Beyond building damage: estimating and understanding non-recovery following disasters

Sabine Loos (sloos@stanford.edu)
Stanford University https://orcid.org/0000-0001-7190-3432

David Lallemant
Earth Observatory of Singapore, Nanyang Technological University

Feroz Khan
Tulane University

Jamie McCaughey
ETH Zürich https://orcid.org/0000-0003-1490-5022

Robert Banick
World Bank Poverty and Equity Global Practice

Nama Budhathoki
Kathmandu Living Labs

Jack Baker
Stanford University https://orcid.org/0000-0003-2744-9599

Article

Keywords: natural disasters, recovery and reconstruction resources, non-recovery, post-disaster decisions

Posted Date: February 25th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-275410/v1

Beyond building damage: estimating and understanding non-recovery following disasters

Sabine Loos[1,2,*], David Lallemant[2,3], Feroz Khan[2,4], Jamie McCaughey[5], Robert Banick[6], Nama Budhathoki[7], and Jack Baker[1]

[1]Stanford University, Civil & Environmental Engineering, Stanford, CA 94305, USA
[2]Nanyang Technological University, Earth Observatory of Singapore, Singapore
[3]Nanyang Technological University, Asian School of the Environment, Singapore
[4]Tulane University, School of Social Work, New Orleans, LA 70112, USA
[5]ETH Zürich, Department of Environmental Systems Science, Institute for Environmental Decisions, Zürich, Switzerland
[6]World Bank Poverty and Equity Global Practice
[7]Kathmandu Living Labs, Kathmandu, Nepal
[*]*corresponding author: sloos@stanford.edu*

ABSTRACT

Following a disaster, crucial decisions about recovery resources often focus on immediate impact, partly due to a lack of detailed information on who will struggle to recover. Here we perform an analysis of surveyed data on reconstruction and secondary data commonly available after a disaster to estimate a metric of *non-recovery* or the probability that a household could not fully reconstruct within five years after an earthquake. Analyzing data from the 2015 Nepal earthquake, we find that non-recovery is associated with a wide range of factors beyond building damage, such as ongoing risks, population density, and remoteness. If such information were available after the 2015 earthquake, it would have highlighted that many damaged areas have differential abilities to reconstruct due to these factors. More generally, moving beyond damage data to evaluate and quantify non-recovery will support effective post-disaster decisions that consider pre-existing differences in the ability to recover.

Introduction

Natural hazards often cause disproportionate impacts on vulnerable populations and amplify inequality for years after an event. Among many examples, multi-family, Hispanic, and linguistically isolated households had inadequate access to loss-based assistance programs following the 1994 Northridge Earthquake[1]. When repeated examples of disaster-exacerbated inequality are evident over time, we must recognize that recovery policies, and the underlying information that supports them, fail to address and prevent deepening inequality.

Housing recovery policies are a powerful tool to prioritize the vulnerable people after an event[2–4]. Early decisions can shape the long-term recovery trajectory of an entire region. Currently, assistance is commonly aimed at restoring housing and is not designed as a means of redistribution[1]. For example, aid is based on losses (or damage) incurred to pre-disaster homes, therefore prioritizing those who had assets before the disaster[1,5]. Alternative "needs-based," "area-targeting," or "subsidiary" approaches exist, where policies prioritize groups who may lack necessary resources to support their own recovery[1,3,6]. However, there is a lack of information that identifies these groups and the factors that will impede their recovery in the weeks following a disaster when crucial decisions are made.

Advances in technology and data availability provide an opportunity to develop information on populations whose recovery may be impeded by factors other than damage to their home. Non-traditional post-disaster data, from remote-sensing or digital crowdsourcing, overwhelmingly focus on quantifying building damage[7], because it is relatively easy to quantify[8] and supports top-down recovery agendas[9]. While damage represents a reduction in housing quality, the focus on *immediate* impact (Figure 1a) is a myopic measure of *long-term* recovery needs (Figure 1b). To identify communities with disproportionate needs long after a disaster, we propose focusing on those who fall behind in recovery over time, or *non-recovery*. We focus on non-recovery since it places attention on those who do not recover rather than delineating the characteristics of good recovery. Here, we develop a specific metric of non-recovery using empirical reconstruction data from the 2015 Nepal earthquake, which is the probability that a household living in a severely damaged house will only partly rebuild or not rebuild at all within five years (Figure 1c). We use a large-scale survey conducted in 2019 in the earthquake-affected districts by the Asia Foundation and local partner Inter-Disciplinary Analysts to assess long-term impacts and recovery patterns[10]. To regionally estimate non-recovery

after a disaster, we use census, remotely-sensed, and modeled data that represent a range of sociodemographic, economic, environmental, and geographic factors likely to affect reconstruction.

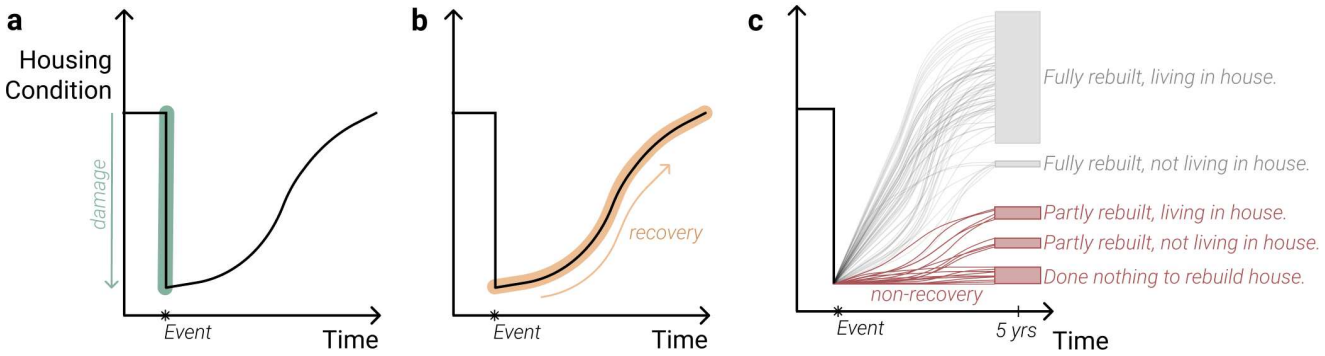

Figure 1. Non-recovery focuses on those who are likely to remain in damaged or destroyed homes over the long-term. Rapidly available post-disaster data often focuses on quantifying building damage, which captures the immediate reduction in housing condition (a), as opposed to long-term recovery needs over time (b). Non-recovery focuses on the impacted households who are not able to fully recover over time, as shown by the red household recovery trajectories in (c). In the Nepal case study, the specific metric of non-recovery is based on the five responses on reconstruction shown in (c), where the sizes of the rectangles represent the relative proportions of each response among the survey sample ($n = 3376$).

Our study shifts attention beyond estimating immediate building damage to identifying predictors of long-term recovery needs. In Nepal, we found the most important predictors of incomplete reconstruction fall into three categories: hazard exposure, rural accessibility and poverty, and reconstruction complexity. The relationships between these predictors and non-recovery are complex and disparate between households, highlighting how a variable like food poverty can be more important for one over another. Notably, the spatial pattern of non-recovery brings to focus regions that were not highlighted by damage alone. The model developed for Nepal can directly guide risk reduction planning in this region. For post-disaster planning in general, evaluating and quantifying non-recovery can build our understanding of disasters beyond damaged buildings and supports decisions that target pre-existing differences in the ability to recover.

Modeling non-recovery in Nepal

We consider non-recovery using the case of the 2015 Nepal earthquake, which is emblematic of a major modern disaster with substantial data produced from sensors, field surveys, and digital crowdsourcing[11]. After Nepal's National Planning Commission led their Post-Disaster Needs Assessment (PDNA) in the first three months after the earthquake, they estimated a total of 350,540 million NPR (∼$3.3 billion USD in 2015) in damages and losses to the housing sector[12]. The PDNA categorized districts by the severity of their impact, largely based on housing sector losses, as shown in Figure 2a. Afterward, the Government of Nepal implemented the Earthquake Housing Reconstruction Program in affected districts, which delivered reconstruction grants to repair or rebuild severely damaged or collapsed homes[13]. Due to data availability, we center our study on those affected districts outside of Kathmandu Valley (Figure 2a).

We relate surveyed non-recovery outcomes to remotely-sensed, modeled, and census-based variables affecting non-recovery (Figure 2b). Here, the non-recovery outcome is reconstruction progress, measured by whether a household fully completed reconstruction five years after the April 2015 Nepal earthquake (Figure 1c). The majority of rural households in the study region owned their homes, so reconstruction outcomes are informative of household recovery in this case. To ensure that the developed model is measuring recovery ability rather than differences in initial damage, we consider only households with damaged or collapsed homes ($n = 3376$).

Initial predictor variables of non-recovery were selected based on interviews with affected community members and reconstruction organizations along with existing literature on recovery in and out of Nepal. To support generalizability, the considered variables are commonly available or easily developed in other countries. Highly correlated variables were also removed. The resulting initial suite of 31 variables covers a wide range of factors with a potential influence on non-recovery and are included in Table **??**. We then reduced this set of variables to improve model parsimony using an automatic selection technique to remove variables less predictive than random noise. A few variables were further removed to increase model interpretability.

By relating reconstruction outcomes to variables affecting non-recovery, we take an empirical and data-driven approach to identify the factors that best predict a household's likelihood to reconstruct. Here, we apply a random forest model to

predict the probability of non-recovery, which is able to capture nonlinear influences and interactions between variables and performed better than traditional modeling methods we also tested. Our model identifies important and realistic factors affecting non-recovery and also predicts a tangible outcome: non-recovery.

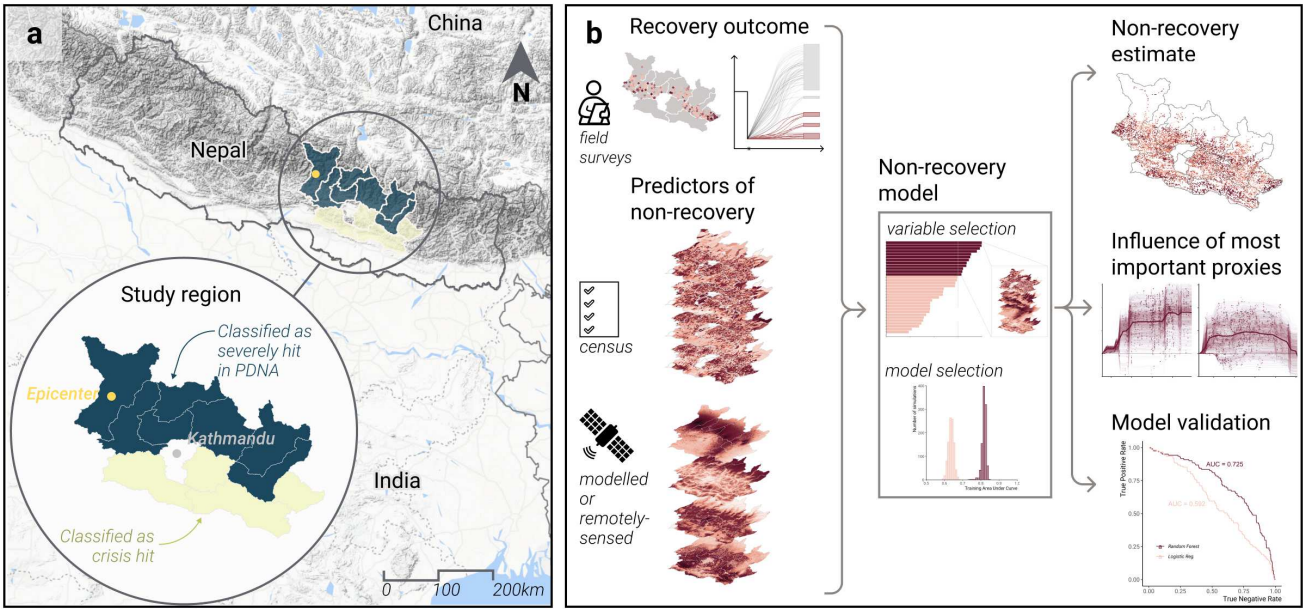

Figure 2. Study area in Nepal and non-recovery estimation approach. (a) The study area considered here are the 11 rural districts outside of Kathmandu Valley affected by the 2015 Nepal earthquake. The areas in blue were originally classified as severely hit (higher impact) and green as lower impact. (b) The model for non-recovery is calibrated on surveyed recovery outcomes, and uses readily available predictor variables representing sociodemographic, environmental, and geographic factors likely to influence recovery capacity. Outputs include a spatial estimate of non-recovery, the relative influence of each variable, and a metric of performance by validating the model on a test set (See Methods for more information).

Results

Our analysis reveals that eight predictors explain the probability of a damaged household completing after the 2015 Nepal earthquake (Table 1). We categorise these predictors into three main categories: 1) hazard exposure, 2) rural accessibility and poverty, and 3) reconstruction complexity. Each of these categories has roughly the same number of predictors, indicating they all are important for predicting non-recovery. These empirically-identified categories linked with impeded recovery are consistent with those defined in other resilience studies in Nepal[14,15] and broader frameworks of vulnerability[16–18]. The range of predictors indicates that reconstruction depends on a collection of socioeconomic, environmental, and geographic factors. Many studies recognize recovery and resilience as a multifaceted process with social and economic dimensions[19–21]. However, existing, rapidly available post-disaster information systems do not clearly acknowledge or account for this.

Influence of predictors on non-recovery

We calculated the marginal effect of each variable to evaluate its relative influence on predicted non-recovery, as shown in Figure 3. This figure allows us to see the average relationship between each variable and reconstruction ability. Each variable generally has a trend where greater values lead to higher probabilities of non-recovery. However, these relationships are not purely monotonic and vary from household to household. This variation points towards the diverse and complex reality of recovery experienced by affected households. Because random forest models capture interactions between variables, these relationships represent the influence of one variable given the inclusion of all the other variables in the model.

Hazard exposure

Hazard exposure includes variables relating to the intensity of the main earthquake or other ongoing or historical hazards that may compound the effects of the earthquake. Since Earthquake Shaking Intensity and Landslide Hazard emerged as significant predictors of non-recovery, our model confirms hazard exposure influences reconstruction. Here, our analysis shows that areas that experienced the most intense shaking from the mainshock (at a Modified Mercalli Intensity of 8.5) are predicted to have an

Table 1. Predictors of non-recovery in Nepal. The final eight predictors fall into three main categories, as indicated in the third column. Variables are ordered from most to least important as identified through the variable selection process.

Variable	Unit	Category
Earthquake shaking intensity	MMI	Hazard exposure
Tree cover	%	Rural accessibility and poverty
Population density	people per km^2	Reconstruction complexity
Remoteness	hours	Rural accessibility and poverty
Rainfall-triggered landslide hazard	index (unitless)	Hazard exposure
Tap water	%	Reconstruction complexity
Topographic slope	°	Reconstruction complexity
Food poverty prevalence	%	Rural accessibility and poverty

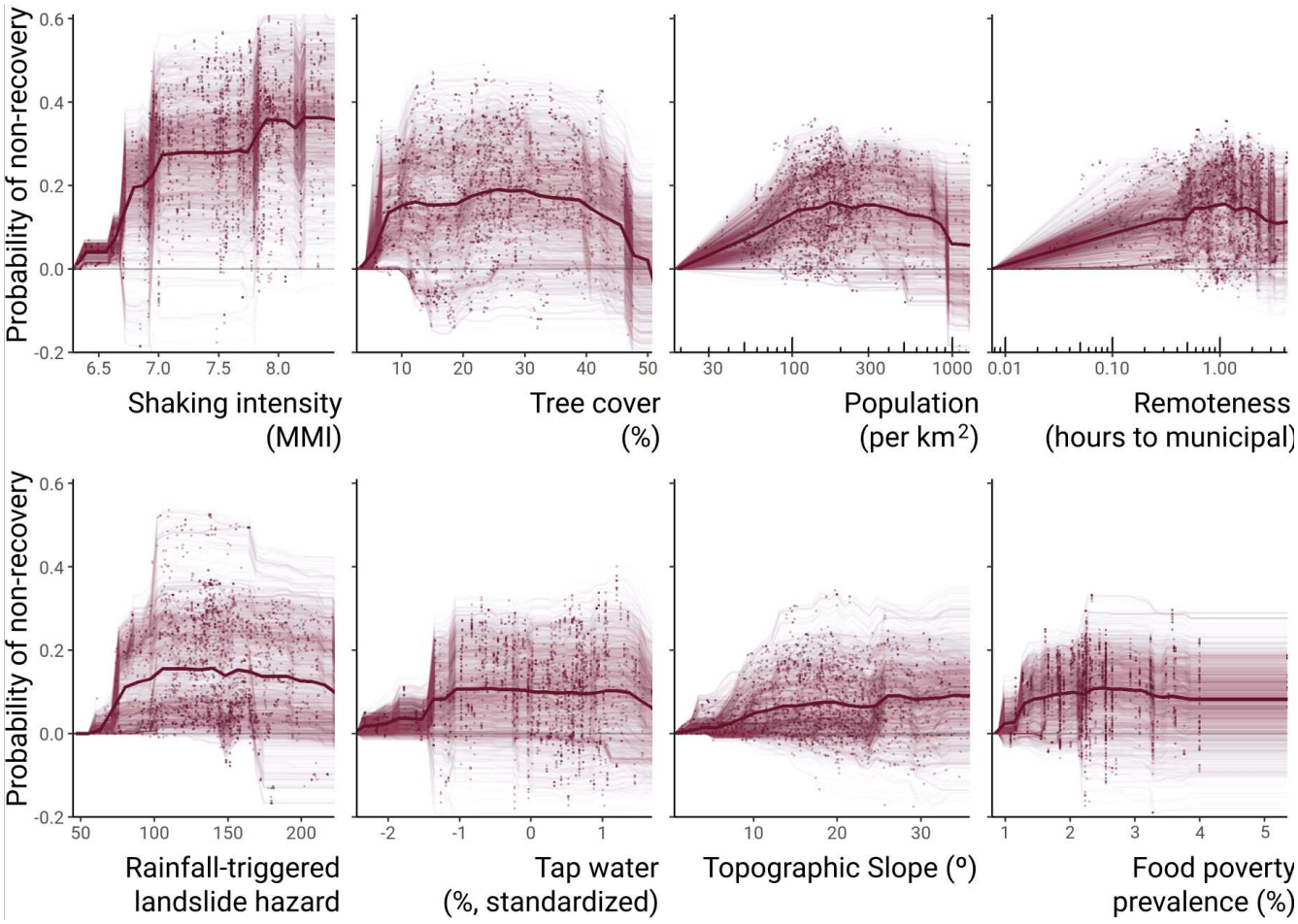

Figure 3. The diverse, relative influence of each predictor on the probability of non-recovery. Each plot shows the values of each of the eight final predictors (x-axis) and the resulting probability of non-recovery from the analysis in Nepal (y-axis). Each point is a household used to develop the model of non-recovery. The thin lines running through these points show how the predicted probability of non-recovery would change for that household when varying the value of the predictor on the x-axis from least to greatest, while keeping the other characteristics of the household fixed. The dark line shows the average relationship among all households. All results are scaled to the predicted probability of the minimum value of each predictor. The top 1% of data was truncated for sample representation.

average of nearly 40% greater probability of impeded recovery independent of the level of damage to the home. Note that we have already controlled for initial damage by considering only damaged houses, so this metric likely quantifies disruption to the surrounding community and the role of aftershocks near the areas closest to the mainshock. Similarly, Landslide Hazard is

associated with up to 20% greater predicted probability of non-recovery.

While we expected areas with higher mainshock Earthquake Shaking Intensity to be less likely to reconstruct due to immediate impacts to surrounding infrastructure, the inclusion of Landslide Hazard reflects the importance of compounding or more frequently occurring hazards on recovery capacity. The mainshock triggered nearly 20,000 identified landslides in Nepal[22], which already faced ongoing landslide risk due to monsoons and urban development[23,24]. Since the earthquake, many rural and remote households faced additional landslides during monsoon season[25]. The relationship we found represents how secondary risks like landslides accumulate pre-existing vulnerabilities of exposed Nepali communities, putting them at greater risk to immediate damage, leading to long-term displacement, and hindering regaining of livelihood[15,26].

Rural accessibility and poverty

Many affected communities in the study area were in rural, geographically isolated, or mountainous regions[10]. The inclusion of Remoteness, Tree Cover, and Food Poverty Prevalence reflect the particular challenges that impede reconstruction for isolated communities. Remoteness captures the travel time to municipal headquarters, which host local markets, services, and government offices[27]. The analysis predicts that the most remote households were nearly 20% less likely to reconstruct. Other studies have found that remoteness complicated the economics of household reconstruction: remote households struggled to attract or afford wage labor in highly competitive post-disaster labor markets; construction materials were much costlier to transport where vehicles could not reach[28,29]; and the lenders, non-profits, and governmental actors supporting recovery tended to neglect difficult-to-reach populations[30]. Tree Cover exhibits a slightly different relationship, where areas with between 10-40% tree cover are predicted to be least likely to reconstruct. While tree cover is natural capital that can promote resilience[14], the opposite relationship found in our model suggests it may be related to another community characteristic and requires further investigation.

Additionally, areas with greater prevalence of pre-existing food poverty were less likely to recover. This relationship provides evidence that already marginalized communities face additional challenges during reconstruction. It also potentially reflects the intertwined relationship between food security, building damage, and reconstruction[25,31], consistent with existing research in these areas.

Reconstruction complexity

The significance of population density, percentage of houses with tap water, and topographic slope likely reflect the logistical complexity of reconstructing. In the case of population density, our model predicts that households in denser areas are less likely to reconstruct. Urban areas in Nepal had unique challenges with reconstruction, such as shared landownership[30] and strict rebuilding requirements for settlements in heritage sites[32], resulting in slower reconstruction progress[10]. While our model captures the differential ability to reconstruct in terms of a gradient of population density, it is reminiscent of the observed discrepancy between urban and rural construction in Nepal and globally[2]. The inclusion of the percentage of households with tap water exhibits a similar relationship—greater prevalence of tap water in a region is associated with higher probability of non-recovery. Again, while infrastructure access can be viewed as promoting resilience, here it seems to be related to slowed reconstruction and warrants further research.

Topographic slope shows an influence on non-recovery beyond its link to hazard and accessibility. It is likely due to the difficulty of reconstructing on steep slopes or increased costs associated with retaining walls necessary in hillside communities[33].

Spatial distribution of non-recovery given damage

The model can be used to map the estimates of non-recovery. Figure 4a shows the probability of a household with a damaged house having not reconstructed within five years. It can be used in addition to the map of building damage in Figure 4b (from an auxiliary eligibility survey by the Government of Nepal). Comparing these maps shows that areas that would have been predicted to face the greatest and most persistent recovery needs are not necessarily those that were most damaged from the 2015 Nepal earthquake. The building damage caused by the earthquake was lowest in the southwest Hill districts of our study area and increased moving north towards the Mountain Districts near the Himalayas, closer to the epicenter and adjacent districts (Figure 4b). This pattern of damage is largely dictated by the high shaking intensity and prevalence of vulnerable construction types in the mountains. In contrast, Figure 4a shows that non-recovery is predicted to be likely scattered throughout the center, west, east, and south of the study region. This shows a pattern of non-recovery dictated by the spatial pattern of the social, geographic, and environmental predictors included (Figure **??**). The map of non-recovery points to areas that were not originally estimated as the most impacted, but that would require support during their recovery due to their socioeconomic and geographic make-up.

Discussion

To shift the focus from damaged buildings to vulnerable communities, we propose emphasizing and quantifying non-recovery which captures the challenges to long-term recovery. The proposed framework employs open data that is readily available after

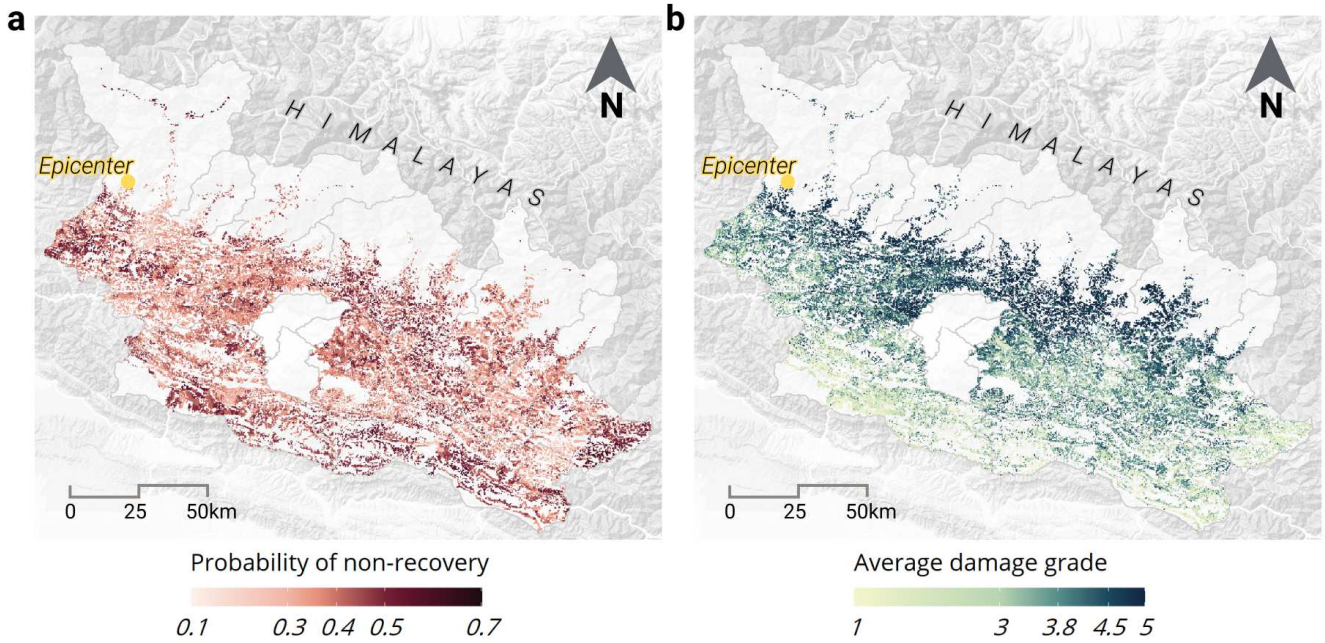

Figure 4. The regions predicted least likely to recover were not necessarily those most damaged from the 2015 Nepal earthquake. (a) The spatial distribution of non-recovery using data from the 2015 earthquake shows areas likely to have impeded reconstruction scattered throughout the center, west, east, and south of the study area. Dark red areas correspond to high probabilities of non-recovery, or top of the y-axis in Figure 3. (b) The pattern of damage was largely concentrated in the north, near the Himalayas. Damage data is from the Central Bureau of Statistics Nepal[34]. Both maps only show locations with buildings and are colored by quantiles of the distribution.

an earthquake to identify communities likely to be unable to recover within five years. The variables that predict non-recovery in Nepal fall in the categories of hazard exposure, rural accessibility and poverty, and reconstruction complexity. Many of the variables associated with non-recovery—such as remoteness and food poverty prevalence—are characteristics of some of the most vulnerable communities in Nepal[35]. By estimating non-recovery, we inherently capture both Nepali-specific and broadly-applicable factors of vulnerability in technical information that can be used as a basis for recovery policies.

Understanding social risks in addition to building damage has the potential to shape recovery policy. Developing rapid data on non-recovery emphasizes a broad range of potential factors to consider during recovery, offering a useful supplement to the myriad of building damage data produced in the weeks after a disaster[7,9,36]. For example, ongoing risks from landslides and food insecurity were identified as important concerns before the 2015 earthquake[24]. Non-recovery information emphasizes these important ongoing factors associated with recovery capacity, which can inform data collection and where recovery organizations focus. An example in Nepal would be the eligibility survey for the household reconstruction grant, which inadequately addressed landslide risks[8]. The reconstruction grant later prioritized reconstruction[35] over resettlement plans that were desired by some communities and would have addressed this landslide risk[26]. In addition, much of the NGO activity supporting reconstruction was concentrated in high damage areas, or near the epicenter in Gorkha, though work was also required in areas with chronic social vulnerability that received less media attention[37]. While the examples listed here are specific to Nepal, similar criteria on the goals and supporting data of recovery policies could be, and have been[9], applied to other regions affected by disasters.

Methodologically, we chose to model non-recovery, because many organizations understand the “most vulnerable" to be those groups who had trouble reconstructing[28]. Compared to index approaches to mapping vulnerability or resilience, our modeling framework provides a direct measure of recovery outcome—whether a house will finish reconstruction—rather than a unitless aggregate of factors of vulnerability that is challenging to validate[38,39]. Indices also rely on prior, place-based frameworks of vulnerability or resilience, which may not exist in most countries, limiting their use to mostly high-income countries where they have been studied. Though out model does require surveyed recovery outcomes of a previous disaster in the region to calibrate the importance of each predictor. Our non-recovery model is also able to capture the complex, often nonlinear relationships between socioeconomic, environmental, and geographic factors and reconstruction. The model we develop in this paper is relevant to Nepal, though the approach of relating recovery outcomes to commonly available data

can and should be applied elsewhere to evaluate its generality in identifying factors that are contextually-relevant to other locations. The approach could also be expanded with other metrics of non-recovery, like nutritional resilience[31] or population displacement[40], especially metrics that address landless populations still in need. Certainly, issues can arise when overly relying on technical disaster information[9,41]—hazards researchers should use data-driven models responsibly when representing complex processes[42,43]. The Area under the ROC Curve shown in Figure **??** demonstrates that the model of non-recovery provides an informative prediction, but remains uncertain. With new data becoming increasingly available, it is expected that the uncertainty could be further reduced in the future. Nonetheless, the model fills an important gap in existing information that is developed rapidly after a disaster in that it addresses long-term and multifaceted recovery needs. More generally, including a holistic and reflective set of initial variables is essential to modeling non-recovery. There were basics factors (e.g. gender or poverty) we would have liked to represent more granularly, though little openly-available data exists, pointing to the need for high-resolution social, economic, and mental well-being data.

Many agencies and NGOs are focusing on harnessing non-traditional data and methods to estimate damage after a disaster[44,45]. There is also great potential to use non-traditional data to estimate inequities in recovery that are just as important to understand when developing long-term plans. The approach presented in this paper can be used to identify unexpected but relevant factors that are important during recovery. Having quantitative data on how to support those least likely to recover can frame recovery actions like how to invest in recovery capacity[14], how best to handle reconstruction versus resettlement[15,46,47], or how to consider the community rather than just reconstructing the building[6]. It is clear that many policymakers and international agencies are moving towards developing data-driven evidence to support disaster decision-making[9,48]. Non-recovery is one crucial mechanism to focus our attention on quantifying metrics that support more nuanced recovery planning sooner after a disaster.

Acknowledgements

We thank The Asia Foundation and supporting field researchers, including Lena Michaels, Pranaya Sthapit, and Carolyn O'Donnell, for providing survey data and feedback on the importance of the predictors included in this model. We also thank Mhairi O'Hara, Robert Soden, Lisa Modifica, Sabin Ninglekhu, Ritika Singh, Jasna Budhathoki, and the rest of the Kathmandu Living Labs team who have all contributed to this project. Finally, we thank the households, community leaders, and other residents of Nepal who were generous with their time, energy, and trust to support this work. This project is supported by the World Bank's Trust Fund for Statistical Capacity Building (TFSCB) with financing from the UK DFID, the government of Korea, and the Dept. of Foreign Affairs and Trade of Ireland. The Disaster Analytics for Society Lab is funded by the Singapore National Research Foundation under the NRF-NRFF2018-06 award, and the Earth Observatory of Singapore. Sabine Loos was partially funded by the Stanford Urban Resilience Initiative, the John A. Blume Earthquake Engineering Center, and by the National Science Foundation Graduate Research Fellowship.

Author contributions statement

S.L., D.L., F.K., J.M., conceptualized model. S.L., N.B., and J.M. conducted exploratory interviews. S.L., D.L., J.B., F.K., and J.M. developed methodology. S.L. and R.B. curated data. S.L. conducted formal analysis and developed software. S.L., D.L., and J.B., acquired funding. S.L. developed visuals. S.L. wrote original draft. D.L. and J.B. provided supervision. All authors reviewed and edited manuscript.

Methods

Study area

The April 25th Nepal earthquake caused extensive damage and loss of life. The location of the earthquake's epicenter meant it affected not only the major cities in Kathmandu Valley, including the capital, but also the surrounding rural and remote villages. After the earthquake, the Government of Nepal established the Earthquake Housing Reconstruction Program (EHRP) to support households, mostly rural and outside of Kathmandu Valley, with reconstructing more safely[13]. This program provided grants of three lakh ($3000 USD) to houses that experienced severe damage or collapsed (damage grades at or above three using the EMS-98 damage scale) based on a detailed eligibility survey conducted in the 18 months after the earthquake. The EHRP was an owner-driven reconstruction program, given the high rates of ownership in these districts. Due to the government's program focus and data availability, we center our study on these less urban districts outside of Kathmandu Valley.

Outside of earthquakes, Nepal faces frequently recurring and ongoing hazards, where the greatest loss of life over the 20 years prior to the 2015 earthquake was caused by landslides and flooding[49]. In the years after the 2015 earthquake, affected communities faced multiple aftershocks, yearly flooding during the monsoon months, and landslides. These ongoing risks mean that Nepal is in a constant cycle of recovery from previous disasters[26].

In addition to Nepal's multihazard risk, the country's geography and changing political landscape make its recovery unique. Rural households face varying levels of remoteness to the nearest municipality, primarily due to the Himalayas' rugged terrain and the inability to access roads. After decades of a monarchist government (which transitioned to a multiparty democracy in the 1990s)[50,51], Nepal underwent a decentralization and devolution process in 2015-2017 that transferred governing power from the central to local governments located in these municipality headquarters throughout the country[52]. Therefore, the importance of local governments for reconstruction increased throughout the recovery period[32,53,54].

Survey data

The field survey data used in this study were collected by The Asia Foundation (TAF) and local partner Inter Disciplinary Analysts as part of their Independent Impact and Recovery Monitoring (IRM) project[10], funded by UKAid. This survey was part of a series of five surveys meant to monitor longer-term impacts, observe recovery patterns, and track the evolving needs of people affected by the earthquake in Nepal. Here, we only used data from their fifth round of surveys ($n = 4854$), conducted between September-October 2019, or four and a half years after the April 2015 earthquake. For this round of the survey, TAF sampled households using a stratified random sampling technique, representative at the district level. Eleven districts were surveyed, five of which were classified as "Severely-Hit" in the Post-disaster Needs Assessment, three as "Crisis-Hit," two as "Hit with Heavy Losses," and one "Hit," in order of most affected to least affected.

In this study, we considered households from the six rural districts classified as severely-hit and crisis-hit since these districts were in the primary region the government focused on during reconstruction ($n = 3484$) after removing all non applicable responses ($n = 83$). The survey question we used as a metric of non-recovery asks "If your house was damaged or completely destroyed by the earthquake, have you done any of the following?". Respondents designated whether they have done nothing to reconstruct their house, have started rebuilding, or have finished rebuilding. Even though this question is conditioned on severe damage, we further ensured this condition by only including households that stated in a separate response that their house was partially or fully damaged ($n = 3376$). Conditioning on damage controlled for the differences in reconstruction rates between damage states and, to some degree, the EHRP reconstruction grant that was geared towards only severely damaged homes. We used these responses as a binary variable for our probability classification models. Households that did not complete reconstruction by the time of the survey were classified as one ($n = 727$), and all other households were classified as zero ($n = 2649$). By classifying the survey data in this way, our model predicts the probability of a household *not completing* reconstruction four and a half years after an earthquake.

Predictor data

We represented factors we expect to influence non-recovery with a set of 31 variables, which come from openly available census, remote-sensing, or modeled datasets. These variables were considered rather than questions from the survey data, because the goal is to implement this model to predict areas of non-recovery in the weeks after an earthquake. Therefore, we used predictor data accessible after an event, whereas survey data would take years to collect. Here, we described predictor data for only those eight variables that were selected through the variable selection process as most important for predicting non-recovery. All other variables that were considered are listed in Table **??**.

Shaking intensity consists of the Modified Mercalli Intensity in the United States Geological Survey's Shakemap developed for the main earthquake on April 25, 2015[55]. The seismic landslide hazard map developed by the British Geological Survey provides an index of relative landslide hazard triggered by extreme 24 hour rainfall[56]. The remoteness variable estimates the time to travel to the nearest municipality headquarters, accounting for walking and driving time if roads are accessible, through a model developed by the World Bank Poverty and Equity Global Practice[27]. While we only calculated remoteness to municipalities, this variable is highly correlated to remoteness to other landmarks (e.g., district headquarters, roads, financial institutions). Original tree cover is derived from Landsat data and shows the per-pixel percentage estimate of tree canopy cover in 2010[57]. To capture the tree cover in the surrounding vicinity of each point in our study area, we took the average percentage within a 30-minute walking distance. Food poverty prevalence is the proportion of individuals living in an local government unit (LGU) who are in households that have a per capita food expenditure that is below the food poverty line. LGUs are a sub-district administrative unit in Nepal that is a collection of multiple villages, similar to a county. Food poverty prevalence per LGU is a small area estimation derived from a statistical model combining surveys and auxiliary data[58]. Population density is the estimate of population per 100 square meters from WorldPop, which we converted to people per square kilometer[59]. Slope was derived in R from the digital elevation model developed by CGIAR[60]. The tap water percentage is from Nepal's 2011 census[34].

Data preparation

Each predictor variable was produced or aggregated to different spatial scales (cells, wards, and LGUs), noted in Table **??**. To merge with the survey data, we extracted the value of each predictor at each household location. Once merged, we split the combined dataset into six folds using stratified random sampling to ensure each fold had roughly the same proportion

of households that are reconstructed and not reconstructed as the full dataset. We also visually inspected whether each fold covered the same spatial distribution of the study area as the full dataset. We used five folds (84%) as the training set to build the model of non-recovery and one fold (16%) as the test set for evaluating how the model would perform on a future dataset. For the spatial prediction of non-recovery over the study region (Figure 4a), we converted each proxy to the same resolution of 300m by 300m by resampling raster data or converting ward and LGU data to cells.

Models to predict probability of non-recovery

We developed a statistical relationship between the surveyed response of non-recovery (Y) and the suite of proxies ($\mathbf{X}$) using the training set. Our goal was to predict the probability that a damaged household has not completed reconstruction given its proxy values ($P(Y = 1|\mathbf{X} = \mathbf{x})$). We used a random forest, which is a non-parametric statistical model that averages the results of many individual, decorrelated decision trees[61]. Here we extended the typical random forest to predict probabilities of each household belonging to each reconstruction outcome (1 = not reconstructed, 0 = reconstructed)[62]. A bootstrapped sample of the training dataset is recursively split into distinct subsets for growing one tree in the random forest. Each split divides the data at that split, or parent node, into two child nodes. The parent node is split using a proxy variable that minimizes the mean squared error over all of a set of randomly selected features ($mtry$). For probability estimation, we continued to grow the tree until we reach the minimum nodesize of 10% of the bootstrapped sample. The probability of each node was the proportion of $Y = 1$'s. This process was repeated for a designated number of trees ($ntree$). For our model, we tuned hyperparameters using a grid search and minimized the mean squared error.

Because the random forest model is non-parametric, it does not require assumptions of the distribution of the data or specification of interaction terms. This is attractive for predicting non-recovery if a sufficient amount of training data is available because it allows for nonlinear relationships between the predictor variables and reconstruction outcome and for unexpected interactions to occur. We found the random forest outperformed (explained below) the standard probability prediction model, the logistic regression, both on the training and test sets (Figure **??**).

Variable selection

To prevent overfitting and for practicality, we reduced the number of variables used in the non-recovery model. We ensured that none of the predictor variables are highly collinear by manually removing all but one variable with a Pearson correlation coefficient greater than 0.75 over the entire study region. Many of these variables tended to be a variation of the same class of predictors (e.g. remoteness to municipality versus remoteness to financial institutions).

The variable selection occurred in two stages—one automatic and one manual. The automatic variable selection for the random forest was done by inserting a simulated noise variable and selecting all the proxy variables with a greater Gini importance[63] than that noise variable. To account for variation in the variable selection due to sample location, we repeated the model building process 1000 times using a bootstrapped sample of the training data. Through this automatic selection, we narrowed down the 31 original predictors to 12 variables that occurred more than 75% of the time in the 1000 models, shown in Supplementary Figure **??**, and retrained a new random forest using these variables.

Once we reduced the variables through this automatic selection, we then manually inspected whether the remaining 12 variables provided predictive relationships that were consistent with other studies in Nepal's reconstruction. We removed an additional four variables (percentage with thatch roof, monsoon month precipitation, dry month precipitation, and percentage Dalit caste), as the trends found here were unexplained in the literature.

Recovery outcome–predictor variable relationships

The partial dependence plots shown in Figure 3 provide insight into so-called "black-box" statistical methods, like the random forest[61]. The dark red line is the average marginal effect of a proxy of interest, X_s, on the random forest function, $f(\mathbf{X})$, when all other complementary proxies, $\mathbf{X}_c$, vary over the training data used to build the model of non-recovery. The resulting partial dependence function on X_s can be estimated with:

$$\hat{f}_{X_S}(X_s) = \frac{1}{N}\sum_{i=1}^{N}\hat{f}(X_S, \mathbf{X}_{Ci}), \tag{1}$$

where $\mathbf{X}_{Ci}$ are the values of the proxy variables in the training data of size N. Here, we show these relationships for the training data, as indicated by the light red lines, which is the partial dependence function $\hat{f}_{X_S}(X_s)^{(i)}$ disaggregated for each household and centered to the minimum value of $X_S^{(i)}$[64]. These plots can be interpreted as the average prediction of the model when varying a proxy of interest—it shows what is happening inside the model. However, it does not indicate causal mechanisms in the real world.

Validation

To evaluate the logistic regression and random forest models' performance, we calculated the area under the receiver operating characteristic (ROC) curve[65]. This curve assesses the trade-off between the rate of true positives versus false positives of our trained model of non-recovery when varying the cutoff used to classify its predictions as reconstructed or not. The closer the area under the ROC curve (AUC) is to one, the better the model is classifying an outcome. Here, we found an average training AUC of 0.817 for the random forest and 0.636 for the logistic regression (Figure **??**). The AUC of the test set (Figure **??**), which indicates performance on a hypothetical future dataset, was 0.725 for the random forest and 0.592 for the logistic regression. Thus, the random forest model's prediction performs better than the logistic regression and was used for our final model.

Model limitations

Several limitations of this model should be noted. The first is the interaction between reconstruction and aid in Nepal. The model attempts to predict differences in reconstruction trajectory that can not be explained by building damage. All damaged and collapsed homes were part of the EHRP's standardized assistance program, but we were unable to control for external, non-governmental assistance that households may have received. The second is the transferability in time of the model to future earthquakes. Since we learn from each disaster, it remains to be seen whether this specific model can be applied to future earthquakes that may occur in Nepal, though it is likely that many of the identified risk factors will continue to be relevant in the future. Finally, the model in Nepal does not include several Nepali-specific sources of vulnerability mentioned in previous vulnerability studies[66], such as gender or caste. This does not mean they are unimportant; rather, the final selected variables had a more representative sample or were more predictive for completing reconstruction in Nepal.

Data and code availability

All predictor data sources are acknowledged in the Methods and Supplementary Information sections. Most predictor data is openly available or otherwise indicated. Because field survey data contains sensitive location information, data can be made available upon request to The Asia Foundation. Data preparation and model building were completed in R version 3.6.1. Modeling packages used include ranger[67] for the random forest and glmnet[68] for the regularized logistic regression. All R scripts are made publicly available at https://github.com/sabineloos/nonrecovery-nepal/.

References

1. Kamel, N. M. & Loukaitou-Sideris, A. Residential assistance and recovery following the Northridge earthquake. *Urban Stud.* **41**, 533–562, DOI: 10.1080/0042098042000178672 (2004).
2. Jha, A. K., Duyne Barenstein, J., Phelps, P., Pittet, D. & Sena, S. *Safer Homes, Stronger Communities: A Handbook for Reconstructing after Natural Disasters* (The International Bank for Reconstruction and Development / The World Bank, Washington, DC, 2010).
3. Tafti, M. T. & Tomlinson, R. Theorizing distributive justice and the practice of post-disaster housing recovery. *Environ. Hazards* **18**, 7–25, DOI: 10.1080/17477891.2018.1435406 (2019).
4. Tierney, K. & Oliver-Smith, A. Social Dimensions of Disaster Recovery. *Int. J. Mass Emergencies Disasters* **30**, 123–146, DOI: 10.1017/CBO9781107415324.004 (2012).
5. Howell, J. & Elliott, J. R. Damages Done: The Longitudinal Impacts of Natural Hazards on Wealth Inequality in the United States. *Soc. Probl.* **66**, 448–467, DOI: 10.1093/socpro/spy016 (2019).
6. Mukherji, A. Post-Disaster Housing Recovery. *Oxf. Res. Encycl. Nat. Hazard Sci.* **1**, 1–35, DOI: 10.1093/acrefore/9780199389407.013.82 (2017).
7. Loos, S. *et al.* G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage. *Earthq. Spectra* DOI: 10.1177/8755293020926190 (2020).
8. Soden, R. & Lord, A. Mapping silences, reconfiguring loss: Practices of damage assessment & repair in post-earthquake Nepal. *Proc. ACM on Human-Computer Interact.* **2**, DOI: 10.1145/3274430 (2018).
9. Liboiron, M. Disaster data, data activism: Grassroots responses to representing Superstorm Sandy. In Leyda, J. & Negra, D. (eds.) *Extreme Weather and Global Media*, chap. 6, 144–162, DOI: 10.4324/9781315756486 (Routledge, New York and London, 2015).
10. The Asia Foundation. Aid and Recovery in Post-Earthquake Nepal: Independent Impacts and Recovery Monitoring Phase Five Quantitative Survey November 2019. Tech. Rep. Phase 5, The Asia Foundation (2019). URL: https://asiafoundation.org/wp-content/uploads/2017/10/Aid-and-Recovery-in-Post-Earthquake-Nepal-Synthesis-Report-Phase-4-1.pdf.

11. Dennison, L. & Rana, P. Nepal's emerging data revolution background paper. Tech. Rep. April, Development Initiatives (2017). URL: http://www.devinit.org/wp-content/uploads/2017/04/Nepals-emerging-data-revolution.pdf.

12. Government of Nepal National Planning Commission. Post Disaster Needs Assessment, Nepal Earthquake 2015. Tech. Rep., Kathmandu (2015).

13. Nepal Earthquake Housing Reconstruction Multi-Donor Trust Fund. Nepal Earthquake Housing Reconstruction Annual Report. Tech. Rep., Nepal Earthquake Housing Reconstruction Multi-Donor Trust Fund, Kathmandu, Nepal (2016). URL: https://www.nepalhousingreconstruction.org/documents/annual-report-nepal-earthquake-housing-reconstruction-mdtf-july-1-2015-july-31-2016.

14. Mishra, A. *et al.* Building ex ante resilience of disaster-exposed mountain communities: Drawing insights from the Nepal earthquake recovery. *Int. J. Disaster Risk Reduct.* **22**, 167–178, DOI: 10.1016/j.ijdrr.2017.03.008 (2017).

15. He, L., Aitchison, J. C., Hussey, K., Wei, Y. & Lo, A. Accumulation of vulnerabilities in the aftermath of the 2015 Nepal earthquake: Household displacement, livelihood changes and recovery challenges. *Int. J. Disaster Risk Reduct.* **31**, 68–75, DOI: 10.1016/j.ijdrr.2018.04.017 (2018).

16. Wisner, B., Blaikie, P., Cannon, T. & Davis, I. *At Risk: Natural Hazards, People's vulnerability and disasters* (2003), 2nd edn.

17. Turner, B. L. *et al.* A framework for vulnerability analysis in sustainability science. *Proc. Natl. Acad. Sci. United States Am.* **100**, 8074–8079, DOI: 10.1073/pnas.1231335100 (2003).

18. Birkmann, J. Measuring vulnerability to promote disaster-resilient societies : Conceptual frameworks and definitions. *Meas. Vulnerability to Nat. Hazards; Towards Disaster Resilient Soc.* **01**, 9–54, DOI: 10.1111/j.1539-6975.2010.01389.x (2006).

19. Comerio, M. C. Disaster Recovery and Community Renewal: Housing Approaches. *Cityscape: A J. Policy Dev. Res.* **16**, 51–68, DOI: 10.1093/jeg/lbs012 (2014).

20. Chang, S. E. Urban disaster recovery: a measurement framework and its application to the 1995 Kobe earthquake. *Disasters* **34**, 303–327 (2010).

21. Aldrich, D. P. & Meyer, M. A. Social Capital and Community Resilience. *Am. Behav. Sci.* **59**, 254–269, DOI: 10.1177/0002764214550299 (2015).

22. Gnyawali, K. R. & Adhikari, B. R. Spatial Relations of Earthquake Induced Landslides Triggered by 2015 Gorkha Earthquake Mw = 7.8. *Adv. Cult. Living with Landslides* DOI: 10.1007/978-3-319-53485-5 (2017).

23. McAdoo, B. G. *et al.* Roads and landslides in Nepal: How development affects environmental risk. *Nat. Hazards Earth Syst. Sci.* **18**, 3203–3210, DOI: 10.5194/nhess-18-3203-2018 (2018).

24. Sudmeier-Rieux, K., Jaquet, S., Derron, M. H., Jaboyedoff, M. & Devkota, S. A case study of coping strategies and landslides in two villages of Central-Eastern Nepal. *Appl. Geogr.* **32**, 680–690, DOI: 10.1016/j.apgeog.2011.07.005 (2012).

25. The Asia Foundation. Aid and Recovery in Post-Earthquake Nepal Independent Impacts and Recovery Monitoring Phase 3 Quantitative Survey: September 2016. Tech. Rep. Phase 3, The Asia Foundation (2016). URL: https://asiafoundation.org/publication/aid-recovery-post-earthquake-nepal-quantitative-survey-september-2016/.

26. He, L. Identifying local needs for post-disaster recovery in Nepal. *World Dev.* **118**, 52–62, DOI: 10.1016/j.worlddev.2019.02.005 (2019).

27. Banick, R. & Kawasoe, Y. Measuring Inequality of Access: Modeling Physical Remoteness in Nepal. Tech. Rep. August, World Bank Group (2019). URL: http://documents.worldbank.org/curated/en/605991565195559324/Measuring-Inequality-of-Access-Modeling-Physical-Remoteness-in-Nepal.

28. Coyle, D. Modelling Vulnerability & Shelter: Hamro Ghar Approach. Tech. Rep., People in Need, Scott Wilson Nepal, Build Change, PHASE Nepal, Kathmandu, Nepal (2019). URL: https://www.hrrpnepal.org/uploads/media/ModelingVulnerabilityandShelter-HamroGharApproach-PIN-March2019_20190318210109.pdf.

29. Action, P. Supply Chain of Construction Materials in Earthquake Aff ected Districts: An Assessment in Nuwakot and Rasuwa. Tech. Rep., Practical Action, Nepal (2016).

30. Limbu, B., Rawal, N., Suji, M., Subedi, P. C. & Baniya, J. Reconstructing Nepal: post-earthquake experiences from Bhaktapur, Dhading and Sindhupalchowk (2019).

31. Thorne-Lyman, A. L. *et al.* Nutritional resilience in Nepal following the earthquake of 2015. *PLoS ONE* **13**, 1–14, DOI: 10.1371/journal.pone.0205438 (2018).

32. Daly, P., Ninglekhu, S., Hollenbach, P., Duyne Barenstein, J. & Nguyen, D. Situating local stakeholders within national disaster governance structures: rebuilding urban neighbourhoods following the 2015 Nepal earthquake. *Environ. Urbanization* **29**, 403–424, DOI: 10.1177/0956247817721403 (2017).

33. Rodgers, J. *et al.* Resilience Challenges in South and Southeast Asian Hill Cities. In *Proceedings of the 11th National Conference in Earthquake Engineering* (Earthquake Engineering Research Institute, Los Angeles, CA, 2018).

34. Central Bureau of Statistics National Planning Commission Secretariat, G. o. N. Nepal - National Population and Housing Census 2011, Marks 100 years of census taking in Nepal (2011). URL: https://nada.cbs.gov.np/index.php/catalog/54.

35. Amnesty International. Building Inequality: The Failure of the Nepali Government to Protect the Marginalised in Post-Earthquake Reconstruction Efforts. Tech. Rep. (2017).

36. Lallemant, D. *et al.* Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake. *Earthq. Spectra* **33**, S435–S451, DOI: 10.1193/120316EQS222M (2017).

37. Lord, A. & Murton, G. Becoming Rasuwa Relief : Practices of Multiple Engagement in Becoming Rasuwa Relief : Practices of Multiple Engagement in Post-Earthquake. *HIMALAYA , J. Assoc. for Nepal Himal. Stud.* **37** (2017).

38. Rufat, S., Tate, E., Emrich, C. T. & Antolini, F. How Valid Are Social Vulnerability Models? *Annals Am. Assoc. Geogr.* **109**, 1131–1153, DOI: 10.1080/24694452.2018.1535887 (2019).

39. Fekete, A. Social Vulnerability (Re-)Assessment in Context to Natural Hazards: Review of the Usefulness of the Spatial Indicator Approach and Investigations of Validation Demands. *Int. J. Disaster Risk Sci.* **10**, 220–232, DOI: 10.1007/s13753-019-0213-1 (2019).

40. van de Lindt Walter Gillis Peacock Judith Mitrani-Reiser, J. W. Community Resilience-Focused Technical Investigation of the 2016 LUMBERTON, NORTH CAROLINA FLOOD MULTI DISCIPLINARY APPROACH. (2018).

41. Shrestha, K. K., Bhattarai, B., Ojha, H. R. & Bajracharya, A. Disaster Justice in Nepal's Earthquake Recovery. *Int. J. Disaster Risk Reduct.* **October**, DOI: https://doi.org/10.1016/j.ijdrr.2018.10.006 (2018).

42. Blumenstock, J. Don't forget people in the use of big data for development. *Nature* **561**, 170–172, DOI: 10.1038/d41586-018-06215-5 (2018).

43. Soden, R., Wagenaar, D., Luo, D. & Tijssen, A. Taking Ethics, Fairness, and Bias Seriously in Machine Learning for Disaster Risk Management. 2–5 (2019).

44. Gupta, R. *et al.* xBD: A Dataset for Assessing Building Damage from Satellite Imagery. 10–17 (2019).

45. Yun, S.-H. *et al.* Rapid Damage Mapping for the 2015 Mw 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO-SkyMed and ALOS-2 Satellites. *Seismol. Res. Lett.* **86**, 1549–1556, DOI: 10.1785/0220150152 (2015).

46. McCaughey, J. W., Daly, P., Mundir, I., Mahdi, S. & Patt, A. Socio-economic consequences of post-disaster reconstruction in hazard-exposed areas. *Nat. Sustain.* **1**, 38–43, DOI: 10.1038/s41893-017-0002-z (2018).

47. Ingram, J. C., Franco, G., Rio, C. R.-d. d. & Khazai, B. Post-disaster recovery dilemmas: challenges in balancing short-term and long-term needs for vulnerability reduction. *Environ. Sci. Policy* **9**, 607–613, DOI: 10.1016/j.envsci.2006.07.006 (2006).

48. Global Facility for Disaster Reduction and Recovery (GFDRR). Machine Learning for Disaster Risk Management. Tech. Rep., Washington, DC (2018). URL: https://www.gfdrr.org/sites/default/files/publication/181222_WorldBank_DisasterRiskManagement_Ebook_D6.pdf.

49. Pokhrel, D., Bhandari, B. S. & Viraraghavan, T. Natural hazards and environmental implications in Nepal. *Disaster Prev. Manag. An Int. J.* **18**, 478–489, DOI: 10.1108/09653560911003679 (2009).

50. Thapa, M. *Forget Kathmandu: An Elegy for Democracy* (Penguin, Viking, 2005), illustrate edn.

51. Sharma, K. The political economy of civil war in Nepal. *World Dev.* **34**, 1237–1253, DOI: 10.1016/j.worlddev.2005.12.001 (2006).

52. Chaudhary, D. The Decentralization, Devolution and Local Governance Practices in Nepal: The Emerging Challenges and Concerns. *J. Polit. Sci.* **19**, 43–64, DOI: 10.3126/jps.v19i0.26698 (2019).

53. Pokharel, T. *et al.* Political Economy Analysis of Post-Earthquake Reconstruction in Nepal: An Assessment of Emerging Role of Local Governments. Tech. Rep. September, Nepal Administrative Staff College and The Asia Foundation, Kathmandu, Nepal (2018). URL: https://asiafoundation.org/publication/political-economy-analysis-of-post-earthquake-reconstruction-in-nepal/.

54. Democracy Resource Center Nepal. The Roles of Local Governments in Disaster Management and Earthquake Reconstruction. Tech. Rep. April, The Asia Foundation, Kathmandu (2019). URL: https://asiafoundation.org/publication/the-roles-of-local-governments-in-disaster-management-and-earthquake-reconstruction/.

55. Worden, C. B., Thompson, E. M., Hearne, M. G. & Wald, D. J. ShakeMap Manual Online: technical manual, user's guide, and software guide. Tech. Rep. (2016). DOI: https://doi.org/10.5066/F7D21VPQ, URL: http://usgs.github.io/shakemap/.

56. British Geological Survey. Modelling Exposure Through Earth Observation Routines (METEOR) (2021). URL: https://meteor-project.org/.

57. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–854 (2013).

58. Haslett, S., Jones, G., Isidro, M. & Sefton, A. Small Area Estimation of Food Insecurity and Undernutrition in Nepal. Tech. Rep. (2014).

59. WorldPop. WorldPop. URL: www.worldpop.org.

60. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled seamless SRTM data V4 (2008). URL: http://srtm.csi.cgiar.org/.

61. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32, DOI: 10.1017/CBO9781107415324.004 (2001).

62. Malley, J., Kruppa, J., Dasgupta, A., Malley, K. & Ziegler, A. Probability Machines: Consistent Probability Estimation Using Nonparametric Learning Machines. *Methods Inf. Medicine* **51**, 1–7, DOI: 10.1038/jid.2014.371 (2012).

63. Nembrini, S., König, I. R. & Wright, M. N. The revival of the Gini importance? *Bioinformatics* **34**, 3711–3718, DOI: 10.1093/bioinformatics/bty373 (2018).

64. Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation (2015). DOI: 10.1080/10618600.2014.907095.

65. Hastie, T. J., Tibshirani, R. J. & Friedman, J. J. H. *The Elements of Statistical Learning* (Springer, 2009).

66. Aksha, S. K., Juran, L., Resler, L. M. & Zhang, Y. An Analysis of Social Vulnerability to Natural Hazards in Nepal Using a Modified Social Vulnerability Index. *Int. J. Disaster Risk Sci.* **10**, 103–116, DOI: 10.1007/s13753-018-0192-7 (2019).

67. Wright, M. N., Wager, S. & Probst, P. A Fast Implementation of Random Forests (2020).

68. Friedman, J. *et al.* glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models (2019). URL: https://cran.r-project.org/web/packages/glmnet/index.html.

Figures

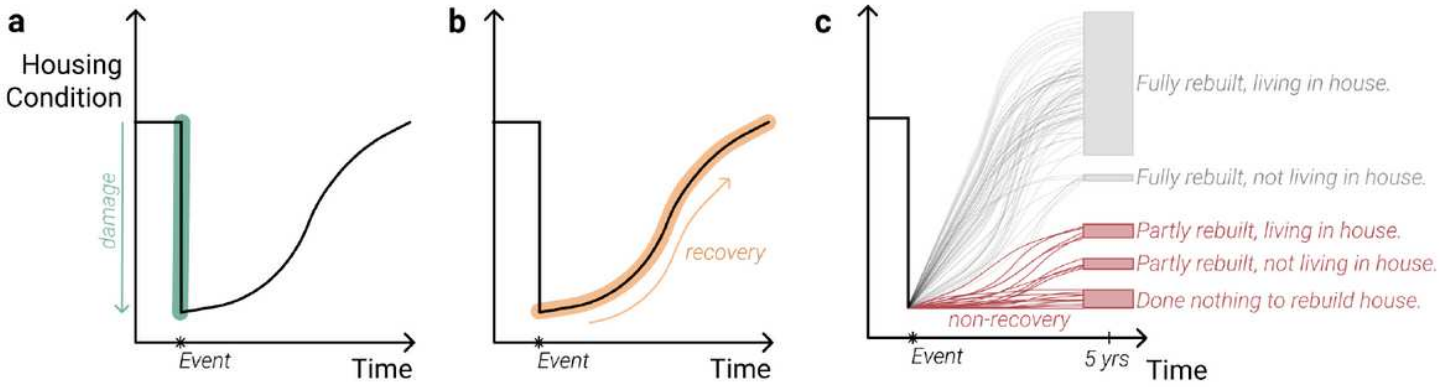

Figure 1

Non-recovery focuses on those who are likely to remain in damaged or destroyed homes over the long-term. Rapidly available post-disaster data often focuses on quantifying building damage, which captures the immediate reduction in housing condition (a), as opposed to long-term recovery needs over time (b). Non-recovery focuses on the impacted households who are not able to fully recover over time, as shown by the red household recovery trajectories in (c). In the Nepal case study, the specific metric of non-recovery is based on the five responses on reconstruction shown in (c), where the sizes of the rectangles represent the relative proportions of each response among the survey sample (n = 3376).

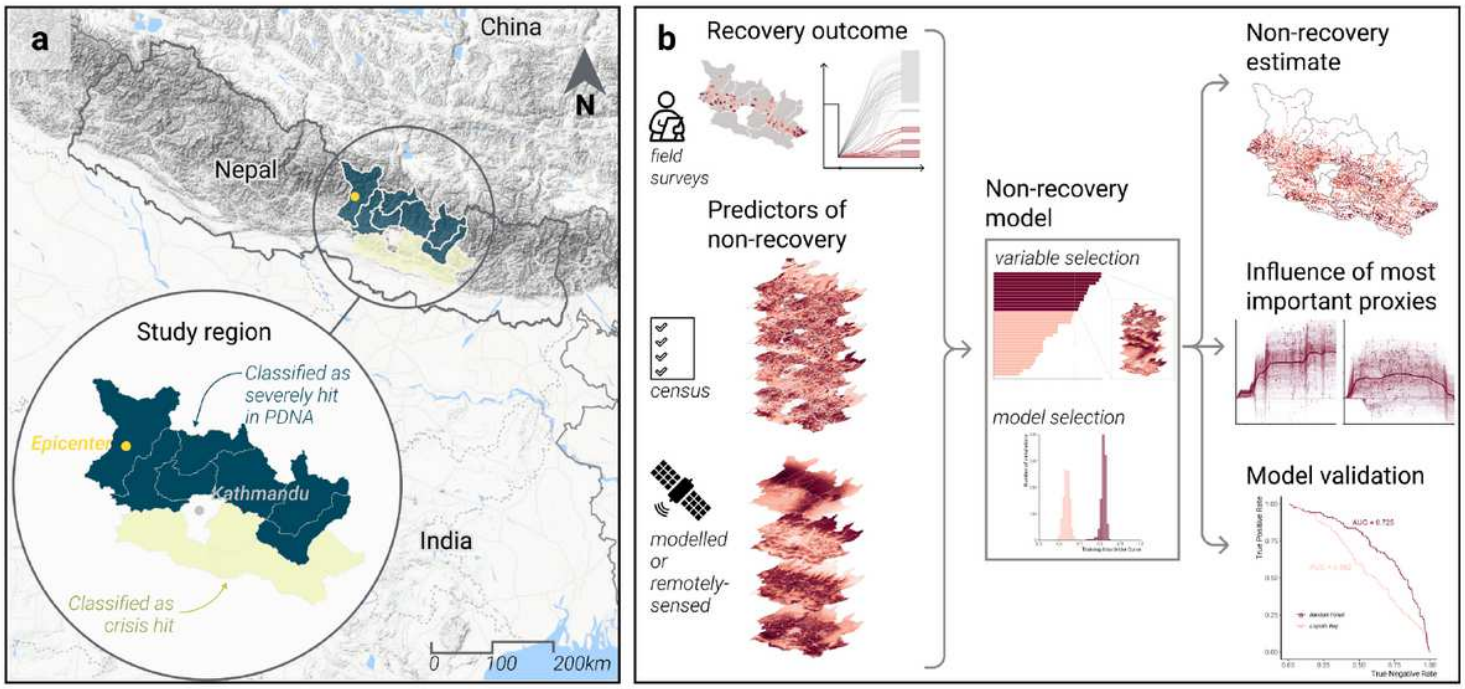

Figure 2

Study area in Nepal and non-recovery estimation approach. (a) The study area considered here are the 11 rural districts outside of Kathmandu Valley affected by the 2015 Nepal earthquake. The areas in blue were originally classified as severely hit (higher impact) and green as lower impact. (b) The model for non-recovery is calibrated on surveyed recovery outcomes, and uses readily available predictor variables representing sociodemographic, environmental, and geographic factors likely to influence recovery capacity. Outputs include a spatial estimate of non-recovery, the relative influence of each variable, and a metric of performance by validating the model on a test set (See Methods for more information). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

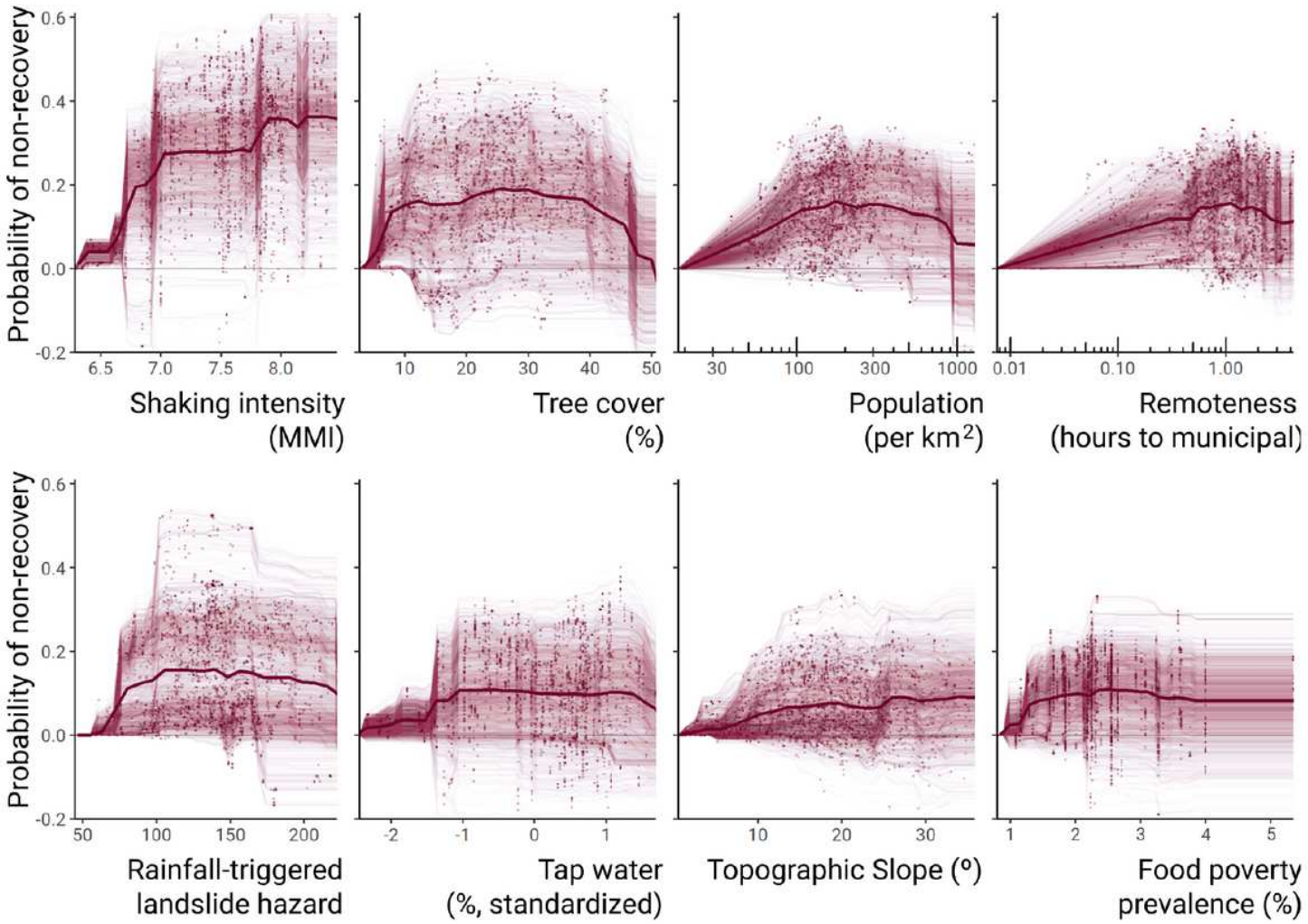

Figure 3

The diverse, relative influence of each predictor on the probability of non-recovery. Each plot shows the values of each of the eight final predictors (x-axis) and the resulting probability of non-recovery from the analysis in Nepal (y-axis). Each point is a household used to develop the model of non-recovery. The thin lines running through these points show how the predicted probability of non-recovery would change for

that household when varying the value of the predictor on the x-axis from least to greatest, while keeping the other characteristics of the household fixed. The dark line shows the average relationship among all households. All results are scaled to the predicted probability of the minimum value of each predictor. The top 1% of data was truncated for sample representation.

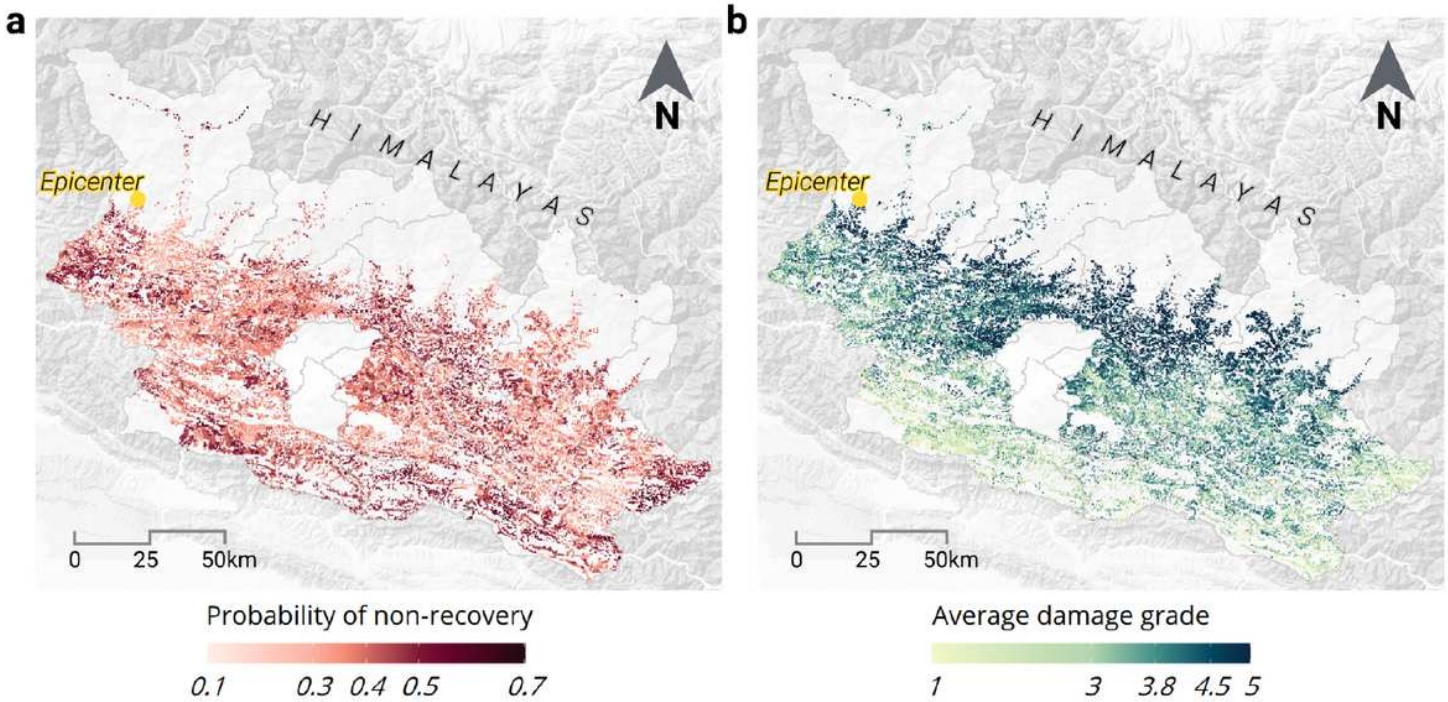

Figure 4

The regions predicted least likely to recover were not necessarily those most damaged from the 2015 Nepal earthquake. (a) The spatial distribution of non-recovery using data from the 2015 earthquake shows areas likely to have impeded reconstruction scattered throughout the center, west, east, and south of the study area. Dark red areas correspond to high probabilities of non-recovery, or top of the y-axis in Figure 3. (b) The pattern of damage was largely concentrated in the north, near the Himalayas. Damage data is from the Central Bureau of Statistics Nepal34. Both maps only show locations with buildings and are colored by quantiles of the distribution. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

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

- Loosnonrecovery20210224SIsubmissionversion.pdf