UNSAFE v0.1 Technical Documentation

Background

UNSAFE modifies a property-level risk assessment framework common in academic research and practice (e.g. Federal Emergency Management Agency (FEMA) loss avoidance studies, United States Army Corps of Engineers (USACE) feasibility studies).

Standard theoretical framework for property-level flood-loss estimation

Flooding reduces the value of affected structures by a damage amount, L_i. L_i is represented as the product of the value of a structure, v_i, and a hazard-damage relationship, D. This relationship, often specified as a depth-damage function (DDF), translates the inundation depth, x_i, faced by an affected structure based on characteristics of the affected structure (given by the vector z_i^s) into the fraction of value lost. U.S. agencies such as FEMA and the USACE develop DDFs using engineering expert judgement to derive depth-damage relationships for so-called archetypes, $a(i)$, that map broad groups of observed house characteristics, $\tilde{z}_{a(i)}^s$, into categories of point-estimate based depth-damage relationships. These DDFs are defined such that depth is related to loss in an increasing, monotonic shape meaning that each additional unit of inundation results in more damage to a structure. Archetypes consist of the foundation type, fnd_type_i, number of stories, num_story_i, and occupancy type, occ_type_i (i.e. residential or commercial occupancy), of a structure. The vector z_i^s also consists of the structure's first-floor elevation. The difference of this vector, ffe_i, from x_i yields the inundation that is the input to a DDF. The standard loss estimation model can be represented as:

$$L_i = D[x_i, \tilde{z}_{a(i)}^s] * v_i$$

Standard way to operationalize this approach

The most common way analysts operationalize the standard theoretical framework is through point-based representations of vulnerability and exposure through

expert-based DDFs, and the USACE National Structure Inventory, respectively. According to its technical documentation, the NSI [**emphasis ours**] "is a system of databases containing structure inventories of varying quality and spatial coverage. The purpose of the NSI databases is to facilitate storage and sharing of **point based** structure inventories used in the assessment and analysis of natural hazards."

The NSI represents each structure's: * (x,y) coordinates in Geographic Coordinate System (GCS) WGS84 which are needed to link structures to flood hazard, and * characteristics needed for flood-loss estimation (occ_type_i, fnd_type_i, ffe_i, num_story_i, v_i)

The NSI records are used as-is even though they are subject to a combination of measurement and/or modeling uncertainty.

This standard approach neglects observation and/or model unceratinty in $\tilde{z}^s_{a(i)}$, v_i, and D. The NSI does not clarify which records in the structure inventories are of high or low quality. In addition, there is a limited understanding of how well expert-based DDFs produced by FEMA or the USACE accurately or consistently reproduce real-world depth-damage relationships.

How UNSAFE accounts for uncertainty

UNSAFE structures and simplifies the process for estimating property-level flood loss under uncertianty. In its current implementation, UNSAFE adds uncertainty to the standard exposure (i.e NSI) and vulnerability (i.e. expert DDFs) inputs for property-level flood risk assessments in the U.S. Currently, UNSAFE is limited to single family residences ($occ\_type\_i =$ 'RES1') with one or two story, and basement, crawl space, or slab foundation. We are working on expanding UNSAFE to every occupancy type included in the NSI. In the future, we plan for UNSAFE to treat the current implementation as a benchmark approach, and allow for arbitrary data input for any component of the loss-estimation framework.

Uncertainty in exposure

UNSAFE calls the NSI API for a single U.S. county, downloads the county's structure inventory, and treats the spatial coordinates and occ_type_i of each record as-is. From this base inventory, UNSAFE generates a J member ensemble of plausible realizations of key variables for each structure. We define the distributions that ensemble members are drawn from based on guidance in the NSI and peer-reviewed academic literature.

Structure value

$$v_i \sim \begin{cases} N(NSI\_val_i, .2 * NSI\_val_i) & \text{if } v_{i,j} \geq 1 \\ 1 & otherwise \end{cases}$$

where NSI_val_i is the structure value provided in the NSI record as-is.

The NSI represents structure value as depreciated replacement values. These are estimated based on an assumed replacement category and a dollar per square footage estimate for that category. According to the NSI documentation, these are informed by an "analysis of survey data, parcel use types, and other source inputs." There is no documentationon what the replacement categories are, or what the dollar per square footage estimates for these categories are. Values are in 2021 price levels. Dollars per square foot are multiplied by square footage estimates from building footprint data to obtain a structure value estimate. Replacement values are depreciated according to a 1% per year schedule for the 1st 20 years and it is assumed there is no further depreciation.

Due to the lack of transparency in the NSI documentation, it is unclear what the underlying distribution for structure values are. We assume a distributional form that produces normally distributed noise around the NSI records. We make the generous assumption that the NSI structure value records are mean unbiased and are as precise as state-of-the-art automated valuation models for property market transactions (Krause, Martin, and Fix 2020). We use a piecewise distribution to left-censor draws from the normal distribution at \$1 because it is illogical to have structure values below \$1. In the case study provided in `examples/` the lowest structure value is \$92,695 and it is extremely unlikely that values below \$1 are drawn (i.e. the probability of drawing a value less than \$25,000 is 0.0001).

We note that there is emerging evidence that the NSI estimates of structure value are mean-biased and noisier than we represent. However, results vary widely on a case-study to case-study basis and case studies are limited to single county studies which makes it challenging to specify a single general distribution for any case study (Shultz Steven 2017b, 2017a; Mostafiz et al. 2021; Sanderson and Cox 2023). Lacking this guidance, we employ a likely over-optimistic approach in the first version of UNSAFE. We hope to refine this approach as more studies characterize the accuracy and precision of the NSI's structure value estimates. It would also be helpful for the NSI technical documentation to provide guidance on the accuracy and precision of their model.

Number of stories

$$num\_story_i \sim binomial(J, p^{tract})$$

where p^{tract} is the proportion of residential structures with 1 story in the census tract that structure i is located in. In the current implementation of UNSAFE, DDFs for residential houses are for 1 and 2 story structures, so we can use the binomial distribution.

According to the NSI technical documentation, the number of stories records are drawn from parcel data where possible. When missing, square footage for single family homes (RES1 category) is estimated by taking 86% of a structure's footprint. This percentage was estimated at a nationwide level using available

parcel, survey, and footprint datasets. If no footprint is available, sq. ft. is randomly assigned "from a distribution that varies by the structure's year built and census region." If the number of stories is not available for RES1, it is estimated by dividing the estimated sq. ft. by the structure's footprint, where any decimal greater than .25 after a digit leads to rounding up the number of story estimate (i.e. 1.25 means 2 story home). If no footprint is available, the number stories is randomly assigned from a distribution that varies by year built and census region. By specifying the UNSAFE distribution at the census tract level, we capture the site-specific assignments that the underlying parcel data provides in addition to the heuristic assignments.

Foundation type

$$fnd_i \sim multinomial(J, p_{slab}^{tract}, p_{crawl}^{tract}, p_{basement}^{tract})$$

where $p_{fnd\_type}^{tract}$ corresponds to the proportion of structures in the census tract that structure i is located in with a given fnd_type. It is uncommon, but single family residences can have other foundation types than those shown above. UNSAFE currently is designed to work in case study environments where only these foundation types are present.

The NSI technical documentation states that foundation type is mapped from parcel data when available. When this data is not available, the foundation types are randomly assigned using the HAZUS structure inventory. The Hazus 6.0 Inventory Technical Documentation states that foundation types are based on General Building Stock distribution tables that contain the proportions of foundation type for different Census Divisions (i.e. New England or South Atlantic) based on surveys conducted in 1993, 1997, or 2000 (depending on riverine or coastal flood hazard zone). The multinomial distribution we define attempts to capture the underlying distribution that the NSI draws these values from, adjusted for records provided by parcel datasets. By specifying the UNSAFE distribution at the census tract level, we capture the site-specific assignments that the underlying parcel data provides in addition to the random assignments from the HAZUS structure inventory. We specify this distribution at the census tract level, instead of block group or block level, so that J_{tract} has a sufficient sample size for specifying We note that foundation type data is relatively uncommon in parcel data sources (Nolte et al. 2023).

First-floor elevation

$$ffe_i = \begin{cases} triangular(0, .5, 1.5), & \text{if } fnd_{i,j} = \text{Slab} \\ triangular(0, .5, 1.5), & \text{if } fnd_{i,j} = \text{Crawl Space} \\ triangular(0, 1.5, 4), & \text{if } fnd_{i,j} = \text{Basement} \end{cases}$$

According to the NSI technical documentation, foundation heights are mapped to each foundation type based on survey estimates from 2021 with each assumed

height "closely matching the median value from the survey." Although the distributions that these are drawn from are not shared in the technical documentation, there is some guidance on the distributions from this repository, which is the basis for a Nature Climate Change study from 2022 with a co-author from the USACE (Wing et al. 2022).

A note on NSI records informed by parcel datasets

According to the NSI documentation, foundation type and number of stories records are occasionally based on parcel datasets. Parcel datasets are likely more accurate (they may contain errors) than the random assignment from the NSI procedure. This suggests that UNSAFE may add unneeded noise to some records, and subsequently flood-loss estimates. Our philosophy is that until the NSI technical documentation provides guidance about which records are of high quality, such as by indicating which records come from parcel data or what distributions certain records are drawn from, all records should be treated as uncertain. We acknowledge that this may lead to noisier flood-loss estimates at the structure level when records are informed by parcel datasets. However, we expect that expected property-level losses, across the ensemble generated by UNSAFE, aggregated to census tract or broader spatial scales will be less biased than if all structure records were taken as-is.

In future versions of UNSAFE, we plan to allow users to specify base structure inventories other than the NSI. In such cases, users can indicate when records should be taken as-is, and when there is uncertainty around records due to measurement and/or model error. For example, appraiser databases may include detailed documentation on the staitical fit of automated value model used to determine depreciated replacement costs.

Uncertainty in vulnerability

As far as we are aware, the DDFs produced from the USACE 2015 North Atlantic Coast Comprehensive Study (NACCS) report provides the only set of expert-based DDFs in the U.S. that are defined with uncertainty characterization. This is the initial characterization of uncertainty in vulnerability in UNSAFE.

Let l_i be the fractional damage output of an expert-based depth-damage function produced by a particular group, D^{group}, and let ind_i be the amount of inundation a structure is exposed to, calculated by $x_i - ffe_i$. For a given archetype, a_i, NACCS specifies three DDFs which are described as corresponding to "min," "most likely," and "max" damages. This reflects a triangular probability distribution, interpreting "most likely" as the "mode." We draw from the NACCS DDFs, available at a Zenodo repoistory, as follows:

$$l_i \sim triangular(DDF_{min}^{NACCS}(ind_i, a_i), DDF_{mode}^{NACCS}(ind_i, a_i), DDF_{min}^{NACCS}(ind_i, a_i))$$

where $DDF_{dam\_cat}^{NACCS}(ind_i, a_i)$ is the point estimate of damage that the NACCS DDF for a given damage category provides for a given inundation and structure archetype.

UNSAFE also allows users to estimate damages using HAZUS damage functions under uncertainty. There are many expert-based DDFs in the HAZUS library. We use damage functions corresponding to the following DDF ids: 106, 108, 114, 116, 129, 130. These are the most recently specified DDFs in the HAZUS library for the structures UNSAFE can estimate losses under uncertainty for. There are six ids because HAZUS uses different damage functions depending on whether a structure is located in or outside of a National Flood Hazard Layer "V" zone. Recent evidence suggests that the HAZUS DDFs for one-story houses without a basement roughly capture mean damage for a given inundation, but there is substantial heteroskedastic variation in the empirical depth-damage relationship (Wing et al. 2020). We reflect this finding with the following functional form:

$$l_i \sim uniform(DDF^{HAZUS}(ind_i, a_i) - \epsilon * DDF^{HAZUS}(ind_i, a_i), DDF^{HAZUS}(ind_i, a_i) + \epsilon * DDF^{HAZUS}(ind_i, a_i))$$

where analysts can choose the parameter ϵ to control the uncertainty bounds (monotonically increasing with depth) around the central DDF^{HAZUS} point estimate. This approach was used in (Zarekarizi, Srikrishnan, and Keller 2020) with $\epsilon = .3$.

Current functionality in UNSAFE

UNSAFE includes several modules. First, we provide an overview of the intended workflow these modules support. Then, we describe aspects of how these modules are implemented. It may be helpful to follow along with one of the case studies in `examples/philadelphia_frd/`.

Using UNSAFE in a property-level flood-loss estimation workflow

UNSAFE v0.1 is designed to estimate flood losses for single family residences with one or two stories, and basement, crawl space, or slab foundation. In the U.S., single family houses are the most common building type and comprise a large proportion of overall flood risk Wing et al. (2022). Expert DDFs exist for one and two story single family houses. Sometimes, the DDFs designed for two stories are interpreted as being for structures with two or more stories, but we avoid this area of uncertainity for our initial implementation.

UNSAFE can esitmate an ensemble of plausible flood losses for exposed structures for a single event. Users can also input so-called design events, which are a standard way to define flood events that correspond to a probability of occurrence, to estimate expected annual losses.

The intended workflow, demonstrated in the `examples/philadelphia_frd/` case studies is:

1) Configure the working directory structure and workflow parameters
2) If you want to make use of the downloading functionality in UNSAFE, download and unzip data
3) Subset the full structure inventory to single-family structures and convert the data to a GeoDataFrame
4) Specify a spatial extent of your study area (i.e. county shapefile or a study boundary) and process reference files through clipping (e.g. census tract data)
5) Process expert DDFs for use in ensembles
6) Process social vulnerability data by linking it with corresponding reference data (e.g. linking Climate and Economic Justice Screening Tool with census tracts)
7) Prepare the National Flood Hazard Layer data for identifying structure location in and outside of the federal floodplain
8) Link structures to all vector spatial data
9) Link structures to inundation data, provided as raster(s)
10) Prepare the structure inventory for loss estimation (this base inventory can be used for estimating losses without uncertainty)
11) Generate an ensemble of plausible structure realizations based on several parameters users can specify
12) Estimate expected annual losses for each ensemble member for a set of design events. Each ensemble member has a unique draw from the DDF distribution. Users can also estimate expected annual losses without uncertainty in exposure and vulnerability.

This workflow produces output data that can be used for visual inspections of each process, and for visualizing results and performing subsequent analyses based on damage estimates.

More details on each step of the workflow

It is easiest to describe the workflow implementation with an example! There are currently two examples in `examples/philadelphia_frd/` The `partial_data_example.ipynb` includes extensive markdown to describe implementation logic and choices. It is "partial" in the sense that it includes a small subset of the hazard data from the `full_data_example.ipynb` so that you don't have to download external data before working through the example. The full data example does not include the extensive markdown documentation, but is included because it better illustrates the implication of accounting for uncertainties in exposure and vulnerability inputs, and includes more visualization examples that are helpful for flood-risk analyaes.

References

Krause, A, A Martin, and M Fix. 2020. "Uncertainty in Automated Valuation Models: Error-Based Versus Model-Based Approaches." *Journal of Property Research* 37 (4): 308–39. https://doi.org/10.1080/09599916.2020.1807587.

Mostafiz, Rubayet Bin, Carol J Friedland, Md Asif Rahman, Robert V Rohli, Eric Tate, Nazla Bushra, and Arash Taghinezhad. 2021. "Comparison of Neighborhood-Scale, Residential Property Flood-Loss Assessment Methodologies." *Front. Environ. Sci. Eng. China* 9. https://doi.org/10.3389/fenvs.2021.734294.

Nolte, Christoph, Kevin J Boyle, Anita Chaudhry, Christopher Clapp, Dennis Guignet, Hannah Hennighausen, Ido Kushner, et al. 2023. "Data Practices for Studying the Impacts of Environmental Amenities and Hazards with Nationwide Property Data." *Land Econ.*, June. https://doi.org/10.3368/le.100.1.102122-0090R.

Sanderson, Dylan, and Daniel Cox. 2023. "Comparison of National and Local Building Inventories for Damage and Loss Modeling of Seismic and Tsunami Hazards: From Parcel-to City-Scale." *International Journal of Disaster Risk Reduction* 93 (July): 103755. https://doi.org/10.1016/j.ijdrr.2023.103755.

Shultz Steven. 2017a. "Accuracy of HAZUS General Building Stock Data." *Nat. Hazards Rev.* 18 (4): 04017012. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000258.

———. 2017b. "Correcting HAZUS General Building Stock Structural Replacement Cost Data for Single-Family Residences." *Nat. Hazards Rev.* 18 (4): 04017015. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000261.

Wing, Oliver E J, William Lehman, Paul D Bates, Christopher C Sampson, Niall Quinn, Andrew M Smith, Jeffrey C Neal, Jeremy R Porter, and Carolyn Kousky. 2022. "Inequitable Patterns of US Flood Risk in the Anthropocene." *Nat. Clim. Chang.* 12 (2): 156–62. https://doi.org/10.1038/s41558-021-01265-6.

Wing, Oliver E J, Nicholas Pinter, Paul D Bates, and Carolyn Kousky. 2020. "New Insights into US Flood Vulnerability Revealed from Flood Insurance Big Data." *Nat. Commun.* 11 (1): 1444. https://doi.org/10.1038/s41467-020-15264-2.

Zarekarizi, Mahkameh, Vivek Srikrishnan, and Klaus Keller. 2020. "Neglecting Uncertainties Biases House-Elevation Decisions to Manage Riverine Flood Risks." *Nat. Commun.* 11 (1): 5361. https://doi.org/10.1038/s41467-020-19188-9.