



Key Points:

- Through a novel framework, we show analytically that hazard grid aggregation leads to bias of key metrics independent of any study region
- This aggregation is shown to always positively bias inundation area when water depth grids are aggregated
- For example, aggregating from 1 to 512 m resolution resulted in a doubling of the inundated area for a 2018 flood in Canada

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Bryant,
seth.bryant@gfz-potsdam.de

Citation:

Bryant, S., Kreibich, H., & Merz, B. (2023). Bias in flood hazard grid aggregation. *Water Resources Research*, 59, e2023WR035100. <https://doi.org/10.1029/2023WR035100>

Received 12 APR 2023
Accepted 25 AUG 2023

Author Contributions:

Conceptualization: Seth Bryant, Heidi Kreibich, Bruno Merz
Data curation: Seth Bryant
Formal analysis: Seth Bryant
Funding acquisition: Heidi Kreibich, Bruno Merz
Investigation: Seth Bryant
Methodology: Seth Bryant
Software: Seth Bryant
Supervision: Heidi Kreibich, Bruno Merz
Validation: Seth Bryant
Visualization: Seth Bryant
Writing – original draft: Seth Bryant
Writing – review & editing: Seth Bryant, Heidi Kreibich, Bruno Merz

¹GFZ German Research Centre for Geosciences, Section 4.4. Hydrology, Potsdam, Germany, ²Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

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 modelers. 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.

1. Introduction

With the increase in flood related disaster damages and the availability of global data sets, 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 modeling; with each step adding uncertainty (de Moel & Aerts, 2011). Vulnerability modeling, 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 & Aerts, 2011; Jongman et al., 2012). These findings are supported by work comparing modeled 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 1,000 m (Hall et al., 2005; Sairam et al., 2021; Ward et al., 2015). 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 (s_1) or coarse (s_2) grid, measured in distance units (where $s_1 < s_2$), 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 & 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 (s_1) and coarse (s_2) 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 (s_2) grid is assembled from some sub-calculation (e.g., the mean statistic) applied to blocks or neighbors

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

of fine (s_1) 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 (s_2) 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 (s_1) and new coarse (s_2) grids with a non-zero mean, the aggregation routine can be said to impart bias or distortion onto the coarse (s_2) grid.

Flood hazards are increasingly modeled 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 (Alipour et al., 2022; Fewtrell et al., 2008; Papaioannou et al., 2016; Savage et al., 2016). 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; Ghimire & Sharma, 2021; Mohanty et al., 2020; Muthusamy et al., 2021; Saksena & Merwade, 2015; 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 & 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 behavior 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 modeling. 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 (Apel et al., 2009; Ghimire & Sharma, 2021; Jongman et al., 2012; Metin et al., 2018; Seifert, Thieken, et al., 2010), but few investigate sensitivity to resolution explicitly (Brussee et al., 2021; Komolafe et al., 2015; 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 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 (Muthusamy et al., 2021; Saksena & Merwade, 2015), 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 50 m 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, that is, overestimating the water depth of the coarser grids. No mention of the aggregation routine is provided or explanation for the behavior observed. Investigating the sensitivity of a flood mortality model to hydrodynamic model resolution, Brussee et al. (2021) compare a 5, 25, and 100 m 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 modeling platforms. Along with testing a 1D and 2D hazard model framework and input data qualities, they investigated alternate DEM constructions with a LiDAR-derived 3 m and two publicly available DEMs at 10 and 30 m resolution. They found the 1D model to be more sensitive

Table 1
Summary of Selected Studies With Paired Grid-Based Models at Fine and Coarse Resolution

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 100 m [...]	1.16
Sieg et al. (2019) and Seifert, Thieken, et al. (2010)	Hazard: random sample of water mask values. 10 m Exposure: 165 businesses (object-scale, aspatial and stochastic) Vulnerability: Random Forest empirical commercial damages	Hazard: interpolation of highwater marks. 25 m Exposure: disaggregated average municipal asset values. 25 m Vulnerability: multi-variate empirical commercial damage	5.68
Sieg et al. (2019) and Seifert, Thieken, et al. (2010)	Hazard: random sample of water mask values. 10 m Exposure: 15 businesses (object-scale, aspatial and stochastic) Vulnerability: Random Forest empirical commercial damages	Hazard: 1D/2D hydrodynamic LISFLOOD-FP. 25 m Exposure: disaggregated average municipal asset values. 25 m Vulnerability: multi-variate empirical commercial damage	8.88
Komolafe et al. (2015)	Hazard: 1D/2D hydrodynamic. 50 m Exposure: remote sensing derived land-use grid. 30 m Vulnerability: multi-variable synthetic direct building damages	[...] 1,000 m upscale (unspecified method) [...] [...]	1.05
Brussee et al. (2021)	Hazard: 2D flexible mesh hydrodynamic. 5 m Exposure: disaggregated neighborhood-scale Vulnerability: multi-variable mortality function	[...] 100 m [...] [...]	1.08
Ghimire and Sharma (2021)	Hazard: 2D hydrodynamic. LiDAR derived 3 m Exposure: buildings (object-scale) Vulnerability: depth-damage curves	[...] unspecified 30 m [...] [...]	1.33
Pollack et al. (2022)	Hazard: 2D hydrodynamic. 30 m Exposure: buildings (object-scale) Vulnerability: uni-variate synthetic	[...] aggregated to census block-group (order 1–100 km) [...]	4.67

Note. 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.

to the different DEMs than the 2D model, with a 25% and 75% increase in damages respectively at 30 m 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 30 m resolution flood depth grids with return periods ranging from 2- to 500-year. 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 data set 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 modeling 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 $\frac{2}{s_1}$ 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 aggregated models used in practice.

Spatial resolution transfers within flood risk models are often implicit and can occur as part of data preparation, modeling, 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 (Bates et al., 2021; Sampson et al., 2015), 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 (Bates et al., 2021; Hall et al., 2005; Sairam et al., 2021). However, some studies aggregate hazard model outputs to facilitate intersection with more coarse exposure data, either through simple averaging (Seifert, Kreibich, et al., 2010; Sieg & Thielen, 2022) or some unspecified method (Jongman et al., 2012; Thielen et al., 2016). Hazard model outputs have also been aggregated to facilitate comparison with more coarse climate re-analysis data (Paprotny et al., 2020). 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 1 through generalizable 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 modeling. To explore scaling effects, we compare fine grids to their coarse analogs using metrics of interest to flood risk modelers. Rather than construct these coarse analogs through hydrodynamic modeling 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.

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:

$$WSE = DEM + WSH \quad (1)$$

Combining Equation 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:

$$WSH \geq 0 \quad \text{and} \quad WSE > DEM \quad (2)$$

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 1 emerges an important distinction for the handling of dry cells:

$$WSH_{i \text{ or } j} = 0 \iff WSE_{i \text{ or } j} = \text{null} \iff \text{“dry”} \quad (3)$$

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 Equations 1 and 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 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. This can be expressed mathematically as:

$$DEM_{s2,j} = \overline{DEM_{s1,i}} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} DEM_{s1,i} \quad (4)$$

$$\overline{WSH_{s1,i}} = \frac{1}{N_{12}} \sum_{i=1}^{N_{12}} WSH_{s1,i} \quad (5)$$

$$\overline{WSE_{s1,i}} = \frac{1}{N_{wet}} \sum_{i=1}^{N_{wet}} WSE_{s1,i} \quad (6)$$

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

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.

3.1. 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 Equations 1 and 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 4 to obtain DEM_{s2} , as this is not affected by the dry cells in Equation 3. Further, both rely on Equation 1 to compute the secondary grid—rather than averaging which would yield a grid set in violation of Equation 1 (this can be seen by comparing the *WSH* grids in Figures 1d and 1e). Figures 1d and 1e provide a graphical summary and toy example of these routines, which are defined mathematically in Text S1 in Supporting Information S1. 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.

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 behavior. We define each of

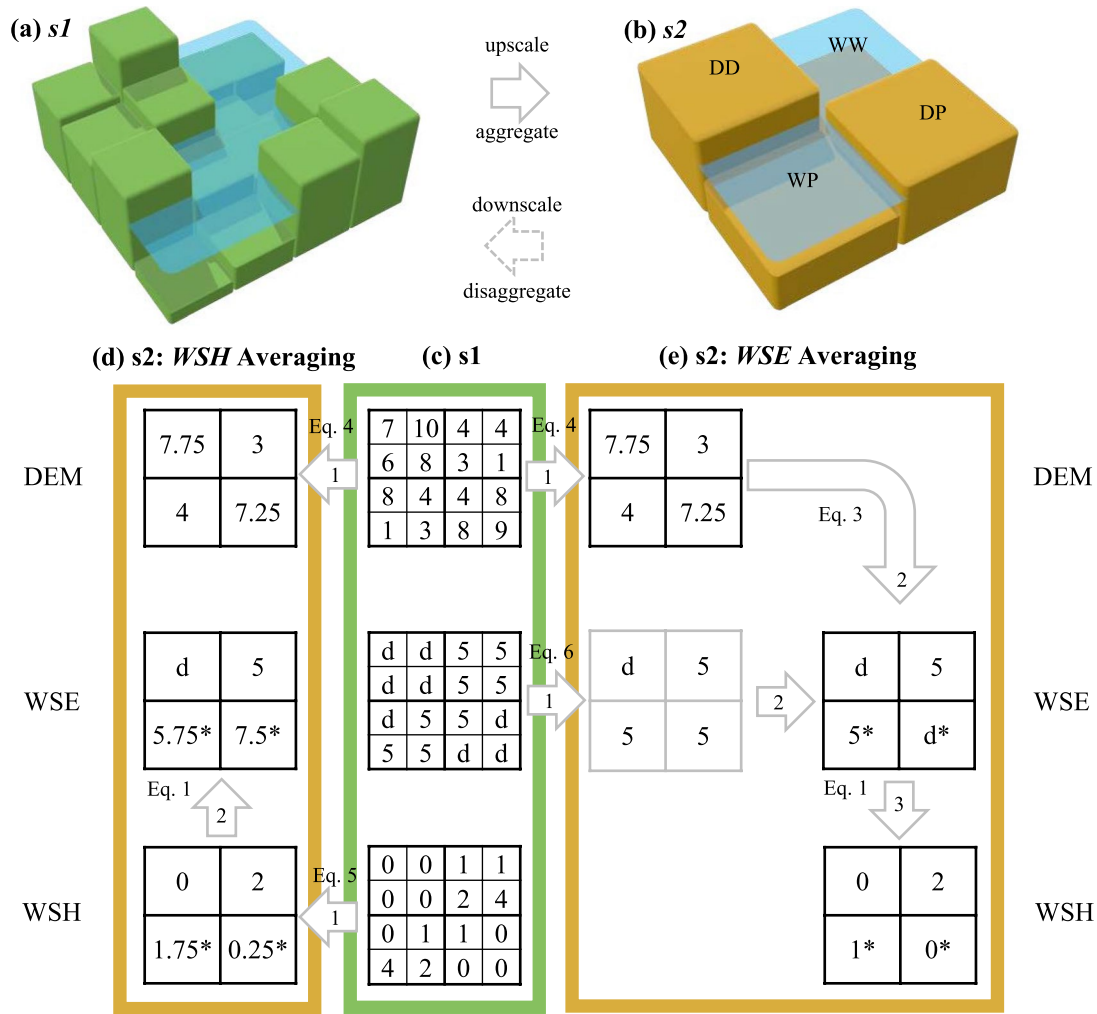


Figure 1. Flood hazard data scaling issues and two aggregation schemes demonstrated with a toy example. Panel (a) is an oblique view of a fine ($s1$) DEM and WSE while panel (b) shows an aggregated coarse ($s2$) analog and corresponding resample case (DD, WW, WP, DP) from Figure 2. Panel (c) shows an example set of $s1$ values for the three grids described by Equation 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 gray grids show intermediate calculations. Discrepancies between resulting $s2$ grids from the two routines are marked with “*.”

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 2 and defined explicitly as:

$$case_j = \begin{cases} DD & \text{if } \max(WSH_{s1,i}) = 0 \\ DP & \text{if not } DD \text{ and } \overline{DEM_{s1,i}} \geq \overline{WSE_{s1,i}} \\ WP & \text{if not } WW \text{ and } \overline{DEM_{s1,i}} < \overline{WSE_{s1,i}} \\ WW & \text{if } \min(WSH_{s1,i}) > 0 \end{cases} \quad (7)$$

where $\overline{WSE_{s1,i}}$ is defined in Equation 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 2b. The quadrants in Figure 1a provide a simple example of four such i groups whose corresponding case labels are shown on Figure 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

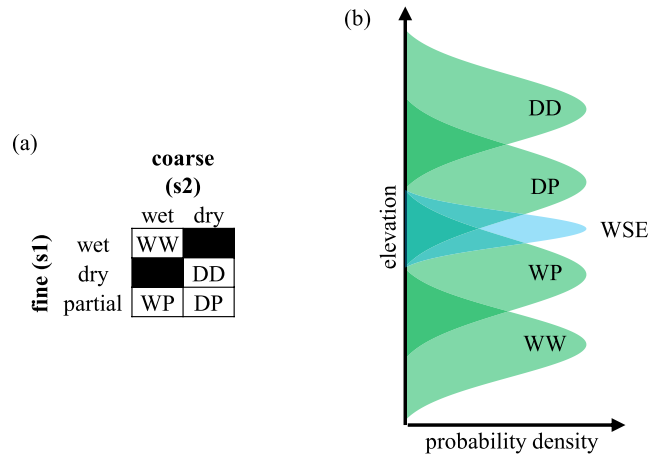


Figure 2. Framework for classification of flood hazard resample case. Panel (a) shows class label acronyms. Panel (b) provides a conceptual diagram showing a hypothetical distribution of WSE_{s1} and four possible DEM_{s1} groups and their resulting resample case. D, W, and P stand for “dry,” “wet,” and “partial” respectively.

discernment of aggregation artifacts (see results below for the *WSE Averaging* routine). Figure 3 shows a fully classified domain where WSH_{s1} has been simulated using a hydrodynamic model built on a 1 m 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.

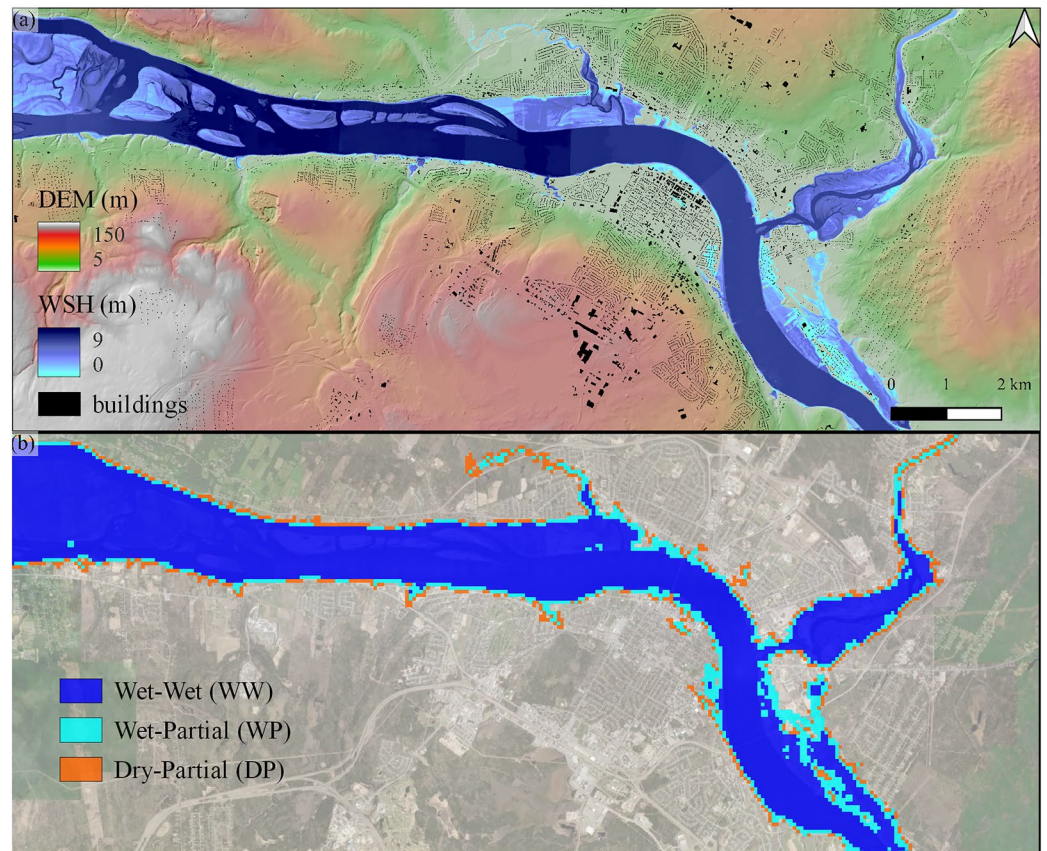


Figure 3. Simulated May 2018 Saint John River flood in Canada. Panel (a) shows DEM_{s1} and WSH_{s1} at 1 m resolution and building footprints from Microsoft (2019). Panel (b) shows corresponding resample case (see Figure 2) for a 1:64 aggregation (*DD* is transparent for clarity).

Such a resample case map is independent of any coarse (s_2) aggregate grids or the routine used to compute them as it is purely a function of the fine source (s_1) grids and the target coarse resolution (i.e., N_{12}). However, by treating each region independently, certain behaviors 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 s_2 grids generated by different aggregation routines may differ in these regions.

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. 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 behavior of an aggregation routine as a function of target resolution (s_2), we define *true* values as those represented in the fine (s_1) 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 (s_1) and coarse (s_2) grids. From this, an important distinction is made between *non-systematic* errors—differences in s_1 and the corresponding s_2 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 Equations 5 and 6, but evaluated against all cells in a region of interest (rather than local groups). For example, $\overline{WSH}_{s_1, WW}$ is the sum of all WSH_{s_1} cells classified as resample case *WW* per Equation 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 s_2 cell versus s_1 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 s_1 and s_2 grids (see Equation 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).

3.4. 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 1 m resolution grids before applying the *resample case framework* to quantify bias arising from the aggregation. The DEM_{s_1} grid was downloaded from GeoNB who constructed the

Table 2
Biases in Two Aggregation Routines Evaluated Analytically for Each Resample Case

Metric	Equation	Resample case			
		DD	DP	WP	WW
WSH Averaging					
$Bias_{global} [WSH]$	$\overline{WSH}_{s2} - \overline{WSH}_{s1}$	0	0	0	0
$Bias_{local} [WSH]$	$\overline{WSH}_{s2} - \overline{WSH}_{s1}$	0	-	-	0
$Bias_{global} [WSE]$	$\overline{WSE}_{s2} - \overline{WSE}_{s1}$	n/a	+	+	0
$Bias_{local} [WSE]$	$\overline{WSE}_{s2} - \overline{WSE}_{s1}$	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					
$Bias_{global} [WSH]$	$\overline{WSH}_{s2} - \overline{WSH}_{s1}$	0	-	-	0
$Bias_{local} [WSH]$	$\overline{WSH}_{s2} - \overline{WSH}_{s1}$	0	n/a	-	0
$Bias_{global} [WSE]$	$\overline{WSE}_{s2} - \overline{WSE}_{s1}$	n/a	n/a	0	0
$Bias_{local} [WSE]$	$\overline{WSE}_{s2} - \overline{WSE}_{s1}$	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

Note. 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 $Bias_{local} [WSH]$ 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.

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. (2022) 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 1 yielding the grids shown in Figure 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 3b) are computed for the WSE Averaging and WSH Averaging routines for a total of 40 grids (4 grid types \times 5 coarse resolutions \times 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.

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 3a). From the centroids of footprints in this layer, each of the aforementioned 40 retrograde grids is sampled to produce a parallel data set from which the same metrics of interest can be computed for the exposed domain.

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 Text S1 in Supporting Information S1 while the results are summarized in Table 2.

Focusing on the non-partial regions (DD and WW), Table 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 behavior deviates from that of continuous grids owing to the presence of dry cells and the inter-grid relations (see Equations 3 and 1).

Examining the bias in these partial regions (WP and DP), Table 2 shows some bias for all metrics except the respective primary grids on the global metric (i.e., WSE Averaging has no $Bias_{global} [WSE]$ bias and WSH Averaging has no $Bias_{global} [WSH]$ 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 Equations 1 and 3.

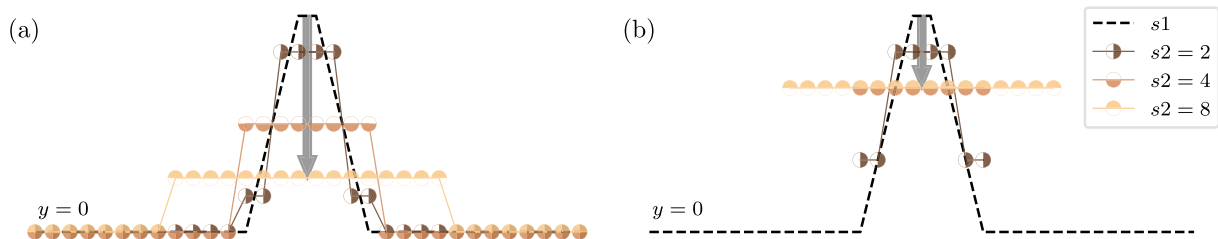


Figure 4. Conceptual diagram showing a cross-section of local bias for three target resolutions generated through two types of averaging: (a) zero-inclusion (as in Equation 5) and (b) zero-exclusion (as in Equation 6). All series within a panel have the same global mean. Black arrow shows the progression of local bias.

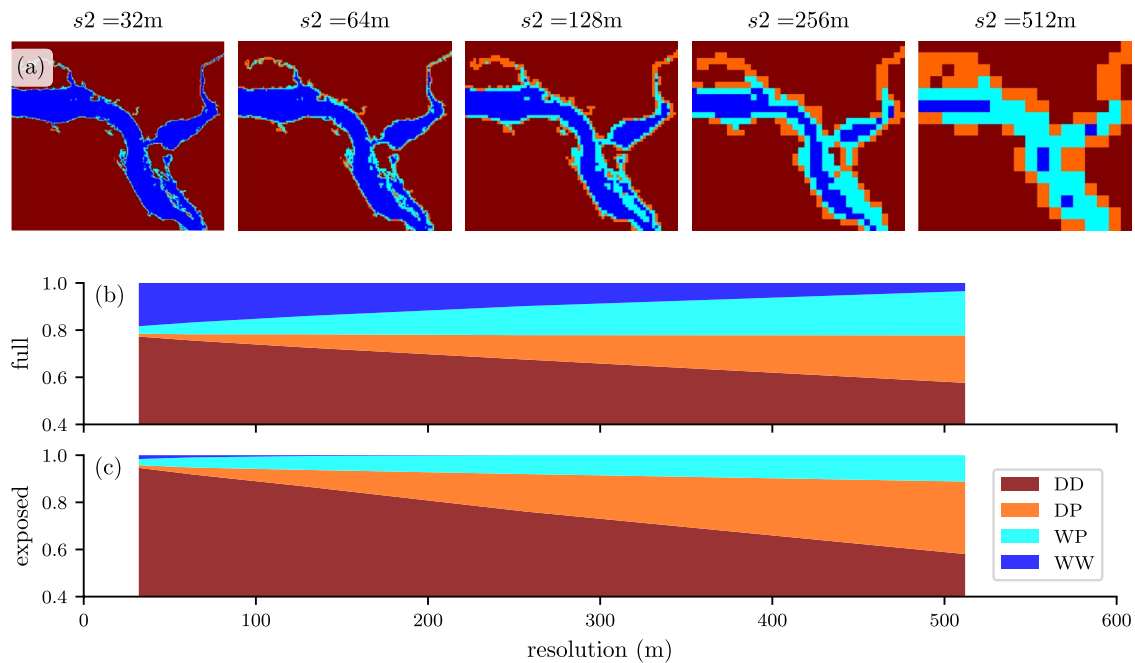


Figure 5. Resample case classification progression for May 2018 Saint John River flood hazard data showing (a) illustrative maps at five target resolutions (s_2); (b) full domain fraction; and (c) exposed domain (i.e., values sampled at buildings) fraction for each case. See Figure 2 for description of legend.

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 4a, where we see the aggregated values have a progressively lower local value (measured at the center), while the global average remains constant. In other words, given a wet s_1 cell with some dry neighbors, 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 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 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 2 shows $Bias[\sum V]$ follows the same behavior as $Bias_{global}[\overline{WSH}]$ (see Text S1 in Supporting Information S1 for derivation), meaning *WSH Averaging* also preserves $\sum V$. This suggests a paradox for hydrodynamic modelers: 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 behavior 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.

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; 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 2.

5.1. Full Domain

Figure 5 shows the resulting change in composition or classification of the domain, computed from the classification map obtained at each s_2 scale. This shows that the portion of partial regions (WP and DP), we call ρ_{s_2/s_1} , increases with aggregation: ranging from $\rho_{32} = 5\%$ to $\rho_{512} = 40\%$ for our case (Figure 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 ρ_{s_2/s_1} is positively correlated with aggregation bias sensitivity or magnitude. In other words, the portion of the domain containing errors of aggregation, ρ_{s_2/s_1} , increases with coarser resolutions or higher s_2/s_1 values. Extending these observations to other study regions, we hypothesize that ρ_{s_2/s_1} is also sensitive to hydraulic regime: with broad-flat floodplains having a higher inundation-area-to-perimeter ratio, and therefore lower ρ_{s_2/s_1} values with less sensitivity to aggregation bias; while narrow-steep floodplains would have higher ρ_{s_2/s_1} 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 5c and discussed below.

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 s_2 grids to the original 1 m resolution s_1 grids. These calculations are performed on the full domain and each resample case as independent regions of interest to develop five magnitude versus resolution series for each metric and routine. Results for four of the key metrics are shown in Figures 6a and 6b while the remaining two metrics are provided in Figure S1.

Comparing Figure 6 and Figure S1 in Supporting Information S1 to Table 2 shows all computations agree with the directional bias derived analytically in the previous section. For the *WSH Averaging* routine, Figure 6a suggests the bias in the DP case is always more severe than the WP case. This is also shown analytically in Text S1 in Supporting Information S1 for certain conditions (e.g., $N_{wet,DP} < N_{dry,WP}$). However, while the conditions favoring 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}_{s_2} (Table 2, Figures 6a0 and 6b0); 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 (s_1) 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 modeling would need to re-calibrate each hydrodynamic model to some observations; an example of which is provided in Text S3 in Supporting Information S1 which shows a stable \overline{WSH}_{s_2} .

For all partial zones, *WSH Averaging* shows a doubling (100% increase) of the inundated area (A) for the $s_2 = 512$ m grids in this case study (Figure 6a2). *WSE Averaging* showed less bias, with the WP and DP global bias nearly balancing, leading to a meager 10% increase for $s_2 = 512$ m (Figure 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; Ghimire &

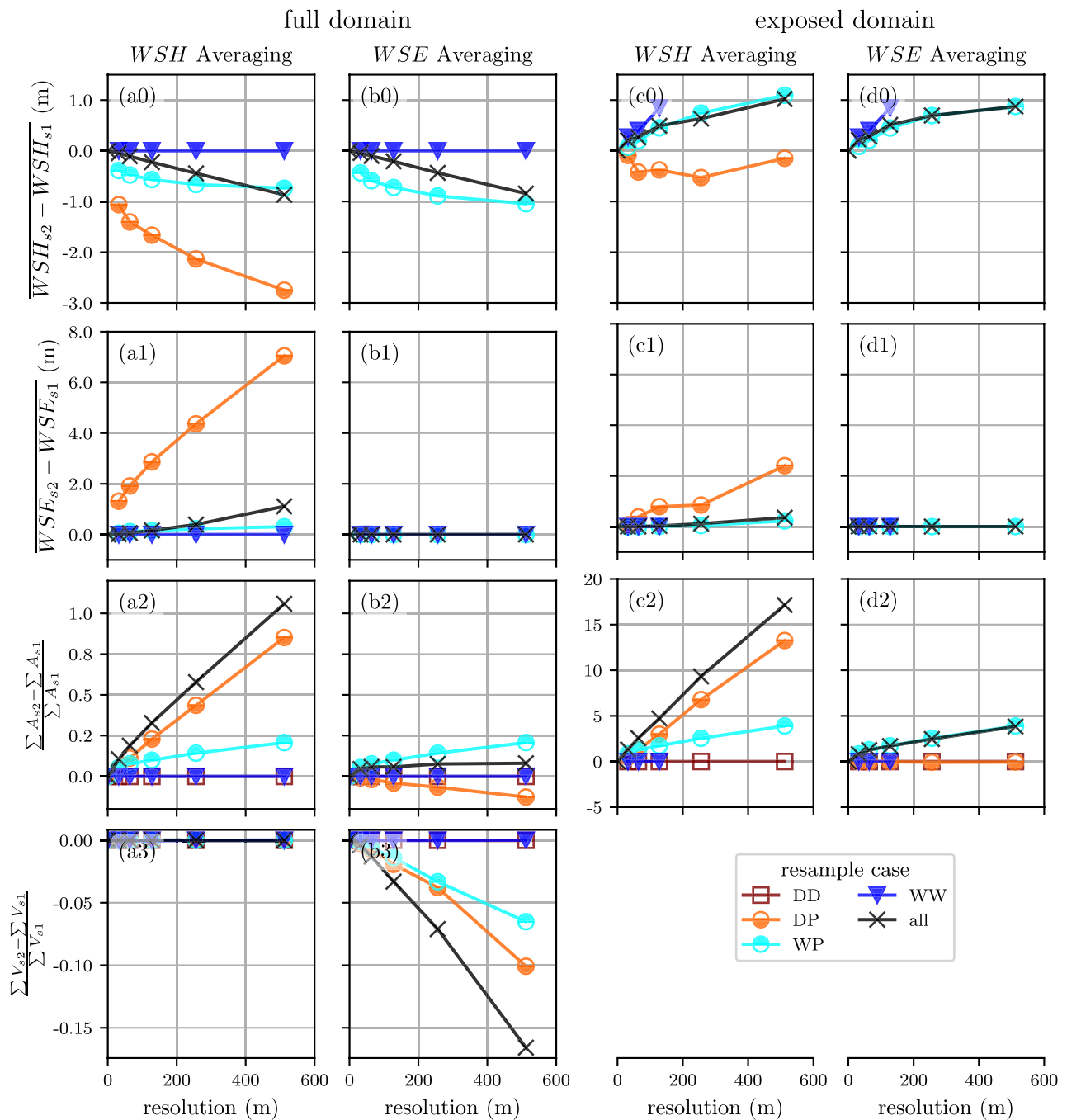


Figure 6. Bias from aggregation of four metrics for two routines sub-sampled by resample case (see Figure 2) for the full domain and the exposed domain. The “all” series uses the complete region of interest, without sub-setting by resample case. Panels (a2) and (b2) show the non-dry or inundated area of the full domain while panels (c2) and (d2) show the count of non-dry or exposed buildings.

Sharma, 2021; Mohanty et al., 2020; Muthusamy et al., 2021; Saksena & Merwade, 2015; Xafoulis et al., 2023). To evaluate the transferability between the aggregation routines considered here and hydrodynamic modeling (i.e., whether aggregation behavior might serve as an analog for coarse hydrodynamic modeling behavior) Text S3 in Supporting Information S1 presents a brief comparison of the aggregation routines against a depth calibrated hydrodynamic model built on a DEM resampled to the coarse resolution ($s_2 = 32$ m) 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 ($s_2 = 32$ m), 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

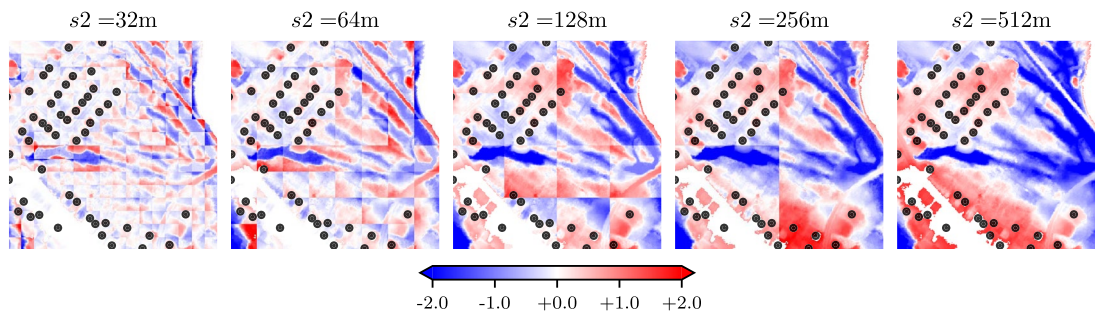


Figure 7. *WSH* difference maps for an example 512 m square region at five resolutions aggregated with the *WSH Averaging* routine showing building centroid locations in black. To compute local errors, WSH_{s_2} grids are downsampled to s_1 then WSH_{s_1} is subtracted, yielding the $WSH_{s_2} - WSH_{s_1}$ values shown in meters on a red-blue color scale.

that some artifacts arrived at through coarse modeling 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.

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 modelers: developed areas or the exposed domain. Figure 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 3b; and second, that buildings are less prevalent within the river channel. Further, Figure 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 512 m, 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 Figures 6c, 6d and Figure S2 in Supporting Information S1. Comparing the elements in Figure 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 $s_2 = 512$ m 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 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 30 m 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 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 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 s_1 cells in a group, while the exposed domain sampling (and therefore the metrics) ignore those cells without exposure. Figure 7 shows a clear

example where each tile has the same \overline{WSH}_{s_2} on the full domain, but within s_2 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 30 m resolution and they aggregate assets to counties which can be on the order of 1–100 km) and on different elements of risk modeling (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 (5–50 m), 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.

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 1–512 m. 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 1 to 512 m) 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 modeling. 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/cefec/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 1) and a growing body of work on hydrodynamic models (Banks et al., 2015; Ghimire & Sharma, 2021; Mohanty et al., 2020; Muthusamy et al., 2021; Saksena & Merwade, 2015). 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 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.

Data Availability Statement

The python scripts used to construct the aggregated grids, sample the grids at building locations, compute the metrics, and generate the plots is provided in Bryant (2023b). An easy-to-use QGIS script with similar tools for aggregating flood hazard grids is provided in Bryant (2023a). The DEM_{s1} grid used in the computation approach is provided by Government of New Brunswick (2016) and the Saint John River 2018 maximum WSH_{s1} data is provided by GeoNB (2019). Building locations are provided by Microsoft (2019).

Acknowledgments

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.

References

- Alipour, A., Jafarzadegan, K., & Moradkhani, H. (2022). Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping. *Environmental Modelling & Software*, 152, 105398. <https://doi.org/10.1016/j.envsoft.2022.105398>
- Apel, H., Aronica, G. T., Kreibich, H., & Thieken, A. H. (2009). Flood risk analyses—How detailed do we need to be? *Natural Hazards*, 49(1), 79–98. <https://doi.org/10.1007/s11069-008-9277-8>
- Banks, J. C., Camp, J. V., & Abkowitz, M. D. (2015). Scale and resolution considerations in the application of HAZUS-MH 2.1 to flood risk assessments. *Natural Hazards Review*, 16(3), 04014025. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000160](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000160)
- Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, 57(2). <https://doi.org/10.1029/2020WR028673>
- Bellos, V., & Tsakiris, G. (2015). Comparing various methods of building representation for 2D flood modelling in built-up areas. *Water Resources Management*, 29(2), 379–397. <https://doi.org/10.1007/s11269-014-0702-3>
- Bierkens, M., Finke, P., & De Willigen, P. (2000). *Upscaling and downscaling methods for environmental research*. Kluwer Academic.
- Brussee, A. R., Bricker, J. D., De Bruijn, K. M., Verhoeven, G. F., Winsemius, H. C., & Jonkman, S. N. (2021). 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(3), e12713. <https://doi.org/10.1111/jfr3.12713>
- Bryant, S. (2023a). Flood grid rescaler. [Software]. Zenodo. <https://doi.org/10.5281/zenodo.8271965>
- Bryant, S. (2023b). Flood hazard grid aggregation scripts. [Software]. Zenodo. <https://doi.org/10.5281/zenodo.8271996>
- Bryant, S., McGrath, H., & Boudreault, M. (2022). Gridded flood depth estimates from satellite-derived inundations. *Natural Hazards and Earth System Sciences*, 22(4), 1437–1450. <https://doi.org/10.5194/nhess-22-1437-2022>
- Degbelo, A., & Kuhn, W. (2018). Spatial and temporal resolution of geographic information: An observation-based theory. *Open Geospatial Data, Software and Standards*, 3(1), 12. <https://doi.org/10.1186/s40965-018-0053-8>
- de Moel, H., & Aerts, J. C. J. H. (2011). Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. *Natural Hazards*, 58(1), 407–425. <https://doi.org/10.1007/s11069-010-9675-6>
- Dimitriadis, P., Tegos, A., Oikonomou, A., Pagana, V., Koukouvinos, A., Mamassis, N., et al. (2016). Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping. *Journal of Hydrology*, 534, 478–492. <https://doi.org/10.1016/j.jhydrol.2016.01.020>
- Fewtrell, T. J., Bates, P. D., Horritt, M., & Hunter, N. M. (2008). Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes*, 22(26), 5107–5118. <https://doi.org/10.1002/hyp.7148>
- GeoNB. (2019). Flood risk areas and historical floods. [Dataset]. Retrieved from <http://www.snb.ca/geonb1/e/DC/floodraahf.asp>
- Ghimire, E., & Sharma, S. (2021). Flood damage assessment in HAZUS using various resolution of data and one-dimensional and two-dimensional HEC-RAS depth grids. *Natural Hazards Review*, 22(1), 04020054. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000430](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000430)
- Government of New Brunswick. (2016). ERD 2015 lidar. [Dataset]. Retrieved from <http://geonb.snb.ca/li/index.html>
- Hall, J. W., Sayers, P. B., & Dawson, R. J. (2005). National-scale assessment of current and future flood risk in England and Wales. *Natural Hazards*, 36(1–2), 147–164. <https://doi.org/10.1007/s11069-004-4546-7>
- Jongman, B., Kreibich, H., Apel, H., Barredo, J., Bates, P., Feyen, L., et al. (2012). Comparative flood damage model assessment: Towards a European approach. *Natural Hazards and Earth System Sciences*, 12(12), 3733–3752. <https://doi.org/10.5194/nhess-12-3733-2012>
- Komolafe, A., Herath, S., & Avtar, R. (2015). 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>
- McGrath, H., Stefanakis, E., & Nastev, M. (2015). 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.ijdr.2015.09.003>
- Merz, B., Kreibich, H., Thieken, A., & Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Sciences*, 4(1), 153–163. <https://doi.org/10.5194/nhess-4-153-2004>
- Metin, A. D., Dung, N. V., Schröter, K., Guse, B., Apel, H., Kreibich, H., et al. (2018). How do changes along the risk chain affect flood risk? *Natural Hazards and Earth System Sciences*, 18(11), 3089–3108. <https://doi.org/10.5194/nhess-18-3089-2018>
- Microsoft. (2019). Microsoft CanadianBuildingFootprints. [Dataset]. Retrieved from <https://github.com/microsoft/CanadianBuildingFootprints>
- Mohanty, M. P., Nithya, S., Nair, A. S., Indu, J., Ghosh, S., Mohan Bhatt, C., et al. (2020). Sensitivity of various topographic data in flood management: Implications on inundation mapping over large data-scarce regions. *Journal of Hydrology*, 590, 125523. <https://doi.org/10.1016/j.jhydrol.2020.125523>

- Molinari, D., & Scorzini, A. R. (2017). On the influence of input data quality to flood damage estimation: The performance of the INSYDE model. *Water*, 9(9), 688. <https://doi.org/10.3390/w9090688>
- Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., et al. (2020). 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>
- Muthusamy, M., Casado, M. R., Butler, D., & Leinster, P. (2021). Understanding the effects of Digital Elevation Model resolution in urban fluvial flood modelling. *Journal of Hydrology*, 596, 126088. <https://doi.org/10.1016/j.jhydrol.2021.126088>
- Papaoiannou, G., Loukas, A., Vasiliades, L., & Aronica, G. T. (2016). Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach. *Natural Hazards*, 83(S1), 117–132. <https://doi.org/10.1007/s11069-016-2382-1>
- Paprotny, D., Voudoukas, M. I., Morales-Nápoles, O., Jonkman, S. N., & Feyen, L. (2020). Pan-European hydrodynamic models and their ability to identify compound floods. *Natural Hazards*, 101(3), 933–957. <https://doi.org/10.1007/s11069-020-03902-3>
- Pollack, A. B., Sue Wing, I., & Nolte, C. (2022). Aggregation bias and its drivers in large-scale flood loss estimation: A Massachusetts case study. *Journal of Flood Risk Management*, 15(4). <https://doi.org/10.1111/jfr3.12851>
- Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V. D., et al. (2021). Process-based flood risk assessment for Germany. *Earth's Future*, 9(10). <https://doi.org/10.1029/2021EF002259>
- Saksena, S., & Merwade, V. (2015). Incorporating the effect of DEM resolution and accuracy for improved flood inundation mapping. *Journal of Hydrology*, 530, 180–194. <https://doi.org/10.1016/j.jhydrol.2015.09.069>
- Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. *Water Resources Research*, 24. <https://doi.org/10.1002/2015WR016954>
- Savage, J., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, 52(11), 9146–9163. <https://doi.org/10.1002/2015WR018198>
- Seifert, I., Kreibich, H., Merz, B., & Thielen, A. H. (2010). Application and validation of FLEMOcs – A flood-loss estimation model for the commercial sector. *Hydrological Sciences Journal*, 55(8), 1315–1324. <https://doi.org/10.1080/02626667.2010.536440>
- Seifert, I., Thielen, A. H., Merz, M., Borst, D., & Werner, U. (2010). Estimation of industrial and commercial asset values for hazard risk assessment. *Natural Hazards*, 52(2), 453–479. <https://doi.org/10.1007/s11069-009-9389-9>
- Sieg, T., & Thielen, A. H. (2022). Improving flood impact estimations. *Environmental Research Letters*, 17(6), 064007. <https://doi.org/10.1088/1748-9326/ac6d6c>
- Sieg, T., Vogel, K., Merz, B., & Kreibich, H. (2019). Seamless estimation of hydrometeorological risk across spatial scales. *Earth's Future*, 7(5), 574–581. <https://doi.org/10.1029/2018EF001122>
- Thielen, A. H., Cammerer, H., Dobler, C., Lammel, J., & Schöberl, F. (2016). 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(3), 343–376. <https://doi.org/10.1007/s11027-014-9602-3>
- Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., et al. (2020). Review article: Natural hazard risk assessments at the global scale. *Natural Hazards and Earth System Sciences*, 20(4), 1069–1096. <https://doi.org/10.5194/nhess-20-1069-2020>
- Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., et al. (2015). Usefulness and limitations of global flood risk models. *Nature Climate Change*, 5(8), 712–715. <https://doi.org/10.1038/nclimate2742>
- Wing, O. E. J., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., et al. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, 12(2), 156–162. <https://doi.org/10.1038/s41558-021-01265-6>
- Xafoulis, N., Kontos, Y., Farsitrou, E., Kotsopoulos, S., Perifanos, K., Alamanis, N., et al. (2023). 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(4), 91. <https://doi.org/10.3390/hydrology10040091>

References From the Supporting Information

- Apel, H., Vorogushyn, S., & Merz, B. (2022). 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(9), 3005–3014. <https://doi.org/10.5194/nhess-22-3005-2022>
- Dietze, M., Bell, R., Ozturk, U., Cook, K. L., Andermann, C., Beer, A. R., et al. (2022, June). More than heavy rain turning into fast-flowing water – A landscape perspective on the 2021 Eifel floods. *Natural Hazards and Earth System Sciences*, 22(6), 1845–1856. <https://doi.org/10.5194/nhess-22-1845-2022>
- Landesamt für Umwelt Rheinland-Pfalz. (2022). Hochwasser im Juli 2021 (Technical report). Landesamt für Umwelt Rheinland-Pfalz. Retrieved from https://ifu.rlp.de/fileadmin/ifu/Wasserwirtschaft/Ahr-Katastrophe/Hochwasser_im_Juli2021.pdf
- LISFLOOD developers. (2020). LISFLOOD-FP 8.0 hydrodynamic model. *Zenodo*. <https://doi.org/10.5281/zenodo.4073011>
- Milan Geoservice GmbH. (2023). Airborne laser scanning (ALS) AHR valley (Technical report no. D-01917 Kamenz).
- OpenStreetMap contributors. (2022). Planet dump. Retrieved from <https://planet.osm.org>, <https://www.openstreetmap.org>
- Shaw, J., Kesserwani, G., Neal, J., Bates, P., & Sharifian, M. K. (2021). LISFLOOD-FP 8.0: The new discontinuous Galerkin shallow-water solver for multi-core CPUs and GPUs. *Geoscientific Model Development*, 14(6), 3577–3602. <https://doi.org/10.5194/gmd-14-3577-2021>
- Szönyi, M., & Roezer, V. (2022). PERC flood event review 'Bernd' (Technical report). Retrieved from <https://www.newsroom.zurich.de/documents/zurich-perc-analysis-bernd-english-version-423750>
- Vorogushyn, S., Apel, H., Kemter, M., & Thielen, A. H. (2022). Analyse der Hochwassergefährdung im Ahrtal unter Berücksichtigung historischer Hochwasser. https://doi.org/10.5675/HyWa_2022.5_2
- Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. (2017). 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(9), 7968–7986. <https://doi.org/10.1002/2017WR020917>