New Empirical Models for Flood Loss Prediction and Implications for the Coterminous United States

Adam Pollack (adam.b.pollack@dartmouth.edu)
Dartmouth College

Ian Sue Wing
Boston University

Nicholas Pinter
University of California, Davis

Christoph Nolte
Boston University  https://orcid.org/0000-0001-7827-689X

Article

Keywords:

Posted Date: October 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2062647/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Title
New Empirical Models for Flood Loss Prediction and Implications for the Coterminous United States

Abstract
Flood-loss estimates are needed for floodplain development and mitigation projects, for setting fair insurance rates, and for guiding climate adaptation policy. Currently, flood-loss models, including depth-damage functions (DDFs) widely used in the U.S., lack empirical validation commensurate with the geographic extent and diversity of structures and flood exposure over which these predictions are needed. Using data from 845,776 U.S. National Flood Insurance Program claims, we validate DDFs and create alternative models grounded in empirical data and validation. These alternative models more accurately predict average observed damages for many types of structures and hazard compared to current DDFs which omit important variables and interactions that drive observed losses. We find that a major bottleneck in flood-loss estimation is the development and validation of flood-loss models for both damaged and undamaged homes, a gap FEMA could help close.
Introduction

Flood risk management relies on estimates of economic losses from flooding, including for a wide range of event types and property characteristics. Obstacles to flood-damage estimation include limited high-resolution data on the characteristics of flood hazards over broad geographic extents, and the complex interplay between these physical flood processes and their impacts upon human infrastructure.

Research has identified several priorities for flood-damage relationships. First is accuracy: after a new flood event, measured damages to structures should reasonably match estimates made before an event. Second is transferability: due to data limitations, damage estimates are often required for settings different from where the damage relationships are derived, calling for extrapolation. A third is parsimony: input data demands should be light enough to facilitate application in a variety of settings and under common data constraints. In addition to the criteria above, we propose interpretability as an additional priority for damage relationships. Since validation data are limited relative to extrapolation applications, such as property-level flood risk estimates for all buildings in the U.S., damage relationships should provide insight into the limitations of predictions. Accuracy can degrade due to differences in validation and extrapolation data that are hard to control for in damage relationship specifications. In settings where prediction occurs under different conditions from the data used in development and validation, explainable models with transparent limitations are often more desirable.

In practice, no damage relationship satisfies all four criteria above. In the U.S., depth-damage functions (DDFs) are widely used for predicting flood damages, for policy analysis, and to inform insurance rates and flood mitigation expenditures. Current DDFs used in the U.S. were created largely from expert judgement and relate the characteristics of structures, often the number of stories and presence of a basement, and inundation depth. Thus, DDFs are parsimonious and interpretable because data on structures’ basic features tend to be available and inundation depths are known or can be modeled. However, the accuracy of DDFs has been questioned: for instance, flood damages have been shown to vary substantially with structural characteristics not accounted for by DDFs.

In contrast to widely used DDFs, damage relationships empirically derived from observed losses, flooding attributes and characteristics of affected structures may capture more of the heterogeneity in losses. Many of these studies rely on sophisticated models such as Random Forests and often include many input variables. However, these models tend to be trained in local contexts that exhibit limited variation in types of floods and damages with respect to expected relevant variables relative to flood types and variation expected in broader geographies. As a result, many existing empirical damage models are neither parsimonious, interpretable, nor can they be transferable to other contexts. Moreover, empirical damage models are not demonstrably more accuracy than current DDFs. Despite their similar performance, predictions from DDFs and empirical alternatives can differ substantially.
There are three key information gaps regarding damage models which we address here. First, comparisons are needed of the predictive ability of DDFs and empirically derived alternatives. Since these are lacking, there is little understanding of the accuracy-parsimony-interpretability tradeoffs between DDFs and current empirical alternatives. Second, the largest previous validation study utilized 7,000 observations from 11 events (Table S2), which is hardly representative of the broad geographic domains for which flood loss estimates are needed. Third, existing assessments of empirical damage models are likely to lead to overestimates of predictive accuracy in extrapolation, as they often employ cross-validation approaches with observations from the same event in both training and validation (Tables S1, S2). Relatedly DDFs are rarely validated against post-flood observations of losses.

In this study, we compare the predictive accuracy and transferability of damage models that differ in parsimony and interpretability. We do so by developing new empirical damage models for the coterminous U.S. (CONUS) and comparing them against current DDFs. Our model training and validation relies on 845,776 insurance claims to the National Flood Insurance Program (NFIP) that were caused by 446 flood events between 1972 and 2015. To rigorously assess the many factors that shape flood losses, we evaluate models for their ability to accurately predict average expected losses across different flood types (depth, intensity, type) and different structures (value, characteristics) using leave-one-event-out cross-validation.

Finally, we apply current DDFs and our preferred alternative model to 72.4 million single-family homes using property-level modeled flood depths. Our goals are (1) to identify differences between prediction accuracy of various models, (2) assess implications of differences in damage relationships and uncertainties of applying them in extrapolation, and (3) determine priorities for research on loss estimation.

**Results**

_Flood losses are long-tailed, uncertain, and associated with a range of attributes_

From a dataset of 2,085,015 claims for damages to main and appurtenant structures of many building types, we extracted 1,099,372 complete records of damages to the main structures of single-family homes (Table S3). Additional filters for compatibility with our large-scale flood risk estimation system and data quality (following a previous analysis of this data) result in a final sample of 845,776 observations. These span 446 unique events spanning 2,581 (85%) U.S. counties, and a wide range of flood types, characteristics of affected structures, and geographic contexts (Fig. 1, Table S1). Further, this data encompasses 120 times more loss observations and 40 times more flood events than any previous study to develop and validate damage models (Table S2).

The primary outcome of flood-loss estimation is percent damage – i.e., damage as a percentage of the value of affected structures. Percent damages are positively skewed with a mean of 23% and standard deviation of 27%. This skew persists across many dimensions, including structure value, basement presence, and number of stories (Fig. 1). However, distributions differ markedly depending on whether flooding is driven by storm or other events (See Methods). A substantial portion of total observations come from two storms, hurricanes Katrina (10%) and Sandy (9%). Damages from hurricane Katrina exhibit a near-uniform distribution for most depth exposures (<4ft.) but are nearly exclusively very high losses (≥90%) at the highest depth exposures (≥4ft.).
Hurricane Katrina disproportionately accounts for high-depth exposures (29% of observations ≥4ft.) and severe losses (63% of observations ≥90% damage), documenting that the presence of levees can reduce the impact of low-magnitude flooding but increase the risk of catastrophic loss, sometimes called the “levee effect”. In contrast, across all other event types, most observations occur at flood depths <4ft. and predominantly incur lower damages.

As flood depth increases, mean percent damage and variability generally increase, conditional on several hazard and structure characteristics (Fig. 2). Some dimensions, such as the age of the structure, make little difference to the shape and spread of this notional depth-damage curve, while others, such as the presence of storms do explain some of the heterogeneity. Nevertheless, the variability and heteroskedasticity of damages persists across all subgroups. Counterintuitively, homes without basements and homes on the coast experience less percent damage at higher depths. This observation suggests that univariate splits of loss data can mask the importance of other influential, correlated variables.

The “levee effect” illustrates one such influential and correlated variable often unaccounted for in damage function validation or loss estimation (Fig. 2). The distribution of percent damages across inundation depths for the Greater New Orleans area reveals the protective role of levees: for non-Katrina events, losses are less common and less severe (Fig. 3A-B). In contrast, levee failures during Katrina resulted in a distribution of flood depth exposures and damages otherwise unseen for these structures. While the storm also caused more severe losses at depths ≥4ft. outside of New Orleans than other storms, the percent damage distributions at lower depths aren’t remarkably different elsewhere in Louisiana or other states in the same FEMA region (Fig. S1).

Another important correlated variable is structure value, as previously observed. At similar flood depths, high-value structures (top quintile of deflated structure values in each state) incur lower percent damage than low-value structures. One plausible explanation for this phenomenon is higher construction quality and more investments in structural defense (e.g., reinforced concrete exterior walls). If true, this would lead to lower relative damages for high-value homes. Consistent with this explanation, we find that both the mean and variability of percent damage are inversely proportional to structure value (Fig. 3C-D). However, structure value is correlated with other variables of interest, such as basement presence and coastal exposure, in ways that may explain seemingly counterintuitive relationships between these characteristics and percent damage. For instance, homes with finished basements or coastal exposure tend to be higher in value (i.e. >=$300K). This relationship would reduce percent damage values for homes with basements and coastal exposure. Further, since the value of basements and damage to basements may not be proportional to the overall value of the structure, homes with basements may experience less relative damage (Fig. S2). In sum, the assumption that percent damage scales linearly with structure value, as implicit in current DDFs, can introduce bidirectional biases as a function of structural characteristics.

Current DDFs omit key variables and interactions

We find that, on average, current DDFs fail to accurately capture observed percent damages across different depths (Fig. 4). First, because most DDFs assign zero (or low) percent damage to flooding below first-floorelevation, which results in basement or foundation damage, current
DDFs tend to underestimate observed losses at these depths (32% of observations). This is also true for regional DDFs recently designed for the Northeast U.S.\(^6\). Even though they include non-zero percent damages at flood depths well below first-floor elevation, they tend to underestimate observed losses that occur in the Northeast U.S. (Fig. S3). This tendency may be affected by the fact that observed losses in our dataset are derived from NFIP claims. Homes that experience low damages, common for lower flood depths, may not file claims, and true losses may be greater than claims data suggest. In contrast, DDFs overestimate losses occurring above first-floor elevation (positive depths) for all flood events other than Katrina. This is also true with DDFs designed specifically for loss prediction in the corresponding FEMA region 6 (Fig. S4). Second, because DDFs ignore structure value and basement presence (Fig. 3), they tend to overestimate damages to homes with basements (Fig. 4A,C).

Overall, current DDFs have a mean bias error (MBE) of 0.005, an explained variance ($R^2$) of 0.062, and a root-mean-squared-error (RMSE) of 0.27, indicating strong mean prediction accuracy but poor ability to explain variation in damages. Low overall mean-bias can be explained by the high density of observations that occur just above first-floor elevation (e.g. depths of 1-2 ft.), which DDFs generally predict accurately. For properties exposed to inundations at or above first-floor elevation (e.g. 0-8 ft.), MBE is 0.008 and $R^2$ is 0.056. In contrast, DDFs predict damages at inundations below first-floor elevation with a MBE of -0.06 and $R^2$ of -0.35. For homes with basements, DDFs have a larger MBE (0.13), and their predictions perform far worse than expected on average ($R^2 = -0.67$). In contrast, for homes without basements MBE is -0.04 and $R^2$ is 0.16. These differences suggest that instead of a single metric to quantify prediction accuracy, accuracy metrics are needed that span various subsets of the data. Based on these observations of DDF inaccuracies we conclude that predictive accuracy should be quantified, at minimum, for subgroups with different inundation depths, structure values, structure types, and event types. This approach also has the benefit of allowing analysts to identify which results may be sensitive to potential selection bias.

**Empirical alternatives improve predictive accuracy only slightly over DDFs**

To which extent can the observed imprecisions and subgroup-specific biases of current DDF-based flood loss prediction be reduced by deriving empirical damage models from the observed loss data? We implemented a nationwide cross-validation analysis to compare the predictive performance of different statistical models in predicting depth-damage relationships. We developed three models that differ in their functional flexibility and in the range of predictor variables used. These are described in detail in the Methods section, but are summarized here:

1. **Depth GAM**: A general additive model (GAM) that predicts percent damage from inundation depth (with respect to first-floor elevation) for each FEMA region. The GAM includes a shape constraint that enforces a monotonically increasing relationship between inundation depth and percent damage. This model leads to damage models that are as parsimonious and interpretable as DDFs.

2. **Stratified GAM**: Building on the Depth GAM, we fit shape-constrained depth-damage functions separately for each relevant subgroup. We split by (i) FEMA region; (ii) whether affected structures have assessed value within the top 80th percentile or not, (iii) a basement vs. no basement, and (iv) one vs. more stories; as well as (v) whether the flood is due to a storm event. This model is less parsimonious than Depth GAMs but is as
interpretable as DDFs.

3. **Extra Trees**: We use tree-based ensemble models (Extremely Randomized Trees\textsuperscript{53}) to regress observed percent damage on the same set of predictors used in the Stratified GAM as well as the age of the structure, whether a home is post-FIRM, and indicators for whether the property is in a zip code that’s adjacent to a river, lake, or coastline. Because Extra Tree models allow for flexible functional forms and high-dimensional interactions, we expect them to do particularly well at prediction. The drawback is that they require more input data, are cumbersome to apply, and have low interpretability (are a “black box”).

We compare all three empirical models in a rigorous leave-one-event-out (LOO) cross-validation framework: for each event (n=446), we develop predictions of expected percent damage based on models trained on observed loss data from all other events. We then quantify prediction error as the difference between predicted and observed percent damage across all loss observations (and, thus, all events). Because of the high leverage of Katrina observations on development and validation, especially at high depths, we first present results of all FEMA regions excluding Katrina in development and validation, and then results for FEMA regions 4 and 6 affected by Katrina. For the latter, we add a correction strategy that controls for the observed “levee effect”, i.e., the influence of high-depth and high-damage observations (See Methods).

The new empirical models developed here capture average losses across all events besides Katrina (Fig. 5). While current DDFs overestimate losses on average for non-storm events (MBE=0.06), ≥4ft. storm-event inundations (MBE=0.21), structure values ≥$750K (MBE=0.15) and one-story homes with basements (MBE=0.16), even simple Depth GAMs improve accuracy for these subsets. MBE generally improves from Depth GAMs to Stratified GAMs or Extra Trees for key subgroups. For instance, for structure values ≥$750K, MBEs are 0.114, 0.037, and 0.008 respectively. On average, Extra Trees and Stratified GAMs have similar prediction performance which suggests a more flexible black-box approach doesn’t always lead to better performance than the more interpretable approach. For example, for high-depth inundations in storms, we obtain a MBE of -0.047 for Extra Trees, -0.023 for Stratified GAMs, and -0.037 for Depth GAMs.

Despite these improvements in average prediction accuracy, predicted flood losses still exhibit high RMSE (i.e., low precision) across all our empirical models. This is especially true at ≥4ft. inundations, where for non-storm (storm) losses, Extra Trees have a RMSE of 0.23 (0.27), Stratified GAMs have a RMSE of 0.26 (0.27), and DDFs have a RMSE of 0.28 (0.29). Similarly, in many data subsets, RMSE suggests no advantage over raw variability in observed losses. For example, although RMSE for Extra Trees and Stratified GAMs appear relatively low at 0.12 each for structures values ≥$750K, the standard deviation of percent damage for these homes is only 0.13.

In FEMA regions 4 and 6, Hurricane Katrina strongly influences model development and validation. Fig. 6A shows that because Katrina composes the vast majority of ≥4ft. inundations, and because these are more damaging than losses from other events (Fig. 2-3), average prediction accuracy of empirical models drops at ≥4ft. inundations regardless of model complexity or sophistication. When Katrina is included in model development but removed from
validation, there is less damage variability in the validation set, and the variability of predictions is greatly reduced. This is because training empirical models on losses from Katrina result in overpredictions of damages at \( \geq 4 \)ft. depths for observations from all other events. Interestingly, although existing DDFs aren’t trained in this fashion, they reveal a large overestimation of \( \geq 4 \)ft. depth losses for observations from events besides Katrina (MBE=0.178). This suggests that their predictions might be biased towards extreme events, like Katrina, at the expense of losses from other events. Adjusting for the levee-effect in the training of Stratified GAMs improves MBE from 0.212 to 0.055 without notably affecting prediction accuracy for other subsets. However, it brings implementation challenges in extrapolation (See Discussion, SI).

Model-based predictions of U.S.-wide flood damages for single-family homes

How do the observed differences in the performance of current DDFs vs. empirical alternatives affect the estimation of U.S.-wide flood risk? To explore this question, we conduct the following analyses. First, we compare observed depths in our sample to modeled depths across 72.4 million single-family homes in an extrapolation analysis (See Methods). The goals of this analysis were to better understand potential selection bias in our data, and to examine how the choice of damage models affects U.S. flood loss estimates. Second, to assess the impact of damage model choice, we scale the large-scale flood loss estimation framework developed by Pollack et al. (2022), which links property level modeled flood hazard, structure locations and characteristics, and other information relevant to loss estimation for use across CONUS (See Methods).

Results in this section are based on comparisons of flood loss estimates derived from the USACE DDFs and Stratified GAMs trained on all data. Stratified GAMs are more accurate in validation than Depth GAMs and are more interpretable and parsimonious than Extra Trees which have similar accuracy. For consistency in comparisons to validation, losses in extrapolation refers to losses within individual return periods weighted by their probability of occurring, analogous to events in the validation data.

Predicting flood losses for CONUS single-family homes requires strong assumptions to address data limitations such as unobserved first-floor elevations and incomplete data on the presence of basements. This includes assumptions on 1) how to translate modeled inundation depths relative to grade (terrain), available for all properties, into depths relative to first-floor elevation, and 2) the probability that a given inundation level leads to damages. Specifically, DDFs and Stratified GAMs are designed and validated for use in settings where damage is known to occur. While this is known for observed losses, NFIP claims data lack precise geolocation. Therefore, we’re unable to link observed losses, and the depths that generated them, to property level flood hazard exposure in our database. Without conditional probabilities of damage occurring for different flood hazard exposures, extrapolation may overestimate the cases in which damage occurs. We assess our extrapolation results in light of this concern, but we apply damage functions to all properties with non-zero flood hazard exposure like other papers that have estimated flood losses over large scales. To explore the sensitivity of aggregate loss estimates to different first-floor elevation adjustments and scenarios where damages occur at low depths, we develop three sets of assumptions that reflect low, benchmark (reference), and high-damage assumptions (see Methods).

The extrapolation analysis reveals striking differences between the distributions of observed inundation depths in the claims data and inundation depths from modeled depths for CONUS
single-family homes (Fig. 7). While observed losses occur three-times more often in the 100-yr FEMA flood zone (SFHA) than outside, predictions of flooding from modeled depths for CONUS single-family homes imply the opposite. This finding suggests that there is a considerable degree of selection bias in observed losses: homes in the SFHA are more likely to hold insurance policies and file claims, whereas U.S. single-family homes may face significant flood hazard outside the SFHA\textsuperscript{56,57}. In addition, modeled inundation depths at or below first floor are estimated to occur much more frequently across CONUS single-family homes than in the claims data. This may reflect another type of selection bias: homes that experience losses from lower flood depths (e.g., from pluvial flooding) might be less likely to be mapped into the SFHA, less likely to hold insurance policies, and less likely to go through the claims process. Another potential explanation for the discrepancy is that the flood hazard model we use may flooding\textsuperscript{16}.

Differences in estimated losses are heavily influenced by the differences in observed and modeled depths discussed above. As current DDFs tend to underestimate observed low-depth losses and overestimate observed high-depth losses relative to those implied by Stratified GAMs, the observed differences in predictions at low-depth losses become more prominent in extrapolation (Fig. 7). For CONUS single-family homes, DDF-based loss predictions in validation are 26% larger than those derived from Stratified GAMs but are only 4.5% larger in extrapolation (for properties where basement type is known). This is consistent in the high damage scenario, but in the low damage scenario where losses can be incurred at even lower depths (See Methods), Stratified GAM predicted losses are slightly larger than DDFs in extrapolation (Fig. S5).

Because the distributions of depth, value and structure characteristics are spatially heterogeneous and correlated (Fig. S6-S7), estimates of expected annual loss (EAL) derived from DDFs and Stratified GAMs diverge across key subgroups of data (Fig. 8) and these differences are asymmetrically distributed at the census tract level (Fig. 9). We estimate $10.2 billion in EAL for 9.1 million homes with Stratified GAMs. In contrast, DDFs estimate EAL of $11.4 billion (+10.5%) from 8.6M homes that incur flood damages (-4.6%). Discrepancies are mostly driven by higher value homes and homes exposed to higher flood depths (Fig. 8), which may be less vulnerable to selection bias issues than findings at low flood depths. Relative to DDFs, Stratified GAMs estimate higher losses in some Northern and mid-Atlantic coastal areas, in the Upper Midwest, and more moderately along the Gulf Coast, with a few scattered locations in the Pacific Northwest and off the Great Lakes (Fig. 8). Another key driver of discrepancies is that the large number of properties which face low flood depths can account for a large portion of estimated overall EAL. Since DDFs estimate very low losses for these homes, Stratified GAMs estimate a relatively larger overall EAL in areas where low depth floods are prevalent (Fig. 8, Fig. S8-S9). These findings are robust across various damage functions and scenarios, such as adjusting for the influence of Katrina on the shape of damage functions (Fig. S10-S13). However, these results are clearly sensitive to the depth discrepancies between observed losses and modeled exposure likely driven by selection bias. Because we expect this bias to exert upwards bias on Stratified GAMs at low depths, the consequence in our extrapolation representation is that we underestimate the overestimation of DDFs relative to Stratified GAMs.

**Discussion**

We show that observed depth-damage relationships are shaped by non-depth attributes of flood...
events and structure relationships that tend to be ignored by current DDFs, which are the prevailing hazard-damage model in the U.S. As a result, current DDFs tend to overestimate damages incurred by homes that are high-value, have basements, or are exposed to deep inundation. Current DDFs may also underestimate damage resulting from inundation below first floor elevation. We also show that a large share (≥87%) of the variance in observed losses cannot be captured even with a comparatively rich dataset of flood attributes and structure characteristics and using highly flexible functional forms (Extra Trees). This suggests that missing data on flood and property attributes remains a key obstacle to further improvement in characterizing the hazard-damage relationships and underscores the difficulty of establishing actuarially fair insurance premiums.

Our findings illustrate the limitations of a “one-size fits all” approach that represents a highly heterogeneous phenomenon (flood damage) using univariate depth-damage curves. To minimize biases in flood loss estimation, analysts will need to better understand how damage models can be calibrated to different subsets of structure and hazard characteristics. Our results suggest that the inclusion of a few key correlates of percent damage, such as structure value and the possibility of catastrophic loss, can improve the predictive accuracy of existing depth-damage functions. More high-dimensional classifications are theoretically possible but will require more data on the attributes of affected structures, as well as non-depth characteristics of the floods affecting them over broad spatial scales.

A key caveat of our analysis is that we find evidence of selection bias that threatens the transferability of our validation results for CONUS-wide loss predictions. An interesting implication of the selection bias discussed here is that the estimated gap between expected losses from Stratified GAMs and DDFs may be larger than the one reported here. An important research goal is, therefore, to better understand what conditions lead to an occurrence of loss and its quantification by assessors, and then develop damage functions that properly account for this selection bias. One county-level study shows that knowledge of conditional probabilities for damage occurrence can support the development of empirical models that more accurately predict losses of events in an extrapolation setting\textsuperscript{20}. While observations of losses and policies made available by FEMA are not linked to property locations, geolocated data could be used internally to close important information gaps. This could enable comparisons of modeled inundations of historic events to their observed extents and damages, provide a better understanding of selection into observed losses and agreement of observed and modeled depths, and produce reliable empirical estimates from conditional probabilities derived from these comparisons. Other important future research directions for improving the accuracy and usefulness of loss estimation include 1) more transparency on the damage functions, probability of damage occurring at various inundation depths and structure characteristics employed and how they are linked to each other in loss estimation; and 2) assessing outcomes with respect to different validated damage functions and incorporating techniques to isolate the role of broadly influential loss variables such as catastrophic events.

High-resolution hazard models made available over large spatial domains have heralded a plethora of large-scale estimates of flood risk to inform efficient and equitable flood risk management policy\textsuperscript{23,38,44}. However, our findings and caveats highlighted above raise questions about the suitability of extrapolation results that don’t consider a large range of uncertainties.
Accessible research directions for overcoming these challenges are outlined here, can be facilitated with technical assistance from FEMA, and represent pathways to improve the lynchpin tools of flood risk management.

**Methods**

**Theoretical Framework**

A flood event reduces the value of an affected structure, which is the value lost through damage, $L$. In the typical model\(^3\), $L$ is the product of the structure value, $v$, and a damage relationship, $D$, that translates the inundation depth, $x$, faced by an affected structure along with characteristics of the hazard and the structure (given respectively by the vectors $z^H$ and $z^S$), into the fraction of value lost. For an affected property, $i$, the true loss is thus

$$L_{i}^{\text{True}} = D^{\text{True}}[x_i; z_i^H, z_i^S]v_i \quad (1)$$

Research elucidates the many factors of $z^H$ and $z^S$, and their complex interactions and functional forms, that must be captured in $D$ to embody the process that generates $L$.\(^{11,12,15}\) The crucial feature of eq. (1) is that heterogeneity in the structure’s characteristics ($z_i^S$) and hazard it faces ($x_i, z_i^H$) make the flood damage function specific to the property at risk. Thus, for different structures, identical hazard exposure can generate markedly different fractional damage. A key problem in practice is that $L_{i}^{\text{True}}$ is unobserved and $D^{\text{True}}$ can’t be derived but estimates of $L_{i}^{\text{True}}$ are needed for a wide variety of flood risk management applications.\(^{3,5,8,23,38,43,44,58}\)

A general approach to handle this problem relies on $L_{i}^{\text{Obs}}$, records of loss linked to $x$ ($x_{i}^{\text{Obs}}$), $v$ ($v_{i}^{\text{Obs}}$) and a limited set of $z^H$ and $z^S$ ($z_i^H, z_i^S$) to derive $\hat{D}$, an empirical damage function that generates loss estimates, $\hat{L}_{i}$, which agree most closely with $L_{i}^{\text{Obs}}$.\(^{11–14,17–20}\) The recommended approach is to take the fraction of damage to the structure affected by flooding, often called relative loss, represented by $rel_{i}^{\text{Obs}} = \frac{L_{i}^{\text{Obs}}}{v_{i}^{\text{Obs}}} \quad (2)$ as the dependent variable for finding $\hat{D}$ \(^3\). Note that this form requires the assumption that $L$ and $v$ scale linearly.

The “best” damage function of a set of candidate modeling approaches to represent the relationship between variables and loss, $k^{\text{Cand}}$, can be represented by:

$$\overline{rel}_{i} = \overline{\hat{D}}[x_i^{\text{obs}}; z_i^{H, \text{obs}}, z_i^{S, \text{obs}}]$$

$$\epsilon = rel_{i}^{\text{Obs}} - \overline{rel}_{i} \quad (3)$$

$$\hat{D} = \arg\min_{D_{\epsilon}} \{g(\epsilon), k \in k^{\text{Cand}}\} \quad (4)$$

where $\epsilon_i$ is the prediction error and $g(\epsilon)$ takes many forms such as absolute error or squared error\(^{12,13,17–19}\).

A common application of $\hat{D}$ is for the estimation of prospective flood losses for structures at risk, called extrapolation. Prospective estimation implies that hazard terms must be modeled, and relevant inputs may not be available (Table S1). Additionally, many structure characteristics used in damage functions, including structure value, are inconsistently observed over the set of structures at risk under consideration (Table S1).\(^{3,6,7,22}\)
In extrapolation, flood characteristics vary by location and return period, \( \rho \), according to the distribution \( f \). An estimated loss to a property from an event corresponding to a particular return period, \( \hat{\rho} \), can be represented by:

\[
\hat{l}(\hat{\rho}) = D\{x_i(\hat{\rho})^{\text{Mod}}; z_i^{H,\text{Mod}}; z_i^{S,\text{Obs}|\text{Mod}}\} v_i^{\text{Obs}|\text{Mod}} \tag{5}
\]

Where \( \text{Obs} \) and \( \text{Mod} \) superscripts indicate if a variable is derived from observations or models, or a combination in practice (Table S1). Aggregating losses over return periods provides the estimated expected annual loss (EAL), the expected damages for a property in any given year:

\[
\hat{EAL}_i = \int \hat{l}(\rho) f\{x_i(\rho)^{\text{Mod}}; z_i(\rho)^{H,\text{Mod}}\} d\rho \tag{6}
\]

A property’s value is a function of its own characteristics (e.g., square footage, architectural style), amenities determined by its location, and the perceived and actual flood risk that it faces. Over the long run, inundation risk can influence decisions to construct or abandon a property of a particular type at a particular site, suggesting that \( z_i^S \) and \( i^H \) are themselves functions of \( x_i \) and \( z_i^H \). These endogeneities and feedbacks are important\(^{52,59,60}\) but excluded from this framework for simplicity because these relationships aren’t fully understood and rarely included in prospective loss estimation applications.

**Validation Data**

The data sources that enable this analysis are a restricted National Flood Insurance Policy (NFIP) claims dataset that consists of 2,085,015 records of loss due to flooding from 1972-2015 (NFIP Loss) and the PLACES database which consists of nearly 150 million land parcels in 3,055 US counties and links precise spatial boundaries of parcels to ownership, buildings, sales, infrastructure, demographics, geophysical characteristics, and a wide range of other variables (Table S4)\(^61\). For the purposes of developing empirical damage functions that can easily be applied in large scale flood loss estimation, property level and aggregate indicators are matched across the NFIP Loss and PLACES data (Table S1, S3).

The NFIP Loss data contains in varying levels of completeness and reliability, variables such as the assessed inundation in feet of flood waters relative to first floor elevation, the appraised damage to structures, the appraised value of the structure, the basement type of the structure, the number of stories in the structure, the occupancy type (i.e. single family residential), the construction date, flood zone, a post Flood Insurance Rate Map (FIRM) indicator, the zip code, and the date of loss. The full list of variables available in the raw data, their definitions and their completeness are shown in Table S1. Notable variables available in the raw data that are not included in the subsequent analysis due to high missingness and difficulty to apply in large scale flood loss estimation include duration of flood water, exterior wall structure, flow velocity, and foundation type. We restrict our analysis to damage to structures as there are no indicators about the types of contents in buildings. These are the appraised damage to structures which are limited to damages covered by the policy, but not limited by the amount of building coverage on a policy. Following the recommendation in Merz et al. (2010), we use the appraised damage to structures in calculating the dependent variable of relative loss, the appraised damage divided by
appraised value of structure. Following the finding of Wing et al. (2020) that loss heterogeneity possibly varies with structure value, we employ Federal Housing Finance Agency housing price deflators for 3-digit zip codes (date back to 1996) and for counties (date back to 1977) in order to evaluate the influence of structure value in terms of 2018 dollars. Further, we link zip codes to their FEMA Community Rating System Community which enables us to identify when the community first was assigned a FIRM which may reveal information about the strength of building codes in an area. To proxy non depth characteristics of flooding, we merge zip code interacted with flood zone indicators for presence of coastal, riverine or lake exposure and we employ the NCEI Storm database to identify which events are influenced by Nor’easters, hurricanes and other storm events. Finally, to emulate the regional approach to loss estimation seen in some FEMA and USACE DDFs, we assign losses within states to the appropriate FEMA region (Fig. S2). We end up with a final sample of 845,776 records of loss to single family homes from 446 month-year combinations that proxy unique flood events. The filters employed, and their impact on the size of the sample, are outlined in Table S3. The most influential filter is limiting depth of flooding relative to first-floor elevation due to data quality concerns.

**Depth Damage Functions**

In light of substantial data constraints, the predominant approach to handle the problem of prospective flood loss estimation in the U.S. omits non-depth characteristics of the hazard and includes limited structural characteristics. U.S. agencies such as FEMA and USACE develop depth-damage functions (DDFs), using engineering expert judgment to derive a single damage relationship for so-called archetypes, \(a(i)\), that map broad groups of observed house characteristics, \(\tilde{z}_a\), into categories of loss relationships:

\[
L_{i}^{FEMA} = D_{FEMA}(x_i, \tilde{z}_a) \nu_i \quad (7)
\]

Importantly, \(D_{FEMA}\) is defined such that depth is related to expected losses in an increasing, monotonic shape meaning each additional unit of inundation results in more damage to the structure.

We employ “generic” USACE DDFs that are recommended as the default damage function for loss estimation in a 2011 FEMA Benefit Cost Analysis report. These take number of stories and presence of basement as inputs resulting in \(\tilde{z}_a\) such as one story without a basement and two or more story with a basement. Generic DDFs are only recommended for loss estimation in non-coastal flood zones, but DDFs produced by FEMA and the USACE in coastal flood zones require data like the duration of flooding, wave heights, or first-floor elevation which are rarely available in validation or extrapolation (Table S1). In all, there are 158 residential curves catalogued in the HAZUS library, 27 which fit the data constraints for validation and extrapolation for single family homes. We also employ FEMA region specific DDFs which are available in FEMA regions 1-6. In FEMA regions 1-3, which consist of thirteen Mid- and North-Atlantic states from Virginia to Pennsylvania to Maine, damage functions that individually incorporate hazard variables like wave height and erosion were produced in addition to DDFs but require data unavailable in validation or extrapolation (Table S1). These DDFs are unique for incorporating information about structure age and other characteristics in order to define min.
most likely and max damage curves for each classic archetype. When these damage functions are employed, we use age of the structure at the time of loss to assign a min, most likely or max damage function based off the specifications outlined in the 2015 North Atlantic Coast Comprehensive Study by the USACE. The USACE DDFs employed in this study are visualized in Fig. S3, but only the most likely damage curve is shown. Others can be found in the 2015 report.

Empirical Modelling Strategies

We derive three empirically based damage functions with varying degrees of parsimony and interpretability to capture heterogeneity in the relative loss dependent variable in the procedure outlined by eq. (2) – (4):

First, we estimate

$$rel_{i}^{obs} = D^{Depth}[x_{i}^{obs}] + \zeta_{i}$$  

via penalized maximum likelihood, minimizing the variance of the random disturbance term, $\zeta$, while enforcing monotonicity in the depth-loss relationship. These shape-constrained, increasing monotonic additive models that relate inundation depth relative to first-floor elevation (Depth GAM) to relative loss are fit using the R package scam$^{66}$. In contrast to our flexible form, some depth-only damage functions in the literature assume a root-based functional form to enforce the increasing, monotonic shapes which requires ad-hoc transformations for depths below first floor$^{12}$. We also estimate Depth GAMs with FEMA region stratification, which is represented by:

$$rel_{i}^{obs} = D^{Depth}[x_{i}^{obs}, \theta_{k(i)}] + \zeta_{i}$$  

where $D$’s shape is characterized by the parameters, $\theta$, that represent a vector of weighting coefficients on thin-plate basis splines, and whose values are stratified according to subsets of the data corresponding to FEMA regions:

$$k(i) = \delta_{FEMA\ Region}(i)$$  

where $\delta$ is the value of the variable denoted in the subscript matches the value of that variable for the observation $i$.

This approach can be extended to include a parsimonious set of property attributes and non-depth flood hazard attributes, which we call Stratified GAMs. For these, we consider FEMA structural archetypes without and with FEMA region, and FEMA structural archetypes with storm and high value indicator without and with FEMA region. The last of these can be represented by:

$$k(i) = \delta_{Basement}(i) \times \delta_{Single-story}(i) \times \delta_{80th\ %ile\ Value}(i) \times \delta_{Storm}(i) \times \delta_{FEMA\ Region}(i)$$  

Finally, we use extremely randomized trees (Extra Trees), a tree-based ensemble method for supervised regression to emulate the preference in the literature for sophisticated and complex models$^{53}$. Compared to tree-based ensemble alternatives like Gradient Boosting or Random
Forest, we find this approach fits quickly and is less sensitive to hyperparameter specifications which makes it suitable for our computationally intensive cross-validation strategy. We the Scikit-learn ExtraTrees implementation in Python with default parameters, 200 trees, the minimum number of samples in a leaf set to 3, and bootstrapped samples. We train and validate Extra Trees with the same four variable specifications as the Stratified GAMs, as well as a specification with the previously described water body proximity indicators, post-FIRM indicator, and age of home. Instead of a high value indicator as input in the Stratified GAMs, value is passed as the raw deflated percentile of structure value in the state.

Cross Validation and Performance Metrics
We evaluate the accuracy and precision of USACE DDFs, and empirically derived alternatives in a cross-validation framework informed by recent research on best practices in spatio-temporal settings.

Commonly implemented random k-fold cross validation (R-CV) can underestimate prediction error in extrapolation. In contrast, spatiotemporally blocked cross-validation strategies (ST-CV) are needed to avoid overconfidence in model predictions. Following the recommendations of Roberts et al., (2017), we block on the flood event grouping structure inherent to the NFIP Loss data to ensure spatial and temporal endogenously varying characteristics are well mixed in training and testing splits. Flood events are defined as month-year groupings of claims which may occasionally overestimate event numbers (i.e. claims from an event that span the end of one month to the beginning of the next) and overestimate event extents (i.e. claims that occur in the same month from different hydrologic circumstances). We employ a leave-one out (LOO) strategy whereby all observations in all events but one are used in training empirical models, and the trained model then predicts on the observations from the held out event. This form of cross-validation ensures no information leaks from training to testing and avoids overstating performance. Note that USACE DDFs are not “trained” in any round of cross-validation. Importantly, there are instances where the Stratified GAMs only have one event in a FEMA region that entails combinations of certain characteristics. In these cases, we ensure that the validation for other models excludes these observations as well.

We evaluate model performance, the agreement of relative loss predictions to observed relative losses, in terms of mean bias error (MBE) to capture the average tendency to over or underpredict (12); Root mean square error (RMSE) to capture a single measure of predictive power (13); and Pearson’s coefficient of determination (R²) to capture a model’s capacity to explain variation in relative loss (14):

\[ MBE = \frac{1}{N} \sum_{i=1}^{N} (\hat{rel}_i - rel_i) \]  (12)

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{N} (rel_i - \hat{rel}_i)^2}{N}} \]  (13)

\[ R^2 = 1 - \frac{\sum_{i=1}^{N} (rel_i - \hat{rel}_i)^2}{\sum_{i=1}^{N} (rel_i - \bar{rel})^2} \]  (14)
where $N$ is the number of observations of loss, $\hat{rel}_i$ is the predicted relative loss, and $\bar{rel}_i$ is the average relative loss.

**Large-Scale Flood Loss Estimation and Evaluation in PLACES**

PLACES accommodates large scale flood loss estimation across a comprehensive set of return periods and flood types from property level hazard through the First Street Foundation Flood Hazard Model and vulnerability inputs linked through ZTRAX and other sources. We scale the methodology in Pollack et al., (2022) to 72.4 million single-family homes in 2,983 U.S. counties and calculate expected annual losses based on the generic USACE DDFs and Stratified GAMs which incorporate FEMA region, number of stories, basement presence, a high value indicator and a storm indicator.

Roberts et al. (2017) outline several valid approaches with various strengths and weaknesses to making ‘final’ predictions in extrapolation after blocked cross validation. Because we are primarily interested in the agreement and disagreement of final predictions across models when applied to PLACES data, we train the ‘final’ Stratified GAMs on all loss data. There are some counties with complete extrapolation (Fig. 1), but the disagreement between the FEMA DDFs and Stratified GAMs are similar in and out of these counties (Fig. 1, Fig. 8).

To account for uncertainty in translating modeled depth relative to grade provided by the First Street Foundation into depth relative to first-floor elevation, we estimate losses under three damage scenarios. These make different assumptions about depth adjustments informed by the HAZUS Technical Methodology manual and detailed in Table S4 (See SI). In short, whether a home is built after its first flood insurance rate map (FIRM), the flood zone it’s built in, the FEMA region it’s in, and its foundation determine its first-floor elevation adjustment. In the “Reference” damage scenario, the combination of those variables determines the first-floor elevation adjustment. In the “Low” and “High” damage scenario, first-floor elevation adjustments are selected to ensure the min or max damage is applied for a particular modeled depth. In any scenario, if foundation is unknown the weighted proportion method used in Pollack et al., (2022) is applied and extended to include pile foundation types.

In addition to first-floor elevation adjustments, there is uncertainty in how to apply Stratified GAMs that take a storm event indicator as input. In the “Reference” damage scenario, we take the history of storm exposure proportions from the NCEI events linked to the NFIP Loss data and calculate the proportion of storm-based losses for each county and all its neighbors. The damages are estimated for storm and non-storm losses, and then the weighted proportion is applied to losses under each scenario for the property. In the “Low” and “High scenarios, the proportion is set to 0 and 1, as storm losses are always higher.

In extrapolation, there is no ground truth data on losses experienced by homes facing the First Street Foundation modeled depths and therefore no ability to estimate an observed bias or error. Instead, we analyze and compare

$$L_i^* = D_{k(i)}[x_i^{FirstStreet}]v_i^{ZTRAX}$$  \hspace{1cm} (15)
where $D^*_k(i)$ is the Stratified GAM applied to a property, $i$, with an estimated loss of $L^*_i$ based off the First Street Foundation modeled depth at the property and the structure value made available from ZTRAX. Specifically, we compare $D^*_k(i)$ and $D^{FEMA}$ at the property level across depth exposures (eq. 5), and $L^*_i$ and $L^{FEMA}_i$ summed over all return periods across depth exposures, structure values and structure types.

We also transform $L^*_i$ and $L^{FEMA}_i$ into EAL (eq. 6), a common representation, as calculated in Pollack et al. (2022). We compare the total EAL across depth exposures, structure, values and structure types. Then, we calculate the percent difference of expected annual loss summed to the county level from Stratified GAM to the USACE DDFs and plot the magnitudes, which we take as a measure of estimation uncertainty, overall and across depth exposure strata. We evaluate these county level discrepancies for various combinations of Stratified GAMs under the reference and high damage scenarios, Stratified GAMs with an adjustment for Hurricane Katrina, and USACE DDFs with and without FEMA region variation (See SI).

Data Availability Statement

The depth-damage functions and empirical alternatives derived are available from the authors upon request. A redacted version of the database used to derive empirical alternatives and validate all functions is available OpenFEMA: https://www.fema.gov/about/reports-and-data/openfema

Open NFIP Policy data is available at: https://www.fema.gov/openfema-data-page/fima-nfip-redacted-policies-v1; FHFA Housing Price Indices: https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx; A link to information on access to PLACES: https://placeslab.org/get-access/; The structure inventory is made available by ZTRAX, but new licenses are no longer available to researchers, and those currently with access will lose it September 30, 2023. (https://www.zillow.com/research/ztrax/); The flood hazard model is made available by the First Street Foundation, and licenses are available for non-commercial purposes: https://firststreet.org/data-access/public-access/; All code is stored in a private GitHub repository, and code can be made available upon request.

References


Figures

**Figure 1**: Spatial distribution of 846K flood loss observations from 1972-2015 (A) and the distribution of percent damage across key subgroups (B). A: Light gray fill indicates counties with no observations in the sample. B: Subgroups are split by structural characteristics (B, left), as well as by depth intensity (< 4 ft. vs. ≥ 4 ft.) across major events, storm events, other flooding events, and all events (B, right). Percent damage is binned into ten bins of equal width. Dark horizontal lines denote the mean percent damage within that group. The "high-value" variable is true when the deflated structure value is within the 80th percentile of all deflated structure values in its state; the "old age" indicator is true for structures that are over 30 years old at the time of loss.
Figure 2: Percent damage distributions (top of each panel) and proportion of observations (bottom of each panel) across inundation depths as a function of univariate subgroup splits. Subgroups are based on whether (blue) or not (orange) a structure was exposed to floods during a storm (A), has a basement (B), is single story (C), has a value within the top 80th percentile of deflated structure values in its state (D), is built after its community joined the national flood insurance program (E), is older than 30 years at the time of loss (F), or is in a zip code adjacent to or containing a river (G), lake (H) or coastline (I). Black diamonds indicate the mean percent damage for all observations; red circles indicates the mean percent damage for observations from Katrina. Boxplot whiskers reflect the upper and lower quartiles +/- 1.5*interquartile range.
**Figure 3:** The role of structure value and extreme events on the percent damage distribution. A and B: percent damage distributions across basement types (A) and coastal exposure (B) with stratifications by binned structure values. Lower panels show the percentage of observations in each group (i.e., basement type, coastal or not) within each value bin. C-D: percent damage distributions across inundation depths for non-Katrina (C) and Katrina (D) losses for observations in different locations in FEMA region 6. Lower bottom panels show the number of loss observations across depths within each location is shown. Black diamonds denote the mean percent damage within each group. Boxplot whiskers reflect the upper and lower quartiles +/- 1.5*interquartile range.
Figure 4: Generic USACE depth-damage functions (DDFs) vs. observed percent damage distributions for four subgroups (number of stories x basement presence) (Panels A, B, C and D). Black diamonds reflect the mean percent damage at each depth for all observations in the stratification, the red circle with white fill reflects the mean for Katrina observations, and the blue circle with white fill reflects the mean for non-Katrina observations. Below each boxplot panel, we plot bars indicating the number of observations due to Katrina (red) and other observations (blue). Boxplot whiskers reflect the upper and lower quartiles +/- 1.5*interquartile range.
Figure 5: Distributions of prediction errors (predicted minus observed percent damage) (top panels) and proportions of observations (bottom panels) for USACE DDFs and our three empirical damage models (Depth GAM, Stratified GAM, and Extra Trees) for relevant subgroups. Subgroups are defined as follows: non-storm (A) and storm events (B) by inundation depth, by structure value (C, all events), and by structure archetype (D, all events). Black diamonds reflect the mean percent damage prediction bias for a damage function within a subgroup. No model is trained or validated on Katrina in the rounds of validation displayed here.
Figure 6: Distributions of prediction errors for losses in FEMA regions 4 and 6 (predicted minus observed percent damage) (top panels) and proportions of observations (bottom panels) for USACE DDFs, our three empirical damage models (Depth GAM, Stratified GAM, and Extra Trees), and Stratified GAMs adjusted for the influence of Katrina (See Methods). Panel A shows the distributions for all storm events, and Panel B shows the distributions for all storm events excluding Katrina in the validation set. Black diamonds reflect the mean percent damage prediction bias for a damage function within a subgroup.
Figure 7 Differences from validation to extrapolation under the reference damage scenario in the depth relative to first-floor elevation distributions, predictions across USACE Generic DDFs and Stratified GAMs, and implications for losses. Panel A (B) shows the proportion of observations in validation (extrapolation for known structures) across different depth relative to first-floor bins and presence in (red) or outside (blue) the Special Flood Hazard Area (SFHA). Panel C (D) shows the mean predicted percent damage in validation (extrapolation for known structures) across these depths, with solid (dashed) lines indicating homes with (without) basements and green (blue) color indicating USACE Generic (Stratified GAM) predictions. Panel E (F) shows the proportion of overall losses in validation (extrapolation for known structures) across depth bins and damage functions. The proportion of observations and losses in extrapolation are weighted by return periods using the trapezoidal approximation method.
Figure 8: Overall expected annual losses in extrapolation. Properties with non-zero expected annual loss by depth bins (Panel A), structure value (Panel B), basement type (Panel C) and overall (Panel D) stratified by Stratified GAM or USACE DDFs. On the bottom row, the same is shown for aggregate expected annual loss within each grouping (Panels E, F, G, and H). Bars denote the estimates under the reference damage scenario and the red (dark blue) dots denote the range from estimates under the minimum to maximum damage scenario.
**Figure 9:** Percent difference in census tract expected annual loss for the Stratified GAM damage functions benchmarked to the USACE DDFs. Panel A shows the results for all properties with losses, and the remaining panels show results for subsets of depth exposures in the 500-yr return period. Panel B shows for depths less than 2 feet, Panel C shows for depths between 2 and 4 feet, Panel D shows for depths between 4 and 8 feet, and Panel E shows for depths at least 8 feet.
Figure 1

Spatial distribution of 846K flood loss observations from 1972-2015 (A) and the distribution of percent damage across key subgroups (B). A: Light gray fill indicates counties with no observations in the sample. B: Subgroups are split by structural characteristics (B, left), as well as by depth intensity (< 4 ft. vs. ≥ 4 ft.) across major events, storm events, other flooding events, and all events (B, right). Percent
damage is binned into ten bins of equal width. Dark horizontal lines denote the mean percent damage within that group. The “high-value” variable is true when the deflated structure value is within the 80th percentile of all deflated structure values in its state; the “old age” indicator is true for structures that are over 30 years old at the time of loss.

Figure 2

Percent damage distributions (top of each panel) and proportion of observations (bottom of each panel) across inundation depths as a function of univariate subgroup splits. Subgroups are based on whether (blue) or not (orange) a structure was exposed to floods during a storm (A), has a basement (B), is single story (C), has a value within the top 80th percentile of deflated structure values in its state (D), is built after its community joined the national flood insurance program (E), is older than 30 years at the time of loss (F), or is in a zip code adjacent to or containing a river (G), lake (H) or coastline (I). Black diamonds
indicate the mean percent damage for all observations; red circles indicates the mean percent damage for observations from Katrina. Boxplot whiskers reflect the upper and lower quartiles +/- 1.5*interquartile range.

Figure 3

The role of structure value and extreme events on the percent damage distribution. A and B: percent damage distributions across basement types (A) and coastal exposure (B) with stratifications by binned structure values. Lower panels show the percentage of observations in each group (i.e., basement type, coastal or not) within each value bin. C-D: percent damage distributions across inundation depths for non-Katrina (C) and Katrina (D) losses for observations in different locations in FEMA region 6. Lower bottom panels show the number of loss observations across depths within each location is shown. Black diamonds denote the mean percent damage within each group. Boxplot whiskers reflect the upper and lower quartiles +/- 1.5*interquartile range.
Figure 4

Generic USACE depth-damage functions (DDFs) vs. observed percent damage distributions for four subgroups (number of stories x basement presence) (Panels A, B, C and D). Black diamonds reflect the mean percent damage at each depth for all observations in the stratification, the red circle with white fill reflects the mean for Katrina observations, and the blue circle with white fill reflects the mean for non-Katrina observations. Below each boxplot panel, we plot bars indicating the number of observations due to Katrina (red) and other observations (blue). Boxplot whiskers reflect the upper and lower quartiles +/-1.5*interquartile range.
Figure 5

Distributions of prediction errors (predicted minus observed percent damage) (top panels) and proportions of observations (bottom panels) for USACE DDFs and our three empirical damage models (Depth GAM, Stratified GAM, and Extra Trees) for relevant subgroups. Subgroups are defined as follows: non-storm (A) and storm events (B) by inundation depth, by structure value (C, all events), and by structure archetype (D, all events). Black diamonds reflect the mean percent damage prediction bias for a damage function within a subgroup. No model is trained or validated on Katrina in the rounds of validation displayed here.
Figure 6

Distributions of prediction errors for losses in FEMA regions 4 and 6 (predicted minus observed percent damage) (top panels) and proportions of observations (bottom panels) for USACE DDFs, our three empirical damage models (Depth GAM, Stratified GAM, and Extra Trees), and Stratified GAMs adjusted for the influence of Katrina (See Methods). Panel A shows the distributions for all storm events, and Panel B shows the distributions for all storm events excluding Katrina in the validation set. Black diamonds reflect the mean percent damage prediction bias for a damage function within a subgroup.
Figure 7

Differences from validation to extrapolation under the reference damage scenario in the depth relative to first-floor elevation distributions, predictions across USACE Generic DDFs and Stratified GAMs, and implications for losses. Panel A (B) shows the proportion of observations in validation (extrapolation for known structures) across different depth relative to first-floor bins and presence in (red) or outside (blue) the Special Flood Hazard Area (SFHA). Panel C (D) shows the mean predicted percent damage in validation (extrapolation for known structures) across these depths, with solid (dashed) lines indicating homes with (without) basements and green (blue) color indicating USACE Generic (Stratified GAM) predictions. Panel E (F) shows the proportion of overall losses in validation (extrapolation for known structures) across depth bins and damage functions. The proportion of observations and losses in extrapolation are weighted by return periods using the trapezoidal approximation method.
Figure 8

Overall expected annual losses in extrapolation. Properties with non-zero expected annual loss by depth bins (Panel A), structure value (Panel B), basement type (Panel C) and overall (Panel D) stratified by Stratified GAM or USACE DDFs. On the bottom row, the same is shown for aggregate expected annual loss within each grouping (Panels E, F, G, and H). Bars denote the estimates under the reference damage scenario and the red (dark blue) dots denote the range from estimates under the minimum to maximum damage scenario.
Figure 9

Percent difference in census tract expected annual loss for the Stratified GAM damage functions benchmarked to the USACE DDFs. Panel A shows the results for all properties with losses, and the remaining panels show results for subsets of depth exposures in the 500-yr return period. Panel B shows for depths less than 2 feet, Panel C shows for depths between 2 and 4 feet, Panel D shows for depths between 4 and 8 feet, and Panel E shows for depths at least 8 feet.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI070822.docx