Loos, S., Levitt, J., Tomozawa, K., Baker, J., Lallemant, D. (2022). "Efficacy of damage data integration: A comparative analysis of four major earthquakes." *ASCE Natural Hazards Review*, 23(4), 04022026.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000581 This material may be downloaded for personal use only. Any other use requires prior permission of the American Society of Civil Engineers.

Efficacy of damage data integration: a comparative analysis of four major earthquakes

Sabine Loos, Jennifer Levitt, Kei Tomozawa, Jack Baker, David Lallemant

June 22, 2022

Abstract

Weeks after a disaster, crucial response and recovery decisions require information on the locations and scale of building damage. Geostatistical data integration methods estimate post-disaster damage by calibrating engineering forecasts or remote sensing-derived proxies with limited field measurements. These methods are meant to adapt to building damage and post-earthquake data sources that vary depending on location, but their performance across multiple locations has not yet been empirically evaluated. In this study, we evaluate the generalizability of data integration to various post-earthquake scenarios using damage data produced after four earthquakes: Haiti 2010, New Zealand February 2011, Nepal 2015, and Italy 2016. Exhaustive surveys of true damage data were eventually collected for these events, which allowed us to evaluate the performance of data integration estimates of damage through multiple simulations representing a range of conditions of data availability after each earthquake. In all case study locations, we find that integrating forecasts or proxies of damage with field measurements results in a more accurate damage estimate than the current best practice of evaluating these input data separately. In cases when multiple damage data are not available, a map of shaking intensity can serve as the only covariate, though the addition of remote sensing-derived data can improve performance. Even when field measurements are clustered in a small area–a more realistic scenario for reconnaissance teams–damage data integration outperforms alternative damage datasets. Overall, by evaluating damage data integration across contexts and under multiple conditions, we demonstrate how integration is a reliable approach that leverages all existing damage data sources to better reflect the damage observed on the ground. We close by recommending modeling and field surveying strategies to implement damage data integration in-real-time after future earthquakes.

1 Introduction

From rapid forecasts to remote sensing-derived maps, novel sources of post-disaster building damage data are needed to make crucial decisions for early recovery. For example, two to four weeks after a disaster, the government of the affected region will often lead a Post-Disaster Needs Assessment (PDNA) to assess metrics such as the number of damaged buildings and cost to reconstruct. The PDNA memorializes the losses from an event and influences the aid a country receives for its recovery. Damage information also underlies shorter-term response activities such as temporary shelter allocation and longer-term recovery policies such as distribution of reconstruction aid (Bhattacharjee et al., 2021). In many cases, potentially useful data, especially derived from satellites, is rapidly available. However, these data were often only used to guide the collection of more precise damage data later on or to inform building safety, as they could not identify lower damage grades necessary to support the PDNA which guides major reconstruction decisions (The European Commission, 2017; Sextos et al., 2018; Eguchi et al., 2010; Government of the Republic of Haiti, 2010).

Post-earthquake damage maps come from a wide range of sources, including remote sensing-derived or forecast-based estimates (Loos et al., 2020). We call these sources secondary datasets, which are

advantageous since they provide a rapid estimate of damage over a large region in less time than it would take to collect primary field surveys of damage. They are highly uncertain, however, usually because they are produced using methods developed for global use. Remote sensing-derived data is based on imagery from any type of remote sensor, including satellites, planes, drones, among many others. Publicly available remote sensing-derived data include NASA JPL-ARIA's Damage Proxy Map (DPM) derived from Interferometric Synthetic Aperture Radar (inSAR) data and the Department of Defense's xView2 challenge, which called for participants to use computer vision with high-resolution imagery to estimate multi-hazard building damage (Yun et al., 2015; Gupta et al., 2019). Additionally, maps from manual interpretation of remote sensing imagery exist, such as the crowdsourcing efforts carried out after the Haiti 2010 earthquake (Ghosh et al., 2011) or damage grading maps from the Copernicus Emergency Management Service (Dorati et al., 2018). Outside of remote sensing-derived maps, engineering forecasts are also produced as soon as a map of shaking intensity becomes available (Erdik et al., 2014; Earle et al., 2009; Trendafiloski et al., 2009; Gunasekera et al., 2018). Engineering forecasts are predictive models of damage that relate the estimated distribution of shaking to consequence metrics, like building collapse, through models of exposure and vulnerability. Alternative machine learning methods that similarly use hazard and building characteristic data to rapidly forecast damage have also been developed (Mangalathu et al., 2020). An example of publicly available engineering forecast is the United States Geological Survey's PAGER system, which aggregates forecast results to country-level estimates of economic loss or casualties (Jaiswal and Wald, 2011).

While abundant data might seem beneficial, three issues exist. First, rapid damage maps are produced at varying resolutions with units that do not necessarily align with the needs of post-disaster planners. In some cases, like with the DPM, the information provided is a proxy of damage, where each pixel contains a unitless integer that indicates change between pre- and post-earthquake imagery, but has inconsistent meanings between earthquakes. Second, many models are developed with data from prior events in other places and therefore still need to be calibrated to the current disaster. Third, because of the fast-moving and haphazard nature of post-disaster decision-making, many response workers or recovery planners use only the data they trust, rather than considering all the available data at once (Liboiron, 2015; Bhattacharjee et al., 2021; Hunt and Specht, 2019).

The Geospatial Data Integration Framework (G-DIF), based on the geostatistical method Regression Kriging, addresses these issues (Loos et al., 2020). G-DIF is a general modeling framework that is agnostic to different types of primary and secondary data, and therefore adapts to different places and new developments in secondary data. The method decomposes the spatial distribution of damage into the trend, or the average gradient in damage over the affected region, and spatially correlated and stochastic residuals around that trend. The estimation of the trend depends on the secondary damage data, while the estimation of the residuals depends on the expected spatial correlation in the residuals from the trend at the field survey locations. Since our initial application of G-DIF to the Nepal 2015 earthquake, others have built upon this idea with alternative models (Sheibani and Ou, 2021; Wilson, 2020).

Three main assumptions were made about the expected performance of G-DIF and alternative damage data integration methods, which we evaluate and address in this paper. The first is that G-DIF will perform better than any alternative secondary dataset alone. Without better performance, the effort of building a G-DIF model would not be justified. The second is that the secondary data available in the earthquake-affected country is of good enough quality to correlate with the damage seen on the ground. This assumption might not be the case after earthquakes in regions with little remote sensing data and few seismic stations to measure shaking intensity. The third assumption is that the field surveys used to calibrate the secondary damage data to the local observations of building damage is collected from a spatially representative sample. Field surveys may not be representative if engineering reconnaissance missions or local survey teams focus on the communities that are easiest to reach immediately following a disaster or the areas where they expect to find damage (resulting in a preferential sample).

In this study, we evaluate these assumptions by applying G-DIF to damage data that became available

after four major earthquakes: Haiti 2010, New Zealand February 2011, Nepal 2015, and Italy 2016. We evaluate whether G-DIF's damage estimate outperforms alternative secondary estimates of damage across various contexts with different patterns of damage and quality of secondary data. Additionally, we examine whether G-DIF is able to produce an accurate damage estimate with different sources of secondary data available or more realistic field surveyed locations. Applying G-DIF to multiple real-world scenarios does require assumptions to perform comparisons. To facilitate comparisons across events, we made several simplifications to develop the inputs and models in G-DIF. While this might somewhat reduce predictive performance, we still find clear and intuitive trends that allow us to understand the general performance of the method.

We find that many of our assumptions hold, indicating that G-DIF generalizes well across contexts under different scenarios of primary and secondary data availability. Overall, this study demonstrates how G-DIF is a reliable approach that can leverage all available damage data after an earthquake to better reflect the damage observed on the ground. Thus, G-DIF is an improvement over the current practice of qualitatively evaluating each input damage data source, whether it be field surveys or remote sensing-derived, on their own. We, therefore, close with both modeling and field surveying strategies to implement damage data integration in-real-time after future earthquakes.

2 Case studies

We consider four major earthquakes from the past decade: 1) Haiti 2010, 2) New Zealand February 2011, 3) Nepal 2015, and 4) Italy 2016. Table 1 summarizes key case study characteristics and Figure 1 shows maps of the true damage obtained from field surveys. These case studies are vastly different in terms of the pattern of damage, spatial scale, available data, and data quality, allowing us to evaluate performance of G-DIF in varied circumstances.

The January 12, 2010 Haiti earthquake is our earliest case study. The M_w7.0 event occurred about 25 km southwest of Haiti's capital of Port-au-Prince and was followed by three major aftershocks in the week afterwards (DesRoches et al., 2011). Haiti had a weakly enforced building code in the dense city of Port-au-Prince composed mostly of unreinforced concrete frame buildings (DesRoches et al., 2011), resulting in an estimated 200,000-300,000 deaths (O'Connor, 2012). The Haiti earthquake was one of the first earthquakes with a proliferation of damage data, pioneering many new techniques to evaluate damage from remote sensing imagery (Corbane et al., 2011; Loos et al., 2020). However, because many of these nontraditional damage datasets were originally tested after this event, Haiti's datasets have relatively poorer quality than the subsequent case studies. In addition, Haiti lacked a seismic network at the time of the earthquake (DesRoches et al., 2011) and thus had a poorly constrained estimate of shaking.

Our next case study is the February 22, 2011 Christchurch, New Zealand Earthquake, the most damaging of the Canterbury Earthquake Sequence (Potter et al., 2015; Comerio, 2014). The M_w6.3 earthquake was an aftershock of the M_w7.1 Darfield earthquake of September 2010 and occurred only 10 km away from downtown Christchurch. It caused damage throughout the Central Business District and residential areas of Christchurch, ultimately leading to 185 deaths (Potter et al., 2015). Unlike the Haiti earthquake, the New Zealand earthquake occurred in a country with relatively good quality secondary data and a strongly enforced building code. The residential houses are predominantly engineered, light timber framed single story homes (Buchanan et al., 2011). Much of the damage was liquefaction-induced, leading to high rates of foundation damage (Van Ballegooy et al., 2014). Liquefaction damage traditionally may not be captured in engineering forecasts but has the potential to be observed through remote sensing.

Our third case study is the April 25, 2015 Nepal earthquake. The M_w7.6 earthquake occurred in the Gorkha district, about 80 km northwest of the capital of Kathmandu. This event and its aftershocks caused nearly 9000 deaths and impacted both urban Kathmandu and the surrounding rural districts (Government

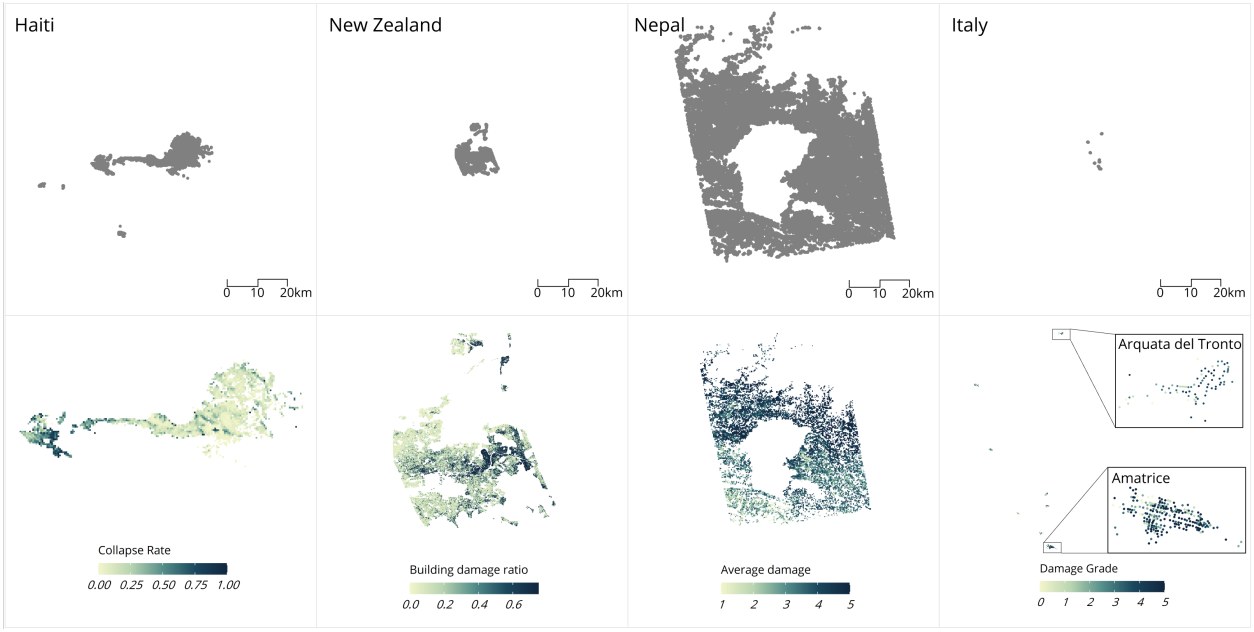

Fig. 1. Field surveys of damage for all four case studies. Row 1 shows the data locations for each case study, plotted on the same scale to indicate their relative spatial extents. Row 2 shows the maps of damage severity as obtained from field surveys. The units of measurement differ for each case study, but for all locations, darker colors indicate higher damage.

of Nepal National Planning Commission, 2015). Like Haiti, this event resulted in a proliferation of non-traditional data (Loos et al., 2020; Dennison and Rana, 2017). Nepal also has a long history of shifting governmental institutions (Thapa, 2005; Sharma, 2006) and an inadequately enforced building code. Many rural houses were low-strength stone masonry structures (Government of Nepal National Planning Commission, 2015). This led to the highest rates of collapse outside of Kathmandu in rural villages, especially near the Himalayas. Because of the rural nature of this earthquake, the spatial extent of damage is much larger in Nepal in other case studies (Figure 1).

Our final case study is the August 24, 2016 Central Italy earthquake. The M_w6.2 event occurred near the village of Accumoli (Stewart et al., 2018; D'Ayala et al., 2019). It caused severe damage to nearby villages including Amatrice and Arquata del Tronto, and ultimately nearly 300 deaths. While Italy is a higher-income country with a steadily improving building code (Liel and Lynch, 2012), many of these towns had historic unreinforced masonry structures that were prone to collapse (Sextos et al., 2018). Variations in building stock led to differing rates of collapse among towns. Figure 1 shows that in Italy the damage is localized in specific towns as opposed to the more continuous pattern of damage in Haiti, New Zealand, and Nepal.

2.1 Data description

Building damage data can generally be categorized into primary field measurements and secondary estimates from remote sensing, engineering forecasts, or related geospatial covariates (Loos et al., 2020). Field measurements of damage are usually obtained through field surveys, where surveyors assign a level of damage to an entire building. Sources for field surveys include research-based reconnaissance teams, government/stakeholder survey teams, and citizen science groups. Field measurements are the most accurate *measurement* of damage, though have limited coverage in the few weeks that are required to make crucial early recovery plans. Secondary damage data come in the form of *inference* from predictive forecasts and

observational *estimates* from remote sensing sources. Damage *inference* from engineering forecasts are often based on an estimate of shaking intensity, exposure, and a function that translates shaking intensity to loss. On the other hand, remote sensing-derived data provide observational *estimates* of damage based on sensors. Another form of secondary data are geospatial covariates that are predictive of building damage, such as shaking intensity itself. Secondary data sources are useful in that they become available in the first week after an earthquake and have dense spatial coverage, though are highly uncertain and require calibration to the locally observed building damage.

We chose these case studies largely because of their exhaustive field surveys that can be used for training and validation, as well as their diversity of secondary data (Table 1). For Haiti, New Zealand, and Nepal, the national governments each coordinated a large-scale field survey census of all buildings in the region to inform recovery planning (MTPTC, 2010; Tonkin and Taylor, 2016; Government of Nepal Central Bureau of Statistics, 2015; Lallemant et al., 2017). In Italy, Fiorentino2018DamageEarthquakes coordinated an assessment of damage for 235 out of 300 buildings in the center of the town of Amatrice, which we supplemented with 425 surveys from the European Commission's Joint Research Centre (The European Commission, 2017).

The *field surveys* in each country used different scales to represent damage. Nepal and Italy used the EMS-98 damage grading system (Grünthal, 1998), Haiti used a modified ATC-13 grading system (Applied Technology Council, 1985), and New Zealand used the building damage ratio, which is the ratio of repair cost to the greater of the replacement cost or valuation of a building (Tonkin and Taylor, 2016).

We include at least one of each category of secondary data (remote sensing-derived, engineering forecast, or geospatial covariate) for each case study, when available. The main *geospatial covariate* for each location is the USGS Shakemap produced for each event (Worden et al., 2016). In addition, we include an estimate of near-surface soil stiffness (Vs30) to represent site-conditions (Allen and Wald, 2009; Foster et al., 2019).

Damage Proxy Maps (DPM) are *remote sensing-derived estimates* of damage that are automatically derived from inSAR data (Yun et al., 2015). The DPM provides a unitless measure of damage per pixel. The Advanced Rapid Imaging and Analysis project started producing DPMs after the February 2011 New Zealand earthquake, so we include them for the most recent three earthquakes (New Zealand, Nepal, and Italy).

The Haiti and Italy earthquakes had remote sensing-derived datasets that were manually produced through crowdsourced and expert interpretation, respectively. After the Haiti 2010 earthquake, the Joint Research Centre of the European Commission (JRC), UNOSAT, and the World Bank coordinated a large scale effort to assess point level damage from remote sensing imagery (Corbane et al., 2011). A team from the JRC assigned building-level EMS-98 damage grades to all buildings in their study area. ImageCat and the World Bank coordinated a crowdsourcing approach to damage assessment with the GEO-CAN (Global Earth Observation–Catastrophe Assessment Network) community, a group of over 600 online engineering and scientific experts (Corbane et al., 2011; Ghosh et al., 2011). The GEO-CAN effort identified heavily damaged and destroyed buildings. We combine the two assessments, as they covered complementary areas. Similarly, after the Italy earthquake, Copernicus assigned damage to individual buildings in satellite imagery using the EMS-98 damage grading system (The European Commission, 2017).

We develop our own *engineering forecasts*, either predicting probability of collapse or average damage grade using the ShakeMap and fragility curves available for each country's housing types. More information on the development of these forecasts is included in Appendix A.

Table 1. Case study context as well as the sources and characteristics of data used for analysis.

		Haiti 2010	**New Zealand 2011**	**Nepal 2015**	**Italy 2016**
Context	**Density**	Urban	Urban	Rural	Rural
	Dominant housing type	Concrete frame single-story	Timber frame single-story	Unreinforced stone with mud mortar single-story	Unreinforced masonry multistory
Original Damage Data	**Field survey metric (numerical scale)**	ATC-13 damage states (1-7)	Building damage ratio (0-0.75)	EMS-98 damage grades (1-5)	EMS-98 damage grades (0-5)
	Geospatial covariates	Shaking intensity Vs30	Shaking intensity Vs30	Shaking intensity Vs30	Shaking intensity Vs30
	Remote sensing-derived - *Automatic*	*N/A*	Damage Proxy Map (DPM)	Damage Proxy Map (DPM)	Damage Proxy Map (DPM)
	Remote sensing-derived - *Manual*	GEO-CAN / JRC assessment	*N/A*	*N/A*	Copernicus damage grading
	Engineering Forecast	Self-developed	Self-developed	Self-developed	Self-developed
Prepared Dataset	**Granularity**	Gridded (100m x 100m)	Building-level	Gridded (300m x 300m)	Building-level
	Predicted value	Collapse rate	Building damage ratio	Average damage grade	Damage grade
	Number of data points	2353	58,426	28,190	660
	Size of region (km^2)	60	1060	45,000	80

We prepared the original damage data for modeling by extracting or transforming each secondary dataset to the same level of *granularity* as the field surveys. In Haiti and Nepal, we translated the building level field assessments to grid-level (100m and 300m, respectively), due to data availability and/or to match primary and secondary data when coordinates did not align. New Zealand and Italy remained at the building-level.

The *predicted value* of G-DIF varied between case studies as well. In Haiti, we predicted collapse rate per grid, since the secondary data from GEO-CAN and JRC focused on collapse. In Nepal, we predicted average damage grade per grid. In New Zealand and Italy, we directly predicted the field surveyed value of each building (i.e. building damage ratio or damage grade). Maps of the field data are included in Figure 2.

The size and scale of the *final* prepared dataset for each location also varies, as shown in Table 1 and visualized in Figure 1. The New Zealand dataset is the largest with 58,426 buildings included. Italy, on

the other hand, only consists of 660 buildings. Because we converted the Haiti and Nepal data to gridded datasets, Haiti and Nepal's final datasets for modeling contain 2,353 and 28,190 grid points, respectively. However, the *size of the region* that Nepal's dataset covers is the largest, at approximately 45,000 km^2. The Haiti, New Zealand, and Italy data cover much smaller regions, with a maximum area of about 1060 km^2.

3 Methods

In this section, we first provide an overview of G-DIF, which is described in more detail in Loos2020G-DIF:Damage. We then introduce the simulations used to evaluate G-DIF's generalizability, the effect of the secondary data, and the effect of the field sample.

3.1 G-DIF: Geospatial Data Integration Framework

The basis for G-DIF is regression kriging, a geostatistical method that uses a sparse sample of field surveys with spatially exhaustive secondary data to predict building damage at all locations.

Consider a region affected by an earthquake that is composed of n grids or buildings, each with location s. The true damage in this region is a function of location, $Z(s)$, and is expected to be spatially correlated due to the factors that drive building damage (shaking intensity, building characteristics, etc.). Therefore, we decompose the true damage into a spatial trend, or the average damage throughout space, with spatially correlated errors:

$$Z(s) = m(s) + \varepsilon(s), \tag{1}$$

where $m(s)$ is the trend and $\varepsilon(s)$ is the error.

After an earthquake, the data that is available is a set of p secondary datasets, $\mathbf{X} = X_1 \dots X_p$, at all n locations and field surveys of damage $\mathbf{Z}$ at a subset of n_{fs} locations. Our goal is to use these data to estimate the damage at an unsurveyed location s_0. Consider the simple example of one unsurveyed location, though the method scales to multiple unsurveyed locations. To estimate the trend, m, we develop a regression function f between field measured damage $\mathbf{Z}$ and the secondary datasets. We then predict the trend at s_0:

$$\hat{m}(s_0) = f(\mathbf{X}(s_0)). \tag{2}$$

A residual will exist between the trend model and the true damage. The residuals at all locations have a mean of zero, but are likely to be spatially correlated because the trend model will not capture all sources of spatial correlation. The residual at s_0 can thus be estimated using the spatial covariance between residuals at all n_{fs} field surveyed locations. We derive the spatial covariance structure based on the semivariance, γ, or dissimilarity in the residual between two field surveyed locations as a function of their separation distance, h:

$$\gamma(h) = \frac{1}{2}\text{var}\big[\varepsilon(s) - \varepsilon(s+h)\big]. \tag{3}$$

Broadly, building damage varies in space, due to both the trend and spatial patterns in variability. Here, we consider the trend through the function, f, which accounts for some of this variation in space. We also assume second-order stationarity, or that the spatial patterns in variability have the same covariance structure across the entire affected region. This allows us to develop a single covariance structure for the entire study area through evaluating an empirical variogram.

An empirical variogram ($\hat{\gamma}(h)$) can be constructed using sample variances of observed residuals (at field-surveyed locations) with separation distance h. A theoretical variogram can then be fit through each $(\hat{\gamma}, h)$ pair. The theoretical variogram is used to solve for the kriging weights, λ, which we implement in

Ordinary Kriging by weighing the known residuals at all field surveyed locations to estimate the unknown residual at s_0:

$$\hat{\varepsilon}(s_0) = \sum_{\alpha=1}^{n_{fs}} \lambda_\alpha(s) \cdot \varepsilon(s_\alpha). \tag{4}$$

To obtain the final damage estimate, $\hat{Z}(s_0)$, we substitute the results from Equation 2 and Equation 4 into Equation 1.

3.1.1 Application to case studies

The above general integration framework is then applied to the case study data of Table 1. In New Zealand and Italy, where building-level data is available, the direct survey of damage is used as the true damage, $Z(s)$. In Haiti, where data is at grid-level, $Z(s)$ is the collapse rate: the percentage of buildings in a grid with a damage state of six or seven. In Nepal, where data is also at grid-level, $Z(s)$ is the average damage grade of buildings in a grid.

To model the trend, we mainly use a linear ordinary least squares (OLS) regression as our function f in this study. This is a common approach in Regression Kriging (Hengl et al., 2007), though assumes independent residuals which is not entirely consistent with the spatial correlation structure of the residuals. Other models, such as generalized least squares, general additive models, regression trees, and artificial neural networks have also been used for Regression Kriging in order to allow for greater flexibility with regard to these features (Chiles and Delfiner, 2012a; Hengl et al., 2003; Grujic, 2017; McBratney et al., 2000; Motaghian and Mohammadi, 2011). Here, we apply OLS to be able to compare models across simulations and across case study locations. All secondary data is standardized to have a mean of zero and standard deviation of one. In cases where high collinearity exists between secondary data–enough so that it impedes fitting of the coefficients of the trend model–we implement mixed stepwise selection (Hastie et al., 2009).

To model the spatially correlated residuals, we consider an exponential, spherical, or Matern theoretical variogram and selected the model with the lowest sum of squared errors. For New Zealand and Nepal, with large field sample sizes, we also apply local kriging to restrict the maximum number of points considered for prediction at s_0.

3.2 Simulation study to evaluate efficacy of G-DIF

We perform a simulation study to evaluate the efficacy of G-DIF in adapting to multiple contexts and damage datasets. Each simulation uses the following procedure:

1. Sample one realization of field surveys.
2. Use the field survey sample to fit the models described above.
3. Use the fitted models to develop the final damage estimate.
4. Calculate performance metrics for the error between damage estimate and true damage at all unsurveyed locations.
5. Repeat Steps 1 through 4 1000 times.

Rows one through three of Figure 2 demonstrate the model building process for one realization of field surveys across all case study locations. By repeating this procedure with 1000 simulations, we estimate and account for the uncertainty in G-DIF's damage estimate due to the field survey sample.

Our goal is to compare G-DIF's damage estimate to alternative damage estimates or alternative configurations of G-DIF using this procedure. We evaluate each option using the error e between the damage

estimate and the true damage at each location i in all n_{val} unsurveyed locations in the study area that were not included in the field survey sample.

$$e_i = \hat{Z}(s_i) - Z(s_i) \tag{5}$$

The distribution of error for one realization of field surveys is shown in row four of Figure 2. The main performance metric is mean squared error (MSE), which measures the overall bias and variance of the error distribution. The MSE is

$$MSE = \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} e_i^2. \tag{6}$$

We also calculate the bias (mean error, ME) and variance (variance of the error, VE) themselves.

$$ME = \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} e_i \tag{7}$$

$$VE = \frac{1}{n_{val}} \sum_{i=1}^{n_{val}} (e_i - ME)^2 \tag{8}$$

Values closer to zero are preferred for all three metrics. A lower ME means the average error is closer to zero, a lower VE means the spread in error is closer to zero, and the MSE captures the combination of these two. It is straight forward to calculate error for G-DIF, as the units are the same as the true damage from the field surveys. However, for some engineering forecasts, the prediction varies from probability of collapse to mean damage ratio. The calculation of error for these secondary datasets is included in Appendix B.

3.2.1 Baseline comparison

We first benchmark the accuracy of G-DIF's damage estimate against that of secondary data alternatives, focusing on sources that produce tangible damage estimates (i.e. manually derived estimates from remote sensing and engineering forecasts). In this initial comparison, we use a set of field surveys that represents the full distribution of damage and is spatially distributed throughout the entire region. We use two sample sizes of field surveys: a consistent sample size of 100 points (i.e. buildings or grids) in all four cases and a sample size that is likely to be collected within the first week after a disaster. By using two sample sizes, we demonstrate how G-DIF's damage estimate changes with different amounts of field surveys for all case study locations, as we previously showed only in Nepal (Loos et al., 2020).

3.2.2 Utility of secondary data sources

The accuracy of G-DIF depends on the secondary data that is included in the integration. Certain secondary data types are more informative than others. To evaluate the utility of each dataset we use only one secondary dataset at a time in the trend model of G-DIF. We then evaluate the MSE of the damage estimate produced by only the trend model (or row one in Figure 2), because the spatial correlation model tends to compensate for secondary datasets that are poor predictors. We repeat this procedure 1000 times with different random samples of field surveys. Within each case study, we use the same 1000 random samples with each secondary dataset, to ensure a fair comparison.

3.2.3 Evaluating the effect of the field survey sample

Beyond the secondary data, the field survey size and sample configuration will also affect the G-DIF estimate. For the previous comparisons, we used a random sample of field surveys used to calibrate the secondary

damage data in G-DIF. This is not realistic, as it is unlikely that surveyors will be able to reach a fully random and spatially distributed set of locations in the aftermath of an earthquake.

We therefore compare G-DIF's damage estimate from a random sample of field surveys to a more realistic scenario where the surveys are spatially clustered in a small sub-region. The spatially clustered sample emulates a situation where surveyors can only reach one neighborhood in the first week after a disaster. We again compare the MSE, ME, and VE of G-DIF's damage estimate using the above simulation procedure, comparing random realizations of samples from the two field survey configurations.

For this comparison where the focus is on the field survey data, we also compare G-DIF to the alternative where no secondary data is available and the field surveyed damage is interpolated directly. We do this by using Ordinary Kriging to spatially interpolate the damage from the field survey sample. Ordinary kriging is an univariate geostatistical prediction method as opposed to the multivariate regression kriging (Chiles and Delfiner, 2012b). Ordinary kriging assumes the average damage throughout space is an unknown constant, whereas in regression kriging the average is varying (and captured with the trend model). Here, we apply Ordinary Kriging by developing a variogram directly with the field surveys of damage, using this to solve for the kriging weights used for spatial interpolation.

4 Results and discussion

In this section, we first apply G-DIF to the four case studies to illustrate the components of the framework. We then provide a benchmark comparison between damage estimates produced from G-DIF and each secondary dataset. Analyzing this result further, we consider which secondary dataset leads to the most accurate prediction of damage within G-DIF. Finally, we evaluate the effect of the configuration of the field surveys on the prediction error of G-DIF. These analyses provide three main takeaways: 1) G-DIF is more accurate than any forecast or remote sensing-derived dataset in each case study, 2) the most predictive secondary source of data varies between case studies but generally a Shakemap can be sufficient as the only covariate, and 3) G-DIF effectively predicts true damage even with spatially clustered field surveys.

4.1 Application of G-DIF to four case studies

We apply G-DIF to the four events' data in Figure 2. This initial application uses the number of field surveys we estimate to be possible to collect within a week. However, this number can vary between events and is not well-documented, so, here, we assume that a field surveyor can carry out 20 damage surveys per day and that there are more field surveyors available after more damaging earthquakes based upon personal reconnaissance experience. In Haiti, we use a field sample of 50 grids (2.1% of all grids). With an average of 145 buildings per grid, this would result in about 7,250 buildings being surveyed, which could be completed by 50 surveyors over 7 days. In New Zealand, we use a field sample of 3,000 buildings (5.1% of all buildings), which could be completed by 30 surveyors over 5 days. In Nepal, we use 500 grids (1.8% of all grids). Nepal has an average of 10 buildings per grid, leading to about 5,000 buildings being surveyed in total, which could be completed by 50 surveyors over 5 days. Finally, in Italy, we use a sample of 60 buildings (9.1% of all buildings), which could be completed by 3 surveyors in 1 day. As an alternative, we also consider a scenario with a consistent field survey sample of 100 buildings or grids across all case study locations. Here, we initially consider a random sample of field surveys.

Figure 2 demonstrates the step-by-step components of G-DIF and the resulting histogram of error when comparing G-DIF's damage estimate to the full set of field surveys. The first row of Figure 2 shows the estimated trend (Equation 2) from a linear regression model after standardizing each secondary dataset predictor to have a mean of zero and standard deviation of one. The second row then shows the residuals at the field survey locations between the true damage and estimated trend, once they have been interpolated using Ordinary Kriging (Equation 4). The third row is the final integrated damage estimate from G-DIF,

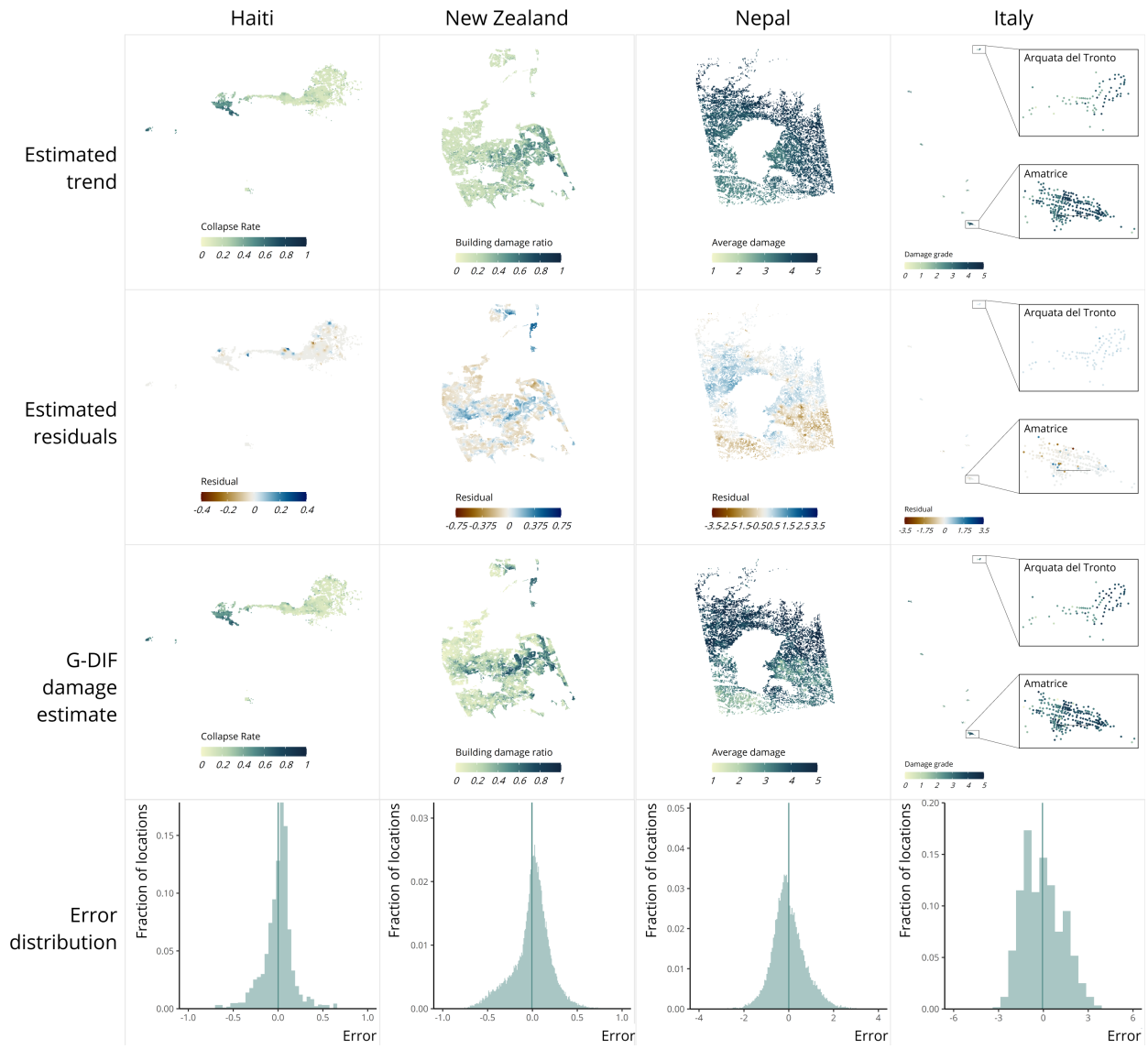

Fig. 2. Example application of G-DIF and associated error for one realization across four case studies. The top three rows show the trend model, estimated residuals, and integrated damage estimate from G-DIF using *one realization* of a random sample of field surveys that could be collected in one week. The fourth row shows a histogram of errors between G-DIF's damage estimate and the true damage in all case study locations, for this one realization, with the mean error indicated as the vertical line.

which is the sum of the estimated trend in the first row and the estimated residuals in the second row (Equation 1). The integrated damage estimate is then compared to the true damage from the full set of field surveyed damage. Note that the prediction unit varies for each case study location depending on the field data: Collapse Rate in Haiti (number of buildings in Damage States 6 or 7 over the total number of buildings), Building Damage Ratio in New Zealand, Mean Damage Grade in Nepal, and Damage Grade in Italy. Finally, the bottom row shows the distribution of error between the integrated damage estimate and the true damage for this one realization. In this error distribution, we highlight the mean error (ME) with the vertical line in the error distribution. In the following sections, we calculate the overall mean squared error (MSE) of each realization's error histogram to capture the change in G-DIF's prediction error with different

field survey samples.

4.2 Baseline comparison of G-DIF to secondary data alternatives

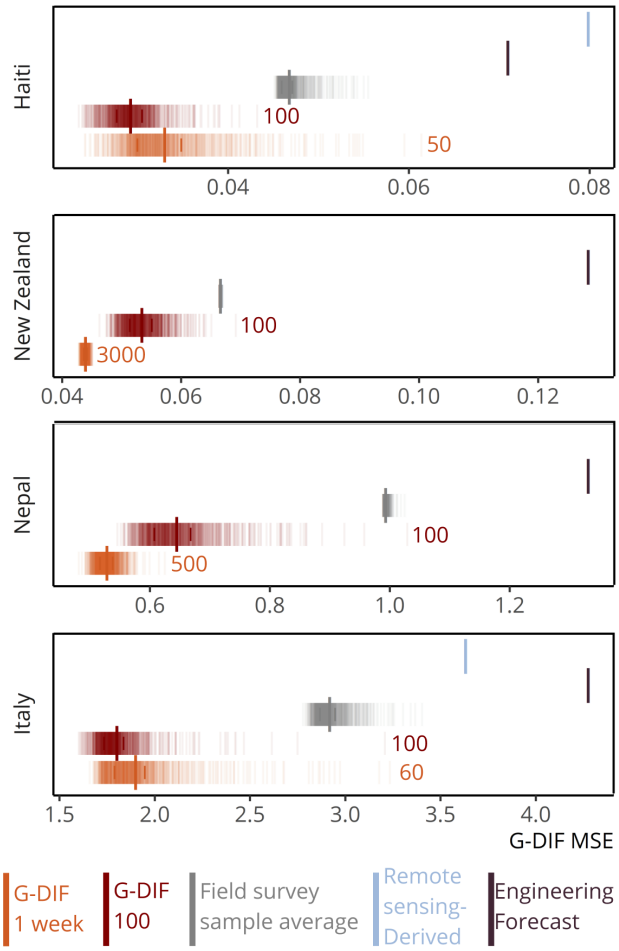

Fig. 3. Baseline comparison of G-DIF to individual secondary damage data. Distribution of mean squared error (MSE) of G-DIF's damage prediction across all case studies is shown in orange and red. G-DIF 1 week in orange uses the amount of field surveys that can be collected in one week and G-DIF 100 in red uses 100 points in the field survey sample. Sample sizes are annotated next to each distribution. Each vertical line is the MSE of G-DIF's damage estimate using one field sample realization, the dark middle line is the average MSE, the left line is the 25th percentile, and right line is 75th percentile. G-DIF's MSE is compared to predicting the average damage of the field surveys, remote sensing-derived estimates, and engineering forecasts. The remote sensing-derived estimate is from crowdsourcing in Haiti and manually-interpreted in Italy. The MSE from the remote sensing-derived estimates and engineering forecasts are single lines since they do not depend on field surveys.

To evaluate the performance of G-DIF relative to any single secondary damage dataset, we compare prediction errors from the two approaches. We compare the MSE's of the G-DIF damage estimate and that from single secondary data predictions. We consider 1000 realizations of random samples of surveys, and compute MSE values for each, as shown in Figure 3. Again, a MSE closer to zero means that the predicted damage is closer to the observed damage.

Figure 3 shows the distribution of the MSE from G-DIF predictions built using 100 buildings/grids in the field survey sample (in red, second from the bottom) or with an amount of field surveys that could be collected within a week (in orange, bottom). Figure 3 shows that G-DIF predictions, with both field survey amounts, result in lower prediction errors than any alternative secondary damage dataset.

G-DIF's lower MSE compared to the secondary data in the four case studies confirms that G-DIF is indeed generalizable to multiple locations when using these secondary datasets and a random set of field surveys. This better performance occurs because G-DIF includes all of the datasets available, but weighs the more predictive datasets as more important to the final prediction. Even with a very predictive set of secondary data, there will always be residuals between the trend and the field surveyed damage. The spatial correlation model addresses this by spatially interpolating those residuals to all locations. Therefore, G-DIF improves upon any secondary damage dataset by combining it with other data and also interpolating the remaining residuals to more closely match the field surveyed damage.

This approach does require enough field surveys to build the trend model and the variogram. With fewer field survey samples, the performance of G-DIF is worse—this can be seen in Figure 3, where the G-DIF MSE's are lower in all four case studies for the row with more field surveys. In addition, for a region with few buildings that are far apart, like in Italy, it may be difficult to build an accurate variogram to capture small-scale spatial correlations. This can affect the performance of G-DIF–a few of the Figure 3 realizations in red and orange for Italy have an MSE similar to the manually-interpreted remote sensing-derived damage estimate.

Notably, we also compare the MSE when predicting one value at all locations, the average damage of the field survey sample, shown in gray in Figure 3. This is a naive comparison, although interestingly, the MSE of the average of the field surveys falls between G-DIF and each secondary dataset. The better performance of G-DIF than the average of the field surveys is to be expected, as G-DIF will be spatially heterogeneous as compared to the single value of the average. G-DIF uses this same set of field surveys as calibration for the secondary datasets in each location. The lower MSE of the average of the field surveys than each secondary dataset indicates that the damage estimates in the secondary data may be over or underpredicting the overall damage, leading to larger overall errors.

4.3 Predictive power of secondary data

Here, we examine the predictive power of each secondary dataset within G-DIF to evaluate which are most useful to collect after an earthquake. For each case study location, we evaluate the error (MSE) of the damage estimate from the trend model when using only one secondary dataset as the predictor, comparing remote sensing-derived damage estimates, forecasts, distributions of shaking, and Vs30. Figure 4 shows the distribution of error (MSE) of the trend estimate; we again vary the locations of the field surveys used to train G-DIF.

The most predictive secondary dataset varies in each case study location. The secondary datasets that show distributions closer to zero on the left in Figure 4 are more predictive of damage, and therefore useful to collect post-earthquake and use within G-DIF. Generally, the shaking intensity provides a trend estimate that is both consistent and has relatively low errors. Engineering forecasts are closely aligned with shaking intensity and perform similarly—especially in Haiti and New Zealand where forecasts were modeled assuming the same structure type across the entire affected region. On the other hand, remote sensing-derived data can be more predictive of true damage than an engineering forecast, as seen in New Zealand and Italy. In Haiti, the GEO-CAN crowdsourced data results in a lower MSE than the engineering forecast in some cases, though it has larger variability.

The observational nature of remote sensing-derived damage data can lead to more accurate estimates of damage compared to the predictive nature of forecasts. Observations from the event based on nadir imagery (imagery seen from above) can capture small-scale variations in damage that cannot be captured in

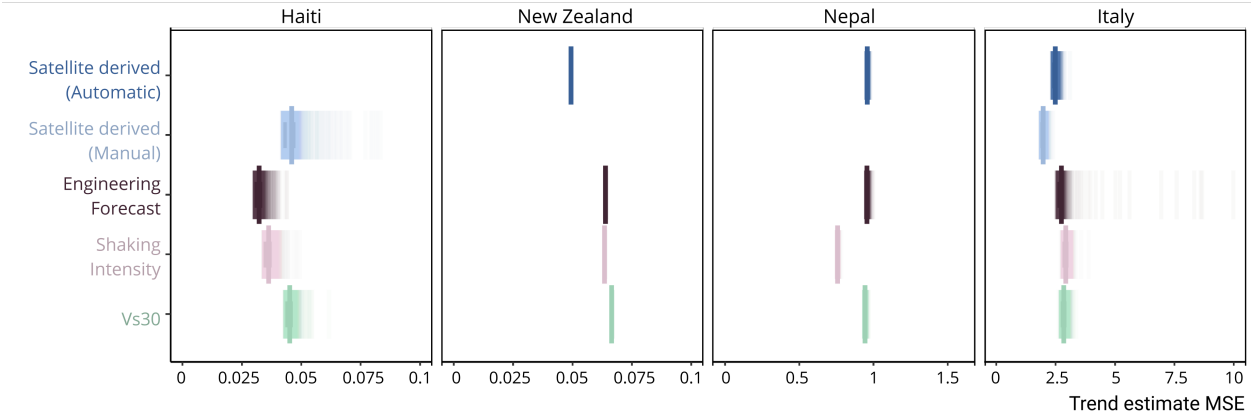

Fig. 4. Comparison of secondary datasets' predictive power in G-DIF. Each vertical line is the MSE of G-DIF's trend model when using each secondary dataset as the only covariate for one field sample realization. Each realization uses a random sample of field surveys that could be collected within a week. The dark middle line is the average MSE, the left line is the 25th percentile, and right line is the 75th percentile. The manual remote sensing-derived damage estimate in Haiti is from crowdsourcing and in Italy is from expert interpretation.

the predictive forecasts. The DPM data in New Zealand strongly correlates with the true damage compared to the engineering forecast, as seen in Figure 5. This is because the DPM detected areas of liquefaction that were not included in the forecast. On the other hand, in Nepal, where there was little liquefaction and damaged rural houses had more potential to be shrouded by dense tree cover, the DPM performed similarly to the engineering forecast. In Haiti, the GEO-CAN estimate shows small-scale patterns of collapse near the center of Port-au-Prince that were not identified in the forecast. However, because the GEO-CAN estimate is based on crowdsourcing, it overestimates the true collapse rates in the center of Port-au-Prince, leading to the large variability in GEO-CAN's resulting MSE for Haiti in Figure 4. Therefore, the method of deriving damage from satellite imagery and the mechanisms of damage influence whether a remote sensing-derived estimate is more predictive of the true damage than a forecast.

Figure 4 provides intuition behind which datasets are most useful to include in G-DIF after an earthquake occurs. In all cases, the shaking intensity from the Shakemap is predictive of damage. This means that after earthquakes where other types of secondary data are not available, this dataset is sufficient as the only predictor in the trend. However, the addition of a remote sensing-derived dataset has the potential to improve the accuracy of the trend estimate due to the increased spatial granularity of the remote sensing-derived damage estimate. Using one field survey sample across all case study locations, we compared a trend model with just shaking intensity to a model with both shaking intensity and a remote sensing-derived dataset as predictors using the F test (Williams, 1959). We found that in each case study location, the model with a remote sensing-derived dataset to the trend model was significantly different than a model without ($P \leq 0.05$ for all five remote sensing datasets, and $P \leq 0.001$ for three). Thus, the addition of a remote sensing-derived dataset will improve the accuracy of the trend model if it is available.

4.4 Effect of field survey configuration

In addition to the secondary datasets, the field surveyed damage data has a large influence on G-DIF's performance. Here we evaluate the effect of the configuration of the field survey sample on G-DIF's damage estimate, focusing on Haiti (the case study with relatively poor quality secondary data).

The prior simulations of G-DIF used a random sample of field surveys with locations scattered throughout

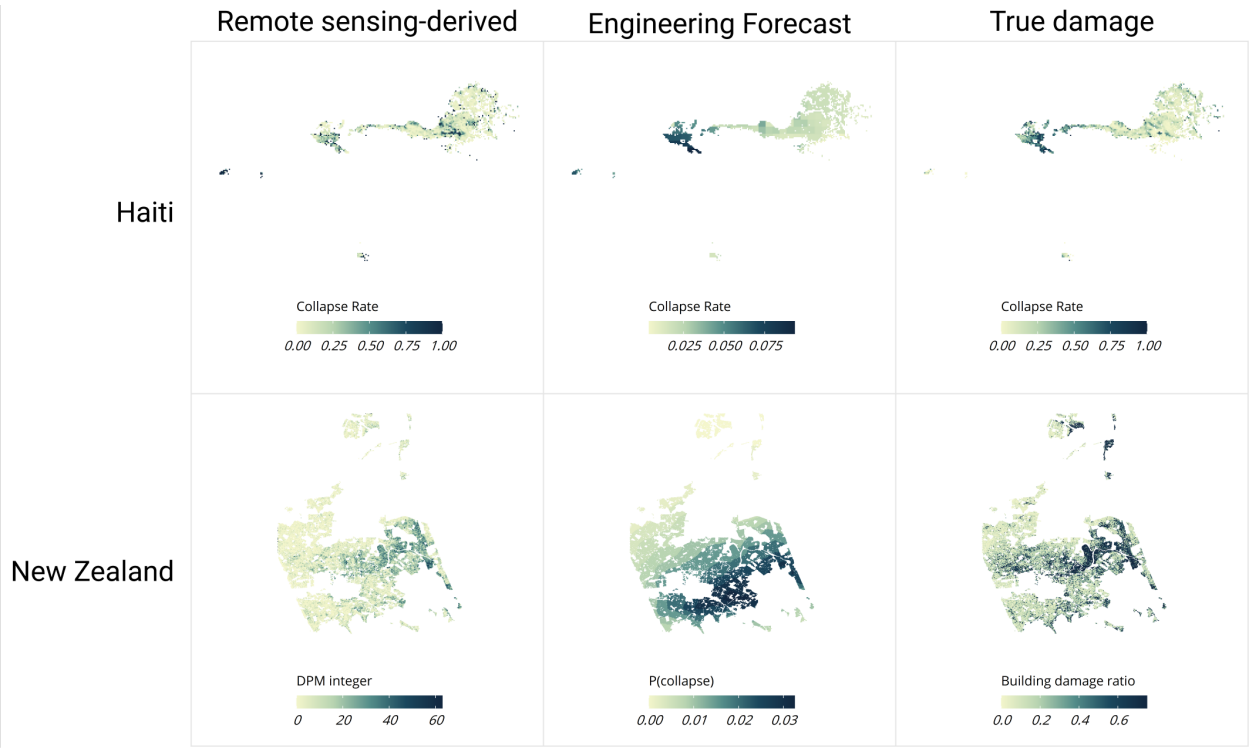

Fig. 5. Maps of remote sensing-derived damage estimates versus engineering forecasts. The remote sensing-derived estimate in Haiti is manually interpreted using crowdsourcing (Ghosh et al., 2011) and in New Zealand is automatically-interpreted (Yun et al., 2015). Haiti's engineering forecast is in units of probability of collapse.

the affected region. With a random sample of surveys, it is possible to directly interpolate the field surveys using Ordinary Kriging without including any secondary datasets. Figure 6 compares the damage estimate from G-DIF, which integrates every damage data available, to the damage estimate from Ordinary Kriging, which only interpolates the field surveys. Figures 6a and b show the damage estimate in two dimensions, whereas Figure 6c shows the damage estimate in one dimension. With a random sample, G-DIF and interpolating the field surveys produce similar performance. This is because with a sufficient number of field surveys at separation distances within the range of spatial correlation of the surveyed building damage, Ordinary Kriging will provide a smooth interpolation.

As mentioned before, it is unlikely that survey teams will be able to reach a random sample of locations within a week after an earthquake. With a random sample, directly interpolating the field surveys using Ordinary Kriging can perform similarly to G-DIF, like in Figure 6. However, in most cases field surveys will only be collected in certain localities. The trend model within G-DIF makes it preferable in these more realistic scenarios, where field surveys are sampled in one area of the affected region. Figure 7 shows the effect of a spatially clustered field sample G-DIF and direct interpolation of the field surveys using Ordinary Kriging. G-DIF's damage estimate is similar when using the clustered sample in Figure 7 and the random sample in Figure 6. In Figure 7, G-DIF's final damage estimate converges to the trend model's damage estimate in the area away from field surveyed data. Conversely, the damage estimate from Ordinary Kriging converges to the average damage, a single value, from the field surveyed sample in this same area.

Ordinary kriging will predict the average field surveyed damage at all locations outside the range of spatial correlation, as seen in Figure 7. Therefore, in a real scenario, Ordinary Kriging will predict the average damage at locations that did not experience shaking, whereas G-DIF has the potential to predict zero to low damage at those locations because of the trend model. Because we do not have data on buildings that were outside of the affected areas in these four events, any calculated error of Ordinary Kriging presented from hereon overestimates the performance of its damage estimate.

We repeat the comparison between G-DIF and Ordinary Kriging, simulating each method's damage estimate with 1000 different field survey samples from the random or clustered field survey configuration. The clustered sample is constrained to the same small area of 54 grids, and we select a different sub-sample of 40 field surveys within this area. The sample of random surveys can be at any location in the affected region. The results of these simulations are shown in Figure 8.

The top row of Figure 8 shows the distributions of MSE, ME, and VE for the two methods' damage estimates when a random field survey sample is used to fit the models. With a random sample, G-DIF and Ordinary Kriging damage estimates have similar average values of MSE, ME and VE, as shown by the middle vertical lines in each distribution. Again, the actual performance for Ordinary Kriging is likely lower than what we are able to calculate with our field survey data. The majority of field sample realizations for G-DIF's damage estimate have lower MSE values than Ordinary Kriging, as indicated by comparing the 25th and 75th percentile vertical lines for each. Though for a very small percentage (0.7%) of field sample realizations, G-DIF does have higher MSE values than Ordinary Kriging. This is most likely due to a poor trend model fit for G-DIF and Ordinary Kriging being a weighted average of nearby points. With the more realistic, spatially clustered sample (bottom row of Figure 8), G-DIF's performance is markedly better than Ordinary Kriging, as seen by the lower interquartile range for G-DIF's MSE distribution in the bottom row. This is because interpolating the field surveys using Ordinary Kriging underestimates the true damage, as shown in Figure 7, resulting in a biased estimate where the mean error is always negative in Figure 8.

These simulations comparing field survey configurations were applied to Haiti, the case study location with poorer quality secondary data. We also used a small percentage (1.7%) of field survey samples to fit the models in G-DIF and Ordinary Kriging. In a case study with more predictive secondary data and with more field surveys, we expect the performance of G-DIF to further improve.

The effect of the field survey sample on G-DIF's damage estimate indicates the need for thoughtful planning of the field sampling strategy immediately after an earthquake. We discuss ways to improve rapid

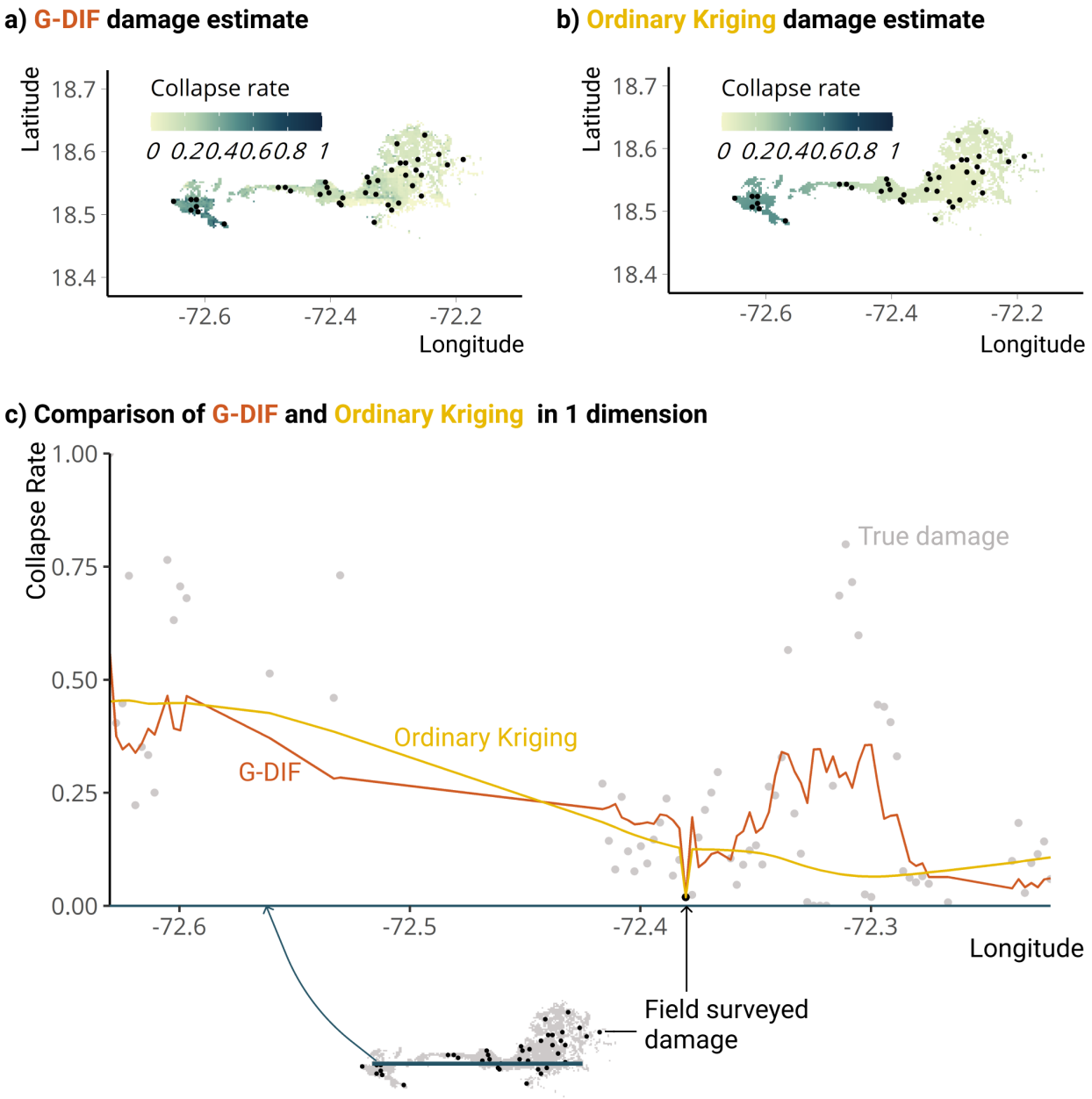

Fig. 6. G-DIF versus Ordinary Kriging using a random sample of field surveys. The estimated distribution of collapse in Haiti, shown in two dimensions, resulting from (a) G-DIF, which integrates field surveyed damage with secondary data, (b) Ordinary Kriging, which only interpolates the field surveyed damage, shown by the black points. Both methods use a *randomly distributed* set of field surveys. The spatial variation in estimated collapse rate from G-DIF and Ordinary Kriging is also shown in one dimension in (c), when plotting collapse along the teal line cutting horizontally across the map shown at the bottom.

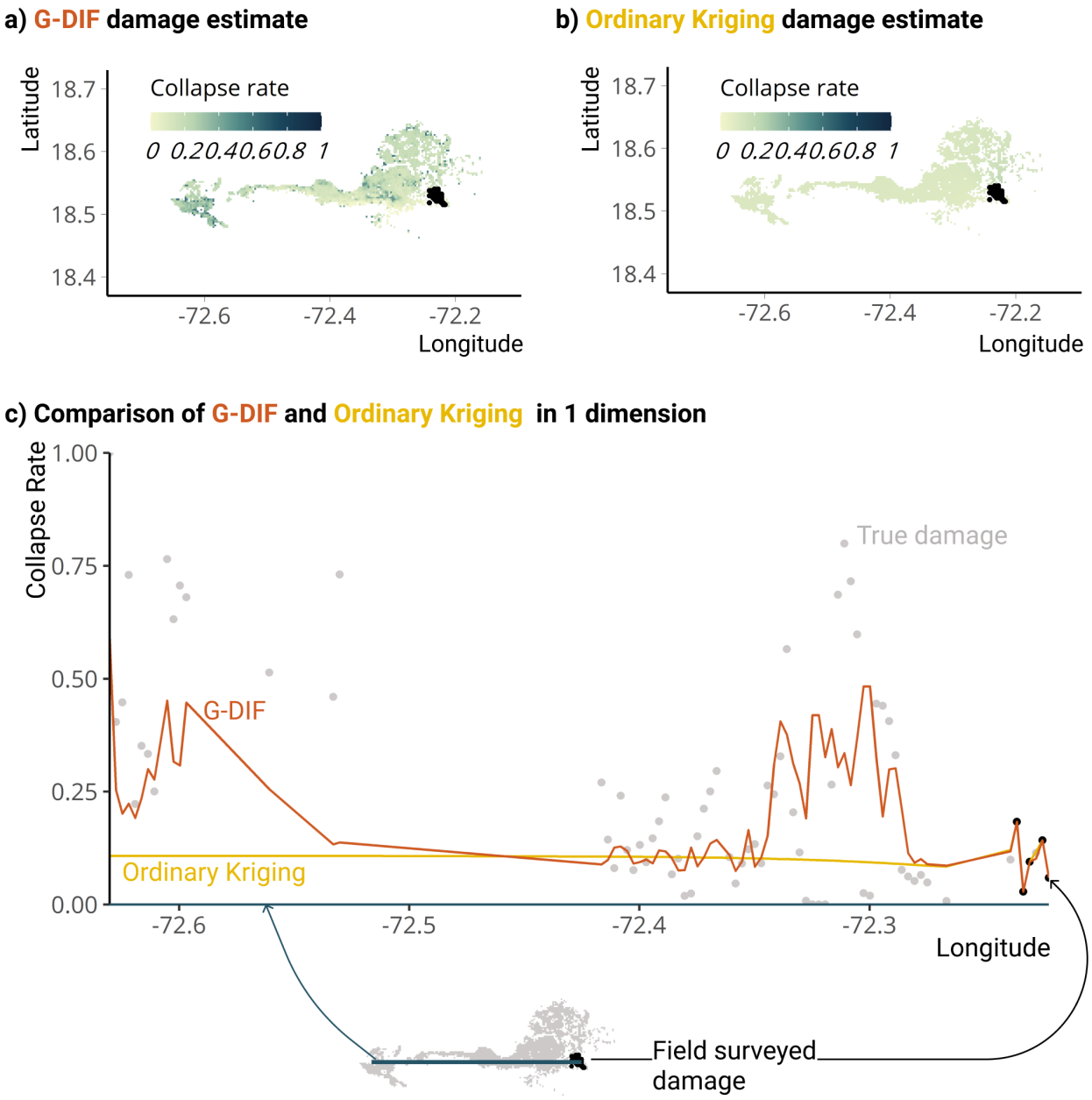

Fig. 7. G-DIF versus Ordinary Kriging using a clustered sample of field surveys. The estimated distribution of collapse in Haiti, shown in two dimensions, resulting from (a) G-DIF, which integrates field surveyed damage with secondary data, (b) Ordinary Kriging, which only interpolates the field surveyed damage. Both methods use a *spatially clustered* set of field surveys, shown by the black points. The spatial variation in estimated collapse rate from G-DIF and Ordinary Kriging is also shown in one dimension in (c), when plotting collapse along the teal line cutting horizontally across the map shown at the bottom.

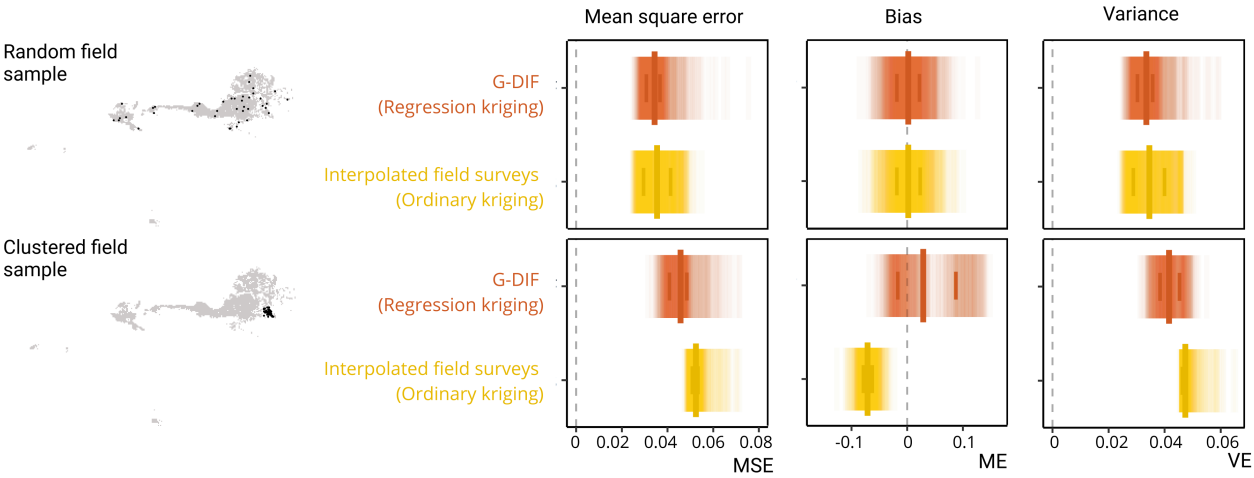

Fig. 8. Effect of field survey configurations on G-DIF versus Ordinary Kriging. Each vertical line is the error of G-DIF (red) or Ordinary Kriging's (yellow) damage estimate for one field sample realization. Error metrics shown include mean squared error (MSE), mean error (ME), or variance in the error (VE). The dashed vertical line plotted at zero is the best possible value for each metric. Each realization uses a sample of 40 field surveys in each field sample configurations, maps of which are shown on the left. The dark middle line is the average MSE, the left line is the 25th percentile, and right line is the 75th percentile.

damage estimates through field samples in the next section.

5 Recommendations for developing G-DIF in real time

The above results point to several recommendations to implement G-DIF in real time. Here, we describe considerations for modelers developing G-DIF damage predictions with the data from a future disaster. We provide an interactive code to support this section as well, which is included in the Data Availability Statement (Loos, 2022). We close this section with a summary of field survey sampling strategies to maximize G-DIF's damage prediction and examples of ways these ideas are being carried out in practice.

5.1 Developing G-DIF

Based on the testing and results summarized above, the following recommendations will enable a modeler to maximize the efficacy of G-DIF in real time after a disaster.

1. Evaluate field survey sample size The size of the field survey sample will dictate how a modeler trains and evaluates the accuracy of G-DIF's damage estimate. Methods such as k-fold cross-validation, leave-one-out cross-validation, and bootstrapping can be used for training and validating the models within G-DIF (Hastie et al., 2009). In addition, with more field data, a modeler can create a separate test set to evaluate G-DIF's potential prediction performance on unsurveyed locations. Generally, more field data will lead to a more accurate damage estimate with less variation as shown in Figure 3.

2. Establish boundary of prediction area In real time, the modeler would establish the spatial boundary for the forward prediction of damage using G-DIF. An initial boundary could be the area of strong shaking in the USGS ShakeMap, and users on the ground can help refine this by considering areas they will focus on during response and recovery. To accommodate datasets that are not available throughout the entire

prediction area, separate trend models can be built for multiple subregions (Loos et al., 2020). Alternatively, one can integrate these datasets using Bayesian methods (Booth et al., 2011; Foster et al., 2019; Lee and Tien, 2018; Noh et al., 2020). However, developing explicit methods to do this in a spatial manner requires future research.

3. Assess spatial distribution of field sample Combining the previous two steps, modelers should evaluate whether the field sample is spatially distributed across the prediction area, like the evaluation completed in Section 4.4. Modelers can also explore the separation distances of the field survey sample. Highly clustered field samples with small separation distances, like in Figure 7, will converge to the trend model in regions outside of the field surveyed area. Separated field survey samples at large distances greater than the expected spatial correlation in trend residuals may result in a "nugget" variogram, or a variogram with constant semivariance at all distances. In both cases, modelers should focus on building a predictive trend model that accurately reflects the relationships between the field survey samples and secondary data (discussed in step 5), since the resulting variogram may not improve the overall performance of G-DIF. In addition, modelers should suggest that additional field surveys be collected at those unobserved separation distances.

4. Compare secondary data at field surveyed sample to full study area The modeler should compare values of secondary data at the field surveyed locations to the full distribution of secondary data at all locations. As much as possible, the variance of the secondary data at the field surveyed locations should reflect the variance over the entire affected region. Otherwise, the trend estimate may not extrapolate well outside of sample distribution's range. This might occur if field surveys are clustered so that there is little variation in the secondary data, as evaluated in Section 4.4. If this occurs, additional field surveys should be collected, if possible, for a wider range of secondary data values.

5. Explore relationships between primary and secondary data When building the trend model, the modeler should incorporate the relationships that exist between each secondary dataset and the field survey sample. In this study, many of the relationships between the secondary data and the field data are close to linear. One can evaluate this by examining the moving average of the field damage at various bins of each secondary data, or creating a "loess" curve. However, nonlinear trends can be considered by transforming the secondary dataset using a nonlinear trend model.

6. Address redundant secondary data Some damage datasets may be collinear (e.g., an engineering forecast may be closely aligned with a ShakeMap). Collinearity can lead to unreliable estimates of the coefficients for each secondary dataset and to overfitting. Ways to evaluate collinearity of the secondary datasets are to evaluate the variance inflation factor and correlation matrix of the secondary data at the field survey locations. If collinearity is found, a modeler can address this by applying variable selection techniques, like mixed stepwise selection which was applied in this study, or aggregating collinear variables using principal component analysis.

7. Build and evaluate the trend model The modeler should choose an appropriate functional form and regression algorithm to implement for the trend model, whether it be Ordinary Least Squares (OLS), Generalized Least Squares, or a more complex regression function (Loos et al., 2020). In the case of OLS, which we implement here, the modeler should check the coefficients for each secondary dataset. A coefficient opposite from expectations (for example, predictions of decreasing damage with increasing ground shaking intensity), may indicate that the secondary dataset is unreliable or that outliers exist, and should be addressed by removing from the model. Otherwise the directions of the coefficients should reflect the relationships seen in Step 5. To evaluate the relative utility of each secondary dataset, one should ensure that each variable is

standardized to ensure that coefficients are comparable. The coefficients and standard errors of each variable in the trend model will provide intuition behind which secondary dataset has the most influence on the trend estimate, similar to the results shown in Figure 4. Finally, the modeler should assess the distribution of the residuals of the trend model to assess whether they meet the Gaussian assumptions for Kriging. If residuals are non-Gaussian, the modeler can explore methods for transforming the residuals (Cecinati et al., 2017).

8. Build and refine the spatial correlation model The choice of variogram and kriging method affects the final spatial pattern of damage, especially in situations where the secondary data are poor predictors of the true damage. In addition to selecting the variogram based on best fit, modelers should also consider the expected spatial pattern that result from the selected variogram. If the fitted variogram exhibits a trend (or the semivariance increases with distance), the field surveyed damage may not have been successfully detrended with the trend estimate or perhaps the selected variogram model is too flexible. In this case, it might be preferable to consider local kriging, where only the closest surveyed points to the unsurveyed location are considered when making the kriging prediction. If a nugget variogram is fit to the detrended survey points, the trend model may have fully captured the spatial correlation in damage. However, the modeler should compare the variogram fit to the detrended field surveys with the variogram fit to the original field surveys, to ensure the nugget variogram is not arising from a lack of closely spaced field data (as discussed in Step 3). Finally, we assume second-order stationarity in this formulation, meaning that semivariances are constant across locations and in all directions, and therefore build a single variogram. The modeler can explore developing several variograms depending on location in the study region, though this requires further research with a full set of ground-truth damage data.

9. Calculate performance metrics Finally, the modeler should calculate the model performance on a test set of field surveys held out from all model fitting. The modeler can use the fitted model to predict damage at the test set locations and calculate the mean error, variance in the error, and mean squared error. Prediction errors for newly acquired field data should be monitored—low errors would confirm the current G-DIF damage estimate, whereas high errors would trigger a model update.

5.2 Strategizing field survey collection

Thoughtful on-the-ground collection of damage data after future disasters can have a meaningful impact on resulting G-DIF estimates, since many of the above recommendations for maximizing G-DIF's damage prediction (including Steps 1, 3, 4, and 9) are influenced by the field survey sample. Decisions for field survey collection can be strategized with respect to *what* measurements are collected, *who* is collecting the measurements, and *where* those measurements are collected.

What: types of damage assessments G-DIF can adapt to different types of field survey assessments, as decided upon by local stakeholders—the EMS-98 damage grading system was employed in Italy and Nepal; a modified ATC-13 in Haiti; and the Building Damage Ratio, or Loss Ratio, in New Zealand. Therefore, it is important that the field surveys that are used to calibrate secondary damage data are in the unit that will ultimately be used for response and recovery planning purposes.

Who: sources of field surveys In this study, we demonstrate G-DIF using field surveyed damage data collected mainly by the governments of each case study location. We use government-collected damage data because the unit of measurement for the damage assessment was used later in the recovery to guide the distribution reconstruction grants and insurance payouts.

However, field surveyed damage data can come from multiple sources outside of government or stakeholder teams, including research-based reconnaissance teams or citizen science. Various groups organize

reconnaissance trips to locations recently affected by disaster including the Earthquake Engineering Research Institute (EERI), Geotechnical Extreme Events Reconnaissance (GEER), the Earthquake Engineering Field Investigation Team, and the Structural Extreme Events Reconnaissance (StEER) network. Reconnaissance teams have the advantage of containing highly trained surveyors who may be able to reach the affected region before a governmental survey is orchestrated. G-DIF makes it possible to integrate reconnaissance-collected field surveys with secondary data to estimate the expected damage at places reconnaissance teams cannot reach. Importantly, if reconnaissance teams conduct assessments in the same unit of measurement as that employed by survey teams for the government—which often go beyond red-yellow-green safety tags—a G-DIF damage estimate calibrated with reconnaissance assessments could be consequential for large-scale planning.

G-DIF is appealing because it also allows for citizen science, or community-based data collection, to be used as the field surveys to calibrate top-down assessments. This could be data collected from mobile phones, as seen in disasters like Haiti (e.g. (Corbane et al., 2012)). Or, community-based disaster preparedness groups like the Community Forest User Groups in Nepal (Gentle et al., 2020) can provide preparedness training on how to collect field surveys of damage. In fact, after the 2021 Haiti earthquake, StEER organized teams in Haiti to take multiple pictures of buildings throughout the affected area, which were then assessed for damage by remote earthquake engineers. Community-based data would be ideal to use in the week after an earthquake, when it is unlikely for government or reconnaissance survey teams to be in-country. In this way, G-DIF is able to combine bottom-up data collection with top-down damage estimates, leading to more participatory damage estimates.

Where: locations of surveys Finally, the locations of the field surveys from these sources have a direct influence on the G-DIF results. The performance of G-DIF improves with field surveys that are distributed throughout the prediction area (Step 3), have nearby separation distances within the range of the expected spatial correlation of trend residuals (Step 3), and that have representative values of secondary data (Step 4). Many of the above groups have developed strategies for field survey collection, which can be used in conjunction to gather a set of field surveys that are adequate for developing G-DIF.

The StEER network advocates for a "Hazard Gradient Survey," a sampling strategy designed to collect an unbiased estimate of damage across all hazard levels, the hazard being shaking intensity for earthquakes (Kijewski-Correa et al., 2021). While not demonstrated in this study, we have found that a biased field survey sample (e.g. when only collapsed buildings are assessed) leads to a biased G-DIF estimate. A "Hazard Gradient" approach would lead to the unbiased sample necessary for developing G-DIF. However, additional guidance could be provided on the sample design of the field data collection concerning other sources of secondary data, such as forecasts or remote sensing-derived.

Practices for strategizing field survey samples in corroboration with alternative damage data have already occurred after past disasters, usually by government or stakeholder survey teams. For example, after the Haiti 2010 earthquake, the JRC, the World Bank, and UNOSAT collected data specifically to corroborate the results from the crowdsourced damage collection (Corbane and Lemoine, 2010; Lemoine et al., 2013). After Typhoon Haiyan in 2013, REACH, in conjunction with the Shelter Cluster and American Red Cross, organized their field sample to validate crowdsourced data from Humanitarian OpenStreetMap Team (Westrope et al., 2014). Finally, after the Italy 2016 earthquake, the local government coordinated their field response based on Copernicus's damage grading map (The European Commission, 2017).

While it may not be possible to collect field surveys at an ideal set of locations in the days after an earthquake, the prediction variance from an initial G-DIF model using early field surveys (for example, from citizen science groups) can inform where to collect additional surveys.

6 Conclusion

In this study, we evaluate the ability of the Geospatial Data Integration Framework (G-DIF) to generalize, or adapt to different contexts, new datasets, and realistic field survey scenarios. Specifically, we consider real damage data from four earthquakes to understand the impacts of differing contexts and available data on G-DIF's damage estimate. We evaluate G-DIF in these four earthquakes and account for the uncertainty in the G-DIF damage estimate by repeatedly building G-DIF with 1000 different field survey samples.

We find that G-DIF is a generalizable framework and predicts damage more accurately than alternative damage datasets under various scenarios of data availability and with realistic field survey strategies. G-DIF's increased accuracy over alternative damage estimates results from the underlying regression kriging model that calibrates secondary data to the true observations of damage from the ground.

Evaluating individual secondary datasets, we find that the shaking intensity from the Shakemap is a reasonably effective initial predictor of the trend, even with poorly constrained Shakemaps. Adding a remote sensing-derived dataset produces more detailed and granular estimates of damage. Though, the added utility of remote sensing-derived data can vary between places and the source of the data. For example, the manually-derived remote sensing estimate from Copernicus in Italy resulted in consistently lower errors than other secondary datasets but the crowdsourced dataset in Haiti had less stable errors. These differences can be evaluated during the model building process through looking at the coefficients in the trend model. Overall, G-DIF's damage estimate is only expected to improve with as remote sensing-derived and forecast-based methods improve.

The accuracy of G-DIF also strongly depends on the field survey locations. G-DIF shines in comparison to interpolating the field surveyed damage using Ordinary Kriging when a spatially clustered set of field surveys has been collected. This means that G-DIF is able to predict damage in realistic post-earthquake field collection scenarios, when it is difficult to reach multiple places due to building debris or damaged infrastructure. In the unlikely case where a random, spatially distributed set of field surveys is collected, G-DIF still performs better than Ordinary Kriging. This conclusion is especially salient given that Ordinary Kriging will perform even worse at locations outside of the range of spatial correlation of damage, since it will predict the average damage everywhere. The overall error of G-DIF's damage estimate when using a random sample is consistently lower than the damage estimate resulting from the clustered sample. These differences illuminate the importance of the field surveys for estimating damage over an entire affected region.

Based on these results, we provide recommendations for collecting field surveys and implementing G-DIF in future disasters. An effective damage estimate benefits from the model forms selected by the analyst and the locations of the field surveys. Fortunately, growing experience with this approach indicates how these issues can be addressed systematically in order to develop confidence in resulting predictions.

By applying G-DIF to multiple case study datasets, we show how this framework can be used by stakeholders to combine all the damage data that is available into a single, accurate estimate of damage. Importantly, G-DIF calibrates secondary data from forecasts and remote sensing to the damage seen on the ground. This means that secondary datasets, which in many cases are derived from global techniques or models, are amended to more accurately reflect the patterns of damage that are specific to that location and that earthquake. The necessity of field data for calibration poses opportunities for the engineering community to strategize where to collect data, so field samples can be used to inform damage estimates produced from G-DIF. More broadly, G-DIF provides a framework to connect top-down damage estimates, like those produced with the PAGER system or NASA-JPL/ARIA, with bottom-up data collection, like crowdsourced estimates of damage. This study shows that G-DIF is a flexible and reliable approach to produce locally-specific damage estimates after future earthquakes.

7 Data Availability Statement

An interactive code to support the "Recommendations" section of this study are available at `https://sabineloos.github.io/GDIF-Gen/Diagnostics.html` and the supporting code generated during this study are available at the following repository: `https://github.com/sabineloos/GDIF-Gen` (Loos, 2022). Additional code and data that support the findings shared in this study are available from the corresponding author upon reasonable request. Field data used during the study were provided by a third party. Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.

8 Acknowledgments

We thank the Ministry of Public works; the Earthquake Commission and Tonkin + Taylor; the Government of Nepal and Kathmandu Living Labs; and the European Commission and the USGS for access to the field data from Haiti, New Zealand, Nepal, and Italy, respectively. Specifically, thanks to Virginie Lacrosse, Sjoerd van Ballegooy, Sang-Ho Yun, Keiko Saito, David Wald, Paolo Zimmaro in preparing and accessing these datasets. Thank you to Kishor Jaiswal and Nicole Paul for providing input on the development of the engineering forecasts used in this study. We also thank three anonymous reviewers who provided valuable feedback that improved this manuscript and shared code. This work was funded by the Stanford Urban Resilience Initiative; the John A. Blume Earthquake Engineering Center; the National Science Foundation Graduate Research Fellowship; and the National Research Foundation, Prime Minister's Office, Singapore under the NRF-NRFF2018-06 award.

A Damage Data Sources

A.1 Development of engineering forecasts for each case study

The process of developing the engineering forecast for each case study is similar in that each combines the spatial distribution of the estimated shaking intensity from the USGS ShakeMap (Worden et al., 2016) with the building stock exposure and the vulnerability of that exposed building stock. Generally, we aimed to replicate a model that could be rapidly produced with openly available datasets in each country. However, due to data limitations, we do not expect these forecasts to be of the same accuracy as those produced by risk modeling companies or agencies.

For all four case studies, we use the estimated distribution of shaking from the USGS ShakeMap, using either the macroseismic intensity, peak ground acceleration, or peak spectral acceleration depending on the intensity measure used in that case study's vulnerability curve. Specific ShakeMaps used for each case study are referenced in Table 2.

Exposure and vulnerability data varied by case study. In all cases, we prioritized using vulnerability and fragility curves that were openly available. In Haiti, most buildings at the time of the earthquake were unreinforced concrete frames with masonry infill and unreinforced masonry. Because of a lack of census data, we assume all buildings correspond to the 'C3' structure type in PAGER's collapse fragilities (Jaiswal et al., 2011), recognizing that these fragilities underestimate the true collapse from the event. In New Zealand, most residential buildings were light timber framed buildings. Again, we assume all buildings correspond to the 'W1' structure type in PAGER's collapse fragilities (Jaiswal et al., 2011). In Nepal, we follow the same process employed in (Loos et al., 2020). Fragility curves come from Nepal's National Society of Earthquake Technology (Japan International Cooperation Agency and Ministry of Home Affairs, His Majesty's Government of Nepal, 2002). We define exposure from Nepal's 2011 census and distribute the exposure according the LandScan 2011 High Resolution Global Population Dataset (Bright et al., 2012).

Finally, in Italy we estimate mean damage ratio using vulnerability curves from the Global Earthquake Model (Martins and Silva, 2020). Here, we acknowledge the building material for each surveyed building and use that as our exposure data, recognizing that this increases the precision of Italy's forecast compared to other locations.

Table 2. Damage data sources

Earthquake	Variable description	Damage data type	Original units	Source
Haiti 2010	Shaking Intensity	Covariate	MMI	USGS (United States Geological Survey, 2010)
Haiti 2010	site characterization - Vs30 (time-averaged shear-wave velocity to 30 m depth)	Covariate	Vs30 per 30 arcsec grid	USGS (Allen and Wald, 2009)
Haiti 2010	% buildings tagged as grade 4 or 5 per grid, crowdsourced	Remote-sensing based	%	UNITAR, UNOSAT, JRC, GEO-CAN (Corbane et al., 2011)
Haiti 2010	Probability of collapse, similar to PAGER	Engineering Forecast	Probability of collapse	Self developed, using PAGER collapse fragilities (Jaiswal et al., 2011)
Haiti 2010	Average central damage factor of buildings per grid, field assessed	Field survey	CDF	MTPTC Haiti (MTPTC, 2010)
New Zealand 2011	Shaking Intensity	Covariate	MMI	USGS (United States Geological Survey, 2011)
New Zealand 2011	Site characterization - Vs30 (time-averaged shear-wave velocity to 30 m depth)	Covariate	Vs30 per 100m grid	Foster and Bradley (Foster et al., 2019)
New Zealand 2011	Damage Proxy Map	Remote-sensing based	DPM value per 30m grid	NASA JPL-ARIA (Yun et al., 2015)
New Zealand 2011	Probability of collapse, similar to PAGER	Engineering forecast	Probability of collapse	Self-developed, using PAGER collapse fragilities (Jaiswal et al., 2011)

Continued on next page

Earthquake	Variable description	Damage data type	Original units	Source
New Zealand 2011	Building damage ratio, field assessed	Field survey	Ratio	Earthquake Commission (Tonkin and Taylor, 2016)
Nepal 2015	Shaking Intensity	Covariate	MMI	USGS (United States Geological Survey, 2015)
Nepal 2015	site characterization - Vs30 (time-averaged shear-wave velocity to 30 m depth)	Covariate	Vs30 per 30 arcsec grid	USGS (Allen and Wald, 2009)
Nepal 2015	Mean damage ratio per grid	Engineering forecast	Ratio	Self developed, using JICA fragilities (Japan International Cooperation Agency and Ministry of Home Affairs, His Majesty's Government of Nepal, 2002)
Nepal 2015	Damage Proxy Map	Remote-sensing based	DPM value per 30m grid	NASA JPL-ARIA (Yun et al., 2015)
Nepal 2015	Average damage grade	Field survey	EMS Damage Grade per building	Government of Nepal (Government of Nepal Central Bureau of Statistics, 2015)
Italy 2016	Shaking Intensity	Covariate	MMI	USGS (United States Geological Survey, 2016)
Italy 2016	Site characterization - Vs30 (time-averaged shear-wave velocity to 30 m depth)	Covariate	Vs30 per 30 arcsec grid	USGS (Allen and Wald, 2009)
Italy 2016	Probability of collapse, similar to PAGER	Engineering forecast	Mean loss ratio	Self developed, using GEM vulnerability curves (Martins and Silva, 2020; Martins, 2020)

Continued on next page

Earthquake	Variable description	Damage data type	Original units	Source
Italy 2016	Damage Proxy Map	Remote-sensing based	DPM value per 30m grid	NASA JPL-ARIA (Yun et al., 2015)
Italy 2016	Damage Grade, assessed from satellite imagery	Remote-sensing based	EMS Damage Grade per building	EC-JRC (The European Commission, 2017)
Italy 2016	Damage Grade, field assessed	Field survey	EMS Damage Grade per building	Fiorentino et al. (Fiorentino et al., 2018), EC-JRC (The European Commission, 2017)

End

B Validation of secondary data

To compare the accuracies of each secondary damage dataset against the field survey data (as shown in Figure 3), necessary transformations and assumptions were made. This is because the secondary data might be in different units than the field survey data, as shown in Table 3. In addition, G-DIF's damage prediction results in real numbers, which requires binning when the field survey data is measured as a positive integer (like a damage grade). We outline the procedures we took to compare each form of secondary data to the full set field validation data in each case study location.

Table 3. Summary of damage estimate translation used for validation for all case study locations. Datasets include damage information from field surveys, engineering forecasts, and remote sensing-derived data. Final units are the units of the dataset after preparing for G-DIF. Possible values are the values of the final units that dataset could take on. For New Zealand, the building damage ratio can take on values of 0-1, though the actual field data was truncated from 0-0.75. Translated? indicates whether the dataset was translated to compare to the field surveyed data.

Case study	Dataset	Final units	Possible values	Translated?
Haiti 2010	Field surveys	Collapse rate	0-1	/
	GEO-CAN/JRC assessment	Collapse rate	0-1	No
	Engineering forecast	Collapse probability	0-1	No
New Zealand 2011	Field surveys	Building damage ratio	0-1*	/

	Engineering forecast	Collapse probability	0-1	No
Nepal 2015	Field surveys	Mean damage grade	1-5	/
	Self developed	Mean damage ratio	0-1	Yes
Italy 2016	Field surveys	Damage grade	0, 1, 2, 3, 4, 5	/
	Copernicus assessment	Damage grade	0, 1, 2, 3, 4, 5	No
	Engineering forecast	Loss ratio	0-1	Yes

Generally, there are limitations in comparing field surveys and forecasted damage estimates due to differences in the damage measured and uncertainties in the exposure and vulnerability models (Silva and Horspool, 2019). Understanding these limitations, we still aim to compare the forecast to the field surveys damage to provide a relative benchmark of performance. Because different vulnerability curves were used for each case study, the resulting engineering forecasts were not always in the same units as the field surveys. For the Haiti earthquake, no translation was required, because our forecast predicts collapse probability which is analogous to the field survey units of collapse rate per grid. In New Zealand, the units of the field surveys and the forecast have the same range of possible values (0-1), so we therefore did not translate the forecast. However, it is important to note that these are fundamentally different values, since the building damage ratio is a measure of overall loss per building whereas the collapse probability is a measure of just collapse. Therefore, the calculated mean squared error for the forecast in New Zealand is likely overestimated. In Nepal, the forecast predicts the mean damage ratio per grid. To compare to the field surveys, we bin these damage ratios according to EMS-98's range of damage ratios per each damage grade (Grünthal, 1998). Similarly, in Italy, we bin the forecast's loss ratio per building to each damage grade using the same approach as in Nepal.

The remote sensing derived proxies of damage, conversely, did not need to be translated to compare with the field surveyed data. In Haiti, the GEO-CAN/JRC assessment derived from crowdsourcing assessed collapse rate per grid, same as the gridded field surveys. In Italy, Copernicus's damage assessment was also in the EMS-98 damage scale and could be directly compared to the field surveys. NASA's damage proxy map for New Zealand, Nepal, and Italy were omitted, as the assessments of damage are unitless and cannot be translated to the field surveyed damage without a model (like G-DIF).

References

Allen, T. I. and Wald, D. J. (2009). "On the Use of High-Resolution Topographic Data as a Proxy for Seismic Site Conditions (Vs30)." *Bulletin of the Seismological Society of America*, 99(2A), 935–943.

Applied Technology Council (1985). "Earthquake damage evaluation data for California (ATC-13)." *Report No. ATC-13*, Applied Technology Council.

Bhattacharjee, G., Soden, R., Barns, K., Loos, S., and Lallemant, D. (2021). "Factors affecting earthquake responders' building damage information needs and use." *Earthquake Spectra*, 87552930211030297.

Booth, E., Saito, K., Spence, R., Madabhushi, G., and Eguchi, R. T. (2011). "Validating Assessments of Seismic Damage Made from Remote Sensing." *Earthquake Spectra*, 27(S1), S157–S177.

Bright, E. A., Coleman, P. R., Rose, A. N., and Urban, M. L. (2012). "LandScan 2011.

Buchanan, A., Carradine, D., Beattie, G., and Morris, H. (2011). "Performance of houses during the Christchurch earthquake of 22 February 2011." *Bulletin of the New Zealand Society for Earthquake Engineering*, 44(4), 342–357.

Cecinati, F., Wani, O., and Rico-Ramirez, M. A. (2017). "Comparing Approaches to Deal With Non-Gaussianity of Rainfall Data in Kriging-Based Radar-Gauge Rainfall Merging." *Water Resources Research*, 53(11), 8999–9018.

Chiles, J.-P. and Delfiner, P. (2012a). *Geostatistics: Modeling Spatial Uncertainty*. Wiley Series in Probability and Statistics, New York, NY, second edition.

Chiles, J.-P. and Delfiner, P. (2012b). "Kriging." *Geostatistics: Modeling Spatial Uncertainty*, D. J. Balding, N. A. C. Cressie, G. M. Fitzmaurice, H. Goldstein, I. M. Johnstone, G. Molenberghs, D. W. Scott, A. F. M. Smith, R. S. Tsay, and S. Weisberg, eds., Wiley Series in Probability and Statistics, Hoboken, NJ, second edition, Chapter 3, 147–237.

Comerio, M. C. (2014). "Disaster recovery and community renewal: Housing approaches." *Cityscape: A Journal of Policy Development and Research*, 16(2), 51–68.

Corbane, C. and Lemoine, G. (2010). *Collaborative Spatial Assessment - CoSA*, Vol. 1. European Commission Joint Research Centre, Luxembourg (December).

Corbane, C., Lemoine, G., and Kauffmann, M. (2012). "Relationship between the spatial distribution of SMS messages reporting needs and building damage in 2010 Haiti disaster." *Natural Hazards and Earth System Science*, 12(2), 255–265.

Corbane, C., Saito, K., Dell'Oro, L., Bjorgo, E., Gill, S. P., Piard, B. E., Huyck, C. K., Kemper, T., Lemoine, G., Spence, R. J., Shankar, R., Senegas, O., Ghesquiere, F., Lallemant, D., Evans, G. B., Gartley, R. A., Toro, J., Ghosh, S., Svekla, W. D., Adams, B. J., and Eguchi, R. T. (2011). "A comprehensive analysis of building damage in the 12 January 2010 Mw7 Haiti earthquake using high-resolution satellite and aerial imagery." *Photogrammetric Engineering and Remote Sensing*, 77(10), 997–1009.

D'Ayala, D., Faure-Walker, J., Mildon, Z., Lombardi, D., Galasso, C., Pedicone, D., Putrino, V., Purugini, P., De Luca, F., Del Gobbo, G., Lloyd, T., Morgan, E. C., Totaro, A., Alexander, D., and Tagliacozzo, S. (2019). "The MW6.2 Amatrice, Italy Earthquake of 24th August 2016: A Field Report by EEFIT." *Report no.*, Earthquake Engineering Field Investigation Team (EEFIT), London (May).

Dennison, L. and Rana, P. (2017). "Nepal's emerging data revolution background paper." *Report No. April*, Development Initiatives (April).

DesRoches, R., Comerio, M., Eberhard, M., Mooney, W., and Rix, G. J. (2011). "Overview of the 2010 Haiti Earthquake." *Earthquake Spectra*, 27(1_suppl1), 1–21.

Dorati, C., Kucera, J., i Rivero, I. M., and Wania, A. (2018). "Product User Manual for Copernicus EMS Rapid Mapping." *JRC Technical Report JRC111889*, European Commission Joint Research Center (May).

Earle, P. S., Wald, D. J., Jaiswal, K. S., Allen, T. I., Hearne, M. G., Marano, K. D., Hotovec, A. J., and Fee, J. (2009). "Prompt Assessment of Global Earthquakes for Response (PAGER): A System for Rapidly Determining the Impact of Earthquakes Worldwide." *USGS Numbered Series 2009-1131*, U.S. Geological Survey.

Eguchi, R. T., Gill, S. P., Ghosh, S., Svekla, W., Adams, B. J., Evans, G., Toro, J., Saito, K., and Spence, R. (2010). "The January 12, 2010 Haiti Earthquake: A Comprehensive Damage Assessment Using Very High Resolution Areal Imagery." *8th International Workshop on Remote Sensing for Disaster Management*, Tokyo, Japan, Tokyo Institute of Technology, 1–8.

Erdik, M., Sesetyan, K., Demircioglu, M., Zulfikar, C., Hancilar, U., Tuzun, C., and Harmandar, E. (2014). "Rapid Earthquake Loss Assessment After Damaging Earthquakes." *Perspectives on European Earthquake Engineering and Seismology*, A. Ansal, ed., Vol. 34 of *Geotechnical, Geological and Earthquake Engineering*, Springer, 53–96.

Fiorentino, G., Forte, A., Pagano, E., Sabetta, F., Baggio, C., Lavorato, D., Nuti, C., and Santini, S. (2018). "Damage patterns in the town of Amatrice after August 24th 2016 Central Italy earthquakes." *Bulletin of Earthquake Engineering*, 16(3), 1399–1423.

Foster, K. M., Bradley, B. A., McGann, C. R., and Wotherspoon, L. M. (2019). "A VS30 map for New Zealand based on geologic and terrain proxy variables and field measurements." *Earthquake Spectra*, 35(4), 1865–1897.

Gentle, P., Maraseni, T. N., Paudel, D., Dahal, G. R., Kanel, T., and Pathak, B. (2020). "Effectiveness of community forest user groups (CFUGs) in responding to the 2015 earthquakes and COVID-19 in Nepal." *Research in Globalization*, 2, 100025.

Ghosh, S., Huyck, C. K., Greene, M., Gill, S. P., Bevington, J., Svekla, W., DesRoches, R., and Eguchi, R. T. (2011). "Crowdsourcing for Rapid Damage Assessment: The Global Earth Observation Catastrophe Assessment Network (GEO-CAN)." *Earthquake Spectra*, 27(1_suppl1), 179–198.

Government of Nepal Central Bureau of Statistics (2015). "2015 Nepal Earthquake: Open Data Portal.

Government of Nepal National Planning Commission (2015). "Post Disaster Needs Assessment, Nepal Earthquake 2015." *Report No. B*, National Planning Commission, Kathmandu.

Government of the Republic of Haiti (2010). "Haiti Earthquake PDNA: Assessment of damage, losses general and sectoral needs." *Report no.*, Port-au-Prince, Haiti.

Grujic, O. (2017). "Subsurface modeling with functional data." Ph.D. thesis, Stanford University, Stanford University.

Grünthal, G. (1998). "European Macroseismic Scale 1998." *European Center of Geodynamics and . . .*, Vol. 15, 100.

Gunasekera, R., Daniell, J., Pomonis, A., Arias, R. A., Ishizawa, O., and Stone, H. (2018). "Methodology Note on the Global RApid Post-Disaster Damage Estimation (GRADE) Approach." *Report no.*, Global Facility for Disaster Reduction and Recovery, Washington, DC.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., and Gaston, M. (2019). "xBD: A dataset for assessing building damage from satellite imagery." *CVPR Workshop*, Computer Vision Foundation, 10–17.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. J. H. (2009). *The Elements of Statistical Learning*. Data Mining, Inference, and Prediction. Springer, New York, NY, second edition (January).

Hengl, T., Heuvelink, G., and Stein, A. (2003). "Comparison of Kriging with External Drift and Regression Kriging." *Technical Note*, ITC (July).

Hengl, T., Heuvelink, G. B., and Rossiter, D. G. (2007). "About regression-kriging: From equations to case studies." *Computers and Geosciences*, 33(10), 1301–1315.

Hunt, A. and Specht, D. (2019). "Crowdsourced mapping in crisis zones: Collaboration, organisation and impact." *Journal of International Humanitarian Action*, 4(1), 1–11.

Jaiswal, K., Wald, D., and D'Ayala, D. (2011). "Developing Empirical Collapse Fragility Functions for Global Building Types." *Earthquake Spectra*, 27(3), 775–795.

Jaiswal, K. and Wald, D. J. (2011). "Rapid Estimation of the Economic Consequences of Global Earthquakes." *U.S. Geological Survey Open File Report 2011-1116*, U.S. Geological Survey, Reston, VA.

Japan International Cooperation Agency and Ministry of Home Affairs, His Majesty's Government of Nepal (2002). "The study on earthquake disaster mitigation in the Kathmandu Valley, Kingdom of Nepal." *Final Report 1*, Japan International Cooperation Agency.

Kijewski-Correa, T., Roueche, D. B., Mosalam, K. M., Prevatt, D. O., and Robertson, I. (2021). "StEER: A community-centered approach to assessing the performance of the built environment after natural hazard events." *Frontiers in Built Environment*, 7(May), 1–27.

Lallemant, D., Soden, R., Rubinyi, S., Loos, S., Barns, K., and Bhattacharjee, G. (2017). "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(1_suppl), 435–451.

Lee, C. and Tien, I. (2018). "Probabilistic framework for integrating multiple data sources to estimate disaster and failure events and increase situational awareness." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(4), 04018042.

Lemoine, G., Corbane, C., Louvrier, C., and Kauffmann, M. (2013). "Intercomparison and validation of building damage assessments based on post-Haiti 2010 earthquake imagery using multi-source reference data." *Natural Hazards and Earth System Sciences Discussions*, 1(2), 1445–1486.

Liboiron, M. (2015). "Disaster data, data activism: Grassroots responses to representing Superstorm Sandy." *Extreme Weather and Global Media*, J. Leyda and D. Negra, eds., Routledge, New York and London, Chapter 6, 144–162.

Liel, A. B. and Lynch, K. P. (2012). "Vulnerability of reinforced-concrete-frame buildings and their occupants in the 2009 L'Aquila, italy, earthquake." *Natural Hazards Review*, 13(1), 11–23.

Loos, S. (2022). "Sabineloos/GDIF-Gen: Submission Release (January).

Loos, S., Lallemant, D., Baker, J., McCaughey, J., Yun, S.-H., Budhathoki, N., Khan, F., and Singh, R. (2020). "G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage." *Earthquake Spectra*, 36(4), 1695–1718.

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., and Burton, H. V. (2020). "Classifying earthquake damage to buildings using machine learning." *Earthquake Spectra*, 36(1), 183–208.

Martins, L. (2020). "Github - global fragility vulnerability." *Github*, <https://github.com/lmartins88/global$_{f}ragility_{v}ulnerability$>.

Martins, L. and Silva, V. (2020). "Development of a fragility and vulnerability model for global seismic risk analyses." *Bulletin of Earthquake Engineering*, (0123456789).

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S., and Shatar, T. M. (2000). "An overview of pedometric techniques for use in soil survey." *Geoderma*, 97(3), 293–327.

Motaghian, H. R. and Mohammadi, J. (2011). "Spatial estimation of saturated hydraulic conductivity from terrain attributes using regression, kriging, and artificial neural networks." *Pedosphere*, 21(2), 170–177.

MTPTC (2010). "Evaluation des batiments.

Noh, H. Y., Jaiswal, K. S., Engler, D., and Wald, D. J. (2020). "An efficient Bayesian framework for updating PAGER loss estimates." *Earthquake Spectra*, 36(4), 1719–1742.

O'Connor, M. R. (2012). "Two years later, haitian earthquake death toll in dispute." *Columbia Journalism Review*, 1–6.

Potter, S. H., Becker, J. S., Johnston, D. M., and Rossiter, K. P. (2015). "An overview of the impacts of the 2010-2011 Canterbury earthquakes." *International Journal of Disaster Risk Reduction*, 14, 6–14.

Sextos, A., De Risi, R., Pagliaroli, A., Foti, S., Passeri, F., Ausilio, E., Cairo, R., Capatti, M. C., Chiabrando, F., Chiaradonna, A., Dashti, S., De Silva, F., Dezi, F., Durante, M. G., Giallini, S., Lanzo, G., Sica, S., Simonelli, A. L., and Zimmaro, P. (2018). "Local site effects and incremental damage of buildings during the 2016 Central Italy Earthquake sequence." *Earthquake Spectra*, 34(4), 1639–1669.

Sharma, K. (2006). "The political economy of civil war in Nepal." *World Development*, 34(7), 1237–1253.

Sheibani, M. and Ou, G. (2021). "The development of Gaussian process regression for effective regional post-earthquake building damage inference." *Computer-Aided Civil and Infrastructure Engineering*, 36(3), 264–288.

Silva, V. and Horspool, N. (2019). "Combining USGS ShakeMaps and the OpenQuake-engine for damage and loss assessment." *Earthquake Engineering and Structural Dynamics*, 48(6), 634–652.

Stewart, J. P., Zimmaro, P., Lanzo, G., Mazzoni, S., Ausilio, E., Aversa, S., Bozzoni, F., Cairo, R., Capatti, M. C., Castiglia, M., Chiabrando, F., Chiaradonna, A., D'Onofrio, A., Dashti, S., De Risi, R., De Silva, F., Della Pasqua, F., Dezi, F., Di Domenica, A., Di Sarno, L., Durante, M. G., Falcucci, E., Foti, S., Franke, K. W., Galadini, F., Giallini, S., Gori, S., Kayen, R. E., Kishida, T., Lingua, A., Lingwall, B., Mucciacciaro, M., Pagliaroli, A., Passeri, F., Pelekis, P., Pizzi, A., Reimschiissel, B., Santo, A., De Magistris, F. S., Scasserra, G., Sextos, A., Sica, S., Silvestri, F., Simonelli, A. L., Spanò, A., Tommasi, P., and Tropeano, G. (2018). "Reconnaissance of 2016 central Italy earthquake sequence." *Earthquake Spectra*, 34(4), 1547–1555.

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

The European Commission (2017). "How the Copernicus Emergency Management Service supported responses to major earthquakes in Central Italy." *Copernicus Emergemcy Management Service - Mapping*.

Tonkin and Taylor (2016). "Practical implications of increased liquefaction vulnerability." *Report No. 52010.140.v2.0*, Tonkin + Taylor, Auckland, NZ (November).

Trendafiloski, G., Wyss, M., and Rosset, P. (2009). "Loss estimation module in the second generation software QLARM." *Second International Workshop on Disaster Casualties*, number June, Cambridge, UK, 1–10.

United States Geological Survey (2010). "M 7.0 - 10 km SE of Léogâne, Haiti." *Earthquake hazards program*.

United States Geological Survey (2011). "M 6.1 - 6 km SE of Christchurch, New Zealand." *Earthquake hazards program*.

United States Geological Survey (2015). "M 7.8 - 36km E of Khudi, Nepal." *USGS Earthquake Hazards Program*.

United States Geological Survey (2016). "M 6.2 - 5 km WNW of Accumoli, Italy." *Earthquake hazards program*.

Van Ballegooy, S., Malan, P., Lacrosse, V., Jacka, M. E., Cubrinovski, M., Bray, J. D., O'Rourke, T. D., Crawford, S. A., and Cowan, H. (2014). "Assessment of liquefaction-induced land damage for residential Christchurch." *Earthquake Spectra*, 30(1), 31–55.

Westrope, C., Banick, R., and Levine, M. (2014). "Groundtruthing OpenStreetMap Building Damage Assessment." *Procedia Engineering*, 78, 29–39.

Williams, E. J. (1959). *Regression Analysis.* Wiley, New York.

Wilson, B. (2020). "Evaluating the INLA-SPDE approach for Bayesian modeling of earthquake damages from geolocated cluster data.

Worden, C. B., Thompson, E. M., Hearne, M. G., and Wald, D. J. (2016). "ShakeMap Manual Online: Technical Manual, User's Guide, and Software Guide." *Report no.*, U.S. Geological Survey.

Yun, S.-h., Hudnut, K., Owen, S., Webb, F., Sacco, P., Gurrola, E., Manipon, G., Liang, C., Fielding, E., Milillo, P., Hua, H., and Coletta, A. (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), 1549–1556.