along a small frontage road travelling underneath an overpass (point E). Because the observed inundation layer was mapped primarily by air (Landesamt für Umwelt Rheinland-Pfalz, 2022), this observation data shows the area underneath the overpass as dry; therefore any simulated inundation in this area is marked False Positive (FP) — supporting our hypothesis that the observed inundation is slightly conservative. Other performance and behavioural differences between the algorithms can also be seen in this area: the lack of treatment for wet-partials (point A) and dry-partials (point B) in the simple algorithms; how *CostGrow* is not limited to a search radius like *Schumann14* for dry-partial

treatment (point C); the isolated inundation artifacts in *Schumann14* (point D); and the blocky result of *Schumann14's* nearest-neighbour resampling in wet-wet (WW) and wet-partial (WP) regions (panel b1).

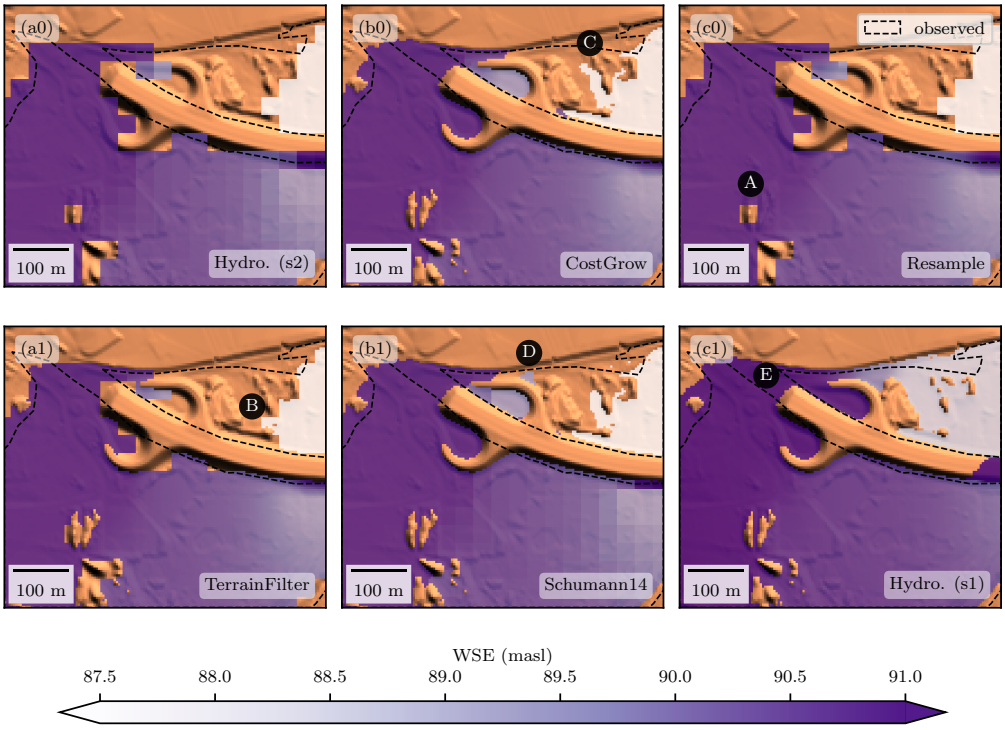


Figure 2.8: Downscale and hydrodynamic model WSE detail results. Observed inundation is shown in black for reference. See Fig. 2.6b0 for location and text for description.

Runtimes for the twin hydrodynamic models and the four downscaling algorithms are shown in Table 2.2. As expected, the algorithms of higher complexity also have higher runtimes and all downscaling algorithms are substantially faster than the hydrodynamic models (which simulate a flood wave of 833 mins). Despite minimal effort being invested in optimization, the novel *CostGrow* algorithm is substantially faster than the state of the art algorithm from Schumann et al. (2014). These runtimes may be improved through parallelization and other programming improvements. Regardless, because Schumann et al. (2014)'s algorithm employs a cell-by-cell nearest-neighbour search, this algorithm is fundamentally less efficient than those like *CostGrow* which employ least-cost mapping.

Comparing alternate approaches to obtain fine-resolution flood grids (4m in this case), the total runtime of a pipeline implementing the *CostGrow* downscaling algorithm (on top of the coarse hydrodynamic model) was roughly 33 seconds versus the 34 minutes necessary for our 4m native hydrodynamic model, a 60-fold improvement. This reduced runtime has a corresponding loss in temporal resolution (only the maximum WSE is downscaled) and inundation accuracy of 0.03 CSI and high water mark accuracy of 0.14 RMSE when comparing the *CostGrow* downscaling algorithm pipeline to the 4m native hydrodynamic model for our study. Were a slower hydrodynamic model used (e.g., a non GPU-parallelized platform) or a larger downscaling ratio (32:4 in our case) the efficiency gain of downscaling over fine hydrodynamic modelling would increase; while shortening the simulation time (833 mins in our case) would reduce the efficiency gain. For example, during study development we implemented similar twin hydrodynamic models in the LISFLOOD-FP 8.1 framework using a second-order discontinuous Galerkin solver which

implements the full shallow water equations (Shaw et al., 2021). Executed on 8 CPU cores the runtime for these models was 7100.00 mins and 0.43 mins for the 4m and 32m discretizations respectively — were a *CostGrow* downscaling algorithm pipeline implemented with this setup we estimate a 16,000-fold improvement in runtime to obtain a comparable $WSE_{s1=4m}$ grid.

The flood grid downscaling algorithms presented here make estimates based on simple hydraulic assumptions and the DEM. Because of this, these algorithms do not consider sub-grid or other hydraulically relevant elements not contained in the DEM like levees, flood-walls, or storm drainage systems. The significance of this limitation will depend on the particular case, but any study where levees or barriers are present but *not* resolved by the DEM should be extra cautious when employing downscaling. Much like hydrodynamic models, sub-grid obstructions like levees could be incorporated into downscaling algorithms, e.g., using a non-neutral cost surface in *CostGrow's* cost-distance routine.

Future work should explore evaluation techniques for flood-related algorithms, where the techniques are less sensitive to study area, domain size, hardware, and software of a particular study. For example, a collection of fully-open data-rich flood events with a wide range of hydraulic character would facilitate more meaningful comparisons of model or algorithm performance across platforms and between researchers. Further, such additional case studies with varied hydraulic character would help to quantify and communicate the benefits and limitations of downscaling in different hydraulic regimes.To better support the emerging needs of impact forecasting, where 2D-velocity grids are sometimes desired, pursuing a method that is also capable of downscaling velocity could be of use; however, this would require blocking-out buildings in the hydrodynamic model. Further performance enhancements of downscaling methods may be found by incorporating machine learning techniques originally developed for image enhancement which have recently been applied to enhance terrain models (Demiray et al., 2021).

2.5 Conclusions

This study has developed, demonstrated, and evaluated the novel *CostGrow* algorithm for resolution enhancement or downscaling of flood water surface grids. This algorithm outperforms the state of the art, with a six-fold improvement in runtime for our case study, a slight improvement in standard performance metrics, and improvements in some fringe areas using qualitative evaluation. When compared to results obtained through fine-resolution hydrodynamic modelling, the proposed downscaling algorithm (in conjunction with coarse-resolution modelling) showed a 60-fold improvement in runtimes with a slight loss of accuracy.

In general, coarse modelling in conjunction with downscaling is shown to be an effective means of obtaining fine-resolution inundation grids at a fraction of the computational cost. However, the utility of employing downscaling to obtain fine-resolution grids is limited by the availability and quality of fine-resolution DEMs. A potential application of downscaling is to facilitate the post-processing of fine-resolution inundation results layers from global models for data-rich regions where a local fine-resolution DEM is available. This could be a cost effective way to deliver fine-resolution inundation maps to any region on the globe without the need for specialized modelling expertise or resources. Regardless, the lack of attention to the subject of downscaling in the academic literature suggests some space remains for downscaling to improve the efficiency of model chains. Towards this, the downscaling algorithms developed in this study have been made open source and are available as QGIS processing scripts (`https://github.com/cefect/FloodRescaler`).

Acknowledgements

The research presented in this article was conducted within the research training group "Natural Hazards and Risks in a Changing World" (NatRiskChange) funded by the Deutsche Forschungsgemeinschaft (DFG; GRK 2043/2). We are thankful to Oliver Wing for his insightful conversations on the topic, Jeff Neal for help with early attempts to implement LISFLOOD-FP, Knut Günther for keeping the section running, Karen Lebek for her brilliant coordination, and the whole NatRiskChange community for support and friendship. And I am particularly indebted to my wife Jody for watching baby Tarn so I could finish this manuscript.

Chapter 3

BIAS IN FLOOD HAZARD GRID AGGREGATION

Bryant, S., Kreibich, H., and Merz, B.: *Bias in Flood Hazard Grid Aggregation*, Water Resources Research, 59, e2023WR035100, https://doi.org/10.1029/2023WR035100, 2023.

Abstract

Reducing flood risk through disaster planning and risk management requires accurate estimates of exposure, damage, casualties, and environmental impacts. Models can provide such information; however, computational or data constraints often lead to the construction of such models by aggregating high-resolution flood hazard grids to a coarser resolution, the effect of which is poorly understood. Through the application of a novel spatial classification framework, we derive closed-form solutions for the location (e.g., flood margins) and direction of bias from flood grid aggregation independent of any study region. These solutions show bias of some key metric will always be present in regions with marginal inundation; for example, inundation area will be positively biased when water depth grids are aggregated and volume will be negatively biased when water surface elevation grids are aggregated through averaging. In a separate computational analysis, we employ the same framework to a 2018 flood and successfully reproduce the findings of our study-region-independent derivation. Extending the investigation to the exposure of buildings, we find regions with marginal inundation are an order of magnitude more sensitive to aggregation errors, highlighting the importance of understanding such artifacts for flood risk modellers. Of the two aggregation routines considered, averaging water surface elevation grids better preserved flood depths at buildings than averaging of water depth grids. This work provides insight into, and recommendations for, aggregating grids used by flood risk models.

3.1 Introduction

With the increase in flood related disaster damages and the availability of global datasets, the pace of development and application of meso- and macro-scale flood risk models has increased in the past decade (Ward et al., 2020). These flood risk models are often conceptualized as a chain of sub-models for the flood hazard, exposure of assets, and vulnerability modelling; with each step adding uncertainty (de Moel and Aerts, 2011). Vulnerability modelling, the last step in the chain where variables describing the assets-at-risk and their flood exposure are related to estimate some flood loss or damage, is generally found to be the most uncertain component in micro- and meso-scale models (de Moel and Aerts, 2011; Jongman et al., 2012). These findings are supported by work comparing modelled damages to those observed during flood events, where large discrepancies are regularly found between different models and against observations (Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020). Further challenges are introduced when such models are applied at the macro-scale, where hazard, exposure, and vulnerability are treated with gridded data of resolutions from 100 to 1000m (Hall et al., 2005; Ward et al., 2015b; Sairam et al., 2021). This process collapses heterogeneities within a grid-cell (like variable flood depth) and poses poorly understood challenges to calculating the exposure of sub-grid assets like buildings.

The terminology of model scaling varies between authors. Here, we use *resolution* of a fine ($s1$) or coarse ($s2$) grid, measured in distance units (where $s1 < s2$), to describe the ground length of a single square

cell or pixel within a grid. We select this term to avoid confusion with the more generic *scale*, which may also refer to the domain extents (Bierkens et al., 2000; Degbelo and Kuhn, 2018). The term *support* is used by some authors to describe the number of pixels per unit of area (Bierkens et al., 2000) but is less well known in the flood literature. We exclusively use *fine* and *coarse* to describe relative differences in resolution to avoid the more ambiguous terms *increase/decrease* which are often confused with the number of pixels/unit rather than the dimension of a pixel.

Operations which transform data or model resolution between fine ($s1$) and coarse ($s2$) are commonly termed *rescaling*, with those that refine resolution called *disaggregating* and those that coarsen called *aggregating*. Alternate terms include *downscaling* and *upscaling* respectively (Bierkens et al., 2000); however, these are less common in the flood literature. Using the term *aggregating* highlights the underlying spatial computation: a new coarse ($s2$) grid is assembled from some sub-calculation (e.g., the mean statistic) applied to blocks or neighbours of fine ($s1$) cells aggregated together (disaggregating is this same process in reverse) resulting in a new discretization of the domain. When an averaging statistic (e.g., the mean) is used for the sub-calculation, this aggregation has the effect of removing extreme values and producing a coarse ($s2$) grid with less variance — or a smoothed grid. While this manuscript exclusively uses averaging statistics that produce a smoothed grid, we apply the more generic term *aggregation* as this better aligns with literature and emphasizes the operation under study rather than the result. When the rescaling sub-calculation yields differences between the original fine ($s1$) and new coarse ($s2$) grids with a non-zero mean, the aggregation routine can be said to impart bias or distortion onto the coarse ($s2$) grid.

Flood hazards are increasingly modelled with 2D grid-based hydrodynamic models or 1D/2D hybrid models, both implementing some simplification of the shallow water equations (Apel et al., 2009; Dimitriadis et al., 2016). For example, Bellos and Tsakiris (2015) study the performance of three different methods for representing buildings in a high resolution model relative to water level measurements from a physical experiment of a flash flood. Because of the computational demands of such models, resolution has been extensively studied and found to be one of the parameters of most importance for accuracy (Fewtrell et al., 2008; Savage et al., 2016; Papaioannou et al., 2016; Alipour et al., 2022). Focusing on the relationship between model resolution and inundation area, many studies of fluvial floods find a positive inundation area and flood depth bias at coarser resolutions (Banks et al., 2015; Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021; Xafoulis et al., 2023) while studies of urban flooding are less conclusive (Fewtrell et al., 2008). For the underlying terrain model grids or digital elevation models (DEM), the resampling method used to generate the coarse analogs is often found to be of little significance (Muthusamy et al., 2021; Saksena and Merwade, 2015) except at high resolutions when buildings are present in the fine DEM (Fewtrell et al., 2008). Comparing fine and coarse models with identical roughness, Muthusamy et al. (2021) use separate resolutions for the channel and floodplain to show that positive bias can be explained by the coarse river channel being poorly defined and a subsequent reduction in conveyance. While these studies provide valuable insight into the behaviour of coarse hydrodynamic models, their utility for practitioners is limited as the coarse models are uncalibrated in these studies (unlike models in practice). Further, the focus of such studies is on a coarse model's (in)ability to reproduce observed high water marks or match some reference model, not on the hazard variables (and their heterogeneity) at asset locations used in risk modelling. In other words, when such studies find high water marks are adequately reproduced by a model at some coarse resolution, this should not be interpreted as that same model adequately reproducing the exposure which

is sensitive to more than just water levels at high water marks.

Many studies investigate flood risk model parameter sensitivity (Metin et al., 2018; Jongman et al., 2012; Apel et al., 2009; Seifert et al., 2010b; Ghimire and Sharma, 2021), but few investigate sensitivity to resolution explicitly (Komolafe et al., 2015; Brussee et al., 2021; Pollack et al., 2022). However, by extracting results from this literature and comparing those candidate fine-coarse model pairs which differ only in the level of aggregation or resolution, a quantitative bias of flood damage from aggregation can be computed from a diverse set of flood risk model experiments. Table 3.1 provides such a comparison that includes all relevant studies (and study pairs) the authors are aware of. This shows a clear positive bias between aggregation and the reported total flood risk metric, albeit of different magnitudes; which is remarkable considering the diverse methods, data, and regions under study. While the positive bias of coarse hazard models is well studied (Saksena and Merwade, 2015; Muthusamy et al., 2021), the implications for risk models have not been explored systematically.

In one of the few studies to investigate risk model sensitivity to grid aggregation specifically, Komolafe et al. (2015) perform a simulation experiment with a model calibrated to the 1996 Ichinomiya river basin flood in Japan. Beginning with 50m gridded asset and flood depth layers, eight additional coarse-resolution models were constructed by aggregating with an unspecified method. Their results show that aggregating or upscaling depth grids introduces a slight positive bias, i.e. overestimating the water depth of the coarser grids. No mention of the aggregation routine is provided or explanation for the behaviour observed. Investigating the sensitivity of a flood mortality model to hydrodynamic model resolution, Brussee et al. (2021) compare a 5, 25, and 100m resolution 2D hydrodynamic model of a densely populated dike ring surrounded by three rivers in the Netherlands. Applying a constant breach width, they find higher discharge and associated mortality in the breach zone at the coarser resolutions and a mortality bias of +8%. Ghimire and Sharma (2021) provide a thorough sensitivity analysis of U.S. focused hazard and vulnerability modelling platforms. Along with testing a 1D and 2D hazard model framework and input data qualities, they investigated alternate DEM constructions with a LiDAR-derived 3m and two publicly available DEMs at 10 and 30m resolution. They found the 1D model to be more sensitive to the different DEMs than the 2D model, with a 25% and 75% increase in damages respectively at 30m with comparable increases in flood footprint. In a recent large-scale study, Pollack et al. (2022) construct a benchmark and aggregated analog models from roughly 800,000 single family dwellings and eight 30m resolution flood depth grids with return periods ranging from 2- to 500-years. When only building attributes were aggregated, a small negative bias was observed (-10%) while when hazard variables were also aggregated a large positive bias was found (+366%) for annualized damage. Given the spatial correlation of building values and flood exposure found in their study area, they conclude that bias would be difficult to predict ex-ante. They also find that errors arising from missing data and damage function uncertainties can be orders of magnitude greater than those arising from aggregation.

Leveraging a rich object-scale dataset of 300 buildings damaged by a 2010 Italian flood, Molinari and Scorzini (2017) provide a non-grid based comparison to investigate the sensitivity of their multi-variate damage modelling framework to input data accuracy. For this, six models were built with different combinations of input data elements either at object-scale or averaged across the census-block (taking the mode or the mean). Results were mixed; however, the model where all inputs were aggregated had a $\frac{s2}{s1}$ bias of 1.51. While this approach is suitable for investigating model sensitivity to input data accuracy, because exposure data was aggregated from object-scale data *after* hazard data sampling (rather than aggregating before sampling) these findings are less relevant to the broader issues of scaling challenging

Table 3.1: Summary of selected studies with paired grid-based models at fine and coarse resolution. The bias is computed from the reported aggregated total damage of the coarse divided by the fine model. "[...]" indicates a coarse ($s2$) model element which is identical to its fine ($s1$) pair.

ref.	**fine (s1) description**	**coarse (s2) description**	**bias (s2/s1)**
Apel et al. (2009)	Hazard: 2D hydrodynamic with triangular finite elements on 25 m DEM. Exposure: building-scale Vulnerability:multi-variate empirical private sector building damage.	[. . .] Exposure: dasymetric land-use grid at best 100m. [. . .]	1.16
Sieg et al. (2019) and Seifert et al. (2010b)	Hazard: random sample of water mask values. 10m. Exposure: 165 businesses (object-scale, aspatial and stochastic) Vulnerability: Random Forest empirical commercial damages.	Hazard: interpolation of highwater marks. 25m. Exposure: disaggregated average municipal asset values. 25m. Vulnerability: multi-variate empirical commercial damage.	5.68
Sieg et al. (2019) and Seifert et al. (2010b)	Hazard: random sample of water mask values. 10m. Exposure: 15 businesses (object-scale, aspatial and stochastic) Vulnerability: Random Forest empirical commercial damages.	Hazard: 1D/2D hydrodynamic LISFLOOD-FP. 25m Exposure: disaggregated average municipal asset values. 25m. Vulnerability: multi-variate empirical commercial damage.	8.88
Komolafe et al. (2015)	Hazard: 1D/2D hydrodynamic. 50m. Exposure: remote sensing derived land-use grid. 30m. Vulnerability: multi-variable synthetic direct building damages.	[. . .] 1000 m upscale (unspecified method) [. . .] [. . .]	1.05
Brussee et al. (2021)	Hazard: 2D flexible mesh hydrodynamic. 5m. Exposure: disaggregated neighbourhood-scale Vulnerability: multi-variable mortality function.	[. . .] 100m [. . .] [. . .]	1.08
Ghimire and Sharma (2021)	Hazard: 2D hydrodynamic. LiDAR derived 3m Exposure: buildings (object-scale) Vulnerability: depth-damage curves	[. . .] unspecified 30m [. . .] [. . .]	1.33
Pollack et al. (2022)	Hazard: 2D hydrodynamic. 30m Exposure: buildings (object-scale) Vulnerability: uni-variate synthetic	[. . .] [. . .] aggregated to census block-group (order 1-100km) [. . .]	4.67

aggregated models used in practice.

Spatial resolution transfers within flood risk models are often implicit and can occur as part of data preparation, modelling, or post-processing for reasons ranging from efficiency to privacy protection. For example, a DEM might be aggregated as part of preprocessing for a hydrodynamic model to improve stability and efficiency (Sampson et al., 2015; Bates et al., 2021), or risk results might be aggregated to adhere to a licensing agreement (Wing et al., 2022). Focusing on the outputs of hazard models (e.g., flood depth grids) and excluding preprocessing (e.g., DEM resampling), most flood risk model studies maintain a single resolution throughout the analysis inherited from some base DEM (Hall et al., 2005; Sairam et al., 2021; Bates et al., 2021). However, some studies aggregate hazard model outputs to facilitate intersection with more coarse exposure data, either through simple averaging (Seifert et al., 2010a; Sieg and Thieken, 2022) or some unspecified method (Thieken et al., 2016; Jongman et al., 2012). Hazard model outputs have also been aggregated to facilitate comparison with more coarse climate re-analysis data (Paprotny et al., 2020c). These often forgotten model manipulations have so far not been investigated.

The goal of this paper is to partially explain the bias shown in Table 3.1 through generalizeable methods (i.e., not bound to the specifics of individual case studies) and thereby improve our understanding of the effects of spatial scale transfers on flood risk models. In this study, we focus on flood hazard data, composed of a set of grids, and their intersection with assets or buildings to calculate exposure — two initial stages of risk modelling. To explore scaling effects, we compare fine grids to their coarse analogs using metrics of interest to flood risk modellers. Rather than construct these coarse analogs through hydrodynamic modelling as has previously been done, we aggregate hazard grids through averaging routines. In this way, we provide the first guidance and explanation for practitioners aggregating or upscaling flood hazard grids, along with an easy-to-use QGIS script (`https://github.com/cefect/FloodRescaler`). Further, we elucidate some endemic scaling effects and provide evidence and explanation to the positive bias common among coarse flood risk models.

3.2 Flood Hazard Grids and Scales

There are three primary hazard grids included in most flood risk models: Water Depths (WSH), Water Surface Elevations (WSE), and the Ground Elevations (DEM) related by the following:

$$WSE = DEM + WSH \tag{3.1}$$

Combining Equation 3.1 with the assumption that the flood hazard grids are constrained to surface water flooding (i.e., ground water is irrelevant), yields the following expectations:

$$WSH \geq 0 \quad \text{and} \quad WSE > DEM \tag{3.2}$$

Alternatively, it can be argued that a WSE grid is still valid when: $WSE = DEM$; however, this relaxation results in a WSE grid with less information as it is no longer possible to determine wet from dry cells without a second companion grid like the DEM. Further, the information provided by the grid becomes less associated with the label "water surface" as many (or potentially all) of the values represent dry ground. For these reasons, we provide no further consideration for this paradigm.

From Equation 3.1 emerges an important distinction for the handling of dry cells:

$$WSH_{i \text{ or } j} = 0 \iff WSE_{i \text{ or } j} = null \iff \text{“dry”} \tag{3.3}$$

where i is the index of a fine ($s1$) and j a coarse ($s2$) grid cell. In other words, because WSE values are on some absolute vertical datum, the grid is undefined in dry regions, whereas WSH, being relative to ground (DEM), should have zero values in these same regions. Absent transformation or resampling, the application of Equation 3.1 and 3.3 is trivial and allows for simple conversion between WSE and WSH or vice versa using the DEM. However, in the presence of dry cells Equation 3.3 leads to inconsistencies when computing the denominator of averaging operations: local averages of DEM and WSH grids use the total count of $s1$ cells contributing to a coarse $s2$ cell (N_{12}), while WSE grids must omit dry cells from the denominator ($N_{wet} = N_{12} - N_{dry}$) where N_{dry} is the count of $s1$ cells described in Equation 3.3. This can be expressed mathematically as:

$$DEM_{s2,j} = \overline{DEM_{s1,i}} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} DEM_{s1,i} \tag{3.4}$$

$$\overline{WSH_{s1,i}} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} WSH_{s1,i} \tag{3.5}$$

$$\overline{WSE_{s1,i}} = \frac{1}{N_{wet}} \sum_{i=1}^{N_{wet}} WSE_{s1,i} \tag{3.6}$$

The remaining sections show how these inconsistencies can lead to bias of key metrics when applying aggregation routines.

3.3 Methods

To investigate any potential bias arising from aggregation of flood hazard grids, we introduce the novel *resample case framework* for classifying the flood hazard grid domain. With this, two typical grid aggregation routines are investigated first analytically, then computationally using data from a 2018 fluvial flood in Canada as an example. Finally, we evaluate regions with exposure (locations with buildings) to provide an analysis of bias particularly relevant to flood risk models.

3.3.1 Aggregation Routines

To demonstrate the application of our *resample case framework*, we consider two routines for yielding a set of $s2$ analog grids from a set of $s1$ grids through averaging local groups with N_{12} cells. Both respect Equation 3.1 and 3.2, but differ in their strategy for preserving averages in the resulting $s2$ analogs: the first preserving water depths (WSH Averaging) and the second water elevations (WSE Averaging). In this way, each routine has a primary grid (WSH or WSE), which is computed through direct averaging, and a secondary grid (WSE or WSH) computed through addition or subtraction with the DEM. Both routines use Equation 3.4 to obtain DEM_{s2}, as this is not affected by the dry cells in Equation 3.3. Further, both rely on Equation 3.1 to compute the secondary grid — rather than averaging which would yield a grid set in violation of Equation 3.1 (this can be seen by comparing the WSH grids in Figure 3.1d and e). Figure 3.1d and e provide a graphical summary and toy example of these routines, which are defined mathematically in Section B.1. Both routines are easily implemented in a few steps using standard spatial software packages or the provided QGIS script (`https://github.com/cefect/FloodRescaler`). While additional aggregation routines are possible, these two were selected as they are the simplest, are amenable to analytical treatment, and yield hydraulically reasonable grids.

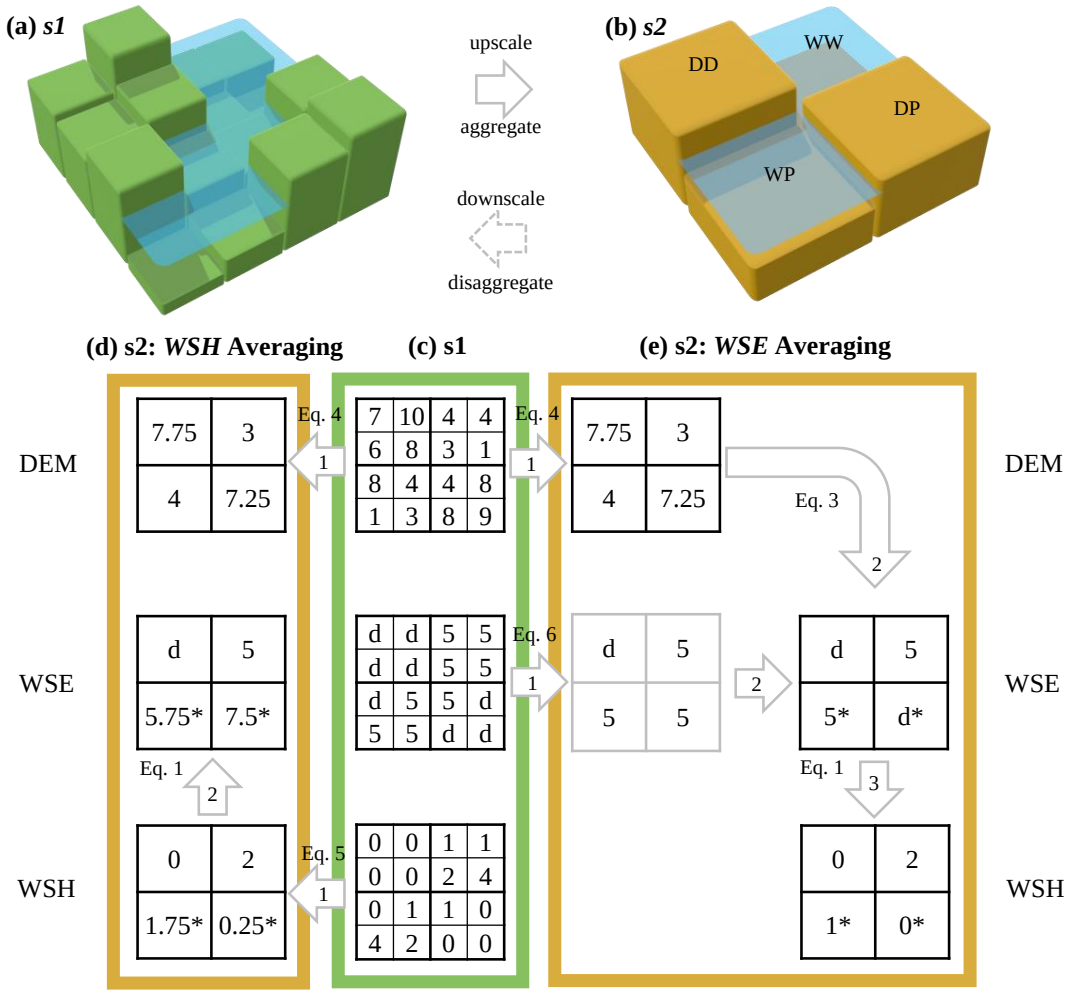


Figure 3.1: Flood hazard data scaling issues and two aggregation schemes demonstrated with a toy example. Panel (a) is an oblique view of a fine ($s1$) DEM and WSE while panel (b) shows an aggregated coarse ($s2$) analog and corresponding resample case (DD, WW, WP, DP) from Figure 3.2. Panel (c) shows an example set of s1 values for the three grids described by Equation 3.1. Panel (d) and (e) show the two aggregation routines described in the text based on averaging the WSH and WSE grid respectively. Numbered arrows indicate phases within each scheme, the "Eq." notes refer to equations from the text, "d" denotes dry or null WSE grid values, and light grey grids show intermediate calculations. Discrepancies between resulting $s2$ grids from the two routines are marked with *.

3.3.2 Resample Case Framework

To understand and spatially attribute the effects of aggregation routines on flood hazard grids, we classify each cell in the $s1$ domain into one of four cases of potentially homogeneous aggregation behaviour. We define each of these resample cases using local relations of the DEM_{s1}, WSH_{s1} and WSE_{s1} fine data grids within a block j of size N_{12} as shown graphically in Figure 3.2 and defined explicitly as:

$$case_j = \begin{cases} DD & \text{if } max(WSH_{s1,i}) = 0 \\ DP & \text{if not } DD \text{ and } \overline{DEM_{s1,i}} \geq \overline{WSE_{s1,i}} \\ WP & \text{if not } WW \text{ and } \overline{DEM_{s1,i}} < \overline{WSE_{s1,i}} \\ WW & \text{if } min(WSH_{s1,i}) > 0 \end{cases} \tag{3.7}$$

where $\overline{WSE_{s1,i}}$ is defined in Equation 3.6, the first letter of the $case_j$ label code is determined by the relative averages of WSH_{s1} and DEM_{s1}, and the second letter by the overlap of extremes between WSE_{s1} and DEM_{s1} grids as shown in Figure 3.2b. The quadrants in Figure 3.1a provide a simple example of four such i groups whose corresponding case labels are shown on Figure 3.1b. A simpler framework could employ only three cases (wet, partial, dry); however, we find sub-dividing the "partial" zone provides more spatial segregation and further discernment of aggregation artifacts (see results below for the *WSE Averaging* routine). Figure 3.3 shows a fully classified domain where WSH_{s1} has been simulated using a hydrodynamic model built on a 1m fine DEM described below. This example shows how partial cases (DP and WP) represent regions of marginal flooding or partial inundation of coarse (j) blocks by the fine source (i) cells.

Such a resample case map is independent of any coarse ($s2$) aggregate grids or the routine used to compute them as it is purely a function of the fine source ($s1$) grids and the target coarse resolution (i.e., N_{12}). However, by treating each region independently, certain behaviours of aggregation routines may be generalized. For example, we expect all aggregation routines to yield fully dry and fully wet grids in DD and WW regions respectively. The partial regions (DP and WP) on the other hand are ambiguous, and we expect $s2$ grids generated by different aggregation routines may differ in these regions.

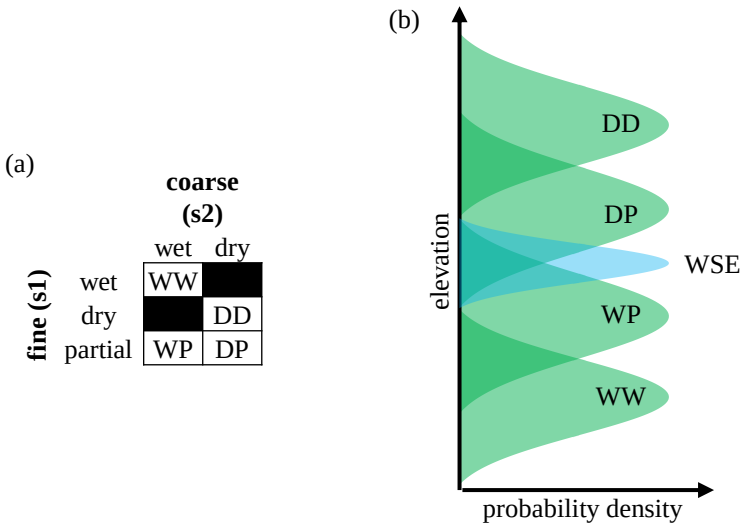


Figure 3.2: Framework for classification of flood hazard resample case. Panel (a) shows class label acronyms. Panel (b) provides a conceptual diagram showing a hypothetical distribution of WSE_{s1} and four possible DEM_{s1} groups and their resulting resample case. D, W, and P stand for "dry", "wet", and "partial" respectively.

This *resample case framework* provides a reproducible approach to advance our understanding of flood grid aggregation in three ways. First, by classifying the resample case of each cell in the domain, bias or artifacts emerging through aggregation can more clearly be communicated. Second, this classification facilitates the study of spatially heterogeneous mechanisms (e.g., inundation area under aggregation may have bias in opposite directions in different regions that cancel absent such classification). Third, and most importantly, this framework facilitates an analytical examination by allowing the problem to be separated into four simpler ones which may yield closed-form solutions of the bias introduced by an aggregation routine — independent of any case study or region.

3.3.3 Analytical Approach

In the first phase of this study, we investigate the two aggregation routines analytically in the pursuit of *generalized* expressions of bias that would apply to any case study region. To accomplish this, each of the four resample cases is investigated separately, which provides the simplifying assumptions that allow closed-form solutions to the bias imparted on each metric of interest. To investigate the behaviour of an aggregation routine as a function of target resolution ($s2$), we define *true* values as those represented in the fine ($s1$) grid used as an input to the aggregation routine. This allows us to investigate the error introduced solely through aggregation by computing, and then comparing, metrics between the fine ($s1$) and coarse ($s2$) grids. From this, an important distinction is made between *non-systematic* errors — differences in $s1$ and the corresponding $s2$ values with a zero-mean — and *systematic* errors which have a non-zero mean. In flood grid aggregation, these *non-systematic* errors are an obvious or even intentional product — generally thought to cancel in larger models (Merz et al., 2004). Systematic errors on the other hand, which we call "bias", are an undesirable artifact of aggregation and the focus of this analysis.

Four metrics, typically of interest to flood-related analysis, are considered: two primary metrics, water depth (WSH) and water surface elevation (WSE), and two derivative metrics, inundation area (A), and volume (V). Primary metrics are computed as grid-wide *global* averages similar to Equation 3.5 and 3.6, but evaluated against all cells in a region of interest (rather than local groups). For example, $\overline{WSH_{s1,WW}}$ is the sum of all WSH_{s1} cells classified as resample case WW per Equation 3.7 divided by the count. The derivative metrics are computed as grid-wide totals: inundation area (A_s) is the count of all non-dry grid cells multiplied by the area of each cell (s^2) and volume (V_s) is the sum of all WSH_s values multiplied by the area of each cell.

To better attribute bias spatially, we also compute a *local* bias for the primary metrics WSH and WSE. This allows us to separate errors owing to the increase in flood footprint, from those attributable to changes in local values. For this, we first calculate the error of each $s2$ cell vs. $s1$ group, before computing the mean of these error values to obtain a single bias metric. For the WSE metric, this local bias can of course only be computed in regions inundated by both $s1$ and $s2$ grids (see Equation 3.2), as the grid is undefined in other regions. For consistency, we apply this same constraint to the WSH metric even though it could be resolved on the full domain. While this obscures the performance of a routine in dry regions, it provides a consistent way to separate the reporting of bias in local variables from bias in inundation area (which is reported as a separate metric).

3.3.4 Computational Approach

In the second phase of the study, we further demonstrate the utility of the novel *resample case framework* by applying it to the May 2018 Saint John River flood in Canada. To elucidate aggregation bias, we aggregate using the two routines on a set of 1m resolution grids before applying the *resample case framework* to quantify bias arising from the aggregation. The DEM_{s1} grid was downloaded from GeoNB who constructed the bare earth terrain model from six aerial LiDAR points per m^2 flown in the summer of 2015 (Government of New Brunswick, 2016). The WSE_{s1} grid was simulated by GeoNB using a hydrodynamic model (on the aforementioned DEM_{s1}) calibrated to field surveyed high water marks and is further described in Bryant et al. (2022a) and provided by GeoNB (2019). This model did not remove building footprints from the DEM; however, given the width and low-density of houses in the flood plain this simplification likely has a negligible influence on the model. The WSH_{s1} grid was computed with Equation 3.1 yielding the grids shown in Figure 3.3a. From these fine ($s1$) grids, a set of five

($s2 = 2^n$m for $n = 3, 6, 7, 8, 9$) aggregated retrograde $s2$ analog grids and the corresponding resample classification maps (e.g., Figure 3.3b) are computed for the *WSE Averaging* and *WSH Averaging* routines for a total of 40 grids (4 grid types x 5 coarse resolutions x 2 routines). Komolafe et al. (2015) take a similar approach, but only for the WSH grid and they do not specify the aggregation routine or report the metrics discussed here.

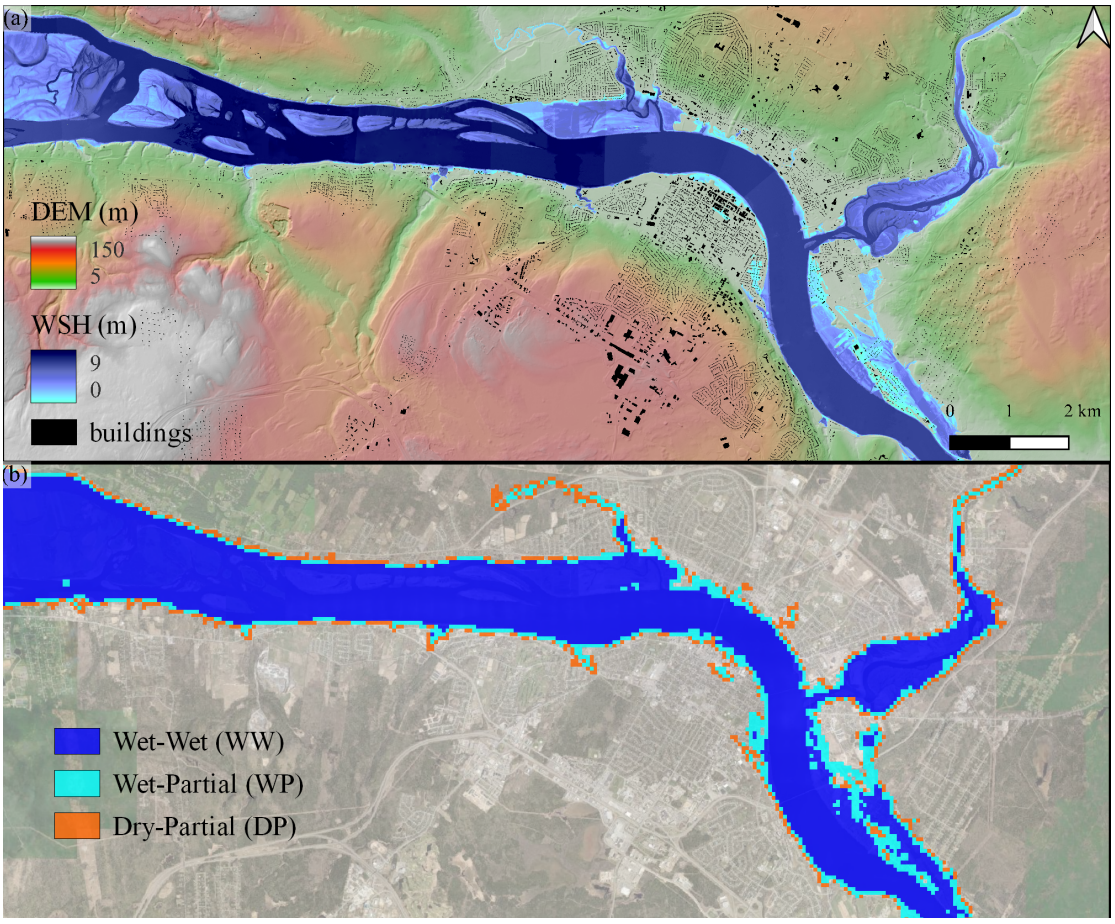


Figure 3.3: Simulated May 2018 Saint John River flood in Canada. Panel (a) shows DEM_{s1} and WSH_{s1} at 1m resolution and building footprints from Microsoft (2019). Panel (b) shows corresponding resample case (see Figure 3.2) for a 1:64 aggregation (DD is transparent for clarity).

While bias in aggregated flood grids is of general interest, flood risk models are particularly concerned with those regions where assets or buildings are present. To explore the significance of this exposed domain (in contrast to the full domain), building locations within the study area were obtained from Microsoft (2019) (see Figure 3.3a). From the centroids of footprints in this layer, each of the aforementioned 40 retrograde grids is sampled to produce a parallel dataset from which the same metrics of interest can be computed for the exposed domain.

3.4 Analytical Results and Discussion

To investigate the six metrics of interest (A, V, and local and global WSH and WSE bias), we apply the *resample case framework* to the two aggregation routines. The analysis and algebra are detailed in Section B.1 while the results are summarized in Table 3.2.

Focusing on the non-partial regions (DD and WW), Table 3.2 shows that aggregation preserves all metrics of interest here. This is intuitive considering our aggregation routines and the selected metrics are commutative and cumulative in the absence of dry cells. Put simply, this is the naive expectation for the aggregation of a continuous grid: averages are preserved. Outside of this — in the partial regions (WP and DP) — flood hazard grid behaviour deviates from that of continuous grids owing to the presence of dry cells and the inter-grid relations (see Equation 3.3 and 3.1). Examining the bias in these partial regions (WP and DP), Table 3.2 shows some bias for all metrics except the respective primary grids on the global metric (i.e., *WSE Averaging* has no $Bias_{global}[\overline{WSE}]$ bias and *WSH Averaging* has no $Bias_{global}[\overline{WSH}]$ bias — or $Bias[\sum V]$, which is discussed below). This suggests that a single aggregation routine which employs averaging will *always* carry bias on some metric in partial regions; another artifact that follows from Equation 3.1 and 3.3.

Table 3.2: Biases in two aggregation routines evaluated analytically for each resample case. For metrics computed from the WSE grid, which has no value for dry cells, "n/a" denotes dry regions. Similarly, the aggregation routine *WSE Averaging*, which resolves dry cells for both DD and DP cases, shows "n/a" for $Bias_{local}[\overline{WSH}]$ as our definition of local requires wet cells on both the $s1$ and $s2$ grids. The remaining "+"/"-" symbols indicate cases where we found the metric calculated with the $s2$ grid to be systematically higher/lower than the $s1$ grid, while "0" indicates the metrics are equivalent.

		resample case			
metric	equation	DD	DP	WP	WW
	WSH Averaging				
$Bias_{global}[\overline{WSH}]$	$\overline{WSH_{s2}} - \overline{WSH_{s1}}$	0	0	0	0
$Bias_{local}[\overline{WSH}]$	$\overline{WSH_{s2} - WSH_{s1}}$	0	-	-	0
$Bias_{global}[\overline{WSE}]$	$\overline{WSE_{s2}} - \overline{WSE_{s1}}$	n/a	+	+	0
$Bias_{local}[\overline{WSE}]$	$\overline{WSE_{s2} - WSE_{s1}}$	n/a	+	+	0
$Bias[\sum A]$	$\sum A_{s2} - \sum A_{s1}$	0	+	+	0
$Bias[\sum V]$	$\sum V_{s2} - \sum V_{s1}$	0	0	0	0
	WSE Averaging				
$Bias_{global}[\overline{WSH}]$	$\overline{WSH_{s2}} - \overline{WSH_{s1}}$	0	-	-	0
$Bias_{local}[\overline{WSH}]$	$\overline{WSH_{s2} - WSH_{s1}}$	0	n/a	-	0
$Bias_{global}[\overline{WSE}]$	$\overline{WSE_{s2}} - \overline{WSE_{s1}}$	n/a	n/a	0	0
$Bias_{local}[\overline{WSE}]$	$\overline{WSE_{s2} - WSE_{s1}}$	n/a	n/a	0	0
$Bias[\sum A]$	$\sum A_{s2} - \sum A_{s1}$	0	-	+	0
$Bias[\sum V]$	$\sum V_{s2} - \sum V_{s1}$	0	-	-	0

Contrary to global bias, the analysis shows the *WSH Averaging* routine has a negative $Bias_{local}[\overline{WSH}]$ in partial regions (WP and DP). A simple explanation for this is illustrated in Figure 3.4a, where we see the aggregated values have a progressively lower local value (measured at the centre), while the global average remains constant. In other words, given a wet $s1$ cell with some dry neighbours, aggregating depths through averaging will produce progressively smaller (i.e., shallower) depth values. *WSE Averaging* on the other hand does not suffer from this as dry cells are omitted from the denominator during averaging (see Figure 3.4b). This has important implications for model scaling. For example, *WSH Averaging*, arguably the simplest aggregation routine, appears to preserve WSH when viewed globally — but in fact imparts a negative bias in partial regions.

For inundation area (A), the analysis shows a positive bias for *WSH Averaging* and a mixed bias for *WSE*

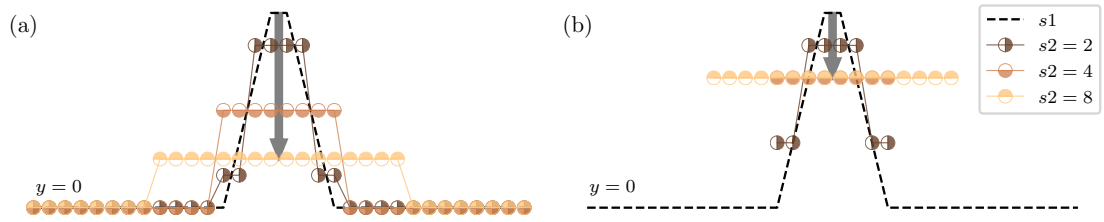


Figure 3.4: Conceptual diagram showing a cross-section of local bias for three target resolutions generated through two types of averaging: (a) zero-inclusion (as in Equation 3.5) and (b) zero-exclusion (as in Equation 3.6). All series within a panel have the same global mean. Black arrow shows the progression of local bias.

Averaging in partial regions. This is consequential for flood risk models, considering changes to flood footprints are expected to lead to changes in flood exposure, a sensitive component in risk calculations (Jongman et al., 2012; Metin et al., 2018). With this in mind, the *WSE Averaging* routine seems preferable considering it at least has the potential to preserve $\sum A$; however, obviously some disparity in local inundation is expected with any routine — this phenomena is explored further below. Finally, Table 3.2 shows $Bias[\sum V]$ follows the same behaviour as $Bias_{global}[\overline{WSH}]$ (see Section B.1 for derivation), meaning *WSH Averaging* also preserves $\sum V$. This suggests a paradox for hydrodynamic modellers: aggregating outputs biases either V, which violates mass conservation, or WSE, which may violate calibration.

This analysis has shown mathematically whether or not a metric will be biased by a given aggregation routine. By employing the *resample case framework*, these bias inequalities become closed-form, independent of grid values, and ubiquitous within their respective regions. In other words, they apply to all grids aggregated with a given routine and *all* cells within that region. These provide definitive, albeit limited, statements about the behaviour of the two aggregation routines applied to any case (assuming segregation into resample cases). However, this does not provide any indication of the magnitude of bias, which is case specific, and provides conditional evidence on the relative magnitude between resample cases (e.g., whether $Bias[WD] > Bias[DP]$). For example, so far we have not provided an evaluation about the prevalence or proportion of each resample case (e.g., a grid set could conceivably have only one resample case, rendering most of the analysis here irrelevant). With this in mind, the following section applies the same resample case framework to a case study of a 2018 flood.

3.5 Computational Results and Discussion

While the analytical approach is the main contribution of our work as it is independent of any case study region, here we demonstrate the application of the novel *resample case framework* on a 2018 flood, and show how it can be used to explain aggregation bias. For this, two domains are considered: first, the complete rectangular or *full domain* shown in Figure 3.3; and second, the *exposed domain*, a sub-set of the full domain including only cells intersecting building centroids. To attribute bias to specific regions, and to compare with the results of the analytical approach, both these domains are further sub-set by the four resample cases defined in Figure 3.2.

3.5.1 Full Domain

Figure 3.5 shows the resulting change in composition or classification of the domain, computed from the classification map obtained at each $s2$ scale. This shows that the portion of partial regions (WP and DP), we call $\rho_{s2/s1}$, increases with aggregation: ranging from ρ_{32} = 5% to ρ_{512} = 40% for our case (Figure 3.5b). This is intuitive considering these partial regions as transition zones between wet and dry cells —– and that these zones must cover an increasing portion of the domain to be resolved as the resolution coarsens. This has significant implications for flood risk models considering the previous section showed these partial regions are those subject to bias of key metrics during aggregation: this suggests $\rho_{s2/s1}$ is positively correlated with aggregation bias sensitivity or magnitude. In other words, the portion of the domain containing errors of aggregation, $\rho_{s2/s1}$, increases with coarser resolutions or higher $s2/s1$ values. Extending these observations to other study regions, we hypothesize that $\rho_{s2/s1}$ is also sensitive to hydraulic regime: with broad-flat floodplains having a higher inundation-area-to-perimeter ratio, and therefore lower $\rho_{s2/s1}$ values with less sensitivity to aggregation bias; while narrow-steep floodplains would have higher $\rho_{s2/s1}$ values and greater sensitivity to aggregation bias. Further, these transition zones, or shorelines, often have a higher density of assets — a phenomena explored in Figure 3.5c and discussed below.

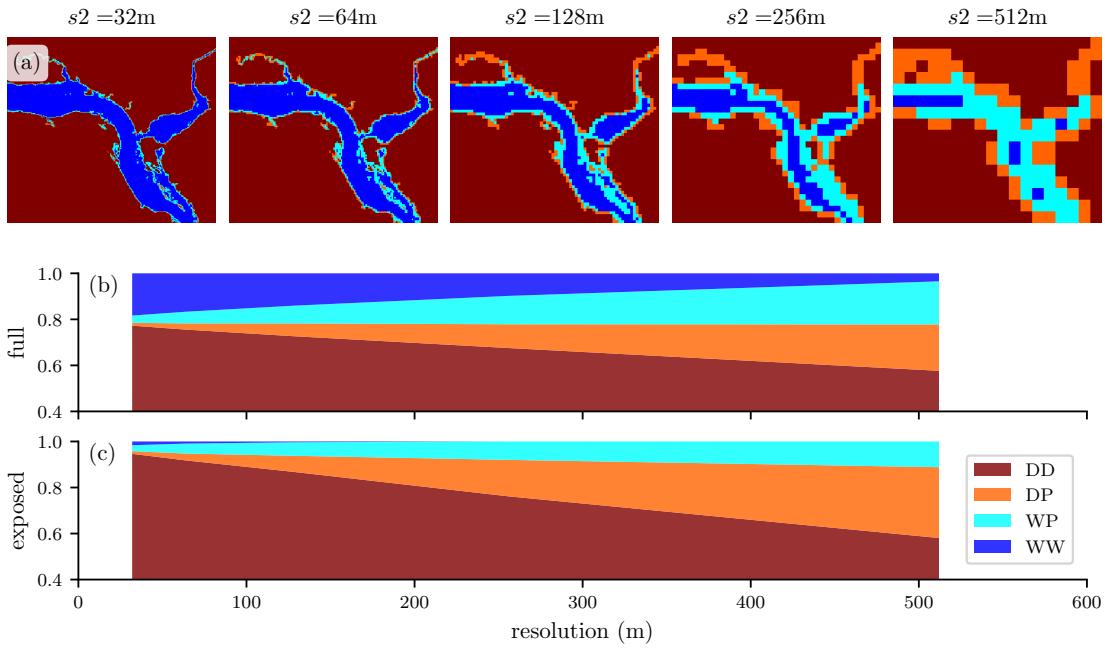


Figure 3.5: Resample case classification progression for May 2018 Saint John River flood hazard data showing (a) illustrative maps at five target resolutions ($s2$); (b) full domain fraction; and (c) exposed domain (i.e., values sampled at buildings) fraction for each case. See Figure 3.2 for description of legend.

To demonstrate how these dynamic regions interact with the grid values calculated by each aggregation routine, the six aforementioned metrics are computed by comparing the analog $s2$ grids to the original 1m resolution $s1$ grids. These calculations are performed on the full domain and each resample case as independent regions of interest to develop five magnitude vs. resolution series for each metric and routine. Results for four of the key metrics are shown in Figure 3.6a and b while the remaining two metrics are provided in Figure B.1.

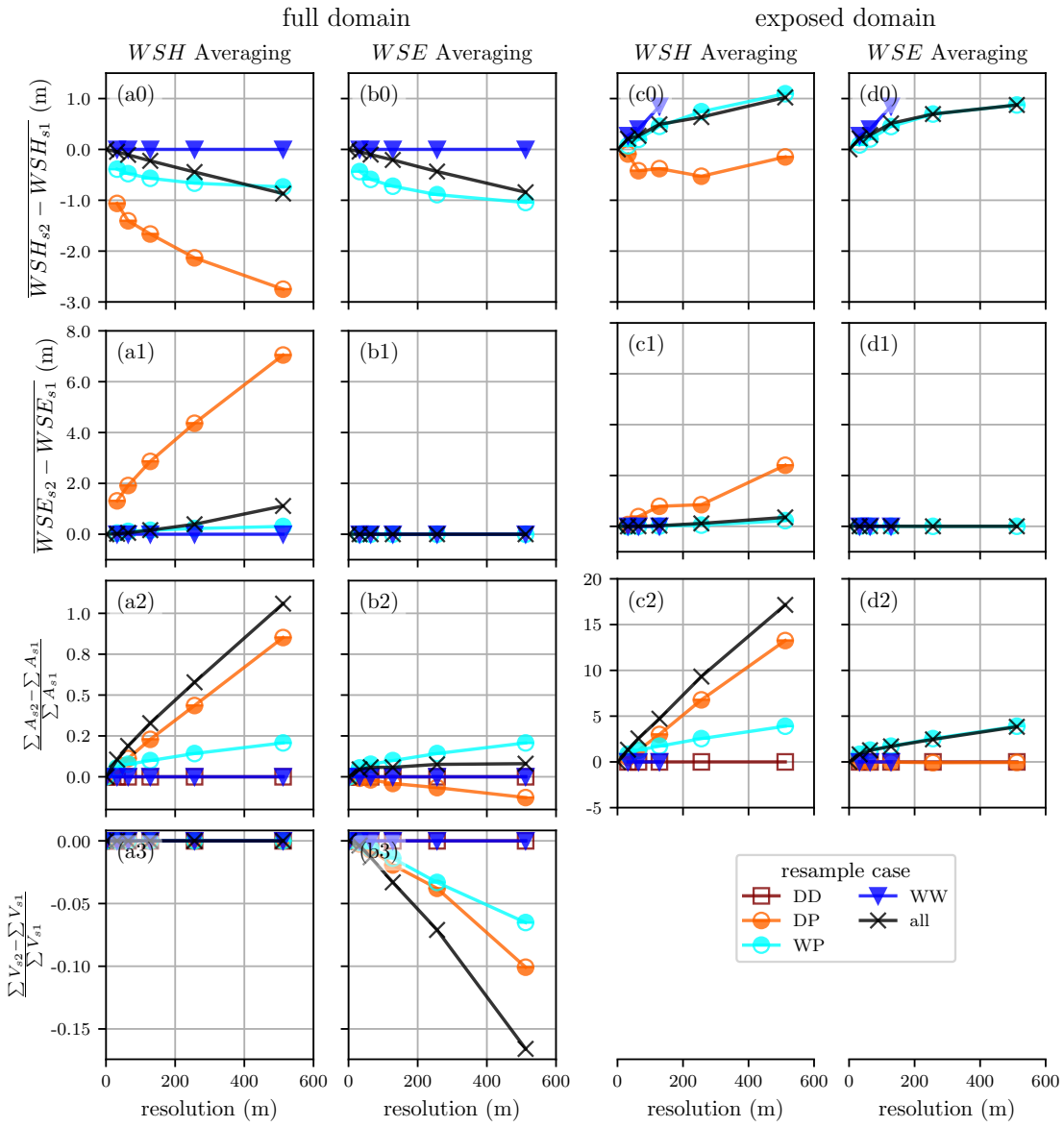


Figure 3.6: Bias from aggregation of four metrics for two routines sub-sampled by resample case (see Figure 3.2) for the full domain and the exposed domain. The "all" series uses the complete region of interest, without sub-setting by resample case. Panels (a2) and (b2) show the non-dry or inundated area of the full domain while panels (c2) and (d2) show the count of non-dry or exposed buildings.

Comparing Figure 3.6 and S1 to Table 3.2 shows all computations agree with the directional bias derived analytically in the previous section. For the *WSH Averaging* routine, Figure 3.6a suggests the bias in the DP case is always more severe than the WP case. This is also shown analytically in Section B.1 for certain conditions (e.g., $N_{wet,DP} < N_{dry,WP}$). However, while the conditions favouring a more severe DP bias are intuitively more common, these conditions are not ubiquitous.

When aggregating, both the analytical and computational results show either decreasing or stable $\overline{WSH_{s2}}$ (Table 3.2, Figure 3.6a0 and b0); opposite of what Muthusamy et al. (2021) find when comparing increasingly coarse hydrodynamic models without adjusting the calibration. Saksena and Merwade

(2015) take a similar approach to Muthusamy et al. (2021) but only report $\overline{WSE}$, which they also find increasing. This contradiction can be explained if we consider the uncalibrated hydrodynamic models are forced by boundary conditions (namely a hydrograph), while the aggregation routines are forced by the fine ($s1$) grid values. To make up for the loss of the deepest cells (i.e., the thalweg), the former achieves balance through increasing depths (and conveyance) while the latter increases volume or area. A more appropriate comparison between aggregation and coarse hydrodynamic modelling would need to re-calibrate each hydrodynamic model to some observations; an example of which is provided in Section B.3 which shows a stable $\overline{WSH_{s2}}$.

For all partial zones, *WSH Averaging* shows a doubling (100% increase) of the inundated area (A) for the $s2 = 512m$ grids in this case study (Figure 3.6a2). *WSE Averaging* showed less bias, with the WP and DP global bias nearly balancing, leading to a meagre 10% increase for $s2 = 512m$ (Figure 3.6b2). However, the reader should note that our selected $\sum A$ metric is *global*, and that while the total areas may nearly balance, a substantial number of falsely inundated cells may be generated in the aggregated grids.

In practice, we recognize scale transfers in flood risk models through grid aggregation generally involve only small changes in resolution. More prevalent is the use of scale transfers via coarse hydrodynamic models, where the friction term is calibrated to observations yielding coarse WSH grids directly from aggregated or coarse DEM grids. Such scaling issues in coarse hydrodynamic models have been studied extensively (Banks et al., 2015; Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021; Xafoulis et al., 2023). To evaluate the transferability between the aggregation routines considered here and hydrodynamic modelling (i.e., whether aggregation behaviour might serve as an analog for coarse hydrodynamic modelling behaviour) Section B.3 presents a brief comparison of the aggregation routines against a depth calibrated hydrodynamic model built on a DEM resampled to the coarse resolution ($s2 = 32m$) for a separate case study where more data is available. While this comparison is limited to the specifics of this study region and a single target resolution ($s2 = 32m$), the results provide an example of the magnitudes of bias for the two routines on a smaller floodplain. However, the differences in depths and bias metrics between the coarse model and the aggregation routines suggests that some artifacts arrived at through coarse modelling are not present in aggregated grids. This is intuitive if we consider that the fine hydrodynamic model (which serves as an input to the aggregation routines) is capable of resolving more flow paths than its coarse twin — leading to frequent disparities between the two results which propagate into disparities between the aggregation routines and the coarse model. In other words, the findings on bias and artifacts in hazard grid aggregation reported thus far do not necessarily extend to coarse hydrodynamic model results.

Finally, given our focus on artifacts introduced through averaging, we treat the high-resolution input grids as *true*, rather than some direct observations (e.g., high water marks). From this, it follows that some artifact or error from aggregation described here may actually provide a more accurate representation of a flood event (e.g., due to errors in the high-resolution input grids); however, any such phenomenon is independent from the computational artifacts we investigate here.

3.5.2 Exposed Domain

Having now demonstrated the character of bias on the full domain, we turn our focus to those regions of particular interest to flood risk modellers: developed areas or the exposed domain. Figure 3.5c shows that WW regions are insignificant for building exposure. This can be explained by: first, that the four cases form roughly concentric rings ($WW > WP > DP > DD$), radiating out from regions of continuous

flooding (i.e., the river channel for fluvial floods) as demonstrated by Figure 3.3b; and second, that buildings are less prevalent within the river channel. Further, Figure 3.5b shows that DP regions are more than twice as prevalent for building exposure, leading to roughly 30% of buildings classified as either WP or DP at a resolution of 512m, compared to 20% on the full domain for this case study. Recalling from the previous section that these partial regions (WP and DP) are those subject to bias suggests that the exposed domain is more sensitive to aggregation bias than the full domain.

The magnitude of increased sensitivity, or relevance, of the exposed domain to aggregation bias for this case study is shown in Figure 3.6c and d and Figure B.2. Comparing the elements in Figure 3.6 row 2 shows that the exposed building count is an order of magnitude more sensitive to aggregation bias than inundation area (note the vertical axis). For example, at $s2 = 512m$ the full domain shows an increase in inundation area of 110% and 8%, while the exposed domain shows an increase of 1800% and 400% for WSH and *WSE Averaging* respectively. This is intuitive if we consider the distribution of buildings: few in regions flooded by the base grids and many immediately adjacent (see Figure 3.5c). In other words, a small increase in flood footprint leads to a large increase in the number of exposed buildings. In their comparison of 3 and 30m hydrodynamic models, Ghimire and Sharma (2021) find a comparable factor of two increase in building exposure.

For water surface elevations (WSE), bias generated in the full and exposed domain have the same direction and relative ranking of resample cases (Figure 3.6 row 1); however, the values show a muted bias in the exposed domain relative to the full domain. In other words, grid cells with the most severe WSE errors tend to have fewer buildings, but this may be specific to our case study. Counter to this, Figure 3.6 row 0 shows a significant difference in the sensitivity to water depth (WSH) errors between the full and exposed domain: with the full domain having a negative (or no) bias and the exposed domain having a positive bias for all but the DP case. This can be explained if we consider that the aggregation routines (and the full domain metrics) include all $s1$ cells in a group, while the exposed domain sampling (and therefore the metrics) ignore those cells without exposure. Figure 3.7 shows a clear example where each tile has the same $\overline{WSH_{s2}}$ on the full domain, but within $s2$ cells the buildings occupy drier ground. In other words, assets exhibit a *dry bias*, so the artifacts leading to systematic grid errors may cease to be systematic when only the exposed subset is considered. Pollack et al. (2022) discuss a mechanism with the opposite bias, where high-value assets tend to be closer to the shoreline and therefore have disproportionately higher risk, imparting a negative bias in the damage estimates for some aggregate blocks. These mechanisms are not contradictory however, as they operate at different scales (Pollack et al. (2022)'s base scenario is 30m resolution and they aggregate assets to counties which can be on the order of 1-100 km) and on different elements of risk modelling (exposure vs. damage). In other words, both may be present in a large model like Pollack et al. (2022)'s.

For our analysis, building centroids were sampled from each flood grid as a simple representation of flood exposure. At grid resolutions near building scale (5-50m), this simple approach may introduce additional artifacts as many pixels are used to represent the building location in the hazard grid but only one is sampled for the exposure value. Further, this problem becomes intractable when buildings are incorporated as obstructions within hydrodynamic models or flood grids. Given the coarse resolutions considered in our case study, we assume the centroid approach is adequate to demonstrate bias from aggregation; however, the challenge of exposure sampling deserves further study.

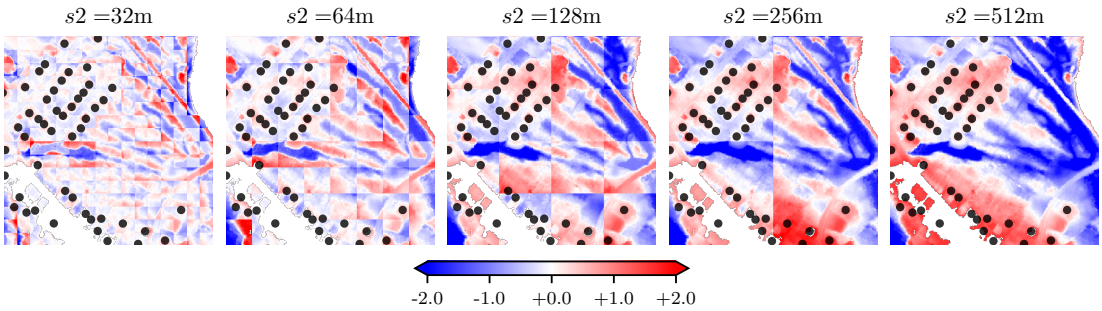


Figure 3.7: WSH difference maps for an example 512m square region at five resolutions aggregated with the *WSH Averaging* routine showing building centroid locations in black. To compute local errors, WSH_{s2} grids are downscaled to $s1$ then WSH_{s1} is subtracted, yielding the $WSH_{s2} - WSH_{s1}$ values shown in meters on a red-blue colour scale.

3.6 Conclusions

In this study, we developed the novel *resample case framework* and used it to analytically demonstrate that aggregation through averaging will always lead to the bias of some metric in regions of marginal inundation. While the direction of this bias will depend on the aggregation routine employed and the metric of interest, we show that inundation extents for example will always increase in marginal regions when simple averaging of water depth grids (WSH) is employed, while flood volume will always decrease when water surface elevation (WSE) grids are averaged. We then applied this framework to a 2018 Canadian flood to spatially attribute and quantify bias from aggregation and showed that inundation area doubled when grids were aggregated from a resolution of 1m to 512m. Finally, this case study was extended to show how those regions with assets or buildings are particularly sensitive to this bias. For example, the number of exposed assets *increased* by 1800% (when aggregating from 1m to 512m) while water depth *decreased* — both would have severe implications on the accuracy of a flood risk model.

Our results provide a framework for evaluating, and a basis for selecting routines to aggregate flood hazard grids while managing distortion or bias. For hazard focused studies where unbiased flood volumes are of primary importance, *WSH Averaging* should be pursued. On the other hand, exposure focused studies where unbiased inundation area is more important should employ the *WSE Averaging* routine. Regardless, some trade-offs will always be required as no routine can preserve all metrics of interest. Practitioners should be aware of such biases introduced by their selected aggregation scheme and work to minimize their effect on conclusions drawn from modelling. For example, in cases where large aggregations are required, practitioners should incorporate alternate aggregation routines as part of a sensitivity analysis to understand the significance and acceptability of the aggregation. For studies where high resolution grids are available and aggregation is still desired, a resample case map can be constructed and used to identify marginal regions more sensitive to error and evaluate the acceptable level of aggregation for the study objectives. Regardless, studies wishing to minimize bias or artifacts should similarly minimize the scale or extent of aggregation, as our results show bias from aggregation can be severe. To support technical implementation of flood hazard grid aggregation and evaluation, an open-source QGIS script was developed as part of this study (`https://github.com/cefect/FloodRescaler`).

The results presented here for the exposed domain show a positive or neutral bias of the six metrics of

interest overall, similar to flood risk model comparison studies (Table 3.1) and a growing body of work on hydrodynamic models (Banks et al., 2015; Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021). While our work stops short of computing risk or impact metrics, the remarkable four-fold increase in exposed assets we find provides a logical, albeit partial, explanation for the bias reported by these studies. Further, we show how the affinity of assets for high ground leads to a systematic over prediction of exposure at coarse scales. Counter to this, we can imagine how hydrodynamic models may miss small channels completely at coarse scales, underestimating flood extent in some areas. Considering this, our findings, and those of similar studies, are likely *somewhat* sensitive to the study area and the flooding mechanism, but *especially* sensitive to the magnitude of the scale or resolution transfer. Regardless, a more comprehensive understanding of these competing biases is needed to fully explain the positive bias of coarse flood risk models found by numerous studies (Table 3.1).

Of equal importance, but not addressed here, is work to understand the role of asset aggregation on flood risk model bias. This longstanding and common practice (Hall et al., 2005; Jongman et al., 2012; Sairam et al., 2021) involves aggregating assets and their attributes, intersecting with hazard grids, then applying these as inputs to damage functions developed for single assets. To attribute and correct for bias which may emerge through such scale transfers, the frameworks and findings developed here could be extended. By studying issues of scale, the accuracy and applicability of large or global flood risk models can be improved — allowing society to better prepare and plan for disasters.

Acknowledgements

The authors thank Jody Reimer, Kai Schröter and Lukas Schoppa for their feedback, suggestions, and discussion. The research presented in this article was conducted within the research training group “Natural Hazards and Risks in a Changing World” (NatRiskChange) funded by the Deutsche Forschungsgemeinschaft (DFG: GRK 2043/2). Open Access funding enabled and organized by Projekt DEAL.

Chapter 4

AGGREGATING FLOOD DAMAGE FUNCTIONS: THE PERIL OF JENSEN'S GAP

Bryant, S., Reimer, J., Kreibich, H., and Merz, B.: *Aggregating Flood Damage Functions: the Peril of Jensen's Gap*, [accepted], Journal of Flood Risk Management, 2024.

Abstract

Flood risk models provide important information for disaster planning through estimating flood damage to exposed assets, such as houses. At large scales, computational constraints or data coarseness often lead modellers to aggregate asset data using a single statistic (e.g., the mean) prior to applying non-linear damage functions. This practice of aggregating inputs to non-linear functions introduces error and is known as *Jensen's inequality*; however, the impact of this practice on flood risk models has so far not been investigated. With a Germany-wide approach, we isolate and compute the error resulting from aggregating four typical concave damage functions under 12 scenarios for flood magnitude and aggregation size. In line with Jensen's 1906 proof, all scenarios result in an overestimate, with the most extreme scenario of a 1 km aggregation for the 500-year flood risk map yielding a country-wide average bias of 1.19. Further, we show this bias varies across regions, with one region yielding a bias of 1.58 for this scenario. This work applies Jensen's 1906 proof in a new context to demonstrate that *all* flood damage models with concave functions will introduce a positive bias when aggregating and that this bias can be significant.

4.1 Introduction

With the increase in flood related disaster damages, the expansion of computation power, and the availability of global data, development and application of meso- and macro-scale flood risk models have increased in the past decade (Ward et al., 2020). These models are composed of a series of sub-models for the flood hazard, exposure of assets, and vulnerability to flooding, where vulnerability modelling, the last step in the chain, is generally found to be the most uncertain component in micro- and meso-scale models (de Moel and Aerts, 2011; Jongman et al., 2012). These findings are supported by work comparing modelled damages to those observed during flood events, where large discrepancies are often found between different models and actual observations (Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020). Further challenges we elaborate below are introduced when such models are transferred to the macro-scale, where many exposed assets are aggregated through averaging into a single unit before vulnerability models are applied (Hall et al., 2005; Ward et al., 2020; Sairam et al., 2021). This process collapses heterogeneities, for instance in the flood depth, within the aggregated unit and generally results in overestimation of the total flood damage (Bryant et al., 2023).

In a flood vulnerability model, flood damage functions (f) provide a mathematical relationship between hazard and vulnerability variables (e.g., flood depth) and the estimated damages from flooding (e.g., building repair costs) for an individual asset (White, 1945). The most basic functions directly relate flood depth to damage — so-called depth-damage curves widely attributed to White (1945). Damage functions are typically categorized based on the model's focus or objective, such as the sector (residential vs. non-residential), tangibility (tangible vs. intangible), damage mechanism (indirect vs. direct), and uncertainty treatment (deterministic vs. probabilistic) (Merz et al., 2010). Further classification considers function

structure such as continuity (discrete vs. continuous) and for tangible economic functions the asset total value relation (relative vs. absolute) (Merz et al., 2010). Gerl et al. (2016) provide a comprehensive review of 47 flood damage functions with the majority deterministic (96%), multi-variable (88%), and expressed loss relative to the total value of the asset (56%). To provide a standardized library of these functions, each function was standardized to a common set of indicator variables, while unique indicators were left as default values. After this standardisation, Gerl et al. (2016) found significant heterogeneity in function shape and magnitude.

Aggregation and scaling issues in flood damage models have only recently been considered systematically (Sieg et al., 2019). Sieg et al. (2019) developed a stochastic framework to seamlessly compute damages at different spatial scales with consistent uncertainty accumulation. For this, a gamma probability distribution was fit to hazard, vulnerability, and exposure variables using a mix of survey data, statistics data, and simulated water levels. From these distributions, 300 samples were drawn to populate assets within the damage model. Results were then aggregated (e.g., total damages within municipalities) together with the accumulated uncertainty. Total damage was then compared to that of an earlier study with 100 m gridded asset aggregation (Seifert et al., 2010b) to show the aggregated study overestimated. Using a more traditional approach, Pollack et al. (2022) developed models for 800,000 single family dwellings and eight flood scenarios from the Fathom US Flood Map product. When only building attributes were aggregated, annualized damage was slightly underestimated (-10%), but when hazard variables were also aggregated, there was a large overestimate (+366%). Focusing on flood hazard grids, Bryant et al. (2023) showed mathematically that common aggregation schemes will always positively overestimate inundation area. We are not aware of any studies specifically investigating the role of aggregation on non-linear damage functions or Jensen's inequality.

The objective of this paper is to explain, demonstrate, and quantify the effects of spatially aggregating non-linear damage functions. In other words, we seek to answer the question of how significant Jensen's inequality is for flood damage models.

4.2 Jensen's Inequality

Scaling and averaging issues are not unique to flood damage models. Many fields find it convenient (or even necessary) to simplify the system under study by averaging or aggregating some variable or computational unit (Denny, 2017). However, the assumption that averaging does not affect system response is incorrect for most real-world system; a conundrum, widely called "Jensen's inequality" (Jensen, 1906) or "the fallacy of the average" which can be formalized as:

$$\overline{g(x)} \neq g\left(\overline{x}\right) \tag{4.1}$$

where g is a non-linear function and x is an independent variable. Applied to flood vulnerability models, which are generally non-linear (Gerl et al., 2016), Equation 4.1 implies that aggregating or averaging assets (e.g., buildings) introduces errors when those aggregate values are used as inputs into a vulnerability model. This difference between modelling entities on an individual basis versus in aggregate is called *Jensen's gap* (see Figure 4.1). Jensen (1906) proved that the magnitude and direction of this gap is related to the variance of the independent variable σ_x^2 and the local shape of the function $g''\left(\overline{x}\right)$. Convex functions result in positive gaps or an underestimating aggregate model $\overline{g(x)} > g\left(\overline{x}\right)$ while concave functions result in negative gaps or an overestimating aggregate model $\overline{g(x)} < g\left(\overline{x}\right)$. The limits of the gap size can

only be determined for certain types of problems (Liao and Berg, 2018; Walker, 2014), such as when *x* follows a mean centred distribution (Gao et al., 2017), or when *g(x)* has a Taylor expansion (Abramovich and Persson, 2016), but are difficult to determine for other classes of problems. In summary, any model containing non-linear functions and variance within the independent variable will introduce some bias when measurements of that variable are aggregated.

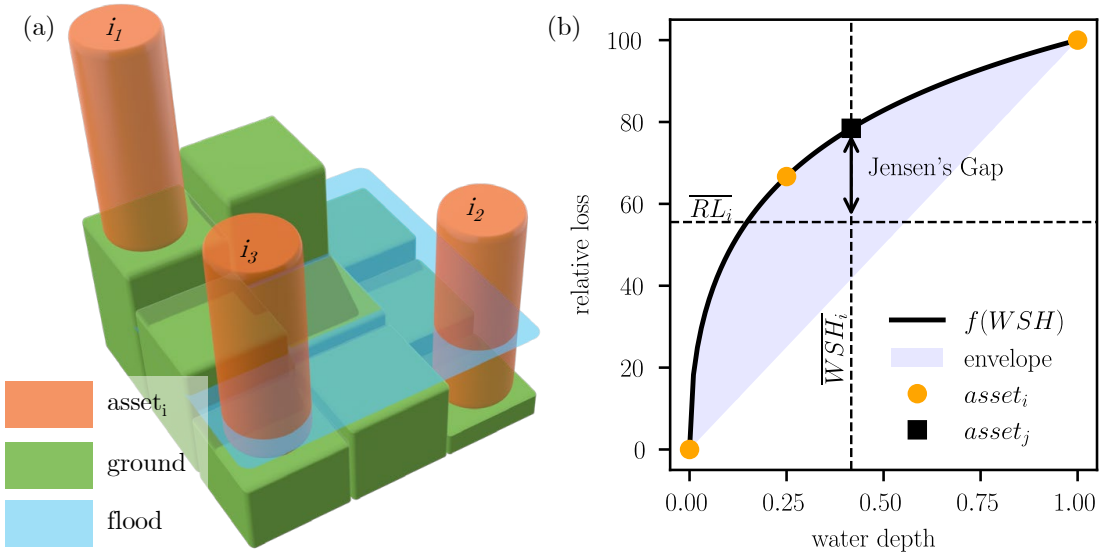


Figure 4.1: Illustrative example of *Jensen's gap* applied to an idealized flood damage function. Panel (a) shows an example of three base or child assets *i* (e.g., buildings) which are to be aggregated into a single parent asset *j* (not shown). The child assets (*i)* are exposed to the flood depths plotted on the x-axis of panel (b), the average of which ($\overline{\mathbf{WSH_i}}$) is taken as the exposure of the aggregate asset (*j)* (black dashed vertical line). Panel (b) also shows the relative losses resulting from the three child assets and the single aggregate asset on the y-axis. Finally, the mean of the child asset losses ($\overline{\mathbf{RL_i}}$; horizontal black dashed line) is compared to the loss of the aggregate asset to demonstrate and quantify Jensen's gap. See later text for discussion of the *envelope*.

4.3 Method

To evaluate the sensitivity of flood damage models to Jensen's inequality, a simulation experiment is used to compare models which differ only in the level of asset aggregation. For this, three different grid sizes are used to construct aggregate models for four direct, tangible, flood-damage functions resulting in a total of 12 scenarios or experiments (3 grid sizes x 4 damage functions). These experiments generate a large spatial dataset of exposed flood depths, relative losses, and total losses (using homogenous replacement values) for the 12 scenarios from which our analysis explores the significance of Jensen's gap on flood damage models.

For this study, we focus on direct tangible economic functions for estimating the relative loss to residential buildings from flood depth or water surface height (WSH) in Germany. While Jensen's gap is a mathematical phenomenon, and therefore indifferent to damage function categories like tangibility (tangible vs. intangible) or damage mechanism (indirect vs. direct), we focus on direct tangible economic functions as this category is the most common (Merz et al., 2010) and has the longest history of application (White, 1945) and therefore provides the most clear and relevant examples for demonstrating Jensen's gap. From the standardized function database of Gerl et al. (2016), three such flood damage functions

were obtained. A fourth more recent function developed from private-household surveys in Germany (Wagenaar et al., 2018) was also pre-processed in a similar manner and included. The complete collection of the database from Gerl et al. (2016) is shown in Figure C.4. All of these are in tabular format and have been simplified as univariate functions of depth by selecting default variable values for the multi-variate functions (FLEMO and BN-FLEMOps). In this form, these functions can be implemented in a damage or vulnerability model as:

$$TL = \sum_{j}^{n} RL_j * V_j \tag{4.2}$$

where TL is the total loss or damage of the flood event, V_j is the replacement value of asset j, and RL_j is the relative loss defined as:

$$RL_j = f\left(WSH_j\right) \tag{4.3}$$

where f is the univariate damage function of the flood depth or water surface height (WSH) at asset j. The two modern and two legacy damage functions selected for this study are summarized in Table 4.1. These were developed from post-flooding records or surveys of damages to individual buildings in Germany, except for IKSE for which the documentation is unclear. The legacy damage functions, IKSE and MURL, were intended as meso-scale damage functions, to be applied on the spatial basis of land use areas; however, how the transformation from per-building records was performed was not documented. The more modern functions, FLEMO and BN-FLEMOps, were intended as micro-scale damage functions, to be applied on the spatial basis of buildings or structures. When flood damage *models* are formulated for the *macro*-scale, it is common to group many buildings into a single computational unit, i.e., aggregate many underlying child-assets i into a single parent-asset j (see Figure 4.1 for illustration). Importantly, while this aggregation implies some transformation of the independent variable WSH (from i to j), neither in practice nor in our study is a similar transformation applied to the damage function f. It is this mismatch, or the application of child-asset derived damage functions to aggregate assets, which is the focus of our study.

Table 4.1: Selected flood damage functions

name	full name	spatial design basis	envelope (cm*%)[1]	ref.
FLEMO	Flood Loss Estimation MOdel	per-building	8978	Thieken et al. (2008)
IKSE	Internationale Kommission zum Schutz der Elbe	per-building	14500	Elbe (2003)
MURL	Ministerium für Umwelt, Raumordnung und Landwirtschaft des Landes Nordrhein-Westfalen	land use	2500	Nordrhein-Westfalen (2000)
BN-FLEMOps	Bayesian Network Based Damage Model for the private sector	land use	6566	Wagenaar et al. (2018)

[1] Envelope of potential error as in Figure 4.1. See Figure C.1 for calculation.

To exclude other artifacts of aggregation that may impact model behaviour (e.g., raster aggregation, exposure sampling), we employ an experimental damage model designed to estimate aggregation error, rather than an applied model designed to estimate flood damage. Our experimental model has two key differences from a more traditional applied model. First, to eliminate any confounding influences from asset categorization, a single damage function is applied in each scenario to all assets, rather than categorizing and applying functions heterogeneously to assets based on some exposure variable (e.g., building use) as is typically done in applied models. Second, the experimental model focuses on relative

losses rather than total losses, thereby eliminating the influence of heterogeneous building replacement values; however, a proxy for total loss using homogeneous replacement values is also provided.

To construct the non-aggregated base-asset (i) control model, building centroid points are extracted Germany-wide from OSM (OpenStreetMap contributors, 2015) for building footprints with area greater than 50 m^2. Flood depth grids are taken from Fathom Global Flood Map 3.0 product with 1 arcsecond (30m) resolution and four fluvial undefended return period scenarios (10, 50, 100, and 500-year) (Fathom, 2023). Relative losses are computed from each of the selected damage functions using linear interpolation for explicitly tabulated WSH-RL pairs, and the maximum RL for exceeding WSH values. All functions are applied on the same datum as the hazard scenario (i.e., it is assumed that the functions start at the terrain surface of the hazard model).

To construct the aggregate (j) models, a square grid is generated for the full Germany-wide domain for three grid sizes: 60, 240, and 1020 m. These three sets of grid polygons are then spatially joined to the building centroids to create a lookup table such that each building is assigned a single grid cell in each of the three grid sizes (i.e., 1 parent grid cell is assigned to many child buildings). Aggregate assets with no child building exposure, i.e., where no buildings are flooded by the most extreme scenario, are then purged from the table yielding varying spatial extents and total building count between the different grid sizes. This lookup table is then used to compute the aggregated asset statistics like the number of buildings or the aggregate flood depth (WSH_j).

To obtain the flood depth on each aggregated asset (WSH_j), a few options are available: 1) a zonal statistic (e.g., mean or mode) of raster depth values within each cell; 2) the raster depth value sampled at the grid polygon centroid; or 3) a statistic of child building depth values within each cell. The first option has been used by some studies; however, Bryant et al. (2023) have shown this introduces some bias that would confound our study of function aggregation — especially in those regions with buildings. Similarly, the second option of centroid sampling also introduces substantial bias as assets and water depths are not evenly distributed across the grid cell. For example, some of our early experiments suggested this centroid sampling generated more bias than the actual aggregation or Jensen's gap. Considering this, we adopt the third option and implement the child depths mean (mean of depths at buildings within an aggregate asset) as the independent variable for each aggregate asset. Obviously, this is impractical for most aggregated applied models as it requires the underlying building location data; however, this aggregation strategy is the only that facilitates isolating the effects of function aggregation and was therefore selected for our study.

The complete data processing pipeline (Bryant, 2023) is written primarily in python and uses PostGIS for most data operations (PostGIS Project Steering Committee, 2018) and Whitebox Tools for raster sampling (Lindsay, 2014).

4.4 Results and Discussion

This section presents results in a sequence similar to the workflow of a classical damage model. First, we start with the child depth statistics which provide the independent variable for the damage functions, then we investigate relative loss effects by intersecting these depths with the selected damage functions, and finally a brief analysis of potential aggregation errors in total loss is presented.

4.4.1 Depth Variance

Considering the variance (σ_2) of the independent variable (WSH_i) are directly related to Jensen's inequality (Denny, 2017), we expect loss calculations on aggregate assets with non-negligible child depths variance to yield some error (when compared to the loss calculated from the child asset depths directly). To show this variance within our aggregations, Figure 4.2 presents the statistics for the child depths for the four hazard scenarios (rows) and three aggregation sizes (columns). Focusing first on the child asset (i) exposure counts (i.e., the total number of buildings within the exposed domain), this plot shows an increase with both hazard severity and grid size, reflecting the increasing footprints of these scenario dimensions. The aggregate (j) exposure counts on the other hand show a similar trend with hazard severity but decrease in count with the larger grid sizes. Aggregated water depths (WSH_j) follow this same trend: the larger grid sizes capture more dry assets thereby reducing the mean.

Looking from left-to-right shows that, with one exception, variance increases with increasing grid size as expected: the greater the number and source area of child buildings, the more likely their depths are to vary. A similar increasing pattern is observed looking from top-to-bottom, where more extreme flood scenarios increase the variance (and depths). While this is less intuitive, we hypothesize this result is caused by the ratio of small-steep river channels activated in the domain. The lower magnitude events are dominated by exposure within broad-flat floodplains while these events do not result in exposed building in the smaller rivers where terrain gradients are more severe. Regardless, Figure 4.2 demonstrates that most aggregate assets have sub-meter standard deviation ($Q_{0.75}[\sigma] < 50$ cm).

Spatially, child depths variances are heterogeneous. Figure 4.3 shows a map of standard deviations for the most extreme 500-year hazard scenario. The regions of high variance correspond to river sections with development on both steep and flooded terrain. From this map, we conclude that sensitivity to Jensen's inequality can differ substantially between regions.

4.4.2 Relative Losses

With the pattern and magnitude of child depths demonstrated, we now feed these into the selected damage functions to compute non-aggregated and aggregated relative losses. By comparing these, the distribution of errors can be quantified as shown in Figure 4.4 for the most extreme hazard scenario where each function is overlaid on the mean loss computed from each child asset ($\overline{RL_{bldg,j}}$). In other words, Figure 4.4 shows the difference between what the aggregate model would calculate ($f(WSH)$) and what a model that included individual buildings would yield. Rather than focus on the absolute value of the computed errors, which are sensitive to the study domain, homogeneous building categorization, and application of the legacy functions at micro-scale, Figure 4.4 is used to understand error trends and behaviour caused by function shape and grid size. Examining general trends shared by all loss functions by looking from left-to-right (a-c) we see both the density plot and the WSH-binned mean line (blue dashes) show a widening gap between per-asset losses and aggregate losses as the grid sizes increases. This follows from Figure 4.2 which shows the same trend on variance. This difference has a similar pattern for all 12 scenarios: for aggregate depths greater than ~200cm, the aggregate relative loss (solid black line) overestimates the per-asset loss (dashed blue line) substantially. This is also reflected in the differences between the population means which show the aggregate ($\overline{RL_{grid,j}}$) overestimating the per-asset ($\overline{\overline{RL_{bldg,j}}}$) losses for all scenarios (with a bias up to 1.19). The extremes are also noteworthy, where we see some aggregate losses are more than double their child-asset mean-loss counterpart.

Comparing across functions, Figure 4.4 shows those with a greater RL range (e.g., function IKSE's

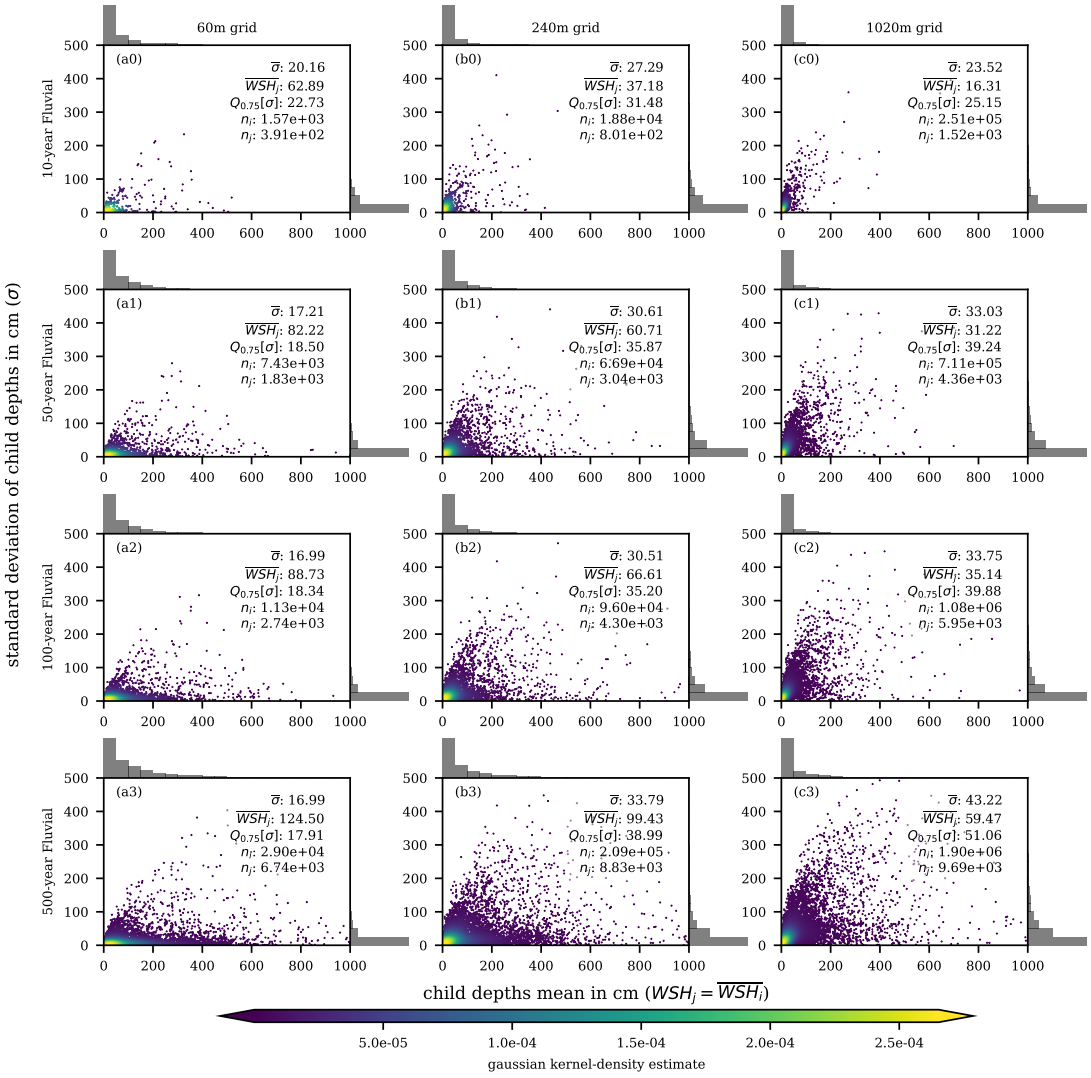


Figure 4.2: Germany-wide child depths mean and standard deviation for three levels of grid-aggregation and four hazard scenarios. Histograms show the distribution of their respective axis. Text in each panel provides the mean ($\overline{\sigma}$) and 0.75 quantile ($Q_{0.75}[\sigma]$) of the child depths standard deviations. The mean of the child depths means ($\overline{WSH_j}$), and the total counts of buildings (n_i) and aggregated assets (n_j) are also provided. Only those aggregate assets with at least 2 child assets are included.

RL ranges from 0 to 40%) yield larger magnitudes of differences and root mean square errors (RMSE) between the aggregate and per-asset losses. To understand this, consider an envelope of potential error bounded by *f(x)* and a line from *f(0)* to *f(x_{max})* (see Figure 4.1 and Figure C.1) where x_{max} is 1000 cm for our hazard scenarios. The greater the area of this envelope the more sensitive the function will be to variance in child depths values. This envelope can also be thought of as a measure of concavity, which yields a negative Jensen gap (Denny, 2017). Because all selected functions are concave (downward curvature), the aggregate functions all overestimate, except for a small portion of the FLEMO function between 0 and 175 cm which is slightly convex (upward curvature). For another example of how curvature relates to this aggregation error, Figure C.2 provides a similar analysis on three idealized functions that

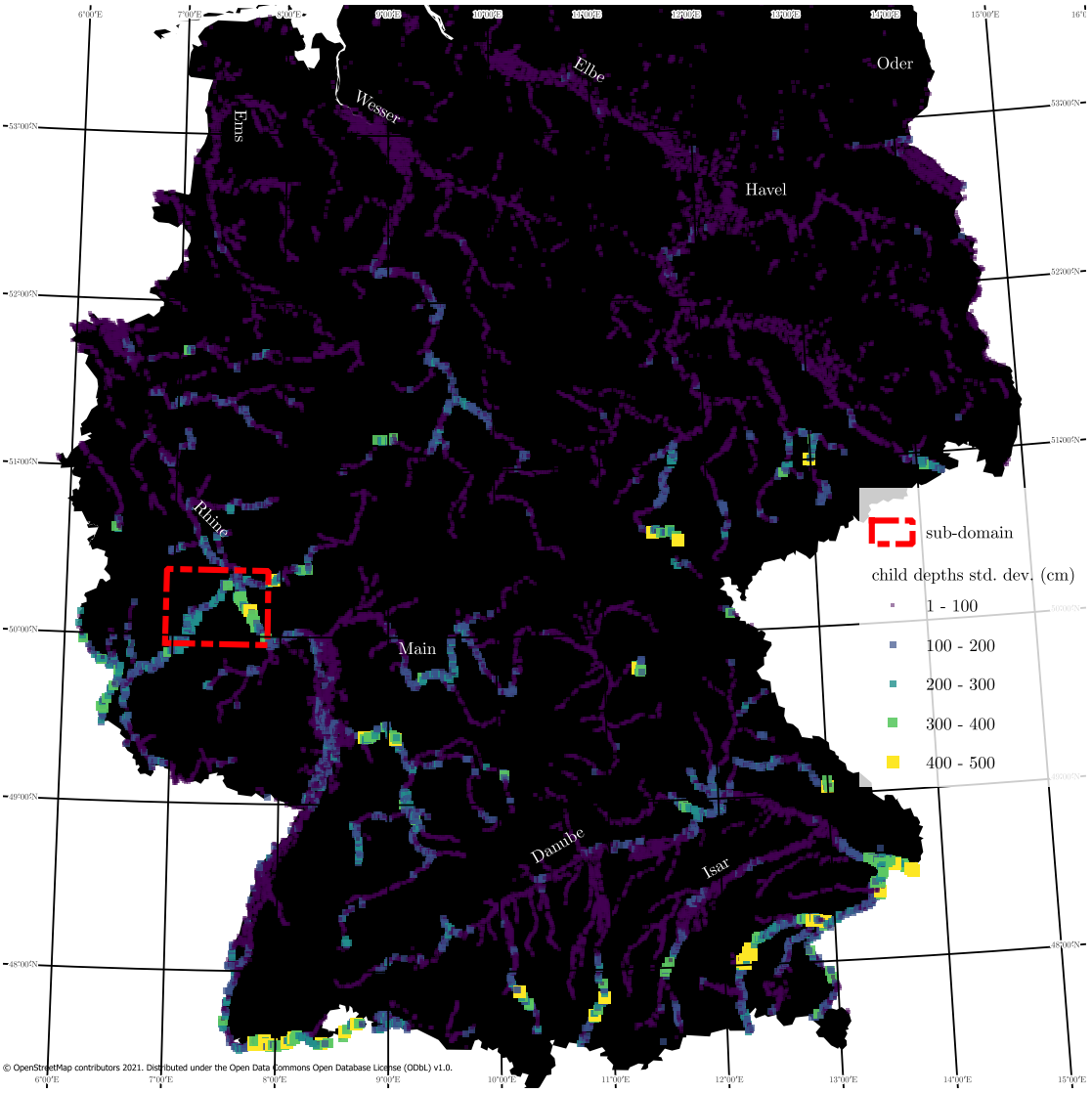


Figure 4.3: Map of major rivers in Germany and standard deviations of child depths for the 500-year hazard scenario with an aggregation grid size of 1020 m. Only those aggregate assets with at least 2 child assets are shown. See text for discussion of sub-domain.

differ only in their curvature parameter, showing the same dependence of error on curvature. Figure C.4 shows there is some diversity of function shape within (Gerl et al., 2016)'s database; however, most have similar concavity and therefore we expect similar overestimation patterns. In summary, Jensen's gap can be significant for a Germany-wide model under the most extreme scenarios; however, Figure 4.3 suggests there may be some local regions where the gap is more pronounced.

To demonstrate the significance of Jensen's gap for a local region with high child depth variance, the sub-domain around Koblenz as shown in Figure 4.3 was selected for further analysis. This area is hydraulically complex encompassing the confluence of the Rhine, Moselle, and Lahn rivers all confined to relatively narrow valleys with dense development along the banks. The mean child depths variance for this region is 1.5 to 4.0 times higher than the Germany-wide averages (see Figure C.3), suggesting Jensen's gap may be more severe here. To illustrate this, Figure 4.5 shows the aggregated and non-aggregated relative losses

for a single example function. Compared to the Germany-wide values, RMSE more than doubles and the bias more than triples (relative to perfect).

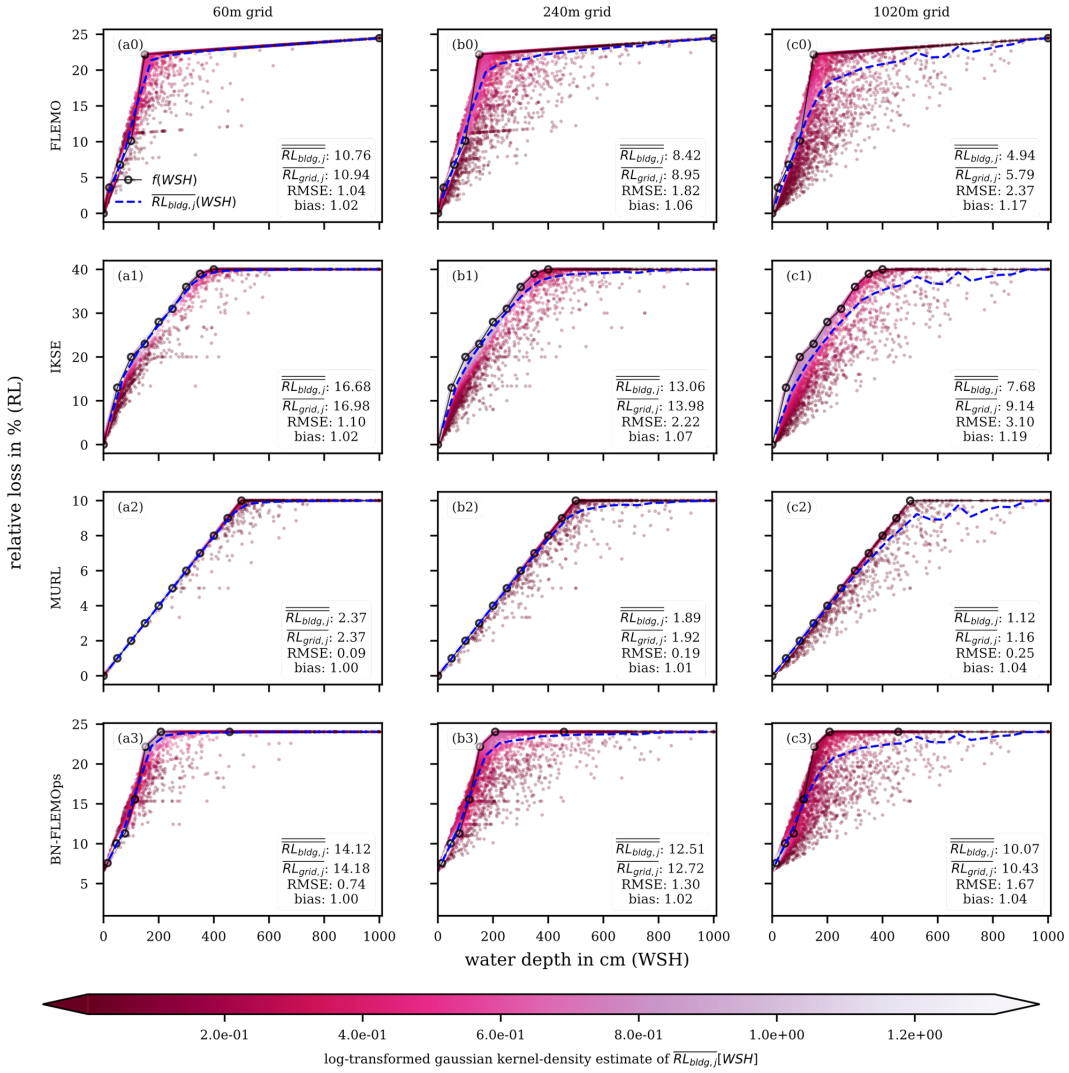


Figure 4.4: Relative losses for three grid aggregations for the 500-year fluvial undefended hazard scenario and the four selected flood damage functions (Table 4.1). Child relative loss means (mean of relative loss of each building within a grid cell) ($\overline{RL}_{bldg,j}$) vs. WSH_j are shown as a density scatter plot. The blue dashed line shows the same $\overline{RL}_{bldg,j}$ values for each aggregate asset with the mean taken again by binning on WSH values. The black dashed line shows the original damage function. On the right-hand-side of these panels the population mean of losses computed from aggregate ($\overline{RL}_{grid,j}$) and child relative loss mean ($\overline{\overline{RL}_{bldg,j}}$) are shown. Bias is computed as $\frac{\overline{\overline{RL}_{bldg,j}}}{\overline{RL}_{grid,j}}$. Only those aggregate assets with at least two child buildings are included.

By implementing a Germany-wide approach, we trade some marginal uncertainty in the absolute RL values of each scenario for breadth and consistency across functions. The two most significant marginal uncertainty sources being the application of residential functions to all buildings and the application of per-building centroid depths (WSH_i) to all functions. The first assumes that the spatial distribution of

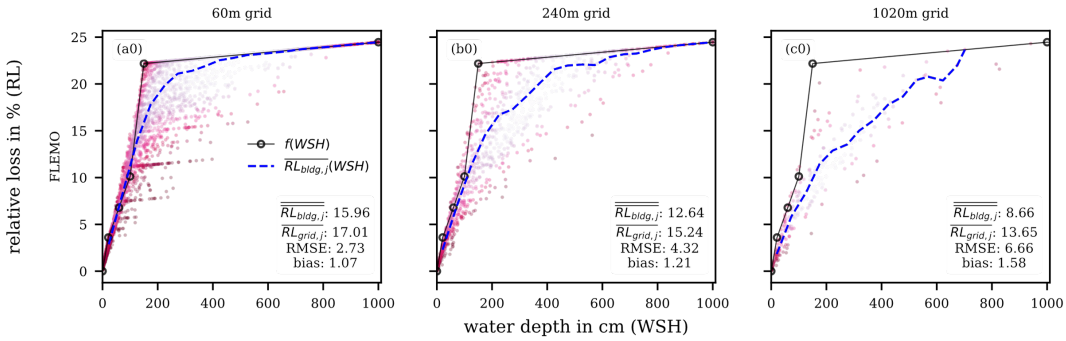


Figure 4.5: Relative losses for a single damage function for the sub-domain shown in Figure 4.3. See Figure 4.4 for additional explanation and legend.

residential and non-residential buildings is roughly equivalent, which likely is untrue in some regions. The second assumes WSH_i is uniform across the footprint of the function's spatial design basis, i.e., for the per-building modern functions WSH_i is assumed uniform across each building footprint and across each land use block for the meso-scale legacy functions. Considering the 30m resolution hazard layers, uniform WSH_i across land use blocks, which are generally 100m square, is untrue in many regions as shown in Figure 4.2. For the two meso-scale legacy functions, this means the RL errors reported in Figure 4.4 (row 1 and 2) have an additional application uncertainty roughly equivalent to the 60m computation (panel a1 and a2). However, considering the magnitude of this uncertainty is an order of magnitude less than the errors computed for these damage functions by Thieken et al. (2008), this application uncertainty is inconsequential to our analysis of aggregation behaviour and an acceptable trade-off for allowing the inclusion of these well-known damage functions in the analysis. Finally, readers should note that even for the extreme cases computed here, the relative loss errors from aggregation are less than those from the initial construction or fitting of the empirical functions themselves. For example, BN-FLEMOps has a Mean Absolute Error around 15% RL (Wagenaar et al., 2018) and FLEMOps 24% RL (Thieken et al., 2008) when comparing modelled to observed losses.

4.4.3 Total Losses

The above analysis, while useful in demonstrating and understanding the mechanics of Jensen's inequality as it applies to flood damage functions, has limited applicability to damage models due to the lack of replacement values (V_j; see Equation 4.2). For example, densely developed regions (e.g., cities) may have different exposure or hazard patterns than less developed regions (e.g., rural areas). These density-dependent effects are obfuscated when the aggregate asset RL is simply compared to the child RL mean (i.e., all aggregate assets are treated equally regardless of their child building counts) as in the previous section. As a proxy to quantify aggregation errors with density dependent effects, we extend the above analysis by include all assets and multiplying the relative losses calculated for the aggregate asset by its building count (B_j) as shown in Figure 4.6.

Looking from left to right again, Figure 4.6 shows the same increase in bias and RMSE with grid size as for the relative loss values in Figure 4.4, as well as the same relative magnitudes between functions. However, the total bias increases for six of the 12 cases with two cases decreasing slightly compared to the relative loss values. This suggests some, generally intensifying, density effects on Jensen's gap in

flood damage models. In other words, aggregate assets with more children (e.g., cities) are more sensitive to Jensen's gap than those with fewer children (e.g., rural areas).

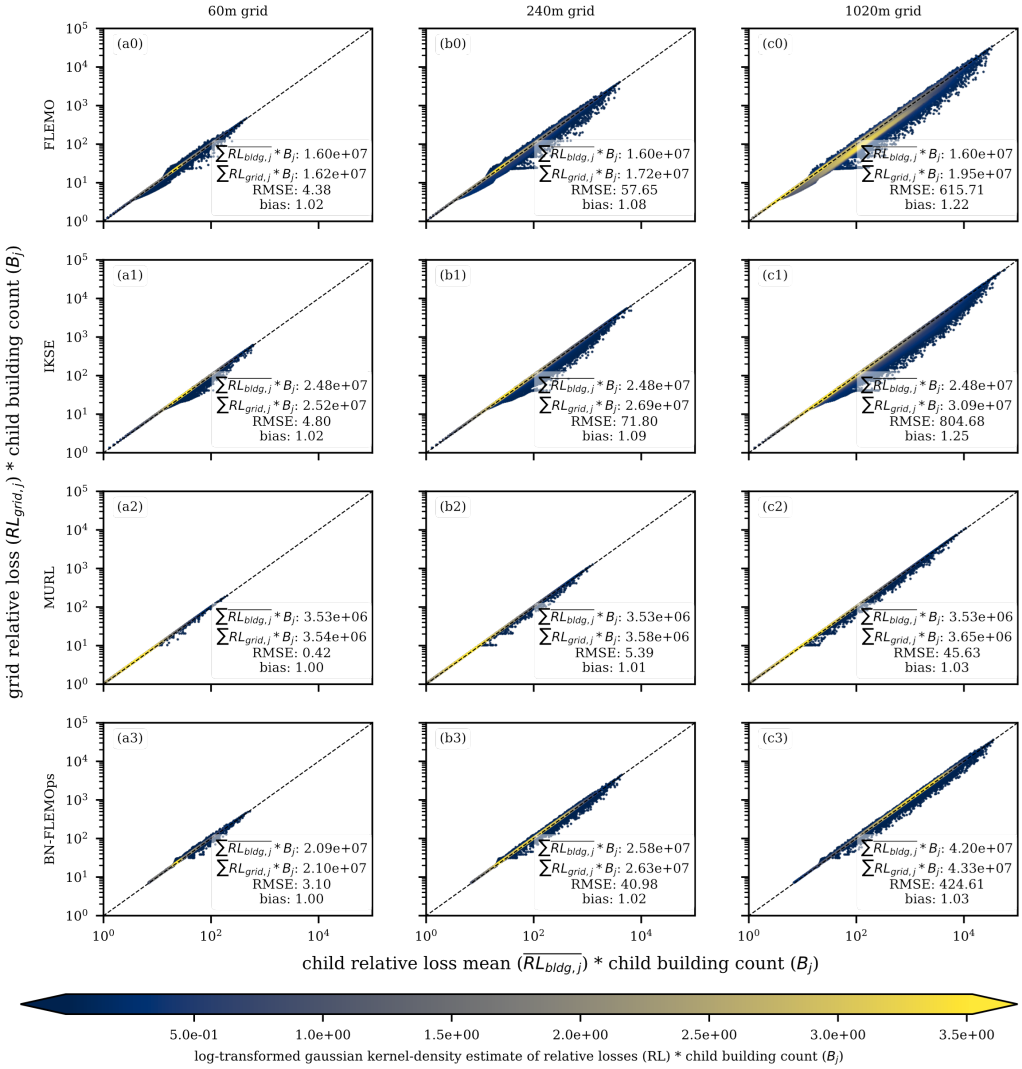


Figure 4.6: Building count weighted relative losses for three grid aggregations and four flood damage functions for the 500-year fluvial undefended hazard scenario. Totals are shown for losses computed with non-aggregated ($\mathbf{\overline{RL_{bldg,j}}}$; x-axis) and aggregated ($\mathbf{RL_{grid,j}}$; y-axis) depths. Bias as in Figure 4.4.

This asset-density study-model serves as a proxy for the sensitivity of applied damage models to non-linear function aggregation errors or Jensen's inequality. In fact, this asset-density study-model is simply a damage model with homogeneous replacement values (V_j), i.e., where each building in the model, regardless of size, costs the same to replace. Because this assumption is false, our asset-density study-model is incapable of yielding total losses; however, when used instead to compare results between scenarios (aggregated vs. non-aggregated in our case) the model can provide a reasonable estimate of the bias of equivalently aggregated damage models. For regions where the variance of replacement values is high and correlated with the hazard within an aggregated asset (e.g., large expensive homes along the river), the estimate provided here is less applicable.

To capture behaviours that emerge at the large scales under which aggregation is commonly implemented, our study domain includes all of Germany. Because of this choice, the resolution of our hazard scenarios is limited to 30 m as this is the highest resolution model at this scale, we are aware of. This coarseness relative to the size of buildings may temper the variance of child depths. In other words, we expect that once higher resolution large-scale models become available, we may learn that the old coarse models overestimate losses even more severely than what is shown here. Similarly, our study only considered univariate damage functions. The aggregation response of multi-variate functions would be similarly sensitive to the spatial variance in the other variables; however, their interactions would be more complex and difficult to quantify.

Focusing on errors specific to non-linear aggregation, our study intentionally excluded other mechanisms that introduce error into aggregated models. For example, the magnitude of overestimate we report is likely tempered by our use of the mean child depth for the aggregate depth given the "dry bias" hypothesis of exposure (Bryant et al., 2023). This hypothesis states that, within an aggregate asset, those regions where buildings are present tend to be less exposed (shallower flood depths) than the average flood depth across the asset. In other words, a model using centroid sampling or grid averaging to obtain the aggregate flood depth will overestimate more severely than one using child mean depths like ours. Errors like this, and the non-linear function aggregation reported here, compound and interact with more classical model uncertainties like spatiotemporal transfers and measurement error.

4.5 Conclusion

This study provides the first demonstration and quantification of errors in aggregate flood damage models resulting from the application of non-linear damage functions. We show that variance in flood depths along with function curvature leads to discrepancies between the aggregated and the underlying non-aggregated model. With our Germany-wide analysis, we find that flood depth variance increases with the magnitude of both aggregation and flood hazard. Further, this variance is spatially heterogeneous, with some local regions exhibiting 1.5 to 4.0 times more variance in flood depth than the Germany-wide averages. Computing the error in relative loss estimates resulting from these variances, we find all selected damage functions overestimate when applied in aggregation. The magnitude of this overestimation is highly dependent on the water depth variances and the amount of curvature, varying from a bias (aggregate result / non-aggregate) slightly above 1.00 to 1.19 when averaged Germany-wide. However, these errors arising from aggregation are generally less significant than the uncertainty within the damage function itself. Finally, we extend the analysis to account for heterogeneous asset density. This shows the same pattern of error but with most cases yielding a larger overestimate when asset density is included (bias up to 1.25).

These averaging artifacts were first described by Jensen (1906) and are well known in other fields as *Jensen's gap*. This study transfers this knowledge to the flood damage modelling domain to demonstrate how *all* damage models applying concave damage functions will overestimate when aggregating. Given the prevalence of aggregation in large-scale models and the use of concave damage functions, this study provides some explanation for the overestimation of coarse models reported elsewhere.

To provide a more consistent evaluation of Jensen's gap, our study only quantified aggregate model overestimates for residential tangible, direct-damage functions applied to buildings exposed to fluvial hazards in Germany. However, considering the magnitude of the overestimate in this case, and that even higher overestimates are probable in some situations, suggests all modellers employing aggregation

should be knowledgeable of the mechanics of Jensen's gap. For example, modellers should be aware that increasing curvature and aggregate depth variance will increase errors. Similarly, this study demonstrates one mechanism that leads to superior accuracy of non-aggregate models over aggregate models. Of course, this mechanism is only one way aggregation introduces error, and these aggregation errors are only one of many sources of error present in flood damage models. To better prepare society for flood disasters, more attention and effort is needed to understand and mitigate model errors and biases like the ones reported here.

Acknowledgements

The research presented in this article was conducted within the research training group "Natural Hazards and Risks in a Changing World" (NatRiskChange) funded by the Deutsche Forschungsgemeinschaft (DFG; GRK 2043/2). Thank you to Fathom for graciously allowing us to use their data for this study. We are thankful to all the GFZ section members for their inputs and camaraderie, esp. Knut Günther for making it all work. We are similarly grateful to Karen Lebek for her tireless coordination, the whole NatRiskChange community for their support and friendship, and Frauke Stobbe for making sure we had visas and housing. And I am especially thankful for baby Tarn for making sure we did not work too hard or sleep too much.

Chapter 5

FLOOD DAMAGE MODEL BIAS CAUSED BY AGGREGATION

Bryant, S., Kreibich, H., and Merz, B.: *Flood damage model bias caused by aggregation*, in: Proceedings of IAHS, 9th International Conference on Flood Management (ICFM9), [under review], 2024.

Abstract

Flood risk models provide important information for disaster planning through estimating flood damage to exposed assets, such as houses. At large scales, computational constraints or data coarseness leads to the common practice of aggregating asset data using a single statistic (e.g., the mean) prior to applying non-linear damage functions. While this simplification has been shown to bias model results in other fields, the influence of aggregation on flood risk models has received little attention. This study provides a first order approximation of such errors in 344 damage functions using synthetically generated depths. We show that errors can be as high as 40% of the total asset value under the most extreme example considered, but this is highly sensitive to the level of aggregation and the variance of the depth values. These findings identify a potentially significant source of error in large-scale flood risk assessments introduced, not by data quality or model transfers, but by modelling approach.

5.1 Introduction

With the increase in flood related disaster damages, the expansion of computation power, and the availability of global data, the development and application of meso- and macro-scale flood risk models has increased dramatically in the past decade (Ward et al., 2020). These flood risk models are often conceptualized as a chain of sub-models for the flood hazard, exposure of assets, and vulnerability to flooding; with each step bringing uncertainty (de Moel and Aerts, 2011). Vulnerability modelling, the last step in the chain, is generally found to be the most uncertain component in micro- and meso-scale models (de Moel and Aerts, 2011; Jongman et al., 2012). These findings are supported by work comparing modelled damages to those observed during flood events, where large discrepancies are regularly found between different models and against observations (Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020). Further challenges are introduced when such models are transferred to the macro-scale, where many exposed assets are aggregated through averaging into a single unit before vulnerability models are applied (Hall et al., 2005; Ward et al., 2015b; Sairam et al., 2021). This process collapses heterogeneities within the aggregated unit (like variable flood depth), and poses poorly understood challenges to the accuracy of flood risk models.

Such scaling issues are not unique to flood risk models. Many fields find it convenient (or necessary) to simplify the system under study by averaging or aggregating some variable or computational unit (Denny, 2017). However, this assumption that system response is unaffected by averaging, is *false* for most real-world systems; a conundrum, widely called "Jensen's inequality" (Jensen, 1906) or "the fallacy of the average" which can be formalized as:

$$\overline{g(x)} \neq g(\overline{x}) \tag{5.1}$$

where g is a non-linear function and x is some response variable. Applied to flood vulnerability models, which are almost never linear (Gerl et al., 2016), Equation 5.1 implies that aggregating or averaging assets

(e.g., buildings) introduces new errors. It can be shown that the magnitude and direction of such errors are related to the variance of the response variable (σ_x^2) and the local shape of the function ($g''(\overline{x})$) (Denny, 2017).

Within a flood vulnerability model, flood damage functions (f) provide the mathematical relation between exposure and vulnerability variables (e.g., flood depth) and estimated flood damages (e.g., building repair costs) for a single asset. The most basic functions directly relate flood depth to damage — so-called depth-damage curves widely attributed to White (1945). Gerl et al. (2016) provides a comprehensive meta-analysis of 47 references containing flood damage functions found in the public literature. These functions cover a wide range of geography and sector including 7 continents and 11 sectors for example. The majority of functions identified were deterministic (96%), multi-variable (88%), and expressed loss relative to the total value of the asset (56%). To provide a standardized library of these functions, each was harmonized to an intrinsic or common set of indicator variables leaving other unique indicators as default values (e.g., inundation duration). Following such a harmonization, Gerl et al. (2016) found significant heterogeneity in function shape and magnitude between authors for supposedly similar models. This partly explains the large discrepancies in risk model results reported by others (Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020).

Aggregation and scaling issues are rarely considered in the flood risk model literature. Jongman et al. (2012) compared results from six aggregated and two object-based asset models against observed damages from a 2002 flood in Germany and a 2005 flood in the U.K.. Asset data was developed from a 100m land cover grid, which required adjusting the two object-based models by 61-88% to reflect the average portion of building footprints within each grid cell. Model performance was mixed between the two studies, with the aggregated models over-predicting by a factor of two and object-based models under-predicting by 5% on average for the German case. For the U.K. case, all models under-predicted.

In a recent large-scale study, Pollack et al. (2022) constructed a benchmark and aggregated analog models from roughly 800,000 single family dwellings and eight 30m resolution flood depth grids with return periods ranging from 2- to 500-years. When only building attributes were aggregated, a small negative bias was observed (-10%) while when hazard variables were also aggregated a large positive bias was found (+366%) for annualized damage. Given the spatial correlation of building values and flood exposure found in their study area, they conclude that bias would be difficult to predict ex-ante.

In this paper we summarize work to improve our understanding of the effect of aggregation on a particular component of flood risk models: the flood damage function. We accomplish this by producing a first order approximation of the potential aggregation error for a general flood risk models, the first attempt of its kind we are aware of. To provide as broad an evaluation as possible, a library of 344 damage functions are evaluated against a single indicator variable: synthetically generated flood depth. These results are then analyzed to elucidate the potential significance and behaviour of aggregation on flood risk models.

5.2 Methods

To evaluate the sensitivity of flood damage functions to the aggregation of input variables, a library of 344 damage functions are evaluated against synthetically generated water depths at various levels of aggregation. Statistics describing the difference between raw function outputs and the aggregated analogues are then computed on each function and each level of aggregation to describe the response of the function to aggregation.

Table 5.1: Summary of flood damage models from Gerl et al. (2016) showing the number of functions (f) per model.

model id	name	f count
3	FLEMO	242
4	IKSE	2
6	MERK	5
7	MURL	2
12	Neyshabouri (2012)	2
16	HAZUS-MH	35
17	MCM	17
23	Luino et al. (2009)	1
27	Toth et al. (2008)	6
37	Budiyono et al. (2015)	24
42	Dutta et al. (2003)	3
44	DSM	5

5.2.1 Flood Damage Functions

For this study, we focus on direct tangible economic functions for estimating the relative loss to buildings from flood depth. From Gerl et al. (2016), 12 such models were identified, each composed of a collection of flood damage functions (f) with the primary independent variable as depth. To simplify the analysis, each of these is discretized by any secondary variables (i.e., variables other than flood depth) to produce a collection of 344 simple functions for relative damage from flood depth as summarized in Table 5.1. Each of these functions is monotonically increasing (i.e., $RL_{i+1} \geq RL_i$) and implemented in python as a lookup table (with linear interpolation applied between discrete depths (x) and their relative loss (RL) pairs). For out-of-range depth values (x) (i.e., those that are greater or less than the most extreme values provided in the lookup tables) either the maximum relative loss value ($f[x] = RL_{max}$) or a value of zero are returned ($f[x] = 0$) following our understanding of common practice. In other words, the function implementation caps the RL values to match those provided in the lookup table.

5.2.2 Synthetic Depths Generation

Each of the 344 selected discrete functions have a distinct shape, depth domain (from -1.22 to 10 m), and relative loss domain (from 0.0 to 116, with a value greater than 100 indicating damages more severe than the estimated replacement value). In practice, each asset (e.g., house) would have its flood damage calculated with a specific function and the exposure variables (e.g., depth) sampled for each flood event. Under aggregation, multiple assets would be collapsed to a single node prior to function computation. The exposure (and vulnerability) indicators of this collapsed asset are generally computed through averaging of the original, singular assets. To provide a comprehensive analysis of the full range of exposure depths supported by each function, a range of depth values are generated using a normal distribution and then sampled and averaged according to the level of aggregation.

To generate the synthetic depth values, the independent variable domain (depths) is discretized from 0.0 to 2.0 to produce 30 anchor depth values (x_b). Each of these x_b values is used as the average or mean depth under evaluation, with lower values representing shallower floods and higher values (closer to 2.0) very

deep floods. To represent the variance in depths within a group of assets for a given flood, raw synthetic depth values are then randomly generated around each x_b by sampling 2000 points from three normal distributions ($N[\sigma = (0.1, 0.55, 1.0), \mu = x_b]$). Higher variances ($\sigma^2$) represent steeper terrain and water surface gradients, while lower values represent flatter areas like a lowland floodplain. These raw values are then aggregated by randomly splitting the 2000 depth values into N_{12} groups of size $s_j = (1, 2, 5, 100)$ and computing the mean. In this way, 360 arrays of synthetic depth values covering a broad range of flood magnitudes, variance, and levels of aggregation are prepared.

5.2.3 Relative Loss Errors

Using each of the 360 synthetic depth arrays on each of the 344 functions, a set of relative loss arrays ($RL_{x_b,s_j,\sigma,f}$) is computed, each associated with an anchor depth value (x_b), a level of aggregation (s), a standard deviation (σ) and a flood damage function (f). To compute the aggregation error potential of each function under various flood conditions, arrays are grouped by function (f), standard deviation (σ), and aggregation level (s). For each of these groups, the mean of the relative loss values is plotted against x_b. Second, for each of these four series, the area against the un-aggregated result is calculated as:

$$e_{s_j,\sigma,f} = \int_{x=0} \overline{RL[x, s_1, \sigma, f]} - \overline{RL[x, s_j, \sigma, f]} \tag{5.2}$$

where e is the aggregation error potential and the integrals are evaluated using the composite trapezoidal rule, a common method for estimating the area under a curve by discretizing into trapezoids (NumPy Developers, 2022).

5.3 Results and Discussion

Three example damage functions (f) are shown in Figure 5.1 along with the mean and two quantiles (q_{95} and q_5) of the resulting loss values computed under four levels of aggregation against the generated depth values. These three damage functions were selected to demonstrate typical behaviour present in all functions. Looking from left-to-right, Figure 5.1 shows how large variances (σ^2) employed in the synthetic depth generation act to spread the resulting relative loss values. Working in a similar but opposite manner, large aggregations (s) reduce the spread of relative loss values. This is intuitive if we consider that the averaging employed in the aggregation works like a filter to reduce variance.

A clock-wise rotation can also be seen in Figure 5.1, where series with less aggregation (e.g., $s = 0$) have positive/negative deviation from the left/right tails of the raw functions. This can be explained by the treatment of out-of-range depth values (x). As explained above, when a synthetic depth value with negative exceedance is calculated (i.e., less than the minimum (x) provided in the lookup table), a relative loss of zero is returned. Away from these x range limits (e.g., $x_b = 1.0$), the normally distributed synthetic depths yield roughly similar positive and negative residuals from the monotonically increasing functions, which balance during the averaging. On the other hand, closer to the x range limits (e.g., $x_b = 0.0$) the residuals become unbalanced as those x values exceeding the range (e.g., $x_b < 0.0$) lose their relation with RL while those within range maintain the relation. At the low end of the range (e.g., $x_b = 0$), this phenomena produces a bias negatively related to aggregation and positively related at the high end (e.g., $x_b = 2.0$). This is intuitive if we consider the aggregated synthetic x values (i.e., $s > 1$) have had their variance reduced (through averaging), meaning fewer extreme x values and therefore the RL imbalance discussed above is less severe with aggregation. Because the depth values considered here are generated with a normal distribution, it is likely that some x values for low anchor depths (e.g., $x_b = 0.0$) would be

below ground in a real flood risk model (some negative x values are realistic as buildings are typically elevated slightly above ground). In cases where such ground water exposure is ignored or negligible (the case of most models), the bias around low anchor depths would be smaller than what we report here. A similar, but weaker argument could be made for large x_b values which traces back to the suitability of a normal distribution for representing exposure depths. This important issue was not investigated as part of our first order approximation.

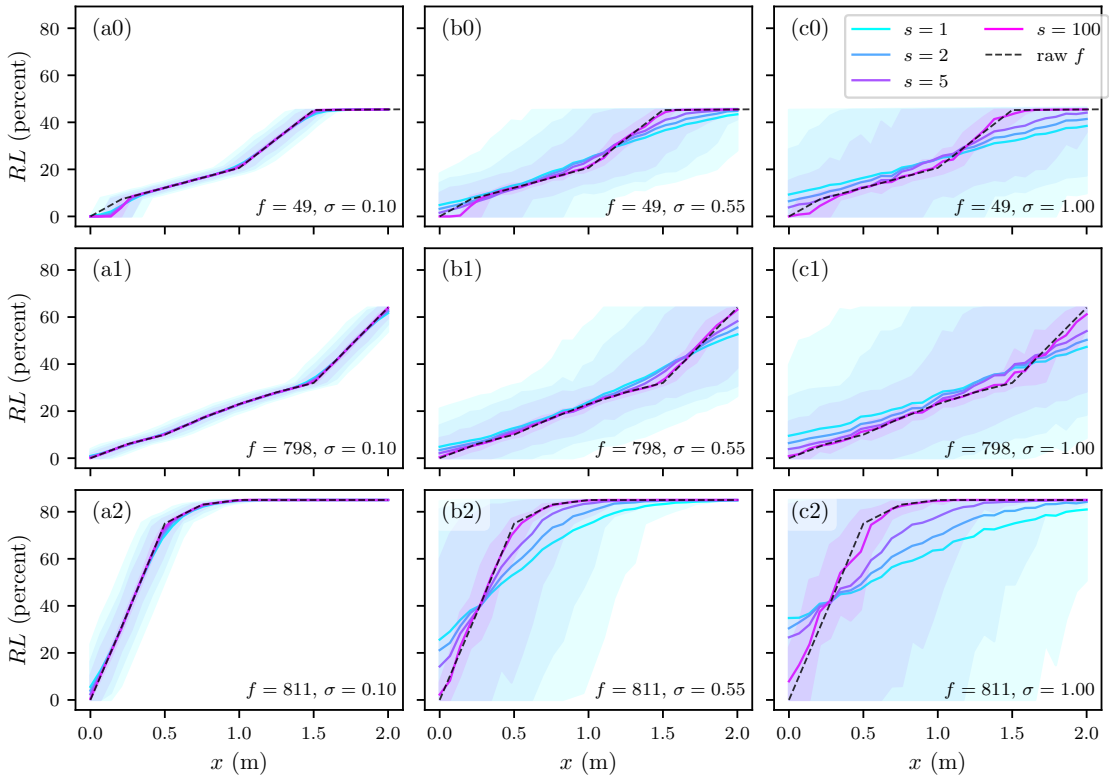


Figure 5.1: Relative loss vs. anchor depth (x_b) for three example functions and three synthetic depth generation standard deviations (σ) computed against the depth values and aggregation (s) described in the text. Shaded areas show the corresponding q_{95} and q_5 for the RL computed for each level of aggregation. Functions are from model 3, 3 and 37 from Table 5.1 for panel a, b, and c respectively.

Figure 5.2 shows the error in relative loss as a function of x_b for two example models from Table 5.1, each containing a set of damage functions. These models were selected to demonstrate behaviour present in all models. This shows that the sensitivity to aggregation errors varies by model or damage function (f). This is intuitive considering the diverse function shapes reported by Gerl et al. (2016). Looking at the general increase in absolute error magnitude from top left to bottom right of Figure 5.2 suggests a positive relation with variance in depth values (σ^2) and the level of aggregation (s). This aligns with our understanding of Jensen's inequality if we consider that larger variances (σ^2) and aggregation both lead to a more severe compression of tail values about the mean. Figure 5.2 also shows that errors can be significant, approaching 40% relative loss for the most extreme case considered here. In contrast, when variance in depth values (σ^2) is low, as in flat floodplains, the majority of functions have relative loss errors less than 5% for the most extreme level of aggregation considered here (s = 100).

Examining the relation of error to x_b, Figure 5.2 suggests generally negative errors for small depths

($x_b < 1.0$) and positive errors for larger depths ($x_b > 1.0$). This can be explained by the treatment of out-of-range depth values (x) and the clock-wise rotation shown in Figure 5.1. The implications of this depth-dependent error could be significant for flood risk analysis aimed at informing flood protection investment decisions where a broad range of flood event magnitudes are assumed to have roughly similar errors (IWR and USACE, 2017). However, the significance of these findings is limited by our first order approach for generating synthetic depths from a normal distribution, which may not represent real exposure depths well.

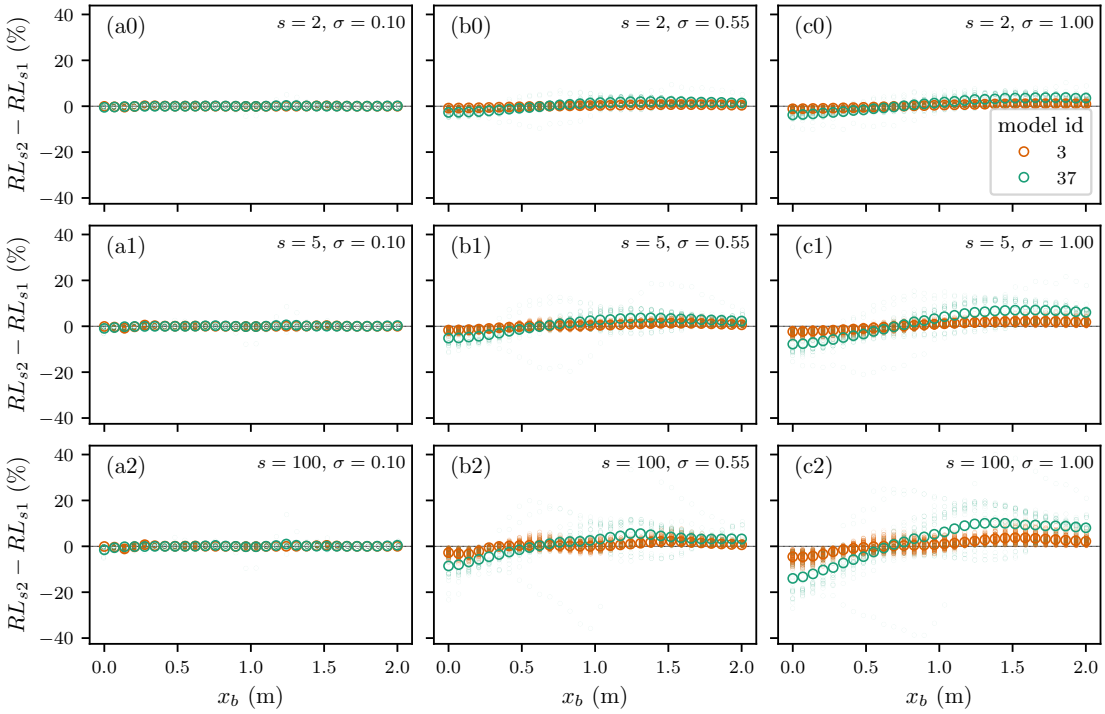


Figure 5.2: Relative loss error for three levels of aggregation (s) and three synthetic depth families (σ) for a selection of the models described in Table 5.1. Faint circles show each function (f) while the bold circles show the function ensemble mean for that model.

To provide a simple metric for the 344 functions covered in Table 5.1, Equation 5.2 was used to compute the total area of each error series (similar to what is shown in Equation 5.2). The resulting aggregation error potentials are shown in Figure 5.3. This shows that the sensitivity to aggregation errors varies widely between models and even between damage functions (f) within the same model. This is intuitive considering the diverse function shapes and the harmonization of variables necessary to directly compare the different models. Further, Figure 5.3 shows the majority of models have mean error (e) values below zero, suggesting an overall negative bias across the full depth domain.

5.4 Conclusion

To better understand the potential and magnitude of errors introduced through averaging of flood risk models, a novel first order evaluation of 344 flood damage functions was performed using synthetically generated depth data. While the character and magnitude of aggregation errors will depend on the specifics of a given flood risk model, the general approach applied here demonstrates that low-depth floods tend to have negative errors while high-depth floods have positive errors. Further, we demonstrate that overall,

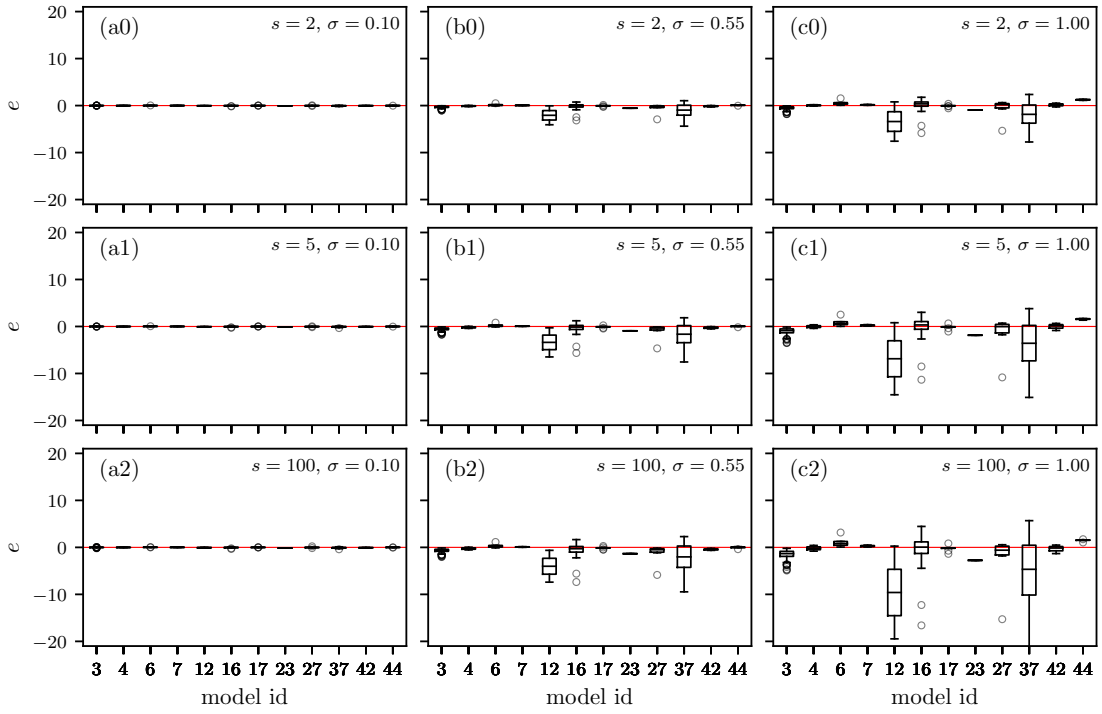


Figure 5.3: Aggregation error potentials calculated with Equation 5.2 for each function (f) of the models described in Table 5.1.

error tends to be negative for the 344 damage functions considered.

The findings reported here provide useful information for flood risk modellers evaluating the appropriateness and extent of aggregation to include in their models. For example, in areas with high depth variance where models with large aggregation error potential (e on Figure 5.3) are to be applied, aggregation should be minimized. While this can be difficult, minimizing aggregation can be achieved by constructing damage models where assets (e.g., buildings) are treated as individual elements within the model (rather than aggregated elements). Finally, this work demonstrates the potential severity of aggregation errors and how poorly these are understood, and therefore the need for further study.

Future work should evaluate the performance of the normal distribution applied here to synthetically generate depths. Also, a more generalizeable measure of curvature (e.g., local derivative) could be explored to more clearly classify and communicate the aggregation error potential of different functions. By extending the first order approximations developed here, the flood risk model domain could be segregated into areas with more or less sensitivity to aggregation errors. In this way, the accuracy of large-scale flood risk models could be improved without drastically increasing the computational requirements.

Chapter 6

DISCUSSION AND CONCLUSION

Within the last decade, macro-scale flood risk models have matured from nascent proof-of-concepts to essential tools for global, national, and — increasingly — sub-national decision makers (Ward et al., 2020). Despite this progress, these models often lack rigorous validation and generally underperform compared to high-resolution local models (Wing et al., 2017). This dissertation investigated implicit scale transfers inherent in such coarse-resolution macro-scale models and their impact on model behaviour. To accomplish this, three sub-questions examining applications of aggregation and disaggregation were explored. The following sections discuss each of these sub-questions in turn, placing them in the overall context of the overarching research question and showing how they fit into the current state of knowledge. The final section addresses the overarching research question directly and discusses how each chapter and sub-question fit together to address it.

6.1 Resolution Enhancement of Flood Inundation Grids

Focusing on hazard downscaling, Chapter 2 addressed the following:

How does enhancing the resolution of flood hazards improve accuracy?

This is the first study to evaluate inundation grid downscaling using observations, comparing the state of the art to a novel algorithm. All downscaling algorithms evaluated resulted in improved accuracy over the coarse hydrodynamic model when compared against observations (Figure 2.6). This finding reinforces the limitations of coarse hydrodynamic models for local-scale applications, as reported in previous studies (Horritt and Bates, 2001; Fewtrell et al., 2008; Savage et al., 2016; Papaioannou et al., 2016; Alipour et al., 2022). While bias remained similar between downscaled and coarse model output, further assessment with more accurate inundation observations is needed to generalize this finding. This study introduced and evaluated a novel algorithm, "CostGrow" for enhancing the resolution of flood water surface grids. The novel CostGrow algorithm not only outperformed the prior state-of-the-art (Schumann et al., 2014) in accuracy, but also reduced computation time by a factor of six. This study underscores the potential of employing coarse modelling with downscaling to produce detailed inundation grids at reduced computational expense, a strategy employed in recent global modelling efforts (Winsemius et al., 2013; Sampson et al., 2015; Bates et al., 2021).

6.2 Bias in Flood Hazard Grid Aggregation

Focusing on hazard aggregation and how this relates to exposure, Chapter 3 addressed the following:

How does the aggregation of flood hazard data affect model behaviour?

This is the first study to provide a formal treatment for the common practice of aggregating flood hazard grids. While such aggregation has been used by past studies, ours is the first to describe and evaluate the aggregation algorithms. To evaluate the behaviour of the algorithms, the novel *resample case framework* was developed and used to demonstrate that aggregation through averaging will always lead to bias in

regions near the edge of inundation. The study first established this finding mathematically, enhancing its generalizability, and then verified the finding with a case study. Importantly, the study demonstrated the direction of bias is sensitive to both the aggregation algorithm and the metric of interest. For example, using simple averaging of water depth grids tended to increase inundation extents in regions near the edge of inundation, while averaging water surface elevation grids typically resulted in decreased flood volume, similar to what Ghimire and Sharma (2021) and Komolafe et al. (2015) found in their case studies.

Expanding the analysis to consider building exposure, our study found specific mechanisms leading to bias; however, these may be sensitive to the particulars of the study region. For example, exposure was found to have a *dry bias* as buildings tended to be on locally higher ground. Pollack et al. (2022) describe a similar mechanism where high-value assets near shorelines have a higher exposure, leading to a negative bias in damage estimates for some aggregate blocks. However, these mechanisms, which operate at different scales and on different elements of risk modelling, are not contradictory and can coexist in large models.

The *dry bias* phenomenon identified by the study provides evidence and explanation for one mechanism that may contribute to the coarse flood risk model overestimation reported elsewhere. More practically, the formalization of different algorithms and the presented behaviours can help modellers select and describe the method most appropriate for their study. During the planning phase of a future flood risk model study, the novel resample case framework and our computed biases can facilitate the selection of the maximum allowable level of aggregation given that future study's error tolerance. In summary, our study provides the first systematic and the first mathematical investigation of the grid aggregation found in many risk models, demonstrating important mechanisms to help flood risk modellers understand model behaviour and better communicate uncertainties.

6.3 Aggregating Flood Damage Functions

Focusing on the aggregation of vulnerability models or flood damage functions, Chapter 4 and 5 addressed the following:

> *How does the aggregation of damage functions affect model behaviour?*

These are the first studies to explicitly consider the implications of the common practice of applying per-asset damage functions to aggregate exposure models. To accomplish this, both studies applied Jensen's inequality to demonstrate that aggregating flood damage functions leads to systematic bias.

For the theory-focused approach in Chapter 5, depths were synthetically generated using normal distributions parameterized based on expert knowledge. These depths were then used to estimate bias from aggregations of up to 100 hypothetical buildings. This first order estimate showed that the error introduced by this aggregation could be as high as 40%. Further, it was found that the direction of this bias is sensitive to relative flood depth magnitude, with large depths driving an overestimate. To provide a broad analysis that included all available damage functions, negative depths were included in the synthetically generated data, leading to the finding that smaller depth magnitudes drive an underestimate. However, this behaviour would only apply to hazard models that include some negative depths (i.e., groundwater flooding or elevated buildings) — something not typically included in macro-scale hazard models like the one employed in Chapter 4. While the synthetic depths implemented in this study make the transferability of bias magnitudes to specific studies difficult, the estimates of error potential for the 344 damage

functions can inform uncertainty communication and error tolerance decisions for studies employing one of these functions.

Chapter 4 leveraged results from a global fluvial-hazard model and 2 million building locations in Germany to provide the first quantification of error from damage function aggregation. Extending the analysis in Chapter 5 that showed most flood damage functions are concave, Chapter 4's application of Jensen's inequality demonstrates how aggregated models systematically overestimate. This has important applications for studies that apply damage functions developed from per-asset data onto aggregate exposure models (Hall et al., 2005; Ward et al., 2020; Sairam et al., 2021) as this artifact has so far not been considered by any study. Further, the analysis showed that bias is not distributed evenly in space, meaning some regions previously identified as "hot spots of risk" may in fact just be host spots of aggregation bias. While this is the first study we are aware of to investigate bias from aggregation and function shape, the findings of positive bias are similar to those reported in scenario-based studies (Seifert et al., 2010b; Pollack et al., 2022).

The obvious extension of this study to consider multi-variate damage functions would be challenging; however, the spatial variance of other, typically weaker damage indicators (e.g., velocity), may be greater than the variances explored here for depth alone. In this case, the aggregation bias reported here for univariate functions may underestimate the bias of more sophisticated multi-variate models. Another limitation of our study is the relatively coarse 30 m resolution of the hazard model. While this is the best available for the macro-scale models of interest to our study, 30 m model resolution may underestimate observed depth variances. Considering urban areas would be the most sensitive to this limitation provides another suggestion that our bias is an underestimate. Taken together, this collection of probable bias underestimates gives us confidence that the coarse model overestimates we report are conservative. Future work could also explore the application of bias correction methods from the statistical sciences to treat the artifacts reported here. Practitioners can use the findings of this study to better understand where and why flood risk may be overestimated, helping inform decisions on model construction and communication.

6.4 Aggregation and Disaggregation in Flood Risk Models

With the four core chapters, this dissertation pursued answers to the following:

> *How do the aggregations and disaggregations currently employed in risk models affect their behaviour?*

The analysis of aggregation or upscaling in Chapters 3, 4, and 5 provided mechanistic explanations for bias production arising from two scale transfers commonly found in macro-scale flood risk models. These studies demonstrated a direct relationship between the degree of aggregation and the magnitude of induced bias. This indicates that coarse aggregations, like those employed in global models, systematically overestimate risk. However, these aggregation errors are generally less than those reported for other model components (Pollack et al., 2022). Chapter 2 demonstrated the computational efficiency gains achievable through downscaling flood hazard grids. This highlights the superior performance and speed of our novel algorithm compared to the current state of the art and the potential for increasing the efficiency of hazard map production.

A focus of this dissertation is to call more attention to model components that are often considered part of pre- or post-processing and therefore not included in typical uncertainty quantification or communication. To accomplish this, Chapter 3 developed the novel *resample case framework* (Figure 3.2) to formalize evaluation of flood hazard aggregation bias. This framework emerges from the properties of water and terrain grids unique to flood hazard rescaling. This development enabled the formalization of grid *dis*aggregation in Chapter 2. This framework could similarly support further investigations into other components of uncertainty and more precise communication of uncertainties.

Taken together, the findings of the core chapters reveal that aggregation and disaggregation can have conditional and complicated influences on the behaviour of risk models and that, especially for large scale-transfers, these effects can be significant. Rising above this however is a simpler overarching finding:

all else equal, coarse models overestimate risk.

A similar phenomenon is well known in the fluvial inundation modelling literature (Saksena and Merwade, 2015; Mohanty et al., 2020; Ghimire and Sharma, 2021; Muthusamy et al., 2021; Banks et al., 2015; Xafoulis et al., 2023) and — while less well known in the risk modelling literature — has been reported by some studies (Komolafe et al., 2015; Brussee et al., 2021; Pollack et al., 2022). Unlike previous studies which simply quantified differences between coarse and fine model results (Komolafe et al., 2015; Brussee et al., 2021; Pollack et al., 2022), this dissertation provides mechanistic explanations for the consistent overestimation bias found by these previous studies. For example, the aggregation of non-linear concave damage functions (Chapter 4) or the dry bias of buildings (Chapter 3).

All three chapters demonstrate that rescaling bias is spatially heterogeneous, suggesting that modellers must understand *how* rescaling may bias their models in order to reduce or communicate uncertainties — knowledge now enabled by this dissertation. To guide this process, this dissertation offers specific recommendations that help modellers avoid problematic rescaling situations in susceptible regions. For example, Chapter 2 and 3 show how regions at the edge of inundation will always experience bias during rescaling while Chapter 4 and 5 show how regions with large water depth variances will be more sensitive to aggregation. Modellers should be wary of both regions when modelling at coarse scale. Or, put another way, modellers may need to employ more fine scale models in these regions.

The studies reported in this dissertation were the first to provide uncertainty analysis for three commonly employed scale transfers. However, macro-scale flood risk models, which are increasingly complex, include more explicit and implicit scale transfers than those addressed in this dissertation. For example, the effect of aggregate exposure models on the exposure calculation (Figure 1.2) has so far not been considered. To accurately quantify and communicate uncertainty, a holistic understanding of all such scale transfers is required.

This dissertation provided the first explanations and magnitude estimates for uncertainties arising from particular scale transfers. With this foundation laid, future studies could explore correcting for these systematic biases. For example, the mean relative loss values reported in Figure 4.4 could be employed as bias corrected damage functions in this particular case. However, further work is needed before such an application (e.g, the calculation of confidence bounds, the extension to more damage functions, and the inclusion of finer resolutions). The flood hazard grid aggregation discussed in Chapter 3 could also

benefit from formal correction to improve risk estimates; however, a more robust approach would likely employ the downscaling of Chapter 2 along with a treatment of exposure aggregation. Regardless, the utility of understanding these scaling uncertainties depends on the popularity of macro-scale flood risk models and the speed at which data and compute allow for accuracies comparable to local-scale models.

Numerous studies have reported on the large uncertainties and errors in components of macro-scale flood risk model chains (de Moel and Aerts, 2011; Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020; Ward et al., 2015a). This dissertation contributes to this literature by quantifying the uncertainty introduced specifically by scale transformations while providing mathematical and mechanistic explanations for the emergence of systematic bias. While errors arising from (reasonable) scale transformations were found to be relatively small, their unique nature warrants close attention. Unlike errors attributable to natural variability, modelling process gaps, or data quality, these errors are systematic, epistemic, and driven by choices within the model structure — artifacts of the modelling process itself. As such, not only is there a greater potential to mitigate such errors, but a greater impetus upon modellers to be aware of where and *how* they manifest.

BIBLIOGRAPHY

Abramovich, S. and Persson, L.-E.: Some new estimates of the 'Jensen gap', Journal of Inequalities and Applications, 2016, 1–9, 2016.

Aires, F., Miolane, L., Prigent, C., Pham, B., Fluet-Chouinard, E., Lehner, B., and Papa, F.: A Global Dynamic Long-Term Inundation Extent Dataset at High Spatial Resolution Derived through Downscaling of Satellite Observations, Journal of Hydrometeorology, 18, 1305–1325, https://doi.org/10.1175/JHM-D-16-0155.1, 2017.

Alipour, A., Jafarzadegan, K., and Moradkhani, H.: Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping, Environmental Modelling & Software, 152, 105 398, https://doi.org/10.1016/j.envsoft.2022.105398, 2022.

Angelakis, A. N., Capodaglio, A. G., Valipour, M., Krasilnikoff, J., Ahmed, A. T., Mandi, L., Tzanakakis, V. A., Baba, A., Kumar, R., Zheng, X., Min, Z., Han, M., Turay, B., Bilgiç, E., and Dercas, N.: Evolution of Floods: From Ancient Times to the Present Times (ca 7600 BC to the Present) and the Future, Land, 12, 1211, https://doi.org/10.3390/land12061211, 2023.

Antwi-Agyakwa, K. T., Afenyo, M. K., and Angnuureng, D. B.: Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools, Water, 15, 427, https://doi.org/10.3390/w15030427, number: 3, Publisher: Multidisciplinary Digital Publishing Institute, 2023.

Apel, H.: hydro / rfm / RIM2D · GitLab, URL `https://git.gfz-potsdam.de/hydro/rfm/rim2d`, 2023.

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H.: Flood risk analyses—how detailed do we need to be?, Natural Hazards, 49, 79–98, URL `http://link.springer.com/article/10.1007/s11069-008-9277-8`, 2009.

Apel, H., Vorogushyn, S., and Merz, B.: Brief communication: Impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany, Natural Hazards and Earth System Sciences, 22, 3005–3014, https://doi.org/10.5194/nhess-22-3005-2022, 2022.

Bach, M., Meigen, T., and Strasburger, H.: Raster-scan cathode-ray tubes for vision research-limits of resolution in space, time and intensity, and some solutions, Spatial vision, 10, 403–414, 1997.

Banks, J. C., Camp, J. V., and Abkowitz, M. D.: Scale and Resolution Considerations in the Application of HAZUS-MH 2.1 to Flood Risk Assessments, Natural Hazards Review, 16, 04014 025, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000160, 2015.

Barendrecht, M. H., Viglione, A., ., and Bloschl, G.: A dynamic framework for flood risk, Water Security, https://doi.org/10.1016/j.wasec.2017.02.001, 2017.

Bates, P. D., Horritt, M. S., and Fewtrell, T. J.: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling, Journal of Hydrology, 387, 33–45, https://doi.org/10.1016/j.jhydrol.2010.03.027, 2010.

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Schumann, G., Giustarini, L., Coxon, G., Porter, J. R., Amodeo, M. F., Chu, Z., Lewis-Gruss, S., Freeman, N. B., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K., Emanuel, K., Ferreira, C. M., Khalid, A., Haigh, I. D., Couasnon, A., Kopp, R., Hsiang, S., and Krajewski, W. F.: Combined Modeling of US Fluvial, Pluvial, and Coastal Flood Hazard Under Current and Future Climates, Water Resources Research, 57, https://doi.org/10.1029/2020WR028673, 2021.

Bedient, P.: Hydrology and Floodplain Analysis. 5ed, Pearson, 5 edn., 2013.

Bellos, V. and Tsakiris, G.: Comparing Various Methods of Building Representation for 2D Flood Modelling In Built-Up Areas, Water Resources Management, 29, 379–397, https://doi.org/10.1007/s11269-014-0702-3, 2015.

Bertola, M., Blöschl, G., Bohac, M., Borga, M., Castellarin, A., Chirico, G. B., Claps, P., Dallan, E., Danilovich, I., Ganora, D., Gorbachova, L., Ledvinka, O., Mavrova-Guirguinova, M., Montanari, A., Ovcharuk, V., Viglione, A., Volpi, E., Arheimer, B., Aronica, G. T., Bonacci, O., Čanjevac, I., Csik, A., Frolova, N., Gnandt, B., Gribovszki, Z., Gül, A., Günther, K., Guse, B., Hannaford, J., Harrigan, S., Kireeva, M., Kohnová, S., Komma, J., Kriauciuniene, J., Kronvang, B., Lawrence, D., Lüdtke, S., Mediero, L., Merz, B., Molnar, P., Murphy, C., Oskoruš, D., Osuch, M., Parajka, J., Pfister, L., Radevski, I., Sauquet, E., Schröter, K., Šraj, M., Szolgay, J., Turner, S., Valent, P., Veijalainen, N., Ward, P. J., Willems, P., and Zivkovic, N.: Megafloods in Europe can be anticipated from observations in hydrologically similar catchments, Nature Geoscience, 16, 982–988, https://doi.org/10.1038/s41561-023-01300-5, number: 11 Publisher: Nature Publishing Group, 2023.

Bierkens, M., Finke, P., and De Willigen, P.: Upscaling and downscaling methods for environmental research, Kluwer Academic, 2000.

Birkmann, J.: Measuring vulnerability to promote disaster-resilient societies: Conceptual frameworks and definitions, Measuring vulnerability to natural hazards: Towards disaster resilient societies, 1, 3–7, 2006.

Birkmann, J., ed.: Measuring vulnerability to natural hazards: towards disaster resilient societies. 2ed., United Nations University Press, 2013.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, Science, 357, 588–590, https://doi.org/10.1126/science.aan2506, publisher: American Association for the Advancement of Science, 2017.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate both increases and decreases European river floods, Nature, 573, 108–111, https://doi.org/10.1038/s41586-019-1495-6, 2019.

Brussee, A. R., Bricker, J. D., De Bruijn, K. M., Verhoeven, G. F., Winsemius, H. C., and Jonkman, S. N.: Impact of hydraulic model resolution and loss of life model modification on flood fatality risk estimation: Case study of the Bommelerwaard, The Netherlands, Journal of Flood Risk Management, 14, e12 713, https://doi.org/10.1111/jfr3.12713, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfr3.12713, 2021.

Bryant, S.: Flood Hazard Grid Aggregation Scripts [Software], https://doi.org/10.5281/zenodo.8271996, 2023.

Bryant, S., McGrath, H., and Boudreault, M.: Gridded flood depth estimates from satellite-derived inundations, Natural Hazards and Earth System Sciences, 22, 1437–1450, https://doi.org/10.5194/nhess-22-1437-2022, 2022a.

Bryant, S., Kreibich, H., and Merz, B.: Bias in Flood Hazard Grid Aggregation, Water Resources Research, 59, e2023WR035 100, https://doi.org/10.1029/2023WR035100, 2023.

Bryant, S., Kreibich, H., and Merz, B.: Flood damage model bias caused by aggregation, in: Proceedings of IAHS, Copernicus GmbH, [In Review], 2024a.

Bryant, S., Reimer, J., Kreibich, H., and Merz, B.: Aggregating Flood Damage Functions: the Peril of Jensen's Gap, Journal of Flood Risk Management, in Review, 2024b.

Bryant, S., Schumann, G., Apel, H., Kreibich, H., and Merz, B.: Technical Note: Resolution enhancement of flood inundation grids, Hydrology and Earth System Sciences, 28, 575–588, https://doi.org/10.5194/hess-28-575-2024, publisher: Copernicus GmbH, 2024c.

Bryant, S. P., Davies, E. G. R., Sol, D., and Davis, S.: The progression of flood risk in southern Alberta since the 2013 flood, Journal of Flood Risk Management, p. 18, https://doi.org/https://doi.org/10.1111/jfr3.12811, 2022b.

Buttle, J. M., Allen, D. M., Caissie, D., Davison, B., Hayashi, M., Peters, D. L., Pomeroy, J. W., Simonovic, S., St-Hilaire, A., and Whitfield, P. H.: Flood processes in Canada: Regional and special aspects, Canadian Water Resources Journal / Revue canadienne des ressources hydriques, 41, 7–30, https://doi.org/10.1080/07011784.2015.1131629, 2016.

Cohen, S., Brakenridge, G. R., Kettner, A., Bates, B., Nelson, J., McDonald, R., Huang, Y.-F., Munasinghe, D., and Zhang, J.: Estimating Floodwater Depths from Flood Inundation Maps and Topography, JAWRA Journal of the American Water Resources Association, 54, 847–858, https://doi.org/10.1111/1752-1688.12609, 2018.

Cohen, S., Peter, B. G., Haag, A., Munasinghe, D., Moragoda, N., Narayanan, A., and May, S.: Sensitivity of Remote Sensing Floodwater Depth Calculation to Boundary Filtering and Digital Elevation Model Selections, Remote Sensing, 14, 5313, https://doi.org/10.3390/rs14215313, number: 21 Publisher: Multidisciplinary Digital Publishing Institute, 2022.

Copernicus Land Monitoring Service: European Union, Copernicus Land Monitoring Service 2018, European Environment Agency (EEA), URL `https://land.copernicus.eu/pan-european/corine-land-cover`, 2018.

Crichton, D.: The risk triangle, Natural disaster management, 102, 102–103, 1999.

de Moel, H. and Aerts, J. C. J. H.: Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates, Natural Hazards, 58, 407–425, https://doi.org/10.1007/s11069-010-9675-6, 2011.

de Moel, H., Van Alphen, J., and Aerts, J. C.: Flood maps in Europe–methods, availability and use, Natural hazards and earth system sciences, 9, 289–301, 2009.

de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E., and Ward, P. J.: Flood risk assessments at different spatial scales, Mitigation and Adaptation Strategies for Global Change, 20, 865–890, https://doi.org/10.1007/s11027-015-9654-z, 2015.

Degbelo, A. and Kuhn, W.: Spatial and temporal resolution of geographic information: an observation-based theory, Open Geospatial Data, Software and Standards, 3, 12, https://doi.org/10.1186/s40965-018-0053-8, 2018.

Demiray, B. Z., Sit, M., and Demir, I.: D-SRGAN: DEM Super-Resolution with Generative Adversarial Networks, SN Computer Science, 2, 48, https://doi.org/10.1007/s42979-020-00442-2, 2021.

Denny, M.: The fallacy of the average: on the ubiquity, utility and continuing novelty of Jensen's inequality, Journal of Experimental Biology, 220, 139–146, https://doi.org/10.1242/jeb.140368, 2017.

Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., Damm, B., Lucia, A., Fauer, F. S., Nissen, K. M., Sieg, T., and Thieken, A. H.: More than heavy rain turning into fast-flowing water – a landscape perspective on the 2021 Eifel floods, Natural Hazards and Earth System Sciences, 22, 1845–1856, https://doi.org/10.5194/nhess-22-1845-2022, publisher: Copernicus GmbH, 2022.

Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., Koutsoyiannis, D., and Efstratiadis, A.: 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, 534, 478–492, https://doi.org/10.1016/j.jhydrol.2016.01.020, 2016.

Dong, C., Loy, C. C., He, K., and Tang, X.: Image Super-Resolution Using Deep Convolutional Networks, URL http://arxiv.org/abs/1501.00092, 2015.

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A., and Feyen, L.: Increased human and economic losses from river flooding with anthropogenic warming, Nature Climate Change, 8, 781–786, https://doi.org/10.1038/s41558-018-0257-z, 2018.

Douris, J. and Kim, G.: The Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970-2019), ., 2021.

Duan, Y., Xiong, J., Cheng, W., Li, Y., Wang, N., Shen, G., and Yang, J.: Increasing Global Flood Risk in 2005–2020 from a Multi-Scale Perspective, Remote Sensing, 14, 5551, https://doi.org/10.3390/rs14215551, 2022.

Eisenberg, D., Seager, T., Alderson, D. L., and others: Rethinking Resilience Analytics, Risk Analysis, 39, 1870–1884, https://doi.org/10.1111/risa.13328, 2019.

Elbe, I. K. z. S. d.: Aktionsplan Hochwasserschutz Elbe, Tech. rep., IKSE, 2003.

Falter, D., Dung, N., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., Apel, H., Theisselmann, F., and Merz, B.: Continuous, large-scale simulation model for flood risk assessments: proof-of-concept: Large-scale flood risk assessment model, Journal of Flood Risk Management, 9, 3–21, https://doi.org/10.1111/jfr3.12105, 2016.

Fathom: Fathom Global Flood Map 3 Specification, Tech. rep., report, 2023.

FEMA: Multi-Hazard Loss Estimation Methodology, Flood Model: Hazus-MH MR2 Technical Manual, Tech. rep., FEMA Washington, DC, URL https://www.fema.gov/media-library-data/20130726-1820-25045-8292/hzmh2_1_fl_tm.pdf, 2012.

Fewtrell, T. J., Bates, P. D., Horritt, M., and Hunter, N. M.: Evaluating the effect of scale in flood inundation modelling in urban environments, Hydrological Processes, 22, 5107–5118, https://doi.org/10.1002/hyp.7148, 2008.

Field, C. B., Barros, V. R., Mach, K. J., Mastrandrea, M. D., van Aalst, R., Adger, W. N., Arent, D. J., Barnett, J., Betts, R. A., Bilir, T. E., and others: Technical summary, Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Worikng Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, publisher: Cambridge University Press, 2014.

Figueiredo, R. and Martina, M.: Using open building data in the development of exposure data sets for catastrophe risk modelling, Natural Hazards and Earth System Sciences, 16, 417–429, https://doi.org/10.5194/nhess-16-417-2016, 2016.

Fluet-Chouinard, E., Lehner, B., Rebelo, L.-M., Papa, F., and Hamilton, S. K.: Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data, Remote Sensing of Environment, 158, 348–361, https://doi.org/10.1016/j.rse.2014.10.015, 2015.

Foltête, J., Berthier, K., and Cosson, J.: Cost distance defined by a topological function of landscape, Ecological Modelling, 210, 104–114, https://doi.org/10.1016/j.ecolmodel.2007.07.014, 2008.

Fraehr, N., Wang, Q. J., Wu, W., and Nathan, R.: Development of a Fast and Accurate Hybrid Model for Floodplain Inundation Simulations, Water Resources Research, 59, e2022WR033 836, https://doi.org/10.1029/2022WR033836, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR033836, 2023.

Gao, X., Sitharam, M., and Roitberg, A. E.: Bounds on the Jensen gap, and implications for mean-concentrated distributions, arXiv preprint arXiv:1712.05267, 2017.

GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library, Open Source Geospatial Foundation, https://doi.org/10.5281/zenodo.5884351, 2022.

GeoNB: Flood Risk Areas and Historical Floods [Dataset], URL `http://www.snb.ca/geonb1/e/DC/floodraahf.asp`, 2019.

Gerl, T., Kreibich, H., Franco, G., Marechal, D., and Schröter, K.: A Review of Flood Loss Models as Basis for Harmonization and Benchmarking, PloS one, 11, e0159 791, 2016.

Ghimire, E. and Sharma, S.: Flood Damage Assessment in HAZUS Using Various Resolution of Data and One-Dimensional and Two-Dimensional HEC-RAS Depth Grids, Natural Hazards Review, 22, 04020 054, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000430, 2021.

Gleick, P. H.: Water in crisis, Pacific Institute for Studies in Dev., Environment & Security. Stockholm Env. Institute, Oxford Univ. Press. 473p, 9, 1051–0761, 1993.

Goudge, T. A., Morgan, A. M., Stucky de Quay, G., and Fassett, C. I.: The importance of lake breach floods for valley incision on early Mars, Nature, 597, 645–649, 2021.

Government of New Brunswick: ERD 2015 Lidar [Dataset], URL `http://geonb.snb.ca/li/index.html`, 2016.

Guo, K., Guan, M., and Yu, D.: Urban surface water flood modelling – a comprehensive review of current models and future challenges, Hydrology and Earth System Sciences, 25, 2843–2860, https://doi.org/10.5194/hess-25-2843-2021, 2021.

Haimes, Y. Y.: Risk modeling, assessment, and management, Wiley series in systems engineering and management, John Wiley & Sons, Hoboken, NJ, 3rd ed edn., oCLC: 232605676, 2009.

Hall, J. W., Sayers, P. B., ., and Dawson, R. J.: National-scale Assessment of Current and Future Flood Risk in England and Wales, Natural Hazards, 36, 147–164, https://doi.org/10.1007/s11069-004-4546-7, 2005.

Heritage, G. L., Milan, D. J., Large, A. R., and Fuller, I. C.: Influence of survey strategy and interpolation model on DEM quality, Geomorphology, 112, 334–344, https://doi.org/10.1016/j.geomorph.2009.06.024, 2009.

Hoeksema, R. J.: Designed for dry feet: flood protection and land reclamation in the Netherlands., American Society of Civil Engineers, 2006.

Horritt, M. S. and Bates, P. D.: Effects of spatial resolution on a raster based model of flood flow, Journal of Hydrology, 253, 239–249, https://doi.org/https://doi.org/10.1016/S0022-1694(01)00490-5, 2001.

Hrabluk, L.: New Brunswick Underwater: The 2018 Saint John River Flood, MACINTYRE & PURCELL, 1-77276-115-X, 2019.

IEC and ISO: Risk management – Risk assessment techniques. 2.0, oCLC: 1112672684, 2019.

IPCC: Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 1 edn., https://doi.org/10.1017/9781009325844, 2023.

IRDR: Peril Classification and Hazard Glossary (IRDR DATA Publication No. 1), 2014.

IWR and USACE: Principles of Risk Analysis for Water Resources, Tech. rep., IWR; USACE, 2017.

Jackson, N. D. and Gunda, T.: Evaluation of extreme weather impacts on utility-scale photovoltaic plant performance in the United States, Applied Energy, 302, 117 508, https://doi.org/https://doi.org/10.1016/j.apenergy.2021.117508, 2021.

Jensen, J. L. W. V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta mathematica, 30, 175–193, 1906.

Jongman, B., Kreibich, H., Apel, H., Barredo, J., Bates, P., Feyen, L., Gericke, A., Neal, J., Aerts, J., and Ward, P.: Comparative flood damage model assessment: towards a European approach, Natural Hazards and Earth System Sciences, 12, 3733–3752, 2012.

Kaplan, S. and Garrick, B. J.: On The Quantitative Definition of Risk, Risk Analysis, 1, 11–27, https://doi.org/10.1111/j.1539-6924.1981.tb01350.x, 1981.

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., and Blöschl, G.: Joint Trends in Flood Magnitudes and Spatial Extents Across Europe, Geophysical Research Letters, 47, e2020GL087 464, https://doi.org/10.1029/2020GL087464, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL087464, 2020.

Keys, R.: Cubic convolution interpolation for digital image processing, IEEE Transactions on Acoustics, Speech, and Signal Processing, 29, 1153–1160, https://doi.org/10.1109/TASSP.1981.1163711, conference Name: IEEE Transactions on Acoustics, Speech, and Signal Processing, 1981.

Knight, F. H.: Risk, uncertainty and profit, New York: Hart, Schaffner and Marx, 1921.

Komolafe, A., Herath, S., and Avtar, R.: Sensitivity of flood damage estimation to spatial resolution: Sensitivity of flood damage estimation to spatial resolution, Journal of Flood Risk Management, 11, S370–S381, https://doi.org/10.1111/jfr3.12224, 2015.

Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C. J. H., Apel, H., Aronica, G. T., Arnbjerg-Nielsen, K., Bouwer, L. M., Bubeck, P., Caloiero, T., Chinh, D. T., Cortès, M., Gain, A. K., Giampá, V., Kuhlicke, C., Kundzewicz, Z. W., Llasat, M. C., Mård, J., Matczak, P., Mazzoleni, M., Molinari, D., Dung, N. V., Petrucci, O., Schröter, K., Slager, K., Thieken, A. H., Ward, P. J., and Merz, B.: Adaptation to flood risk: Results of international paired flood event studies, Earth's Future, 5, 953–965, https://doi.org/10.1002/2017EF000606, 2017.

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., Alvarez-Garreton, C., Aznar, B., Balkhi, L., Barendrecht, M. H., Biancamaria, S., Bos-Burgering, L., Bradley, C., Budiyono, Y., Buytaert, W., Capewell, L., Carlson, H., Cavus, Y., Couasnon, A., Coxon, G., Daliakopoulos, I., de Ruiter, M. C., Delus, C., Erfurt, M., Esposito, G., François, D., Frappart, F., Freer, J., Frolova, N., Gain, A. K., Grillakis, M., Grima, J. O., Guzmán, D. A., Huning, L. S., Ionita, M., Kharlamov, M., Khoi, D. N., Kieboom, N., Kireeva, M., Koutroulis, A., Lavado-Casimiro, W., Li, H.-Y., LLasat, M. C., Macdonald, D., Mård, J., Mathew-Richards, H., McKenzie, A., Mejia, A., Mendiondo, E. M., Mens, M., Mobini, S., Mohor, G. S., Nagavciuc, V., Ngo-Duc, T., Thao Nguyen Huynh, T., Nhi, P. T. T., Petrucci, O., Nguyen, H. Q., Quintana-Seguí, P., Razavi, S., Ridolfi, E., Riegel, J., Sadik, M. S., Savelli, E., Sazonov, A., Sharma, S., Sörensen, J., Arguello Souza, F. A., Stahl, K., Steinhausen, M., Stoelzle, M., Szalińska, W., Tang, Q., Tian, F., Tokarczyk, T., Tovar, C., Tran, T. V. T., Van Huijgevoort, M. H. J., van Vliet, M. T. H., Vorogushyn, S., Wagener, T., Wang, Y., Wendt, D. E., Wickham, E., Yang, L., Zambrano-Bigiarini, M., Blöschl, G., and Di Baldassarre, G.: The challenge of unprecedented floods and droughts in risk management, Nature, 608, 80–86, https://doi.org/10.1038/s41586-022-04917-5, 2022.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K., and Sherstyukov, B.: Flood risk and climate change: global and regional perspectives, Hydrological Sciences Journal, 59, 1–28, https://doi.org/10.1080/02626667.2013.857411, 2014.

Kundzewicz, Z. W., Su, B., Wang, Y., Wang, G., Wang, G., Huang, J., and Jiang, T.: Flood risk in a range of spatial perspectives – from global to local scales, Natural Hazards and Earth System Sciences, 19, 1319–1328, https://doi.org/10.5194/nhess-19-1319-2019, publisher: Copernicus GmbH, 2019.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C.: A new method to assess the risk of local and widespread flooding on rivers and coasts, Journal of Flood Risk Management, 3, 323–336, https://doi.org/10.1111/j.1753-318X.2010.01081.x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1753-318X.2010.01081.x, 2010.

Landesamt für Umwelt Rheinland-Pfalz: Hochwasser im Juli 2021, Tech. rep., Landesamt für Umwelt Rheinland-Pfalz, URL https://lfu.rlp.de/fileadmin/lfu/Wasserwirtschaft/Ahr-Katastrophe/Hochwasser_im_Juli2021.pdf, 2022.

Li, S., Sun, D., Goldberg, M., Kalluri, S., Sjoberg, B., Lindsey, D., Hoffman, J., DeWeese, M., Connelly, B., Mckee, P., and Lander, K.: A downscaling model for derivation of 3-D flood products from VIIRS imagery and SRTM/DEM, ISPRS Journal of Photogrammetry and Remote Sensing, 192, 279–298, https://doi.org/10.1016/j.isprsjprs.2022.08.025, 2022.

Liao, J. and Berg, A.: Sharpening Jensen's inequality, The American Statistician, 2018.

Lindsay, J.: The whitebox geospatial analysis tools project and open-access GIS, in: Proceedings of the GIS Research UK 22nd Annual Conference, The University of Glasgow, pp. 16–18, URL https://jblindsay.github.io/ghrg/pubs/LindsayGISRUK2014.pdf, 2014.

LISFLOOD developers: LISFLOOD-FP 8.0 hydrodynamic model, Zenodo, https://doi.org/10.5281/zenodo.4073011, publisher: Zenodo, 2020.

Löwe, R., Urich, C., Sto. Domingo, N., Mark, O., Deletic, A., and Arnbjerg-Nielsen, K.: Assessment of urban pluvial flood risk and efficiency of adaptation options through simulations – A new generation of urban planning tools, Journal of Hydrology, 550, 355–367, https://doi.org/10.1016/j.jhydrol.2017.05.009, 2017.

Mallapaty, S.: Why are Pakistan's floods so extreme this year?, Nature, https://doi.org/10.1038/d41586-022-02813-6, 2022.

McBain, W.: Twenty-first century flood risk management, in: Flood Risk: Planning, Design and Management of Flood Defence Infrastructure, pp. 7–22, ICE Publishing, 2012.

McGrath, H., Stefanakis, E., and Nastev, M.: Sensitivity analysis of flood damage estimates: A case study in Fredericton, New Brunswick, International Journal of Disaster Risk Reduction, 14, 379–387, https://doi.org/10.1016/j.ijdrr.2015.09.003, 2015.

Merz, B. and Thieken, A. H.: Flood risk analysis: Concepts and challenges, Österreichische Wasser-und Abfallwirtschaft, 56, 27–34, publisher: Springer, 2004.

Merz, B., Kreibich, H., Thieken, A., and Schmidtke, R.: Estimation uncertainty of direct monetary flood damage to buildings, Natural Hazards and Earth System Sciences, 4, 153–163, https://doi.org/10.5194/nhess-4-153-2004, 2004.

Merz, B., Elmer, F., and Thieken, A.: Significance of" high probability/low damage" versus" low probability/high damage" flood events, Natural Hazards and Earth System Sciences, 9, 1033–1046, https://doi.org/https://doi.org/10.5194/nhess-9-1033-2009, 2009.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Review article "Assessment of economic flood damage", Natural Hazards and Earth System Sciences, 10, 1697–1724, https://doi.org/10.5194/nhess-10-1697-2010, 2010.

Merz, B., Aerts, J., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., Blöschl, G., Bouwer, L. M., Brauer, A., Cioffi, F., Delgado, J. M., Gocht, M., Guzzetti, F., Harrigan, S., Hirschboeck, K., Kilsby, C., Kron, W., Kwon, H.-H., Lall, U., Merz, R., Nissen, K., Salvatti, P., Swierczynski, T., Ulbrich, U., Viglione, A., Ward, P. J., Weiler, M., Wilhelm, B., and Nied, M.: Floods and climate: emerging perspectives for flood risk assessment and management, Natural Hazards and Earth System Sciences, 14, 1921–1942, https://doi.org/10.5194/nhess-14-1921-2014, 2014.

Messner, F.: FLOODSite: Evaluating flood damages: guidance and recommendations on principles and methods, Tech. Rep. T09-06-01, Helmholz Unweltforschungszentrum (UFZ), URL http://repository.tudelft.nl/view/hydro/uuid:5602db10-274c-40da-953f-34475ded1755/, 2007.

Messner, F. and Meyer, V.: Flood damage, vulnerability and risk perception–challenges for flood damage research, in: Flood risk management: hazards, vulnerability and mitigation measures, pp. 149–167, Springer, URL http://link.springer.com/chapter/10.1007/978-1-4020-4598-1_13, 2006.

Mester, B., Vogt, T., Bryant, S., Otto, C., Frieler, K., and Schewe, J.: Human displacements from Tropical Cyclone Idai attributable to climate change, Natural Hazards and Earth System Sciences, 23, 3467–3485, https://doi.org/10.5194/nhess-23-3467-2023, publisher: Copernicus GmbH, 2023.

Metin, A. D., Dung, N. V., Schröter, K., Guse, B., Apel, H., Kreibich, H., Vorogushyn, S., and Merz, B.: How do changes along the risk chain affect flood risk?, Natural Hazards and Earth System Sciences, 18, 3089–3108, https://doi.org/10.5194/nhess-18-3089-2018, 2018.

Metin, A. D., Dung, N. V., Schröter, K., Vorogushyn, S., Guse, B., Kreibich, H., and Merz, B.: The role of spatial dependence for large-scale flood risk estimation, Natural Hazards and Earth System Sciences, 20, 967–979, https://doi.org/10.5194/nhess-20-967-2020, 2020.

Microsoft: microsoft CanadianBuildingFootprints [Dataset], URL https://github.com/microsoft/CanadianBuildingFootprints, 2019.

Milan Geoservice GmbH: Airborne Laser Scanning (ALS) Ahr Valley, Tech. Rep. D-01917 Kamenz, report, 2023.

Mitchell, H., Hamilton, T., Steggerda, F., and Bean, H.: The chemical composition of the adult human body and its bearing on the biochemistry of growth, Journal of Biological Chemistry, 158, 625–637, 1945.

Mohanty, M. P., Nithya, S., Nair, A. S., Indu, J., Ghosh, S., Mohan Bhatt, C., Srinivasa Rao, G., and Karmakar, S.: Sensitivity of various topographic data in flood management: Implications on inundation mapping over large data-scarce regions, Journal of Hydrology, 590, 125 523, https://doi.org/10.1016/j.jhydrol.2020.125523, 2020.

Mohor, G. S.: Exploring the transferability of flood loss models across flood types, Ph.D. thesis, Universität Potsdam, https://doi.org/10.25932/publishup-55714, 2022.

Molinari, D. and Scorzini, A. R.: On the Influence of Input Data Quality to Flood Damage Estimation: The Performance of the INSYDE Model, Water, 9, 688, https://doi.org/10.3390/w9090688, 2017.

Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., Gallazzi, A., Galliani, M., Grelot, F., Kellermann, P., Kreibich, H., Mohor, G. S., Mosimann, M., Natho, S., Richert, C., Schroeter, K., Thieken, A. H., Zischg, A. P., and Ballio, F.: Are flood damage models converging to reality? Lessons learnt from a blind test, Natural Hazards and Earth System Sciences, https://doi.org/10.5194/nhess-2020-40, 2020.

Muthusamy, M., Casado, M. R., Butler, D., and Leinster, P.: Understanding the effects of Digital Elevation Model resolution in urban fluvial flood modelling, Journal of Hydrology, 596, https://doi.org/10.1016/j.jhydrol.2021.126088, 2021.

National Research Council: Levees and the National Flood Insurance Program: Improving Policies and Practices, National Academies Press, Washington, D.C., https://doi.org/10.17226/18309, 2013.

Nguyen, V. D., Metin, A. D., Alfieri, L., Vorogushyn, S., and Merz, B.: Biases in national and continental flood risk assessments by ignoring spatial dependence, Scientific Reports, 10, 19 387, https://doi.org/10.1038/s41598-020-76523-2, 2020.

Nievas, C. I., Pilz, M., Prehn, K., Schorlemmer, D., Weatherill, G., and Cotton, F.: Calculating earthquake damage building by building: the case of the city of Cologne, Germany, Bulletin of Earthquake Engineering, 20, 1519–1565, https://doi.org/10.1007/s10518-021-01303-w, 2022.

NISO CRediT Working Group: ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy, https://doi.org/10.3789/ansi.niso.z39.104-2022, 2022.

Nocedal, J. and Wright, S. J.: Numerical Optimization, Springer Series in Operations Research and Financial Engineering, Springer New York, https://doi.org/10.1007/978-0-387-40065-5, 2006.

Nones, M. and Caviedes-Voullième, D.: Computational advances and innovations in flood risk mapping, Journal of Flood Risk Management, 13, https://doi.org/10.1111/jfr3.12666, 2020.

Nordrhein-Westfalen, M. f. U. R. u. L. d. L.: Potentielle Hochwasserschäden am Rhein in NRW, Tech. rep., MURL Report Düsseldorf, 2000.

NumPy Developers: numpy.trapz — NumPy v1.26 Manual, URL `https://numpy.org/doc/stable/reference/generated/numpy.trapz.html`, 2022.

OpenStreetMap contributors: Planet dump [Data file from 2023 07 25] retrieved from https://download.geofabrik.de/, published: https://www.openstreetmap.org, 2015.

OpenStreetMap contributors: Planet dump retrieved from https://planet.osm.org, published: https://www.openstreetmap.org, 2022.

O'Connor, J. E., Baker, V. R., Waitt, R. B., Smith, L. N., Cannon, C. M., George, D. L., and Denlinger, R. P.: The Missoula and Bonneville floods—A review of ice-age megafloods in the Columbia River basin, Earth-Science Reviews, 208, 103 181, https://doi.org/10.1016/j.earscirev.2020.103181, 2020.

Papaioannou, G., Loukas, A., Vasiliades, L., and Aronica, G. T.: Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach, Natural Hazards, 83, 117–132, https://doi.org/10.1007/s11069-016-2382-1, 2016.

Paprotny, D., Sebastian, A., Morales-Nápoles, O., and Jonkman, S. N.: Trends in flood losses in Europe over the past 150 years, Nature Communications, 9, 1985, https://doi.org/10.1038/s41467-018-04253-1, 2018.

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Castellarin, A., Carisi, F., and Schröter, K.: Exposure and vulnerability estimation for modelling flood losses to commercial assets in Europe, Science of The Total Environment, 737, 140 011, https://doi.org/10.1016/j.scitotenv.2020.140011, 2020a.

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., and Schröter, K.: Estimating exposure of residential assets to natural hazards in Europe using open data, Natural Hazards and Earth System Sciences, 20, 323–343, https://doi.org/10.5194/nhess-20-323-2020, publisher: Copernicus GmbH, 2020b.

Paprotny, D., Vousdoukas, M. I., Morales-Nápoles, O., Jonkman, S. N., and Feyen, L.: Pan-European hydrodynamic models and their ability to identify compound floods, Natural Hazards, 101, 933–957, https://doi.org/10.1007/s11069-020-03902-3, 2020c.

Parker, D. J., Green, C. H., and Thompson, P. M.: Urban flood protection benetits: A project appraisal guide, Gower technical press, 1987.

Parker, J. A., Kenyon, R. V., and Troxel, D. E.: Comparison of Interpolating Methods for Image Resampling, IEEE Transactions on Medical Imaging, 2, 31–39, https://doi.org/10.1109/TMI.1983.4307610, conference Name: IEEE Transactions on Medical Imaging, 1983.

Paterson, D. L., Wright, H., and Harris, P. N. A.: Health Risks of Flood Disasters, Clinical Infectious Diseases, 67, 1450–1454, https://doi.org/10.1093/cid/ciy227, 2018.

Peduzzi, P., Dao, H., Herold, C., and Mouton, F.: Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index, Natural Hazards and Earth System Sciences, 9, 1149–1159, https://doi.org/10.5194/nhess-9-1149-2009, publisher: Copernicus GmbH, 2009.

Penning-Rowsell, E., Priest, S., Parker, D., Morris, J., Tunstall, S., Viavattene, C., Chatterton, J., and Owen, D.: Flood and coastal erosion risk management - Manual, Routledge, URL https://www.mcm-online.co.uk/manual/, 2013.

Petrucci, O.: Factors leading to the occurrence of flood fatalities: a systematic review of research papers published between 2010 and 2020, Natural hazards and earth system sciences, 22, 71–83, 2022.

Petsch, D. K., Cionek, V. d. M., Thomaz, S. M., and dos Santos, N. C. L.: Ecosystem services provided by river-floodplain ecosystems, Hydrobiologia, 850, 2563–2584, https://doi.org/10.1007/s10750-022-04916-7, 2023.

Pollack, A. B., Sue Wing, I., and Nolte, C.: Aggregation bias and its drivers in large-scale flood loss estimation: A Massachusetts case study, Journal of Flood Risk Management, https://doi.org/10.1111/jfr3.12851, 2022.

Pomeroy, J. W., Stewart, R. E., ., and Whitfield, P. H.: The 2013 flood event in the South Saskatchewan and Elk River basins: Causes, assessment and damages, Canadian Water Resources Journal / Revue canadienne des ressources hydriques, 41, 105–117, 2016.

Poussard, C., Dewals, B., Archambeau, P., and Teller, J.: Environmental Inequalities in Flood Exposure: A Matter of Scale, Frontiers in Water, 3, URL https://www.frontiersin.org/journals/water/articles/10.3389/frwa.2021.633046, 2021.

Pregnolato, M., Ford, A., Glenis, V., Wilkinson, S., and Dawson, R.: Impact of climate change on disruption to urban transport networks from pluvial flooding, Journal of Infrastructure Systems, 23, 04017 015, 2017.

Ranke, U.: Natural Disaster Risk Management, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-20675-2, 2016.

Reed, C., Anderson, W., Kruczkiewicz, A., Nakamura, J., Gallo, D., Seager, R., and McDermid, S. S.: The impact of flooding on food security across Africa, Proceedings of the National Academy of Sciences, 119, e2119399 119, 2022.

Rokaya, P., Budhathoki, S., and Lindenschmidt, K.-E.: Trends in the timing and magnitude of ice-jam floods in Canada, Scientific Reports, 8, 5834, 2018.

Rytwinski, T., Harper, M., Taylor, J. J., Bennett, J. R., Donaldson, L. A., Smokorowski, K. E., Clarke, K., Bradford, M. J., Ghamry, H., Olden, J. D., et al.: What are the effects of flow-regime changes on fish productivity in temperate regions? A systematic map, Environmental Evidence, 9, 1–26, 2020.

Röthlisberger, V., Zischg, A. P., ., and Keiler, M.: A comparison of building value models for flood risk analysis, Natural Hazards and Earth System Sciences, 18, 2431–2453, https://doi.org/10.5194/nhess-18-2431-2018, 2018.

Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V. D., Lüdtke, S., Merz, B., Schröter, K., Vorogushyn, S., and others: Process-Based Flood Risk Assessment for Germany, Earth's Future, 9, https://doi.org/10.1029/2021EF002259, publisher: Wiley Online Library, 2021.

Saksena, S. and Merwade, V.: Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping, Journal of Hydrology, 530, 180–194, publisher: Elsevier, 2015.

Salman, A. M. and Li, Y.: Flood Risk Assessment, Future Trend Modeling, and Risk Communication: A Review of Ongoing Research, Natural Hazards Review, 19, 04018 011, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000294, 2018.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E.: A high-resolution global flood hazard model, Water Resources Research, p. 24, 2015.

Savage, J., Pianosi, F., Bates, P., Freer, J., and Wagener, T.: Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model, Water Resources Research, 52, 9146–9163, https://doi.org/10.1002/2015WR018198, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR018198, 2016.

Sayers, P. B., ed.: Flood risk: planning, design and management of flood defence infrastructure, ICE Publishing, London, oCLC: 817521040, 2012.

Schanze, J., Zeman, E., and Marsalek, J., eds.: Flood risk management: hazards, vulnerability and mitigation measures, no. vol. 67 in NATO science series. Series IV, Earth and environmental sciences, Springer, Dordrecht, oCLC: ocm71425441, 2006.

Scheuer, S., Haase, D., and Meyer, V.: Spatial explicit multi criteria flood risk—Fundamentals and semantics of multi criteria flood risk assessment, Flood risk and flood management, ed. TSW Wong, pp. 1–34, 2012.

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., and Merz, B.: How useful are complex flood damage models?, Water Resources Research, 50, 3378–3395, https://doi.org/10.1002/2013WR014396, 2014.

Schumann, G. J.-P., Andreadis, K. M., and Bates, P. D.: Downscaling coarse grid hydrodynamic model simulations over large domains, Journal of Hydrology, 508, 289–298, https://doi.org/10.1016/j.jhydrol.2013.08.051, 2014.

Seifert, I., Kreibich, H., Merz, B., and Thieken, A. H.: Application and validation of FLEMOcs – a flood-loss estimation model for the commercial sector, Hydrological Sciences Journal, 55, 1315–1324, https://doi.org/10.1080/02626667.2010.536440, 2010a.

Seifert, I., Thieken, A. H., Merz, M., Borst, D., and Werner, U.: Estimation of industrial and commercial asset values for hazard risk assessment, Natural Hazards, 52, 453–479, https://doi.org/10.1007/s11069-009-9389-9, 2010b.

Shaw, J., Kesserwani, G., Neal, J., Bates, P., and Sharifian, M. K.: LISFLOOD-FP 8.0: the new discontinuous Galerkin shallow-water solver for multi-core CPUs and GPUs, Geoscientific Model Development, 14, 3577–3602, https://doi.org/10.5194/gmd-14-3577-2021, publisher: Copernicus GmbH, 2021.

Shi, P.: Disaster Risk Science, IHDP/Future Earth-Integrated Risk Governance Project Series, Springer Singapore, Singapore, https://doi.org/10.1007/978-981-13-6689-5, 2019.

Sieg, T. and Thieken, A. H.: Improving flood impact estimations, Environmental Research Letters, 17, 064 007, https://doi.org/10.1088/1748-9326/ac6d6c, 2022.

Sieg, T., Vogel, K., Merz, B., and Kreibich, H.: Seamless estimation of hydrometeorological risk across spatial scales, Earth's Future, 7, 574–581, URL https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018EF001122, publisher: Wiley Online Library, 2019.

Simonovic, S.: Floods in a Changing Climate: Risk Management, Cambridge University Press, 2013.

Steinhausen, M., Paprotny, D., Dottori, F., Sairam, N., Mentaschi, L., Alfieri, L., Lüdtke, S., Kreibich, H., and Schröter, K.: Drivers of future fluvial flood risk change for residential buildings in Europe, Global Environmental Change, 76, 102 559, https://doi.org/10.1016/j.gloenvcha.2022.102559, 2022.

Sun, Q., Zhang, X., Zwiers, F., Westra, S., and Alexander, L. V.: A Global, Continental, and Regional Analysis of Changes in Extreme Precipitation, Journal of Climate, 34, 243 – 258, https://doi.org/10.1175/JCLI-D-19-0892.1, 2021.

Szönyi M. and Roezer V.: PERC Flood event review 'Bernd', Tech. rep., report, URL https://www.newsroom.zurich.de/documents/zurich-perc-analysis-bernd-english-version-423750, 2022.

Talbot, C. J., Bennett, E. M., Cassell, K., Hanes, D. M., Minor, E. C., Paerl, H., Raymond, P. A., Vargas, R., Vidon, P. G., Wollheim, W., and Xenopoulos, M. A.: The impact of flooding on aquatic ecosystem services, Biogeochemistry, 141, 439–461, https://doi.org/10.1007/s10533-018-0449-7, 2018.

Tariq, M. A. U. R., Farooq, R., and van de Giesen, N.: A Critical Review of Flood Risk Management and the Selection of Suitable Measures, Applied Sciences, 10, 8752, https://doi.org/10.3390/app10238752, number: 23 Publisher: Multidisciplinary Digital Publishing Institute, 2020.

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., and Slayback, D. A.: Satellite imaging reveals increased proportion of population exposed to floods, Nature, 596, 80–86, https://doi.org/10.1038/s41586-021-03695-w, number: 7870 Publisher: Nature Publishing Group, 2021.

Thieken, A. H., Müller, M., Kleist, L., Seifert, I., Borst, D., and Werner, U.: Regionalisation of asset values for risk analyses, Natural Hazards and Earth System Sciences, 6, 167–178, https://doi.org/10.5194/nhess-6-167-2006, 2006.

Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., and Merz, B.: Development and evaluation of FLEMOps – a new Flood Loss Estimation MOdel for the private sector, in: Flood Recovery, Innovation and Response I, vol. I, pp. 315–324, WIT Press, London, England, URL http://library.witpress.com/viewpaper.asp?pcode=FRIAR08-030-1, iSSN: 1743-3541, 1746-448X, 2008.

Thieken, A. H., Cammerer, H., Dobler, C., Lammel, J., and Schöberl, F.: Estimating changes in flood risks and benefits of non-structural adaptation strategies - a case study from Tyrol, Austria, Mitigation and Adaptation Strategies for Global Change, 21, 343–376, https://doi.org/10.1007/s11027-014-9602-3, 2016.

Tramblay, Y., Mimeau, L., Neppel, L., Vinet, F., and Sauquet, E.: Detection and attribution of flood trends in Mediterranean basins, Hydrology and Earth System Sciences, 23, 4419–4431, https://doi.org/10.5194/hess-23-4419-2019, publisher: Copernicus GmbH, 2019.

UNISDR: Technical Guidance for Monitoring and Reporting on Progress in Achieving the Global Targets of the Sendai Framework for Disaster Risk Reduction, Tech. rep., ., URL https://www.preventionweb.net/files/54970_techguidancefdigitalhr.pdf, 2017.

UNISDR, M.: UNISDR Terminology for Disaster Risk Reduction, 2009.

URS: Delta Risk Management Strategy (DRMS) Phase 1 - Risk Analysis Report - Final, Tech. rep., California Department of Water Resources, 2008.

US Army Corps of Engineers: Risk-Based Analysis for Flood Damage Reduction Studies, Tech. Rep. EM 1110-2-1619, US Army Corp of Engineers Washington, DC, URL https://www.publications.usace.army.mil/Portals/76/Publications/EngineerManuals/EM_1110-2-1619.pdf, 1996.

Vestby, J., Schutte, S., Tollefsen, A. F., and Buhaug, H.: Societal determinants of flood-induced displacement, Proceedings of the National Academy of Sciences, 121, e2206188 120, 2024.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., and Contributors), R. S. .: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

VNK: Flood Risk in the Netherlands. VNK2 The method in brief. TECHNICAL BACKGROUND, Tech. rep., VNK2 project office, 2012.

Vorogushyn, S., Bates, P. D., de Bruijn, K., Castellarin, A., Kreibich, H., Priest, S., Schröter, K., Bagli, S., Blöschl, G., Domeneghetti, A., Gouldby, B., Klijn, F., Lammersen, R., Neal, J. C., Ridder, N., Terink, W., Viavattene, C., Viglione, A., Zanardo, S., and Merz, B.: Evolutionary leap in large-scale flood risk assessment needed, Wiley Interdisciplinary Reviews: Water, 5, e1266, https://doi.org/10.1002/wat2.1266, 2018.

Vorogushyn, S., Apel, H., Kemter, M., and Thieken, A. H.: Analyse der Hochwassergefährdung im Ahrtal unter Berücksichtigung historischer Hochwasser, https://doi.org/10.5675/HyWa_2022.5_2, 2022.

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., and Kreibich, H.: Regional and Temporal Transferability of Multi-Variable Flood Damage Models, Water Resources Research, https://doi.org/10.1029/2017WR022233, 2018.

Walker, S. G.: On a lower bound for the Jensen inequality, SIAM Journal on Mathematical Analysis, 46, 3151–3157, 2014.

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., Beek, R. v., Bierkens, M. F. P., Ligtvoet, W., and Winsemius, H. C.: Assessing flood risk at the global scale: model setup, results, and sensitivity, Environmental Research Letters, 8, 044 019, https://doi.org/10.1088/1748-9326/8/4/044019, publisher: IOP Publishing, 2013.

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., and Winsemius, H. C.: Usefulness and limitations of global flood risk models, Nature Climate Change, 5, 712–715, https://doi.org/10.1038/nclimate2742, 2015a.

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., and Winsemius, H. C.: Usefulness and limitations of global flood risk models, Nature Climate Change, 5, 712–715, https://doi.org/10.1038/nclimate2742, 2015b.

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G. A., Schäfer, A., Stanley, T., Veldkamp, T. I. E., and Winsemius, H. C.: Review article: Natural hazard risk assessments at the global scale, Natural Hazards and Earth System Sciences, 20, 1069–1096, https://doi.org/10.5194/nhess-20-1069-2020, 2020.

White, G. F.: Human Adjustment to Floods. A Geographical Approach to the Flood Problem in the United States, Ph.D. thesis, The University of Chicago, Chicago, 1945.

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., and Erickson, T. A.: Validation of a 30 m resolution flood hazard model of the conterminous United States: 30 m RESOLUTION FLOOD MODEL OF CONUS, Water Resources Research, 53, 7968–7986, https://doi.org/10.1002/2017WR020917, 2017.

Wing, O. E. J., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., Neal, J. C., Porter, J. R., and Kousky, C.: Inequitable patterns of US flood risk in the Anthropocene, Nature Climate Change, 12, 156–162, https://doi.org/10.1038/s41558-021-01265-6, 2022.

Winsemius, H. C., Van Beek, L. P. H., Jongman, B., Ward, P. J., and Bouwman, A.: A framework for global river flood risk assessments, Hydrology and Earth System Sciences, 17, 1871–1892, https://doi.org/10.5194/hess-17-1871-2013, publisher: Copernicus GmbH, 2013.

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P., and Ward, P. J.: Global drivers of future river flood risk, Nature Climate Change, 6, 381–385, https://doi.org/10.1038/nclimate2893, number: 4 Publisher: Nature Publishing Group, 2016.

Wyncoll, D. and Gouldby, B.: Integrating a multivariate extreme value method within a system flood risk analysis model, Journal of Flood Risk Management, 8, 145–160, https://doi.org/10.1111/jfr3.12069, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfr3.12069, 2015.

Wünsch, A., Herrmann, U., Kreibich, H., and Thieken, A. H.: The Role of Disaggregation of Asset Values in Flood Loss Estimation: A Comparison of Different Modeling Approaches at the Mulde River, Germany, Environmental Management, 44, 524–541, https://doi.org/10.1007/s00267-009-9335-3, 2009.

Xafoulis, N., Kontos, Y., Farsirotou, E., Kotsopoulos, S., Perifanos, K., Alamanis, N., Dedousis, D., and Katsifarakis, K.: Evaluation of Various Resolution DEMs in Flood Risk Assessment and Practical Rules for Flood Mapping in Data-Scarce Geospatial Areas: A Case Study in Thessaly, Greece, Hydrology, 10, 91, https://doi.org/10.3390/hydrology10040091, 2023.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resources Research, 47, publisher: Wiley Online Library, 2011.

Yarveysi, F., Alipour, A., Moftakhari, H., Jafarzadegan, K., and Moradkhani, H.: Block-level vulnerability assessment reveals disproportionate impacts of natural hazards across the conterminous United States, Nature Communications, 14, 4222, https://doi.org/10.1038/s41467-023-39853-z, number: 1 Publisher: Nature Publishing Group, 2023.

Yu, D., Yin, J., Wilby, R. L., Lane, S. N., Aerts, J. C., Lin, N., Liu, M., Yuan, H., Chen, J., Prudhomme, C., et al.: Disruption of emergency response to vulnerable populations during floods, Nature Sustainability, 3, 728–736, 2020.

Zevenbergen, C., Cashman, A., Evelpidou, N., Pasche, E., Garvin, S., and Ashley, R.: Urban flood management, CRC Press, 2010.

Appendix A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 Inundation Performance Metrics

Quantitative evaluation of flood inundation grids is commonly accomplished using a diverse set of metrics that communicate and quantify over- and under-predictions and their proportions. To compute these metrics, simulations for maximum inundation are evaluated against some observed binary data grid of wet and dry cells. First, each cell is classified according to Table A.1 by comparing the simulated to the observed data grids to generate a confusion map. From this confusion map, the total counts of each of the four classifications is computed. These total counts are then used to calculate the domain-wide inundation metrics commonly used in flood inundation evaluation shown in Table A.2.

Table A.1: Inundation confusion matrix. For a given simulation, each cell in the domain is compared to the corresponding cell in the observed grid and classified according to this table. Adapted from Wing et al. (2017).

		Simulated	
		Wet	Dry
Observed	Wet	True Positive (TP)	False Negative (FN)
	Dry	False Positive (FP)	True Negative (TN)

Table A.2: Flood inundation performance metrics. See Table A.1 for acronyms. Adapted from Wing et al. (2017).

Metric	Equation	Poor	Perfect	Description
Critical Success Index	$\frac{TP}{TP+FP+FN}$	0	1	ratio of accurate wet cells to total wet cells and missed wet cells
Hit Rate	$\frac{TP}{TP+FN}$	0	1	portion of observed wet cells reproduced by the model
False Alarms	$\frac{FP}{TP+FP}$	1	0	portion of modelled wet cells which are erroneous
Error Bias	$\frac{FP}{FN}$	0 or inf	1	ratio of over-predictions to under-predictions

A.2 Hydrodynamic Model Calibration

To obtain accurate water level grids at coarse ($s2 = 32m$) and fine ($s1 = 4m$) resolutions, twin hydrodynamic models are constructed in the RIM2D platform and calibrated using a mix of brute force and scipy's implementation of the Newton-Conjugate Gradient algorithm (Nocedal and Wright, 2006; Virtanen et al., 2020). Roughness values for built-up and channel/floodplain are treated as two (independent) free parameters for the optimization. A single performance metric, Critical Success Index (CSI) defined in Table A.2, is calculated against the observed inundation for each iteration and used to optimize with the free parameters. Optimization trials were undertaken on a Tesla P100 GPU using python scripts.

Results of the two calibration trials are shown in Fig. A.1 and A.2. The performance metrics shown in Table A.2 are also shown; however, only CSI was used for optimization. In general, the fine ($s1 = 4m$) model replicates the target inundation with over- and under-predictions roughly balancing (Error Bias = 1.2) while the coarse model ($s2 = 32m$) generally under-predicts when CSI is optimized (Error Bias = 0.33). Focusing on water surface elevations (which were not part of the optimization), the fine ($s1 = 4m$) model has lower WSE values upstream and higher WSE values downstream when compared to the coarse ($s2 = 32m$); likely owing to the difficulties in modelling the narrower channel in this region at the coarser resolutions. While this difference would be problematic for some hydrodynamic model applications, here we focus on inundation extents – not elevations. If elevations and flow dynamics were the focus, the coarse hydrodynamic model would be an inappropriate choice, because the model resolution is about 3 times larger than the width of the river, which results in the observed deviations in water slope profile compared to the fine resolution model. For an estimation of the flood extent, however, the coarse model can provide useful results despite the deficiencies in simulating the flow dynamics. Because of our focus on flood extent, we use a simple calibration of two roughness parameters to optimize the Critical Success Index which is a measure of fit to the observed inundation extents. It is therefore not surprising that inundation is reproduced well by both models while water elevations are less satisfactory. A more sophisticated (e.g., multi-metric optimization) calibration could have been pursued to try and address this; however, as our paper focuses on downscaling (not model calibration) we felt this would be distracting. Note the performance metrics reported in the manuscript are computed on a smaller domain.

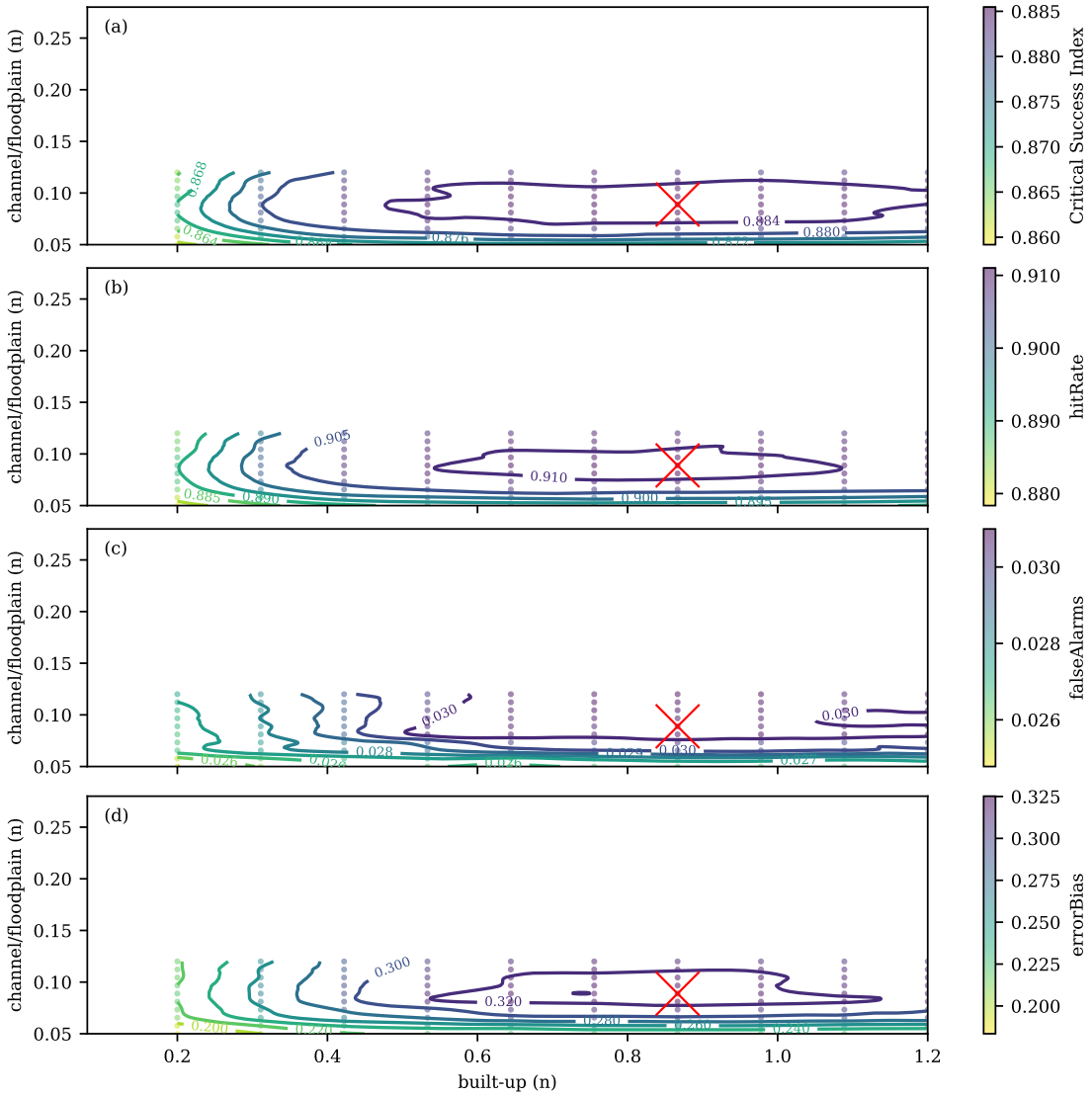


Figure A.1: Calibration results for 32 m hydrodynamic model showing the four metrics from Table A.2. Points denote individual model runs (at the shown roughness) and contours are computed via interpolation of the metric value at each point. Red 'X' marks the optimal (using the maximum CSI) and the parameterization used for downscaling.

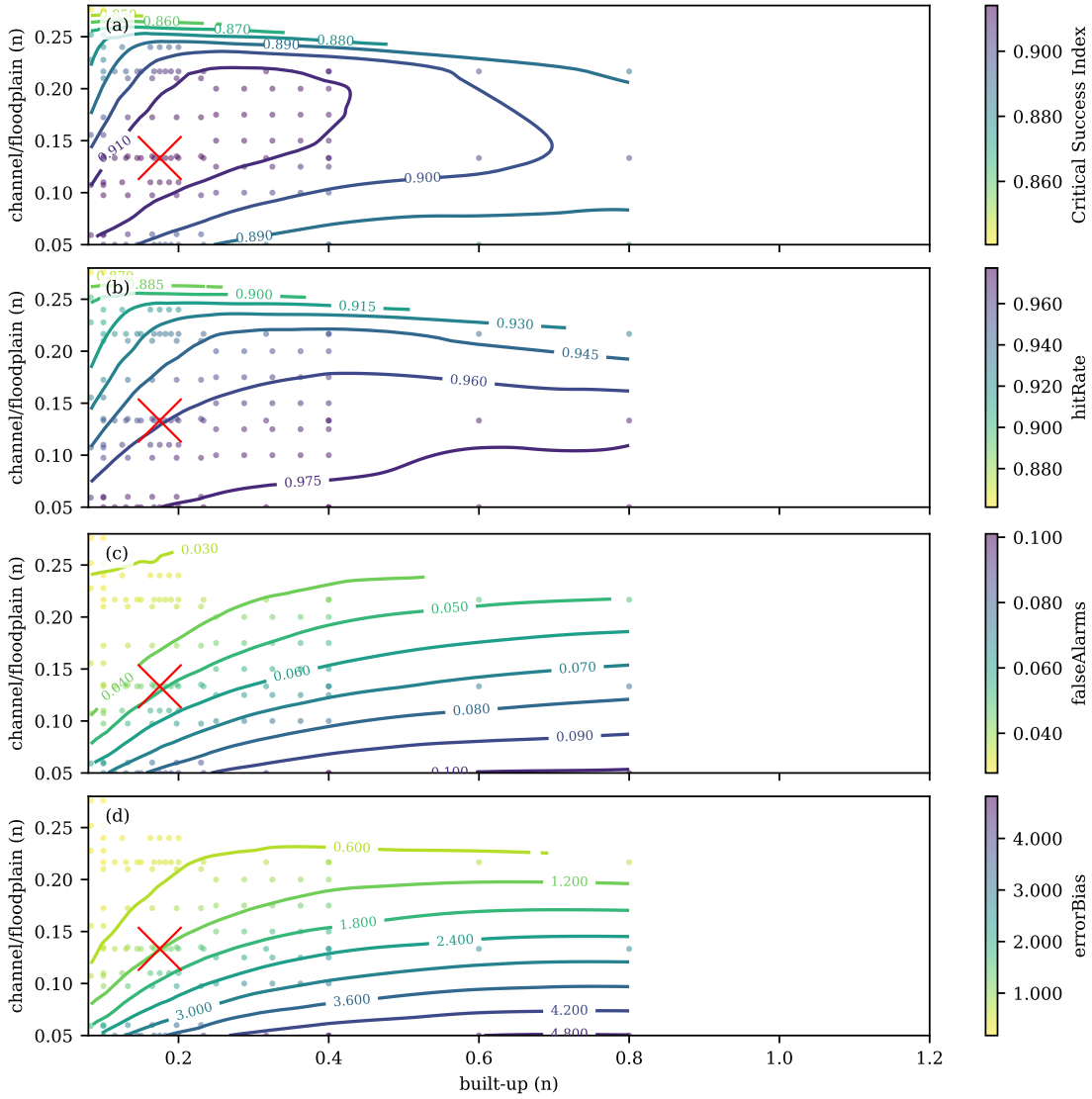


Figure A.2: Calibration results for 4 m hydrodynamic model similar to Figure A.1.

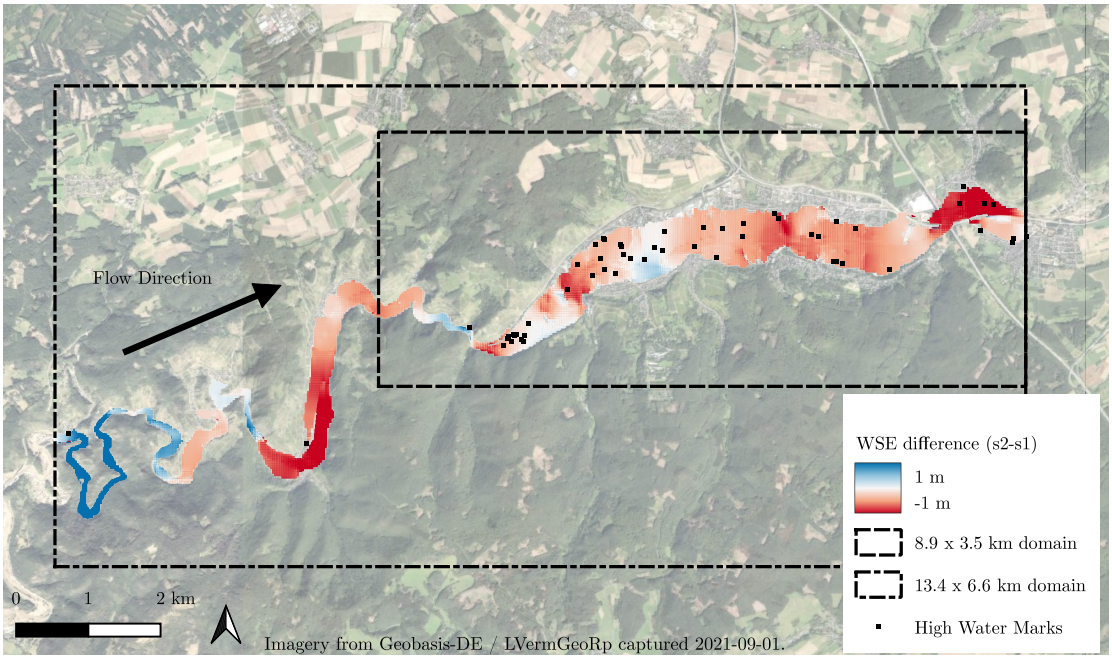


Figure A.3: WSE max difference between coarse ($s2 = 32m$) and fine ($s1 = 4m$) models at their respective optimum roughnesses clipped to intersecting inundation region. Red denotes regions where the fine ($s1 = 4m$) solution yielded higher or larger water depths than the coarse ($s2 = 32m$). Domain used for hydrodynamic modelling (13.4 x 6.6 km) and subset used for downscaling analysis (8.9 x 3.5 km) shown in black for reference.

A.3 Additional Results Figures

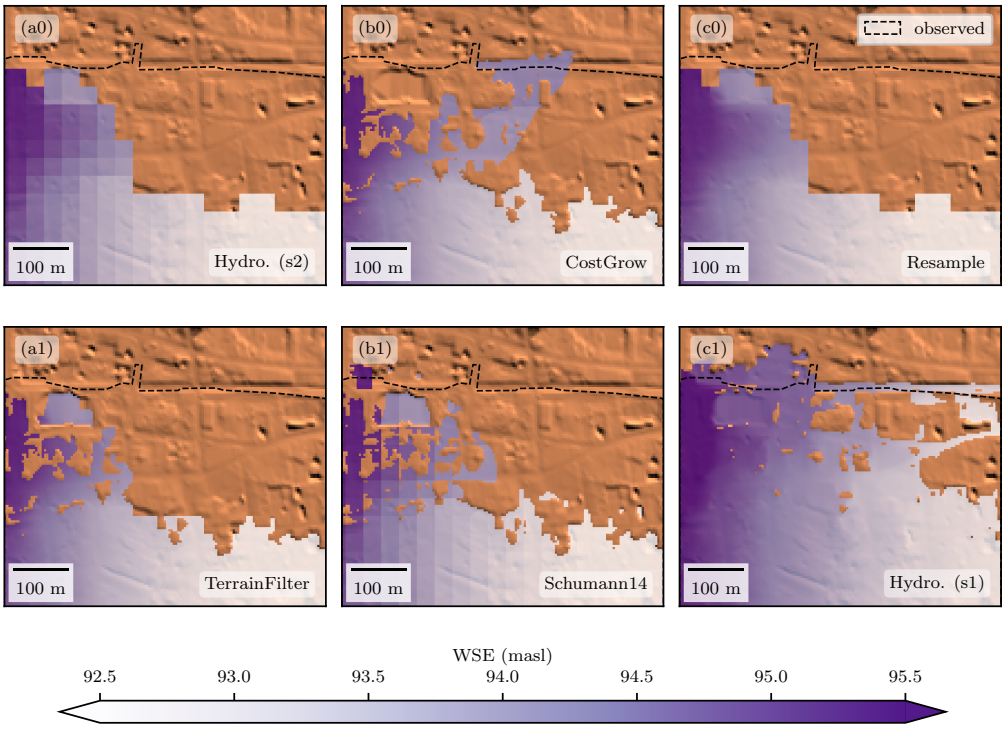


Figure A.4: Downscale and hydrodynamic model WSE detail results as in Figure 8 for a separate location.

Appendix B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 Analytical Results: Detail

This supplement provides an analytical treatment of errors introduced by flood hazard data aggregation described in the main text. This employs a novel *resample case framework* to investigate two typical aggregation routines. Aggregating or upscaling involves transferring data grids from fine ($s1$) to coarse ($s2$) resolutions where the resolutions of the two domains can be expressed as: $s1 < s2$, often expressed in meters. From this, and the fact that both grids have the same extents, we can say:

$$\frac{s2^2}{s1^2} = \frac{N_1}{N_2} \tag{B.1}$$

where N is the total cell count of the corresponding grid. See Section 3.2 of the main text for further context and equations.

B.1.1 Aggregation Routines

The two aggregation routines considered here are summarized in the main text. These routines can generally be formulated as:

$$DEM_{s2}, WSH_{s2}, WSE_{s2} = f[DEM_{s1}, WSH_{s1}, WSE_{s1}, s2] \tag{B.2}$$

where f is some aggregation routine. All aggregation routines first act on local groups of $s1$ cells, who we index with i, to obtain a new $s2$ value with index j. In this way, each i cell maps to a j index with a many:1 relation. The following sections elaborate on the two routines.

First Routine: WSH Averaging

In the *WSH Averaging* routine, local $s1$ groups of DEM and WSH grids are simply averaged to yield new $s2$ cells, using Equations 3.4 and 3.5. Applying these to the full $s1$ domain yields aggregated DEM_{s2} and WSH_{s2} grids. Equation 3.1 is then used to compute WSE_{s2}:

$$\begin{aligned} DEM_{s2,j} &= \overline{DEM_{s1,i}} \\ WSH_{s2,j} &= \overline{WSH_{s1,i}} \\ WSE_{s2,j} &= DEM_{s2,j} + WSH_{s2,j} \end{aligned} \tag{B.3}$$

Second Routine: WSE Averaging

To satisfy Equation 3.3, *WSE Averaging* is more complicated, requiring a two step process: first, a wet averaging via Equation 3.6, then the cells violating Equation 3.3 are masked before computing WSH_{s2}

via a modified Equation 3.1:

$$
\begin{aligned}
DEM_{s2,j} &= \overline{DEM_{s1,j}} \\
WSE_{s2,j} &= \begin{cases} null & \text{if } \overline{WSE_{s1,j}} \leq \overline{DEM_{s1,j}} \\ \overline{WSE_{s1,j}} & \text{else} \end{cases} \\
WSH_{s2,j} &= \begin{cases} 0 & \text{if } WSE_{s2,j} = null \\ WSE_{s2,j} - DEM_{s2,j} & \text{else} \end{cases}
\end{aligned}
\tag{B.4}
$$

Both routines are summarized in Figure 3.1.

B.1.2 Bias Types

Using the *resample case framework* defined in Section 3.3.2 and summarized in Figure 3.2, we investigate bias or systematic errors introduced by the two aggregation routines presented above on four metrics of importance to flood models: two primary metrics, water depth (WSH) and water surface elevation (WSE), and two derivative metrics, inundation area (A), and volume (V).

Global Bias

To attribute errors to some aggregation routine, we define *global* bias as the difference between some metric computed with the aggregated vs. the raw grid. This can be formulated for some metric M, which is a reducing function of grid G (e.g., $M[G] = mean[G]$) and the aggregation routine f as:

$$
Bias_{global}[M, f, s2] = M[f[G_{s1}, ...]] - M[G_{s1}] \tag{B.5}
$$

Local Bias

For the primary grid metrics (WSH and WSE), Equation B.5 can alternatively be computed locally, by first calculating the difference between each i and its corresponding j cell, before applying the reducing function M to obtain the grid bias:

$$
Bias_{local}[M, f, s2] = M[\sum_{i=1}^{N1}(f[G_{s1}, ...]_j - G_{s1,i})] \quad \text{where} \quad i \neq null \tag{B.6}
$$

For metrics computed from the WSE grid, this local bias can of course only be computed in regions inundated by both $s1$ and $s2$ grids (see Equation 3.2). For consistency, we apply this same constraint to WSH metrics (i.e., cells where $WSH = 0$ are excluded). While this masks the performance of a routine in dry regions, it provides a consistent way to separate the reporting of bias in local variables from bias in inundation area (which is reported as a secondary metric). Extending the *resample case* framework to these two definitions of bias, and assuming that M is linear in the wet domain, it follows that:

$$
Bias_{local} = \begin{cases} n/a & \text{if } DD \\ \neq Bias_{global} & \text{if } DP \\ \neq Bias_{global} & \text{if } WP \\ = Bias_{global} & \text{if } WW \end{cases} \tag{B.7}
$$

In other words, when computed on specific *resample case* regions, $Bias_{local}$ may only differ from $Bias_{global}$ in partial regions (DP and WP) and is undefined in DD regions.

B.1.3 Bias in Flood Depths (WSH)

For computing bias in flood depths (WSH), we focus on the grid average at resolution $s = s1$ or $s = s2$:

$$M[s] = \frac{1}{N_s} \sum_{i=1}^{N_s} WSH_{s,i} \tag{B.8}$$

where N_s is the count of cells i within the grid at resolution s, which is the global version of Equation 3.5. Expanding the global bias in Equation B.5 with this yields:

$$Bias_{global}[f, s2] = \frac{1}{N_2} \sum_{j=1}^{N_2} WSH_{s2,j} - \frac{1}{N_1} \sum_{i=1}^{N_1} WSH_{s1,i} \tag{B.9}$$

A similar expansion for the local bias in Equation B.6 yields:

$$Bias_{local}[f, s2] = \frac{1}{N_2} \sum_{j=1}^{N_2} (WSH_{s2,i,j} - WSH_{s1,i,j}) \tag{B.10}$$

First Routine: WSH Averaging

Comparing Equation B.9 with our definition of the *WSH Averaging* routine (Equation B.3) shows this routine has no bias in global WSH. To explore the local bias of this routine, we first examine regions classified by the WW resample case defined by Equation 3.7, where $min(WSE_{s1,i}) > max(DEM_{s1,i})$. This can be re-written using Equation 3.1 in terms of WSH for convenience as:

$$WW \equiv min(WSH_{s1,i}) > 0 \tag{B.11}$$

In other words, all i cells within j are wet. Expanding Equation B.9 for a j group of i cells, then substituting with Equation B.3 yields:

$$\begin{aligned} Bias[f, s2] &= \frac{1}{N_2} \sum_{j=1}^{N_2} \left(\frac{1}{N_{12}} \sum_{i=1}^{N_{12}} WSH_{s1,j,i} - WSH_{s1,j,i}\right) \\ &= \frac{1}{N_2} \sum_{j=1}^{N_2} \left(\frac{N_{12}}{N_{12}} (WSH_{s1,j,1} + WSH_{s1,j,2} + ...) - (WSH_{s1,j,1} + WSH_{s1,j,2} + ...)\right) \\ &= 0 \end{aligned} \tag{B.12}$$

where N_{12} is the count of $s1$ cells (indexed with i) contributing to a new $s2$ cell with index j. The terms in line two cancel because in the WW region (Equation B.11) i blocks and j blocks are equivalent. This is intuitive if we consider the absence of *null* values in these WW regions.

Taking a similar approach to evaluate the WP region, where:

$$WP \equiv min(WSH_{s1,i}) = 0 \quad \text{and} \quad \overline{DEM_{s1,i}} < \overline{WSE_{s1,i}} \tag{B.13}$$

in other words, some i cells now are dry, but the group average is still higher than the DEM average. Following our definition of local bias (Equation B.6), the domain of computation for this metric is further constrained to cells which are non-*null* in both the $s1$ and $s2$ grids:

$$\text{for } i \text{ where } WSH_{s1,i} > 0 \text{ and } WSH_{s2,i} > 0 \tag{B.14}$$

Starting from Equation B.12 for the combined domain of Equation B.13 and B.14 and adopting $i = 2$ as an illustrative dry cell yields:

$$\begin{aligned} Bias_{local}[f, s2] &= \frac{1}{N_2}\sum_{j=1}^{N_2}(\frac{N_{wet,j}}{N_{12}}(WSH_{s1,j,1} + \cancelto{0}{WSH_{s1,j,2}} + ...) - (WSH_{s1,j,1} + ...)) \\ &= \frac{1}{N_2}\sum_{j=1}^{N_2}(\frac{N_{wet,j}}{N_{12}} - 1) \\ &< 0 \end{aligned}$$

because $N_{wet} < N_{12}$ by definition. In other words, if the calculation is limited to the wet domain (per our definition of local bias in Equation B.6), the $s2$ values are systematically low, as the aggregation does *not* have the same limitation, and the $s2$ value is therefore pulled down by the dry neighbours. The same result holds for DP regions, with the bias likely being more severe as we assume that $N_{wet,DP} < N_{wet,WP}$.

Second Routine: WSE Averaging

To derive the global bias introduced by the *WSE Averaging*, we first substitute Equation B.4 into Equation B.9:

$$Bias_{global}[f, s2] = -\frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} + \begin{cases} 0 & \text{if } WSE_{s2,j} = null \\ \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i}} - \overline{DEM_{s1,i}}) & \text{else} \end{cases} \tag{B.15}$$

Starting with the DD case ($WSH = 0$ and $WSE = null$), all terms reduce to zero. For the DP case, we can re-write the domain condition from Equation 3.7 by substituting in WSH using Equation 3.1:

$$DP \equiv max(WSH_{s1,i}) > 0 \text{ and } \overline{DEM_{s1,j}} \geq \overline{WSE_{s1,j}} \tag{B.16}$$

In other words, there are some wet i cells, but their (wet) average is less than the (wet+dry) average of the terrain. Equation B.4 states that $WSE_{s2,j} = null$ for this condition, reducing Equation B.15 to:

$$\begin{aligned} Bias_{global}[f, s2] &= -\frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} + 0 \\ &< 0 \end{aligned}$$

because $max(WSH_{s1,i}) > 0$ implies $\sum_{i=1}^{N_1} WSH_{s1,i} > 0$. In other words, because this routine always yields at least one dry $WSH = 0$ value in DP_i cells, the bias is always negative in this region.

For the WP case, the domain condition is stated above in Equation B.13. This is the most interesting case as $WSE_{s2,j}$ is non-null and the second part of Equation B.15 therefore reduces to the non-zero term:

$$Bias_{global}[f, s2] = \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i}} - \overline{DEM_{s1,i}}) - \frac{1}{N_1}\sum_{i=1}^{N_1} WSH_{s1,i} \tag{B.17}$$

To evaluate this, we separate DEM averaging into *wet* and *dry* regions for later comparison knowing

$\overline{DEM_{all}} = \overline{DEM_{wet}} + \overline{DEM_{dry}}$:

$$Bias_g[f, s2] = \frac{1}{N_2}\sum_{j=1}^{N_2}(\overline{WSE_{s1,i,wet}} + \cancelto{0}{\overline{WSE_{s1,i,dry}}} - \overline{DEM_{s1,i,wet}} - \overline{DEM_{s1,i,dry}}) \tag{B.18}$$

$$- \frac{1}{N_{wet}}\sum_{i=1}^{N_{wet}} WSH_{s1,i} - \cancelto{0}{\frac{1}{N_{dry}}\sum_{i=1}^{N_{dry}} WSH_{s1,i}}$$

$$= \frac{1}{N_2}\sum_{j=1}^{N_2}((\overline{WSE_{s1,i,wet}} - \overline{DEM_{s1,i,wet}}) - \cancelto{0}{\frac{1}{N_{wet}}\sum_{i=1}^{N_{wet}} WSH_{s1,i}})$$

$$- \frac{1}{N_2}\sum_{j=1}^{N_2}\overline{DEM_{s1,i,dry}}$$

$$= -\frac{1}{N_2}\sum_{j=1}^{N_2}\overline{DEM_{s1,i,dry}}$$

$$< 0$$

in other words, in WW regions this routine introduces a negative bias equivalent to the average value of the dry DEM cells. For the WW case and the domain condition ($min(WSH_{s1,i}) > 0$), Equation B.17 reduces to zero.

Examining the local bias (Equation B.10) of WSH produced by the *WSE Averaging* routine, Equation B.7 states WW will also have no local bias. For DP and DD regions, recall that this routine (Equation B.3) returns *dry* values for j cells, therefore these are excluded per our definition of local bias (Equation B.6). For the remaining WP regions, Equation B.18 still holds; however, the summation domain will differ and therefore so will the magnitude.

B.1.4 Bias in Water Surface Elevation (WSE)

Like WSH, WSE is a primary variable and we therefore focus on the grid average at resolution $s = s1$ or $s = s2$:

$$M[s] = \frac{1}{N_{s,wet}}\sum_{i=1}^{N_{s,wet}} WSE_{s,i}$$

where dry values are ignored (as in Equation 3.6). Expanding the global bias with this as in Equation B.9 yields:

$$Bias_{global}[f, s2] = \frac{1}{N_{2,wet}}\sum_{j=1}^{N_{2,wet}} WSE_{s2,j} - \frac{1}{N_{1,wet}}\sum_{i=1}^{N_{1,wet}} WSE_{s1,i} \tag{B.19}$$

And for the local bias:

$$Bias_{local}[f, s2] = \frac{1}{N_{2,wet}}\sum_{j=1}^{N_{2,wet}} (WSE_{s2,i,j} - WSE_{s1,i,j}) \tag{B.20}$$

First Routine: WSH Averaging

Substituting our definition of the *WSH Averaging* routine from Equation B.3 into Equation B.19 to obtain the global bias yields:

$$Bias_{global}[f, s2] = \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\overline{DEM_{s1,i}} + \overline{WSH_{s1,i}}) - \frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} (DEM_{s1,i} + WSH_{s1,i}) \tag{B.21}$$

For the DD case, WSE is not defined; while for the WW all terms cancel to zero. For the DP case, the domain is provided in Equation B.16. Expanding Equation B.21 with this and separating into wet and dry regions again yields:

$$\begin{aligned} Bias_{global}[f, s2] &= \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\overline{DEM_{s1,i,wet}} + \overline{DEM_{s1,i,dry}} + \overline{WSH_{s1,i,wet}} + \cancelto{0}{\overline{WSH_{s1,i,dry}}}) \\ &\quad - \frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} (DEM_{s1,i,wet} + WSH_{s1,i,wet}) \\ &= \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} (\overline{DEM_{s1,i,wet}} - \cancelto{0}{\frac{1}{N_{1,wet}} DEM_{s1,i,wet}} \\ &\quad + \overline{WSH_{s1,i,wet}} - \cancelto{0}{\frac{1}{N_{1,wet}} WSH_{s1,i,wet}} + \overline{DEM_{s1,i,dry}}) \\ &= \frac{1}{N_{2,wet}} \sum_{j=1}^{N_{2,wet}} \overline{DEM_{s1,i,dry}} \\ &> 0 \end{aligned} \tag{B.22}$$

Similar to the $Bias_{global}[WSH]$ of the *WSE Averaging* routine derived above (Equation B.18), the magnitude of the bias is related to $\overline{DEM_{s1,dry}}$, but in the opposite direction. A similar result holds for the WP case; however, at a lesser magnitude assuming $\overline{DEM_{s1,dry,WP}} < \overline{DEM_{s1,dry,DP}}$.

For the local bias of WSE (Equation B.20) produced by the *WSH Averaging* routine, again Equation B.7 shows WW will also have no local bias (and DD cells are *null*). Similarly, Equation B.22 holds for the partial regions (DP and WP).

Second Routine: WSE Averaging

To derive the global bias introduced by the *WSE Averaging* routine, we substitute Equation B.4 into Equation B.19:

$$Bias_{global}[f, s2] = -\frac{1}{N_{1,wet}} \sum_{i=1}^{N_{1,wet}} WSE_{s1,i} + \begin{cases} null & \text{if } \overline{WSE_{s1,j}} = null \\ \overline{WSE_{s1,j}} & \text{else} \end{cases} \tag{B.23}$$

For DD and DP regions, all terms are *null*. For WW regions, all terms reduce to zero. For WP regions, all terms also reduce to zero per Equation 3.6. Given that both local bias and WSE are only defined in wet regions, WSE global bias is equivalent to local bias for the *WSE Averaging* routine.

B.1.5 Bias in Inundation Area (A)

Inundation area (A) is an important secondary metric for flood-related analysis and can be simply computed with a binary transformation from either the WSE or the WSH grid using Equation 3.2 multiplied by the pixel area (s^2):

$$A_k = \begin{cases} 0 & \text{if } WSH = 0 \text{ or } WSE = null \\ s^2 & \text{else} \end{cases} \tag{B.24}$$

where $k = i$ or j. For computing bias from aggregation routines, we focus on the total grid inundation area:

$$M[s] = \sum_{k=1}^{N} A_k \tag{B.25}$$

We select this metric, rather than average area, to align with standard metrics in flood literature. However, because the grid sizes do not change, total and average area only differ by a scalar ($\frac{1}{N_s}$). By combining Equation B.24 with the *resample case framework* (Equation 3.7), the total inundation area of some group j, computed on the original fine ($s1$) grid, can be written as:

$$A_{s1,j} = \begin{cases} 0 & \text{if } DD \\ 0 < x < s2^2 & \text{if } DP \\ 0 < x < s2^2 & \text{if } WP \\ s2^2 & \text{if } WW \end{cases} \tag{B.26}$$

In other words, partial regions have some dry cells, DD regions have all dry cells, and WW regions have no dry cells. With this, we can compare against the area $A_{s2,j}$ which is computed on the aggregated grids to calculate the bias of this metric. For this metric, local bias and global bias are equivalent by definition.

First Routine: WSH Averaging

Combining Equation B.3 and Equation 3.7, inundation area for this routine can be written for each j in terms of *resample case* as:

$$A_{s2,j} = \begin{cases} 0 & \text{if } DD \\ s2^2 & \text{if } DP \\ s2^2 & \text{if } WP \\ s2^2 & \text{if } WW \end{cases}$$

Comparing this to Equation B.26 shows that the partial zones have a positive bias while WW and DD have no bias.

Second Routine: WSE Averaging

Combining Equation B.4 and Equation 3.7, inundation area can be written as:

$$A_{s2,j} = \begin{cases} 0 & \text{if } DD \\ 0 & \text{if } DP \\ s2^2 & \text{if } WP \\ s2^2 & \text{if } WW \end{cases}$$

The only difference with the previous routine being the DP region. From this, it follows that DP has a negative bias and WP has a positive bias, while the remaining have no bias.

B.1.6 Bias in Flood Volume (V)

Flood volume (V) is a metric of interest to hydrodynamic models which assume volume conservation (as an extension of mass conservation). For our evaluation, flood volume is computed from the depth grid (WSH) and the geospatial area (A):

$$\begin{aligned} V_k &= WSH_k * \cancelto{1 \text{ for wet}}{A_k} * s^2 \\ &= WSH_k * s^2 \end{aligned}$$

Like inundation area, here we focus on total grid volume:

$$\begin{aligned} M[s] &= \sum_{k=1}^{N_s} V_k \\ &= s^2 \sum_{k=1}^{N_s} WSH_k \end{aligned}$$

Expanding Equation B.5 with this yields:

$$Bias[f, s2] = s2^2 \sum_{j=1}^{N_2} WSH_{s2,j} - s1^2 \sum_{i=1}^{N_1} WSH_{s1,i} \tag{B.27}$$

First Routine: WSH Averaging

To evaluate volume (V) bias for the *WSH Averaging* routine, we substitute Equation B.3 into Equation B.27 which yields:

$$\begin{aligned} Bias[f, s2] &= s2^2 \sum_{j=1}^{N_2} \overline{WSH_i} - s1^2 \sum_{i=1}^{N_1} WSH_i \\ &= s2^2 (N_2 \overline{WSH_i}) - s1^2 (N_1 \overline{WSH_i}) \\ &= \overline{WSH_i} \cancelto{0}{\left(N_2 s2^2 - N_1 s1^2 \right)} \end{aligned}$$

which cancels to zero following Equation B.1.

Second Routine: WSE Averaging

As discussed above, the *WSE Averaging* routine has no WSH bias in the WW and DD domains, so it follows V bias is similarly absent. For the DP and WP case, WSH bias is negative, so it follows V bias will also be negative.

B.2 Computational Results: Additional Figures

Additional figures for the computational analysis are provided below. See the main text for details.

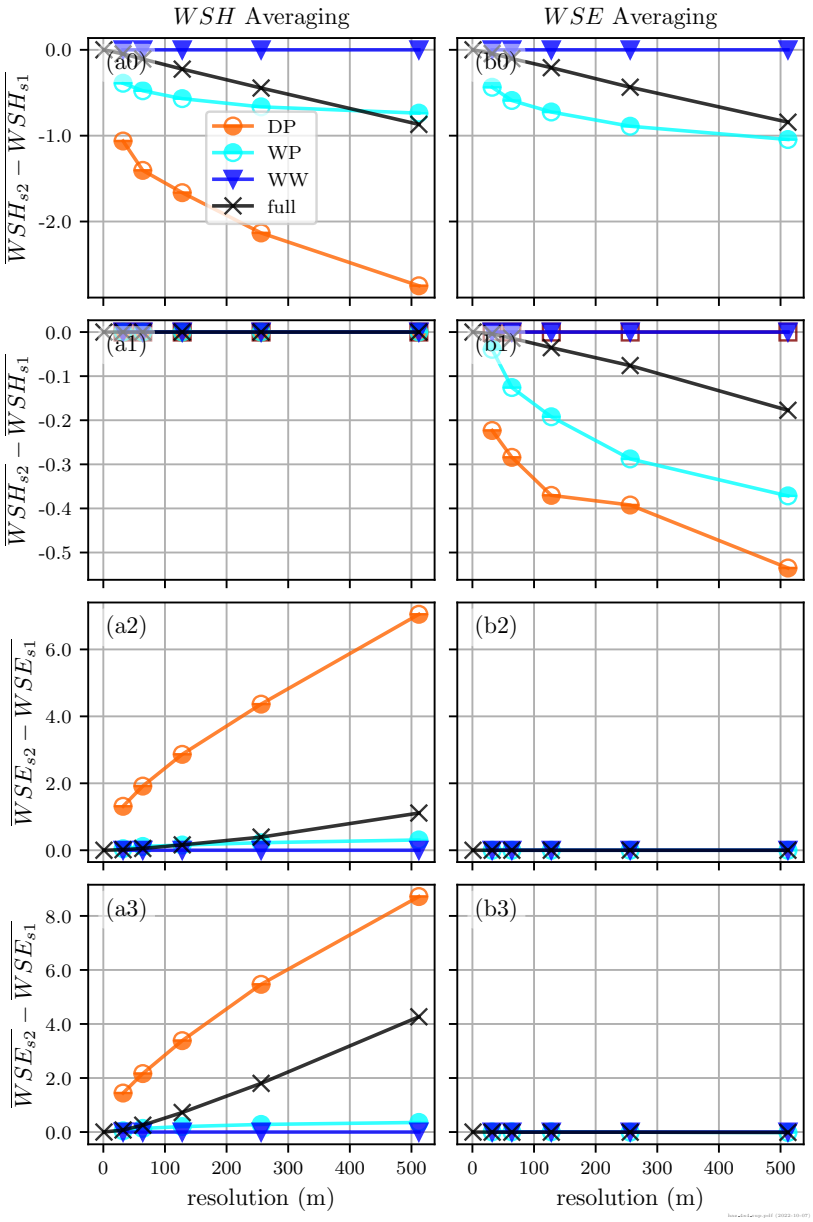


Figure B.1: Full domain computation results. See main text for details

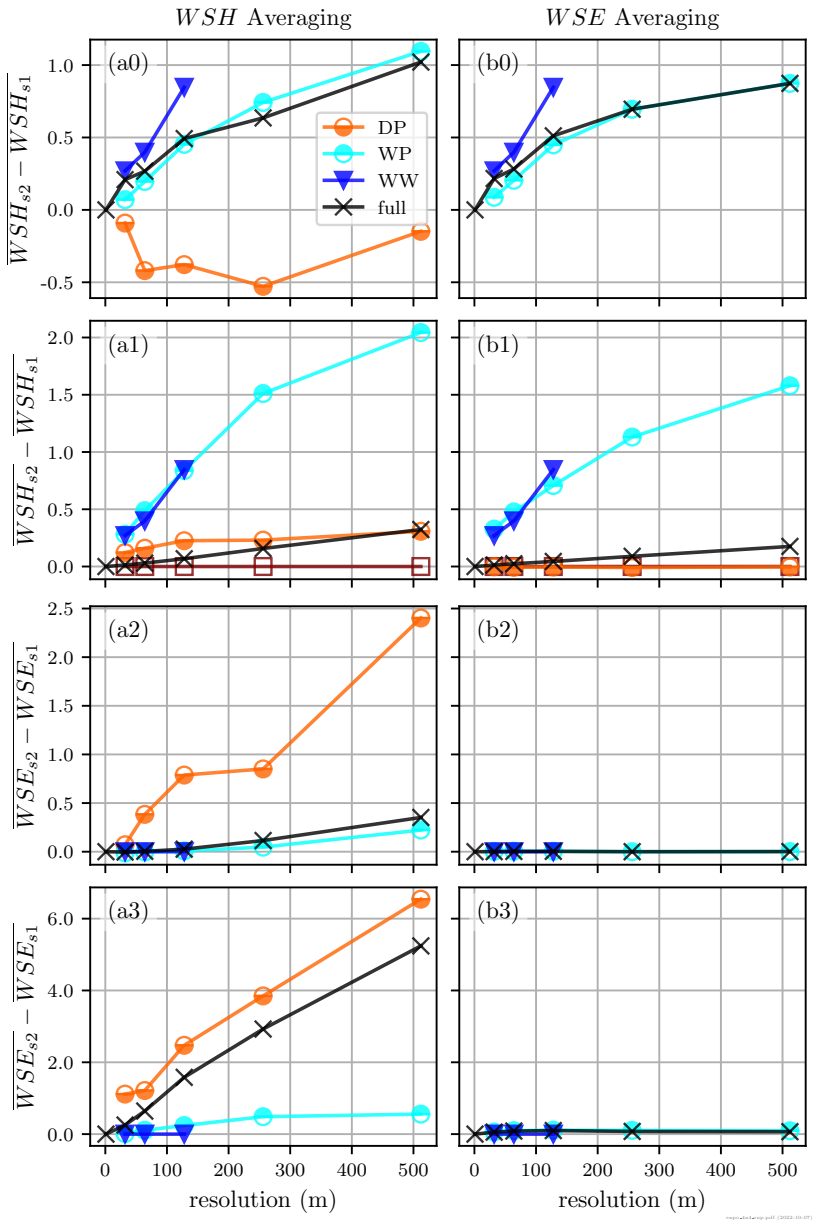


Figure B.2: Exposed domain computation results. See main text for details

B.3 Aggregation as an Analog for Hydrodynamic Modelling

This supplement compares coarse hydrodynamic modelling to the two aggregation routines to evaluate the transferability of the behaviour and bias observed during aggregation to coarse modelling. In other words, we seek to establish whether or not the study of artifacts and bias which emerge during aggregation provides an analog through which to study artifacts and bias which may emerge during coarse hydrodynamic modelling.

B.3.1 Methods

In the first part of the analysis, four sets of WSH and WSE grids are constructed using twin fine ($s1 = 4m$) and coarse ($s2 = 32m$) calibrated hydrodynamic models along with two aggregation routines. In the second part, these grids are compared to observations of high water marks and inundation extents to establish the accuracy of the respective methods. In the third part, the two aggregation routines are compared to the coarse hydrodynamic model results to evaluate the extent to which aggregation (of fine model results) can reproduce coarse model results. In the final part of the evaluation, the same metrics reported in the manuscript are computed on the four grid sets to ascertain whether the same bias observed through aggregation is present in the coarse modelling.

Aggregation Routines

We consider the same two routines as the manuscript for transforming fine-resolution ($s1 = 4m$) grids to a coarse-resolution ($s1 = 32m$) through averaging or aggregating local groups (see Figure 3.1). Respective fine-resolution grids from the hydrodynamic modelling described below are used as inputs.

Case Study, Data, and Hydrodynamic Modelling

To obtain the modelled flood grids, coarse and fine resolution twin hydrodynamic models are calibrated to high water marks obtained during the July 2021 flooding of the Ahr River in Germany. This was the most extreme flood event to hit the region in living memory, with precipitation exceeding a 500-year return period (Dietze et al., 2022), peak discharge estimated between a 2,600- and 58,700-year return period (Vorogushyn et al., 2022), and 134 casualties in the Ahr valley (Szönyi M. and Roezer V., 2022).

The twin hydrodynamic models were constructed in the LISFLOOD-FP 8.1 framework (LISFLOOD developers, 2020) and run on a Tesla K40m GPU using the second-order discontinuous Galerkin solver which implements the full shallow water equations (Shaw et al., 2021). The data used to construct and calibrate the models is summarized in Table B.1. The 0 5m resolution DEM was resampled to obtain the two target resolutions (4m and 32m) through averaging. A single Manning's roughness (n) parameter was applied to the domain and calibrated through trial-and-error to minimize the root-mean-square error (RMSE) of the maximum simulated depth against the high water marks. For the coarse ($s2 = 4$) and fine ($s1 = 32$) model, this yielded a value of $n = 0.035$ and $n = 0.025$ respectively with the RMSE values shown in Figure B.3.

Table B.1: Summary of data used for the 2021 flood

type	metadata	ref.
DEM	0.5 m resolution bare earth DEM created from aerial LiDAR survey from September 22 to October 24, 2021 in twelve sessions with a RIEGL scanner LMS-VQ780i with 20 points/m^2 achieved.	(Milan Geoservice GmbH, 2023)
High water marks	75 high water marks at buildings reported by residents.	(Apel et al., 2022)
Inflow hydrograph	30 hour hydrograph at Altenahr gauge with maximum depth of 10.2 m reconstructed by Environmental Office of the federal state Rhineland-Palatinate.	(Apel et al., 2022)
Building locations	Building footprint polygons downloaded from OSM on 2022-11-14.	(OpenStreetMap contributors, 2022)
Observed inundation	Polygon of maximum flood extents compiled from an aerial survey on July 16th and 20th and a second survey on July 24th and 29th.	(Landesamt für Umwelt Rheinland-Pfalz, 2022)

B.3.2 Results and Discussion

This section first treats the performance of the various methods against observations, before turning to the differences and similarities between coarse hydrodynamic modelling and aggregation.

Performance Against Observations

The performance of the various methods at reproducing depth values at a sub-set of buildings is evaluated by comparing the observed high water marks (HWM) described in Table B.1 with the values sampled at the HWM locations for the four grids (Figure B.3). In general, the hydrodynamic model results and the aggregated grids reproduce the HWMs fairly well; albeit with a slight negative bias which suggests imperfect calibration or some oversimplification of the model (e.g., absence of bridge blocking). The hydrodynamic model results slightly out-perform the aggregated grids on some metrics; however, the performance of the aggregated grids are similar.

The performance of the flood grids in reproducing the observed *inundation extents* is shown in Figure B.3.2 using common inundation performance metrics like the Critical Success Index (CSI). Figure B.3.2b.0 and b.1 show the aggregated grids have some skill at reproducing the flood, with CSI values similar to the coarse modelling. The twin hydrodynamic models (Figure B.3.2b.2 and b.3) reproduce the extents well, with both under- more than over-predicting (Error Bias = $FP/FN < 1$). For this case, the inundation from *WSH Averaging* actually out-performs both hydrodynamic models on some metrics ($CSI = 0.88$). This could be attributed to imperfect calibration (here we calibrated against HWMs, not inundation) or imperfect observations (Figure B.4 suggests the observed inundation is conservative); however, the difference in skill is marginal.

To illustrate the differences and similarities between the methods, Figure B.4 and B.5 show the WSE and WSH respectively for a complex region of the domain where flood waters encountered an underpass of an elevated highway. In this region, the aerial observation of the extents recorded the underpass as dry; however, this low-lying area was likely inundated — as shown by the fine-hydrodynamic model (Figure B.4.b1). South of the highway, the hydrodynamic and the aggregation methods generally agree

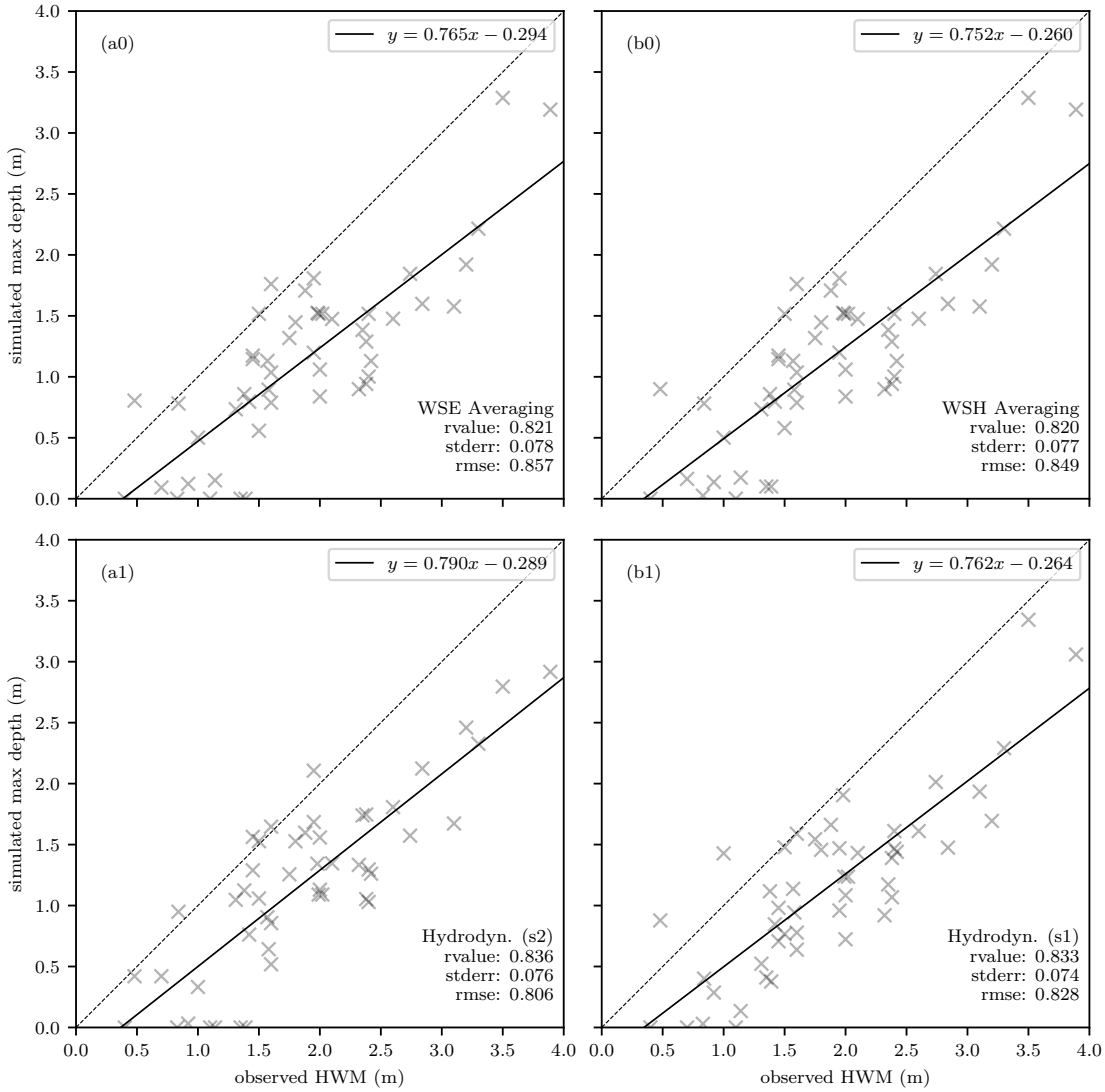
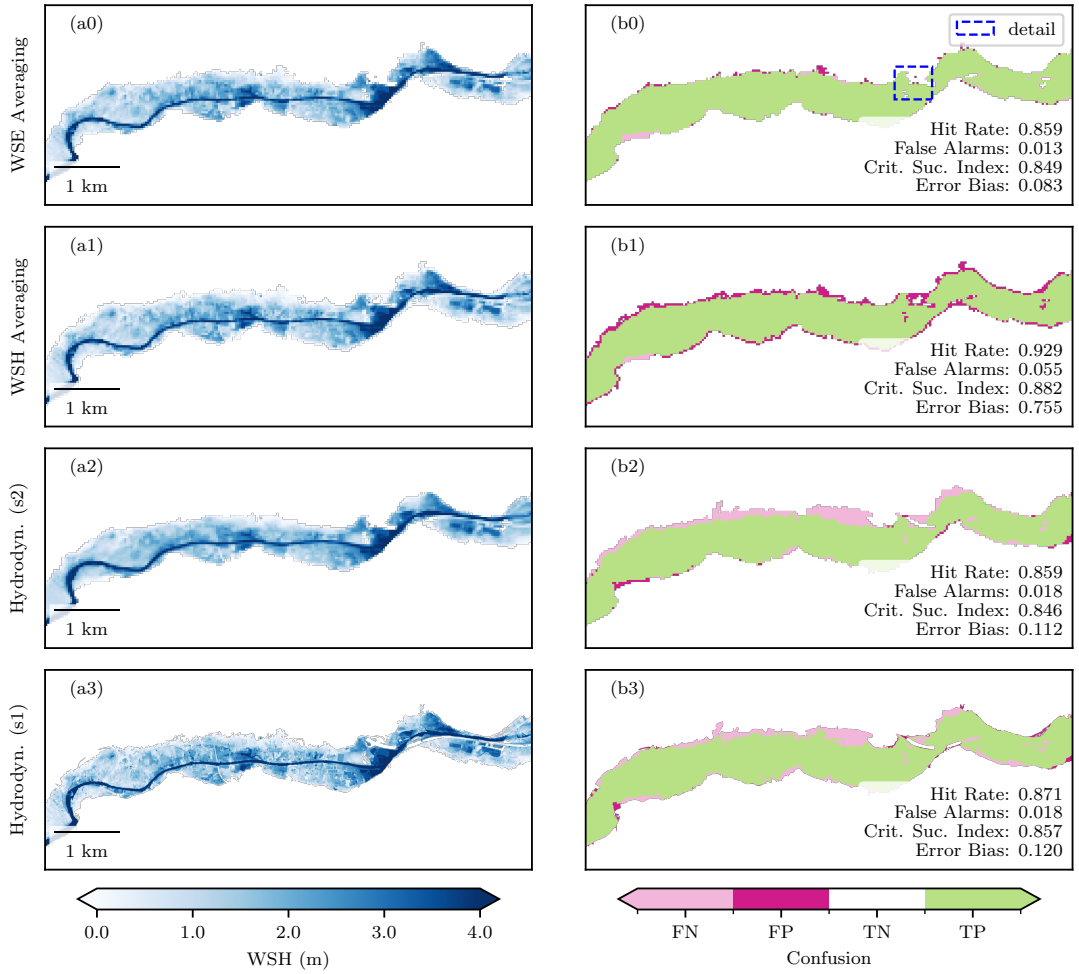


Figure B.3: WSH performance against the high water marks

and reproduce the extents well. North of the highway, where the coarse hydrodynamic model could not resolve any flow-paths, both aggregation methods replicate some of the inundation identified by the fine hydrodynamic model (see black arrow) — suggesting aggregation routines can be more accurate than equivalent coarse modelling in isolated areas. Comparing the aggregation methods within this region, the *WSH Averaging* routine shows substantially more inundation — this is expected as this routine propagates all volume in the fine-grid.

Aggregation and Modelling Comparison

Returning to the full domain, Figure B.6 compares inundation of the two aggregation routines to the coarse hydrodynamic result directly (rather than the observations). Similar to the comparison against the observations, this shows the aggregation routines have some agreement with the coarse result ($CSI = 0.95$ and $CSI = 0.88$). However, the aforementioned inundation over-estimation of the *WSH Averaging* routine



causes a clear departure from the coarse model results (Error Bias = 16.343). Figure B.7 shows the difference in depths between the aggregated grids, the fine model, and the coarse model. Differences are generally small ($< 0.5m$) and contrary to inundation similarity, depth values from the *WSH Averaging* routine (Figure B.7a1) are more similar to the coarse model than the *WSE Averaging* routine — highlighting the previous finding that some trade-off is required when aggregating. Regardless, Figure B.7a2 shows that these WSH differences are largely explained by differences between the coarse and fine hydrodynamic models. In other words, aggregating a fine model result does not capture the same differences in WSH as modelling explicitly on the coarse domain.

To compare the computational results of this case study to the findings of the analytical section of the manuscript, Table B.2 reports three grid-wide metrics computed on the full domain. For the two aggregation routines, all calculated metrics agree with those predicted by the methods described in the manuscript (Table 3.2). The coarse hydrodynamic model ($s2 = 32m$) on the other hand shows a reduction in area relative to the fine model (also shown in Figure B.3.2) but an increase in $\overline{WSH}$ and volume. This suggests the biases derived for aggregation routines (Table 3.2) do not map well onto twin hydrodynamic models calibrated on depths. It is likely that alternate calibration strategies (e.g., calibrating to WSE) would yield lesser or greater similarity to aggregation.

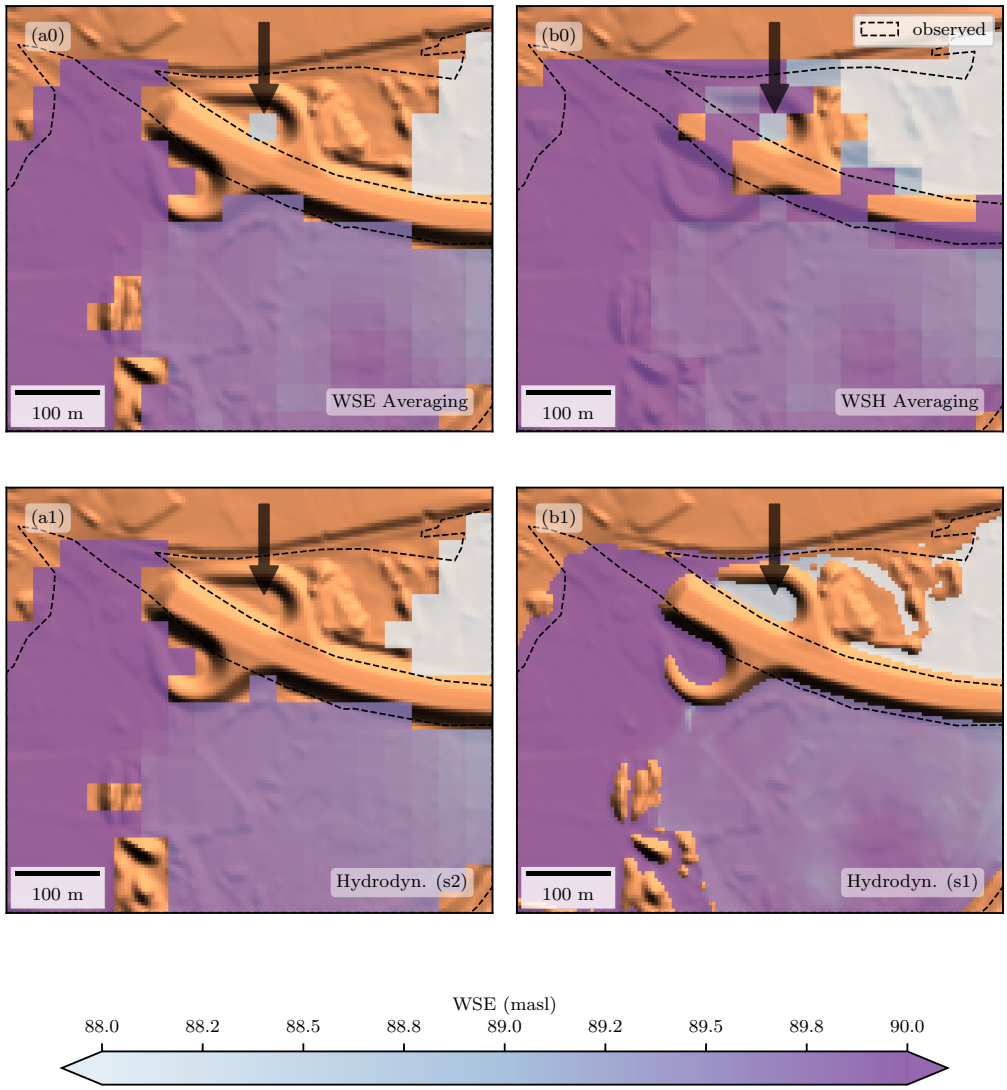


Figure B.4: Water Surface Elevation (WSE) detail. See Figure B.3.2 for location and text for description. Dashed line shows observed inundation from Table B.1.

B.3.3 Conclusions

This supplement compared flood grids generated with two aggregation, twin hydrodynamic models, and observations of both high water marks and extents from a 2021 flood. For both types of observations, all methods considered performed well. However, one aggregation routine performed less well at matching high water marks while the other over-predicted inundation — suggesting some trade-off in accuracy is required when aggregating grids. While the skill for both model results and aggregated results was similar, aggregation performed slightly better at reproducing the observed inundation than coarse modelling. This can be partly explained by the inability of the coarse model to resolve some sub-grid flow paths to isolated areas.

Comparing the aggregated grids directly to the coarse model result showed small ($< 0.5m$) spatially varied differences in depths, suggesting artifacts arrived at through coarse modelling may not be present in aggregation. Similarly, grid-wide bias metrics were computed and shown to agree with the analytical

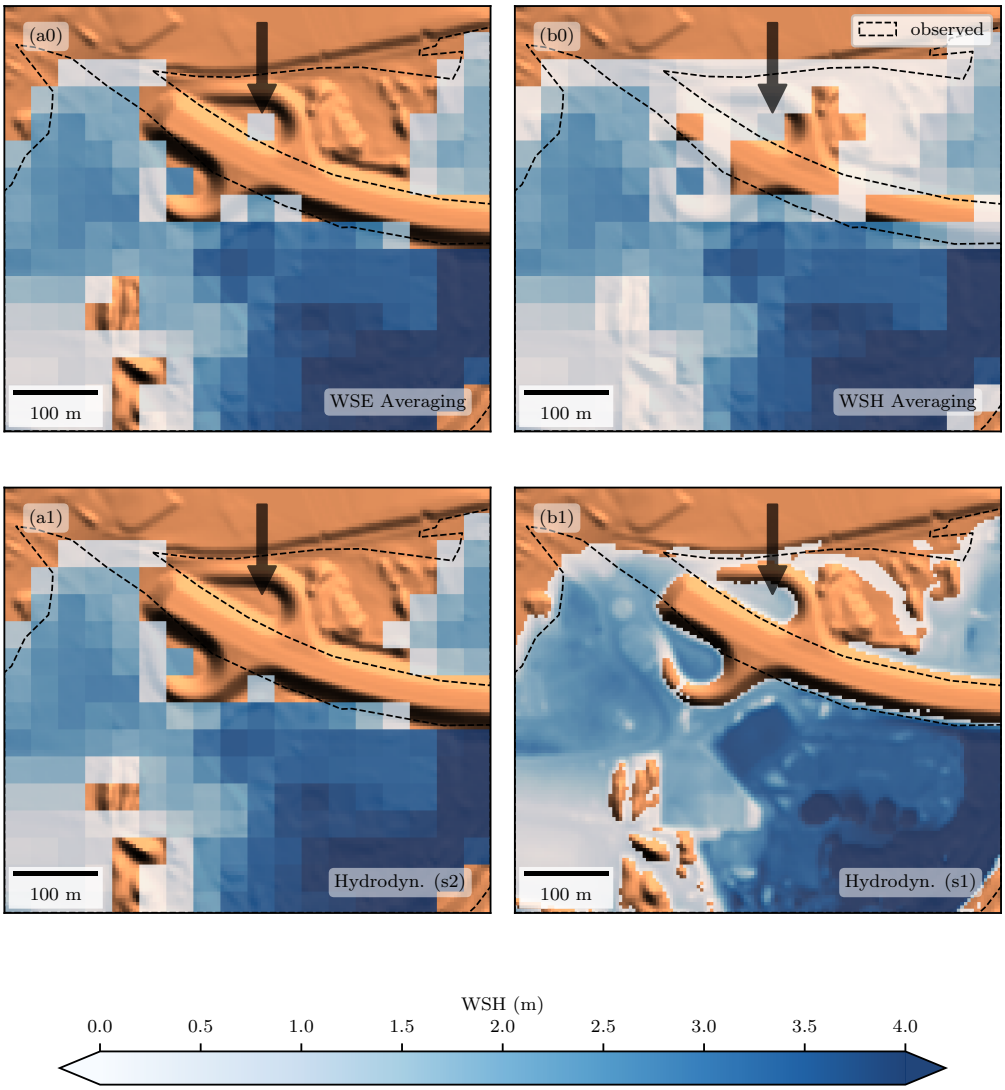


Figure B.5: Water Surface Height (WSH) detail. See Figure B.4 for description.

results of the manuscript for the two aggregation routines. However, the bias computed by comparing the coarse to the fine model differed in direction and magnitude from the aggregation routines — suggesting aggregation bias behaviour is a poor proxy for depth calibrated coarse modelling in this case. While the findings reported here are likely sensitive to the calibration strategy and the hydraulic character of the domain, we suspect the study of aggregation behaviour to be a poor analog for coarse modelling behaviour in general.

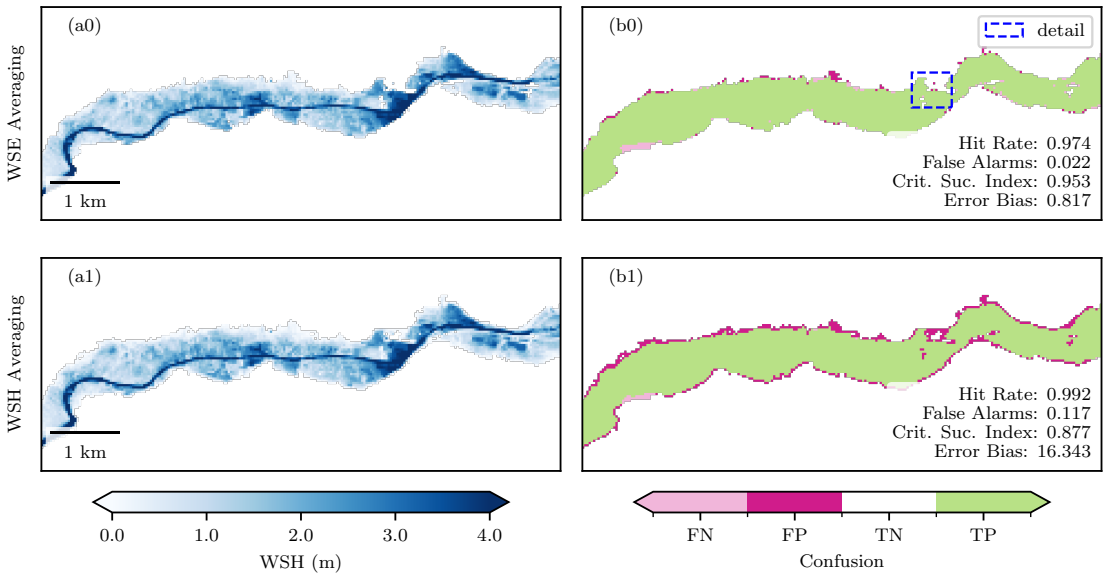


Figure B.6: Flood grids and corresponding performance against *coarse hydrodynamic* inundation (WSH_{s2}). See Figure B.3.2 for further description.

Table B.2: Grid-wide bias metrics for the two aggregation routines and the coarse hydrodynamic model ($s2 = 32m$) computed relative to the fine hydrodynamic model ($s1 = 4m$).

WSH Averaging	
$Bias_{global}[\overline{WSH}]$	0.000
$\frac{Bias[\sum A]}{\sum A_{s1}}$	0.108
$\frac{Bias[\sum V]}{\sum V_{s1}}$	0.000
WSE Averaging	
$Bias_{global}[\overline{WSH}]$	-0.006
$\frac{Bias[\sum A]}{\sum A_{s1}}$	-0.018
$\frac{Bias[\sum V]}{\sum V_{s1}}$	-0.018
Hydrodyn. (s2)	
$Bias_{global}[\overline{WSH}]$	0.007
$\frac{Bias[\sum A]}{\sum A_{s1}}$	-0.013
$\frac{Bias[\sum V]}{\sum V_{s1}}$	0.020

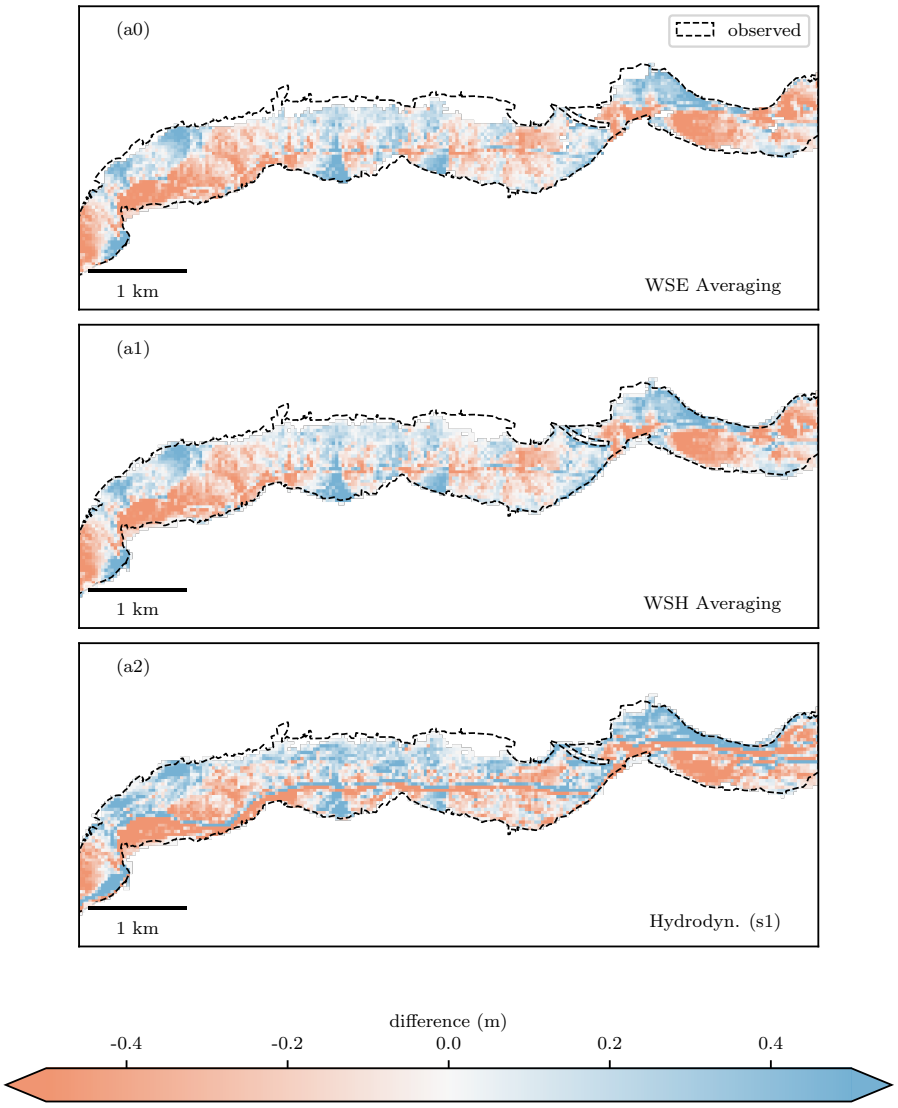


Figure B.7: Aggregated and fine hydrodynamic model (resampled) WSH_{s2} grid minus coarse hydrodynamic model difference grid.

Appendix C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

C.1 Function Curvature

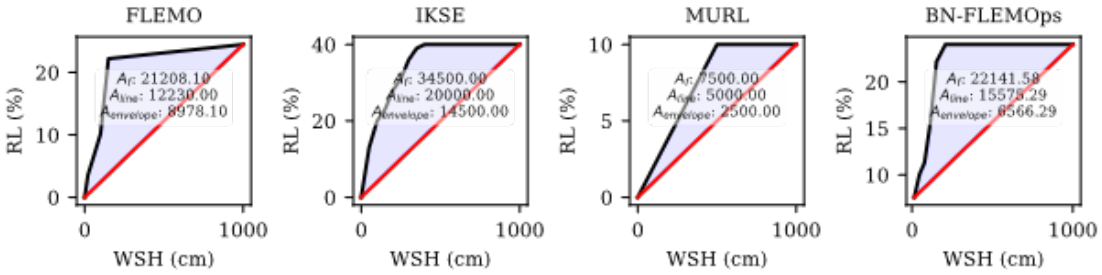


Figure C.1: Envelope of potential error for the four selected functions showing the area under the function (A_f), the area of the bounding line (A_{line}) and the area of the envelope ($A_{envelope}$=A_f-A_{line}).

C.2 Conceptual Functions

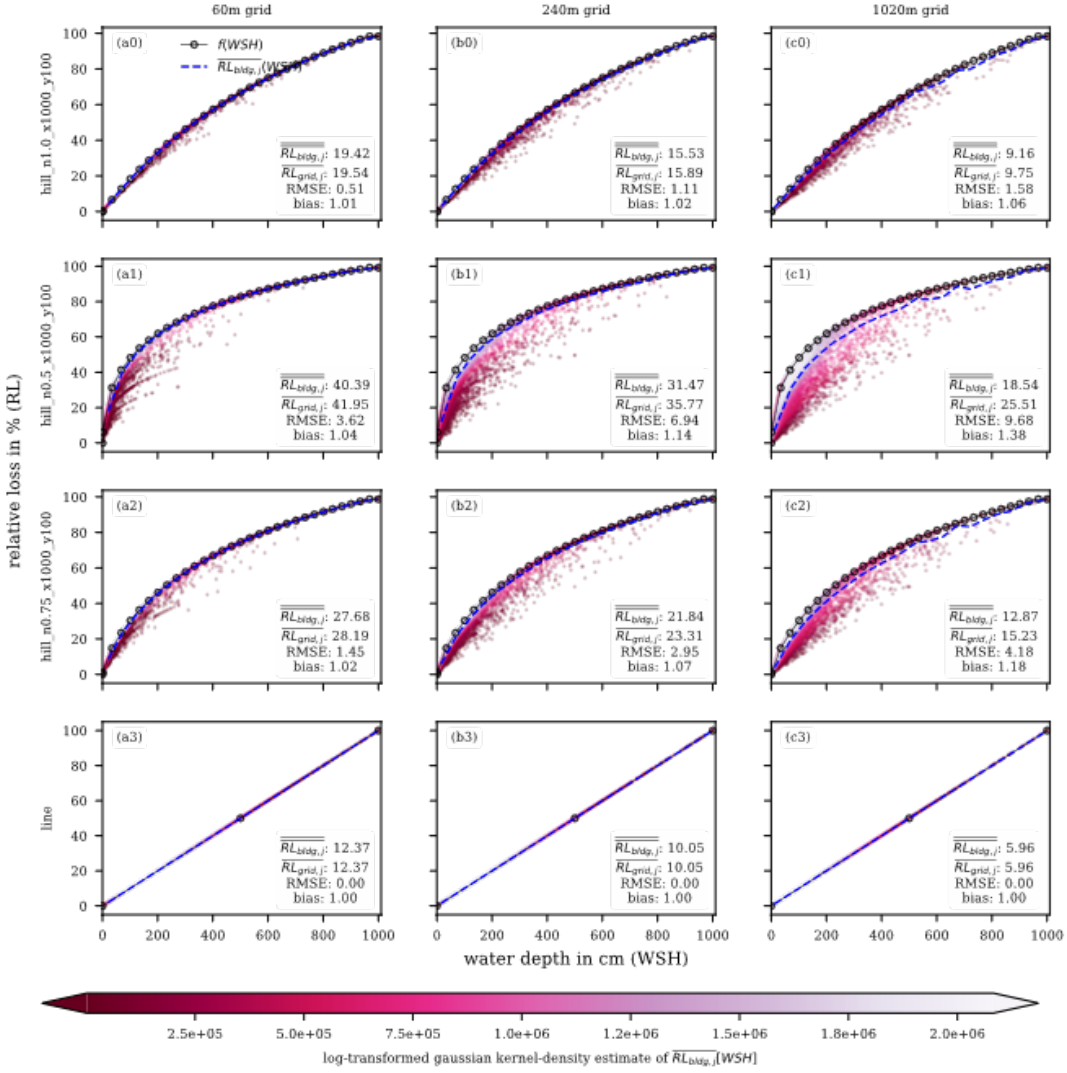


Figure C.2: Relative losses for three grid aggregations and four idealized functions as in Figure C.4.

C.3 Koblenz Sub-Domain

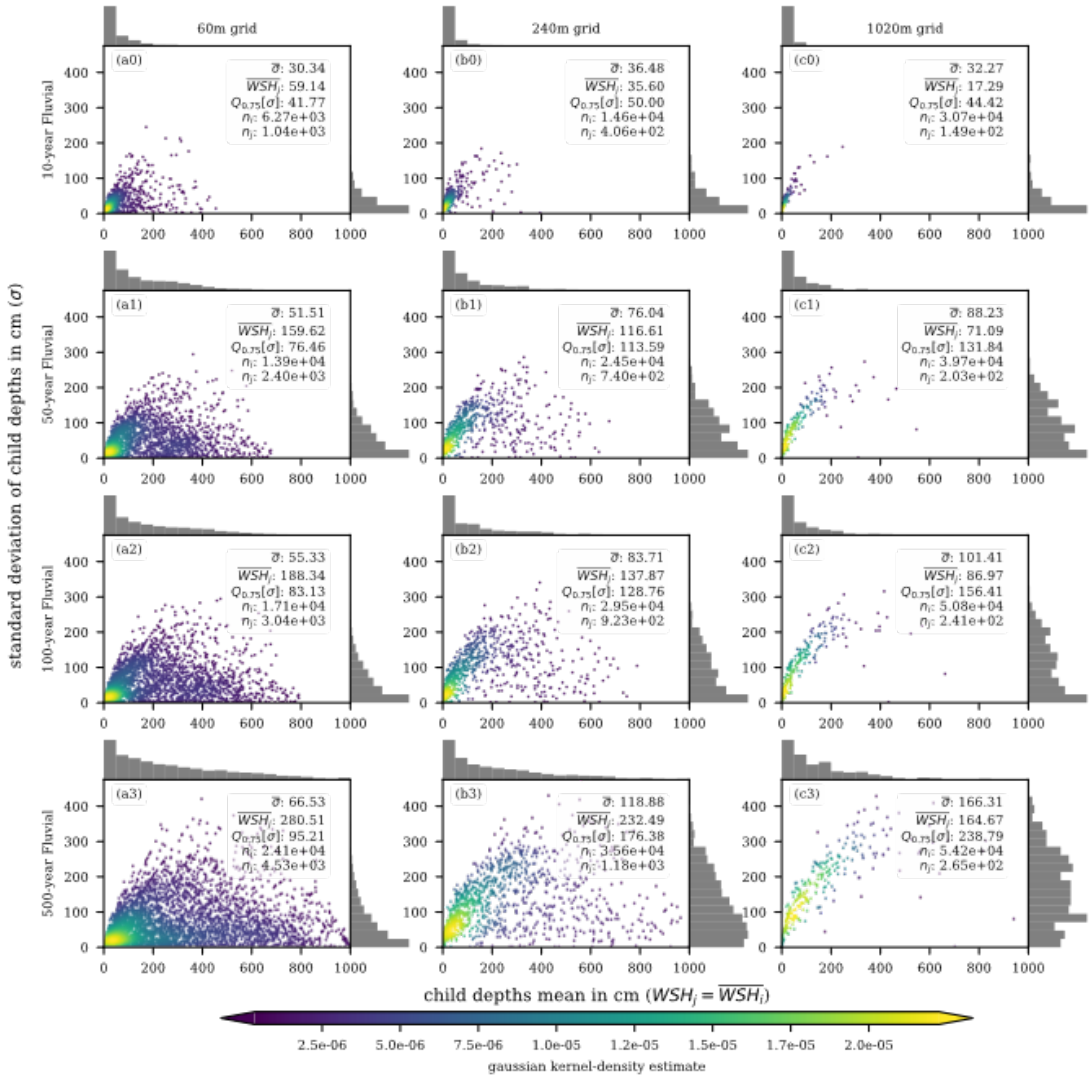


Figure C.3: Child depths mean and standard deviation for the Koblenz sub-domain as in Figure C.1.

C.4 Function Database

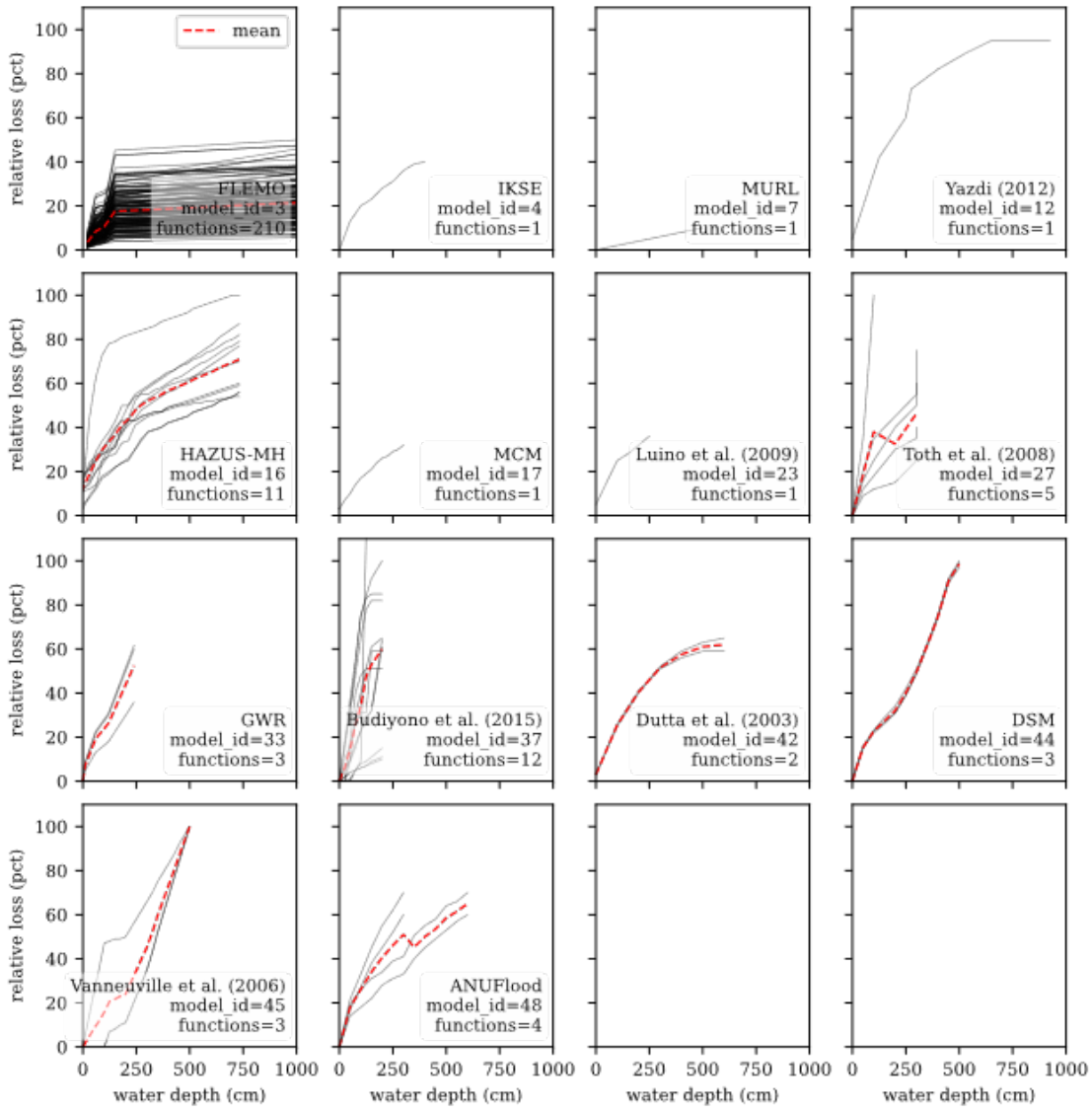


Figure C.4: All tabular, relative loss, uni-variate functions from Gerl et. al., (2016).