Check for updates

Received: 8 October 2023 | Revised: 12 January 2024 | Accepted: 5 February 2024

DOI: 10.1111/jfr3.13053

SPECIAL ISSUE PAPER

CIWEM Chartered Institution of Water and Environmental Management | Journal of Flood Risk Management | WILEY

Aggregating flood damage functions: The peril of Jensen's gap

Seth Bryant[1,2] | **Jody Reimer**[3] | **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

[3]Department of Mathematics & School of Biological Sciences, University of Utah, Salt Lake City, Utah, USA

Correspondence
Seth Bryant, GFZ German Research Centre for Geosciences, Section 4.4. Hydrology, Potsdam, Brandenburg, Germany.
Email: bryant.seth@gmail.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: GRK 2043/2

Abstract

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

KEYWORDS

flood damages, modelling, risk assessment, uncertainty analysis

1 | INTRODUCTION

With the increase in flood-related disaster damages, the expansion of computation power, and the availability of global data, development and application of meso- and macroscale flood risk models have increased in the past decade (Ward et al., 2020). These models are composed of a series of sub-models for the flood hazard, exposure of assets, and vulnerability to flooding, where vulnerability modeling, the last step in the chain, is generally found to be the most uncertain component in micro- and mesoscale 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 often found between different models and actual observations (Jongman et al., 2012; McGrath et al., 2015; Molinari et al., 2020). Further challenges we elaborate below are introduced when such models are transferred to the macroscale, where many exposed assets are aggregated through averaging into a single unit before vulnerability models are applied (Hall et al., 2005; Sairam et al., 2021;

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2025 The Author(s). *Journal of Flood Risk Management* published by Chartered Institution of Water and Environmental Management and John Wiley & Sons Ltd.

Ward et al., 2020). This process collapses heterogeneities, for instance in the flood depth, within the aggregated unit and generally results in overestimation of the total flood damage (Bryant et al., 2023).

In a flood vulnerability model, flood damage functions (f) provide a mathematical relationship between hazard and vulnerability variables (e.g., flood depth) and the estimated damages from flooding (e.g., building repair costs) for an individual asset (White, 1945). The most basic functions directly relate flood depth to damage—so-called depth-damage curves widely attributed to White (1945). Damage functions are typically categorized based on the model's focus or objective, such as the sector (residential vs. nonresidential), tangibility (tangible vs. intangible), damage mechanism (indirect vs. direct), and uncertainty treatment (deterministic vs. probabilistic) (Merz et al., 2010). Further classification considers function structure such as continuity (discrete vs. continuous) and for tangible economic functions the asset total value relation (relative vs. absolute) (Merz et al., 2010). Gerl et al. (2016) provide a comprehensive review of 47 flood damage functions, the majority of which are deterministic (96%), multi-variable (88%), and express loss relative to the total value of the asset (56%). To provide a standardized library of these functions, each was standardized to a common set of indicator variables, while unique indicators were left as default values. After this standardization, Gerl et al. (2016) found significant heterogeneity in function shape and magnitude.

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

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

2 | JENSEN'S INEQUALITY

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

$$\overline{g(x)} \neq g(\overline{x}), \tag{1}$$

where g is a non-linear function and x is an independent variable. Applied to flood vulnerability models, which are generally nonlinear (Gerl et al., 2016), Equation 1 implies that aggregating or averaging assets (e.g., buildings) introduces errors when those aggregate values are used as inputs into a vulnerability model. This difference between modeling entities on an individual basis versus in aggregate is called *Jensen's gap* (see Figure 1). Jensen (1906) proved that the magnitude and direction of this gap is related to the variance of the independent variable σ_x^2 and the local shape of the function $g''(\overline{x})$. Convex functions result in positive gaps or an underestimating aggregate model $\overline{g(x)} > g(\overline{x})$ while concave functions result in negative gaps or an overestimating aggregate model $\overline{g(x)} < g(\overline{x})$. The limits of the gap size can only be determined for certain types of problems (Liao & Berg, 2018; Walker, 2014), such as when x follows a mean centered distribution (Gao et al., 2017), or when $g(x)$ has a Taylor expansion (Abramovich & Persson, 2016), but are difficult to determine for other classes of problems. In summary, any model containing non-linear functions and variance within the independent variable will introduce some bias when measurements of that variable are aggregated.

3 | METHODS

To evaluate the sensitivity of flood damage models to Jensen's inequality, a simulation experiment is used to

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CIWEM Chartered Institution of Water and Environmental Management Journal of Flood Risk Management—WILEY

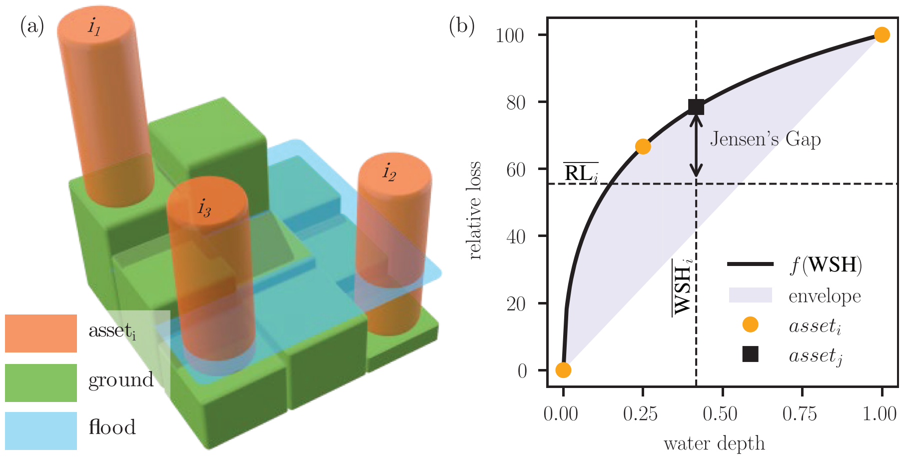

FIGURE 1 Illustrative example of *Jensen's gap* applied to an idealized flood damage function. (a) An example of three base or child assets i (e.g., buildings) which are to be aggregated into a single parent asset j (not shown). The child assets (i) are exposed to the flood depths plotted on the x-axis of panel (b), the average of which ($\overline{\mathrm{WSH}_i}$) is taken as the exposure of the aggregate asset (j) (black-dashed vertical line). (b) The relative losses resulting from the three child assets and the single aggregate asset on the y-axis. Finally, the mean of the child asset losses ($\overline{\mathrm{RL}_i}$; horizontal black-dashed line) is compared to the loss of the aggregate asset to demonstrate and quantify Jensen's gap. See later text for a discussion of the *envelope*.

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

For this study, we focus on direct tangible economic functions for estimating the relative loss to residential buildings from flood depth or water surface height (WSH) in Germany. While Jensen's gap is a mathematical phenomenon, and therefore indifferent to damage function categories like tangibility (tangible vs. intangible) or damage mechanism (indirect vs. direct), we focus on direct tangible economic functions as this category is the most common (Merz et al., 2010) and has the longest history of application (White, 1945) and therefore provides the most clear and relevant examples for demonstrating Jensen's gap. From the standardized function database of Gerl et al. (2016), three such flood damage functions were obtained. A fourth more recent function developed from private-household surveys in Germany (Wagenaar et al., 2018) was also preprocessed in a similar manner and included. The complete collection of the database from Gerl et al. (2016) is shown in Figure S4. All of these are in tabular format and have been simplified as univariate functions of depth by selecting default variable values for the multivariate functions (FLEMO and BN-FLEMOps). In this form, these functions can be implemented in a damage or vulnerability model as:

$$\mathrm{TL} = \sum_{j}^{n} \mathrm{RL}_j{}^{*} V_j, \tag{2}$$

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

$$\mathrm{RL}_j = f\left(\mathrm{WSH}_j\right), \tag{3}$$

where f is the univariate damage function of the flood depth or water surface height (WSH) at asset j. The two modern and two legacy damage functions selected for this study are summarized in Table 1. These were developed from post-flooding records or surveys of damages to individual buildings in Germany, except for IKSE for which the documentation is unclear. The legacy damage functions, IKSE and MURL, were intended as meso-scale damage functions, to be applied on the spatial basis of land use areas; however, how the transformation from per-building records was performed was not documented. The more modern functions, FLEMO and BN-FLEMOps, were intended as micro-scale damage functions, to be applied on the spatial basis of buildings or structures. When flood damage *models* are formulated for the *macro*-scale, it is common to group many buildings into a single computational unit, that is, aggregate many

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Selected flood damage functions.

Name	Full name	Spatial design basis	Envelope (cm × %)[a]	Reference
FLEMO	Flood Loss Estimation MOdel	Per-building	8978	(Thieken et al., 2008)
IKSE	Internationale Kommission zum Schutz der Elbe	Per-building	14,500	(Elbe, 2003)
MURL	Ministerium für Umwelt, Raumordnung und Landwirtschaft des Landes Nordrhein-Westfalen	Land use	2500	(Nordrhein-Westfalen, 2000)
BN-FLEMOps	Bayesian Network Based Damage Model for the private sector	Land use	6566	(Wagenaar et al., 2018)

[a]Envelope of potential error as in Figure 1. See Figure S1 for calculation.

underlying child-assets i into a single parent-asset j (see Figure 1 for illustration). Importantly, while this aggregation implies some transformation of the independent variable WSH (from i to j), neither in practice nor in our study is a similar transformation applied to the damage function f. It is this mismatch, or the application of child-asset-derived damage functions to aggregate assets, which is the focus of our study.

To exclude other artifacts of aggregation that may impact model behavior (e.g., raster aggregation, exposure sampling), we employ an experimental damage model designed to estimate aggregation error, rather than an applied model designed to estimate flood damage. Our experimental model has two key differences from a more traditional applied model. First, to eliminate any confounding influences from asset categorization, a single damage function is applied in each scenario to all assets, rather than categorizing and applying functions heterogeneously to assets based on some exposure variable (e.g., building use) as is typically done in applied models. Second, the experimental model focuses on relative losses rather than total losses, thereby eliminating the influence of heterogeneous building replacement values; however, a proxy for total loss using homogeneous replacement values is also provided.

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

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

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

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

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CIWEM Chartered Institution of Water and Environmental Management Journal of Flood Risk Management—WILEY

4 | RESULTS AND DISCUSSION

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

4.1 | Depth variance

Considering the variance (σ_2) of the independent variable (WSH_i) is directly related to Jensen's inequality (Denny, 2017), we expect loss calculations on aggregate assets with non-negligible child depths variance to yield some error (when compared with the loss calculated from the child asset depths directly). To show this variance within our aggregations, Figure 2 presents the statistics for the child depths for the four hazard scenarios (rows)

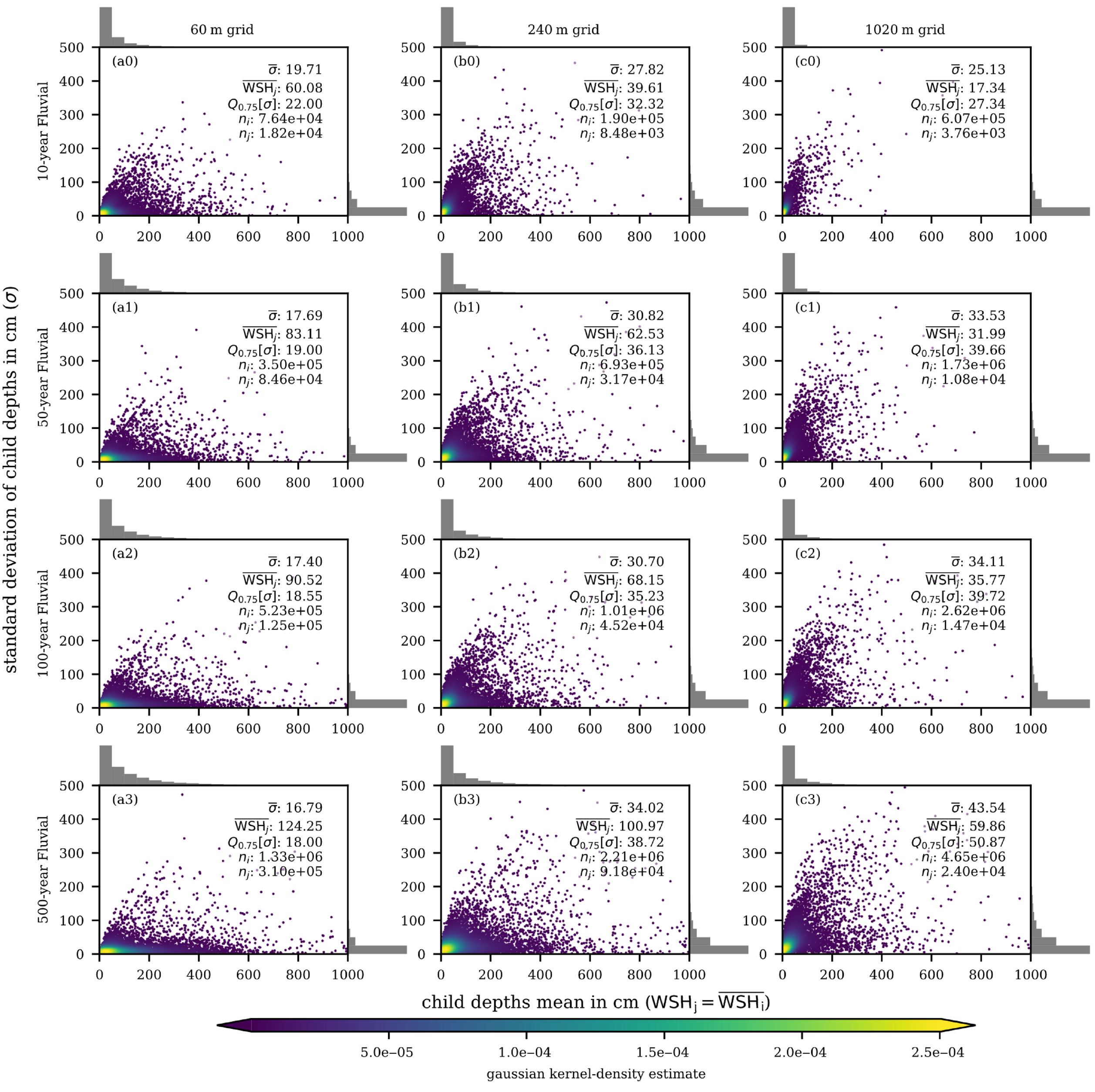

FIGURE 2 Germany-wide child depths mean and standard deviation for three levels of grid-aggregation and four hazard scenarios. Histograms show the distribution of their respective axis. Text in each panel provides the mean ($\overline{\sigma}$) and 0.75 quantiles ($Q_{0.75}[\sigma]$) of the child depths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

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and three aggregation sizes (columns). Focusing first on the child asset (i) exposure counts (i.e., the total number of buildings within the exposed domain), this plot shows an increase with both hazard severity and grid size, reflecting the increasing footprints of these scenario dimensions. The aggregate (j) exposure counts on the other hand show a similar trend with hazard severity but decrease in count with the larger grid sizes. Aggregated water depths (WSH_j) follow this same trend: the larger grid sizes capture more dry assets thereby reducing the mean.

Looking from left-to-right shows that, with one exception, variance increases with increasing grid size as expected: the greater the number and source area of child buildings, the more likely their depths are to vary. A similar increasing pattern is observed looking from top-to-bottom, where more extreme flood scenarios increase the variance (and depths). While this is less intuitive, we hypothesize this result is caused by the ratio of small-steep river channels activated in the domain. The lower magnitude events are dominated by exposure within broad-flat floodplains while these events do not result in exposed building in the smaller rivers where terrain gradients are more severe. Regardless, Figure 2 demonstrates that most aggregate assets have a submeter standard deviation ($Q_{0.75}[\sigma] < 50$ cm).

Spatially, child depth variances are heterogeneous. Figure 3 shows a map of standard deviations for the most extreme 500-year hazard scenario. The regions of high variance correspond to river sections with development on both steep and flooded terrain. From this map, we conclude that sensitivity to Jensen's inequality can differ substantially between regions.

4.2 | Relative losses

With the pattern and magnitude of child depths demonstrated, we now feed these into the selected damage functions to compute non-aggregated and aggregated relative losses. By comparing these, the distribution of errors can be quantified as shown in Figure 4 for the most extreme hazard scenario where each function is overlaid on the mean loss computed from each child asset ($\overline{RL_{bldg,j}}$). In other words, Figure 4 shows the difference between what the aggregate model would calculate ($f(WSH)$) and what a model that included individual buildings would yield. Rather than focus on the absolute value of the computed errors, which are sensitive to the study domain, homogeneous building categorization, and application of the legacy functions at micro-scale, Figure 4 is used to understand error trends and behavior caused by function shape and grid size. Examining general trends shared by all loss functions by looking from left-to-right (Figure 4a–c) we see both the density plot and the WSH-binned mean line (blue dashes) show a widening gap between per-asset losses and aggregate losses as the grid sizes increases. This follows from Figure 2 which shows the same trend on variance. This difference has a similar pattern for all 12 scenarios: for aggregate depths >200 cm, the aggregate relative loss (solid black line) overestimates the per-asset loss (dashed blue line) substantially. This is also reflected in the differences between the population means which show the aggregate $\left(\overline{RL_{grid,j}}\right)$ overestimating the per-asset $\left(\overline{\overline{RL_{bldg,j}}}\right)$ losses for all scenarios (with a bias up to 1.19). The extremes are also noteworthy, where we see some aggregate losses are more than double their child-asset mean-loss counterpart.

Comparing across functions, Figure 4 shows those with a greater RL range (e.g., function IKSE's RL ranges from 0% to 40%) yield larger magnitudes of differences and root mean square errors (RMSE) between the aggregate and per-asset losses. To understand this, consider an envelope of potential error bounded by $f(x)$ and a line from $f(0)$ to $f(x^{max})$ (see Figures 1 and S1) where x^{max} is 1000 cm for our hazard scenarios. The greater the area of this envelope the more sensitive the function will be to variance in child depths values. This envelope can also be thought of as a measure of concavity, which yields a negative Jensen's gap (Denny, 2017). Because all selected functions are concave (downward curvature), the aggregate functions all overestimate, except for a small portion of the FLEMO function between 0 and 175 cm which is slightly convex (upward curvature). For another example of how curvature relates to this aggregation error, Figure S2 provides a similar analysis but with three idealized functions that differ only in their curvature parameter, showing the same dependence of error on curvature. Figure S4 shows there is some diversity of function shape within Gerl et al.'s (2016) database; however, most have similar concavity and therefore we expect similar overestimation patterns. In summary, Jensen's gap can be significant for a Germany-wide model under the most extreme scenarios; however, Figure 3 suggests there may be some local regions where the gap is more pronounced.

To demonstrate the significance of Jensen's gap for a local region with high child depth variance, the subdomain around Koblenz as shown in Figure 3 (red box) was selected for further analysis. This area is hydraulically complex, encompassing the confluence of the Rhine, Moselle, and Lahn rivers, all confined to relatively narrow valleys with dense development along the banks. The mean child depths variance for this region is 1.5–4.0 times higher than the Germany-wide averages (see Figure S3), suggesting Jensen's gap may be more severe

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

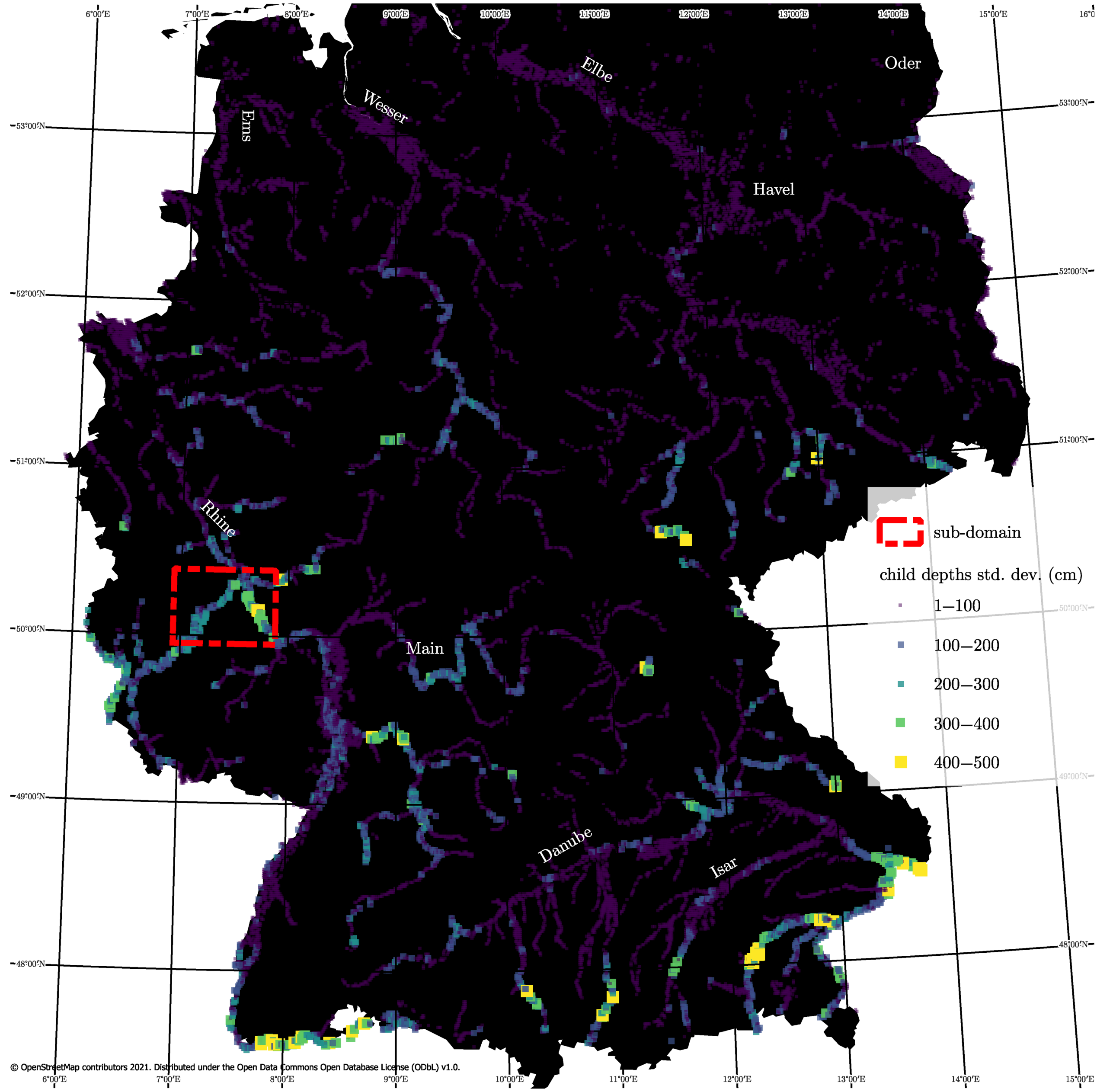

FIGURE 3 Map of major rivers in Germany and standard deviations of child depths for the 500-year hazard scenario with an aggregation grid size of 1020 m. Only those aggregate assets with at least two child assets are shown. See text for a discussion of sub-domain.

here. To illustrate this, Figure 5 shows the aggregated and non-aggregated relative losses for a single example function. Compared with the Germany-wide values, RMSE more than doubles and the bias more than triples (relative to perfect).

By implementing a Germany-wide approach, we trade some marginal uncertainty in the absolute RL values of each scenario for breadth and consistency across functions. The two most significant marginal uncertainty sources being the application of residential functions to all buildings and the application of per-building centroid depths (WSH_i) to all functions. The first assumes that the spatial distribution of residential and nonresidential buildings is roughly equivalent, which likely is untrue in some regions. The second assumes WSH_i is uniform across the footprint of the function's spatial design basis, that is, for the per-building modern functions WSH_i is assumed uniform across each building footprint and across each land use block for the meso-scale legacy functions. Considering the 30 m resolution hazard layers, uniform WSH_i across land use blocks, which are generally 100 m square, is untrue in many regions as shown in Figure 2. For the two mesoscale legacy functions, this means the RL errors reported in

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

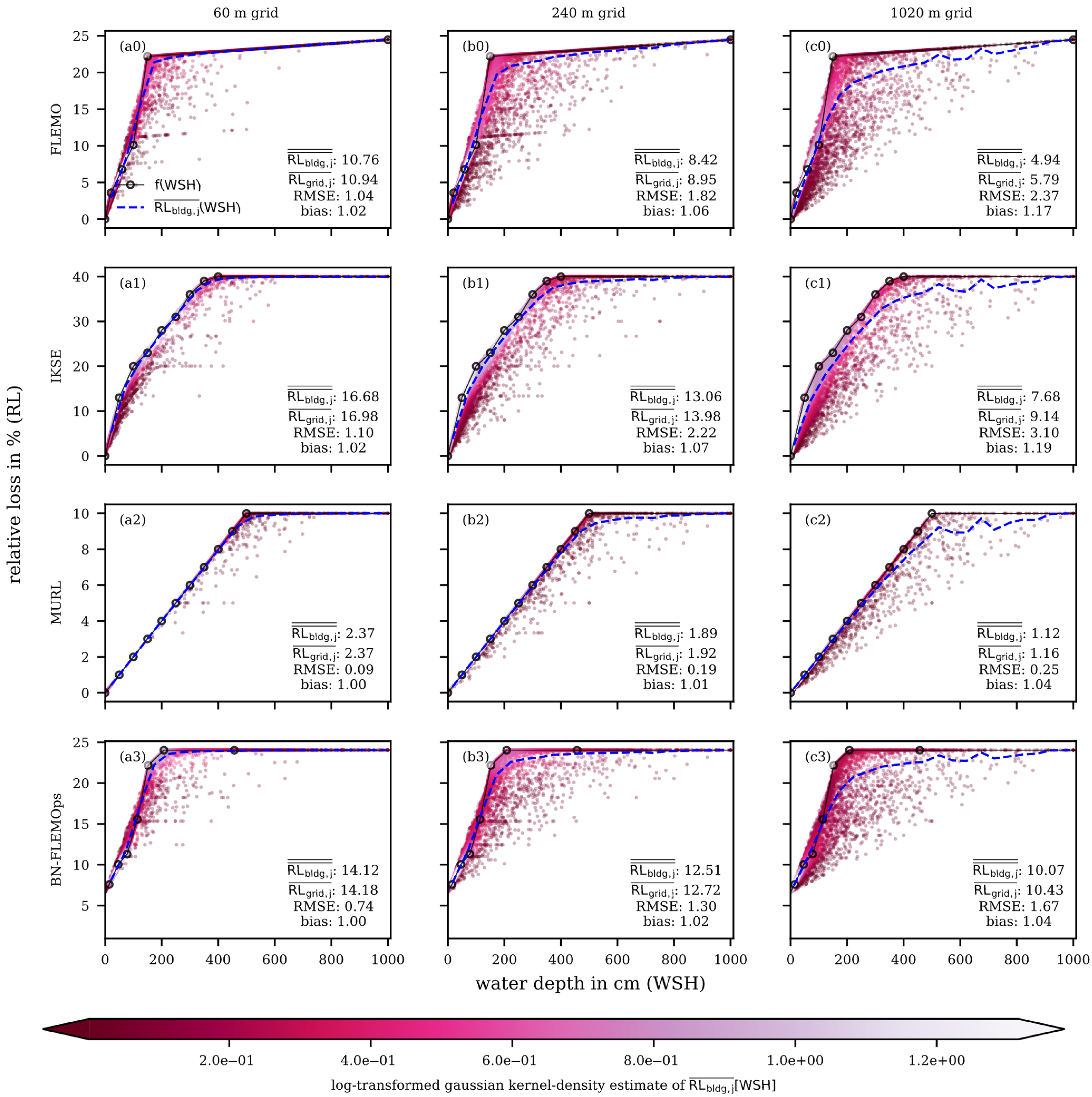

FIGURE 4 Relative losses for three grid aggregations for the 500-year fluvial undefended hazard scenario and the four selected flood damage functions (Table 1). Child relative loss means (mean of relative loss of each building within a grid cell) $(\overline{RL_{\text{bldg},j}})$ versus WSH_j are shown as a density scatter plot. The blue-dashed line shows the same $\overline{RL_{\text{bldg},j}}$ values for each aggregate asset with the mean taken again by binning on WSH values. The black-dashed line shows the original damage function. On the right-hand-side of these panels the population mean of losses computed from aggregate $(\overline{RL_{\text{grid},j}})$ and child relative loss mean $\overline{\overline{RL_{\text{bldg},j}}}$ are shown. Bias is computed as $\overline{\overline{RL_{\text{bldg},j}}} / \overline{RL_{\text{grid},j}}$. Only those aggregate assets with at least two child buildings are included.

Figure 4 (Rows 1 and 2) have an additional application uncertainty roughly equivalent to the 60 m computation (Figure 4a1,a2). However, considering the magnitude of this uncertainty is an order of magnitude less than the errors computed for these damage functions by Thieken et al. (2008), this application uncertainty is inconsequential to our analysis of aggregation behavior and an acceptable tradeoff for allowing the inclusion of these well-known damage functions in the analysis. Finally, readers should note that even for the extreme cases computed here, the relative loss errors from aggregation are less than those from the initial construction or fitting of the empirical functions themselves. For example, BN-FLEMOps has a mean absolute error around 15% RL (Wagenaar et al., 2018) and FLEMOps 24% RL (Thieken et al., 2008) when comparing modeled to observed losses.

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

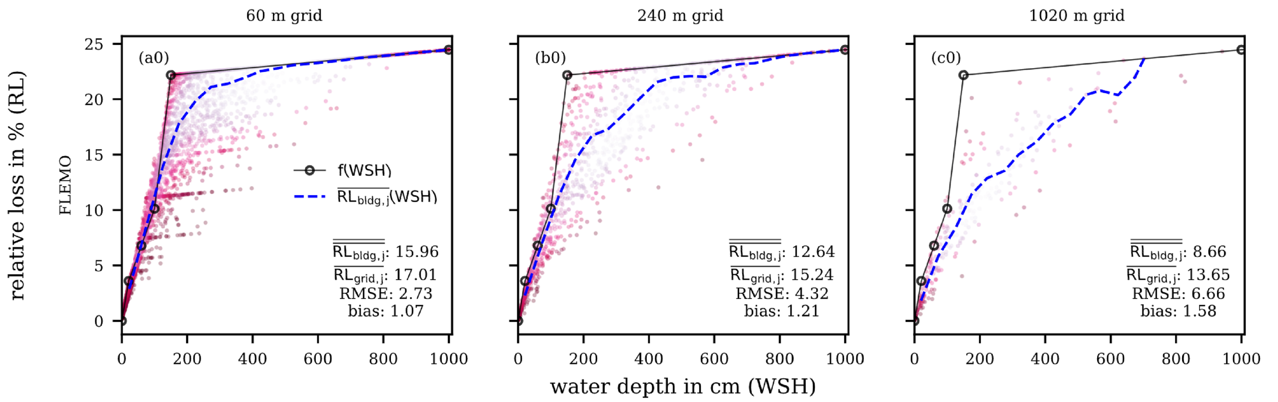

FIGURE 5 Relative losses for a single damage function for the sub-domain shown in Figure 3. See Figure 4 for additional explanation and legend.

4.3 | Total losses

The above analysis, while useful in demonstrating and understanding the mechanics of Jensen's inequality as it applies to flood damage functions, has limited applicability to damage models due to the lack of replacement values (V_j; see Equation 2). For example, densely developed regions (e.g., cities) may have different exposure or hazard patterns than less developed regions (e.g., rural areas). These density-dependent effects are obfuscated when the aggregate asset RL is simply compared to the child RL mean (i.e., all aggregate assets are treated equally regardless of their child building counts) as in the previous section. As a proxy to quantify aggregation errors with density dependent effects, we extend the above analysis by including all assets and multiplying the relative losses calculated for the aggregate asset by its building count (B_j) as shown in Figure 6.

Looking from left to right again, Figure 6 shows the same increase in bias and RMSE with grid size as for the relative loss values in Figure 4, as well as the same relative magnitudes between functions. However, the total bias increases for 6 of the 12 cases with two cases decreasing slightly compared to the relative loss values. This suggests some, generally intensifying, density effects on Jensen's gap in flood damage models. In other words, aggregate assets with more children (e.g., cities) are more sensitive to Jensen's gap than those with fewer children (e.g., rural areas).

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

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

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

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

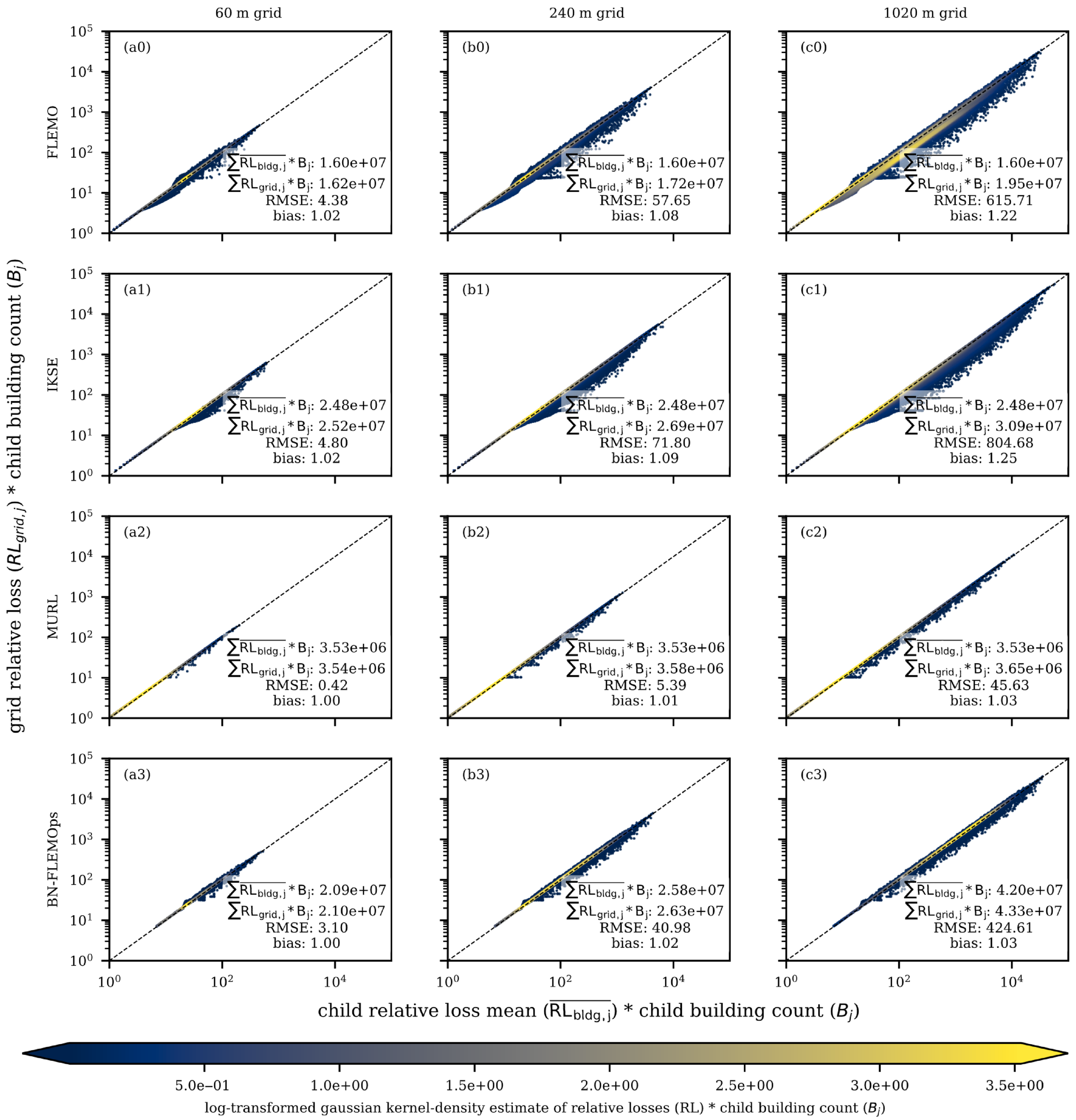

FIGURE 6 Building count weighted relative losses for three grid aggregations and four flood damage functions for the 500-year fluvial undefended hazard scenario. Totals are shown for losses computed with non-aggregated ($\overline{RL_{bldg,j}}$; x-axis) and aggregated ($RL_{grid,j}$; y-axis) depths. Bias as in Figure 4.

than one using child mean depths like ours. Errors like this, and the non-linear function aggregation reported here, compound and interact with more classical model uncertainties like spatiotemporal transfers and measurement error.

5 | CONCLUSION

This study provides the first demonstration and quantification of errors in aggregate flood damage models resulting from the application of nonlinear damage functions. We show that variance in flood depths, along with function curvature, leads to discrepancies between the aggregated and the underlying non-aggregated model. With our Germany-wide analysis, we find that flood depth variance increases with the magnitude of both aggregation and flood hazard. Further, this variance is spatially heterogeneous, with some local regions exhibiting 1.5–4.0 times more variance in flood depth than the Germany-wide averages. Computing the error in relative loss estimates resulting from these variances, we find all

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

selected damage functions overestimate when applied in aggregation. The magnitude of this overestimation is highly dependent on the water depth variances and the amount of curvature, varying from a bias (aggregate result/non-aggregate) slightly above 1.00–1.19 when averaged Germany-wide. However, these errors arising from aggregation are generally less significant than the uncertainty within the damage function itself. Finally, we extend the analysis to account for heterogeneous asset density. This shows the same pattern of error but with most cases yielding a larger overestimate when asset density is included (bias up to 1.25).

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

To provide a more consistent evaluation of Jensen's gap, our study only quantified aggregate model overestimates for residential tangible, direct-damage functions applied to buildings exposed to fluvial hazards in Germany. However, considering the magnitude of the overestimate in this case, and that even higher overestimates are probable in some situations, suggests all modelers employing aggregation should be knowledgeable of the mechanics of Jensen's gap. For example, modelers should be aware that increasing curvature and aggregate depth variance will increase errors. Similarly, this study demonstrates one mechanism that leads to superior accuracy of non-aggregate models over aggregate models. Of course, this mechanism is only one-way aggregation introduces error, and these aggregation errors are only one of many sources of error present in flood damage models. To better prepare society for flood disasters, more attention and effort are needed to understand and mitigate model errors and biases like the ones reported here.

AUTHOR CONTRIBUTIONS

Conceptualization: S.B. *Data curation*: S.B. *Formal analysis*: S.B. *Methodology*: S.B. and J.R. *Software*: S.B. *Validation*: S.B. *Funding acquisition*: B.M and H.K. Writing—original draft: S.B. *Writing—review and editing*: All authors. *Visualization*: S.B. *Supervision*: B.M. and H.K.

ACKNOWLEDGMENTS

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

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Building data were obtained from OSM (OpenStreetMap contributors, 2015). Flood damage functions were obtained from Gerl et al. (2016). Flood hazard data were provided by SSBN UK LIMITED (a.k.a. Fathom) under academic license and is therefore not shareable by the authors.

ORCID

Seth Bryant https://orcid.org/0000-0002-8799-3919

REFERENCES

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

Bryant, S. (2023). *Flood Hazard Grid Aggregation Scripts [software]*. Zenodo. https://doi.org/10.5281/zenodo.8271996

Bryant, S., Kreibich, H., & Merz, B. (2023). Bias in flood Hazard grid aggregation. *Water Resources Research*, *59*(9), e2023WR035100. https://doi.org/10.1029/2023WR035100

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

Denny, M. (2017). The fallacy of the average: On the ubiquity, utility and continuing novelty of Jensen's inequality. *Journal of Experimental Biology*, *220*(2), 139–146. https://doi.org/10.1242/jeb.140368

Fathom. (2023). *Fathom Global Flood Map 3 Specification* (Technical Report).

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

Gerl, T., Kreibich, H., Franco, G., Marechal, D., & Schrter, K. (2016). A review of flood loss models as basis for harmonization and benchmarking. *PLoS One*, *11*(7), e0159791.

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

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Internationale Kommission zum Schutz der Elbe. (2003). *Aktionsplan Hochwasserschutz Elbe* (Technical Report). IKSE.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les ingalits entre les valeurs moyennes. *Acta Mathematica*, *30*(1), 175–193.

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

Liao, J., & Berg, A. (2018). Sharpening Jensen's inequality. *The American Statistician*, *73*, 278–281.

Lindsay, J. (2014). The whitebox geospatial analysis tools project and open-access GIS. In *Proceedings of the GIS research UK 22nd annual conference* (pp. 16–18). *The University of Glasgow*. Retrieved from April 11, 2022. https://jblindsay.github.io/ghrg/pubs/LindsayGISRUK2014.pdf

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.ijdrr.2015.09.003

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., Gallazzi, A., Galliani, M., Grelot, F., Kellermann, P., Kreibich, H., Mohor, G. S., Mosimann, M., Natho, S., Richert, C., Schroeter, K., Thieken, A. H., Zischg, A. P., & Ballio, F. (2020). Are flood damage models converging to reality? Lessons learnt from a blind test. *Natural Hazards and Earth System Sciences*, *20*, 2997–3017. https://doi.org/10.5194/nhess-2020-40

Nordrhein-Westfalen, M. f. U. R. u. L. d. L. (2000). *Potentielle Hochwasserschden am Rhein in NRW* (Techinncal Report): MURL Report Dsseldorf.

OpenStreetMap Contributors. (2015). *Planet dump*. Retrieved form July 25, 2023. https://download.geofabrik.de/. https://www.openstreetmap.org

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*, e12851. https://doi.org/10.1111/jfr3.12851

PostGIS Project Steering Committee. (2018). *PostGIS*. Retrieved from December 23, 2024. https://postgis.net

Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V. D., Lüdtke, S., Merz, B., Schröter, K., Vorogushyn, S., & Kreibich, H. (2021). Process-based flood risk assessment for Germany. *Earth's Future*, *9*(10), e2021EF002259. https://doi.org/10.1029/2021EF002259

Seifert, I., Thieken, 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., 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

Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., & Merz, B. (2008, June). Development and evaluation of FLEMOps – A new flood loss estimation MOdel for the private sector. In *Flood recovery, innovation and response I* (Vol. I, pp. 315–324). WIT Press. Retrieved from July 13, 2022. http://library.witpress.com/viewpaper.asp?pcode=FRIAR08-030-1

Wagenaar, D., Ldtke, S., Schrter, K., Bouwer, L. M., & Kreibich, H. (2018). *Regional and temporal transferability of multi-variable flood damage models*. Water Resources Research. https://doi.org/10.1029/2017WR022233

Walker, S. G. (2014). On a lower bound for the Jensen inequality. *SIAM Journal on Mathematical Analysis*, *46*(5), 3151–3157.

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G. A., Schäfer, A., Stanley, T., Veldkamp, T. I. E., Winsemius, H. C., & Winsemius, H. C. (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

White, G. F. (1945). Human adjustment to floods. A geographical approach to the flood problem in the United States [Unpublished doctoral dissertation]. The University of Chicago, Chicago.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bryant, S., Reimer, J., Kreibich, H., & Merz, B. (2025). Aggregating flood damage functions: The peril of Jensen's gap. *Journal of Flood Risk Management*, *18*(1), e13053. https://doi.org/10.1111/jfr3.13053

1753318x, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jfr3.13053 by University Of Utah, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License