Household displacement and return after disasters

Research supervised by Prof. Carmine Galasso, Prof. Jack Baker, and Prof. Vitor Silva

World Federation of Engineering Organizations (WFEO)

Committee on Disaster Risk Management (CDRM) Webinar

7 March 2025

NICOLE PAUL

Outline

Introduction and key context

Paul, Nicole, Carmine Galasso, and Jack Baker. 2024. "Household Displacement and Return in Disasters: A Review." Natural Hazards Review 25 (1): 03123006. https://doi.org/10.1061/NHREFO.NHENG-1930

Understanding the role of housing damage

Paul, Nicole, Carmine Galasso, Vitor Silva, and Jack Baker. 2024. "Population Displacement after Earthquakes: Benchmarking Predictions Based on Housing Damage." Seismica 3 (2). https://doi.org/10.26443/seismica.v3i2.1374.

Understanding displacement durations

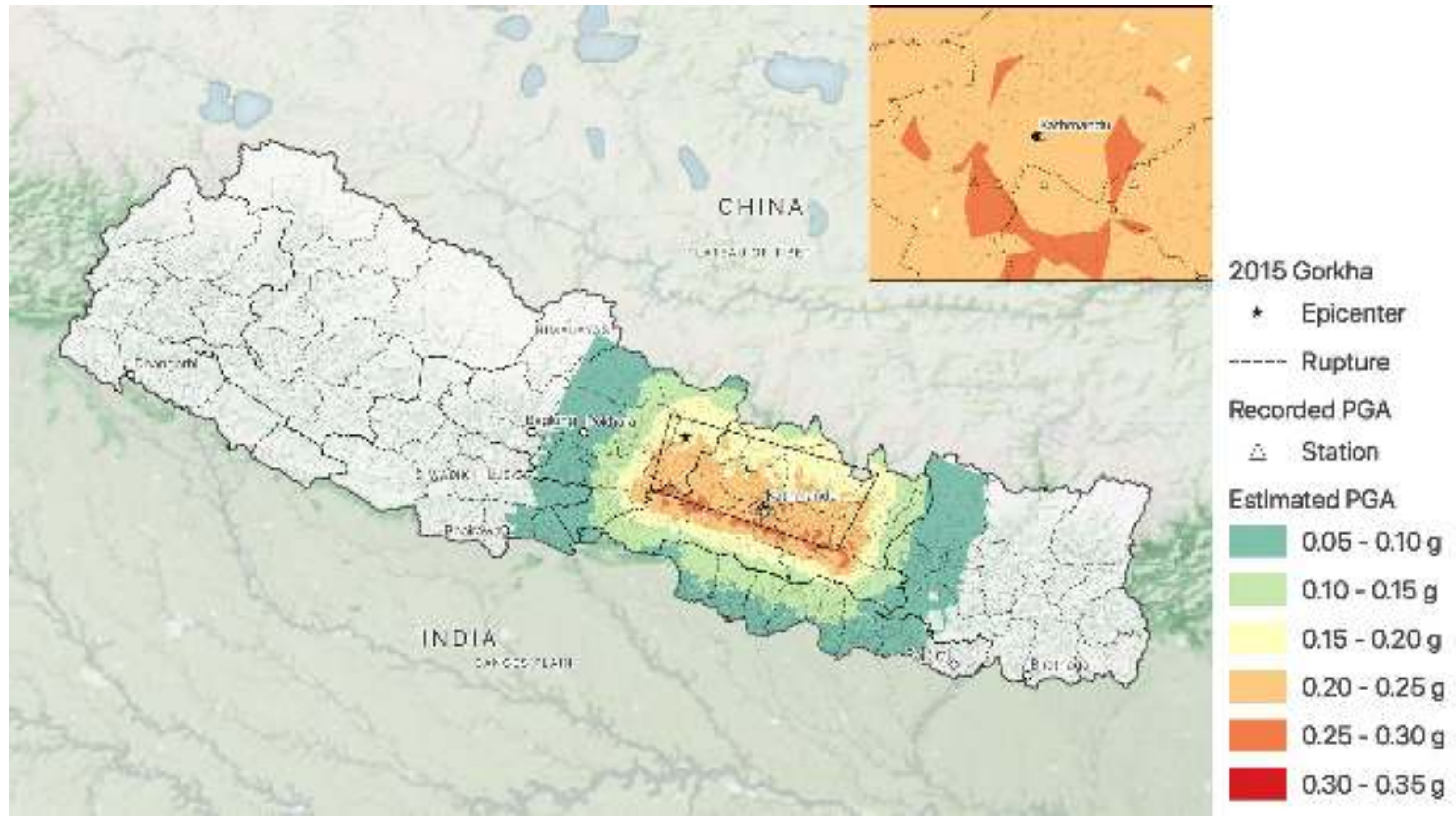
Paul, Nicole, Carmine Galasso, Jack Baker, and Vitor Silva. 2025. "A predictive model for household displacement duration after disasters." Risk Analysis. https://onlinelibrary.wiley.com/doi/10.1111/risa.17710

What's coming up next?

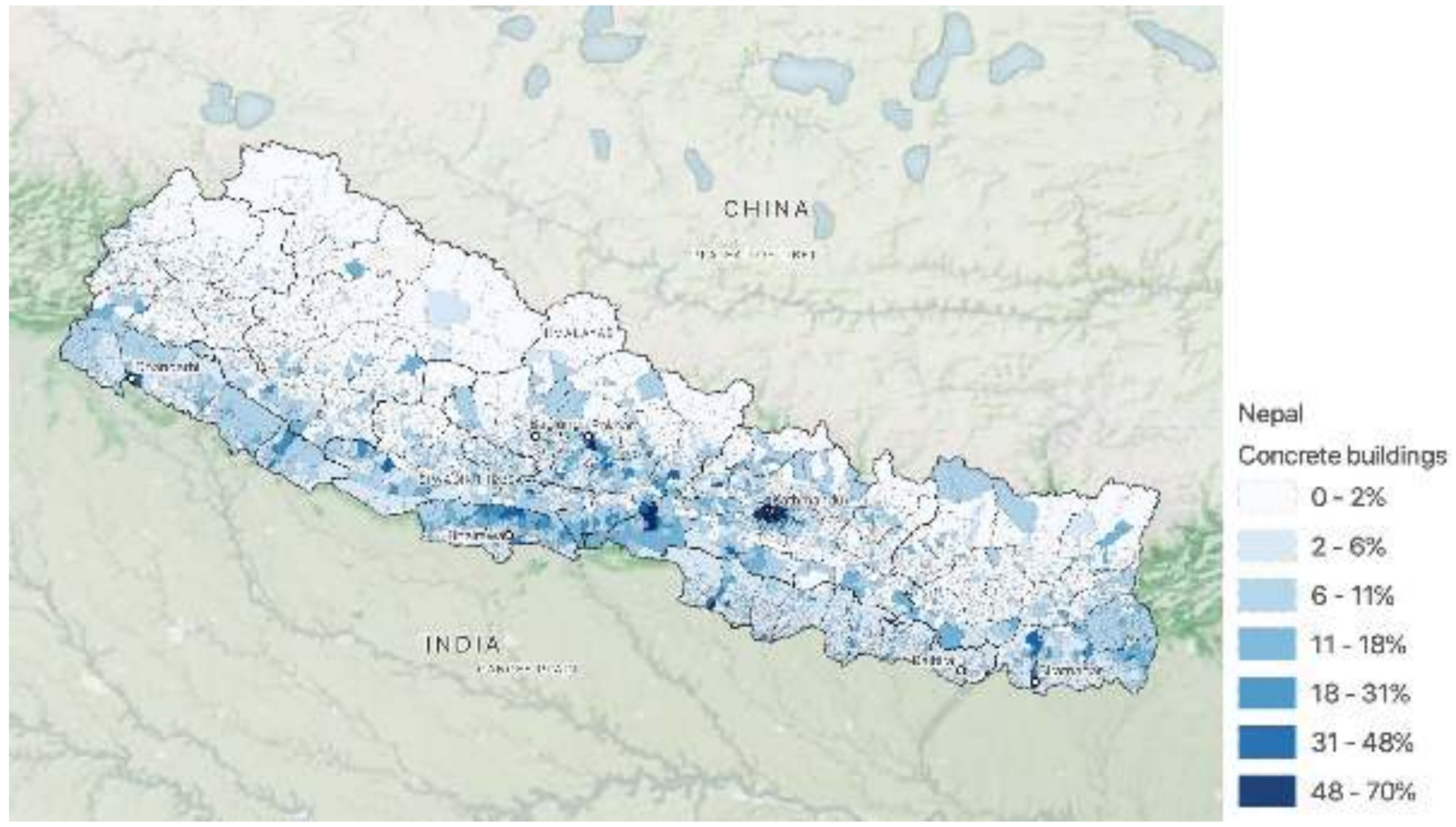
All papers are open access and are linked on my website: nicolepaul.io

A primer on disaster risk modeling

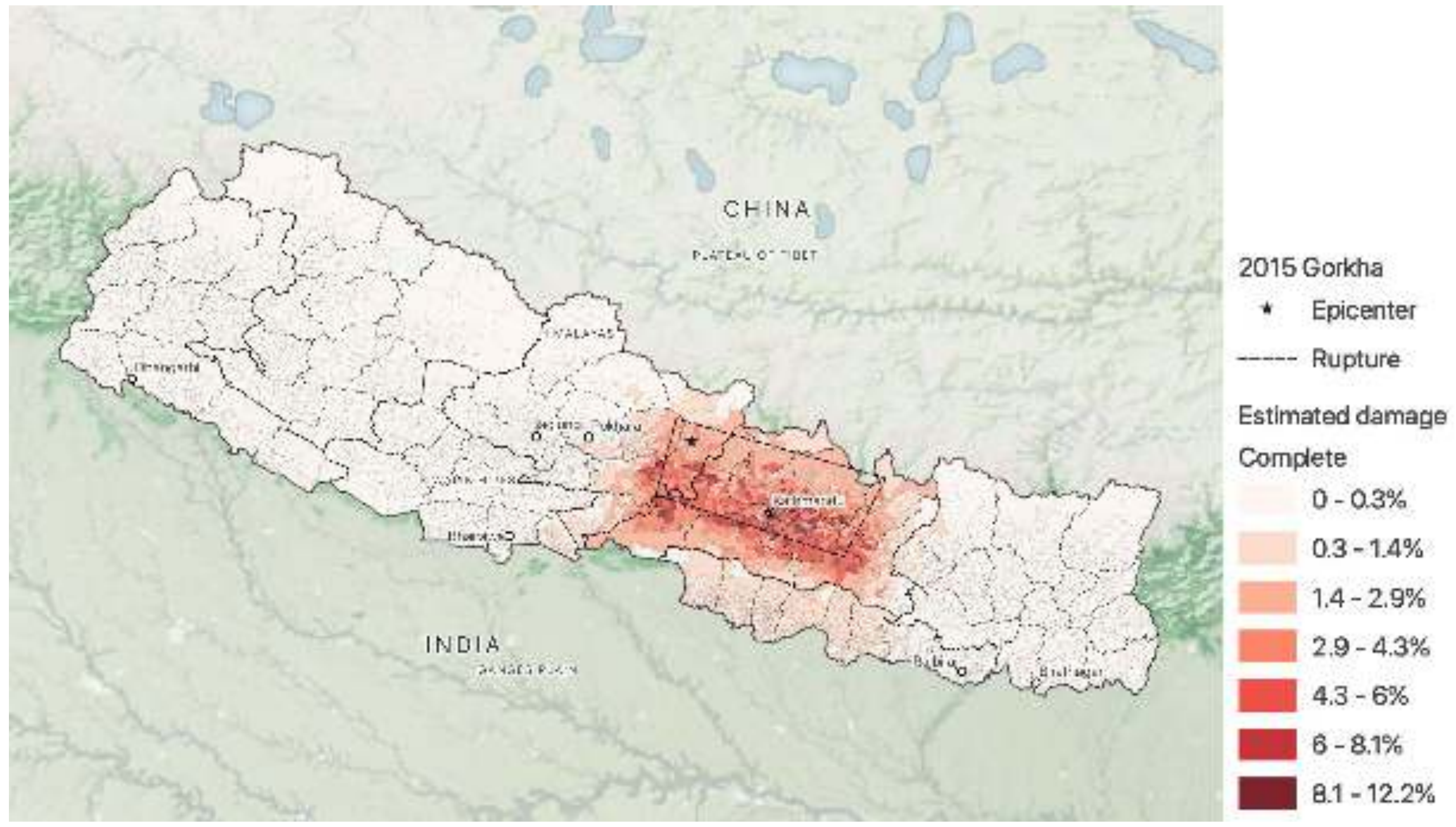
Hazard



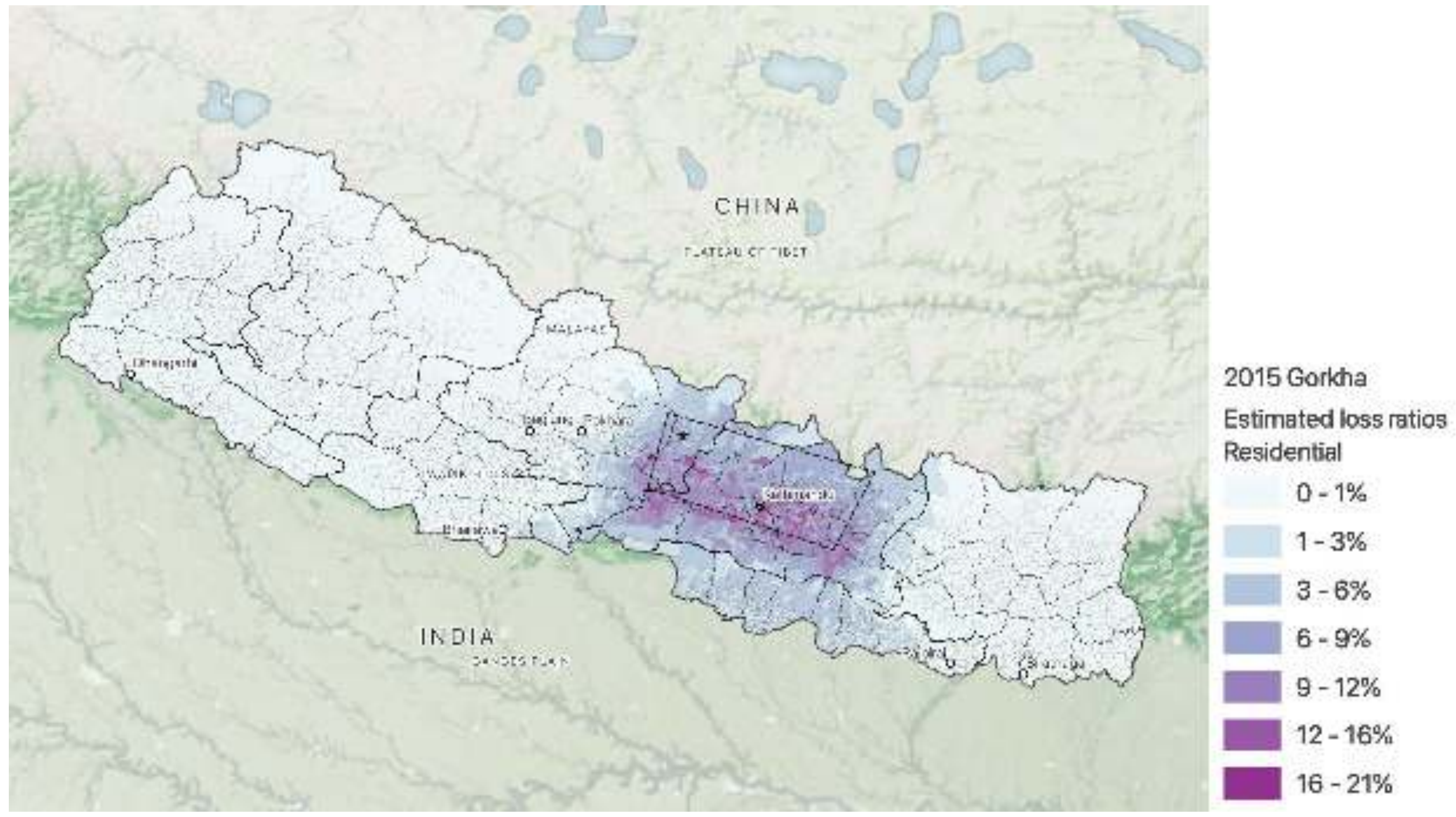
Exposure



Damage



Loss



Introduction

Paul, Nicole, Carmine Galasso, and Jack Baker. 2024. "Household Displacement and Return in Disasters: A Review." Natural Hazards Review 25 (1): 03123006. https://doi.org/10.1061/NHREFO.NHENG-1930

Why study population displacement?

The scale of human impact

In 2020 alone, at least 30 million were displaced due to natural hazards. By the end of the year, at least 7 million were still displaced (IDMC 2021)

Increasing risk under current trends

The annual number displaced from disasters is expected to increase, driven by population growth in hazard-prone areas and exacerbated by climate change

A more equitable risk metric

Existing disaster risk assessments tend to focus on economic loss, a metric that often highlights the wealthiest as the most at-risk

Scope and definitions

Residential migration

The movement of households from their habitual residence to other housing

Displacement (Involuntary migration)

The forced movement of households from their habitual residence,

as triggered by:

Voluntary migration

The voluntary movement of households (e.g., economic opportunity)

Disasters

due to geophysical and weather-related hazards

Conflict & violence

due to bombardments, threats of armed attack, gang harassment

Other categories

due to development projects, evictions, policies, etc.

Sudden onset

Emerges quickly (e.g. earthquakes)

Slow onset

Emerges gradually (e.g. droughts)

The importance of duration

Event occurs

Minutes to days ahead → **Days to months after** → **Months to years after**

Early warning

People are pre-emptively evacuated and lives are potentially saved

Emergency

People are temporarily displaced and seek refuge with family and friends, in rentals or hotels, and in public shelters

Recovery

People consider waiting while housing and infrastructure are repaired or resettling elsewhere

Consequences of protracted displacement

Protracted displacement is associated with negative consequences for households and the community

- Loss of livelihood
- Disrupted education
- Reduced supports
- Psychosocial issues
- Worsened housing

Household decisions to return

Physical damage

Habitability of housing (damage, weather, utilities)
Housing type
Community damage
Reconstruction progress

Psychological & social phenomena

Acceleration of ongoing trends
Attachment to place
Social capital (networks, family/friends)
Perceived risk

Household demographics

Socioeconomic status (e.g., income)
Housing and land tenure
Race/ethnicity/caste
Age

Pre- and post-disaster policies

Pre-existing housing conditions (e.g., vacancies)
Housing reconstruction approaches
Other disaster assistance policies

The role of housing damage

Paul, Nicole, Carmine Galasso, Vitor Silva, and Jack Baker. 2024. "Population Displacement after Earthquakes: Benchmarking Predictions Based on Housing Damage." Seismica 3 (2). https://doi.org/10.26443/seismica.v3i2.1374.

What is the standard practice?

Despite the range of factors that influence household return identified in the literature review, standard practice is to consider just housing damage

Destroyed homes × **Average household size** = **Displaced population**

Selecting past earthquake scenarios

Recency of the earthquake event

The exposure models used for the scenario analysis are representative of the year 2021 and would not be representative for older events

Diverse geographic coverage

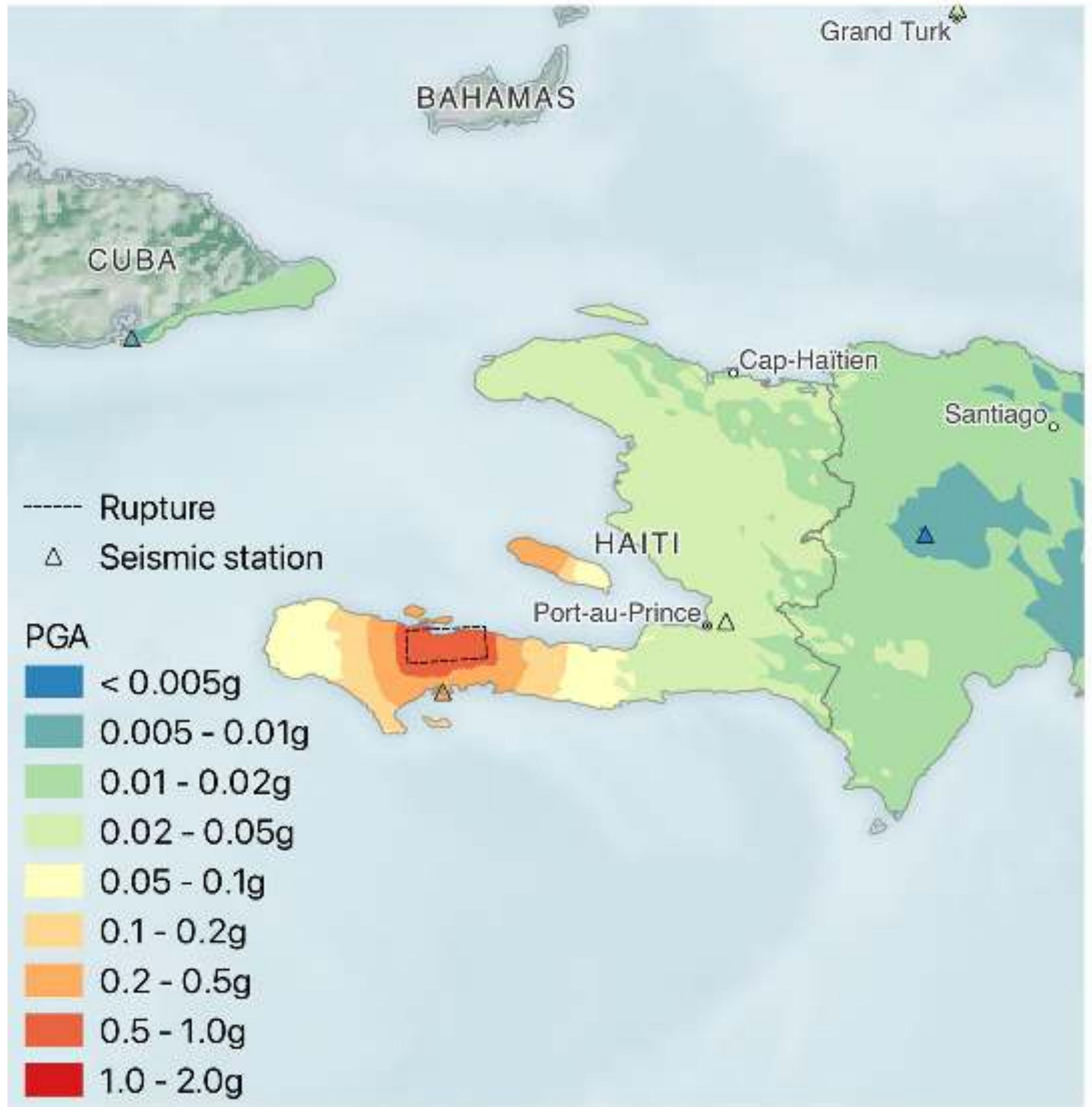
Locations were chosen to cover different tectonic regions, standard construction practices, and levels of data availability

Availability of mobile location data-based estimates

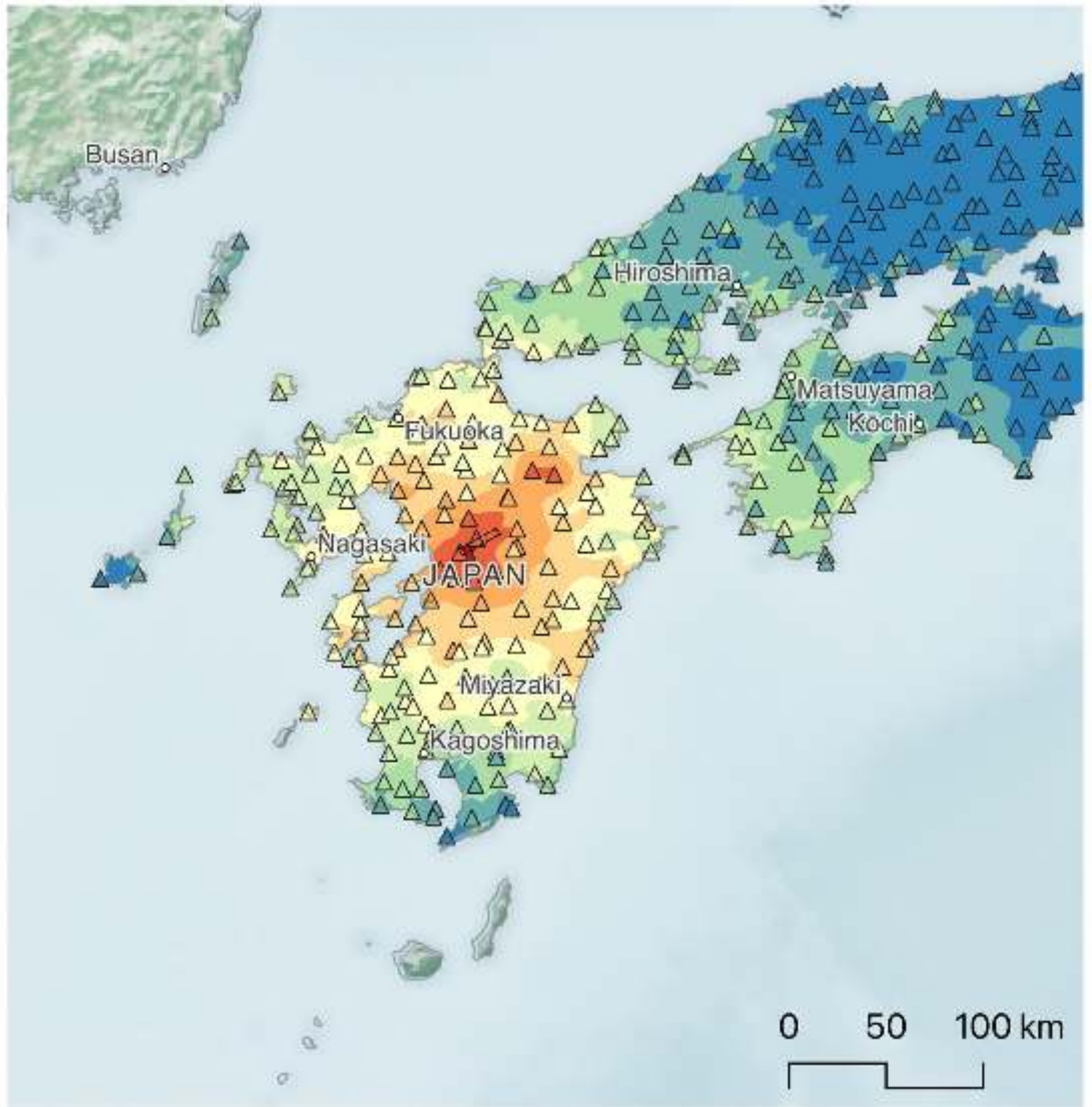
Most estimates assume housing destruction as the primary driver, whereas mobile location estimates do not rely on this assumption

Selecting past earthquake scenarios

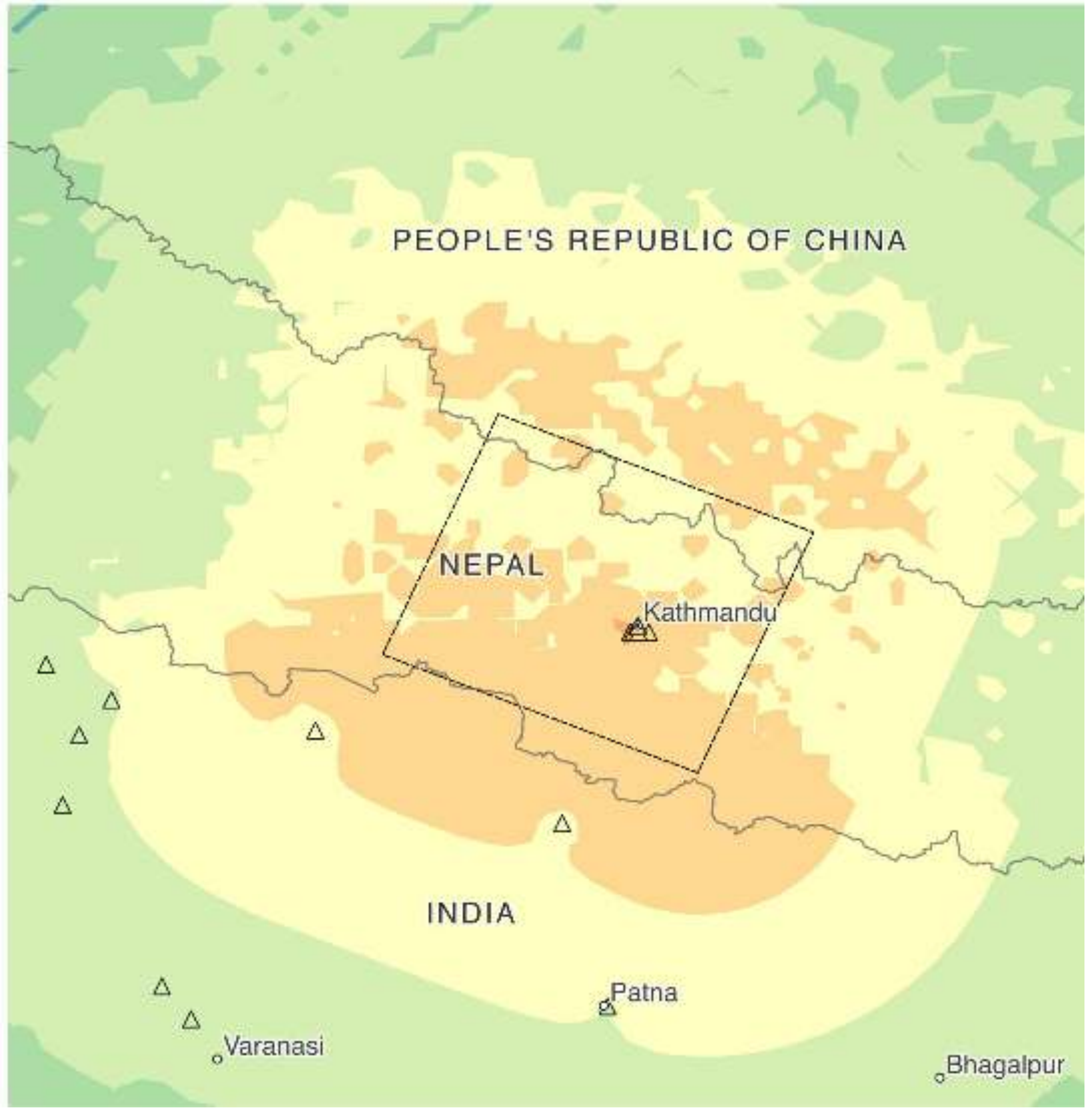
Haiti
2021 M_W7.2 Nippes



Japan
2016 M_W7.0 Kumamoto



Nepal
2015 M_W7.8 Gorkha

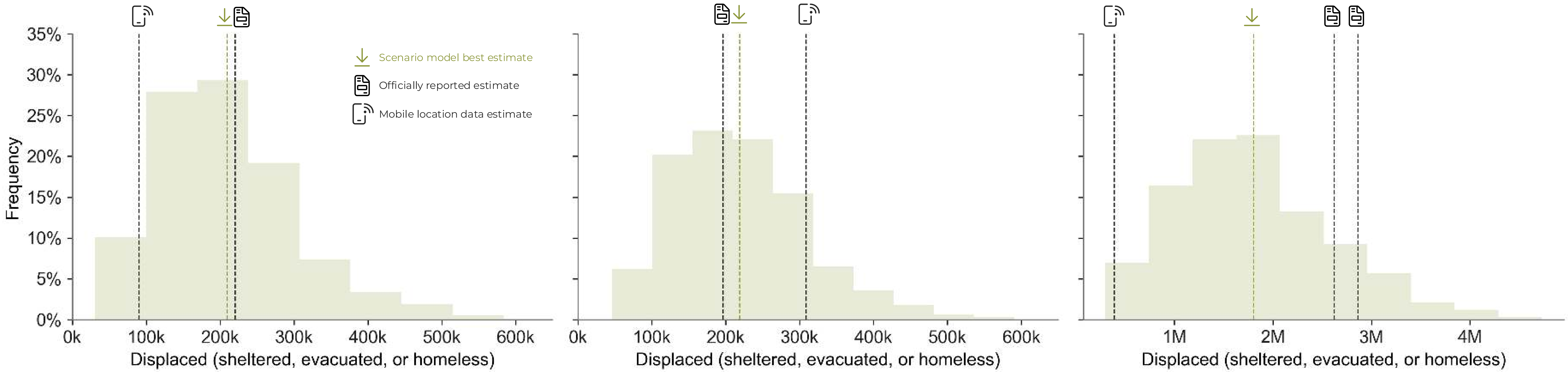


Benchmarking displaced estimates

Haiti 2021 M_W7.2 Nippes

Japan 2016 M_W7.0 Kumamoto

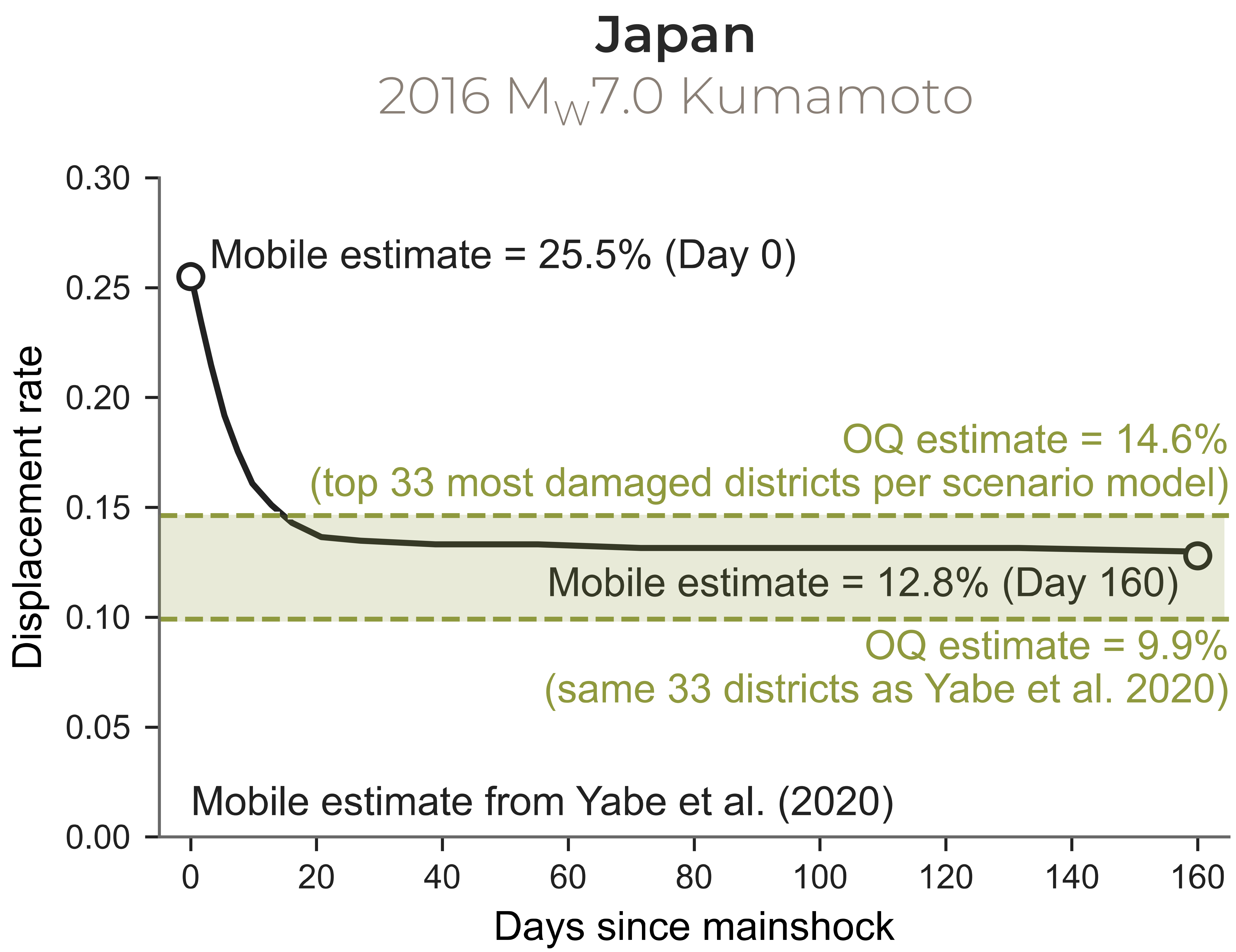
Nepal 2015 M_W7.8 Gorkha



The scenario model estimates were largely consistent with the official reports, albeit with a broad range of uncertainty.

The mobile location data estimates were closer to the distribution tails

Benchmarking displaced estimates over time



Estimating displacement using only housing damage seems to estimate potential long-term housing needs realistically

However, this approach offers no view on displacement duration or return

Understanding the role of damage

Scenario model estimates based on housing damage show some promise

The models using damage as a displacement driver were consistent with official reports and long-term mobile location estimates, but have large uncertainty

Quantification of displacement duration remains a challenge

Official reports lack information on displacement over time, and the model estimates using housing damage similarly lack a time component

Mobile location data-based estimates require further benchmarking

Mobile location data could fill the data gap on duration, but further investigation on the displacement criterion & sample representativeness is needed

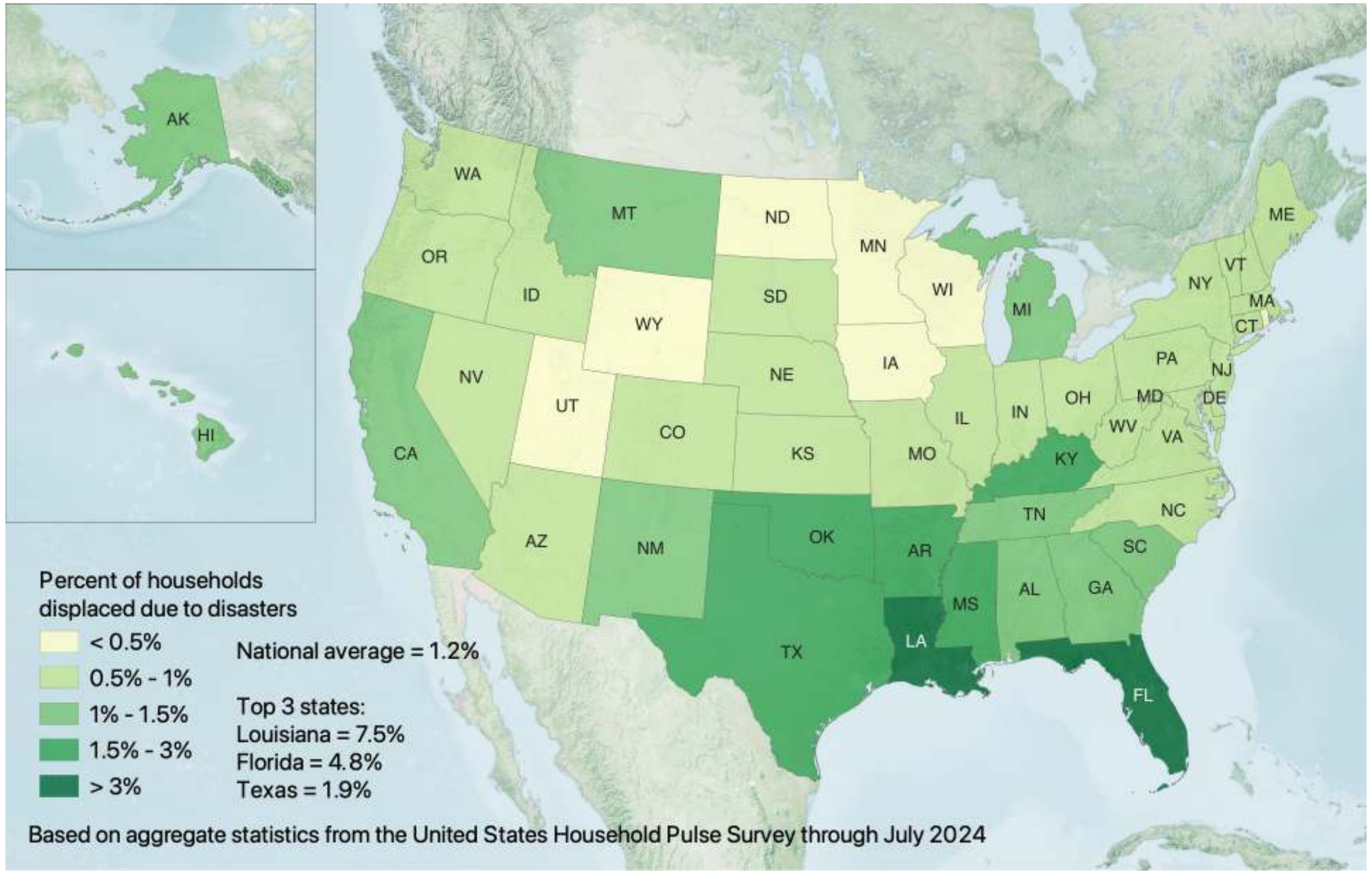
The duration of displacement

Paul, Nicole, Carmine Galasso, Jack Baker, and Vitor Silva. 2025. "A predictive model for household displacement duration after disasters." Risk Analysis.
https://onlinelibrary.wiley.com/doi/10.1111/risa.17710

Household displacement in US disasters

Since 2021, 1.1% of households have been displaced by disasters in the US

Proportion of households that were displaced



Household displacement in US disasters

Since 2021, 1.1% of households have been displaced by disasters in the US

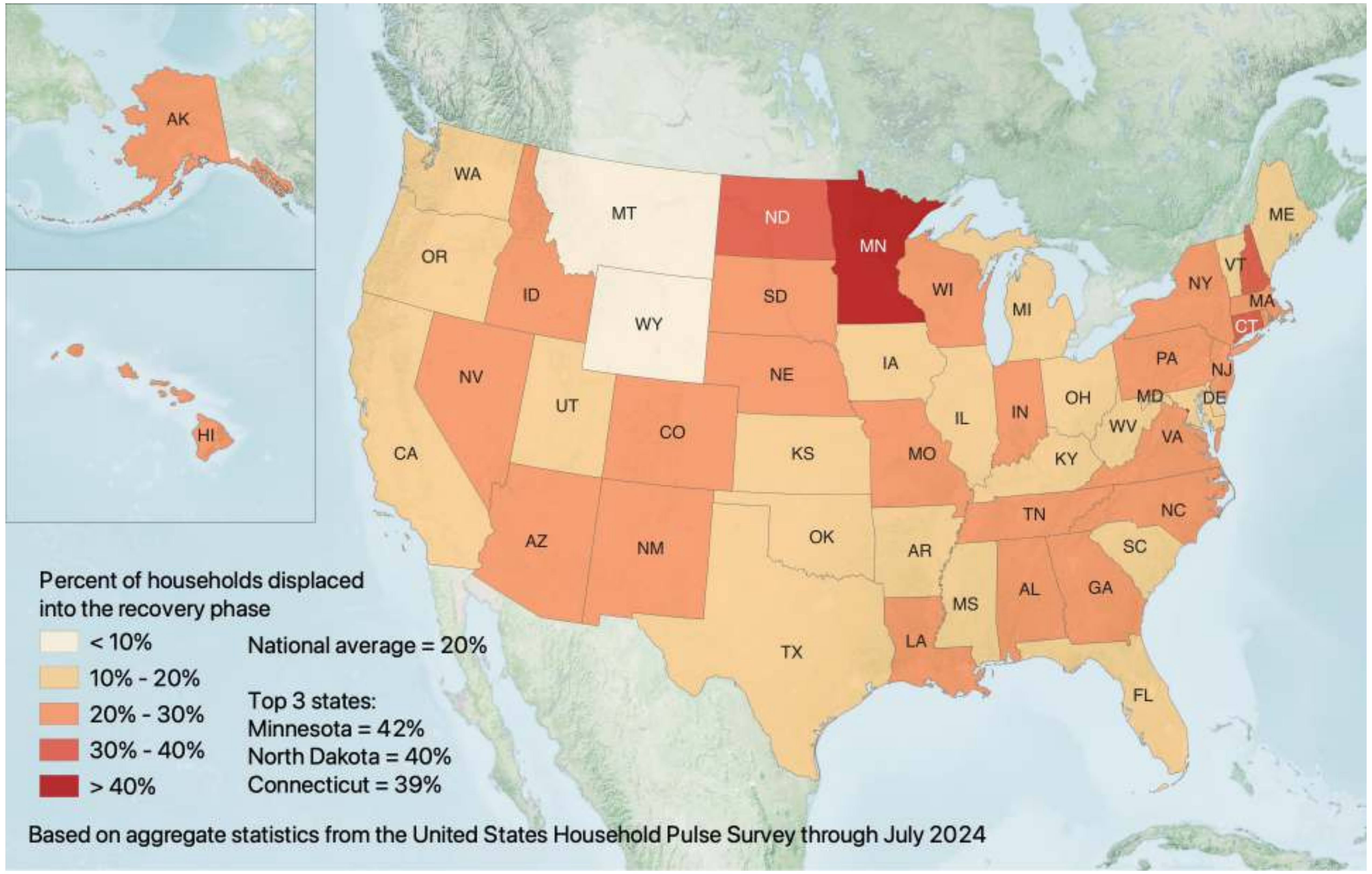
Most households returned quickly

- Within a week: 43%
- Within a month: 23%

Others faced protracted displacement

- One to six months: 12%
- Over six months: 8%
- Not returned: 14%

Proportion of displaced households that took >1mo to return



Household displacement in US disasters

Since 2021, 1.1% of households have been displaced by disasters in the US

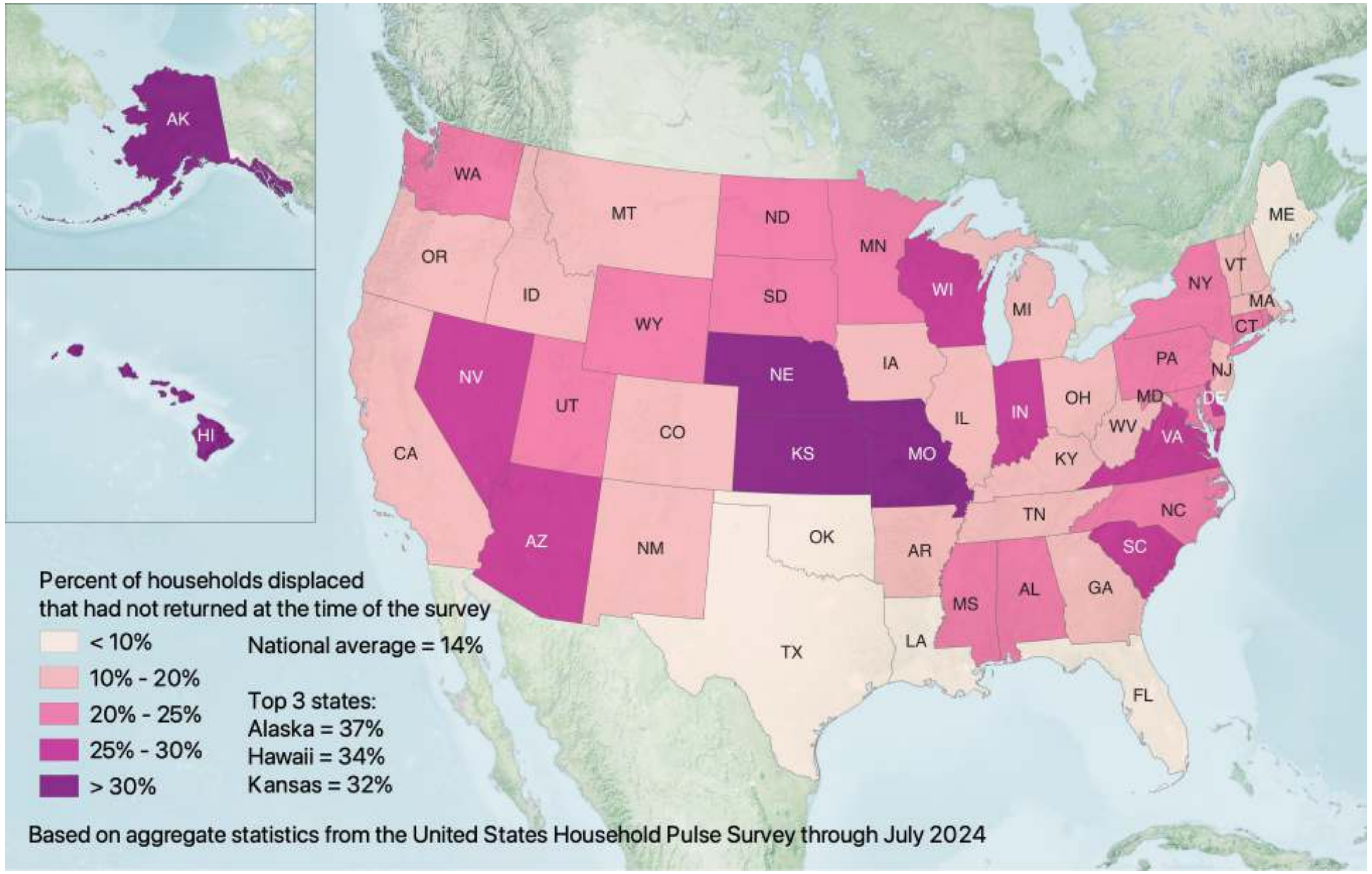
Most households returned quickly

- Within a week: 43%
- Within a month: 23%

Others faced protracted displacement

- One to six months: 12%
- Over six months: 8%
- Not returned: 14%

Proportion of displaced households that have not returned

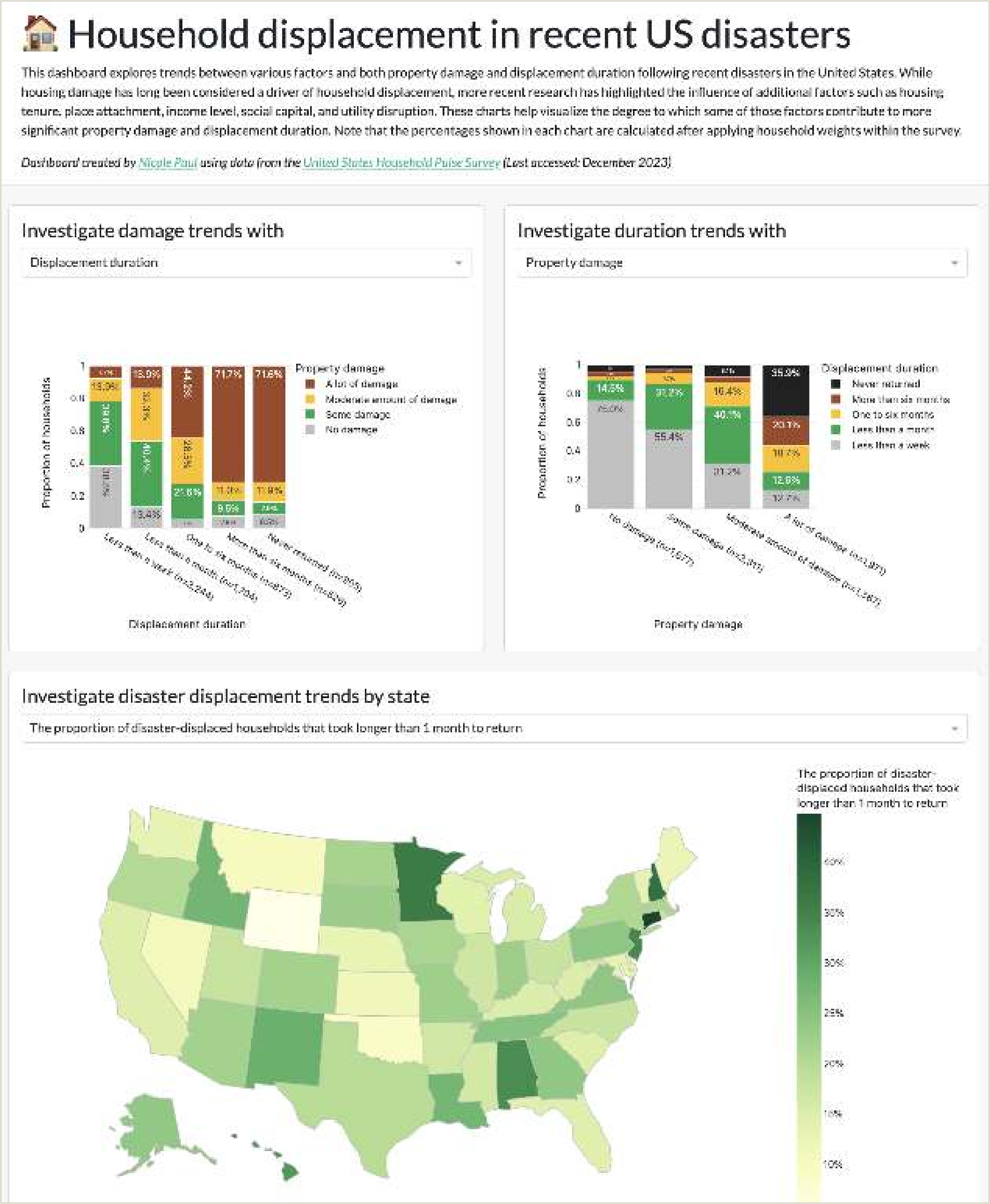


Exploring trends and fitting predictive models

The availability of microdata allows us to explore trends various factors have with displacement durations:

- Property damage
- Lifeline disruption
- Household demographics
- Area-based attributes

We can also fit predictive models for household displacement durations and evaluate their performance



hps.nicolepaul.io

Predicting household return after disasters

Output variable

Displacement duration

1. Emergency phase
   (return in less than 1 month)
2. Recovery phase
   (return after 1 month)
3. Not returned

Input variables

Physical factors

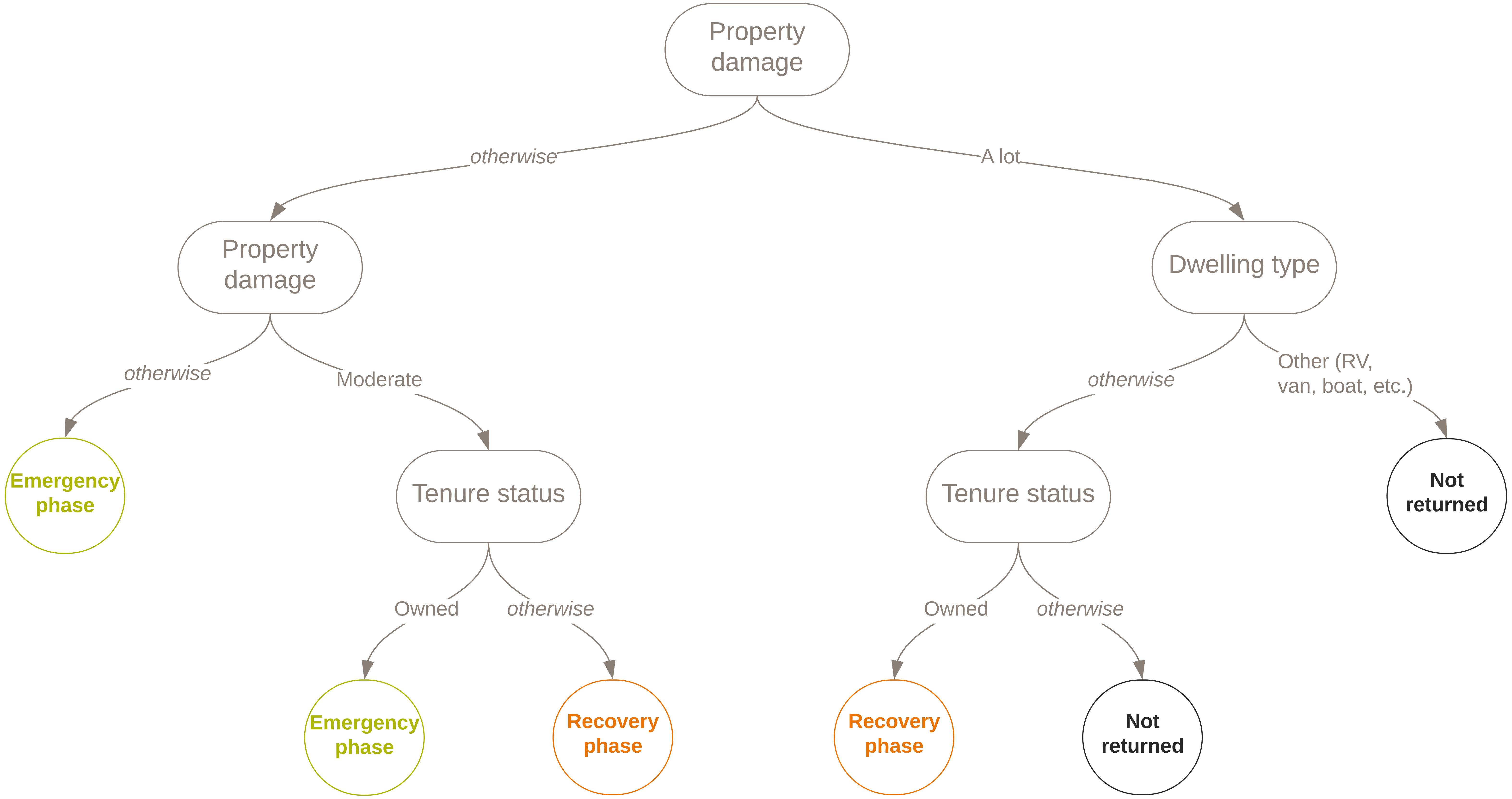
- Property damage
- Lifeline disruption (electricity loss, water shortage, unsanitary conditions, food shortage)
- Dwelling type
- Hazard type

Socioeconomic factors

- Household demographics (e.g., income level, tenure, race/ethnicity, age, size)
- Area-based statistics (e.g., vacancy rates, number of disaster declarations, unemployment rate)

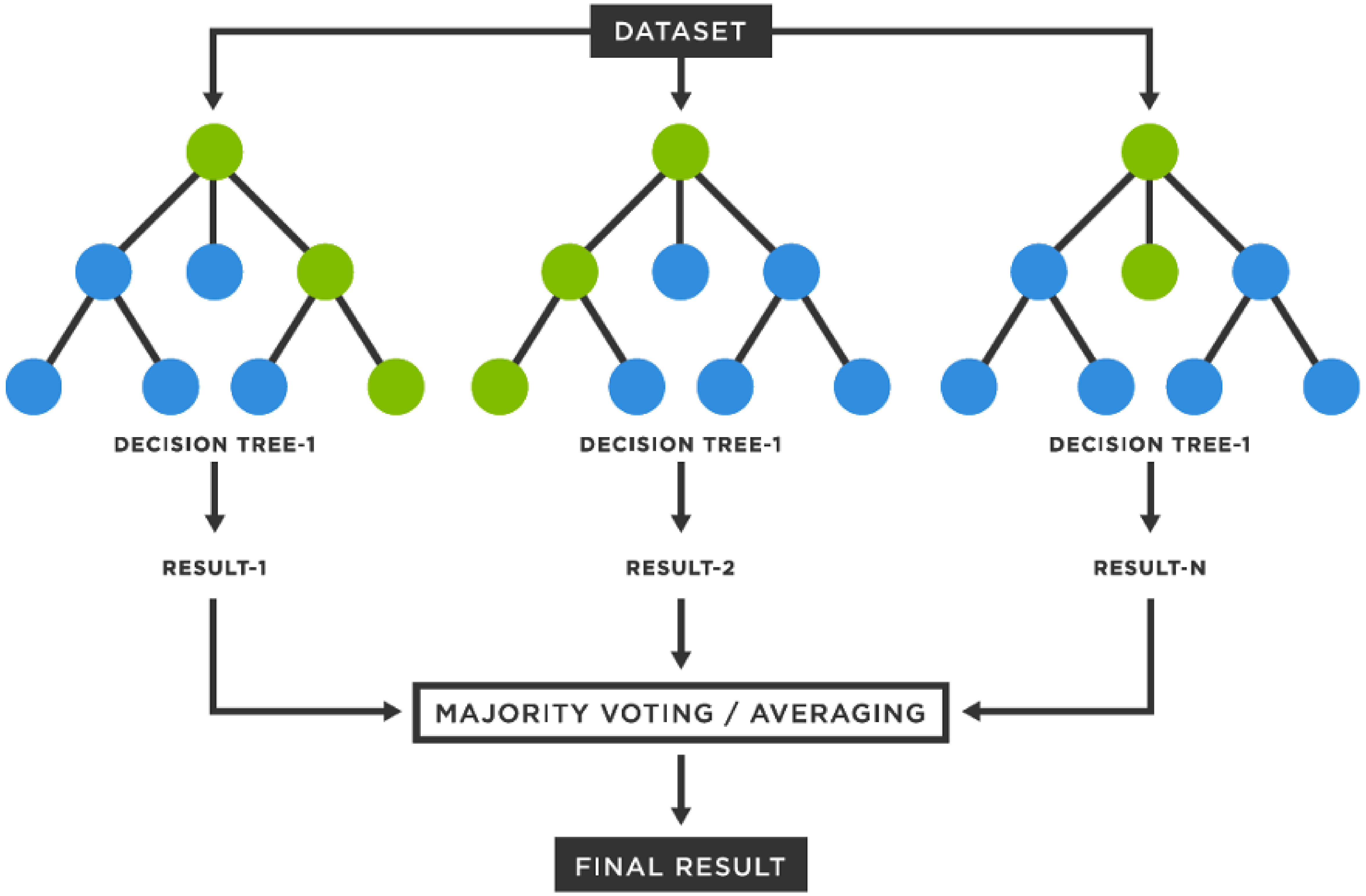
Classification tree | Simple model for risk analysis

Classification trees allow straightforward implementation within disaster risk analyses, allowing us to restrict the number of predictors



Random forest | Adding complexity

Random forest models can incorporate all considered factors and improve predictions relative to individual trees



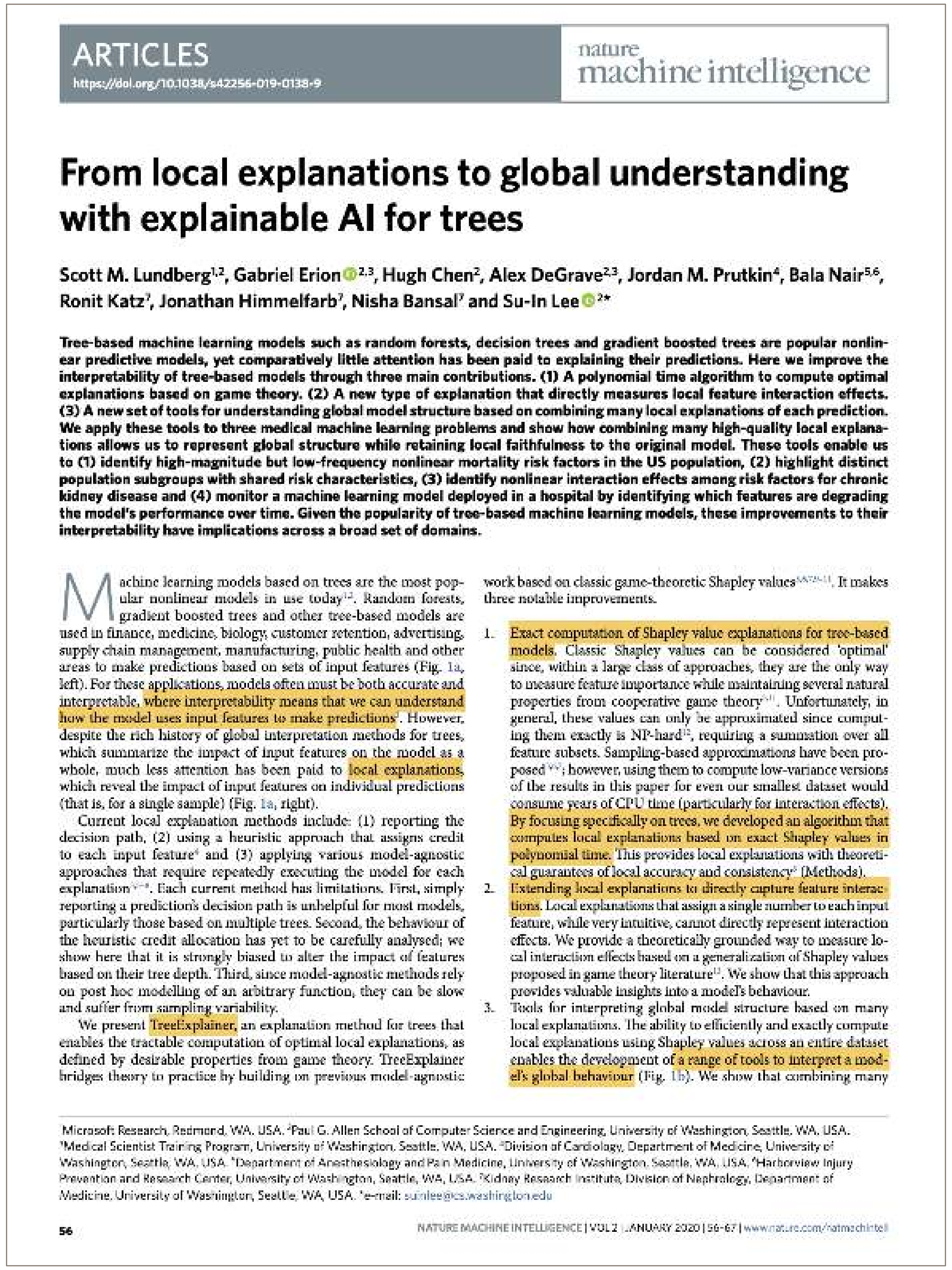
Source: https://www.spotfire.com/glossary/what-is-a-random-forest

Random forest | Explaining complexity

Recent advances in explainable AI have aimed to improve the interpretability of machine learning models

Shapley Additive exPlanations (SHAP) have been proposed to quantify the marginal contribution of individual features on model predictions.
SHAP can also account for interaction effects.

We calculate SHAP values for each variable, such that
Final prediction = Baseline prediction + Σ(SHAP values)



ARTICLES
https://doi.org/10.1038/s42256-019-0138-9
nature machine intelligence

From local explanations to global understanding with explainable AI for trees

Scott M. Lundberg[1,2], Gabriel Erion[2,3], Hugh Chen[2], Alex DeGrave[2,3], Jordan M. Prutkin[4], Bala Nair[5,6], Ronit Katz[7], Jonathan Himmelfarb[7], Nisha Bansal[7] and Su-In Lee[2*]

Tree-based machine learning models such as random forests, decision trees and gradient boosted trees are popular nonlinear predictive models, yet comparatively little attention has been paid to explaining their predictions. Here we improve the interpretability of tree-based models through three main contributions. (1) A polynomial time algorithm to compute optimal explanations based on game theory. (2) A new type of explanation that directly measures local feature interaction effects. (3) A new set of tools for understanding global model structure based on combining many local explanations of each prediction. We apply these tools to three medical machine learning problems and show how combining many high-quality local explanations allows us to represent global structure while retaining local faithfulness to the original model. These tools enable us to (1) identify high-magnitude but low-frequency nonlinear mortality risk factors in the US population, (2) highlight distinct population subgroups with shared risk characteristics, (3) identify nonlinear interaction effects among risk factors for chronic kidney disease and (4) monitor a machine learning model deployed in a hospital by identifying which features are degrading the model's performance over time. Given the popularity of tree-based machine learning models, these improvements to their interpretability have implications across a broad set of domains.

Machine learning models based on trees are the most popular nonlinear models in use today[1,2]. Random forests, gradient boosted trees and other tree-based models are used in finance, medicine, biology, customer retention, advertising, supply chain management, manufacturing, public health and other areas to make predictions based on sets of input features (Fig. 1a, left). For these applications, models often must be both accurate and interpretable, where interpretability means that we can understand how the model uses input features to make predictions[3]. However, despite the rich history of global interpretation methods for trees, which summarize the impact of input features on the model as a whole, much less attention has been paid to local explanations, which reveal the impact of input features on individual predictions (that is, for a single sample) (Fig. 1a, right).

Current local explanation methods include: (1) reporting the decision path, (2) using a heuristic approach that assigns credit to each input feature[4] and (3) applying various model-agnostic approaches that require repeatedly executing the model for each explanation[5–8]. Each current method has limitations. First, simply reporting a prediction's decision path is unhelpful for most models, particularly those based on multiple trees. Second, the behaviour of the heuristic credit allocation has yet to be carefully analysed; we show here that it is strongly biased to alter the impact of features based on their tree depth. Third, since model-agnostic methods rely on post hoc modelling of an arbitrary function, they can be slow and suffer from sampling variability.

We present TreeExplainer, an explanation method for trees that enables the tractable computation of optimal local explanations, as defined by desirable properties from game theory. TreeExplainer bridges theory to practice by building on previous model-agnostic work based on classic game-theoretic Shapley values[5,6,7,9–11]. It makes three notable improvements.

1. Exact computation of Shapley value explanations for tree-based models. Classic Shapley values can be considered 'optimal' since, within a large class of approaches, they are the only way to measure feature importance while maintaining several natural properties from cooperative game theory[9–11]. Unfortunately, in general, these values can only be approximated since computing them exactly is NP-hard[12], requiring a summation over all feature subsets. Sampling-based approximations have been proposed[5,7]; however, using them to compute low-variance versions of the results in this paper for even our smallest dataset would consume years of CPU time (particularly for interaction effects). By focusing specifically on trees, we developed an algorithm that computes local explanations based on exact Shapley values in polynomial time. This provides local explanations with theoretical guarantees of local accuracy and consistency[5] (Methods).
2. Extending local explanations to directly capture feature interactions. Local explanations that assign a single number to each input feature, while very intuitive, cannot directly represent interaction effects. We provide a theoretically grounded way to measure local interaction effects based on a generalization of Shapley values proposed in game theory literature[13]. We show that this approach provides valuable insights into a model's behaviour.
3. Tools for interpreting global model structure based on many local explanations. The ability to efficiently and exactly compute local explanations using Shapley values across an entire dataset enables the development of a range of tools to interpret a model's global behaviour (Fig. 1b). We show that combining many

[1]Microsoft Research, Redmond, WA, USA. [2]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. [3]Medical Scientist Training Program, University of Washington, Seattle, WA, USA. [4]Division of Cardiology, Department of Medicine, University of Washington, Seattle, WA, USA. [5]Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, WA, USA. [6]Harborview Injury Prevention and Research Center, University of Washington, Seattle, WA, USA. [7]Kidney Research Institute, Division of Nephrology, Department of Medicine, University of Washington, Seattle, WA, USA. *e-mail: suinlee@cs.washington.edu

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." Nature Machine Intelligence 2 (1): 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Explaining individual return predictions

Household #202

	Emergency phase (return within 1 month)	**Recovery phase** (return after 1 month)	**Not returned**
Baseline probability	33%	33%	33%
Property damage = Some	+20%	-12%	-8%
Income per HH member = $50-100k	+3%	--	-3%
Tenure status = Owner	+1%	+1%	-2%
Dwelling type = Single-family	+1%	--	-1%
Food shortage = Not at all	+2%	--	-2%
Hazard type = Hurricane	+2%	--	-2%
All other factors	+1%	+3%	-3%
Final probability	63%	25%	12%

Explaining individual return predictions

Household #12

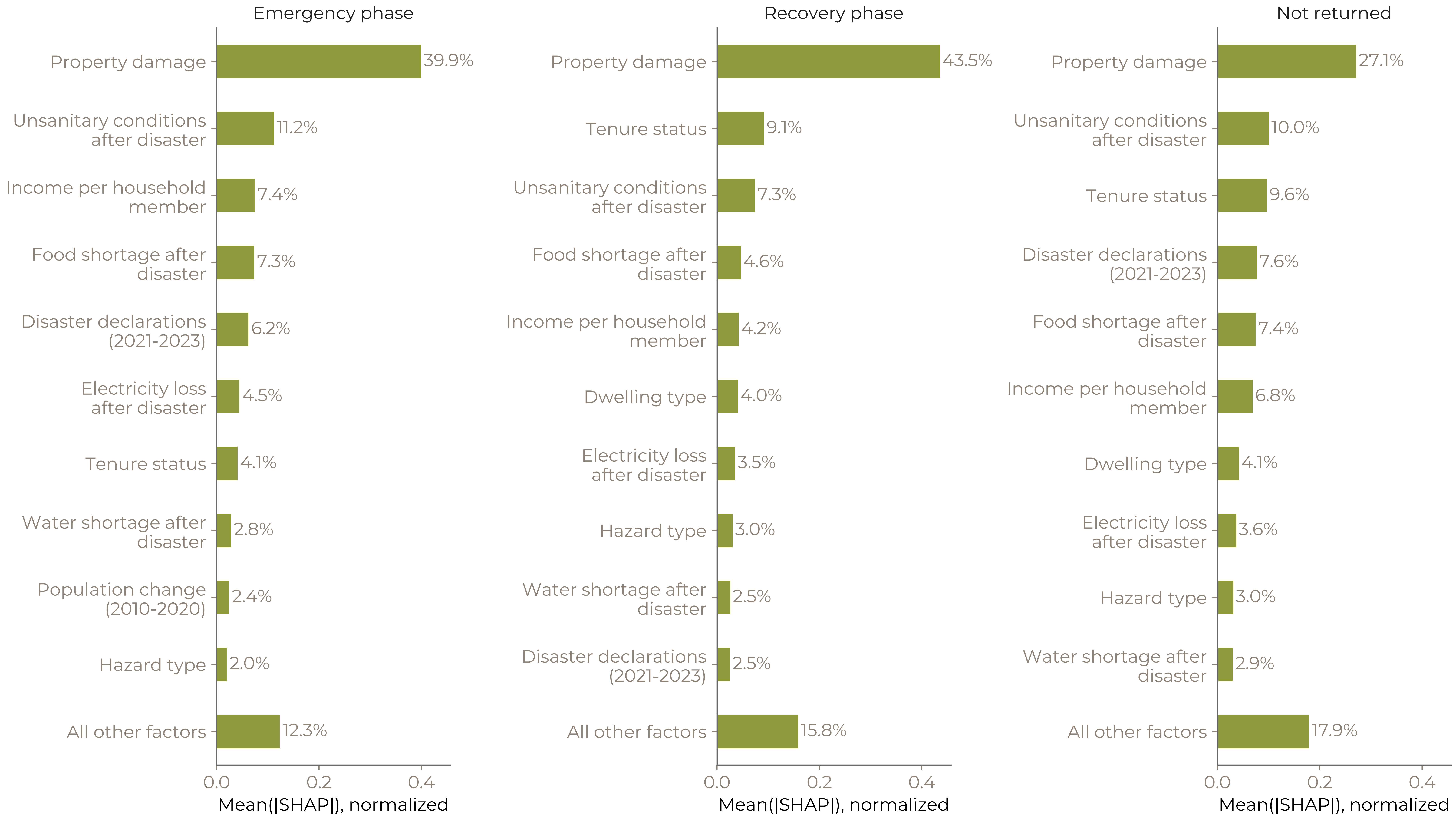
	Emergency phase (return within 1 month)	**Recovery phase** (return after 1 month)	**Not returned**
Baseline probability	33%	33%	33%
Property damage = A lot	-19%	+11%	+8%
Food shortage = A lot	-3%	--	+3%
Hazard type = Multiple	-3%	+2%	+1%
Tenure status = Owner	+1%	+2%	-3%
Homeowner vacancy rate = 1.3%	+2%	+1%	-3%
Income per HH member = $50-100k	+1%	+2%	-3%
All other factors	+11%	+3%	-13%
Final probability	23%	54%	23%

Explaining individual return predictions

Household #34

	Emergency phase (return within 1 month)	**Recovery phase** (return after 1 month)	**Not returned**
Baseline probability	33%	33%	33%
Property damage = A lot	-16%	+4%	+12%
Physical mobility = A lot of difficulty	-2%	-4%	+6%
Race = Other/mixed	-1%	-3%	+4%
Disaster declarations = 27	-3%	-1%	+4%
Tenure status = Renter	-1%	-1%	+2%
Dwelling type = Single-family	+1%	+1%	-2%
All other factors	--	-7%	+8%
Final probability	11%	22%	67%

Explaining aggregate return predictions



Understanding displacement durations

Property damage is a primary driver of displacement outcomes

Consistent with disaster literature, property damage is the number one predictor of displacement duration and return outcomes

Socioeconomic factors become more important in the recovery phase

However, some socioeconomic factors require consideration to understand the duration of household displacement, particularly tenure status and income level

Aggregate findings can obscure significant differentials at the margins

Some factors (e.g., physical immobility, large household sizes) are associated with negative outcomes, even if they are not top factors at the aggregate level

What’s coming up next?

2018 Central Sulawesi earthquake and tsunami

Collaborators

- Tayo Opabola (UC Berkeley)
- Sukiman Nurdin (Tadulako University)
- Dicky Pelupessy (University of Indonesia)
- Aulia, Reval, Shafitri, Sifa (field researchers)

Case study topics

- Drivers of relocation decisions
- Displacement duration, severity, distance
- Consequences of protracted displacement across dimensions of:
  - Standard of living
  - Livelihood
  - Wellbeing

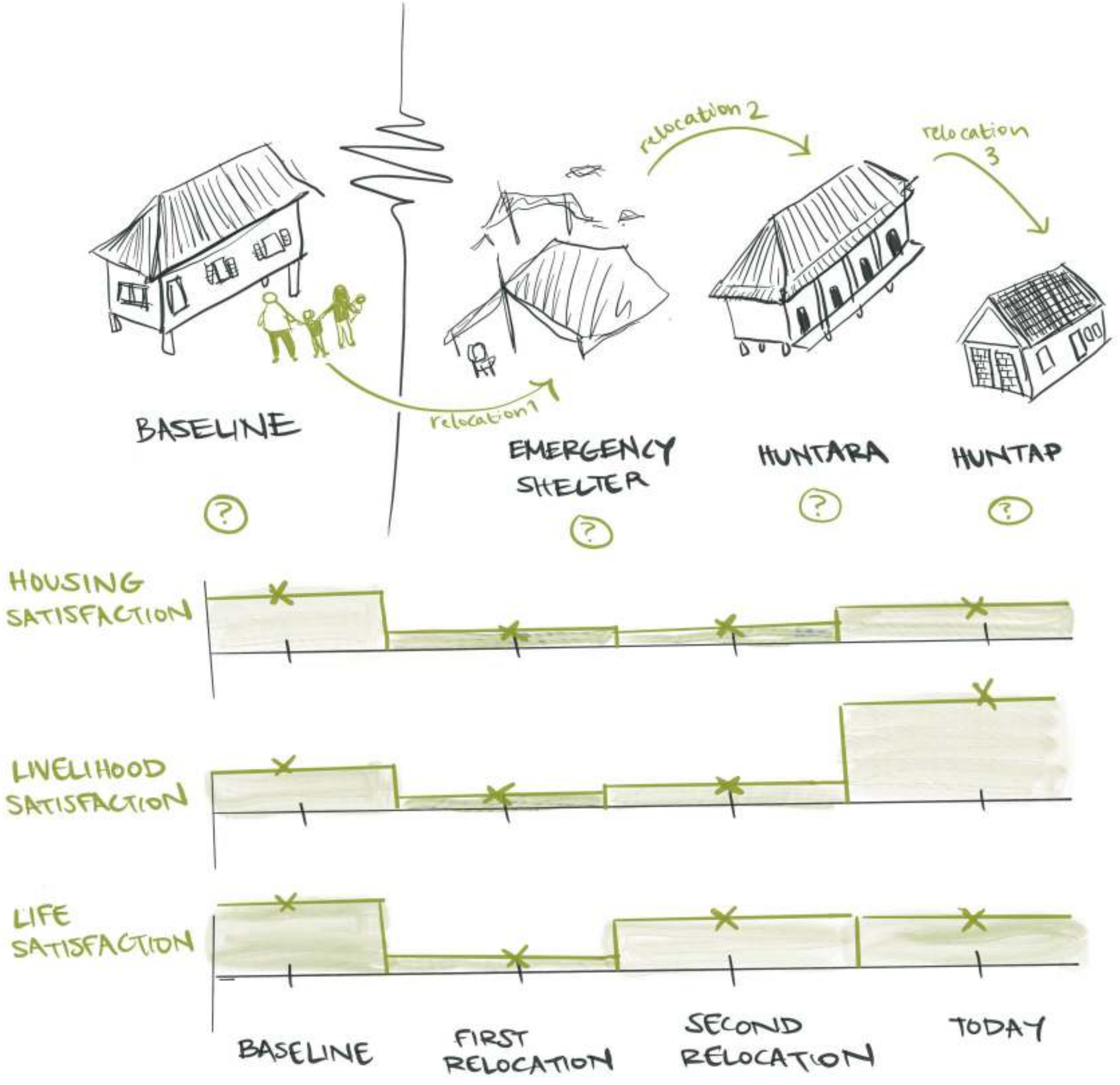
2018 Central Sulawesi earthquake and tsunami

Survey features

- Retrospective longitudinal
- ~250 households, half of which rebuilt in-situ & half of which permanently relocated

Survey components

- Baseline characteristics
- Immediate impacts (first week)
- Transitional period (each relocation until today)
- Policy and assistance



Preliminary findings

Property damage is still a primary driver of displacement outcomes

Households with moderate damage or less usually found permanent housing within a month, while households with heavy or complete damage took 2+ years

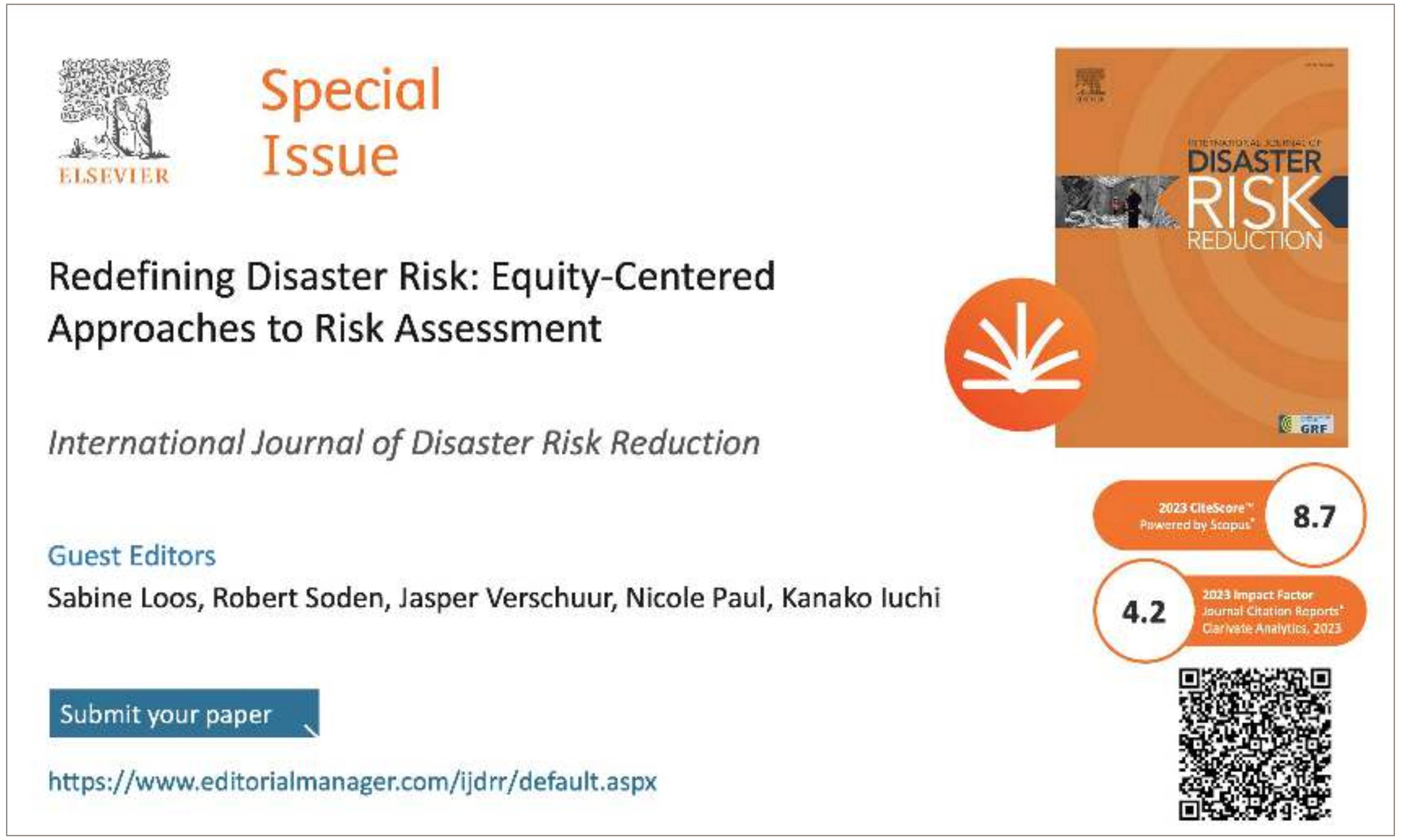
Housing recovery is not the same as household recovery

Several households that have since found permanent housing still perceive that they are only half-recovered or less, such as those who faced income decline

The drivers of household relocation decisions vary over time

Concerns about disaster risks drove initial moves, but today these concerns are tertiary compared to a much broader range of influencing factors

Consider submitting to our special issue in IJDRR!


Special Issue

Redefining Disaster Risk: Equity-Centered Approaches to Risk Assessment

International Journal of Disaster Risk Reduction

Guest Editors
Sabine Loos, Robert Soden, Jasper Verschuur, Nicole Paul, Kanako Iuchi

Submit your paper

https://www.editorialmanager.com/ijdrr/default.aspx

Thank you!

Nicole Paul

PhD Candidate, UCL RDR

nicole.paul.22@ucl.ac.uk

nicolepaul.io