Systematic Error in Flood Hazard Aggregation

Seth Bryant[1,2], **Heidi Kreibich**[1], **Bruno Merz**[1,2]

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

Key Points:

- Flood hazard aggregation is shown to carry systematic error for a specific case study and some general cases
- A new framework is introduced to spatially attribute systematic aggregation errors
- Evidence and explanation is provided for the positive dependence between scale coarseness and systematic error recurring in the literature

Corresponding author: Seth Bryant, `seth.bryant@gfz-potsdam.de`

Abstract

Reducing flood risk through improved disaster planning and risk management requires accurate and reliable estimates of flood damages. Models can provide such information by calculating the costs of flooding to exposed assets, such as buildings within a community. Computational or data constraints often lead to the construction of such models from coarse aggregated data, the effect of which is poorly understood. Through the application of a novel spatial segregation framework, we are able to show mathematically that aggregating flood grids through averaging will always introduce a systematic error in a particular direction in partially inundated regions. By applying this framework to a case study we spatially attribute these errors and demonstrate how the exposure of buildings can be an order of magnitude more sensitive to these errors than uninhabited regions. This work provides insight into, and recommendations for, upscaling grids used by flood risk models. Further, we demonstrate a positive dependence of systematic error magnitude on scale coarseness, suggesting coarse models be used with caution and greater attention be paid to issues of scale.

1 Introduction

With the increase in flood related disaster damages, the expansion of computation power, and the availability of global datasets, 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 modelling; with each step adding uncertainty (de Moel & 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 & 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 hazard, exposure, and vulnerability are treated with gridded data of resolutions from 100 to 1000m (Hall et al., 2005; Ward et al., 2015; 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 model or grid "support" of a fine ($s1$) or coarse ($s2$) grid (where $s1 < s2$) to avoid confusion with the more generic "scale" which can also refer to the related spatial extents (Bierkens et al., 2000). This is closely related to the resolution ($\lambda_{s1} < \lambda_{s2}$) of the corresponding square grid cells. 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. This transformation between resolutions is generally employed on flood hazard grids to improve model stability or to satisfy some computational constraints (Sampson et al., 2015).

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; Sampson et al., 2015). 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 be-

tween model coarseness 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 & Merwade, 2015; Mohanty et al., 2020; Ghimire & Sharma, 2021; Muthusamy et al., 2021) 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 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) used 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.

Many studies investigate flood risk model parameter sensitivity (Metin et al., 2018; Jongman et al., 2012; Apel et al., 2009; Seifert et al., 2010; Ghimire & Sharma, 2021), but few investigate sensitivity to scale 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, 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, 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 & Merwade, 2015; Muthusamy et al., 2021), the implications for risk models have not been explored systematically.

Table 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. (2010)	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. (2010)	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

In one of the few studies to investigate risk model sensitivity to grid aggregation specifically, Komolafe et al. (2015) performed a simulation experiment with a model cal-

ibrated 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 depths introduces a slight positive bias, while aggregating assets introduces a strong negative bias. 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) compared 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 scales and a mortality bias of +8%. Ghimire and Sharma (2021) provides 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) 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. 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 aggregated models used in practice.

The goal of this paper is to partially explain the bias shown in Table 1 through generalizeable methods (i.e., not bound to the specifics of individual case studies) and thereby improve our understanding of the effects of scale 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 scale 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; a less common practice, but one that is more amenable to analytical investigation. 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 Depth (WSH), Water Surface Elevation (WSE), and the Ground Elevations (DEM)

related by the following:

$$WSE = DEM + WSH \tag{1}$$

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:

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

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

$$WSH_{i \text{ or } j} = 0 \iff WSE_{i \text{ or } j} = null \iff \text{"dry"} \tag{3}$$

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), has a zero value in these same regions. Absent transformation or resampling, the application of Equation 1 and 3 is trivial and allows for simple conversion between WSE and WSH or vice versa using the DEM as shown in Figure 1c. However, in the presence of dry cells Equation 3 leads to inconsistencies when computing the denominator of averaging operations:

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

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

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

where N_{12} is the count of $s1$ cells contributing to a coarse $s2$ cell, and $N_{wet} = N_{12} - N_{dry}$ where N_{dry} is the count of $s1$ cells described in Equation 3. Later, we show how these inconsistencies can lead to systematic errors in aggregation routines.

3 Methods

To investigate the potential for systematic errors to be introduced through aggregating of fluvial flood hazard data, we introduce a 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 flood in Canada as an example. This analysis is then extended to consider only exposed regions (locations with buildings) to provide an analysis of systematic errors particularly relevant to flood risk models.

3.1 Aggregation Routines

To demonstrate the application of our framework, we consider two routines for yielding a set of $s2$ analog grids from a set of $s1$ grids through averaging local groups of size N_{12}. Both respect Equation 1 and 2, but differ on the strategy for preserving averages in the resulting $s2$ analogs: the first preserving WSH ("WSH Averaging") and the second WSE ("WSE Averaging"). In this way, each routine has a primary grid (WSH or WSE), which is computed through direct averaging, and a secondary grid (WSH or WSE). 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 Figure 1d and e). Figure 1d and e provide a graphical summary and toy example of these routines, which are defined mathematically in the Supplement. Both routines are easily implemented in a few steps using standard spatial software packages (GDAL, Whitebox Tools, QGIS, rasterio, etc.) 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 provide a reasonable approximation of analog grids built with hydrodynamic models.

3.2 Resample Case

To understand and spatially attribute the effects of such 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 2 and defined explicitly as:

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

where 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 groups whose corresponding case labels are shown on Figure 1b. Figure 3 shows a fully classified domain where WSH_{s1} has been simulated using a hydrodynamic model built on a 1m fine DEM described below. Such a resample case map is independent of any $s2$ grids resulting from a specific aggregation routine. However, this classification provides simplifying assumptions for the investigation of aggregation behaviour by considering each case region independently. For example, the DD and WW regions we expect to be fully dry and fully wet respectively in the $s2$ grids, regardless of the aggregation routine. The partial regions (DP and WP) on the other hand are ambiguous, and we expect $s2$ grid results can differ based on the routine applied.

3.3 Analytical Approach

For this evaluation, we define error as the difference between a "true" value, a physical property, and the modelled value, taken here as the corresponding grid value. For example, the "true" WSH of a flood event could be measured at a discrete point in space and time (say 1m), and compared to the value at the corresponding location in the WSH grid (say 1.5m) to quantify the grid error (+0.5m in this case). For the purposes of this analysis, we assume "true" values are represented in the fine ($s1$) grid. 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 *random* errors, i.e. 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 *random* 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 from here on we call "bias", are an undesirable artifact of aggregation and the focus of this analysis.

Biases in the two aggregation routines are first investigated analytically to derive inequalities between metrics computed on the fine ($s1$) and coarse ($s2$) grids. To accomplish this, each of the four "resample cases" is investigated separately, which provides the simplifying assumptions that allow closed-form solutions to the errors in each metric of interest. Bias is evaluated in four metrics of interest to flood models: 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 5 and 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 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 (λ_s^2) and volume (V_s) is the sum of all WSH_s values multiplied by the area of each cell (λ_s^2).

To better attribute bias spatially, we also compute a "local" bias for 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 $s2$ vs. $s1$ cell, 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 2), as the grid is undefined in other regions. For consistency, we apply this same constraint to the WSH metric. 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 separate metric).

3.4 Computational Approach

To demonstrate the application of the novel "resample case" framework, we apply the aggregation routines to a set of 1m resolution grids from the May 2018 Saint John River flood in Canada. 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 described in Bryant et al. (2022). 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 ($\frac{1}{s2} = 2^n$ 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 x 5 coarse scales x 2 routines). Komolafe et al. (2015) takes 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 black "buildings"). From the centroids of 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.

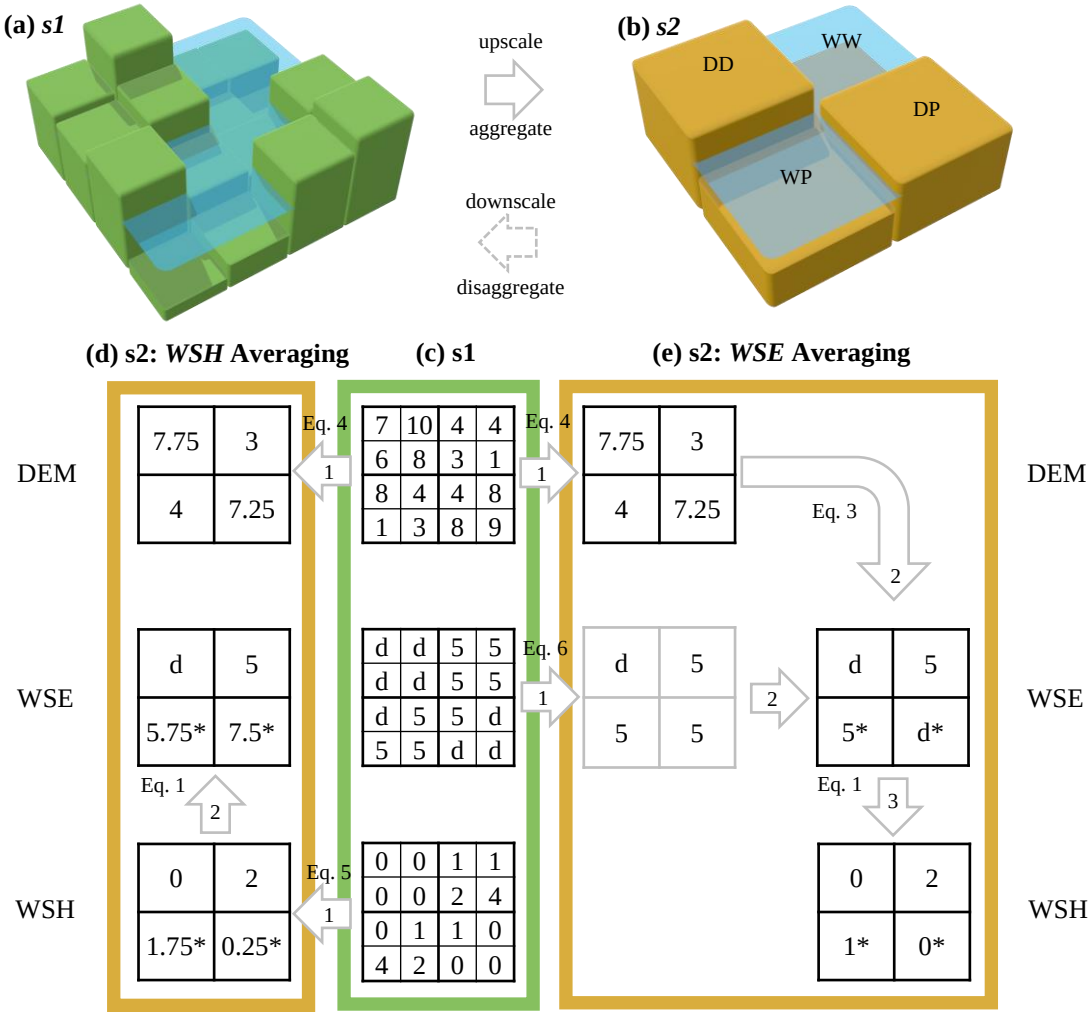

Figure 1. Flood hazard data scaling issues and two aggregation schemes demonstrated with a toy example. Panel (a) is an oblique view of a fine ($s1$) DEM and WSE while panel (b) shows an aggregated coarse ($s2$) analog and corresponding resample case (DD, WW, WP, DP) from Figure 2. Panel (c) shows an 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 different phases within these schemes, 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 *.

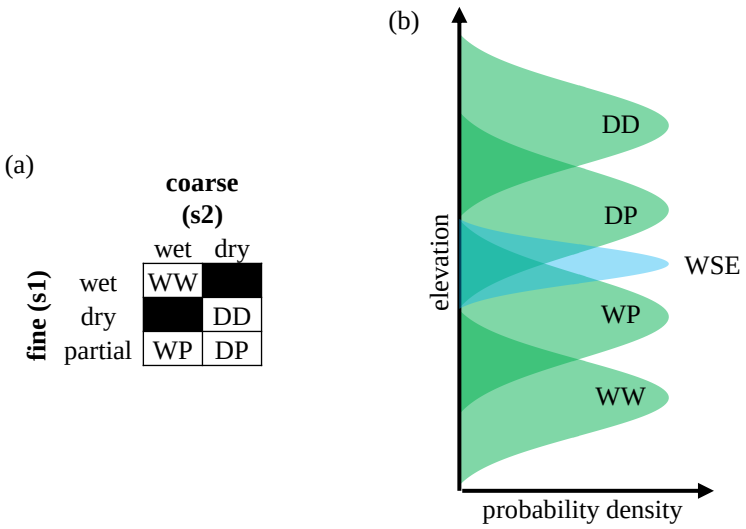

Figure 2. Framework for classification of flood hazard resample case. Panel (a) shows class label acronyms. Panel (b) provides a conceptual diagram showing a hypothetical distribution of WSE_{s1} and four possible DEM_{s1} groups and their resulting resample case. D, W, and P stand for "dry", "wet",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

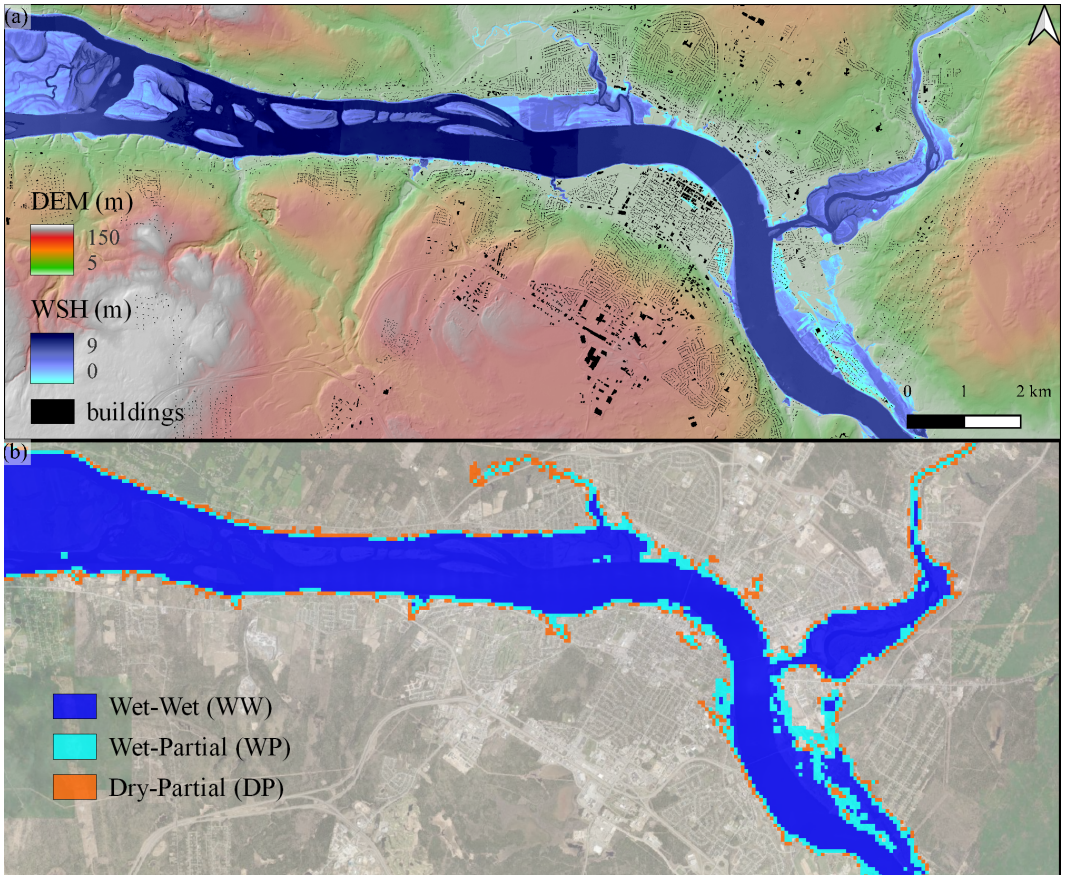

Figure 3. Simulated May 2018 Saint John River flood in Canada. Panel (a) shows DEM_{s1} and WSH_{s1} at 1m resolution and building footprints from Microsoft (2019). Panel (b) shows corresponding resample case (see Figure 2) for a 1:64 upscale (DD is transparent for clarity).

4 Analytical Results and Discussion

To investigate the six metrics of interest (A, V and local and global WSH and WSE), we applied the "resample case" framework to the two aggregation routines (details in the Supplement). Results are summarized in Table 2.

Focusing on the non-partial columns (DD and WW), Table 2 shows that aggregation preserves all our metrics of interest in these regions. 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 — 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 and 1). Examining the bias in 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 $Bias_{global}[\overline{WSE}]$ 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 and 1.

Table 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	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				
$Bias_{global}[\overline{WSH}]$	0	-	-	0
$Bias_{local}[\overline{WSH}]$	0	n/a	-	0
$Bias_{global}[\overline{WSE}]$	n/a	n/a	0	0
$Bias_{local}[\overline{WSE}]$	n/a	n/a	0	0
$Bias[\sum A]$	0	-	+	0
$Bias[\sum V]$	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 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 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.

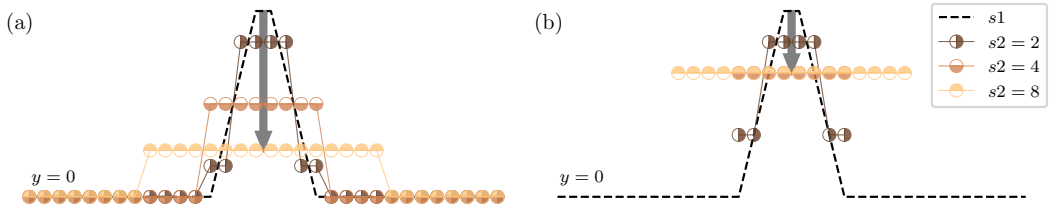

Figure 4. Conceptual diagram showing a cross-section of local bias generated through two types of averaging: (a) zero-inclusion (as in Equation 5) and (b) zero-exclusion (as in Equation 6). All series within a panel have the same global mean. Black arrow shows the progression of local bias.

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 highly consequential for flood risk models, considering changes to flood footprints are expected to lead to changes in flood exposure, a highly sensitive component (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 behaviour as $Bias_{global}[\overline{WSH}]$ (see Supplement 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 violates the calibration.

This analysis has shown mathematically whether or not a metric will be biased by a given routine aggregating a hypothetical grid. By employing the "resample case" framework, these bias solutions 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 (see below), 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 (i.e., a grid could conceivably have only one resample case, rendering most of the analysis here irrelevant). With this in mind, the following section applies a similar analysis computationally to a case study. Further, the reader should note that requiring the "resample case" segregation is a significant limitation, as this requires the original $s1$ grids.

5 Computational Results and Discussion

To evaluate aggregation bias, the resample case framework and the two aggregation routines are applied to a case study of the May 2018 Saint John River flood in Canada. 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 of 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 $s2$ scale. This shows that the portion of partial regions (WP and DP) increases with aggregation. This is intuitive if we consider 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 if we consider the previous section showed these partial regions are those which generate bias during aggregation. In other words, the portion of the domain subject to aggregation bias increases with resolution. Further, these transition zones, or shorelines, often have a high-density of assets — a phenomena explored in Figure 5c and discussed below.

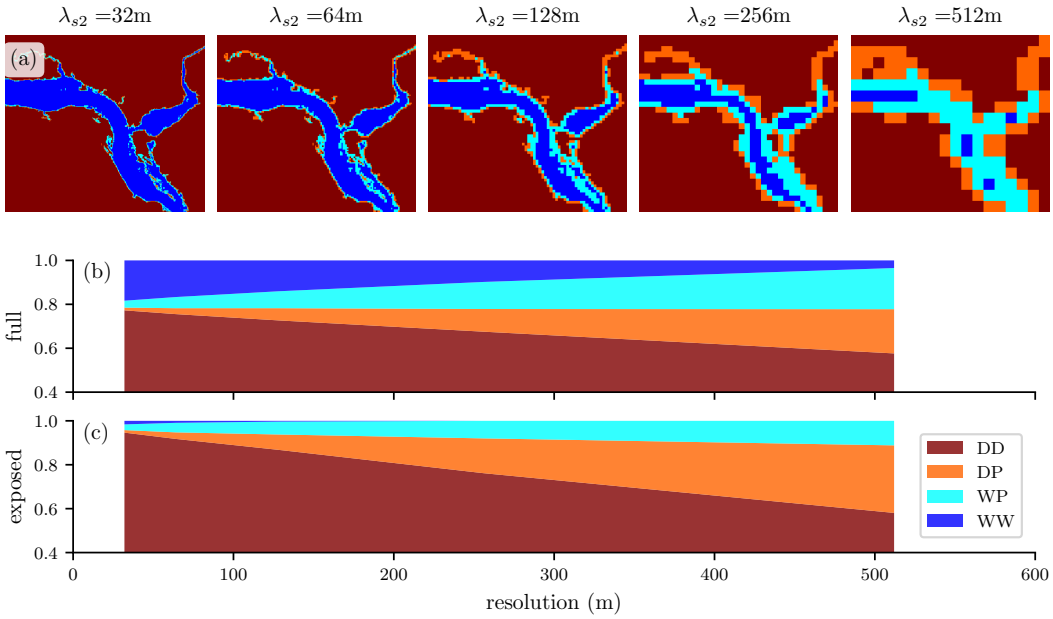

Figure 5. Resample case classification progression for May 2018 Saint John River flood hazard data showing (a) illustrative maps at five scales; (b) full domain fraction; and (c) exposed domain (i.e., values sampled at buildings – see text) for each case. See Figure 2 for description of legend.

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 of four

key metrics are shown in Figure 6a and b and the remaining two metrics are provided in Figure S1.

Comparing Figure 6 and 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 the Supplement 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 2, Figure 6a0 and b0 and Figure S1); 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 but only report $\overline{WSE}$, which they also find increasing. This contrast 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. Of more value would be a comparison against a similarly coarse hydrodynamic model calibrated to high water marks.

For all partial zones, "WSH Averaging" shows a doubling (100% increase) of the inundated area (A) for the $\lambda_2 = 512m$ grids for this case study. "WSE Averaging" fared better, with the WP and DP global bias nearly balancing, leading to a meagre 10% increase for $\lambda_2 = 512m$. 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. These increases in flood footprint are in line with those reported by coarse hydrodynamic model comparisons (Banks et al., 2015; Saksena & Merwade, 2015; Mohanty et al., 2020; Ghimire & Sharma, 2021; Muthusamy et al., 2021).

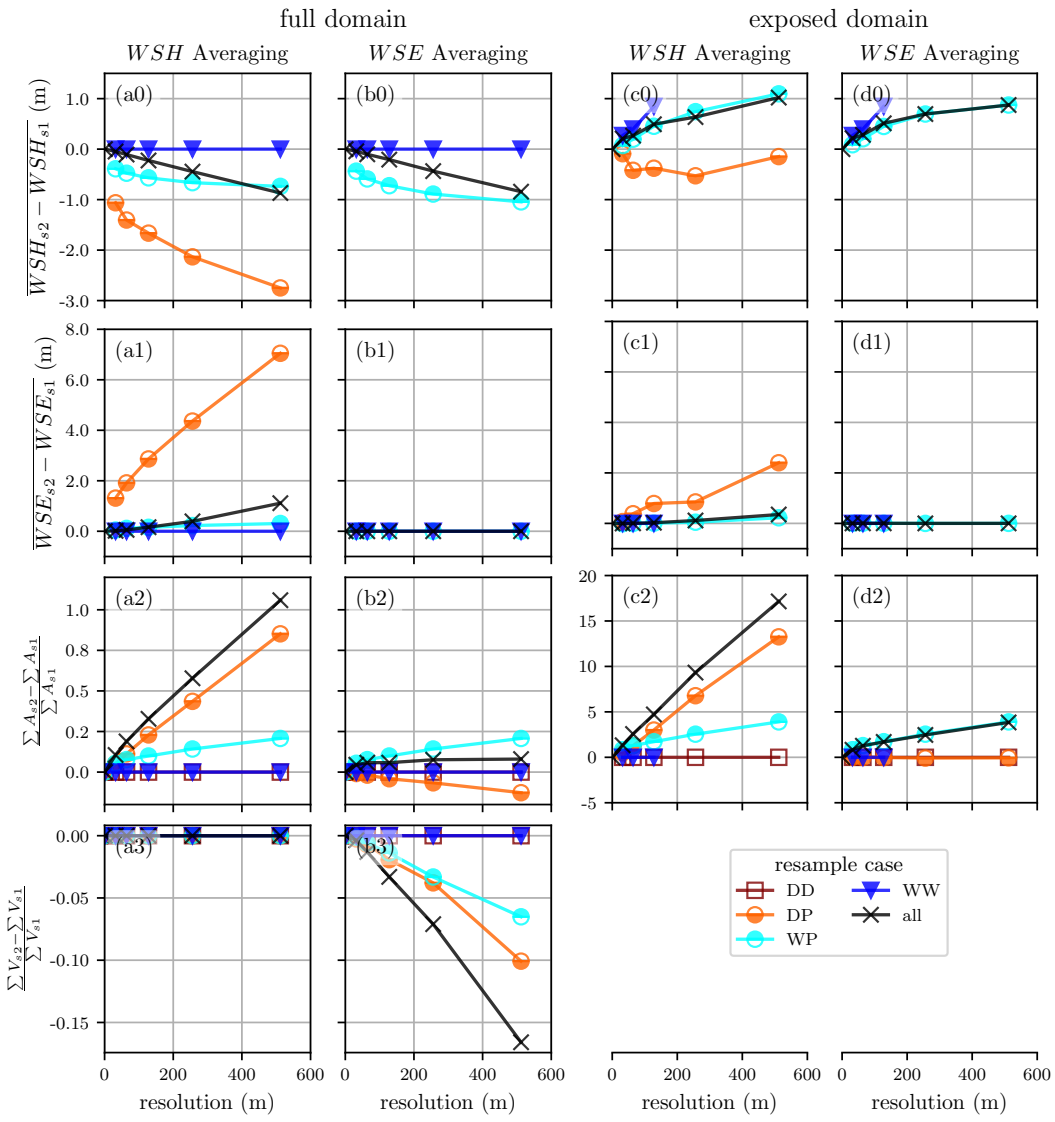

Figure 6. Bias from aggregation of four metrics for two routines sub-sampled for the full domain and the exposed domain (i.e., values sampled at buildings – see text) by resample case. See Figure 2 for description of resample cases. The "all" series uses the complete region of interest, without sub-setting by resample case. $\sum A$ is the non-dry area of the full domain (panels (a2) and (b2)) and the count of non-dry (i.e., exposed) buildings (panels (c2) and (d2)).

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 models: developed areas or areas with exposure. Figure 5c shows that WW regions are insignificant for building exposure. This is intuitive if we consider: 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 512m, compared to 20% on the full domain for this case study. Recalling that these partial regions (WP and DP) are those responsible for the bias produced by aggregation suggests that exposure 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 6c and d and Figure S2. 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). This is intuitive if we consider the distribution of buildings: few in regions flooded by the base grids and many immediately adjacent. 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) found a comparable factor of 2 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; 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 6a 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 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. We suspect this "dry bias" is equally relevant for coarse hydrodynamic models although we can find no such discussion in the literature; however, this mechanism should be present in Ghimire and Sharma (2021) and Pollack et al. (2022). In fact, Pollack et al. (2022) discusses a counter mechanism, 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.

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 partially inundated regions. We then applied this framework to a case

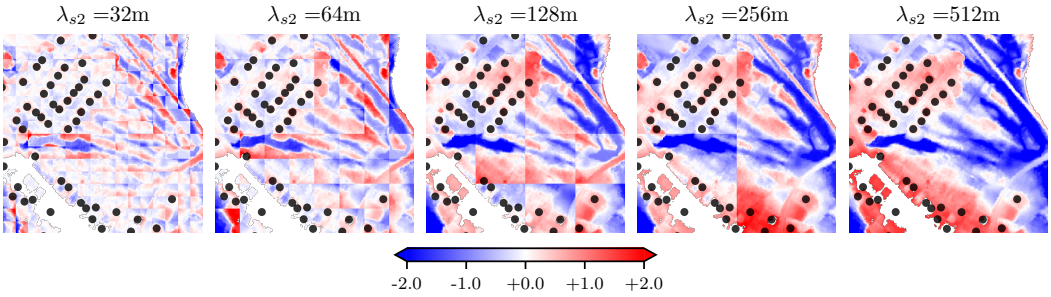

Figure 7. Maps of 512m example region at five resolutions aggregated with the "WSH Averaging" routine showing building centroid locations in black. To compute local errors, WSH_{s2} grids are downscaled to $s1$ then WSH_{s1} is subtracted, yielding the $WSH_{s2} - WSH_{s1}$ values shown in meters on a red-blue colour scale

study of a 2018 Canadian flood to spatially spatially attribute the biases to the flood fringes or edges. Using this case study we provide example magnitudes of these biases for each metric showing, for example, inundation area can double at an aggregation of 2^9. Finally, this case study was extended to show how those regions with assets or buildings are particularly sensitive to this bias – sometimes in counter-intuitive ways.

Through attributing and deriving errors, these results have direct utility for those seeking to aggregate or upscale flood hazard grids. In addition to formally defining two routines, in Table 2 we have shown to what extent and in which regions metrics are preserved: providing a framework for evaluating additional routines and enabling practitioners to make more informed decisions when selecting a routine. For example, in a hazards focus analysis where flood volume and average depth are of importance: a routine similar to the "WSH Averaging" should be pursued. However, we suspect most flood risk modellers would place more emphasis on exposure accuracy, suggesting the "WSE Averaging" routine. Regardless, some trade offs must be considered when selecting the appropriate routine. To support technical implementation, an open-source QGIS script is provided here (`https://github.com/cefect/FloodRescaler`).

In practice, we recognize scale transfers in flood risk models through grid aggregation generally involve only small changes in scale; and the errors introduced are minor compared to other sources of uncertainty (Pollack et al., 2022; Ghimire & Sharma, 2021). More prevalent is the use of hydrodynamic models, where the friction term is calibrated to observed high water marks, to develop WSH grids from aggregated or coarse DEM grids. These practices however are less amenable to the analytical methods employed here. Considering this, our exploration of grid aggregation may provide a simplified analog through which to better understand systematic errors in hydrodynamic models, especially in regions with buildings or assets. However, additional work is required to understand the limits of this comparison.

The results presented here for the exposed domain all show a positive bias (Figure 6c and d), like the previous studies summarized in Table 1 and a growing body of work on hydrodynamic models (Banks et al., 2015; Saksena & Merwade, 2015; Mohanty et al., 2020; Ghimire & Sharma, 2021; Muthusamy et al., 2021). While our work stops short of computing risk or impact metrics like those in Table 1, the remarkable four-fold increase in exposed assets shown in Figure 6d2 provides a logical, albeit partial, explanation for the bias shown. Figure 7 provides a graphical demonstration of 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, introducing a negative bias. Considering this, our findings and those of similar studies are likely somewhat sensitive to the study area and the flooding mechanism, and especially sensitive to the magnitude of the scale transfer. Regardless, a more comprehensive understanding of these competing biases is sorely needed to fully explain the biases shown in 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 these with the aggregated grids explored here, then applying these as inputs to damage functions developed on single assets. To attribute and correct for systematic errors which may emerge through such scale transfers, the frameworks and findings developed here could be extended to study such processes. 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.

Open Research Section

The python scripts used to construct the aggregated grids, sample the grids at building locations, compute the metrics, and generate the plots are provided here: `https://github.com/cefect/2112_agg_pub`. An easy-to-use QGIS script for aggregating flood hazard grids is provided here: `https://github.com/cefect/FloodRescaler`. The DEM_{s1} grid used in the computation approach is hosted by GeoNB (`http://geonb.snb.ca/li/index.html`) and the Saint John River 2018 maximum WSH_{s1} data is also hosted by GeoNB (`http://www.snb.ca/geonb1/e/DC/floodraahf.asp`) under the "GeoNB Open Data License" (`http://www.snb.ca/e/2000/data-E.html`).

Acknowledgments

Jody Reimer, Kai Schröter, Lukas Schoppa, NatRiskChange

References

Alipour, A., Jafarzadegan, K., & Moradkhani, H. (2022, June). Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping. *Environmental Modelling & Software*, *152*, 105398. Retrieved 2022-05-06, from `https://linkinghub.elsevier.com/retrieve/pii/S1364815222001049` doi: 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. Retrieved 2016-10-01, from `http://link.springer.com/article/10.1007/s11069-008-9277-8`

Banks, J. C., Camp, J. V., & Abkowitz, M. D. (2015, August). Scale and Resolution Considerations in the Application of HAZUS-MH 2.1 to Flood Risk Assessments. *Natural Hazards Review*, *16*(3), 04014025. doi: 10.1061/(ASCE)NH.1527-6996.0000160

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. Retrieved 2022-09-07, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12713`

(_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfr3.12713) doi: 10.1111/jfr3.12713

Bryant, S., McGrath, H., & Boudreault, M. (2022, April). Gridded flood depth estimates from satellite-derived inundations. *Natural Hazards and Earth System Sciences*, *22*(4), 1437–1450. Retrieved 2022-05-03, from https://nhess.copernicus.org/articles/22/1437/2022/ doi: 10.5194/nhess-22-1437-2022

de Moel, H., & Aerts, J. C. J. H. (2011, July). Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. *Natural Hazards*, *58*(1), 407–425. Retrieved 2016-10-01, from http://link.springer.com/10.1007/s11069-010-9675-6 doi: 10.1007/s11069-010-9675-6

Fewtrell, T. J., Bates, P. D., Horritt, M., & Hunter, N. M. (2008, December). Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes*, *22*(26), 5107–5118. Retrieved 2022-05-06, from https://onlinelibrary.wiley.com/doi/10.1002/hyp.7148 doi: 10.1002/hyp.7148

Ghimire, E., & Sharma, S. (2021, February). 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. doi: 10.1061/(ASCE)NH.1527-6996.0000430

Government of New Brunswick. (2016). *ERD 2015 Lidar.* Retrieved from http://geonb.snb.ca/li/index.html

Hall, J. W., Sayers, P. B., ., & Dawson, R. J. (2005, September). National-scale Assessment of Current and Future Flood Risk in England and Wales. *Natural Hazards*, *36*(1-2), 147–164. Retrieved 2018-11-13, from http://link.springer.com/10.1007/s11069-004-4546-7 doi: 10.1007/s11069-004-4546-7

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

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. Retrieved 2018-03-05, from http://doi.wiley.com/10.1111/jfr3.12224 doi: 10.1111/jfr3.12224

McGrath, H., Stefanakis, E., & Nastev, M. (2015, December). Sensitivity analysis of flood damage estimates: A case study in Fredericton, New Brunswick. *International Journal of Disaster Risk Reduction*, *14*, 379–387. Retrieved 2016-10-25, from http://linkinghub.elsevier.com/retrieve/pii/S2212420915300625 doi: 10.1016/j.ijdrr.2015.09.003

Merz, B., Kreibich, H., Thieken, A., & Schmidtke, R. (2004, March). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Sciences*, *4*(1), 153–163. Retrieved 2022-07-13, from https://nhess.copernicus.org/articles/4/153/2004/ doi: 10.5194/nhess-4-153-2004

Metin, A. D., Dung, N. V., Schröter, K., Guse, B., Apel, H., Kreibich, H., ... Merz, B. (2018, November). How do changes along the risk chain affect flood risk? *Natural Hazards and Earth System Sciences*, *18*(11), 3089–3108. Retrieved 2021-05-04, from https://nhess.copernicus.org/articles/18/3089/2018/ doi: 10.5194/nhess-18-3089-2018

Microsoft. (2019). *microsoft/CanadianBuildingFootprints.* Retrieved 2022-04-11, from https://github.com/microsoft/CanadianBuildingFootprints

Mohanty, M. P., Nithya, S., Nair, A. S., Indu, J., Ghosh, S., Mohan Bhatt, C., ... Karmakar, S. (2020, November). Sensitivity of various topographic data in flood management: Implications on inundation mapping over large data-scarce regions. *Journal of Hydrology*, *590*, 125523. Retrieved 2022-05-06, from

https://linkinghub.elsevier.com/retrieve/pii/S0022169420309835 doi: 10.1016/j.jhydrol.2020.125523

Molinari, D., & Scorzini, A. R. (2017, September). On the Influence of Input Data Quality to Flood Damage Estimation: The Performance of the INSYDE Model. *Water*, *9*(9), 688. Retrieved 2022-03-22, from http://www.mdpi.com/2073-4441/9/9/688 doi: 10.3390/w9090688

Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., ... Ballio, F. (2020, February). Are flood damage models converging to reality? Lessons learnt from a blind test. *Natural Hazards and Earth System Sciences*. Retrieved 2020-11-10, from https://nhess.copernicus.org/preprints/nhess-2020-40/nhess-2020-40.pdf doi: 10.5194/nhess-2020-40

Muthusamy, M., Casado, M. R., Butler, D., & Leinster, P. (2021, May). Understanding the effects of Digital Elevation Model resolution in urban fluvial flood modelling. *Journal of Hydrology*, *596*, 126088. Retrieved 2022-05-06, from https://linkinghub.elsevier.com/retrieve/pii/S0022169421001359 doi: 10.1016/j.jhydrol.2021.126088

Papaioannou, G., Loukas, A., Vasiliades, L., & Aronica, G. T. (2016, October). Flood inundation mapping sensitivity to riverine spatial resolution and modelling approach. *Natural Hazards*, *83*(S1), 117–132. Retrieved 2022-05-06, from http://link.springer.com/10.1007/s11069-016-2382-1 doi: 10.1007/s11069-016-2382-1

Pollack, A. B., Sue Wing, I., & Nolte, C. (2022, August). Aggregation bias and its drivers in large-scale flood loss estimation: A Massachusetts case study. *Journal of Flood Risk Management*. Retrieved 2022-09-15, from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.12851 doi: 10.1111/jfr3.12851

Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V. D., ... others (2021). Process-Based Flood Risk Assessment for Germany. *Earth's Future*, *9*(10). (Publisher: Wiley Online Library) doi: 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. (Publisher: Elsevier)

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.

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. Retrieved 2022-05-06, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR018198 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR018198) doi: 10.1002/2015WR018198

Seifert, I., Thieken, A. H., Merz, M., Borst, D., & Werner, U. (2010, February). Estimation of industrial and commercial asset values for hazard risk assessment. *Natural Hazards*, *52*(2), 453–479. Retrieved 2019-11-04, from http://link.springer.com/10.1007/s11069-009-9389-9 doi: 10.1007/s11069-009-9389-9

Sieg, T., Vogel, K., Merz, B., & Kreibich, H. (2019). Seamless estimation of hydrometeorological risk across spatial scales. *Earth's Future*, *7*(5), 574–581. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018EF001122 (Publisher: Wiley Online Library)

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., ... Winsemius, H. C. (2020, April). Review article: Natural hazard risk assessments at the global scale. *Natural Hazards and Earth System Sciences*, *20*(4), 1069–1096. Retrieved 2020-11-02,

from https://nhess.copernicus.org/articles/20/1069/2020/ doi: 10.5194/nhess-20-1069-2020

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., ... Winsemius, H. C. (2015, August). Usefulness and limitations of global flood risk models. *Nature Climate Change*, *5*(8), 712–715. Retrieved 2022-07-14, from http://www.nature.com/articles/nclimate2742 doi: 10.1038/nclimate2742