Supporting equitable disaster recovery through mapping and integration of social-vulnerability into post-disaster impact assessments

INFORMATICS *for*
EQUITABLE RECOVERY PROJECT

EARTH OBSERVATORY OF SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

STANFORD URBAN RESILIENCE INITIATIVE

ETH zürich

Kathmandu Living Labs

Humanitarian OpenStreetMap Team

Report date
29 February 2020

Lead organisation
Earth Observatory of Singapore

Collaborators
Stanford University,
ETH Zurich,
Kathmandu Living Labs,
Humanitarian OpenStreetMap Team

SDGs covered by project scope
This project addresses issues and challenges linked to the following United Nations Sustainable Development Goals (SDGs):

- SDG 11: Make cities and human settlements inclusive, safe, resilient and sustainable
  - 11.1 By 2030, ensure access for all to adequate, safe and affordable housing and basic services and upgrade slums
  - 11.5 By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations
- SDG 10: Reduce inequality within and among countries
- SDG 1: End poverty in all its forms everywhere
  - 1.5 By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters.

Countries, data types, & technologies used

- Countries: Nepal
- Data types:
  - Geospatial data on physical impacts
  - Census & other population data
  - Qualitative data (interviews)
  - Quantitative field survey data (household questionnaire)

Project objective: To enable more equitable, resilient disaster recovery by improving post-disaster information systems so that agencies can understand, identify and respond to the differential needs of diverse groups of people after a disaster.

Suggested citation: Lallemant, David, Sabine Loos, Jamie McCaughey, Nama Budhathoki, Feroz Khan. 2020. "Informatics for Equitable Recovery: Supporting equitable disaster recovery through mapping and integration of social vulnerability into post-disaster impact assessments."

Acknowledgements

This report was written by a team led by David Lallemant (Earth Observatory of Singapore, Nanyang Technological University), and composed of Sabine Loos (Stanford Urban Resilience Initiative), Jamie McCaughey (ETH-Zurich), Feroz Khan (Earth Observatory of Singapore, Nanyang Technological University), and Nama Budhathoki (Kathmandu Living Labs). The project benefited from important contributions by Mhairi O'Hara (Humanitarian Openstreetmap Team), Ritika Singh (Kathmandu Living Labs) and Jasna Budhathoki (Kathmandu Living Labs). Photographs and figures used in this report were produced by Sabine Loos. Paras Shrestha designed the report.

For their valuable contributions and support, the team thanks former and current officials of the Government of Nepal: Dr. Swarnim Wagle, former Vice Chairman of the National Planning Commission, Mr. Yubaraj Bhusal, former CEO of the National Reconstruction Authority, and Mr. Manohar Ghimire, Under Secretary of the National Reconstruction Authority.

We are also grateful for the important guidance and advice provided by Sabin Ningleku, Nagendra Amatya, Patrick Daly, Shin Bin Tan and David Wald. The team thanks Robert Soden for early advising and insights and Robert Banick for data inputs and insights on recovery processes in Nepal.

Special thanks are also due to:
The exceptional staff of Kathmandu Living Labs (KLL), especially Bhawak Pokharel, Roshan Paudel, Arogya Koirala, and Manoj Thapa, who have contributed to this project at different points.

Field researchers for their exceptional work collecting household surveys throughout severely earthquake hit districts of Nepal: Kajal Shrestha, Shamjhana Thapa, Manoj Maharjan, Sajal Khadka, Asmita Banjara, Geeta Bhujel, Pratibha Poudel and Bindu Khanal.

The Blume Center for Earthquake Engineering at Stanford University for supporting the development of the tools described in this report. Special thanks are due to Dr. Jack Baker for providing advice and ideas and the Baker Research Group for continuous feedback.

Sang-Ho Yun from the Jet Propulsion Laboratory (JPL) team at NASA for data and technical advice on Damage Proxy Maps.

Members of the Disaster Analytics for Society Lab (DASL) at the Asian School of the Environment of Nanyang Technological University for technical feedback, regular project inputs and advice.

Chen Qi and Nuraishah Binte Kasmadi of the Earth Observatory of Singapore for administrative support of the project team.

We are especially grateful to the World Bank Data Innovation Fund team members Leslie Ricketts, Siddhesh Kaushik and Sun Hwa Song for their support and regular feedback throughout the project process. Finally, we thank the 815 households, numerous community leaders, and other residents of Nepal we surveyed and interviewed. We are grateful for their time, generosity and insight, and their trust that we will reflect their experiences thoughtfully. We hope that their and our contributions will lead to a more just, equitable and resilient future.

For funding and support, we acknowledge:
This project, Supporting Equitable Disaster Recovery Through Mapping (Nepal), submitted in response to the 2017 call for proposals by the World Bank's Development Data Group (DECDG) and the Global Partnership for Sustainable Development Data (GPSDD), is supported by the World Bank's Trust Fund for Statistical Capacity Building (TFSCB) with financing from the United Kingdom's Department for International Development (DFID), the Government of Korea, and the Department of Foreign Affairs and Trade of Ireland.

Other funding support was provided by the Earth Observatory of Singapore, and the National Research Foundation of Singapore through Fellowship grant NRF-NRFF2018-06.

This report does not constitute an endorsement by any project advisor, institutional contributor or funding agency.

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

DECDG
Trust Fund for Statistical Capacity Building
TFSCB

Global Partnership for Sustainable Development Data

UKaid
from the British people

Irish Aid
Rialtas na hÉireann
Government of Ireland

KWPF
KOREA-WORLD BANK PARTNERSHIP FACILITY

Executive Summary

Disasters continue to present tremendous obstacles to sustained development progress and the wellbeing of communities around the world. Key to mitigating the long-term impacts of disasters is the ability to rapidly respond and recover in ways that build resilience and protect hard-fought development gains. However, the information systems needed to support such resilient and equitable recovery are currently lacking, such that decision makers often end up gathering evidence for recovery decisions in ad-hoc ways rather than systematically.

In the hours, days, and weeks following a disaster, governments of affected countries, international institutions, the humanitarian sector and other stakeholders try to make sense of the catastrophe with information and data that is limited in scope and considerably uncertain. In a relatively short time frame, these stakeholders must grapple with making urgent response and recovery decisions with limited and uncertain data, knowing that their early plans will have long-term impacts on recovery processes. These decisions are further complicated by the many ways that disaster impacts are entangled with pre-disaster inequities and vulnerabilities, often meaning that the poorest and most disempowered bear the brunt of damage and impacts.

This study focuses on the information and the data that influence early decision-making processes after a disaster. We call this project **Informatics for Equitable Recovery** because we have built frameworks for developing information systems that focus on identifying not only those who are impacted but also those who are least able to recover so that we can inform *more equitable long-term recovery processes.*

The main outcomes of this study are the following:

1. **We propose and provide new tools to estimate and map post-disaster building damage more accurately, integrating the multiple data-sources produced after a disaster.**
   Our Geospatial Data Integration Framework (G-DIF) produces high-resolution estimates and maps of building damage which leverage and integrate numerous data sources, mobilizing recent advances in remote sensing and big data technologies. G-DIF relies on a limited amount of ground-truth data to locally calibrate and integrate the increasing number of damage proxies produced after a disaster into a single estimate and map of damage and its uncertainty. When validated against the full data-set of building damage across the entire affected area, we demonstrated that the model more accurately captures the scale and spatial distribution of damage. Importantly, through the integration of numerous data sources, local calibration from limited ground truth data, and accounting for the spatial relation in data, G-DIF is a very flexible, scalable and generalizable framework for estimating and mapping post-disaster damage. Implementing the framework can further provide context-specific damage estimates irrespective of new technological developments or the specific data that is produced after an event.

2. **We provide new insights into the diversity of recovery trajectories and the characteristics that link pre-disaster vulnerability to post-disaster recovery and need.**
   While defining a normative standard of equity in recovery is beyond the scope of this

study, we highlight salient sources of differential vulnerability and need that can be incorporated into early disaster assessments. While building damage is often the simplest measurement of disaster impact, it is insufficient to capture the differential needs and obstacles to recovery. Through extensive household surveys, we identify social, economic and geographical characteristics before and immediately after the earthquake that are predictive of household recovery progress years later. Importantly, recovery can follow many different paths, often involving difficult trade-offs. Even if households reconstruct to higher standards, common narratives around 'building back better' can conceal complex recovery trajectories fraught with trade-offs between rapidity of reconstruction, housing safety, taking on debt, and much more. Resilience gains in one dimension, like safer construction, can come at the expense of another dimension, like increased debt. For any large crisis, recovery trajectories are not linear, lack a clear end-point and rarely converge to the pre-disaster state. Such emperical evidence can support more effective and wholistic community recovery.

3. **We developed a metric to estimate differential post-disaster need in order to inform early decision making that fosters more effective and equitable recovery processes.** We demonstrate that factors contributing to post-disaster presistent need (PDPN) are to some extent predictable. As such, we propose and share a framework to indicate which communities are likely to have persistent needs and face obstacles to recovery. While probabilistic and uncertain, such additional sources of information are important when planning recovery activities and monitoring systems. Unlike the building damage mapping (conducted through G-DIF), which is an estimate of the immediate state of impact, the PDPN metric is a prediction of a future outcome of recovery. It highlights the fact that even though two communities might have been impacted equally, one might recover more quickly and to different standards because it benefits from certain socioeconomic or geographic factors that put it at an advantage over the other. Tools such as the PDPN metric are important because basing decisions solely on estimates of physical damage tends to prioritize addressing damage over addressing most critical needs. This can conceal potentially adverse tradeoffs and exacerbate pre-existing inequalities by supporting those with the greatest assets rather than those with the greatest needs.

Insights and tools developed are based on the Nepal 2015 Gorkha earthquake as our case study. In the aftermath of the earthquake, significant amounts of data was collected or produced to facilitate the response and recovery. For this reason, we are able to validate our methods with multiple existing datasets that have been produced for this event.

We openly provide the tools and frameworks developed in this study to the community of potential users. In designing them, we focused on making them flexible to diverse data inputs but context-specific to enable generalizability across regions and hazards, while also ensuring local validation through incorporating ground truth data inputs. We close this report with recommendations for use of the tools and ideas we share throughout. Most clearly, the tools to rapidly estimate damage and persistent needs can inform important post-disaster decision-making processes and existing planning processes such as the Post-Disaster Needs Assessment. Looking beyond physical impacts to buildings and infrastructure uncovers important elements of social vulnerability, around which recovery activities can engage more deeply. As such, the information systems underpinning early recovery planning can lead to more effective, resilient and equitable recovery.

Table Of Contents

Background & Motivations

PROBLEM

Disasters have complex, multi-faceted impacts in the short and long term. The information systems we currently use to understand disaster impacts are highly uncertain and leave us with an incomplete picture. We need new ways to estimate immediate and long-term effects of disasters that give us a more holistic understanding of their impacts.

1. Background & Motivations

Post-disaster recovery is a challenging process. Those empowered to lead this process – most often national governments, international financial institutions, and the humanitarian sector – must address at least three interlinked challenges. First, that disaster recovery decisions must be made with great urgency, limited information, and under considerable uncertainty[1]; second, that disasters disproportionately affect the poorest and most disempowered; and third, that disaster recovery is a complex and multi-faceted process where early recovery decisions will have long-term impacts[2].

Our project, Informatics for Equitable Recovery, develops tools to address each of these challenges by improving how stakeholders can map damage and integrate social vulnerability into post-disaster information systems.

1.1: Informatics: more transparent and reflective impact assessments to inform decisions

Effective recovery decisions require clear evidence. In particular, nearly all post-disaster recovery decisions rely on the availability of **disaster impact information** to understand how and where communities have been impacted. These sources and systems of information include risk models, crowdsourced maps, early field surveys, and more. Such informatics are also key to reducing the long-term impacts of disasters by supporting response and recovery efforts that build resilience in impacted communities.

In recent years, the amount of data rapidly available after a disaster has increased dramatically, especially with advances in technologies such as remote sensing and crowdsourced mapping. This proliferation of post-disaster information means that stakeholders must often sift through different types of data with varying levels of reliability to inform their decisions. Furthermore, current informatics tend to overemphasise losses to physical assets, thus leading recovery activities to be concentrated in the areas with the most **damage** rather than areas with the most **need** after a disaster.

New frameworks are therefore needed that are transparent, well-contextualised, and reflective of both damage and differential needs. Our project addresses these gaps by developing tools to produce **more accurate and transparent spatial estimates of impact**, by combining existing informatics and extending measures of impact from purely physical damage to differential need.

1.2: Equity: reckoning with differential needs and obstacles to recovery

No two households experience a disaster the same way. Households, communities, and even nations are affected by disasters in ways that reflect pre-disaster vulnerabilities, which are themselves products of past decisions[6]. In any community, factors like community infrastructure, access to markets, livelihood diversity, and social capital all shape vulnerability to natural hazards long before an earthquake or other event happens.

Despite this differential vulnerability being recognised[7], post-disaster recovery decisions often do not account for, or may at times exacerbate, pre-disaster inequalities[8,9]. This can happen despite the best intentions of recovery planners and stakeholders: in the urgency of post-disaster action, it can be challenging to identify which households or communities will have the greatest persistent needs or predict how policies might affect vulnerable groups differently. Tools that can help identify persistent post-disaster needs — essentially, identifying groups that face exceptionally intractable obstacles to recovering after a disaster — can inform how stakeholders plan, target, and monitor recovery programs.

New methods and tools are therefore needed to better identify, explain, and make actionable the diverse needs of households after a disaster. Obstacles to recovery also depend heavily on context, so close attention must be paid to the experiences of communities to identify early predictors of need that can be useful rapidly after an event. Our project illustrates a framework to estimate early trajectories of recovery accounting for pre-existing vulnerabilities, enabling local stakeholders to promote more equitable recovery outcomes.

1.3: Recovery: developing a holistic understanding of long-term disaster impacts

What exactly does it mean to 'recover' after a disaster? Post-disaster recovery is exceptionally challenging to define, since disasters can affect regions with very different pre-disaster characteristics, levels of wellbeing, and development trajectories. It is not always clear to what extent households facing such challenges can really be said to have 'recovered'[10]. Even so, scholars and practitioners continue to emphasise actions to link recovery processes with development, risk reduction, and resilience-building efforts, often referred to as 'building back better'[11].

While academic and policy debates about recovery and related concepts continue, recovery planners and stakeholders still need to define specific, actionable and reportable metrics of impact and recovery in order to carry out processes like Post-Disaster Needs Assessments (PDNAs), donor conferences for international support, and recovery monitoring after an event. Stakeholders already use some metrics of impact to inform recovery, but these often focus too narrowly on immediate damage to physical infrastructure[1]. Much more needs to be done to expand the assessment of disaster impacts beyond physical damage to incorporate non-physical aspects and reflect different

household recovery trajectories. Measures of impact and recovery should investigate more of the social and economic aspects of livelihood disruption caused by disasters — including factors like mental and physical health, incomes and debt, and issues like nutrition and sanitation.

Our project therefore develops new approaches to account for the multidimensional impacts of disaster and obstacles to recovery. Such approaches are needed for researchers and stakeholders to be able to inform recovery programs and interventions more holistically.

1.4: Human-centric informatics: debates linking values and practice

Our consortium is not the first to notice issues of equity in disaster recovery or to propose new frameworks to develop more human-centric disaster informatics. We draw on a rich literature from the social sciences, engineering, data science, science and technology studies, geography, human rights, development studies, and other fields to inform our approach to modelling, surveying, and understanding the differential impacts of disaster.

While we do not explicitly define a normative standard of equity here — as others have explored in numerous disasters and contexts [6, 12a, 12b] — we do highlight salient sources of differential vulnerability and need that can be incorporated into a stakeholder's assessment. We see our categories and informatics as wholistic but not exhaustive. For example, questions about differential access to resources and housing reconstruction outcomes are highlighted both in our data and in the literature, but we also incorporated other emergent factors such as debt and mental health based on our exploratory research.

Given that recovery, equity and resilience are complex and multi-dimensional problems, our approach to disaggregate and describe — rather than aggregate and advocate — enables stakeholders to more readily use sector-specific information to improve recovery plans. This approach allows us to contribute to debates on equity in disaster recovery while also producing informatics that are useful to stakeholders making decisions after an event. We welcome and encourage further debate on these topics, especially questions of equity in recovery and the use of evidence in post-disaster decisions. Other references and resources on this important topic include those listed in [6, 12a, 12b, 12c, 12d, 12e, 12f, 12g].

1.5: 'Informatics for Equitable Recovery': a new, transdisciplinary perspective on post-disaster information

Combining these three motivations, our project Informatics for Equitable Recovery brings together diverse disciplinary and methodological backgrounds including engineering, data science, applied statistics, mapping and geographical information systems, and the qualitative and quantitative social sciences. We combine these perspectives to develop tools that stakeholders can use to estimate post-disaster impacts and understand persistent recovery needs.

Study Region: Our team looks at post-earthquake impact estimation and recovery in Nepal, after the April 2015 Mw 7.8 Gorkha earthquake — the country's largest earthquake since 1934. The earthquake struck the Gorkha district near Nepal's capital city, Kathmandu, thus affecting both urban areas and remote rural regions. Massive amounts of data were collected and produced in the immediate response and recovery phases by local and international groups, precipitating a key moment in what some observers call Nepal's 'data revolution'[13, 14]. Notably, the Government of Nepal orchestrated a large-scale building-by-building survey to identify beneficiaries for its reconstruction assistance program focused on rebuilding safer homes[15]. We focus our study on the same districts where this census of surveys were carried out--the 11 rural districts of the 14 "severely hit" and "crisis hit" districts[16].

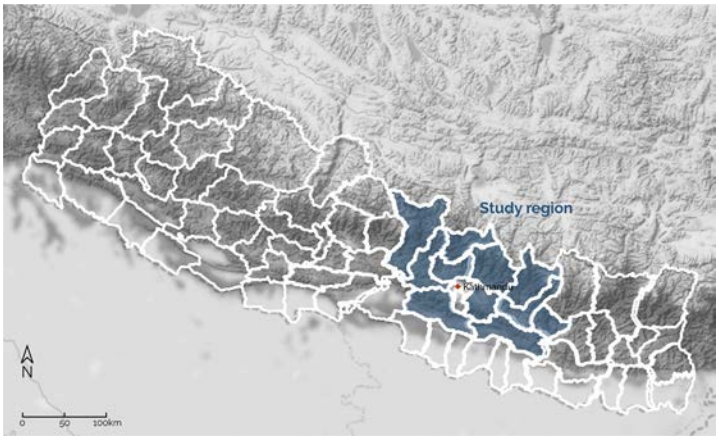
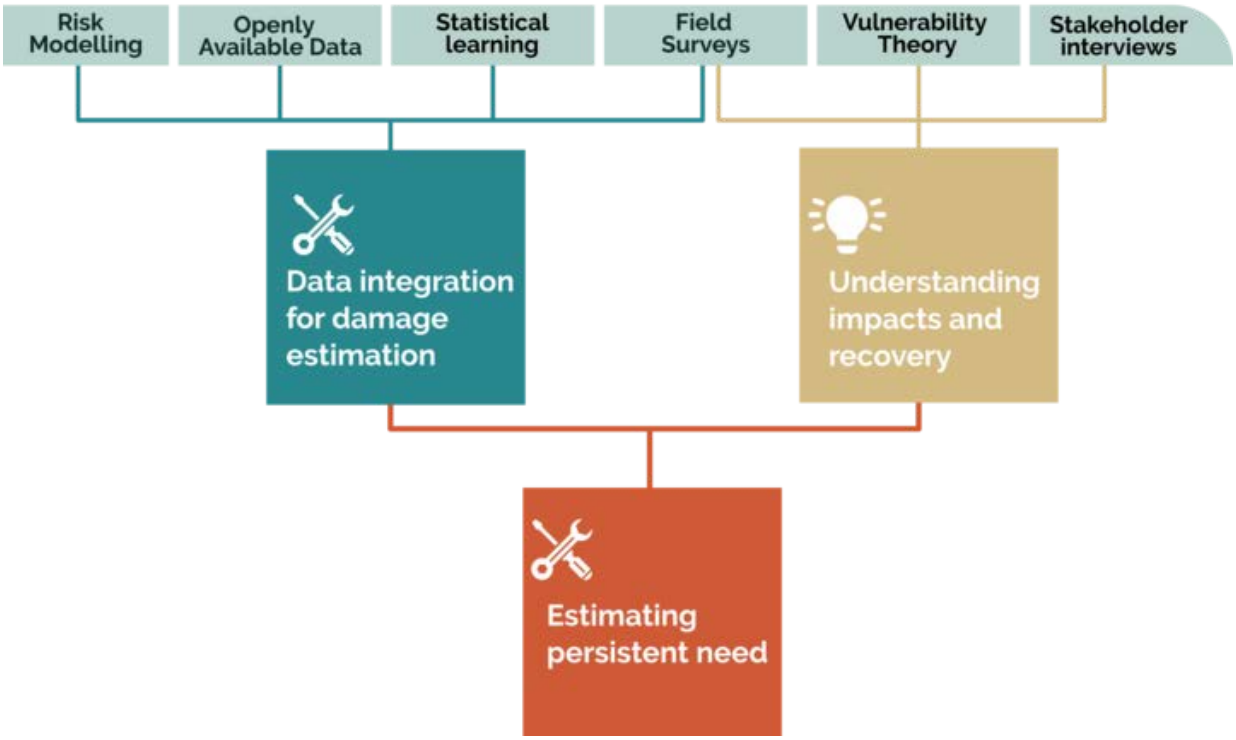


Figure 1: Region of Nepal where this study is focused. These are the 11 districts outside of Kathmandu Valley that were "most affected", as categorized by the Government of Nepal and the Nepal Post Disaster Needs Assessment.

Working closely with our team members in Nepal who were directly involved in post-disaster data crowdsourcing[17] and collection[18], we developed three work streams to develop informatics for equitable recovery.

The Informatics for Equitable Recovery Approach



The following sections describe our approach to developing methods to rapidly assess impact and recovery needs for disaster recovery planning. We develop two new models for (1) more accurate estimation and mapping of post-disaster damage (Section 4) and (2) prediction and mapping of post-disaster persistent need accounting for pre-existing community vulnerability characteristics (Section 6). Through extensive field interviews of households impacted by disaster, we further uncover and detail the complexity of disaster impacts and recovery trajectories (Section 5). We close this report with recommendations for use of the tools and the ideas we share throughout.

References

[1]Lallemant, David, Robert Soden, Steven Rubinyi, Sabine Loos, Karen Barns, and Gitanjali Bhattacharjee. "Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake." Earthquake Spectra 33, no. S1 (September 19, 2017): S435–51. https://doi.org/10.1193/120316EQS222M.

[2]Hallegatte, Stephane, Jun Rentschler, and Brian Walsh. Building Back Better: Achieving Resilience through Stronger, Faster, and More Inclusive Post-Disaster Reconstruction. World Bank, 2018.; Wisner, Ben, Piers Blaikie, Terry Cannon, and Ian Davis. At Risk: Natural Hazards, People's Vulnerability and Disasters. 2nd edition. London ; New York: Routledge, 2003.

[3]Tierney, Kathleen, and Anthony Oliver-Smith. "Social Dimensions of Disaster Recovery." International Journal of Mass Emergencies & Disasters 30, no. 2 (2012); Olshansky, Robert B., Lewis D. Hopkins, and Laurie A. Johnson. "Disaster and Recovery: Processes Compressed in Time." Natural Hazards Review 13, no. 3 (August 1, 2012): 173–78. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000077.

[4]Loos, Sabine, David Lallemant, Jack Baker, Jamie McCaughey, Sang-Ho Yun, Nama Budhathoki, Feroz Khan, and Ritika Singh. "G-DIF: A Geospatial Data Integration Framework to Rapidly Estimate Post-Earthquake Damage (In Review)." Earthquake Spectra, 2019.

[5]Lallemant et al., "Post-Disaster Damage Assessments as Catalysts for Recovery."; Hunt, Amelia, and Doug Specht. "Crowdsourced Mapping in Crisis Zones: Collaboration, Organisation and Impact." Journal of International Humanitarian Action 4, no. 1 (2019): 1–11.

[6]\Wisner et al., At Risk; Oliver-Smith, Anthony. "Anthropological Research on Hazards and Disasters." Annual Review of Anthropology 25, no. 1 (1996): 303–28. https://doi.org/10.1146/annurev.anthro.25.1.303.

[7]UNISDR. "Sendai Framework for Disaster Risk Reduction 2015 - 2030," 2015, 37.

[8]Ingram, Jane C., Guillermo F., Cristina Rumbaitis-del Rio, and Bjian Khazai. "Post-Disaster Recovery Dilemmas: Challenges in Balancing Short-Term and Long-Term Needs for Vulnerability Reduction." Environmental Science & Policy 9, (2006): 607–613.

[9]Peacock, Walter Gillis, Shannon Van Zandt, Yang Zhang, and Wesley E. Highfield. "Inequities in Long-Term Housing Recovery after Disasters." Journal of the American Planning Association 80, no. 4 (2014): 356–371.

[10]Smith, Gavin P., and Dennis Wenger. "Sustainable Disaster Recovery: Operationalizing An Existing Agenda." In Handbook of Disaster Research, edited by Havidán Rodríguez, Enrico L. Quarantelli, and Russell R. Dynes, 234–57. Handbooks of Sociology and Social Research. New York, NY: Springer, 2007. https://doi.org/10.1007/978-0-387-32353-4_14.

[11]Hallegatte, Rentschler, and Walsh, Building Back Better; Hallegatte, Adrien Camille Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. "Unbreakable : Building the Resilience of the Poor in the Face of Natural Disasters." The World Bank, November 24, 2016.

[12a] Anderson, William A. "Mobilization of the Black Community Following Hurricane Katrina: From Disaster Assistance to Advocacy of Social Change and Equity." International Journal of Mass Emergencies and Disasters 26, no. 3 (2008): 197–217

[12b] Jerolleman, Alessandra. Disaster Recovery Through the Lens of Justice. Palgrave Pivot, 2019. https://doi.org/10.1007/978-3-030-04795-5.

[12c] Doorn, N., Resilience indicators: opportunities for including distributive justice concerns in disaster management. J. Risk Res. 20, 711–731. 2017. doi:10.1080/13669877.2015.1100662

[12d] Doorn, N., Gardoni, P., & Murphy, C. (2019). A multidisciplinary definition and evaluation of resilience: The role of social justice in defining resilience. Sustainable and Resilient Infrastructure, 4(3), 1–12.

[12e] Rufat, S., Tate, E., Burton, C. G., & Maroof, A. S. (2015). Social vulnerability to floods: Review of case studies and implications for measurement. International Journal of Disaster Risk Reduction, 14, 470–486. doi: 10.1016/j.ijdrr.2015.09.013

[12f] Thaler, T., Fuchs, S., Priest, S., & Doorn, N. (2018). Social justice in the context of adaptation to climate change—reflecting on different policy approaches to distribute and allocate flood risk management. Regional Environmental Change, 18(2), 305–309. doi: 10.1007/s10113-017-1272-8

[12g] Muñoz, C., & Tate, E. (2016). Unequal recovery? Federal resource distribution after a Midwest flood disaster. International Journal of Environmental Research and Public Health, 13(5), 507. doi: 10.3390/ijerph13050507

[13] Ghimire, Subhash. "Data Revolution." MyRepublica. May 27, 2016. http://archive.myrepublica.com/2015-16/the-week/story/43132/banking-on-data.html.

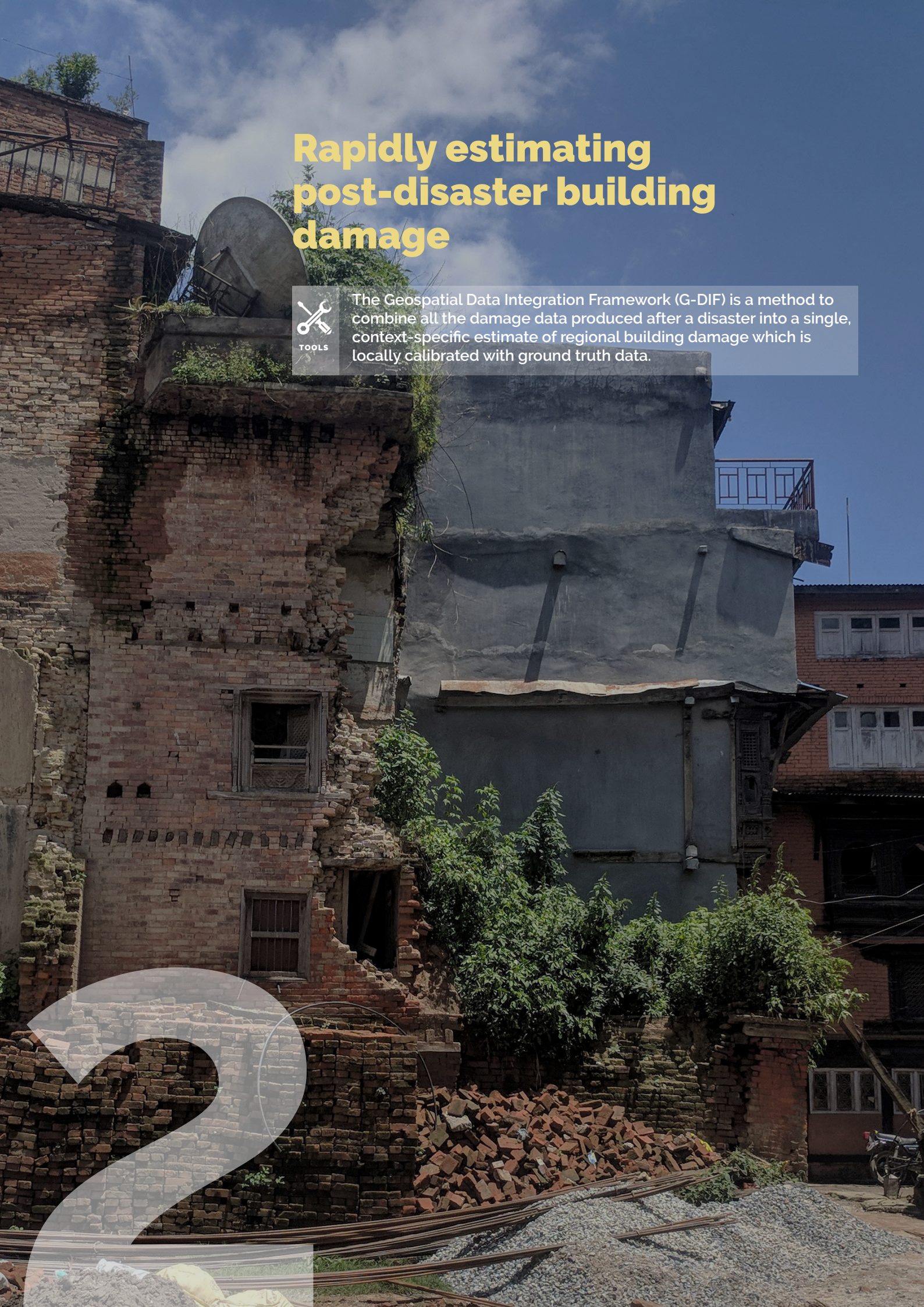
[14]Dennison, Louisa, and Pavitra Rana. "Nepal's Emerging Data Revolution," 2017.

[15]This survey was supported by Kathmandu Living Labs. Data is openly available at http://eq2015.npc.gov.np/#/about

[16]As classified by: Government of Nepal, "Nepal Earthquake 2015 Post Disaster Needs Assessment: Executive Summary," Kathmandu, Nepal, 2015.

[17]Kathmandu Living Labs, "Nepal Earthquake: Update from KLL Situation Room, June 1 2015"

[18]Kathmandu Living Labs, "National Housing Reconstruction Programme Project"



Rapidly estimating post-disaster building damage

TOOLS

The Geospatial Data Integration Framework (G-DIF) is a method to combine all the damage data produced after a disaster into a single, context-specific estimate of regional building damage which is locally calibrated with ground truth data.

2

2. Rapidly estimating post-disaster building damage

One of the most essential pieces of information after a disaster is the scale and extent of building damage. To prepare for the scale and need of the recovery process, the national government of the affected country will often conduct a Post Disaster Needs Assessment (PDNA) 4-6 weeks after the disaster, part of which requires an estimation of the "damage to infrastructure and physical assets"[19]. The PDNA is one example of a post-disaster sense-making process[20], and represents a crucial moment for stakeholders from different sectors — including housing, education, public health, heritage, and many others — to collectively gather evidence and understand a disaster's impacts.

For earthquakes and numerous other hazards, damage to the housing sector often comprises the largest portion of total losses. However, the method to estimate housing damage in the PDNA can be ad-hoc and potentially inaccurate due to the difficulty of synthesizing multiple datasets the government might receive from the ground and from outside data producers[21].

Many data producers recognize that data on building damage is required after a disaster. We see an increasing number of post-disaster damage maps, especially with the increasing availability of remote sensing data. Institutional examples include the USGS's PAGER estimate[22], Copernicus's manually conducted rapid emergency mapping products[23], and NASA's Damage Proxy Map[24]. The popularity of developing post-disaster damage maps is also reflected by multiple attempts to crowdsource building damage[25], recent competitions to use machine learning to identify building damage[26], and the focus within academia to develop similar damage identification models[27].

Generally, these rapid damage datasets that are produced after a disaster fall into three categories:

1. **Damage prediction models** are near-real-time predictions of regional impact available within hours, as soon as a map of shaking intensity is available, and usually based on simplified engineering risk analysis methods. Examples include USGS's PAGER system or the World Bank's GRADE[29].

2. **Remote sensing-derived damage data** are observations related to damage, retrieved from earth observation technologies such as sensors mounted on satellites, aircraft, or unmanned aerial vehicles. These observations are either interpreted automatically through an algorithm or manually by humans.

3. **Field surveys** surveys can take multiple forms, but most include at least an assignment of a damage state or damage grade to individual buildings on the ground. These surveys are typically carried out by engineering reconnaissance teams (e.g. EERI's Learning from Earthquake program), commissioned by the government, and/or carried out for the PDNA.

Both the damage prediction models and the remote sensing-derived damage datasets are useful, because they provide damage estimates over a large region in a relatively short timescale (most are available within the first 2 weeks). Unfortunately, these datasets can be inaccurate because they are either predictions that use global estimation methodologies or are observations from above (nadir) that cannot capture all types of damage (such as when the bottom stories of a building collapse on themselves yet the roof remains mostly intact[29]).

2.1: The Geospatial Data Integration Framework

The Geospatial-Data Integration Framework (G-DIF) capitalizes on all these existing data sources by using them as inputs to rapidly estimate building damage[30]. G-DIF requires a limited sample of primary damage data from field surveys, which are accurate but have low spatial coverage, to predict damage using secondary data from damage prediction models and remote sensing, which have lower accuracy but higher spatial coverage. In G-DIF, the data sources that are most accurate--or most related to the "true" building damage from the field surveys--get automatically weighted more importantly than those with inaccurate damage information. G-DIF results in a single map of building damage that is more accurate than any single damage dataset, especially at local-levels. This damage map can be updated as more field surveys (or other data) are collected, resulting in increasingly accurate estimates of damage.

Requirements for implementing G-DIF: *G-DIF requires a limited sample of primary damage data from field surveys, which are accurate but have low spatial coverage, to predict damage using secondary data from engineering forecasts and remote sensing, which have lower accuracy but higher spatial coverage.*

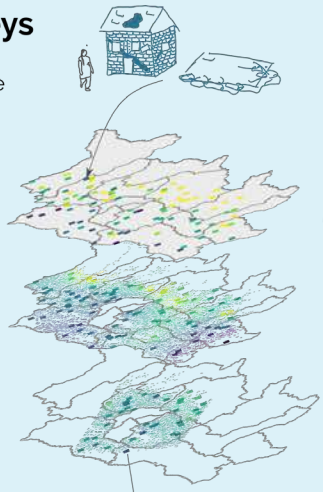
- A set of field surveys scattered throughout the area where you need to map damage
- Secondary damage data produced after the event. For earthquakes, this will likely include the USGS PAGER estimate and NASA's Damage Proxy Map. Other options include crowdsourced maps and alternative products that will likely be openly available on online databases (e.g. reliefweb.com).
- Other geospatial covariates that might inform building damage (example: USGS Shakemap, topographic slope, geological data, etc).
- Damage data integration model implemented in the source code using the R statistical computing software is available through the https://disaster-analytics.com/projects/ier-nepal/

As an example, we implement G-DIF as if we were to do it in the month after the April 25, 2015 Nepal earthquake. We use an example set of field survey assessments in our study region, taken from the survey coordinated by the Government of Nepal's Central Bureau of Statistics to identify beneficiaries for the government's National Housing Reconstruction Program (NHRP), hereon referred to as the NHRP dataset[31]. Again, this survey was carried out for all households in the 11 "most affected[32]" districts outside Kathmandu Valley, so we focus our analysis in these 11 districts. We combine a random sample of these surveys (100 grids, consisting of about 1000 buildings) and combine it with the following datasets:

1. **Shaking intensity -** A map of the shaking intensity produced by USGS's Shakemap
2. **Remote-sensing derived dataset -** The Nepal 2015 Damage Proxy map produced by NASA-JPL/ARIA
3. **Engineering forecast -** A prediction of total damage we developed using the USGS Shakemap and Nepali building fragility curves
4. **Elevation -** Another covariate that could be related to building damage (and ends up falling out of the prediction)

1 Collect field surveys

Trained researchers survey buildings at a random sample of locations throughout the affected area



2 Connect with secondary data

Merge field surveys with other damage data produced for the same event

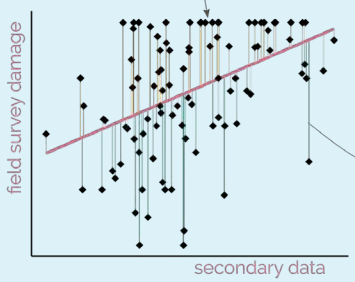
The Geospatial Data Integration Framework (G-DIF)

G-DIF is a *framework* that can be implemented in the weeks after a disaster to combine all the damage data that is produced after a disaster into a single map.

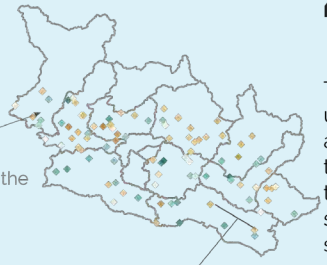
It is a framework because it is a set of models that can adapt to the data specific to each disaster.

3 Build trend model

The trend model relates the field surveys to all other secondary data. It's called the "trend" because it predicts the average spatial trend in damage



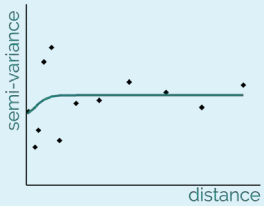
Calculate the error between the trend and field surveys



For every pair of two field surveyed points, calculate the distance and variance. Use this to make the spatial correlation model

4 Build spatial error model

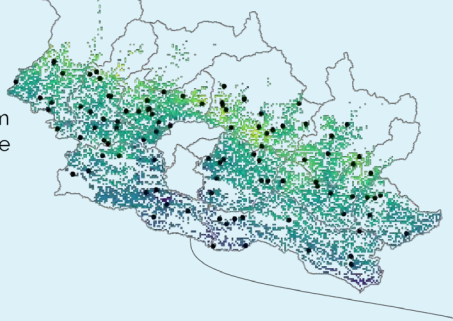
The spatial error model is used to estimate the error at all locations. "Error" refers to the difference between the trend and true damage. It is spatial because it is based on spatial correlation in the error.



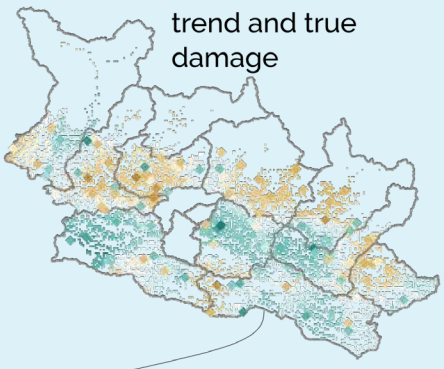
5 Predict the trend and error at all locations

Using the secondary data from (2) and the models from (3) and (4), we can predict the trend and error for all the locations we didn't survey.

Trend Average damage throughout space

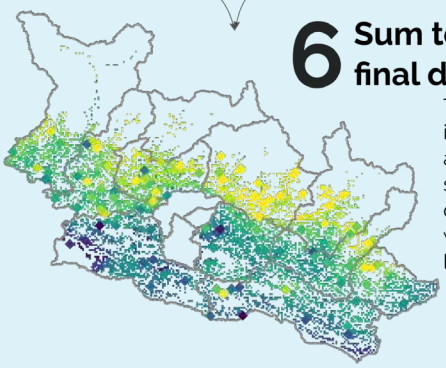


Error Difference between trend and true damage



6 Sum together to get final damage estimate

The final estimate of damage is simply the sum of the trend and the error. All units are the same as the field surveyed damage, so it's similar to if we were to field survey all locations.



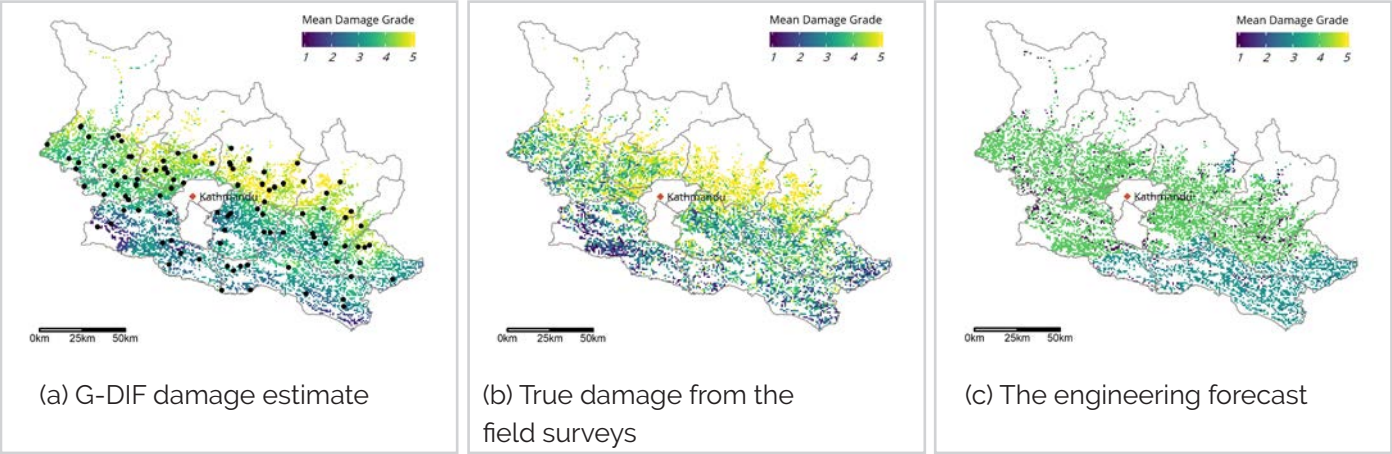
Advantages of G-DIF

G-DIF is ideal to implement for post-disaster impact assessments because

1. It can adapt to the data that is available
2. It automatically downweights inaccurate damage data inputs
3. It easily fits into existing impact assessment processes

2.2: Damage estimates with increased accuracy

Implementing G-DIF results in a map of predicted damage that is both visually and analytically more accurate than other datasets. As we can see below, the G-DIF estimate better captures local spatial patterns of damage than alternative estimates.



(a) G-DIF damage estimate

(b) True damage from the field surveys

(c) The engineering forecast

Figure 2: More accurate maps of damage. *A comparison of the map of damage from the (a) Geospatial Data Integration Framework (black points are locations of field surveys) with (b) the true damage from the NHRP field surveys and (c) the damage estimate from a damage prediction model. All maps show the average damage grade of all buildings within a grid for visual comparison. Simple visual inspection shows that the G-DIF damage estimate better captures damage and its spatial distribution compared the alternative damage prediction model.*

The close agreement between G-DIF and the true damage is expected, since G-DIF uses local measurements from field surveys to simultaneously calibrate and supplement datasets that might be prepared using global methods. The improvement in damage prediction is even further highlighted when looking at the errors at the local or district scale (Figure 3). For each grid (at whichever resolution we compute our prediction), we compute the error (or difference) between the predicted damage from G-DIF and the observed damage from the government led field survey. We can then develop histograms of error for all grids in a district or other area of interest. A well-performing model should result in an error histogram centered at zero (near-zero average error, also known as having little 'prediction bias') and with little spread (high precision, also known as having little 'prediction variance'). Looking at all 11 districts, G-DIF's damage estimate has lower bias and variance than the estimate from a damage estimation model--this increased accuracy becomes even more pronounced for local districts.

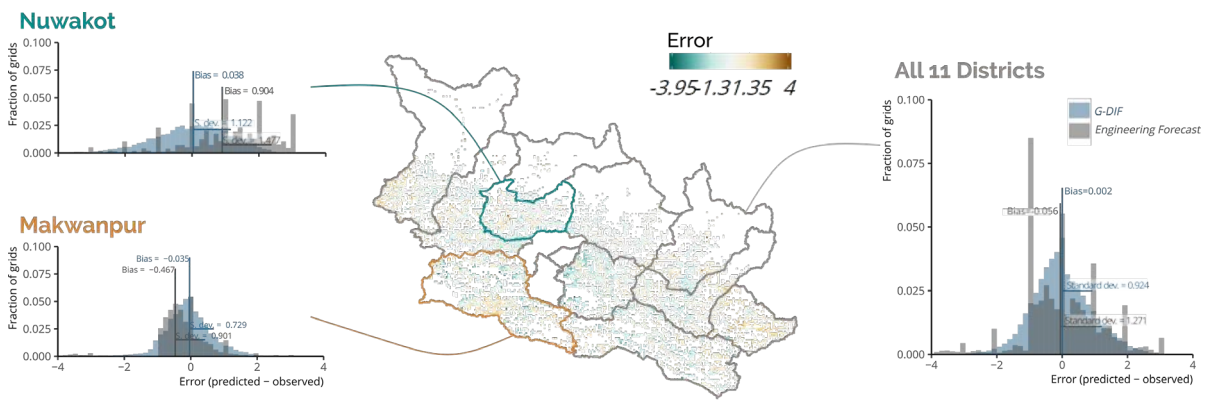


Figure 3: Enhanced Local Accuracy. *The damage estimate from G-DIF has errors (i.e. the difference between the true damage and the G-DIF estimate) that are centered at zero (low bias) and have a narrow spread (low variance) to their distribution. G-DIF has lower bias and variance than alternate damage prediction models.*

We also tested G-DIF with different amounts and locations of field surveys. With just 50 field surveyed locations (which is 0.06% of the total number of grids and between 250-1150 buildings), the G-DIF estimate results in a lower total error (mean squared error) than that of a damage prediction model (Figure 4). So, if just a small percentage of the buildings in the affected region are surveyed on the field, it's likely that the results from G-DIF will be more accurate than the typical damage prediction model in most data-limited regions.

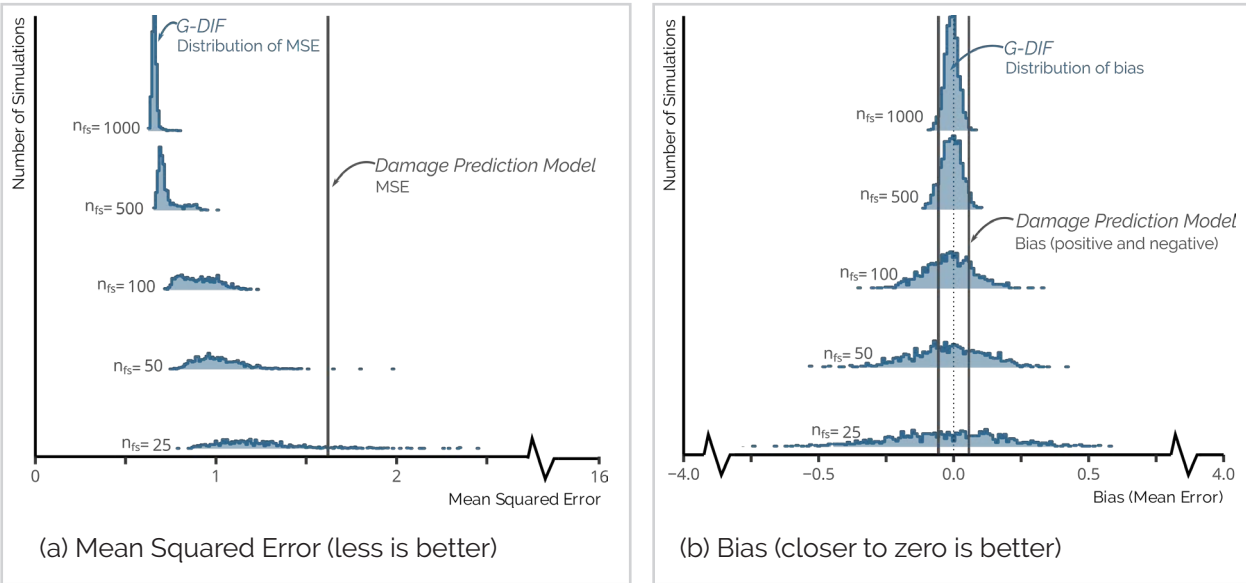


Figure 4: Improvement with more field surveys. *G-DIF's damage estimate has lower error as the number of field surveys (nfs) increases. In fact it can be updated over time as more data is collected.*

2.3: Advantages

The Geospatial Data Integration Framework can produce more accurate and higher-resolution estimates of damage after a disaster. It can be integrated in and inform numerous existing impact assessments, such as the Post Disaster Needs Assessment. The main advantages of G-DIF are that it's a flexible framework in terms of:

1. **Data:** It can integrate most data made available in a particular post-disaster context,
2. **Technology:** It is not reliant on any particular technology and could be used with future technologies (e.g. new remote-sensing technologies), and
3. **Context:** It is generalizable to other contexts since it is always locally calibrated with available field data.

While accurate estimates of physical damage are key for informing early response and recovery activities, such an impact assessment is only one factor in understanding and assessing the total needs for recovery.

References

[19]The European Commission, The United Nations Development Group, and The World Bank. 2013. "Post-Disaster Needs Assessments Guidelines."

[20] Soden, Robert, and Austin Lord. "Mapping Silences, Reconfiguring Loss: Practices of Damage Assessment & Repair in Post-Earthquake Nepal." Proc. ACM Hum.-Comput. Interact. 2, no. CSCW (November 2018): 161:1–161:21. https://doi.org/10.1145/3274430.

[21]D. Lallemant, R. Soden, S. Rubinyi, S. Loos, K. Barns, and G. Bhattacharjee, "Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake," Earthq. Spectra, vol. 33, no. S1, pp. S435–S451, 2017.

[22]Earle, Paul S, David J Wald, Kishor Jaiswal, Trevor I Allen, Michael G Hearne, Kristin D Marano, Alicia J Hotovec, and Jeremy M Fee. 2009. "Prompt Assessment of Global Earthquakes for Response (PAGER): A System for Rapidly Determining the Impact of Earthquakes Worldwide." U.S. Geological Survey Open-File Report 2009-1131, 15p.

[23]Dorati, Chiara, Jan Kucera, and Annett Wania. 2018. "Product User Manual for Copernicus EMS Rapid Mapping." https://emergency.copernicus.eu/mapping/ems/rapid-mapping-portfolio.

[24]Yun, Sang-Ho, Kenneth Hudnut, Susan Owen, Frank Webb, Patrizia Sacco, Eric Gurrola, Gerald Manipon, et al. 2015. "Rapid Damage Mapping for the 2015 Mw 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO-SkyMed and ALOS-2 Satellites." Seismological Research Letters 86 (6): 1549–56. https://doi.org/10.1785/0220150152.

[25]S. Loos et al., "The Development and Uses of Crowdsourced Building Damage Information based on Remote-Sensing," Blume Cent. Tech. Reports, pp. 1–130, 2018.

[26]R. Gupta et al., "xBD: A Dataset for Assessing Building Damage from Satellite Imagery," pp. 10–17, 2019.

[27]L. Dong and J. Shan, "A comprehensive review of earthquake-induced building damage detection with remote sensing techniques," ISPRS J. Photogramm. Remote Sens., vol. 84, pp. 85–99, Oct. 2013.

[28]R. Gunasekera, J. Daniell, A. Pomonis, R. A. D. Arias, O. Ishizawa, and H. Stone, "Methodology Note on the Global RApid post-disaster Damage Estimation (GRADE) approach," Washington, DC, 2018.

[29]Earthquake Engineering Research Institute. 2015. "Learning From Earthquake (LFE) Program." Oakland, CA. https://www.eeri.org/projects/learning-from-earthquakes-lfe/.

[30]Loos, Sabine, David Lallemant, Jack Baker, Jamie McCaughey, Sang-Ho Yun, Nama Budhathoki, Feroz Khan, and Ritika Singh. 2019. "G-DIF: A Geospatial Data Integration Framework to Rapidly Estimate Post-Earthquake Damage (In Review)." Earthquake Spectra.

[31]Openly available at the earthquake data portal: http://eq2015.npc.gov.np/#/about

[32]The Government of Nepal categorized 14 districts as"Most Affected" (originally categorized as "Severely Hit" and "Crisis Hit" in the PDNA), three of which were in Kathmandu Valley and not included in this study.

Going beyond building damage - understanding recovery

UNDER-STANDING

In rebuilding their houses, people face difficult choices between the pace of reconstruction, building more earthquake-resistant houses, avoiding taking on debt that will be difficult to pay off, and many other tradeoffs. The earthquake has also had important impacts on mental health, education, livelihoods, water supplies, and food security.

3

3. Going beyond building damage — understanding recovery

Improved rapid post-disaster estimates of housing damage, as outlined in the G-DIF framework above, can help national governments, international financial institutions, the humanitarian sector, and other stakeholders more effectively mobilize aid, as well as plan and implement recovery activities. This is a crucially important first step in helping affected people recover.

Key themes of qualitative interviews and household survey questions with residents and community leaders

Physical impacts: impacts on housing, water supply systems, roads, and community infrastructure.

Building back better: factors leading to whether, how, and at what cost people rebuilt houses that would be safer in any future earthquake

Socioeconomic impacts: impacts on livelihoods, mental health, food security and nutrition, education.

Community life and cooperation: the level of interaction and involvement with other members of the community

Home reconstruction: factors that helped or hindered the reconstruction of people's homes

Livelihood recovery: factors leading to whether people were able to re-establish their livelihoods, including family members living elsewhere and sending remittances

Our project also sought to understand the adequacy and/or limitations of measuring disaster impact solely in terms of building damage, and measuring disaster recovery solely in terms of housing reconstruction. We further aimed to improve our understanding of how recovery trajectories unfold over time, beyond building damage, and where diverse recovery needs may remain unmet in the medium- to longer-term. Toward this end, we carried out a field study in 2019, around four years after the earthquake. Physical housing recovery was a major focus, but we also looked broadly at diverse aspects of household well-being, including displacement, food security, education, and mental health. We began with in-depth interviews with residents, community leaders, and other stakeholders in the recovery process. Based on these qualitative findings, we developed and carried out a quantitative household survey (n=815) in the most-affected districts, accompanied by interviews with ward leaders in each surveyed community.

To enable comparisons with the NHRP survey completed in the first year after the earthquake, we sampled households from the 11 most affected districts in our study region. Surveyed areas are therefore mostly rural. The following summary synthesizes findings and reflections from both the interviews and survey.

3.1: Recommendations for conducting field surveys

We developed and implemented a survey of 815 households in 11 earthquake-affected districts in rural Nepal. This was a major undertaking that required the expertise, collaboration, and energy of field researchers, app developers, trainers, and social scientists. These nine reflections from our colleagues in Kathmandu Living Labs (KLL) describe replicable best practices for those considering deploying similar surveys for collecting recovery data.

Nine tips for recruiting, training, and mobilizing field researchers: Reflections on field data collection

1. **Identify people with energy and passion.** Field work is physically challenging, especially in difficult terrain. So, do we look for people with tons of experience in field work or do we recruit people who have little or no experience but are full of energy and passion? There is a trade-off. Also, keep in mind that the former group may have fatigue from their field experiences. In our project, we decided to recruit young people with energy and passion.

2. **Reconceptualize the predominant view of the field data collector.** Often, field data collectors are narrowly viewed as 'enumerators' or 'surveyors'. Field work encompasses much more than mere data collection. Before collecting data, researchers have to understand local power dynamics, identify key influencers and take them in confidence, and develop rapport with the responder. Collecting data is only one part of the puzzle. So, we rightfully called them 'field researcher'.

3. **Help field researchers understand other benefits of field work.** In many cases, field mobilisers and field workers see the exchange of a certain number of data records and their remuneration as the sole contract. In reality, field researchers gain much more than the money they earn. They develop different skills: technological skills, social skills, research skills, and get an opportunity to visit a new place, expand their networks, enrich their CVs, and more. For example, our field researchers shared that they developed mapping skills and improved their ability to communicate through their field experiences. In order to help them understand these additional benefits, we integrated discussions on personal and professional development as part of the training.

4. **Emphasize the contribution that field work makes.** People are often looking for the meaning of their work. When field researchers understand the meaning of their field work, both the quantity and quality of data increase. Completing field work well requires both extrinsic and intrinsic motivation. Each piece of data is collected to answer certain research questions, inform policy decisions, or change the lives of people. During their training process, we were careful to situate their contributions in the scheme of things, including explaining the motivations of our project and the scope of our research.

5. **Let the field researchers know early on that this is going to be a different experience for them.** The field researchers need to know from the very beginning that the work they are going to undertake is going to be different from their experiences in the field. Fieldwork in post-disaster contexts requires exceptional nuance and sensitivity, and we emphasised these practices in our training, check-in, and debriefing processes.

6. **Do not just ask for numbers of records; ask also about researchers' wellbeing in the field.** In the field, people may encounter adversity. They may have been going through all sorts of difficulties such as lack of food, sleep and other anxieties. It is difficult to build rapport with new people every day, and it is difficult to ask about a disaster that for many was also a tragedy. Researchers need to feel that their difficulties are heard and their wellbeing is cared for, especially by the organisers of the study. After all, field researchers are human. To ensure their wellbeing, we constantly stayed in touch with field researchers throughout the survey process. We also encouraged them to push themselves but never at the cost of their own wellbeing.

7. **Stand on the shoulders of giants.** It is always important to have an authorised entry into the field areas so as to maximise local support and minimise adversities. We secured support from NRA in the form of a letter and also encouraged all the field researchers to get in touch with local governments of their sample areas. This radically changed their field experience, granting them smooth access to many survey sites.

8. **Build on past experiences.** Our local collaborating team, Kathmandu Living Labs (KLL), had already done several field surveys in the past, some of which were large-scale mobilisations. Many of the tips listed above were, in fact, inspired by KLL's prior work. We used those lessons in this survey and it worked.

9. **Encourage them to think of themselves as field researchers beyond the field experience.** It was important for us to not only train the researchers in field work, but in overall research methods, research ethics, questionnaire development and sampling. Furthermore, we also tasked them with reflecting on their overall experience and presenting their learning and challenges with the entire team. We developed a skilled team that will sustain itself beyond this one field work.

3.2: The factors that shape reconstruction

> **"The earthquake destroyed something that people had spent all their savings on: their houses."**
> – Municipality Mayor

Destruction of housing is an especially tangible and far-reaching impact of an earthquake. Even as our interviews tried to explore more indirect impacts beyond building damage alone, we found that residents and community leaders tended to steer conversations back toward the damage to houses, the difficulties living in temporary shelters, and the difficulties in starting and completing the reconstruction or repair of their homes. Housing destruction is a major impact with both immediate and longer-term consequences.

We examined three main aspects related to housing reconstruction: how quickly people were able to rebuild, the quality of their rebuilt homes (size, comfort, and earthquake resistance), and whether people were able to rebuild without compromising their future well-being (by selling assets or taking on burdensome debt).

Reconstruction was fastest for those whose houses had suffered the greatest damage in the earthquake. We found that most of those with high levels of earthquake damage have finished reconstruction, while less than half of those with low levels of earthquake damage have finished their repairs. At first this would seem counterintuitive: a complete rebuild is more expensive and requires more labor than minor repairs. However, households at higher levels of damage would find their house uninhabitable and families would typically live in a temporary shelter, adding impetus to finishing reconstruction. In addition, the repair of an earthquake-damaged house is typically complex and unfamiliar to local construction workers, whereas complete reconstruction is more straightforward.

Most families in rural Nepal lived in stone-and-mud houses at the time of the 2015 earthquake. Such structures h ave very low earthquake resistance; tragically most of these houses collapsed partially or totally in the earthquake. One community leader described the earthquake destruction as *"like the effect of a bomb – smoke was rising as mud houses were destroyed".*

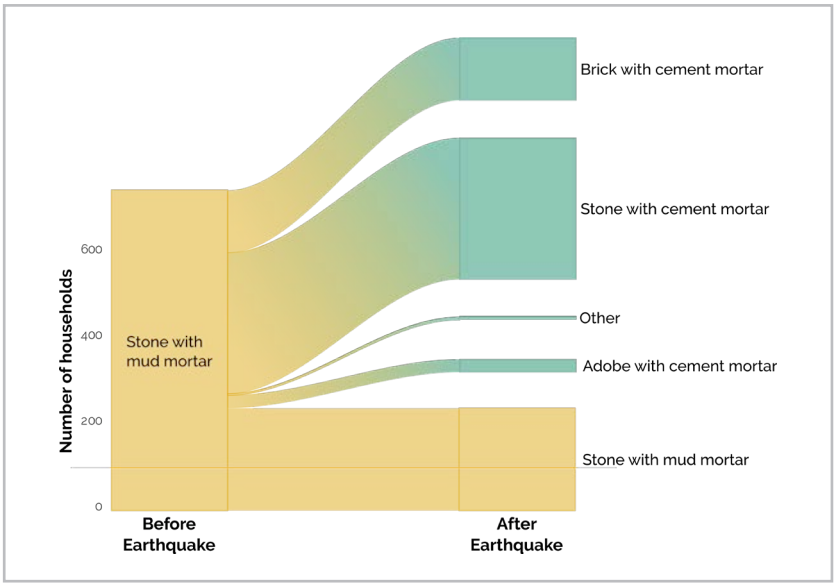


Figure 5: Most households rebuilt with safer materials. *After the earthquake, more than two-thirds of households in our survey rebuilt their house using materials that were safer than the original stone with mud mortar.*

The destruction and trauma wrought by the earthquake brought the earthquake safety of buildings to the forefront of people's minds. In addition to this, the National Reconstruction Authority (NRA) endeavored to provide technical assistance for safer reconstruction by sending field engineers to affected areas and by conditioning the distribution of financial assistance on reconstruction progress following their minimum standards. More than two-thirds of households in our survey rebuilt a house with incrementally safer materials – replacing mud with cement and in some cases replacing stone with brick. Having had a visit from a field engineer was instrumental in the decision to rebuild using safer materials.

The cash grant from the NRA helped people get started but was not enough to rebuild an entire house[33]. People appreciated receiving this assistance and appreciated receiving the technical advice on home design, but there remained a major cost burden that each household had to cover. This cost burden was greater for those who had decided to rebuild by buying safer materials – cement and brick – rather than re-using local stone and mud. These households faced additional difficulties in procuring building materials that were in high demand and transporting these to their home sites.

Consequently, many households have gone into debt, often at high interest rates. Households that are rebuilding with safer combinations of materials tend to take longer to finish reconstruction, tend to build houses that are smaller than their pre-earthquake houses, and tend to go into debt by doing so. People are concerned about being able to repay their debts. While many households have been able to recover their income to pre-earthquake levels, they no longer consider this income sufficient for their needs, as they now have more debt.

> **"I had a lot of difficulties. It feels like this one life has already been three lives."**
> – Resident who lost his wife to the earthquake and lost his farmland to landslides

3.3: Socioeconomic impacts and recovery

Beyond damage to buildings and the costs of rebuilding, an earthquake, like any disaster, has far-reaching effects. Many of these further effects are not as immediately visible as physical damage. Many residents and community leaders told us that there remain substantial unmet needs from the mental trauma related to the earthquake. Nearly a tenth of our survey sample had two or more indicators of post-traumatic stress symptoms (on a four-item scale). Women were more likely to have post-traumatic stress symptoms, as were those who had lost a family member in the earthquake, those who experienced an additional disaster after the earthquake, those who had not yet finished reconstructing their home, and those who lived in more geographically isolated areas. Though our survey interviewed adults, community leaders told us that school children have also had psychological difficulties.

Education was interrupted as damaged schools needed repair or replacement, while teachers were themselves busy rebuilding their own homes and

> **"Even now, students are afraid if they hear an abrupt noise and think of the earthquake. They have not been able to study well."**
> – Community leader

everyone was coping with the aftermath of the disaster. Water supplies were disrupted in many communities as springs dried up and piping systems were damaged. Hunger slightly increased and nutrition quality slightly decreased as attention turned from farming toward cleaning up and rebuilding after the earthquake. Prices for many goods temporarily increased as demand went up and, in some cases, bridges, roads, and paths were disrupted by the earthquake and resulting landslides. These diverse impacts negatively affected people's well-being, especially in the first year after the earthquake. Four years on, there has generally been recovery in these areas, though in many areas the recovery remains incomplete.

COMMUNITY LEADER PERSPECTIVES

When asked what could be done to help recovery from an earthquake or another disaster to be more swift and complete, **community leaders** suggested a wide variety of needs in areas that bridged from disaster recovery to development in general:

More aid, more rapidly delivered:

- Larger reconstruction grants for homes
- More technical assistance for building design
- Faster provision of these forms of aid

New forms of recovery support:

- Low-interest loans for homeowners and businesses
- Psychosocial support

Support for development and resilience-building:

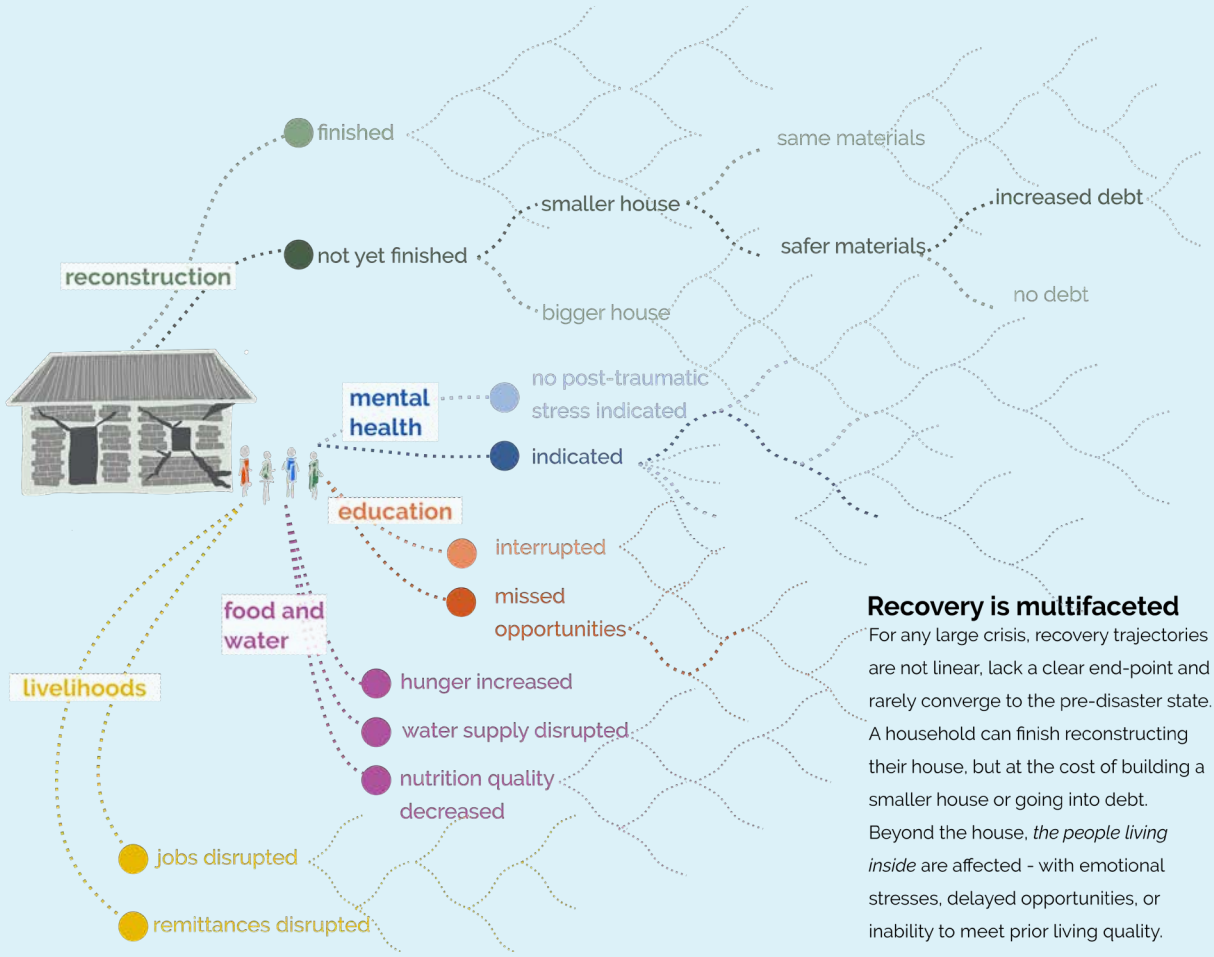
- Better infrastructure: roads, water supply, community facilities that are less vulnerable to earthquakes and landslides
- More construction following earthquake-resistant design
- Vocational, sanitation, and disaster preparedness training
- Support for vulnerable populations
- More facilities and support for health and education

Despite the broad impacts, many aspects of life did not fundamentally change because of the earthquake. People generally recovered their livelihoods, though not without difficulty. At the time of the earthquake, many family members were living abroad and sending remittances. These families faced the difficult tradeoff between having their relative return home, to help with the recovery and provide emotional support, and having their relative stay abroad, to maintain the much-needed remittance income. Though there were many cases of return migration earlier than planned, many families have since sent a family member abroad again to restore that source of income, though at the cost of additional travel and fees.

There were also areas of improvement: though the earthquake damaged roads, many communities say their accessibility is now better than it was before the earthquake, reflecting a general trend of road building and improvement in Nepal. Sanitation conditions improved in some communities where programmes to build better toilets were carried out.

The earthquake was a physical event at one moment in time. The most immediate, tangible, and at times tragic impact was the destruction of people's homes. From this one form of impact, over the following years many other difficulties arose indirectly in many other aspects of people's well-being. Some indirect impacts may have intergenerational consequences, especially disruption to education, nutrition, and the taking on of burdensome debt.

Despite the broader set of impacts that arise indirectly from building damage, finishing rebuilding their homes remains a major focus for residents. It can be difficult to attend to other problems when the challenges of rebuilding a home and the discomfort and insecurity of living in a temporary shelter all occupy so much of people's attention, effort, and investment. For these reasons, in the next section we return the focus to the completion of home reconstruction. We turn to the question of whether it may be possible, at the outset of a post-disaster reconstruction, to use pre-existing datasets to identify geographical areas where people may have greater difficulty rebuilding their homes.



Recovery is multifaceted

For any large crisis, recovery trajectories are not linear, lack a clear end-point and rarely converge to the pre-disaster state. A household can finish reconstructing their house, but at the cost of building a smaller house or going into debt. Beyond the house, *the people living inside* are affected - with emotional stresses, delayed opportunities, or inability to meet prior living quality.

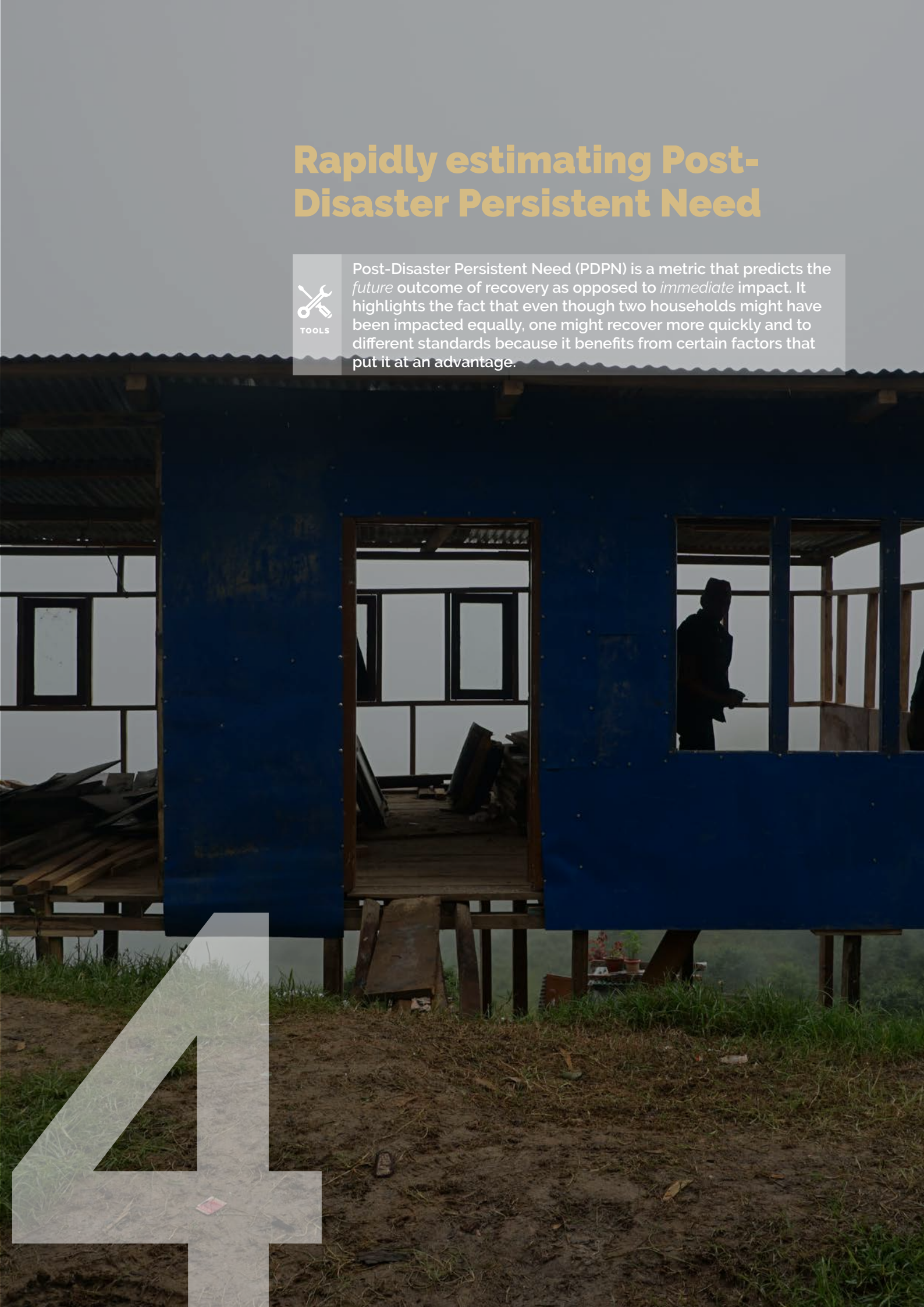
References

[33]D. Coyle, "Modelling Vulnerability & Shelter: Hamro Ghar Approach," Kathmandu, Nepal, 2019.

Rapidly estimating Post-Disaster Persistent Need

TOOLS

Post-Disaster Persistent Need (PDPN) is a metric that predicts the *future* outcome of recovery as opposed to *immediate* impact. It highlights the fact that even though two households might have been impacted equally, one might recover more quickly and to different standards because it benefits from certain factors that put it at an advantage.



4

4. Rapidly estimating Post-Disaster Persistent Need

As previously explored, assessing building damage alone is insufficient to capture the complexity of disaster impacts and resulting needs for recovery. Two equally impacted households might recover at entirely different rates and to different standards due to different pre-existing socio-economic factors —such as prior saved assets or access to community resources and infrastructure—that put one at an advantage over the other.

4.1: The meaning of Post-Disaster Persistent Need

We develop a new metric, called **Persistent Recovery Need (PRN)**, that attempts to capture those combinations of factors leading to consistent and continued inability to recover from a disaster. While building damage maps are crucial information for planning and prioritizing recovery assistance, Post-Disaster Persistent Need attempts to identify populations at high risk of being unable to recover despite standard assistance. It serves as valuable and actionable information to be mapped in conjunction with building damage.

As opposed to the building damage assessment (conducted through G-DIF), which is an estimate of the immediate state of impact, the PDPN metric is a prediction of a *future outcome of recovery.* The PDPN metric shows the likelihood of not recovering given that a community has suffered a certain level of impact and possesses certain pre-existing vulnerability indicators.

Post-Disaster Persistent Need can be explored and predicted for different metrics of recovery depending on available data and interests of stakeholders (e.g. housing recovery, livelihood recovery, etc). Here, we look at the likelihood to reconstruct one's house as our recovery outcome due to its importance in the recovery process and its influence on various other household well-being factors. Given the level of damage at the time of the disaster (point A), what is the probability that a house will stay damaged (point B1) versus reconstructed (point B2) in t years after the disaster? In this case, damage is the measure of impact and housing reconstruction is the measure of recovery. Factors that are obstacles to a household's ability to reach full reconstruction could be their access to certain construction materials or socioeconomic status (in orange).

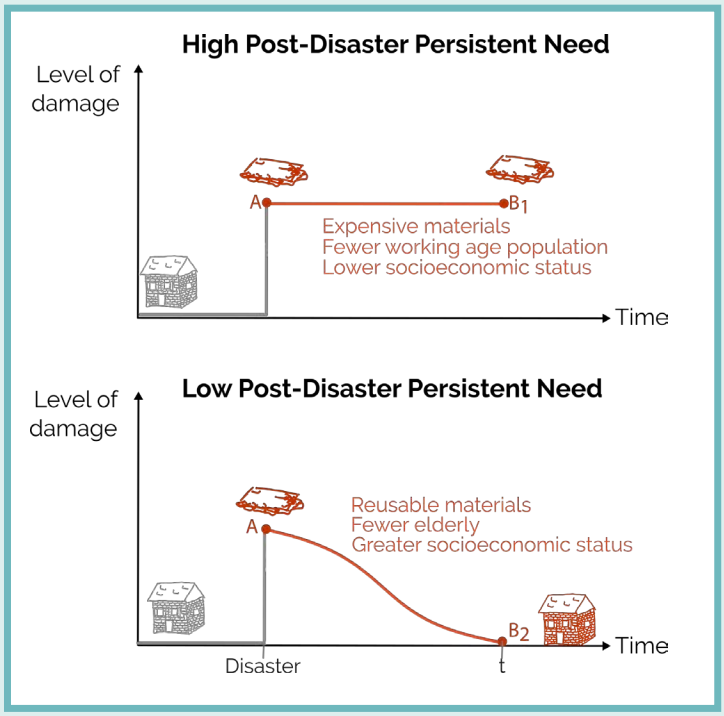


Figure 6: The meaning of Post-Disaster Persistent Need for housing recovery. *Two housholds that have been equally impacted may face different needs during the recovery process. A household that makes no reconstruction progress (top) will have greater Post-Disaster Persistent Need.*

While G-DIF is calibrated in near-real-time based on field observations of damage collected rapidly after a disaster, PDPN is applied to the current disaster using a model that is calibrated with recovery data from a previous disaster in a similar context. This is because we *weigh the importance of different obstacles to recovery based on their past relationships to recovery progress.*

4.2: Approach to modeling need

For Nepal, we develop the model to predict the PDPN metric using responses to questions from our field surveys as measurements of damage and reconstruction progress. We use these field surveys to calibrate a model that predicts the likelihood to recover (only for severely damaged houses, ~400) by relating these recovery outcomes to spatial proxy data that represent different factors of recovery. The spatial proxy data we use would be available rapidly after an event in most developing contexts, and includes data from census, remote-sensing, or geospatial models[34]. Here, we examined the utility of a suite of 38 spatial proxies (see Annex) in terms of predicting the likelihood to reconstruct. The overall model to predict the PDPN metric is shown in Figure 7.

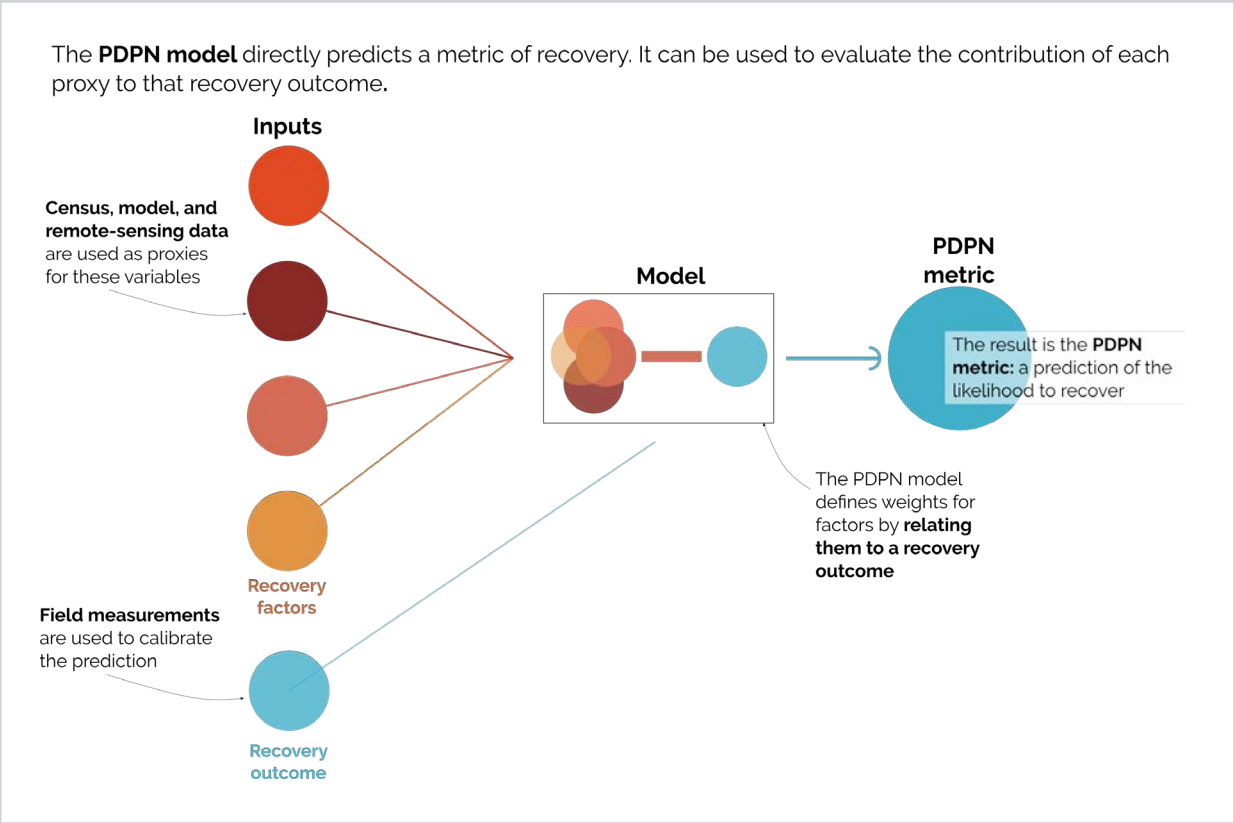


Figure 7: Predicting Post-Disaster Persistent Need. *We develop a model that relates recovery outcomes from a previous disaster with proxies for obstacles to recovery. In doing so, the PDPN metric is a direct metric of recovery, predicting the likelihood of a certain outcome years after the disaster.*

Our approach to predict Post-Disaster Persistent Need is different from typical resilience or vulnerability indicators which aggregate factors of resilience together into a weighted average (where weights can be defined through multiple methods) without relating them to recovery outcomes of interest. We can see the difference by comparing our PDPN model to the model to develop resilience indicators in Figure 8.

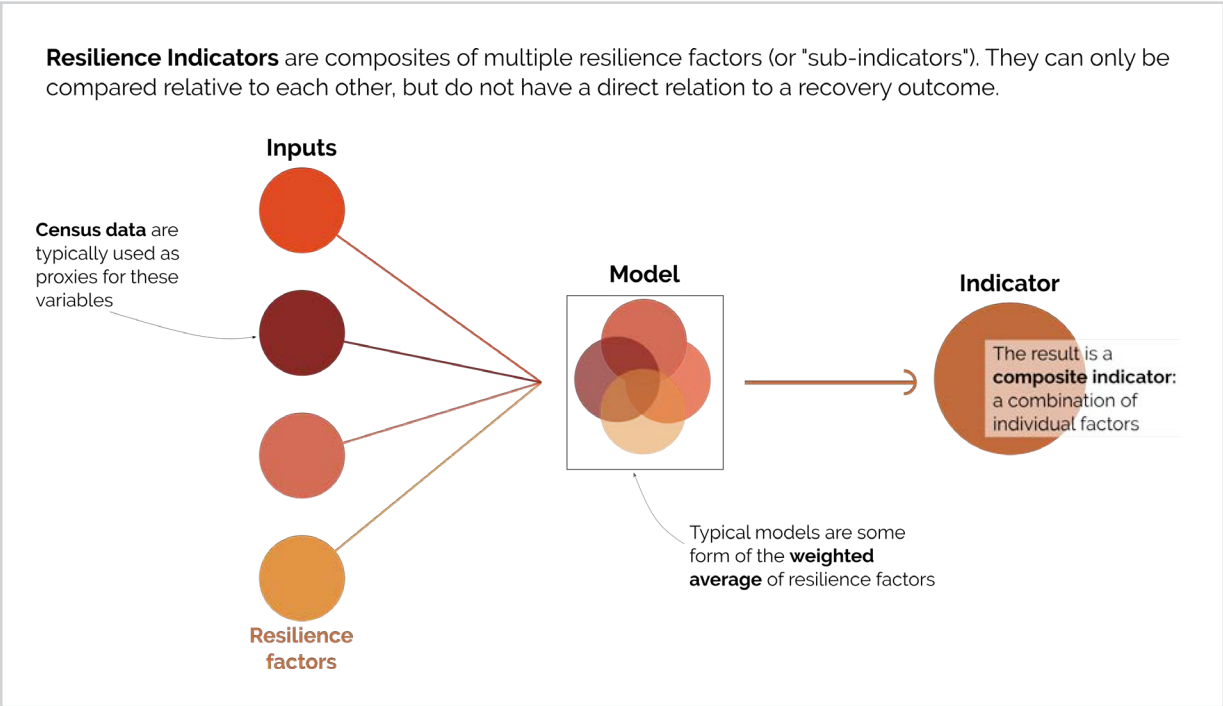


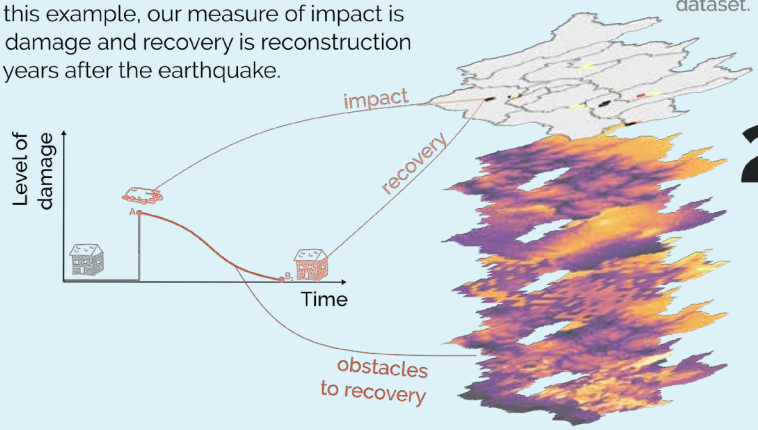
Figure 8. The indicator method to quantify resilience. *Typical resilience indicators quantify resilience by aggregating multiple factors of resilience together. While this method is useful to track indicators over time or compare them to each other, they can not directly related to metrics of recovery (such as reconstruction or livelihood recovery).*

This framework relates proxies to the probability of a household recovering in a certain amount of time (here t=4 years) after a disaster. Since we are predicting the probability of a household being in one of two classes--constructed or not--we compare classification models that can predict probabilities. For Nepal, we compare two machine learning models to predict the probability of a damaged or collapsed home reconstructing in 4 years, and find that both random forests[35] and logistic regression[36] have comparable results. The random forests prediction performs slightly better (as evaluated by the area under the receiver operating characteristic curve through cross-validation), and provides more realistic probabilities, so we show results from this model for the rest of this section. For more information on this modeling process see the Annexes for links to our project code, or to more generally understand the use of machine learning for disaster risk management, see The World Bank GFDRR's Machine Learning for Disaster Risk Management guidance note[37].

1 Collect impact and recovery data

To build a model to predict persistent recovery need, we first need high-resolution measurements (from the field) of household impact and recovery from a previous disaster. In this example, our measure of impact is household damage and recovery is reconstruction progress 4 years after the earthquake.

Connect recovery measurements to spatial proxies to build training dataset.

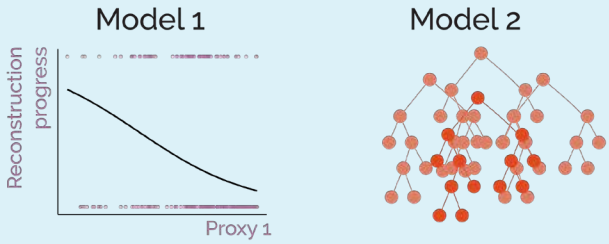


2 Gather proxies for obstacles to recovery

Our goal is to predict the likelihood to recover at all locations in the affected region. To do this, we use spatial proxies for the obstacles to recovery. In Nepal, we tested 38 different proxies from spatial models, remote-sensing, and census

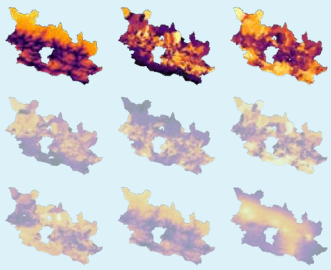
3 Train PRN model(s)

Once we connect the data from 1 and 2 together, we can train models to predict the likelihood to recover (y) using data on the obstacles to recovery (x). In this example we test two models: a logistic regression and a probability random forest.



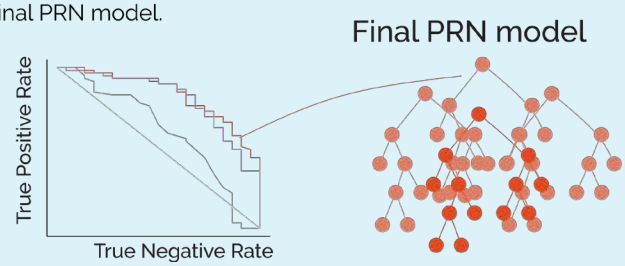
4 Select most important proxies

To ease data collection after future disasters, we want to remove proxies that are non-informative or redundant. The selection of proxies should be semi-automated to remove modeler bias, but also evaluated by the modeler to ensure the selected variables make sense.



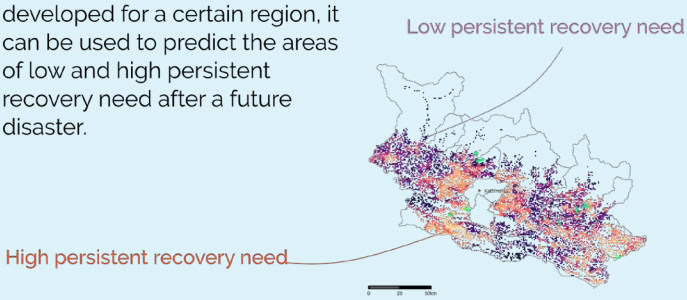
5 Retrain, evaluate, & select best model

Once the most important variables are selected, each model should be retrained and evaluated on a validation set. Since we are predicting probabilities, we recommend using the model with the highest area under the ROC curve as the final PRN model.



6 Use PRN model in next disaster

Once the PRN model is developed for a certain region, it can be used to predict the areas of low and high persistent recovery need after a future disaster.



4.3: Mapping the spatial distribution of need

While our approach reduces the generalizability of individual PDPN models, **it simply reflects the reality that obstacles to recovery are context specific.** For example, we found that remoteness[38] is a significant factor in the ability of communities to recover in rural Nepal, but we might expect this to differ in dense urban environments, or island contexts. The model developed here is therefore tailored to the context of mountainous rural areas in South-Asia, with the most important proxies for prediction shown in Figure 9.

Here, we estimate the PDPN metric (the probability of not finishing reconstruction) for all grids in the affected region[39]. We can interpret the final estimates shown in Figure 10a as the probability of every household within a grid not completing reconstruction four years after the earthquake. model developed here is therefore tailored to the context of mountainous rural areas in South-Asia, with the most important proxies for prediction shown in Figure 9.

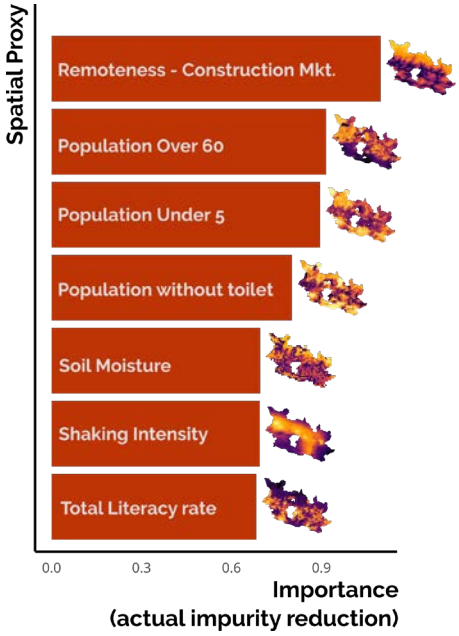
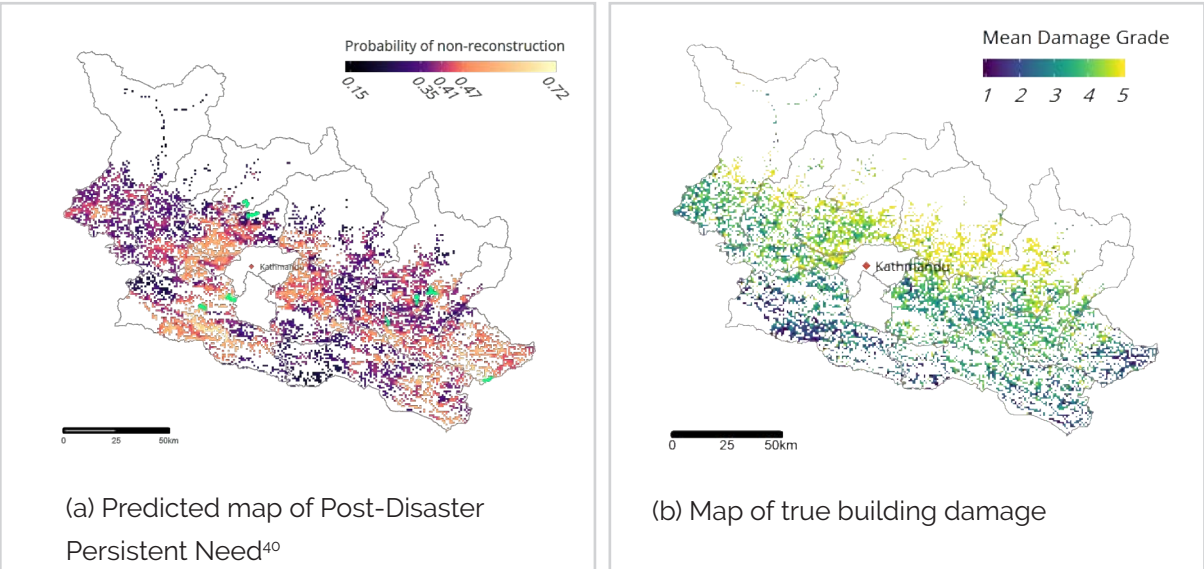


Figure 9. Important proxies to predict housing recovery in Nepal.



(a) Predicted map of Post-Disaster Persistent Need[40]

(b) Map of true building damage

Figure 10. The map of Post-Disaster Persistent Need is not the same as that of damage. *The difference in these two maps exemplifies the fact that maps of building damage alone may be insufficient to identify populations that will experience the greatest obstacles to recovery years after a disaster.*

Shown side-by-side with the true damage from the earthquake, we can see that the two maps highlight different spatial distributions of impact. The map of PDPN on the left shows areas with greater persistent recovery need towards the southeast, while the damage map highlights the areas with the greatest impact in the areas towards the north. The difference between these two maps clearly show that **using building damage as the only basis for recovery decisions may be insufficient to help those that will experience persistent needs throughout the recovery process.**

4.4: Advantages

This framework to develop a more needs-based metric for impact was built and tested using data from Nepal and should be tested for multiple locations and recovery outcomes. The general model-building process can also be implemented for other sectors, as long as a measure of impact and recovery can be collected. Though, given the prevalence of reconstruction progress data over other measurements of recovery, it may be simplest for stakeholders to start with housing recovery and then move to other sectors.

Requirements to develop a PRN prediction model in your region: *The PDPN model requires a scattered set of field surveys of a recovery metric (i.e. reconstruction progress) from a previous disaster and spatial proxies for the obstacles to that recovery (for example, proportion of elderly people)*

- Field surveys of a **recovery outcome** distributed over a region of interest
- A **measure of impact** at those same locations (either from field surveys or another proxy)
- Spatial proxies for likely **factors affecting recovery** from either models, remote-sensing, or census data
- PDPN model implemented in the source code using the R statistical computing software is available through the project resources page

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[41]. Using these two maps together, we are able to identify areas with high damage that will likely experience persistent recovery needs over time after the disaster happens. This expands our definition of impact informatics to be forward-looking and holistic, because it predicts future recovery given all the multifaceted characteristics that influence it.

Importantly, the same model used to predict recovery need in a post-disaster situation can also be used for risk reduction prior to a disaster. Given predicted damage from a standard earthquake hazard model, the PDPN framework can identify communities likely to struggle in recovery. These can be targeted for seismic risk reduction interventions and social programs to increase their resilience prior to an inevitable earthquake. As such the model can be used to promote equitable recovery following an earthquake as well as promote equitable (and reduced) risk prior to it.

References

[34] The selection of proxies that go into our model was collectively decided and driven by theory and our interviews with community leaders. To balance this with data-driven evidence, we let the model decide which variables are most predictive of recovery.

[35]Random forests are an ensemble statistical learning method which averages the results of a large number of individual, decorrelated decision trees. For more information see: Breiman, 36Leo. 2001. "Random Forests." Machine Learning 45 (1): 5–32. https://doi.org/10.1017/CBO9781107415324.004.

[37]Logistic regression models predict the log likelihood of an outcome given a set of predictor variables. For more information see: Hastie, Trevor J, Robert J Tibshirani, and J Jerome H Friedman. 2009. The Elements of Statistical Learning. Springer.

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

[39]R. Banick and Y. Kawasoe, "Measuring Inequality of Access: Modeling Physical Remoteness in Nepal," 2019.

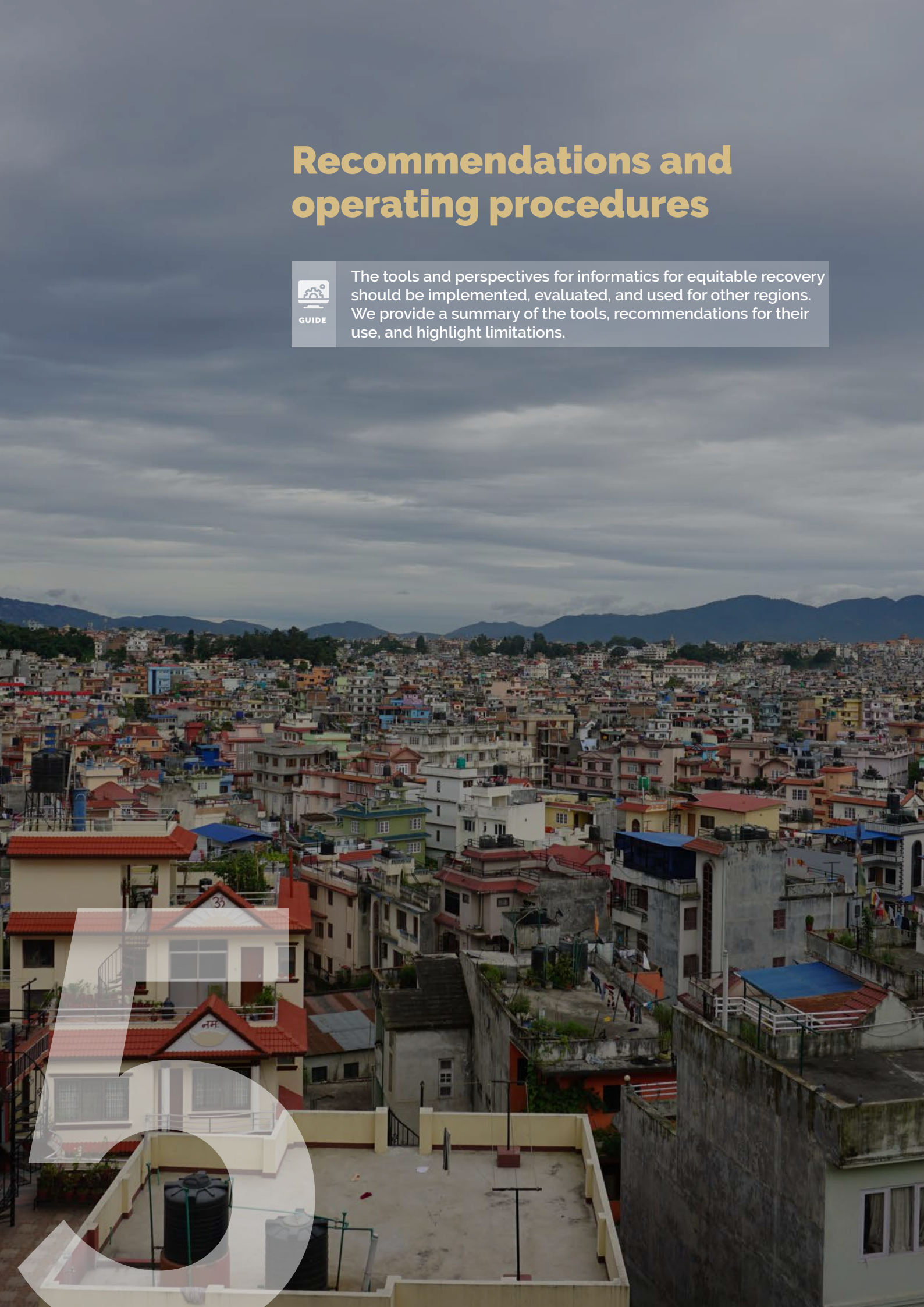
[40]Note - this map is subject to change based on future updates to the model, to see the latest prediction of Post-Disaster Persistent Need, visit https://disaster-analytics.com/projects/ier-nepal.

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

Recommendations and operating procedures

GUIDE

The tools and perspectives for informatics for equitable recovery should be implemented, evaluated, and used for other regions. We provide a summary of the tools, recommendations for their use, and highlight limitations.



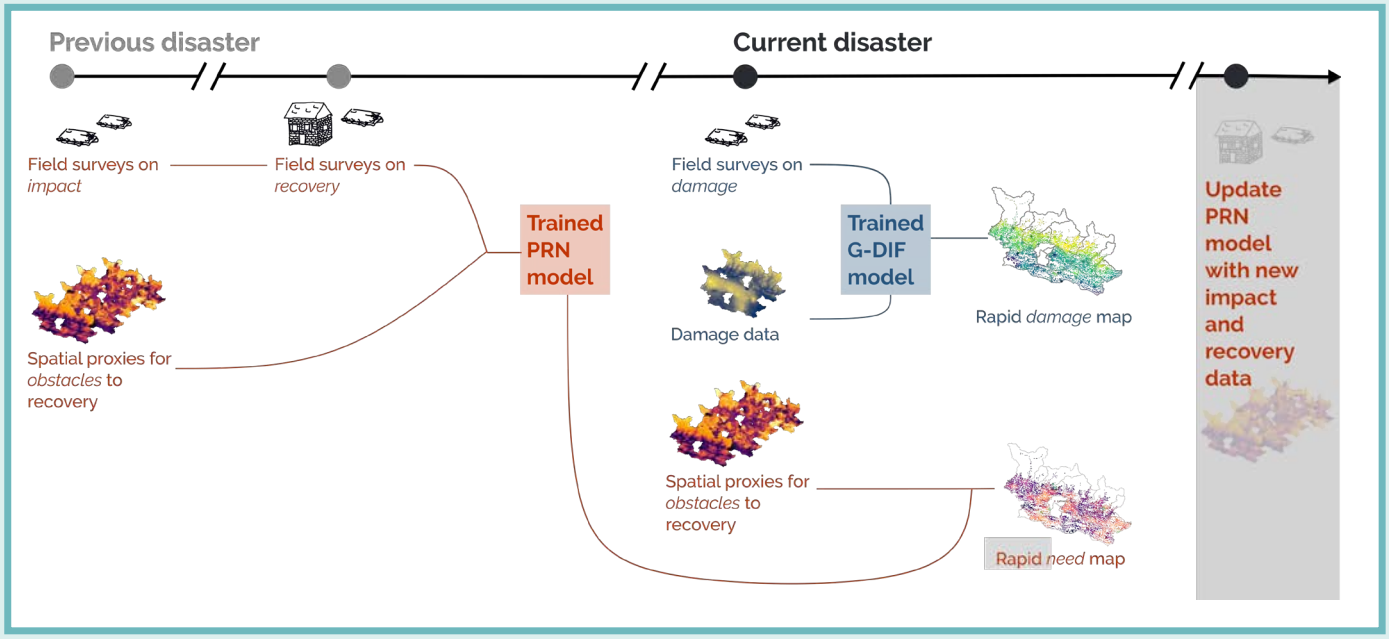
5. Recommendations and operating procedures

Both the Geospatial Data Integration Framework (G-DIF) and the Post-Disaster Persistent Need (PDPN) framework are meant to be implemented immediately after a disaster to estimate high-resolution maps of impact and likely areas with significant recovery need due to pre-existing vulnerability characteristics. We have provided examples of how these frameworks were developed and tested using data from the 2015 Nepal earthquake. However, we have formulated these frameworks so that they can be both implemented, examined, improved and used in future disasters in other regions of the world[42]. This section will outline

1. the procedure for implementing these frameworks in future disasters,
2. the uses and limitations of these frameworks and
3. the implications of using these frameworks.

5.1: Summary of the tools and how they work together

The G-DIF and PDPN frameworks can be implemented together in the weeks after a disaster. The main requirements are a limited sample of field-based data (or other validation data that is highly accurate) to use as training/calibration data to calibrate each model. The key difference between G-DIF and PDPN is that the model to predict PDPN is trained before the next disaster, while the model to predict damage is trained immediately after the disaster. This is because damage data are measurements of immediate impact and so can be collected in the weeks after a disaster, while the recovery data required for the PDPN framework can only be collected years after a disaster. If there is no data from a previous disaster in a comparable context, only G-DIF can be used. Nonetheless, monitoring recovery would enable PDPN to be used for risk reduction measures or immediately after a future disaster in a comparable context.



The key features of the G-DIF are that it:

1. integrates multiple datasets that are expected to be available in most post-disaster scenarios;
2. is locally calibrated to the specific context using relatively few field-based data;
3. leverages the spatial nature and correlation that exists in damage data for improved damage estimation.

Conceptually, this framework is generalizable to other contexts and other hazards as long as we have some sample of field-based data (or other calibration data that is highly accurate) so that it can be recalibrated to any new local context. Maps of estimated damage can then be produced by combining these field data with other damage proxies (such as shakemap or remote-sensing based assessments) using G-DIF. By nature of the model, any proxies that are uninformative will have lower weights in the final damage estimate based on how well they predict the observed damage from the field-based data. Importantly, the maps of damage can be improved (and uncertainty reduced) over time as more field-based data is added, or additional damage proxies are integrated in the model.

While field data collection in post-disaster settings is often complex and time-consuming, such data collection is critical to building understanding of the disaster event, and is in fact already part of numerous standard humanitarian and early recovery activities. The field collection of damage data for G-DIF is also an opportunity to identify households to collect recovery data to build the PDPN metric for a future disaster.

Beyond building damage, this study demonstrates the importance of non-damage factors in household's ability to recover from disaster. The PDPN framework is designed to account for such non-damage factors and provide early estimates of areas at high risk of significant and persistent needs for recovery. The key features of the PDPN framework are that it

1. predicts a forward-looking metric of persistent recovery need over time rather than immediate damage;
2. uses readily available and spatially explicit proxies for the obstacles to recovery;
3. is applicable to other recovery outcomes of interest as long as it can be measured in the field,
4. and can be used for pre-disaster estimation and mapping of communities likely to possess significant vulnerability characteristics.

The G-DIF and PDPN models provide flexible frameworks for estimating and mapping immediate post-disaster damage and areas likely to have significant recovery challenges. While they were developed and demonstrated for the particular context of Nepal following the 2015 earthquake, the models are easily adaptable to other contexts, hazards, data-sources and metrics of relevant impact. They are not limited to specific datasets--or even metrics of impact or recovery--since these frameworks are meant to adapt with time and from place-to-place. Current and future applications of the model, model updates, source code and additional resources are described in the annex.

5.2: Guidelines for using tools

There are numerous use-cases for these tools to develop high-resolution maps of impact, both for Nepal and other contexts.

5.2.1: Rapid Post-Disaster Damage Estimation using the Geospatial Data Integration Framework

The Post Disaster Needs Assessment (PDNA) requires an estimate of the "disaster effects", or the damages and losses to each sector, one of which is housing. Two of the main key steps of carrying out the PDNA is data collection--field visits and gathering of secondary quantitative data-- and consolidation/analysis--compilation of the collected data[43]. The G-DIF can easily be implemented within the already existing PDNA process and the expected timeline of a few weeks following a disaster.

To implement G-DIF, those carrying out the sectoral damage assessment could collect any and all available field-based data that exists from field reconnaissance teams, civil defence groups, and PDNA field evaluation teams.

A desk review can be carried out to collect all available secondary damage datasets. This would include damage forecasts based on hazard source models (like the USGS's Earthquake ShakeMap/PAGER estimates or hurricane track and wind-speed based models) or remote-sensing based damage proxies (like NASA's Damage Proxy Map or other estimates made available). **The accuracy of each secondary dataset should not be of much concern, since the model in G-DIF will self-evaluate and weight individual secondary datasets based on their performance in predicting observed damage from the field-based data.**

As an example, the map in Figure 2a could have been implemented for the 2015 Nepal earthquake. This estimate can be translated to economic losses using information on housing types and their cost to reconstruct.

Beyond the Post Disaster Needs Assessment, the resulting damage estimate can be updated over time as more field-based data is collected. These estimates can inform more recovery-oriented decisions such as the Post-Disaster Recovery Framework, a document that outlines the structure of recovery policies that should be put in place over the years following the disaster[44].

5.2.2: Estimating Post-Disaster Persistent Need with the PDPN Framework.

This study has demonstrated that **physical damage alone is an insufficient metric to capture the variation in post-disaster recovery needs.** It is important to focus on persistent need, because those **households that are unable to recover are likely to be the most vulnerable.**

Beyond damage, we have identified several other factors that are important to predict a household's persistent need in the years after a disaster. While the importance of each individual predictor is likely to be context-specific, our study in Nepal confirms the importance of using socio-economic indicators as additional information in early decision making surrounding recovery and assistance. Some likely important indicators identified empirically through this study include a household's distance to the nearest market, the percentage of young children and elderly, and a proxy for socioeconomic status (as reflected by the percentage of homes without toilets in their home).

Factors likely to be important determinants of difficulties in recovery are often surmised during the PDNA process (and many other processes in the post-disaster phase) through pre-existing studies or local expert-based knowledge . The framework to develop the PDPN metric identifies those factors that correlate with housing non-recovery and provides a spatially explicit prediction of where recovery is likely to be particularly challenging. If a model to develop the PDPN metric is already trained before implementing a PDNA, this could provide an additional quantitative estimate of recovery needs. If a model to develop the PDPN metric does not already exist, individual socio-economic indicators that are identified during the PDNA process as obstacles to recovery could be compiled and mapped, so they are used more directly in the recovery process.

Regular multidisciplinary surveys of recovery can be used to build PDPN models to estimate which locations are likely to have persistent needs across the whole of the affected area. Outside of building the PDPN models, these surveys also serve the purpose of monitoring recovery progress. Collecting recovery data with multidisciplinary teams ensures that multiple aspects of recovery (e.g. housing, livelihood, nutrition, etc) are monitored and, later, modeled. Representing recovery holistically would facilitate more effective and equitable plans that support the multifaceted needs of affected communities.

5.3: Model limitations

Both the Geospatial Data Integration Framework and the Post-Disaster Persistent Needs Framework contain prediction models, as such, they provide estimates of damage and of the likelihood to recover. As with any model, estimates are a useful indicator but are not completely correct predictions, and can in fact be incorrect. Key questions for users of a model include: **what is the uncertainty in model prediction? What level of uncertainty is acceptable? What decisions are appropriate and not appropriate based on a model with particular uncertainties?**

5.3.1: Limitations of G-DIF

We attempt to validate these models and be transparent about any models that we create. One of the strengths of the G-DIF is that it locally calibrates other damage data using field data that are observed measurements of damage for a specific context and specific disaster. While this is an advantage, it does present a danger for model bias if the field data used

to calibrate the prediction model is biased or otherwise not representative of the overall damage. To mitigate this risk, field surveyors should ensure that the field sample covers

- a wide range of damage levels (including no damage at all) and
- the full spectrum of typologies of buildings and affected environments (example: urban, rural, peri-urban).

The models contained in G-DIF predict the best estimate of damage and at the highest resolution possible based on available information. This means that within an individual estimate (here, a pixel of about 300m x 300m), the prediction is of the average damage of all buildings within that area. Thus, some buildings within an area might have collapsed even if the overall average damage is quite low, or vice versa. G-DIF aims to provide estimates averaged over small areas, but cannot predict the damage to any particular building. Despite everything, error and uncertainty will still exist. Some areas may be predicted as undamaged while they are in fact significantly damaged due to processes that are not captured in the model or data inputs.

Minding this, the G-DIF estimate is not a replacement for field-surveys of damage, which are absolutely necessary. The model is geared at providing best possible estimates in the short term, filling an important information need before extensive field surveys are conducted. Users should use this model as a first best estimate, knowing that it can be incorrect at the local level. Over time, the model can be improved by integrating more field and secondary data, and should eventually be supplanted by ground-based assessments.

5.3.2: Limitations of the model to predict PDPN

The map of Post-Disaster Persistent Need should be seen as a starting point as well. It highlights areas that are likely to not recover given that they have been impacted. Reasons why certain areas are highlighted should be further investigated through site visits throughout the recovery process.

This specific model of PDPN was built for one context, one type of disaster, and one definition of recovery (reconstruction progress). The main assumption here is that the recovery trajectories from a past disaster can be predictive of future disaster recoveries. To the degree that this may not hold – if something significant has changed, such as the injection of external aid after one disaster but not another – the model will make less reliable predictions. We note that similar limitations apply to any other predictive recovery or resilience model, especially those models that apply aggregation methods to quantify resilience (discussed more in Section 6) .

The PDPN model also depends on proxy data for the different obstacles to recovery. Broadly, the modeler is responsible to select proxy variables informed by theory, on-the-ground evidence, and multiple perspectives. Furthermore, a model is only as good as the data it is built upon, so it is the modeler's interest to ensure all data inputs are of good quality. This would improve if census data were collected more regularly and in more detail and as more high resolution estimates of obstacles to recovery (for example estimates of poverty[45]) are developed further. That being said, the map of predicted Post-Disaster Persistent Need would benefit from additional validation by developing a similar model with an alternative recovery dataset from the same disaster.

Finally, by no means should the map shown in this report be interpreted as a map of current recovery rates. That data is available on various public sources[46,47]. Our model shown here represents the spatial distribution of recovery had the entire affected area been heavily and equally impacted (it is independent of damage).

Overall, both the damage and PDPN estimates shown here are focused in rural Nepal, in 11 districts where a detailed post-disaster survey enabled us to develop these models effectively. These 11 districts were heavily impacted by the disaster, but so were some areas in neighboring districts. Importantly, recovery needs from the 2015 earthquake extend beyond the 11 districts that were the focus of this study.

We have highlighted how damage alone is insufficient to map persistent needs, because it does not capture holistic impacts or future recovery likelihoods. Similarly, **the map of persistent recovery need should be used in conjunction with the map of damage so as to highlight the areas of high impact and lowest likely recovery rates.** When possible, both maps should be developed after a disaster; the combination of the two provide information on both short and long-term impacts imposed by the disaster.

References

[42]Source code for these frameworks developed in Nepal has been made publicly available. To adapt these frameworks to future disasters, users will need basic experience with coding, data cleaning, and geographic information systems (GIS). Additional resources on code are detailed in the annex.

[43]The European Commission, The United Nations Development Group, and The World Bank, "Post-Disaster Needs Assessments Guidelines," 2013.

[44]Global Facility for Disaster Reduction and Recovery (GFDRR), World Bank Group, The European Commission, and United Nations Development Programme, "Guide to Developing Disaster Recovery Frameworks," 2015.

[45] Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." Science 353 (6301): 790–94. https://doi.org/10.1126/science.aaf7894.

[46]https://www.hrrpnepal.org/

[47]http://nra.gov.np/np

Policy implications and conclusions

6

6. Policy implications and conclusions

Disasters continue to hinder sustained development progress. Key to mitigating the long-term impact of disasters on vulnerable communities is the ability to rapidly respond and recover from disaster, in ways that build further resilience and protect hard fought development gains. Through our project, we have developed new information systems aligned with this goal.

Specifically, our study has produced two new models for (1) more accurate estimation and mapping of post-disaster damage (G-DIF, Section 2), and (2) prediction and mapping of post-disaster persistent need that accounts for pre-existing community vulnerability characteristics (PDPN, Section 4). Through extensive field interviews of households impacted by disaster, our work has further described the complexity of disaster impacts and recovery trajectories experienced by disaster-affected households (Section 3).

We believe that disaster data providers, recovery planners and policymakers, nongovernmental organisations (NGOs), the disaster risk reduction (DRR) community, and disaster-affected groups can all benefit from applying the tools or perspectives developed through our project. Fundamentally, we argue for expanding the understanding of post-disaster evidence, treating recovery needs as evolving and dynamic, leveraging recovery data for risk reduction efforts, and designing policies that are sensitive to differential needs.

6.1: Expand the understanding of post-disaster evidence, inputs, and processes

Current post-disaster information systems tend to emphasise physical damage and losses, produce estimates of damage without estimates of uncertainty, and tend not to be calibrated against locally collected damage information (such as field surveys). One crucial implication of tools like G-DIF and PDPN is therefore that **response and recovery stakeholders now have rapidly deployable methods to combine multiple types of data inputs and estimate context-specific impacts and recovery outcomes.**

Rather than having to choose between competing and uncertain estimates, recovery planners and agencies can use G-DIF to synthesize an expanding slate of data inputs to map building damage, and use the PDPN framework to understand how recovery needs vary across space. Concretely, this may mean something as simple as incorporating tools like G-DIF and PDPN in preparedness training, or as complex

as codifying processes for rapid data collection and incorporate data integration principles in organisational best practices. At the very least, a platform to both identify post-disaster data requirements and to share disaster datasets would organize and make usable the multiple datasets that are produced after an event.

While stakeholders can use these tools to better organise post-disaster information that they receive, **disaster data providers** should also ensure that the estimates they provide are transparent about inputs and uncertainties, and amenable to being combined with and validated against other data sources. For G-DIF, this means that all damage datasets should clearly document any underlying datasets and methods, use damage metrics that are consistent with those used on the ground, and cover the region of interest that stakeholders need. More generally, this implication highlights how important it is for **data providers — such as international science agencies and remote sensing organisations —** to understand how response and recovery agencies will actually use the data they provide. We see considerable potential here for data providers to engage more actively with recovery stakeholders through well-understood processes like user-centered design and participatory research. Programs that seek to bridge across these groups would be crucial for better tailoring data products to user needs.

Furthermore, the use of geospatial proxies for socioeconomic characteristics — a key aspect of the PDPN approach — also broadens the scope of inputs in disaster information systems. The PDPN model relies on context-specific proxies for obstacles to recovery, such as remoteness indices, environmental datasets, and census variables. These datasets should be further developed and made openly available to be used for disaster modeling purposes (for instance, Nepal has a geonode which hosts open spatial datasets, including models developed specifically for Nepal[48]). These implications are particularly salient for **disaster and risk modelers** who currently rely on very specific types of exposure and vulnerability data, so that they can incorporate other types of insightful data into their models.

6.2: Treat recovery needs as evolving and dynamic

A core issue highlighted by PDPN and by our household recovery survey is that damage assessments and numerous recovery plans are currently based on information collected at a single, specific window in time: the immediate response and early recovery phase. While this is understandable given the structure of disaster aid — where needs assessments usually serve as inputs into a donor conference to secure aid pledges — there is good reason to believe that recovery needs can evolve over time. Some households and communities may be exposed to subsequent disasters, struggle to receive technical assistance, have less access to economic opportunities, or be affected by other obstacles to recovery. **In essence: a one-shot disaster impact and needs assessment process is not enough to convincingly describe the recovery needs of a region through the years after a disaster.**

Recognising these realities, we see potential for **disaster reduction organisations and international aid groups** to develop methods for monitoring and updating disaster impacts and recovery data over time. There is already strong interest and action in this direction, especially in Nepal: multiple NGOs and aid groups have published[49, 50, 51] analyses year-on-year documenting the recovery process and the effects of interventions. **A platform to share, consolidate and consistently format such data using a clear methodology could greatly enrich our understanding of recovery obstacles in specific contexts.**

6.3: Leverage recovery data collection for risk reduction efforts

Researchers and practitioners have long emphasised the value of linking recovery efforts with disaster risk reduction and resilience-building. Our tools suggest several implications for such efforts to build back better, especially for **local and national recovery planners.** For one, collecting recovery data allows us to measure, and therefore support, improvements in household safety and building stock. Detailed recovery data from household surveys allow us to identify where, and to what degree, households are able to build back better or otherwise recover in ways that boost their resilience. Agencies could cross-reference this data with their own knowledge of targeted interventions and technical assistance to identify areas where specific resilience-building initiatives were effective.

An important corollary of this is that PDPN and household recovery surveys allow us to identify and name **obstacles to recovery,** which could have implications for risk reduction efforts. Since our project's scope and tools focus on more than just building damage, narratives around 'building back better' quickly become more complicated **— recovery planners and community leaders** would benefit from knowing when resilience gains in one dimension (e.g. safe construction) have come at the expense of another dimension (e.g. increased debt). Similarly, the PDPN tool could highlight and therefore allow for **international risk reduction practitioners** to target less conventional obstacles to recovery that emerge from the PDPN metric, especially prior to the next disaster.

Additionally, we see our tools as opportunities to build links and capacity between local recovery efforts and international disaster risk reduction practice. In designing tools like G-DIF and PDPN which are **input-agnostic but context-specific** (i.e. flexible to diverse data inputs while locally calibrated), our team carefully made decisions balancing applicability (i.e. ensuring that our tools were truly usable) with generalisability (i.e. developing insights from Nepal's recovery that can inform disaster reduction globally). One implication of our approach is that resulting maps of impact and need are tailored to Nepal's context in the Gorkha 2015 earthquake recovery process, but **the framework, models and process of mapping and calibrating against ground truth data is broadly applicable in other contexts and for other hazard and disaster types.** We envision that salient lessons will be developed as G-DIF and PDPN — or indeed, other tools that combine multiple types of data with ground truth — are applied in different contexts.

The PDPN tool will become more robust the more that it is used. PDPN assumes that past experiences with recovery from a disaster can inform our understanding of recovery **obstacles for a subsequent disaster in a similar context.** While this assumption has support from previous research, it would best be tested by application in many contexts. To this end, it would be helpful for **risk reduction practitioners** to gather recovery data and experiences in other contexts and regions to develop tools like the PDPN, even before the next disaster. This will help **recovery planning stakeholders** gauge how well PDPN may perform in the context where they work. This not only provides a useful input for recovery planning following a particular disaster, but also enables a deeper understanding of recovery obstacles that are common across multiple disaster events.

6.4: Design recovery policies sensitive to differential needs

Our surveys and tools illustrate how household experiences of recovery can diverge, how stakeholders can predict or estimate differential needs, and how physical damage alone is an insufficient measure of disaster impact. Since aid decisions are highly motivated by contextual factors and resource constraints, we do not prescribe any specific model of aid delivery to address inequities in pre-disaster vulnerability and recovery outcomes. Rather, an implication of tools like G-DIF and PDPN is that vulnerable groups — groups that will persistently struggle with recovery — can also be defined and understood using empirical categories like recovery outcomes.This evidence can validate, or in some cases nuance, other pre-existing portrayals of vulnerability. These tools can complement and extend our understanding of vulnerability – to help **make invisible inequities visible** – toward more equitable recovery.

6.5: Towards reflective, human-centric post-disaster evidence

The tools and perspectives developed by our project — informatics for equitable recovery — ultimately contribute to a more human-centric understanding of post-disaster impact. Current informatics focus on buildings, lost assets, and other metrics that conceal the nuances of how disasters leave households with persistent unmet needs. Even within categories such as housing, we have highlighted how closely issues like reconstruction can be intertwined with livelihood and social aspects of recovery. By distinguishing between estimates of **damage** (which are addressed by G-DIF) and **need** (which PDPN focuses on), we illustrate how disaster informatics can incorporate new and relevant types of information that enable more equitable recovery.

Fundamentally, our aim is to enable stakeholders to use evidence in post-disaster contexts in systematic ways that incorporate ground realities and are sensitive to issues of equity. We make a case for collecting data on recovery systematically and openly, even after a disaster event, so that stakeholders and users of post-disaster data can understand persistent needs and obstacles to recovery more holistically. We also invite deeper engagement with questions of equity in disaster informatics more broadly, especially given the role of new technologies and data sources in commonly implemented processes.

References

[48]http://geonepal.info/

[49]https://www.hrrpnepal.org/

[50]The Asia Foundation. 2015. "Independent Impacts and Recovery Monitoring."https://asiafoundation.org/tag/independent-impacts-and-recovery-monitoring-nepal

[51]Interagency Common Feedback Project. 2018. "Community Perception Report: Reconstruction, Food Security, & Livelihood and Protection." https://www.alnap.org/help-library/inter-agency-common-feedback-project-community-perception-report-reconstruction-food.

Annexes

RE-SOURCES

Additional information and documentation about the tools, methods and processes developed in this project.

7

7. Annexes

7.1: Project code and further reading

All additional material to both understand and implement these tools are hosted on our project website at disaster-analytics.com/projects/ier-nepal/. Here you can find:

1. Source code to develop the estimates shown in this report
2. Full academic publications behind each section
3. Blogs
4. Visualizations
5. Media coverage
6. Additional resources

7.2: Datasets used in this study

The datasets used to develop both G-DIF and PDPN models are listed below. Updated links to each dataset can be found on the project website.

Tool	Dataset	Description	Source
G-DIF	GT_mnDG	Average damage grade per grid of 290m x 290m from field surveys	NHRP field surveys
G-DIF	DPM_bldg_med	Damage Proxy map	NASA JPL-ARIA
G-DIF	MMI_mn	Shaking intensity	USGS
G-DIF, PDPN	dem_cgiar	Digital elevation model	CGIAR
PDPN	Av_HH_Size	Average Household Size per VDC	2011 Nepal Census
PDPN	edu_sec9.10_plus	Percentage of pop. with education greater than grade 9&19	2011 Nepal Census
PDPN	evaptot_ici	Total evaporation	ICIMOD
PDPN	EVI_mod	Enhanced vegetation index	MODIS
PDPN	EVI_buff	Enhanced vegetation index with a buffer	Self-developed
PDPN	Healthfacdel_dhs	Percentage of births delivered at a health facility	DHS
PDPN	Lang_other	Percent of population that speaks a language other than Nepali	2011 Nepal Census
PDPN	Menlit_dhs	Male literacy rates	DHS
PDPN	MMI_usgs	Modified Mercalli intensity shaking intensity	USGS
PDPN	NDVI_mod	Normalized difference vegetation index	MODIS

PDPN	NDVI_buff	Normalized difference vegetation index with a buffer	Self-developed
PDPN	No_toilet	Percentage of population without a toilet	2011 Nepal census
PDPN	Ntlt_rad_bin	Binary nightlight radiance	
PDPN	Pop_00.04	Percentage of population under 5	2011 Nepal census
PDPN	Pop_05.14	Percentage of population between 5-15	2011 Nepal census
PDPN	Pop_15.25	Percentage of population between 15-25	2011 Nepal census
PDPN	Pop_25.60	Percentage of population between 25-60	2011 Nepal census
PDPN	Pop_60plus	Percentage of population over 60	2011 Nepal census
PDPN	Popn_2015_wp	Population density 2015	Worldpop
PDPN	Prec_ici	Precipitation in April 2015	ICIMOD
PDPN	Pregnancies_wp	Number of pregnancies	Worldpop
PDPN	RC_wall	Percentage of households with reinforced concrete wall	2011 Nepal census
PDPN	Remote_avg	Average remoteness	World Bank
PDPN	Remote_n1hr	Number of facilities within 1 hour	World Bank
PDPN	remoteDHQ	Remoteness from district headquarter	World Bank
PDPN	remoteHealthF	Remoteness from a health facility	World Bank
PDPN	remoteHRRPmrkt	Remoteness from HRRP defined construction market	World Bank
PDPN	Sex_Ratio	Sex Ratio	2011 Nepal census
PDPN	Slopedeg	Slope in degrees	Self-developed
PDPN	Soilm_ici	Soil moisture in April 2015	ICIMOD
PDPN	Stunt_dhs	Stunting	DHS
PDPN	Tap_Water	Percentage of households whose main drinking supply is from the tap	2011 Nepal census
PDPN	Tempavg_ici	Average April temperature	ICIMOD
PDPN	Tot_lit_rate	Total literacy rate	2011 Nepal census
PDPN	Urm_wall	Percentage of households with unreinforced masonry wall	2011 Nepal census
PDPN	Womlit_dhs	Women's literacy rate	DHS
PDPN	Wood_Wall	Percentage of households with timber wall	2011 Nepal census