17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020
Paper N° C002455
Registration Code: S- A02526

Beyond building damage: modeling post-disaster need

S. Loos(1), D. Lallemant(2), J. McCaughey(3), N. Budhathoki(4), F. Khan(5), R. Singh(6), J. Baker(7)

(1) *Ph.D. Candidate, Stanford University, sloos@stanford.edu*
(2) *Assistant Professor, Earth Observatory of Singapore, Nanyang Technological University, dlallemant@ntu.edu.sg*
(3) *Postdoctoral Researcher, ETH Zurich, jamie.mccaughey@usys.ethz.ch*
(4) *Executive Chairman, Kathmandu Living Labs, nama.budhathoki@kathmandulivinglabs.org*
(5) *Research Assistant, Earth Observatory of Singapore, Nanyang Technological University, feroz.khan@ntu.edu.sg*
(6) *Research Assistant, Kathmandu Living Labs, ms.singhritika@gmail.com*
(7) *Professor, Stanford University, bakerjw@stanford.edu*

Abstract

During the months after an earthquake, government and humanitarian agencies scramble to collect information on the scale of its impact to make crucial decisions that shape the trajectory of that region's recovery, including how to allocate funding from domestic sources and international donors. In this study, we develop a model that can be used to rapidly estimate the spatial distribution of a novel metric of impact—persistent recovery need—using geospatial data that would be readily available in most countries.

As opposed to vulnerability or resilience indicators based solely on theory and demographics, the persistent recovery need metric also recognizes specific household obstacles to reaching recovery based on empirical evidence. We conducted a household survey (N=815) four years after the 2015 Nepal earthquake throughout the 11 most-affected districts outside of Kathmandu Valley. The goal of this survey was to assess several social, economic, and geographic characteristics that lead to differential recovery outcomes, such as housing reconstruction. Through this survey, we were able to identify household obstacles to recovery and therefore identify sub-factors that shape different household recovery trajectories.

We use these identified sub-factors to inform the data inputs into our spatial model of the persistent recovery need metric. The data we use are either readily available or can be developed from censuses, regular longitudinal surveys, remote sensing, and other geospatial data. The combination of these publicly available sources of data allows for high resolution and spatially distributed estimates that can be developed for most data-limited countries.

Our developed model to estimate the spatial distribution of persistent recovery need can supplement traditional impact estimates based on building damage alone. This metric would allow stakeholders to address a significant source of inequity in current post-disaster recovery decisions by supporting those vulnerable households that experienced the highest relative—rather than absolute—loss.

Keywords: Disaster Impact; Recovery; Vulnerability; Machine learning; Geographic Information Systems

1. Introduction

Some view post-disaster recovery as a catalyst for positive change—in fact, the entire rhetoric for building back better is that in the aftermath of a disaster, affected communities can theoretically become more resilient by rebuilding stronger, equitably, or otherwise aligned with community values [1]. Unfortunately, a disaster and the policies put in place afterward can also deepen existing inequities during the recovery phase and beyond[2], [3]. Two equally impacted households could recover at entirely different rates and to different standards because one household benefits from societal factors—such as prior saved assets or greater market accessibility—that put them at an advantage over the other [4], [5]

Post-disaster recovery decisions by the government and multilateral agencies have the potential to reduce these inequities[2], [6]. Early-on—concurrent with the response phase—stakeholders make policies that influence individual decisions households make on how they will proceed with reconstruction, shaping the entire trajectory of recovery for the affected region. For example, after many major disasters, the national government leads a Post Disaster Needs Assessment (PDNA), which determines the amount of external funding received and frames the recovery plans for years to come. This process has guided and provided resources for numerous disaster recoveries since its inception in 2005, including the 2010 Haiti earthquake, 2015 Nepal earthquake, and most recently at the time of this writing, the 2019 Mozambique Cyclone [7]–[9]. Even when PDNA's are not initiated, like after the 2010 earthquake in Chile, national governments must make early decisions on the eligibility requirements for households to receive aid from internal funds [10].

Stakeholders base these early recovery decisions on information about the scale of impact to estimate the total requirements for recovery. Typically, in the cases like the ones described above, disaster impact is equated with damage to buildings and infrastructure—the number and locations of damaged or collapsed buildings and its associated costs. However, housing recovery depends on much more than just building damage—it depends on factors on the local economy or the social fabric, among others. An overemphasis on damage-induced losses in decisions like the PDNA leads to housing recovery programs that focus solely on replacing the building stock rather than meeting the holistic needs of the community [6].

In this study, we develop a method to rapidly map a supplementary metric to building damage, termed, Persistent Recovery Need (PRN), which represents those factors that put households at more disadvantage when it comes to recovery. To develop the PRN metric, we conducted a household survey four years after the 2015 Nepal earthquake to assess several recovery outcomes and social, economic, and geographic obstacles to that recovery. We then define the PRN metric by empirically relating recovery to spatial proxies for these obstacles, which differs from existing indicators in terms of both methodology and data [11], [12]. In this study, we demonstrate the development of the PRN metric using housing reconstruction as a specific example of a recovery outcome in this framework.

The PRN metric can be used in conjunction with rapid metrics on building damage (e.g. PAGER, or the DPM) to provide a more holistic view of impact [13], [14]. The coupling of rapid maps of damage and rapid maps of need can inform recovery programs that are becoming increasingly focused on holistic needs, rather than solely on physical losses [6].

2. Defining need and its relation to resilience

Persistent Recovery Need is modified from and inversely related to resilience. While resilience in the hazards and vulnerability literature is often broadly defined to encompass the ability to *both minimize and recover from an impact* [15], here we examine Persistent Recovery Need as the *inability to recover*, given that a household has experienced a known *impact*.

In this paper, we demonstrate the development of the PRN metric and mapping its spatial distribution by looking at physical reconstruction at the household level. A household that cannot reconstruct to pre-event quality is less resilient, and therefore has greater persistent need. The relationship between resilience and PRN for housing condition is shown in Figure 1, highlighting different levels of reconstruction progress t years after the disaster. In this example, we develop the PRN metric for those who have been impacted, considering only those households that have experienced equal and high damage and require reconstruction, based on findings from our survey results described in Section 3.

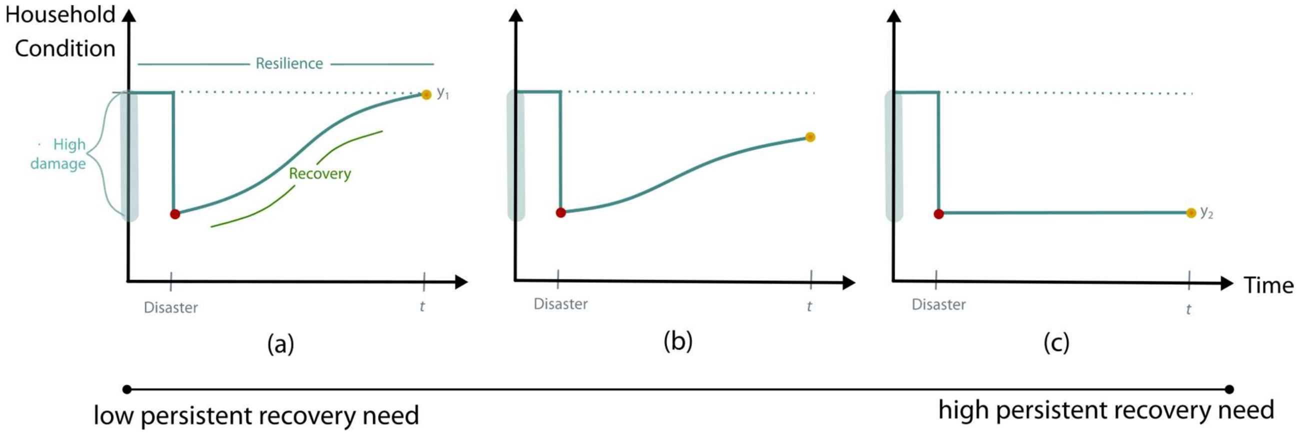


Figure 1. The relationship between need and resilience in recovery. Given a certain level of damage, a household that cannot recover, has low resilience and, therefore, greatest need.

Persistent Recovery Need is a composite of the multiple factors that are obstacles to recovery—what keeps a household in the curve in Figure 1c as opposed to that in Figure 1a? In the case of household reconstruction, these influencing factors could include physical obstacles like the sourcing of construction material, but also include social factors such as social capital, economic factors like prior savings, political factors like involvement in decision-making, and geographic factors like market access. Similar to previous frameworks for resilience, some of these factors overlap with those that influence both resilience and vulnerability [11]. The degree of influence of these obstacles on recovery may be a global relationship or be locally-specific. For example, lower institutional preparation and disaster planning is likely to lead to lower recovery [16]. Certain factors, like demographics (e.g. gender, ethnicity, or age), could affect recovery differently depending on the region or context [2]. Rather than treating these relationships as universal, weighting of these obstacles in relation to recovery should be treated differently for each context when quantifying the PRN metric.

2.1. Quantification of need vs. resilience

In our conceptual definition, those households with the greatest need are those that do not reach full recovery within a period of interest (Figures 1b and 1c). Since Persistent Recovery Need is definitively tied to the ability to recover, we quantify the PRN metric in a manner that explicitly ties household obstacles, x, to a recovery outcome, y. We, therefore, combine these multiple household obstacles by developing weights for each obstacle through a regression function that relates y to x.

There are three main differences between this quantification method may seem similar to the standard method of developing an indicator to quantify resilience—or vulnerability.

1. **Aggregation method.** Many indicators use aggregation methods to combine the factors of resilience, x, such as the weighted average, principal component analysis, k-means analysis, or expert weighting [11], [12], [17]–[19]. These aggregation methods result in a composite indicator that groups factors based on their similarity to each other, rather than their relationship to recovery. Instead, we define weights for each factor, x, by developing a function that relates each variable to a household recovery outcome, y. Another study that explicitly develops a function between the factors of resilience and recovery was created after the 2014 Napa earthquake—a more affluent context than many low- and middle-income countries around the world [20]. We describe our specific choice of function to relate recovery to obstacles to recovery in Section 4.3.
2. **Proxies for x, the obstacles to recovery.** Factors that promote or obstruct recovery are not always available at a detailed household level. Furthermore, the importance of each factor can vary from place to place. Typical resilience indicators select factors based on conceptual frameworks built on theory or prior empirical studies and then represent these factors using census data that is aggregated

to some administrative boundary (e.g. block-level, ward, district) [4], [19]. A few resilience indicators have been built using raster data [21], [22]. Using low-resolution data as proxies for these factors leads to a low-resolution quantification of resilience. Conversely, we identify the important obstacles to recovery for the PRN metric through a context-specific survey. We then represent these obstacles with a combination of census and remotely-sensed variables to disaggregate need to a higher resolution than an administrative boundary level.

3. **Choice of *y*, the recovery outcome.** A small subset of resilience indicators use metrics of recovery for validation [23], [24]. These recovery metrics tend to focus on objective physical indicators of reconstruction, defined by researchers or engineers. In this study, we define a recovery outcome that is based on the homeowner's self-evaluation of their recovery—a satisfactory metric for the nuanced definition of recovery since objective reconstruction rates may not accurately reflect recovery quality or housing adequacy [3], [6]. Alternatively, the demonstrated framework for developing the PRN metric can also be applied using other recovery outcomes, such as return to permanent housing.

Through these three main differences in our method to quantify PRN, we develop a metric of impact that 1) explicitly predicts a recovery outcome of interest, 2) uses high-resolution spatial proxies of the obstacles to recovery and 3) defines context-specific relationships between the recovery outcome and those proxies.

3. The structure of recovery after the 2015 Nepal earthquake

The 2015 Nepal earthquake is a fitting case study to develop the PRN metric because the structure of the recovery aid process led to owner-driven reconstruction and there was ample data collection after the event.

The April 25, 2015 earthquake triggered a Post-Disaster Needs Assessment (PDNA) which was completed exactly two months after the event [25]. The PDNA reported the total damages and losses to the housing sector (plus other sectors) then described the needs for recovery. This PDNA ultimately led to a commitment of $4 billion by the international donor community and the subsequent development of the Earthquake Housing Recovery Project (EHRP)—an owner-driven retrofit and reconstruction program aimed at mobilizing a portion of this aid to homes in the 11 most affected districts outside of Kathmandu Valley [26]. To identify beneficiaries to receive a grant for reconstruction, the Government of Nepal collected information on households impact for over 700,000 buildings. Nepal's National Reconstruction Authority (NRA) structured the grant to provide NPR 300,000 (~USD3000) for reconstruction of heavily damaged houses (damage grades 3-5 under the European Macroseismic Scale) and NPR 100,000 (~USD1000) for retrofit of lower damaged homes (damage grades 2-3) [27]. Since the structure of the reconstruction grant was consistent across these 11 districts, and households were empowered to organize their own reconstruction, the different household trajectories of recovery are not attributable to the amount of aid received. Rather, a household's recovery trajectory depends on other obstacles to recovery, allowing us to develop recovery models within each tier for the grant disbursement.

4. A geospatial model to predict Post Disaster Need

We develop a function to spatially predict PRN in the weeks following a future disaster. To develop this function, we use a field metric of recovery that we surveyed on the ground four years after the earthquake and a suite of spatial proxies for the obstacles to recovery that were available in Nepal in 2015.

4.1 Survey on household recovery, impact, and pre-existing characteristics

We conducted a household survey (N=815) four years after the 2015 earthquake to assess metrics of long-term recovery, immediate impact, and pre-existing factors. Given that the EHRP aid program was structured to last five years, four years ($t = 4$) allows for a sufficient amount of time to identify households that have persistent recovery needs [26].

The purpose of this survey is to define those household factors and impacts that are obstacles to multiple aspects of recovery, including household reconstruction, and should be identified either before or immediately after an event. Questions covered dimensions of well-being including housing, savings, health, nutrition, and education, among others. This paper does not focus in-depth on these specific empirical relationships between recovery, impact, and pre-existing characteristics found from the entire survey, but we supplement our discussion of the modeling results with those findings.

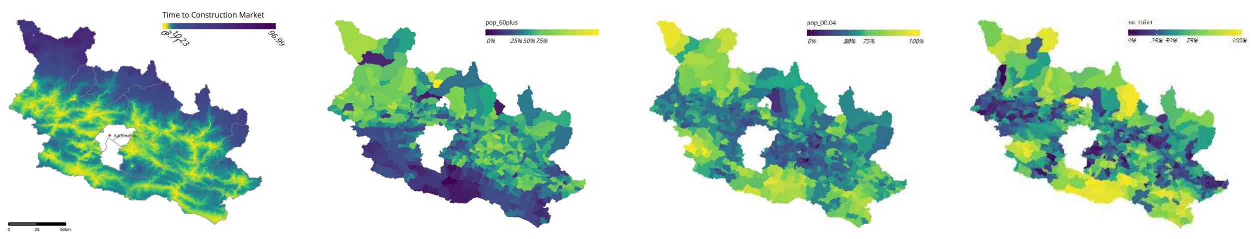
Field researchers conducted the survey for those 11 rural districts classified in the PDNA as severely impacted after the earthquake and were eligible for the reconstruction grant. We used a multistage cluster sampling design that kept logistics feasible while providing representative distributions of poverty rates, damage level, accessibility, and shaking intensity. We sampled nine village development communities (VDC), then three wards within each VDC. In each ward we mapped all houses then randomly selected one. Field researchers interviewed that household and its surrounding nearest neighbors. This sampling strategy resulted in a spatially distributed field dataset over all 11 rural districts, with individual locations clustered within a village, as shown in green in Figure 3.

We illustrate the development of the PRN metric by using responses to one of the recovery outcomes we surveyed respondents about: reconstruction progress. Each participant answered on a combined verbal and numerical scale of how close they felt to finishing the reconstruction of their house. Rather than an objective measurement of reconstruction, each respondent was able to self-assess their recovery progress. A subjective measure of reconstruction completion is a reasonable metric for the very nuanced definition of recovery since each homeowner can self-define their personal progress—especially since the use of objective measures of reconstruction progress, such as number of homes rebuilt, has obscured negative aspects that coincide with quick progress [3].

The rates of finishing reconstruction from our survey are higher for high damaged buildings than low damaged buildings. This may arise from lack awareness of the NRA support for retrofitting, the difficulty of retrofitting versus reconstructing [28], [29], and/or the necessity to complete reconstruction in cases where the previous home was uninhabitable. We therefore reduce our dataset to only those households that were severely damaged or collapsed (damage grade 4 or 5, N = 426).

4.2 Spatial proxies for the obstacles to recovery

The model to predict PRN requires data that could be openly available and accessible in the weeks after an earthquake. This includes census data, which is typically aggregated to some administrative unit, and raster data available at the pixel-level. The choice of spatial proxies to include in the model for PRN depends not only on its likely availability after a disaster, but also its spatial resolution, temporal resolution, and the reliability of any underlying methods to develop the proxy. Ideally, these proxies would also relate to important obstacles to recovery that we identify from our survey.



(a) Remoteness from Construction Market

(b) Percentage of pop. over 6o years old

(c) Percentage of pop. under 5 years old

(d) Percentage of households without toilet

Figure 2. Examples of spatial proxies for the obstacles to recovery, x. The three proxies shown here are the most influential for predicting Persistent Recovery Need. All other considered proxies, and their sources, are included in the appendix.

With the above constraints, we collected a suite of 38 potential spatial proxies to use as independent variables, x, for predicting nonrecovery. We show the four most important spatial proxies for prediction in

Figure 2. The full list of considered spatial proxies are in the appendix. To use these spatial proxies for prediction, we overlay each spatial proxy with the survey dataset. We also resample each spatial dataset to be the same resolution of about 290m by 290m.

4.3 Training the model

We develop a functional relationship between our reconstruction outcome and the suite of spatial proxies by training two different models to predict the probability of nonrecovery. In this case study, we want to predict the probability of a severely damaged or collapsed household not completing reconstruction within t = 4 years, and therefore having the greatest persistent need**.** We turn the responses to the reconstruction progress question into a binary variable (1 = not reconstructed, 0 = reconstructed) to use as our recovery outcome, the dependent variable y. As a result, the model for PRN will predict the probability of being at the yellow point in Figures 1a and 1c, given that a household starts at the red point. We used stratified random sampling to develop a balanced training set with 81% of the dataset, which ensures that our training data is distributed among all surveyed municipalities and contains a similar balance of the two reconstruction outcomes as the original dataset.

The choice of statistical model should allow for the prediction of probabilities of belonging to a certain class and indicate which variables are significant, and therefore, important to collect after a disaster. Here, we compare a linear and nonlinear method: 1) lasso regularized logistic regression and 2) probability random forest, respectively. These two models are sufficient for modeling after a disaster since both require little tuning of hyperparameters, handle multicollinearity, and provide a measure of variable importance or significance. Non-regularized logistic regression has previously been applied for relating resilience to reconstruction outcomes after the 2014 Napa earthquake [20]. Since we did not observe significant spatial correlation in the reconstruction outcome, we did not consider any spatial regression methods, although others have used Bayesian geostatistical models for similar purposes [30]–[32].

A common method to predict class probabilities is through logistic regression. With logistic regression, we predict the conditional probability that a reconstruction outcome, $Y = y_1$, given that the input proxies for the obstacles to recovery, $X = x$. To ensure that probabilities lie between 0 and 1, we apply the log-odds transformation to the dependent reconstruction outcome and set equal to a linear combination of the independent spatial proxy variables, as shown in Eq. (1).

$$log\left(\frac{P(Y = y_1|X = x)}{1 - P(Y = y_1|X = x)}\right) = \beta_0 + \beta^T x \tag{1}$$

in the case of two class probability estimation. To solve for β, we minimize the negative log likelihood,

$$L = -log\left(\prod_{i:Y_i = y_1} p(Y = y_1|X = x_i) \prod_{j:Y_j = y_2} \left(1 - p(Y = y_2|X = x_j)\right)\right) \tag{2}$$

or the product of the probabilities in Eq. (1). Logistic regression, however, can be affected by multicollinearity in the independent variables, so we first apply the lasso shrinkage method for variable selection. The lasso shrinkage method adds a penalty term, λ, to the function in Eq. (2) to shrink the coefficients on the independent variables that contribute least to the negative log likelihood to exactly zero [33]:

$$L + \lambda \sum|\beta|. \tag{3}$$

Once the coefficients for a certain set of variables are shrunken to zero, we then refit the logistic regression with the remaining variables, estimating the final coefficients for the selected variables without shrinkage.

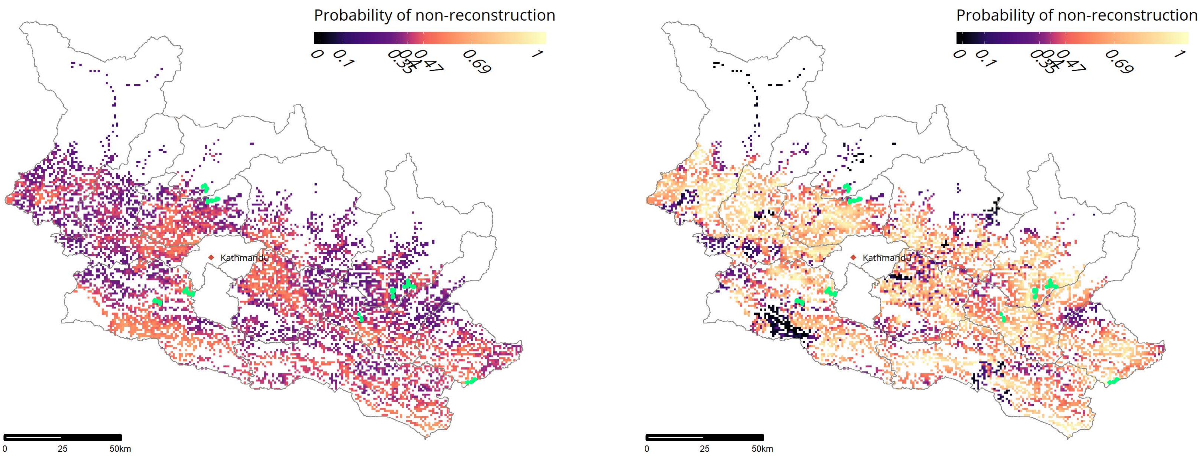
Here, we use the R package **glmnet** for the initial variable selection using lasso regularization [34]. We use a five-fold cross-validation to select λ, choosing the value one standard error away from that which provides the minimum area under the receiver operating characteristic (ROC) curve to prevent overfitting [35], [36]. We then refit a logistic regression using only the selected variables.

An alternative method to predict class probabilities is with random forest—an ensemble statistical learning method which averages the results of a large number of individual, decorrelated decision trees. First proposed for regression and classification, they have been extended to predict conditional probabilities of belonging to a class given a set of input variables [37]–[40]. Probability random forest are attractive for predicting need, because they can handle multiple types of input variables; allow for nonlinear relationships between the dependent and independent variables; and do not force probabilities to reach 0 and 1.

Here, we use the **ranger** package in R to fit the probability random forest to predict the conditional probability of each house belonging to a grid given the set of spatial proxies. Here, the probability of an observation falling into one class is the estimated proportion of that class averaged over all decision trees [38]. We employ a grid search using five-fold cross-validation to fit hyperparameters for the minimum node size in each tree and the number of variables available for splitting at each tree.

5. A rapid map of Persistent Recovery Need

With the spatial proxies that are comprehensively available, we can use our two trained models to predict Persistent Recovery Need at all locations in our area of interest to understand the benefits of each model. Figure 3 shows predicted probabilities of every household within a grid *not* completing reconstruction four years after the earthquake.



(a) Probability random forest (selected model) (b) Lasso regularized logistic regression

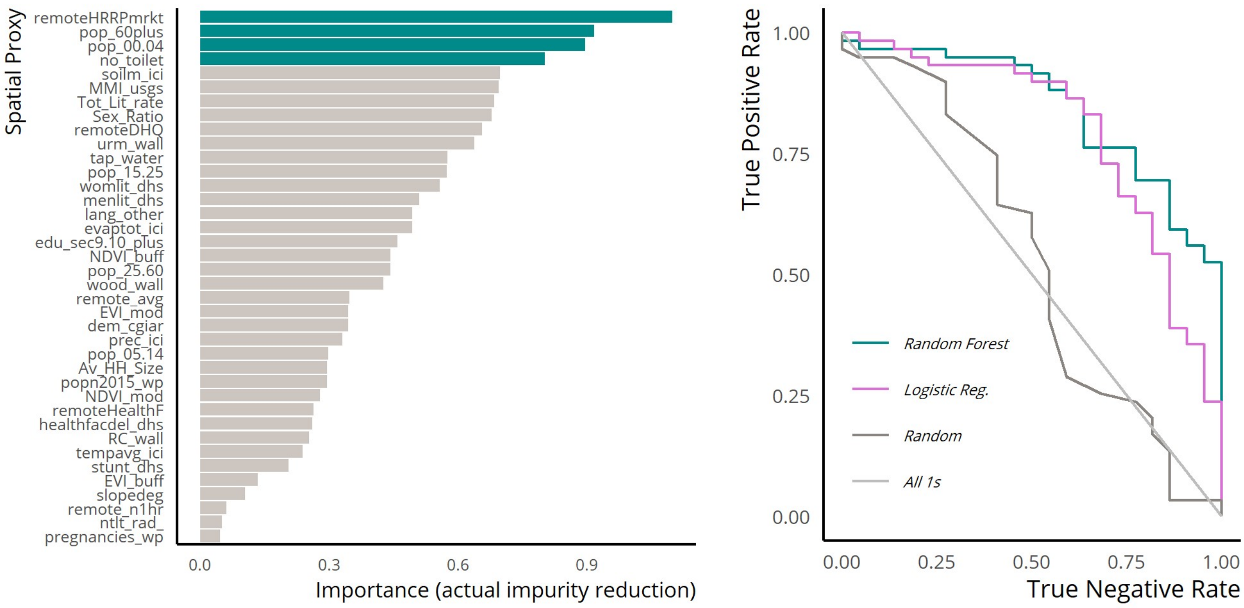
Figure 3. Prediction of Persistent Recovery Need using two alternative models. Areas in yellow have high need, or high probability of not reaching full recovery. The probability random forest model was chosen for the final prediction because of higher area under the curve during the training process.

The two maps of predicted need exemplify the differences in the outcome based on the choice of functional relationship between recovery outcome and spatial proxies. We selected the best model to use during the training step by comparing the area under the ROC curve from the five-fold cross-validations used for parameter tuning. Here, the probability random forest results in the greater area under the curve during training. This may arise from a more nonlinear relationship between spatial proxies and reconstruction and/or from possible underfitting by the logistic regression model. The random forest model also does not ensure probabilities of 0 and 1—values that would be highly unlikely to be predicted with such certainty in the immediate aftermath of a disaster—but predictions could be limited if the training data does not include samples over the full distribution of each covariate. For this application, we select the random forest model for prediction and interpretation.

5.2 Variable interpretation

The spatial trend of the PRN metric reflects the spatial proxies that are most important for prediction. The probability random forest model provides a variable importance measure that ranks each spatial proxy in terms of their contribution to the prediction performance of PRN. In this model, we use the corrected impurity importance measure—the sum of the total decrease in impurity of all nodes in the forest, normalized by the number of trees, and corrected for bias[33]. The most important variables for the prediction of nonrecovery are 1) remoteness from the construction market, 2) percentage of elderly population, 3) percentage of children under 5, and 4) percentage of households without a toilet (Figure 4a).

While the prediction of PRN does not indicate a causal relationship between spatial proxies and nonrecovery, reasons for evident spatial trends are supported by empirical analysis of our survey and similar findings in the literature. In the random forest prediction, surprisingly, areas closer to reconstruction markets are less likely to have completed reconstruction. Our survey, as well as other studies of reconstruction in Nepal [5] corroborate this finding—indicating that areas further away from construction markets have chosen to rebuild their homes using local materials and have finished reconstruction sooner than those who must save up to buy materials.



(a)Variable importance (b) ROC curves

Figure 4. Model outputs and validation for the random forest prediction of need. The variable importance (a) shows the most important spatial proxies for predicting need and the ROC curve (b) shows that the random forest has greater prediction accuracy than alternative methods.

5.3. Validity and utility

To understand the prediction performance of the random forest prediction of need, we can calculate the area under the ROC curve for predictions made on the test set. The area under the curve is highest for the random forest at 0.84, compared to 0.81 for the logistic regression, 0.60 for randomly generated probabilities, and 0.5

for predicting only nonrecovery at all locations. The ROC curve shown in Figure 4b can be further used to identify thresholds for classification of nonrecovered and recovered.

The maps shown in Figure 3 provide insight on the benefits of each model type, which use spatial proxy data from 2015. To map this metric after a future disaster in Nepal, we would use the developed functional relationship from the tuned random forest model shown here to predict the probability of non-recovery with spatial proxy data from the year of the disaster. Using these predicted probabilities, a national government, planning agency, or other stakeholders could use this spatial estimation of need alongside other impact metrics, such as regional damage, to focus on those areas at risk of persistent recovery need.

5. Conclusion

Maps of Persistent Recovery Need reflect and acknowledge the need for more holistic rapid metrics of impact, in addition to and beyond building damage, that support early plans to address households' differential needs for recovery. The proposed PRN framework here is novel in that it

1) predicts a forward-looking metric of persistent recovery need rather than immediate damage;
2) uses readily available and spatially explicit proxies for obstacles to recovery to allow for higher-resolution mapping;
3) and is applicable to other recovery outcomes of interest.

Here, we demonstrate the development of the Persistent Recovery Need (PRN) metric by using housing reconstruction from the 2015 Nepal earthquake as a case study. We developed a survey to assess this reconstruction outcome, among others, for 815 distributed households through the most affected rural districts of Nepal. We then compared the performance of two models to functionally relate housing reconstruction to spatial proxies for obstacles to recovery, including the probability random forest model and a lasso regularized logistic regression. The probability random forest outperformed the logistic regression in both the training and test sets, potentially due to a more nonlinear relationship. This model provides a measure of variable importance which can be interpreted and also used to understand which datasets are most important to collect after a disaster.

This framework for mapping a human-centric impact metric should be developed further for other regions and recovery outcomes. Plausibly, the relationships between reconstruction progress and spatial proxies may be similar in other rural areas in Nepal and possibly in other similar contexts elsewhere. Further study of spatially distributed data on disaster impacts and housing recovery would help to test the broad applicability of the relationships found in this study. To enable this further development, spatially distributed data on impacts and multiple recovery outcomes of interest (i.e. housing reconstruction, displacement, livelihoods sufficiency) should be collected after a future disaster to develop similar functions for that region.

6. Acknowledgements

Field survey and census data for this project were made available by the Government of Nepal and remoteness data was provided by the World Bank. The sources of all other datasets are included in the appendix and available online. We thank Robert Banick, Mhairi O'Hara and the entire Kathmandu Living Labs team, especially Bhawak Pokharel, Roshan Paudel, Manoj Thapa, Arogya Koirala, and Jasna Budhathoki, who have contributed to this project, as well as the many field researchers who collected our survey data. 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 Singapore's National Research Foundation, Nanyang Technological University, and the Earth Observatory of Singapore.

7. References

[1] S. Hallegatte, J. Rentschler, and B. Walsh, “Building Back Better: Achieving resilience through stronger , faster , and more inclusive post-disaster reconstruction,” Washington, DC, 2018.

[2] W. G. Peacock, S. Van Zandt, Y. Zhang, and W. E. Highfield, “Inequities in long-term housing recovery after disasters,” *J. Am. Plan. Assoc.*, vol. 80, no. 4, pp. 356–371, 2014.

[3] J. C. Ingram, G. Franco, C. R. Rio, and B. Khazai, “Post-disaster recovery dilemmas : challenges in balancing short-term and long-term needs for vulnerability reduction,” vol. 9, pp. 607–613, 2006.

[4] B. Wisner, P. Blaikie, T. Cannon, and I. Davis, *At Risk: Natural Hazards, People’s vulnerability and disasters*, 2nd ed. 2003.

[5] R. Banick, “Understanding the economic geography of disaster housing reconstruction: Measuring remoteness’s impact on post-earthquake housing reconstruction costs and outcomes in Nepal,” University of London, Centre for Development, Environment and Policy, School of Oriental and African Studies, 2019.

[6] A. Mukherji, “Post-Disaster Housing Recovery,” *Oxford Res. Encycl. Nat. Hazard Sci.*, vol. 1, no. April, pp. 1–35, 2017.

[7] Government of the Republic of Haiti, “Haiti Earthquake PDNA : Assessment of damage , losses , general and sectoral needs,” Port-au-Prince, Haiti, 2010.

[8] Government of Nepal National Planning Commission, “Post Disaster Needs Assessment, Nepal Earthquake 2015,” Kathmandu, 2015.

[9] Post Cyclone Idai Cabinet for Reconstruction, “Mozambique Cyclone Idai - Post Disaster Needs Assessment,” 2019.

[10] M. C. Comerio, “Housing recovery lessons from Chile,” *J. Am. Plan. Assoc.*, vol. 80, no. 4, pp. 340–350, 2014.

[11] S. L. Cutter, C. G. Burton, and C. T. Emrich, “Disaster Resilience Indicators for Benchmarking Baseline Conditions : Journal of Homeland Security and Emergency Management,” *J. Homel. Secur. {…}*, 2010.

[12] S. L. Cutter, B. J. Boruff, and W. L. Shirley, “Social vulnerability to environmental hazards,” *Soc. Sci. Q.*, vol. 84, no. 2, pp. 242–261, 2003.

[13] P. S. Earle *et al.*, “Prompt Assessment of Global Earthquakes for Response (PAGER): A System for Rapidly Determining the Impact of Earthquakes Worldwide,” *U.S. Geol. Surv. Open-File Rep. 2009-1131*, p. 15p, 2009.

[14] S.-H. Yun *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.*, vol. 86, no. 6, pp. 1549–1556, 2015.

[15] S. L. Cutter *et al.*, “A place-based model for understanding community resilience to natural disasters,” *Glob. Environ. Chang.*, vol. 18, no. 4, pp. 598–606, 2008.

[16] M. C. Comerio, “Disaster Recovery and Community Renewal: Housing Approaches,” *Cityscape A J. Policy Dev. Res.*, vol. 16, no. 2, pp. 51–68, 2014.

[17] N. S. N. Lam, M. Reams, K. Li, C. Li, and L. P. Mata, “Measuring Community Resilience to Coastal Hazards along the Northern Gulf of Mexico,” *Nat. Hazards Rev.*, vol. 17, no. 1, pp. 1–12, 2016.

[18] O. D. Cardona and M. L. Carreño, “Updating the Indicators of Disaster Risk and Risk Management for the Americas,” *J. Integr. Disaster Risk Manag.*, vol. 1, no. 1, pp. 27–47, 2011.

[19] C. Burton *et al.*, “Social Vulnerability : Conceptual Foundations and Geospatial Modeling To cite this version : HAL Id : hal-02001407 Social Vulnerability :,” 2019.

[20] V. Despotaki, L. Sousa, and C. G. Burton, “Using resilience indicators in the prediction of earthquake recovery,” *Earthq. Spectra*, vol. 34, no. 1, pp. 265–282, 2018.

[21] A. E. Frazier, C. S. Renschler, and S. B. Miles, “Evaluating post-disaster ecosystem resilience using MODIS GPP data,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 21, no. 1, pp. 43–52, 2012.

[22] M. A. Sherly, S. Karmakar, D. Parthasarathy, T. Chan, and C. Rau, “Disaster Vulnerability Mapping for a Densely Populated Coastal Urban Area: An Application to Mumbai, India,” *Ann. Assoc. Am. Geogr.*, vol. 105, no. 6, pp. 1198–1220, 2015.

[23] H. Cai, N. S. N. Lam, Y. Qiang, L. Zou, R. M. Correll, and V. Mihunov, “A synthesis of disaster resilience measurement methods and indices,” *Int. J. Disaster Risk Reduct.*, vol. 31, no. April, pp. 844–855, 2018.

[24] C. G. Burton, “A Validation of Metrics for Community Resilience to Natural Hazards and Disasters Using the Recovery from Hurricane Katrina as a Case Study,” *Ann. Assoc. Am. Geogr.*, vol. 105, no. 1, pp. 67–86, 2015.

[25] Government of Nepal, *Nepal Earthquake 2015 Post Disaster Needs Assessment*. 2015.

[26] Nepal Earthquake Housing Reconstruction Multi-Donor Trust Fund, “Nepal Earthquake Housing Reconstruction Annual Report,” Kathmandu, Nepal, 2016.

[27] National Reconstruction Authority, “Grant Disbursement Procedures for Private Houses Destroyed by the Earthquakes, 2073 (2016),” Kathmandu, Nepal, 2016.

[28] The Asia Foundation, "Independent Impacts and Recovery Monitoring Phase 4: April 2017," no. April, 2017.
[29] A. K. Jha, J. Duyne Barenstein, P. Phelps, D. Pittet, and S. Sena, *Safer Homes, Stronger Communities: A Handbook for Reconstructing after Natural Disasters*. Washington, DC: The International Bank for Reconstruction and Development / The World Bank, 2010.
[30] C. Bosco *et al.*, "Exploring the high-resolution mapping of gender-disaggregated development indicators," *J. R. Soc. Interface*, vol. 14, no. 129, p. 20160825, 2017.
[31] J. E. Steele *et al.*, "Mapping poverty using mobile phone and satellite data supplementary information," *J. R. Soc. Interface*, vol. 14, no. 127, 2017.
[32] N. Golding *et al.*, "Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals," *Lancet*, vol. 390, no. 10108, pp. 2171–2182, 2017.
[33] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
[34] J. Friedman, T. Hastie, and R. Tibshirani, "glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models." 2019.
[35] G. James, D. Witten, T. Hastie, and R. J. Tibshirani, *An Introduction to Statistical Learning*. New York, NY: Springer, 2013.
[36] T. J. Hastie, R. J. Tibshirani, and J. J. H. Friedman, *The Elements of Statistical Learning*. Springer, 2009.
[37] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[38] J. Malley, J. Kruppa, A. Dasgupta, K. Malley, and A. Ziegler, "Probability Machines: Consistent Probability Estimation Using Nonparametric Learning Machines," *Methods Inf. Med.*, vol. 51, no. 1, pp. 1–7, 2012.
[39] M. A. Olson and A. J. Wyner, "Making Sense of Random Forest Probabilities: a Kernel Perspective," 2018.
[40] S. Nembrini, I. R. König, and M. N. Wright, "The revival of the Gini importance?," *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, 2018.

8. Appendix

Proxy variable name	Description	Source
Av_HH_Size	Average Household Size per VDC	2011 Nepal Census
dem_cgiar	Digital elevation model	CGIAR
edu_sec9.10_plus	Percentage of pop. with education greater than grade 9&19	2011 Nepal Census
evaptot_ici	Total evaporation	ICIMOD
EVI_mod	Enhanced vegetation index	MODIS
EVI_buff	Enhanced vegetation index with a buffer	Self-developed
Healthfacdel_dhs	Percentage of births delivered at a health facility	DHS
Lang_other	Percent of population that speaks a language other than Nepali	2011 Nepal Census
Menlit_dhs	Male literacy rates	DHS
MMI_usgs	Modified Mercalli intensity shaking intensity	USGS
NDVI_mod	Normalized difference vegetation index	MODIS
NDVI_buff	Normalized difference vegetation index with a buffer	Self-developed
No_toilet	Percentage of population without a toilet	2011 Nepal census
Ntlt_rad_bin	Binary nightlight radiance	
Pop_00.04	Percentage of population under 5	2011 Nepal census
Pop_05.14	Percentage of population between 5-15	2011 Nepal census
Pop_15.25	Percentage of population between 15-25	2011 Nepal census
Pop_25.60	Percentage of population between 25-60	2011 Nepal census
Pop_60plus	Percentage of population over 60	2011 Nepal census

Popn_2015_wp	Population density 2015	Worldpop
Prec_ici	Precipitation in April 2015	ICIMOD
Pregnancies_wp	Number of pregnancies	Worldpop
RC_wall	Percentage of households with reinforced concrete wall	2011 Nepal census
Remote_avg	Average remoteness	World Bank
Remote_n1hr	Number of facilities within 1 hour	World Bank
remoteDHQ	Remoteness from district headquarter	World Bank
remoteHealthF	Remoteness from a health facility	World Bank
remoteHRRPmrkt	Remoteness from HRRP defined construction market	World Bank
Sex_Ratio	Sex Ratio	2011 Nepal census
Slopedeg	Slope in degrees	Self-developed
Soilm_ici	Soil moisture in April 2015	ICIMOD
Stunt_dhs	Stunting	DHS
Tap_Water	Percentage of households whose main drinking supply is from the tap	2011 Nepal census
Tempavg_ici	Average April temperature	ICIMOD
Tot_lit_rate	Total literacy rate	2011 Nepal census
Urm_wall	Percentage of households with unreinforced masonry wall	2011 Nepal census
Womlit_dhs	Women’s literacy rate	DHS
Wood_Wall	Percentage of households with timber wall	2011 Nepal census