Dynamic Volcanic Ash Damage Forecasting for Aotearoa New Zealand

Prepared by: Josh Hayes[1], Niamh Stratton[2], Thomas Wilson[2], Heather Craig[2], James Williams[2], Christina Magill[1], Graham Leonard[1]

Organisation: [1]Earth Sciences New Zealand, [2]University of Canterbury

Email: Josh.hayes@earthsciences.nz

Date: 30 April 2026

With funding support from

Natural Hazards Commission
Toka Tū Ake

Acknowledgements

We thank the Natural Hazard Commission Toka Tū Ake for providing the funding for this project.

Executive Summary

This report examines how to improve rapid post-eruption volcanic ash building damage assessment in Aotearoa New Zealand when decisions are needed before complete observations are available. There were three core components to the project: (1) a review of impact-data reliability that was used to inform a data reliability framework with five criteria (completeness, quality, consistency, timeliness, and assessor certainty); (2) development of a method for updating initial impact modelling estimates with new observed data as it becomes available after a volcanic eruption; and (3) testing of the updating approach with a hypothetical Taupō eruption at pre-defined post-eruption time-steps of: 12 hours, 24 hours, 48 hours, 1 week, and 1 month.

In the experiments conducted in this project, updating was found to improve damage-state estimates in 38 of 45 cases (84%) relative to a no-update baseline. Results also show that observation quality is as important as observation volume. Early in response, high-volume but low-reliability data can temporarily increase estimation error if not appropriately weighted. This also indicates that post-eruption data collection strategies are critical to ensuring that model updates are not overly biased.

The initial model was also compared with empirical building damage data from the 2015 Calbuco volcano eruption, Chile, and the 1994 eruption impacting Rabaul, Papua New Guinea. This indicated that the initial vulnerability model tends to underestimate lower-intensity damage, supporting the need for updates as observations arrive. This also suggests the need to develop New Zealand-specific vulnerability models for volcanic ash induced damage to buildings.

We suggest that Bayesian updating is a viable approach to rapidly calibrate vulnerability models. For practical deployment, we recommend that explicit data quality weighting be used to ensure that unreliable data does not have an undue influence on model estimates. Whilst the approach shows some promise, more work is needed to better refine optimal post-event impact data mixtures that appropriately balance wide coverage with reliable information.

Abstract

Rapid post-eruption impact assessment requires actionable building-damage estimates before complete field data are available. This report develops and tests a practical framework for rapid post-eruption building damage assessment under incomplete and variable data conditions relevant to emergency decision-making in Aotearoa New Zealand. We undertook a review of data-reliability issues in the context of post-disaster impact data collection. This review was then used to develop a data reliability framework that could be used to inform how to weight post-eruption observational data within a Bayesian model updating framework. A Bayesian updating workflow was developed for a common residential building typology in New Zealand (timber-framed buildings), with weighted likelihoods, adaptive prior weighting, and tail regularisation. This was built into a RiskScape model and tested using a hypothetical Taupō eruption using pseudo-observations from credible scenarios of field, aerial/photo, satellite, and crowd-sourced data. Pre-defined time-steps (12hr, 24hr, 48hr, 1 week, 1 month) to evaluate how sensitive the model updating process was to variable pseudo-observational data. Across 45 scenario combinations, Bayesian updating outperformed a no-update baseline in 38 cases (84%), with mean absolute error improvements of about 0.30-0.36 damage-state units where improvements occurred. Results also show that early high-volume, low-reliability data can temporarily degrade estimates, suggesting that careful and strategic data collection is important. We propose that the workflow is suitable as an adaptive damage estimation method for quickly improving impact estimates after volcanic eruptions, so long as data are appropriately weighted and potential sampling biases are avoided.

Keywords

Volcanoes, Quantifying hazards and risk, vulnerability, rapid damage assessment, Bayesian updating, tephra, fragility.

Introduction

In the aftermath of a disaster, developing an understanding of the impacts to society (scale, dimensions, and likely most affected areas) is required to identify priority areas for rescue and relief, and sectors requiring recovery. For example, emergency managers need to know how many people might require welfare provisions or the number of casualties, and infrastructure providers need to know where to send resources to restore disrupted services. Businesses and the insurance industry also rely upon the data from such assessments for activating their own response activities. Detailed and accurate assessments often take months to prepare and require considerable resources to be deployed. However, rapid impact assessments usually need to be conducted within 48 hours of a disaster for emergency management needs. This can be difficult, especially when the impact is widespread across a large area. This is a common occurrence during volcanic eruptions, as volcanic ash can be deposited hundreds of kilometres away from a volcano and still cause considerable disruption to socio-economic systems (Wilson et al. 2012). A recent survey of emergency management and risk science stakeholders in Aotearoa-New Zealand highlighted real-time and rapid impact assessment methodologies as an important and necessary development for Aotearoa-New Zealand's disaster response and recovery needs (Thomas et al. 2020).

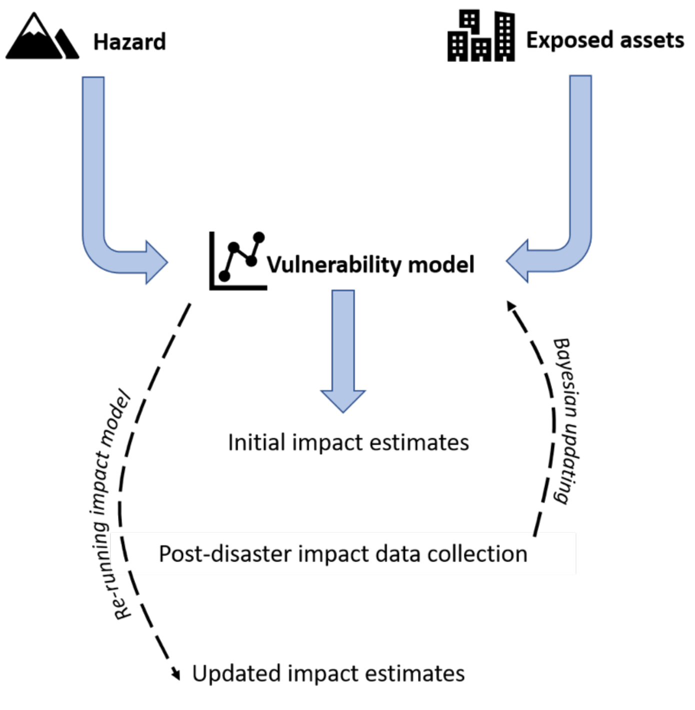

Figure 1: Impact assessment process. Generic approach indicated by blue arrows. Bayesian updating indicated by dashed arrows.

Disaster impact assessments are commonly used to estimate the potential consequences of natural hazards. Impact assessments can be undertaken pre-event for planning purposes (Zuccaro et al. 2013; Scaini et al. 2014; Deligne et al. 2017; Biass et al. 2017). Post-event impact assessments focus on characterising the actual impacts that have occurred from a specific event or events (Baxter et al. 2005; Wilson et al. 2011; Magill et al. 2013; Blake et al. 2015). Pre-disaster impact assessments often rely on modelling approaches that combine three components: hazard, exposure, and vulnerability (Figure 1). Hazard characterises the footprint and spatial representation of hazard intensity (e.g. ground acceleration, volcanic ash thickness, flood depth). Exposure characterises the assets that might be affected by the hazard. Vulnerability characterises the specific effects (e.g. damage or disruption severity) an asset or assets are subjected to given exposure to a hazard. Fundamental to this are the statistical relationships between a given hazard intensity and the corresponding impacts that result. Such relationships are usually defined using fragility (probability of a level of damage being exceeded) or vulnerability (probability of a number of losses) functions, which are developed using expert elicitation, empirical relationships derived from laboratory or field studies, and theoretical/analytical studies (henceforth referred to as vulnerability models). Post-disaster impact assessments often require a combination of intelligence from a variety of sources to obtain a holistic view of the impacts that have occurred. It can take weeks to months in some situations to obtain some impact information. However, often decisions need to be made rapidly in the immediate aftermath of a disaster, which means there may be a reliance on poorly constrained information.

Rapid impact modelling approaches have been used to fill the information gap early in a disaster response, when information is limited. Rapid and updatable impact methodologies have been developed and applied for earthquakes and tropical cyclones (Gunasekera et al. 2018; de Bruijn et al. 2022). Rapid impact modelling uses modelling approaches similar to those used for pre-event planning to build a picture of the impacts that have likely occurred. To successfully implement rapid impact modelling methods it is necessary that robust hazard, vulnerability, and exposure data is able to be leveraged (Weir et al. 2024). Both earthquakes and tropical cyclones have a well-developed and comprehensive suite of vulnerability models to draw upon (Pita et al. 2013, 2015; Yepes-Estrada et al. 2016; Silva et al. 2019; Wilson et al. 2022). Comparatively, for volcanic hazards these are still in their infancy (Deligne et al. 2022; Fitzgerald et al. 2023; Hayes et al. 2024). For example, globally, there are currently only five published datasets on post-eruption damage from volcanic ash fall (Spence et al. 1996; Blong 2003; Hayes et al. 2019; Jenkins et al. 2024; Tennant et al. 2024). Of these, the examples from Rabaul, Papua New Guinea, 1994 (Blong 2003) and Calbuco, Chile, 2015 (Hayes et al. 2019) are the closest to the typologies seen in New Zealand residential building stock. This limited empirical evidence presents a challenge when implementing vulnerability models within rapid impact assessments, especially when local contextual vulnerability models are lacking (Weir et al. 2024).

Bayesian updating is a way of improving estimates by combining prior knowledge with new evidence to produce a more accurate estimation. It provides a formal quantitative framework for representing uncertainty and updating model parameters as new information becomes available, and weighting incoming observations according to their quantity and reliability. This makes it particularly well suited to situations where data are sparse, noisy, or incrementally acquired, as is typical during an unfolding volcanic eruption response. Estimates can be progressively refined as additional observations are collected, without requiring complete or fully consistent datasets. Therefore, using Bayesian updating to improve rapid building damage and loss estimates following a volcanic eruption may be a useful method to close the gap between uncalibrated globally generalised vulnerability models and sparsely observed local impact data.

The objectives of this work are to:

- Review issues associated with post-disaster impact data reliability and develop a framework to effectively evaluate data reliability for impact assessment purposes.
- Develop a methodology that can produce rapid impact estimates following an eruption that are dynamically updated as and when new data becomes available.
- Test the methodology using the Calbuco 2015 building damage dataset from Hayes et al. (2019).
- Apply the framework to a hypothetical New Zealand eruption scenario to evaluate how this could be applied in a New Zealand context.

Methodology

The conceptual approach to achieving the objectives listed above is outlined in Figure 2. In the below subsections, we detail how each of these components was undertaken.

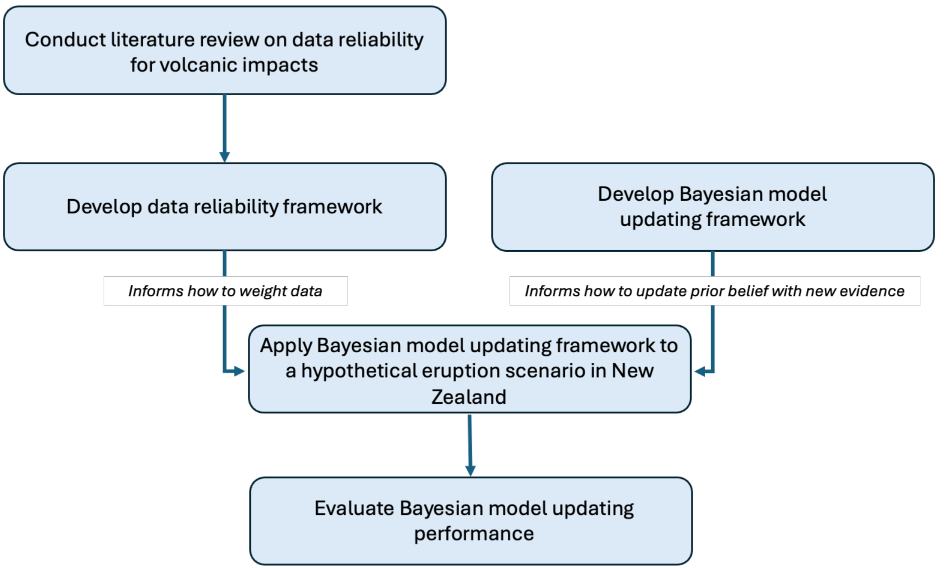

Figure 2: Conceptual workflow used in this report to achieve the projects aims

Reviewing post-eruption impact data reliability and developing a data reliability framework

Stratton (in prep) undertook a detailed literature review and development of the reliability framework. Here, we briefly summarise the approach, but suggest reviewing Stratton (in prep) for detailed methodological evaluation. A review of the literature addressing data reliability was undertaken. Two foundational studies (Hayes et al. 2019; Meredith et al. 2022) on post-event volcanic impact data reliability were used as anchor papers, from which snowball forward and backward citation tracking was employed to identify and review relevant secondary literature. A database search was also employed to identify potentially relevant literature from fields adjacent to volcanic impacts that may have useful information on data reliability assessment and evaluation (e.g., emergency health). Google Scholar was used with two different search terms: “Disaster impact data reliability assessments” and “Assessment of data quality”. Studies were then filtered further as to whether they focussed evaluating data reliability. From the identified studies, data reliability attributes and definitions were extracted, thematically grouped, and synthesised to inform the design and structure of an idealised impact data reliability assessment framework.

Development of the reliability framework then went through an iterative design process to consider how the framework would work in practice. This was done through workshops amongst the report authors to discuss potential challenges with implementing it. Once a framework was settled on, a comparison between the estimated data reliability from the proposed framework and the data quality indicators published in Hayes et al. (2019) were undertaken as a final sense check on the framework to ensure it did not produce unreasonable results.

Bayesian model updating within an impact assessment framework

The overall workflow couples prior vulnerability modelling, observational updating, and loss estimation within a single framework (Figure 2). An existing vulnerability model is first used to define a prior relationship between ash loading and expected damage/loss. Post-event observations are then used to update this relationship, producing an event-informed vulnerability model that reflects both prior knowledge and empirical evidence.

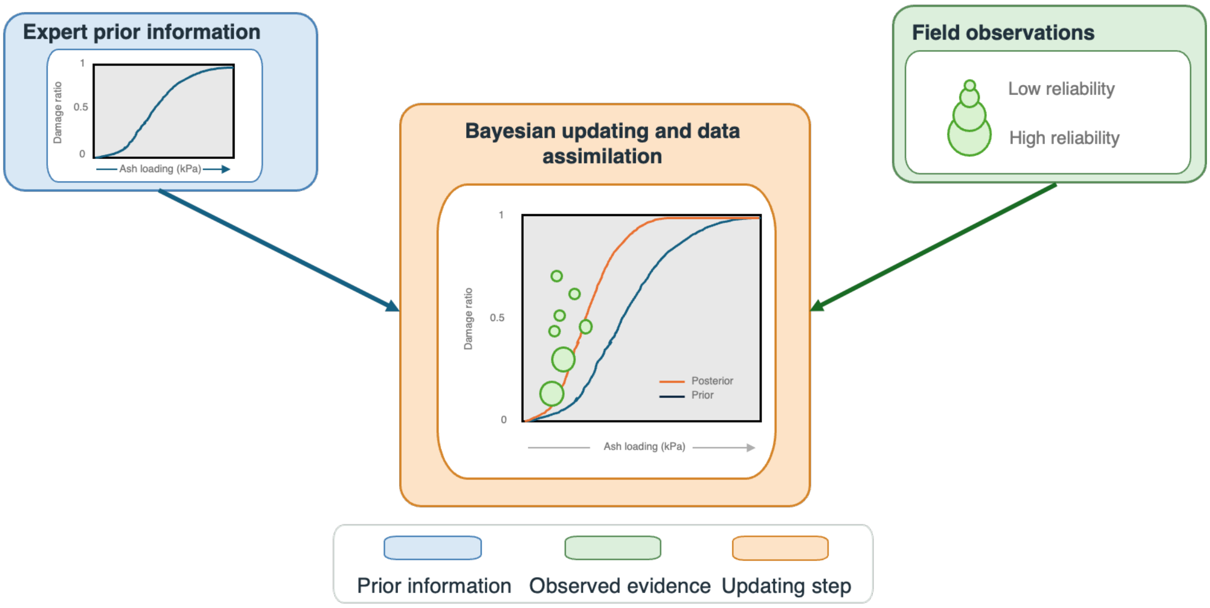

Figure 3: Conceptual diagram of the vulnerability model updating workflow. Expert vulnerability information defines a prior vulnerability curve, while event-specific field observations are used as evidence for updating. These two information sources are combined within a Bayesian updating step to produce a posterior vulnerability relationship that reflects both prior knowledge and observed event impacts.

The prior vulnerability model is a lognormal cumulative damage curve derived from expert elicitation during the development of the 2015 Global Assessment of Risk (GAR) (Maqsood et al. 2014). The selected volcanic ash vulnerability function is for building typology "W1", which is described as "Wood, Light Frame 1-2 stories". This prior expresses the expected relationship between ash loading and damage ratio before any event-specific observations are introduced. When observational data are available, the model updates the two lognormal parameters, mu and sigma, using Bayesian inference. Here, mu controls the horizontal position of the curve along the loading axis and sigma controls the spread of the transition from low to high damage (Eq. 1 and Fig 1).

$$DR(IM|\mu, \sigma) = \phi\left(\frac{\ln(IM) - \mu}{\sigma}\right) \tag{1}$$

Where $\phi(.)$ is the standard normal cumulative distribution, DR is the expected damage ratio, IM is the hazard intensity measure (ash load in kPa), μ and σ are parameters controlling the location and dispersion of the damage function in log-intensity space.

Since observational damage data are generally reported as damage states or descriptions of damage, not damage ratios, it is necessary to make some conversions between damage ratios and damage states (DS), which is more likely to be the form of damage observations. Each observed damage state is converted into a plausible continuous damage-ratio value by sampling uniformly within a prescribed interval for that state:

- 0 for state DS0,

- 0.01-0.05 for DS1,
- 0.05-0.20 for DS2,
- 0.20-0.50 for DS3,
- 0.50-0.90 for DS4, and
- 0.90-1.00 for DS5.

For each update run, DR values are sampled from these intervals and compared to model-predicted DR values at corresponding ash loads through a weighted likelihood, so higher-confidence observations have greater influence on posterior parameter estimation. For Calbuco, where ash thickness is often reported as ranges (for example, 20-30 cm), intensity uncertainty is also propagated by sampling within each reported thickness band during update runs. The weighted likelihood contribution for observation is:

$$\mathcal{L}_i(\mu, \sigma) \propto [DR(IM_i|\mu, \sigma)]^{\omega_i dr_i}[1 - DR(IM_i|\mu, \sigma)]^{\omega_i(1-dr_i)} \quad (2)$$

Where dr_i is the sampled damage ratio for observation i, IM_i is the observed intensity measure (ash load kPa), and ω_i is a weight reflecting the confidence in the observation.

The posterior combines three terms. First, a prior term favours parameter values close to the expert model, scaled by standard deviation of the mean prior curve. Second, a likelihood term rewards fit to observations. Third, a tail-regularisation term penalises strong divergence from the prior outside the observed intensity range to reduce unstable extrapolation. Conceptually, this balances responsiveness (learning from data) against stability (avoiding physically implausible curve shifts when observations are sparse or concentrated in a narrow intensity window). In log form:

$$\log p\,(\mu, \sigma|D) \propto \alpha \log p_{prior}(\mu, \sigma) + \sum_i \log \mathcal{L}_i(\mu, \sigma) + \lambda\; R_{tail}(\mu, \sigma) \quad (3)$$

Where D is the set of observations, α is a prior weight, λ controls the strength of the tail regularisation, and $R_{tail}(\mu, \sigma)$ is a penalty term that increases with divergence from the prior damage function outside the range of observed intensities.

Prior influence is adjusted according to evidence volume and how much the observations cover the full ash loading spectrum. Prior weight is stronger when effective observation weight is low or when observations span only a narrow portion of intensity space, and weaker when observations are numerous and span more bins. This makes updates more conservative when data are sparse or poorly distributed. The adaptive prior weight used in this work is:

$$\alpha = max\left(0.1, \frac{10}{\sqrt{W} + 1}\sqrt{1 - c}\right) \quad (4)$$

Where W is the total effective observation weight and C is the fraction of intensity bins covered by at least one observation.

Posterior sampling uses an ensemble Markov chain Monte Carlo algorithm implemented through the emcee python package (Foreman-Mackey et al. 2013). The sampler used 50 walkers, each run for 3,000 steps. The first 500 steps per walker were discarded as burn-in, and every tenth subsequent sample was retained, producing 12,500 posterior samples. The sampler used 50 walkers, each run for 3,000 steps. The first 500 steps per walker were discarded as burn-in, and every tenth subsequent sample was retained, producing 12,500 posterior samples. Sampling diagnostics were calculated from the complete, unthinned chains. Across the five primary time-stepped updates, mean acceptance fractions ranged from 0.707 to 0.712, with individual-walker values ranging from 0.681 to 0.736. Parameter-specific integrated autocorrelation times ranged from 25.1 to 29.6 steps. The retained chain length of 2,500 steps per walker therefore represented at least 84.5 integrated autocorrelation times for every parameter, exceeding the approximately 50-autocorrelation-time criterion recommended in the emcee guidance. Approximate effective sample sizes ranged from 4,225 to 4,986 per parameter, indicating that the sampling duration was adequate. Final updated parameter estimates were calculated as the means of the retained posterior samples. This process estimated a distribution of plausible vulnerability curves, rather than a single deterministic curve, allowing uncertainty bounds to be reported.

Exact implementation details affecting numerical replication, including random seeds, chain settings, DS-to-DR conversion, prior-weighting rules, regularisation logic, and construction-type assignments, are documented in the associated code (see Outputs section).

Modelling a hypothetic rapid data collection exercise for a New Zealand eruption scenario

The workflow has five key steps: (1) generate ash intensity at building locations; (2) assign a pseudo-true building damage state; (3) simulate time-stepped, source-specific observations with realistic operational constraints and classification errors; (4) update prior vulnerability parameters using weighted Bayesian inference as observations accumulate; and (5) evaluate updated damage estimates against the pseudo-true reference. This design allows direct testing of how observation timing, quality, and spatial coverage affect forecast performance through time. The RiskScape Engine (v1.12.0) was used to run the impact modelling (Paulik et al. 2023). The RiskScape model: (1) takes a hazard raster file (GeoTiff – see next section on for how the scenario was constructed) and a building points layer (GeoPackage from Scheele et al. (2023)); (2) filters the building layer for only points that overlap with the hazard raster; (3) samples ash intensity at each building location using centroid sampling; (4) a risk function is then used to assess the damage given the buildings ash hazard value. Below we detail specifics of how the ash hazard and building damage scenario was developed and how the Bayesian updating process works.

Ashfall deposition and building damage scenario

The Bayesian updating model was applied to a hypothetical eruption scenario from Taupō volcano (Figure 3). The eruption scenario is taken from a set of ashfall simulations developed by Barker et al. (2019). The select eruption scenario had a volume (Dense Rock Equivalent) of 1km^3 (~Volcanic Explosivity Index (VEI) 5), which is considered to be a moderate size eruption for Taupō. An eruption of this size is estimated to have an annual probability of ~0.03% (Stirling and Wilson 2002). To ensure we sufficiently stress test our updating approach, we selected a simulation (simulation 00190 in Barker et al. (2019) that contained a wind field that would result in a substantial amount of the ashfall to deposit on the North Island, rather than taken offshore.

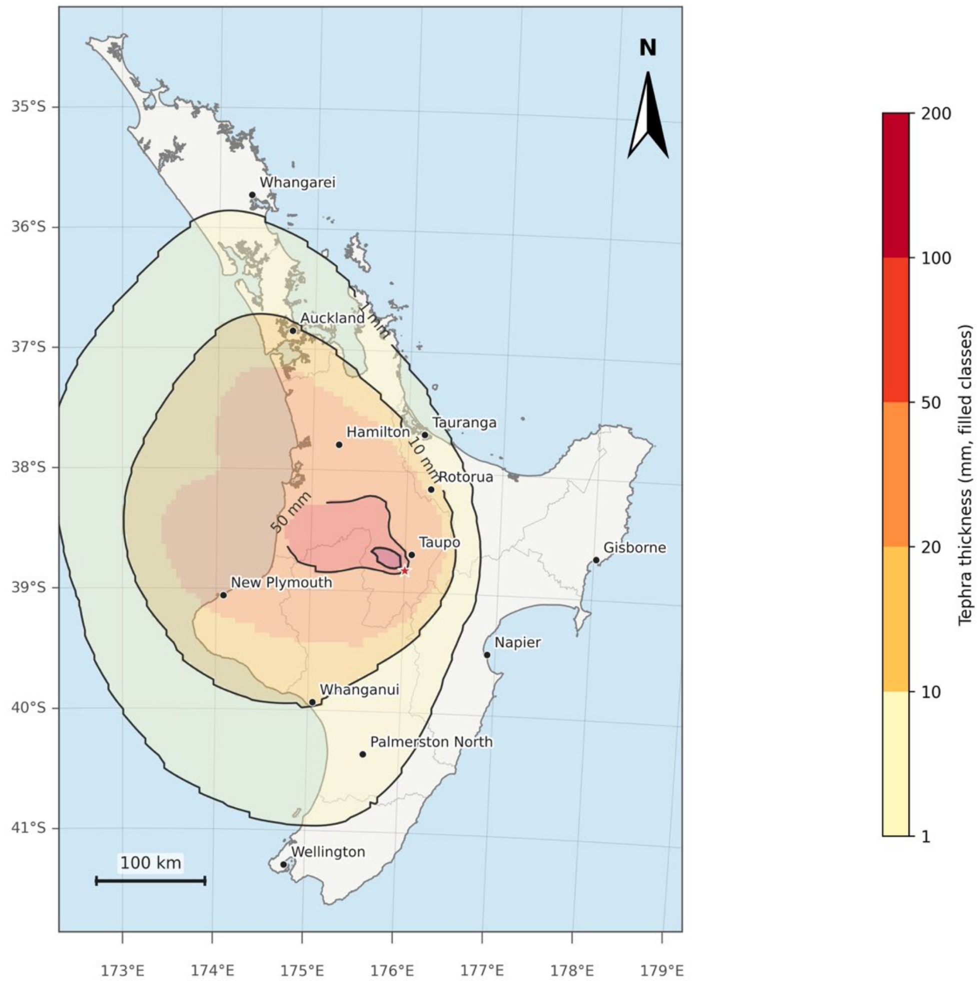

Figure 4: Spatial distribution of modelled ashfall from the selected Taupō eruption scenario.

Developing pseudo damage and pseudo-observational datasets

New Zealand has not had a major ash-producing eruption in over 30 years, since the 1995-96 Ruapehu eruptions. This eruption was relatively minor and there was relatively minimal damage caused to buildings. As there is a lack of local experience to use to evaluate the Bayesian updating model with, we need to construct pseudo-building damage and damage observation datasets. Our intention with these is merely as a credible example of how such an assessment could occur should an eruption damage assessment need to occur in the near term and evaluate if and how Bayesian updating could be of use. Below we describe how we constructed pseudo building damage (i.e., hypothetical actual building damage) and pseudo-observation (i.e., the different data sources likely to be used to inform Bayesian updating) datasets.

To generate the pseudo-true reference dataset for the Taupō scenario, estimated conditional probabilities of building damage state given ash-thickness band, from the

Calbuco timber-framed building analogue dataset and then used these probabilities to stochastically assign “true” damage states at each exposed building. For each thickness band, probabilities were computed over DS0–DS5 as weighted relative frequencies:

$$\widehat{p_{d,b}} = \frac{\sum_{i\in b} w_i 1(DS_i = d)}{\sum_{i\in b} w_i} \tag{5}$$

Where w_i is a weighted score for the data point based on the data quality indicators presented in Hayes et al. (2019).

This weighted form was used for the default pseudo-true table so that higher-confidence analogue records had greater influence than low-confidence records. Exceedance probabilities were constrained to be monotonic with increasing ash thickness then converted into damage state probabilities, which were sampled for the Taupō case study (Table 1). This is because vulnerability severity should not decrease with increasing load, and so applying this smoothing removes non-physical crossovers caused by sampling noise while preserving the empirical relationship as closely as possible.

Table 1: Calbuco-derived conditional damage-state probabilities used to generate the pseudo-true Taupō building damage state dataset

Thickness band (cm)	DS0	DS1	DS2	DS3	DS4	DS5
0	1.0	0	0	0	0	0
0.1-3	0.260	0.288	0.205	0.247	0	0
3-10	0.059	0.260	0.280	0.401	0	0
10-20	0.059	0.260	0.171	0.360	0.1	0.051
20-60	0.038	0.152	0.300	0.225	0.152	0.133

With pseudo-true reference dataset established, the next step is to represent what information would plausibly be available at different time steps throughout the eruption response and the potential errors associated with it. To illustrate this, “pseudo-observations” were generated from four post-eruption data streams: field surveys, crowd-sourced reports, aerial imagery (e.g., helicopters/drones)/ ground photos, and satellite data. These streams differ by temporal availability, spatial coverage, and reliability. Pseudo-observations were generated at five fixed timesteps (12 hours, 24 hours, 48 hours, 1 week, and 1 month), accumulated through time, and allowed to transition in source type for the same building (for example, aerial early and field later). Source-specific error models were applied to perturb observed DS values from pseudo-true DS values (Table 2). The rates of data collection at the property level for each data source was also estimated

(Table 3). Field surveys were assumed to provide the most reliable but slowest observations, crowd-sourced reports the fastest but noisiest observations, and aerial or satellite interpretation an intermediate source with broad coverage but more limited interpretability of damage.

Table 2: Synthetic observations error generation model. Sequence of numbers is: Correct exact damage state, Minor error (± 1 damage state), major error rate (± 2 damage state).

Scenario component	Field	Aerial/photos	Satellite	Crowd-sourced
Low error	95%, 5%, 0%	85%, 13%, 2%	75%, 21%, 4%	62%, 30%, 8%
Moderate error	90%, 9%, 1%	78%, 19%, 3%	67%, 27%, 6%	52%, 36%, 12%
High error	84%, 14%, 2%	68%, 26%, 6%	58%, 31%, 11%	42%, 40%, 18%

Field-survey observations were generated at three fixed post-eruption timesteps (48 hours, 1 week, 1 month). It was assumed that no surveys would be conducted within the first 24 hours due to both safety and logistics. Candidate buildings were restricted to locations outside of a hypothetical exclusion zone around the vent (20 km through 1 week; 10 km at 1 month). We emphasize this exclusion zone is not intended as a recommended zone for future eruptions at Taupō, nor is it formal policy. We use it in this instance to restrict our simulated pseudo-field surveys based on similar exclusion zones implement at overseas eruptions and it is used here to acknowledge that it is unlikely that engineering level damage surveys will be conducted within areas of potential life safety risk. Within the allowable zones, sampled surveys are biased towards assessing buildings with true damage states (DS2–DS4). The reason for this is that we anticipate that the initial focus of damage surveys will be to determine whether buildings are safe for habitation. At each timestep, a fixed number of hub locations was sampled from eligible buildings with probability weighted by local ash thickness, and each hub defined a local survey footprint using a timestep-specific radius (6 km at 48 h, 10 km at 1 week, 14 km at 1 month) (Figure 4). Buildings within these footprints were then sampled to match timestep observation capacities:

- 12 hours: Satellite = 1500, crowd = 400, aerial/photos = 0, field = 0
- 24 hours: Satellite = 5000, crowd = 1500, aerial/photos = 800, field = 0
- 48 hours: Satellite = 15,000, crowd = 4000, aerial/photos = 2500, field = 10
- 1 week: Satellite = 40,000, crowd = 10,000, aerial/photos = 7000, field = 50
- 1 month: Satellite = 70,000, crowd = 12,000, aerial/photos = 10,000, field = 120

Table 3: Synthetic observations collection rate model used to multiply by the base values above. Sequence of numbers is: field, aerial/photo, satellite, crowd-sourced.

Scenario component	12 hours	24 hours	48 hours	1 week	1 month
Crowd heavy early	0, 0.7, 1.2, 1.8	0, 0.7, 1.2, 1.8	0.8, 1, 1.1, 1.4	0.8, 1, 1.1, 1.4	1.2, 1, 0.9, 0.9
Field-surge	0, 0.8, 1.1, 1.3	0, 0.8, 1.1, 1.3	1.3, 1, 1, 1	1.3, 1, 1, 1	1.8, 1.1, 0.9, 0.8

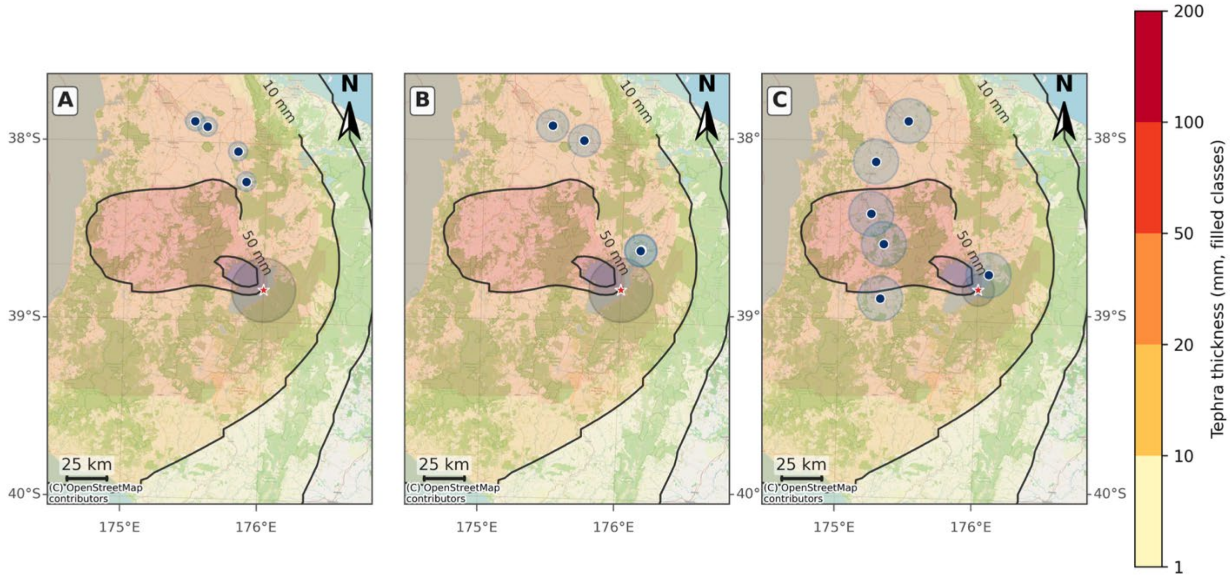

Figure 5: Simulated field-survey deployment through time for the Taupō eruption scenario (48 h (A), 1 week (B), 1 month (C)). Blue points show selected field-survey hubs and blue buffered polygons show the corresponding operational survey radius around each hub. Grey filled circles indicate the exclusion zone applied at each timestep (20 km at 48 h and 1 week; 10 km at 1 month).

Weighting observational data for Bayesian updating

The weighting of pseudo-observational data was designed to reflect the ordinal structure of the reliability framework (presented later in this report), in which confidence increases substantially as data become more complete, better quality, and more certain in terms of damage-state assignment (Table 4). Across the framework, movement from low reliability to high is not a small incremental improvement. Rather, each step represents a material increase in evidential value: from sparse or ambiguous information with low confidence in the assigned damage state, through partially constrained and moderately reliable evidence, to comprehensive, internally consistent, site-specific observations with high assessor certainty. For that reason, the Bayesian updating weights were chosen to increase by orders of magnitude rather than linearly.

Under this logic, crowd-sourced data were assigned a weight of 0.1 because they are most likely to fall within the lower reliability grades. Such data are often incomplete, variably located, inconsistently described, and prone to uncertainty, spatial and thematic bias, or misclassification (Zhang and Zhu 2018; Xing et al. 2021; Ba et al. 2025). However, they may still provide useful information, particularly for identifying whether damage is present at all, but the framework indicates low to moderate confidence in the correctness of the damage-state assignment. Assigning a weight below 1 ensures that these observations can still contribute to updating, but only weakly, so they do not unduly distort posterior estimates when more reliable evidence is available.

Satellite data were assigned a weight of 1. This is because satellite observations commonly provide a more systematic and objective basis than crowd-sourced material, but still often have important limitations relative to the higher reliability grades. They may offer good spatial coverage and consistency, but damage detail, site visibility, and building type characterisation are often constrained (Williams et al. 2020; Biass et al. 2021). In the framework, this aligns broadly with the middle reliability grades, depending on resolution, acquisition timing, and interpretability. A weight of 1 therefore represents sufficiently credible information to directly influence the posterior.

Aerial imagery and photo-based assessments were assigned a weight of 10. Compared with satellite or crowd-sourced data, aerial and close-range photographic observations usually provide better site visibility (e.g., more angles to see varying damage types), stronger spatial confidence (geolocated) resulting in fewer ambiguities in interpretation. The Spence et al. (1996) damage assessment from the Pinatubo 1991 eruption was entirely based on photographs. Drone images were also heavily used by Tennant et al. (2024) with good success. A weight of 10 reflects that these data should have substantially greater influence than baseline observations, but still remain distinct from the strongest evidence class.

Field surveys were assigned the highest weight of 100. The framework defines the highest reliability data as requiring clear information on hazard, assets, and impact descriptions, which would come from engineering level field survey data. Particularly where trained assessors undertake direct inspection and use correct engineering terminology to assign damage states. A weight of 100 therefore reflects that field observations should dominate the updating process when available, because they represent the most credible approximation to ground truth within the evidence hierarchy.

Table 4: Weights used in the Bayesian updating for different data sources

Data source	Weight
Crowd-sourced	0.1
Satellite	1
Aerial imagery	10
Field surveys	100
20-60	0.038

Performance metrics and interpretation framework

To evaluate whether updating improves decision-relevant forecasting performance, model outputs were assessed against pseudo-true damage states using multiple complementary metrics. Model performance was evaluated against the pseudo-true damage state using mean absolute error (MAE) because it is directly interpretable in damage-state units (Eq. 6).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{D}S_i - DS_i^{true}\right| \quad (6)$$

Where $\widehat{D}S_i$ is the predicted damage state, DS_i^{true} is the pseudo-true damage state for building i, and N is the number of buildings.

To contextualise the best achievable performance under the assumed observation-error structure, a scenario-specific MAE_{floor} was estimated. MAE_{floor} represents the lowest expected MAE that could be achieved if inference were otherwise ideal, given the assumed error characteristics and the intensity-space covered by observations. It should therefore be interpreted as a conditional lower bound under the experimental assumptions. This is expressed as:

$$MAE_{floor} \approx \frac{1}{B}\sum_{b=1}^{B}\left[I_b m_b + \ (1 - I_b)m_{base}\right] \quad (7)$$

Where b indexes intensity bins, $I_b \in \{0,1\}$ indicates whether bin b is observed, m_b is the minimum achievable error given observational uncertainty, m_{base} is the baseline error for uncovered bins, and B is the total number of bins.

The MAE_{floor} is used as a diagnostic benchmark. Its role is to distinguish two situations: (1) poor performance because model assumptions or source mix are weak, versus (2) performance near the practical limit implied by noisy and incomplete observations. This distinction is important for interpreting whether additional data collection is likely to materially improve assessment quality and for identifying when model refinement, rather than more of the same data, is likely to produce the next major improvement.

Results and discussion

Impact data reliability from the literature

Here we summarise the key findings from the literature review from Stratton (in prep), which found that data reliability is multi-dimensional comprising of:

- Data quality: The degree to which the data accurately represents reality. This includes attributes such as accuracy, precision, validity, and internal consistency, as well as the absence of discrepancies or conflicting information. High-quality data provide a faithful and credible representation of impacts and are typically supported by expert validation or comparison against known patterns or models.
- Completeness: The extent to which a dataset captures the full scope of relevant information required for assessment. This includes both the breadth (coverage across affected areas or systems) and depth (level of detail) of the data, as well as the representativeness of the sample. Incomplete datasets may omit critical impacts or introduce bias, even if the available data are internally consistent.
- Temporal relevance. How current and temporally appropriate the data are for the assessment context. In post-disaster settings, where conditions evolve rapidly, delays between data collection and analysis can significantly reduce reliability. This dimension also includes periodicity, or how frequently data are updated, which is critical for tracking dynamic impacts and recovery processes.
- Accessibility and usability: How easily data can be obtained, interpreted, and applied. This includes availability, readability, interpretability, and understandability, as well as practical constraints such as access restrictions or data formats. Even high-quality data may be of limited value if they are difficult to access or require specialised knowledge to interpret.
- Governance: The trustworthiness of the data source and the processes used to generate and manage the data. This includes data source reputation, integrity, adherence to standards, and objectivity, as well as the extent to which data have been subject to verification or quality control. Reliable datasets are typically produced by credible institutions and follow established data management practices.

- Uncertainty and error encompass the presence and treatment of inaccuracies, biases, and limitations within the data. This includes measurement error, omission, commission, representativeness, and repeatability, as well as the transparency with which these uncertainties are communicated. Explicit consideration of uncertainty is critical in disaster contexts, where incomplete or rapidly collected data are common and decision-making must account for potential error.

Each dimension comprises a set of underlying attributes identified from the literature. Data quality includes accuracy, precision, validity, and internal consistency. Completeness encompasses dataset coverage, level of detail, and representativeness. Temporal relevance includes timeliness and periodicity of data collection. Accessibility and usability include availability, readability, interpretability, and understandability. Governance captures data source credibility, integrity, adherence to standards, and objectivity. Uncertainty and error include measurement error, omission, commission, and representativeness of the data sample. Table 5 applies these synthesised categories to the reviewed studies, showing both the presence of each dimension and the proportion of underlying attributes considered, enabling comparison of the breadth of data reliability across studies.

Our review found that Post-Disaster Impact Assessment (PDIA) studies tend to emphasise dataset completeness (e.g., volume, coverage, and representation) while giving comparatively limited attention to other dimensions such as accessibility, governance, and uncertainty. In contrast, other fields (e.g., public health, data science) adopt a more holistic perspective, routinely incorporating considerations of accuracy (precision, validity), consistency, accessibility, and source credibility. This suggests it is important when using impact observational data to interrogate whether those data are correct, interpretable, or trustworthy. If not, there is a risk that quantitative assessments provide a potentially misleading representation of impacts.

The attribute framework further reinforces that reliable post-disaster data must extend beyond completeness to include fitness for purpose, temporal relevance, and transparency of uncertainty. Key dimensions such as timeliness are particularly critical in disaster contexts, where delays in data acquisition or processing can rapidly degrade reliability as conditions evolve. Similarly, data source credibility, objectivity, and error management (including omission, bias, and representativeness) are essential for ensuring that derived impact estimates reflect reality rather than artefacts of data limitations or processing choices. The findings point to a clear need for a unified, impact-centric data reliability framework that explicitly integrates completeness, accuracy, accessibility, governance, and uncertainty. Such a framework would support more robust and transparent post-disaster impact assessments by ensuring that data are not only available, but also credible, interpretable, and appropriate for decision-making under uncertainty.

The impact data reliability framework developed in this study is designed as a categorical, site-based framework that integrates the dimensions of data reliability identified in the literature. These are translated into five practical assessment criteria: completeness, quality, consistency, timeliness, and assessor certainty. This structure allows complex and overlapping reliability concepts to be simplified into a form that can be consistently applied in post-disaster contexts. In particular, consistency captures aspects of both data quality and internal coherence, timeliness directly reflects temporal relevance, and assessor certainty acts as a catch-all criterion to account for governance, accessibility, and residual uncertainty that are difficult to measure explicitly at the site level.

The assessment is applied at the individual building scale, where each is evaluated independently based on the available impact data. Each site is then scored against the five criteria using a graded scale, reflecting how well the available data meets the defined reliability requirements. The grades range from RG0 (no reliable data) to RG4 (highly reliable data), providing a clear and interpretable hierarchy. A minimum threshold is also introduced to exclude very low-reliability data (RG0–RG1) from further analysis, recognising that poor-quality data can undermine subsequent applications such as vulnerability modelling. The framework allows flexibility in how a final overall grade is estimated when there are multiple datapoints for the same site (e.g., averaging or minimum grade).

This framework provides a practical implementation of the multi-dimensional concept of data reliability, bridging the gap between theoretical definitions and applied post-disaster impact assessment. By consolidating a wide range of underlying attributes into a set of criteria, the assessment enables consistent, transparent, and context appropriate evaluation of impact data. This serves as the foundation for considering how to weight impact data for the Bayesian model updating process.

Table 5: Data reliability dimensions and proportion of attributes considered across reviewed studies. Numbers and colours represent the proportion of underpinning attributes for the reliability dimensions that were considered in the study. Coloured by quantile. PDIA = Post Disaster Impact Assessment, DS = Data science, EM/PH = Emergency Health / Public Health. Adapted from Stratton (in prep).

Study	Field of study	Data quality	Completeness	Temporal	Accessibility & usability	Governance & trust	Errors & uncertainty
Hayes et al. (2019)	PDIA	0	1	0.5	0	0.17	0
Meredith et al. (2022)	PDIA	0	0.5	1	0	0.17	0
Tennant et al. (2024)	PDIA	0.8	0.75	1	1	0	0.75
Mashoufi et al. (2018)	EM/PH	0.4	0.75	1	1	0.83	0.5
Cai & Zhu (2015)	DS	0.8	0.75	1	1	0.83	0.5
Chen et al. (2014)	EM/PH	1	1	1	1	0.67	0.75
Pipino et al. (2002)	DS	0.8	0.75	0.5	0.8	0.5	0.5
Romão & Paupério (2016)	PDIA	0.4	0.75	1	0.4	0.33	1
Hao & Wang (2021)	PDIA	0.2	1	0.5	0.4	0.17	0
Eckhardt et al. (2019)	PDIA	0.2	0.75	0.5	0.4	0.33	0
Corbane et al. (2015)	PDIA	0.8	0.75	0.5	0.8	0.67	0.25
Moriyama et al. (2018)	PDIA	0.4	1	0.5	0.4	0.17	0.5
Ladds et al. (2017)	PDIA	0.4	0.5	1	0.4	0.33	0.5
Chen & Lim (2021)	PDIA	0.2	0.25	0.5	0.4	0.33	0.25
Xing et al. (2021)	PDIA	0.2	0.5	0.5	0.4	0.17	0
Calantropio et al. (2021)	PDIA	0.2	0.5	0.5	0.4	0.33	0
Asadzadeh et al. (2017)	PDIA	0.2	0.25	0.5	0.2	0.5	0.25

Reliability Grade	Criteria				
	Completeness	Quality	Consistency	Time	Assessor Certainty
RG0	None across all sources	None or no level of detail	Not applicable Or No consistency	No date/time information	No confidence that DS is correct (particularly DS2-4). Extensive signs of uncertainty, bias, misinformation, modification, contamination, and lack of credibility.
RG1	Data presents no useful information OR/AND Scarce data across all available event data sources.	**H:** No measurement, not visible or unrelated to site. **A:** Limited detail – inferred typology from geographic context, regional-level spatial location, site may not be visible. Potential incorrect site in dataset – no spatial reference. **I:** Minimal details – theoretical nature and estimated extent from models. No site-specific details, at best adjacent site data.	Highly conflicting data Significant inconsistency No rational explanation	> 2 Years	Low confidence that DS is correct. Numerous signs of uncertainty, bias, misinformation, modification, contamination, and lack of credibility Low confidence in RG value.
RG2	Data only presents information on one of the three components OR/AND Limited data across all available event data sources	**H:** No scientific measurement, however, a measure is inferred from the data by a secondary source (assessor). Extent inferred across site. **A:** Moderate detail – basic typology, general spatial location, sections of site not visible. Might be inferred from geographical context. **I:** Semi-scientific details – generic damage nature and full extent uncertain. Vague terminology with substantial proportion of site absent from data.	Moderate conflicts Minor inconsistencies Limited rational explanation	Months – Years	Average confidence that DS is correct (~50% certainty). Some signs of uncertainty, bias, misinformation, modification, contamination, and lack of credibility Average confidence in RG value.
RG3	Data only presents information on two of the three components OR/AND Majority of available event data sources contribute to site dataset.	**H:** No scientific measurement for intensity, however an estimated measure has been provided from the primary data source (method/collector). Extent should be known at site. **A:** Good detail - general typology, confident spatial location, majority of site visible. **I:** Scientific details – nature and potential extent of damage, incomplete description, with variation in terminology. Not known across full site.	Limited conflicts Mostly consistent Rational explanation available	Weeks – Months	Good confidence that DS is correct. Limited signs of uncertainty, bias, misinformation, modification, contamination, or lack of credibility Good confidence in RG value.
RG4	Data MUST present information on all three components AND ALL available event data sources contribute to site dataset.	**H:** Scientific measurements for hazard extent and intensity. Known intensity distribution from overlap of hazard intensity metrics or observation across entire site. **A:** Comprehensive details – known typology, known spatial location, full site visibility, potential internal surveys. **I:** Engineering-level details – nature and extent of damage with correct terminology.	No conflicts Completely consistent Rational explanation comprehensive (where needed)	Hours – Days	High confidence that DS is correct. No signs of uncertainty, bias, misinformation, modification, contamination, or lack of credibility. High confidence in RG value.

Figure 6: Proposed impact data reliability assessment framework, operationalising key dimensions of data reliability into five criteria and five reliability grades (RG0–RG4) for consistent site-level evaluation of post-disaster impact data. H = hazard data, A = Asset data, I = impact/damage data. Data falling within the grey shaded area with bounding box generally considered unsuitable or of very limited value for further analysis. Adapted from Stratton (in prep).

Prior model behaviour relative to empirical evidence

There are considerable discrepancies between the GAR vulnerability function and observed building damage data from the Calbuco eruption. This appears to be most important at intermediate damage states (e.g., DS2-4), where the GAR function appears to have a nearly 0 damage before very quickly ramping up compared to the Calbuco dataset. There were considerable variances between experts during the expert elicitation process, which may influence this. Alternatively, the Calbuco dataset may be biased towards instances of damage. The general quality of data was classified as poor-moderate by Hayes et al. (2019). Allen (2021) investigated ashfall loading on New Zealand style metal sheet roofs on a timber-framed building and found that extremely high accumulations of ash were required before failure was observed (16kPa). Allen (2021) suggests that their experimental setup may have contributed to this outcome and that they were not able to observe non-failure damage states that could have occurred prior to full failure. This highlights the important of improving building vulnerability models for volcanic ash, and further improving and expanding the empirical evidence of building damage from ashfall.

Bayesian model update performance

Given the observation error assumptions used in this study, the best-case achievable MAE is 0.46 to 0.50 (approximately half a damage state) across the scenarios (Figure 7). This provides a hypothetical lower bound for expected forecast error under the simulated data-generating conditions and should be treated as conditional on those assumptions. Thus, even with an ideal Bayesian updating scheme, expected error cannot reasonably approach zero when observations are noisy, partially misclassified, and unevenly distributed through intensity space.

At 1 month in the base input simulation, source-informed updates remained better than no updating (for example, MAE ~1.10 versus baseline 1.33), but still well above the scenario-specific floor range (~0.46-0.50) (Figure 8). This indicates meaningful improvement over prior-only/no-update performance, while also confirming substantial remaining error under realistic observation noise. The gain from weighting was modest in many cases because later timesteps were dominated by high-volume lower-confidence sources (satellite and crowd) relative to sparse high-confidence field observations, limiting leverage from the highest-trust data.

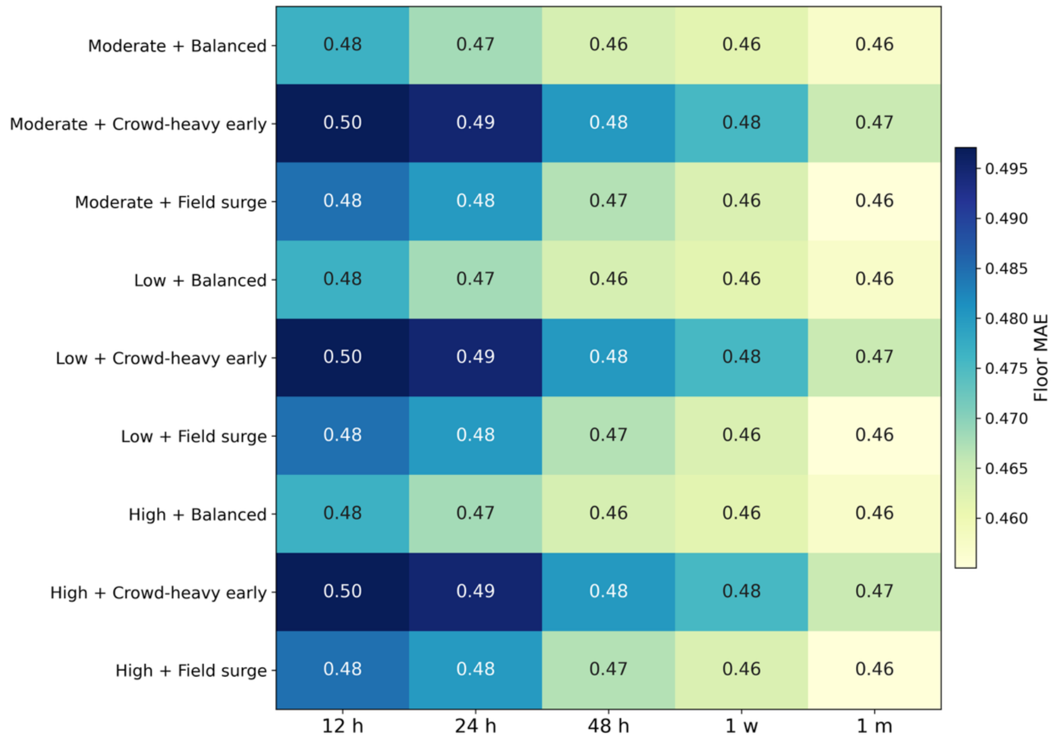

Figure 7: Lowest plausible MAE for each scenario

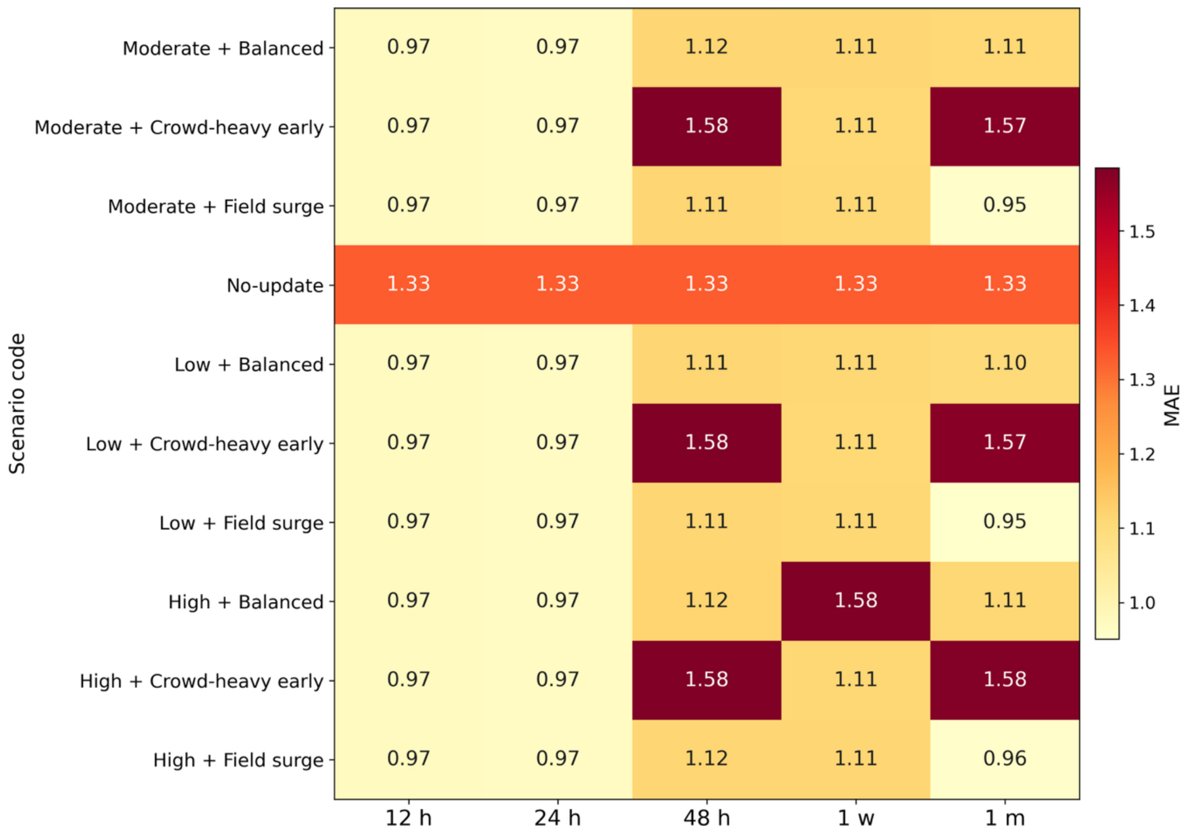

Figure 8: MAE at each timestep for different pseudo-observational settings

The models provide an improvement on the no-update model in 84% (38/45) of the time-steps for the scenarios (Figure 8). The median improvement in MAE across the scenarios where an improvement is observed is between 0.3 and 0.36. The majority of the scenarios where there is no improvement on the no-update model are when there is heavy influence of the crowd-sourced data, indicating that high-volume noisy data can temporarily pull the posterior away from the true state. Interestingly, the strongest gains are observed early in the sequence (12-24 hrs). This all points towards the importance for obtaining an appropriate mixture of data that does not bias the observations. Most misclassifications occur in the intermediate damage states (DS2-4). MAE could possibly be reduced through time by identifying buildings that are likely to be wrong and target these for improved ground-truthing. However, more work is required to identify a reliable approach to evaluate buildings that are likely to be wrong and not just uncertain.

An important consideration when interpreting our results is the potential influence of the chosen weighting scheme. The source-specific weights used here should be interpreted as ordinal assumptions about relative reliability rather than empirically calibrated measures of observation error. Although weighting was intended to prevent large volumes of lower-confidence satellite and crowd-sourced observations from overwhelming fewer higher-confidence field observations, the selected values directly influence the effective contribution of each source and, consequently, the posterior estimates. Alternative weight ratios could therefore produce different update trajectories and conclusions about the relative value of each observation stream. Future work should undertake systematic sensitivity testing across plausible weight combinations and source volumes, ideally informed by empirical estimates of classification accuracy. This would help identify whether the reported improvements are robust, determine when high-volume observations dominate despite low assigned weights, and establish operational weighting schemes appropriate for different stages of an eruption response.

Case-study implications for New Zealand

Comparative posterior checks with the two volcanic ash building damage datasets for timber-framed buildings (Calbuco 2015 and Rabaul 1994) show that both datasets shift inferred vulnerability away from the prior toward higher damage at lower intensity (Figure 9). Rainfall that occurred following both eruptions could partially be the reason for this by increasing loading on roofs. It is not immediately clear how or if rainfall were incorporated into the vulnerability curves from Maqsood et al. (2014). Further, some of the observed damage (e.g., roof denting) is not obviously influenced by rainfall. It is possible that this no-update vulnerability curve could under-estimate building damage in both case-study contexts. This finding is relevant for volcanic risk in New Zealand, where building exposure is often dominated by relatively low ash loadings due to the distance of most development from volcanoes. As a result, relying on the prior alone could lead to systematic under-estimation of ash-related building losses. Therefore, Bayesian updating provides a defensible interim mechanism to reduce prior bias while improved empirically grounded vulnerability models are developed.

An important question is whether the observed increase in low-intensity damage reflects systematic inter-event differences (e.g., event-specific characteristics such as ash properties, duration of loading, or regional building vulnerability) or intra-event variability and compounding processes that are not well captured in the prior model. For example, the Calbuco dataset may reflect local building stock (e.g., roof slopes) or particular ash characteristics that promote non-structural damage such as larger grain sizes leading to roof denting (Osman et al. 2019), or higher roof slopes promoting ash sliding and increased gutter damage (Osman et al. 2023). This level of detail is not captured in the underpinning datasets. Alternatively, the discrepancy may indicate that the prior model under-represents the range of processes contributing to low-level damage, which tend to be more variable and less threshold-controlled than structural collapse. Distinguishing between these explanations is challenging with limited datasets but has important implications for model transferability and generalisation. This suggests low-damage states may be more sensitive to both event-specific conditions (e.g., grainsize, chemistry) and unresolved variability, whereas higher damage states are more strongly governed by structural capacity limits.

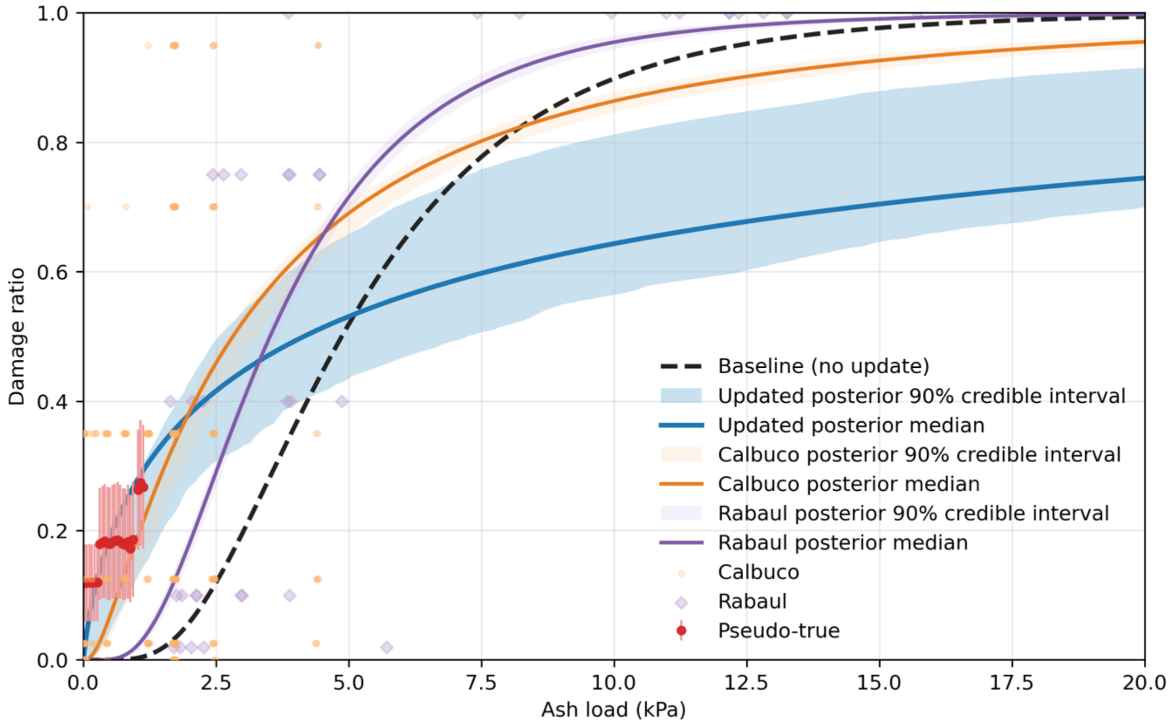

Figure 9: Bayesian updated vulnerability curves for the hypothetical Taupō eruption scenario with pseudo-observations compared with the prior model (GAR 2015 vulnerability functions) and similarly updated vulnerability functions using Calbuco 2015 and Rabaul 1994 datapoints.

The findings of this work highlight the importance of improved vulnerability models for volcanic ash induced damage. Low-end damage is dominated by non-structural and cosmetic effects, while higher damage reflects structural failure, each with different triggers and mechanics. Therefore, separate component or regime-specific curves may be worth exploring. For example, separate curves for non-structural and structural damage types. Another option would be to explore a two-stage hurdle model that first assesses whether any meaningful damage has occurred and then a second stage evaluates how severe that damage is conditional on passing the initial hurdle. This would allow for low-damage onset and high damage severity to have different shapes and possibly different hazard intensity metrics (e.g., grainsize, chemistry, duration of loading), which could be important for assessing building damage from volcanic ash.

Conclusions

This study shows that dynamic Bayesian updating can produce meaningful improvements in rapid volcanic ash damage assessment and provide a stronger evidence base for time-critical response decisions. However, it is important to note that since there has not been a major ash producing eruption in New Zealand for 30 years, our study findings are necessarily inferential rather than direct. The case-study datasets used are among the few available empirical building-damage datasets for ash fall and are consistent with the timber-framed buildings in New Zealand, but the limitations with this approach are acknowledged.

The following three conclusions are most relevant from this work:

- Prior vulnerability models should be treated as starting points rather than fixed when conducting rapid impact assessments where locally specific and validated models are not available, particularly in low-to-moderate intensity ranges.
- Observation weighting is useful when assimilating mixed-quality sources, but performance also depends on coverage of high-confidence observations through the hazard intensity space. Therefore, developing strategic approaches for collection of high reliability but resource intensive data sources is critical.
- Better understanding of low-level building damage and how to incorporate it into damage and loss modelling frameworks is necessary.

Whilst this work shows promising results, more work is required before such a framework could be operationally used. Recommended next steps are to:

- Develop New Zealand-specific building vulnerability models to improve prior models. In particular, the development of fragility functions for the range of New Zealand building types is a key priority. This is because present approaches require the application of a globally generalised vulnerability function, which introduces issues with its applicability to the New Zealand building stock and applying it when building damage (not just loss) is the key requirement. Exploring regime-specific/hurdle model structures to obtain a better characterisation of minor-moderate damage states and better uncertainty characterisation is needed.
- Explore methods for better treatment of mixed data sources in Bayesian updating. For example, is it possible to identify buildings that are likely to have the wrong damage state assigned by a low reliability data source and prioritise ground-truthing for this?
- Validate this approach for a real eruption case-study, should the opportunity arise.

Outputs and dissemination

The model and associated functions used in this study are available at:

- Active repository: https://github.com/josh-hayes/Rapid_building_damage_updating
- Initial release: https://doi.org/10.5281/zenodo.19692910

Publications, communications and engagement

Student projects

Stratton N, In prep. Evaluating the Reliability of Post-Volcanic Eruption Building Damage Data Collection Methods. MSc University of Canterbury. Intended submission: 30 April 2026

Manuscripts in preparation

Two manuscripts are current in preparation to be submitted to international peer reviewed academic journals:

1. Improving post-eruption volcanic ash building damage estimation through Bayesian updating with heterogeneous observational data.
2. Evaluating the Reliability of Post-Volcanic Eruption Building Damage Data Collection Methods

Media and other communications

The project featured in a number of media publications in late September 2024:

- https://www.rnz.co.nz/news/national/528723/hundreds-of-thousands-of-north-island-homes-could-be-covered-in-ash-in-next-volcanic-eruption
- https://www.newstalkzb.co.nz/news/national/north-island-risk-how-scientists-plan-to-forecast -volcanic-ashfall-from-next-major-nz-eruption/
- https://www.insurancebusinessmag.com/nz/news/catastrophe/new-zealand-prepares-for-volcanic-ashfall-with-new-damage-model-507006.aspx
- https://www.naturalhazards.govt.nz/news/volcanic-ashfall-what-damage-could-be-caused-after-an-eruption-new-research-to-assist-authorities-and-communities/
- https://www.gns.cri.nz/news/new-research-assesses-damage-from-volcanic-ashfall/
- TVNZ News Bulletin Sunday September 22, 2024 (no link available)

NHC featured project member MSc student Niamh Stratton in an online LinkedIn Post and on their website:

- https://www.naturalhazards.govt.nz/news/research-to-cut-through-crowdsourced-noise-after-ashfall/
- https://www.linkedin.com/posts/naturalhazardscommission_after-a-volcanic-eruption-updates-pour-in-activity-7437286047404756992-TUjQ?utm_source=share&utm_medium=member_desktop&rcm=ACoAABAlVYsBwQUuFVe3dQgplNcQC4BOyctTc2o

NHC LinkedIn Post on project:

- https://www.linkedin.com/posts/naturalhazardscommission_tsunami-volcanoes-activity-7311185703625732097-shFs?utm_source=share&utm_medium=member_desktop&rcm=ACoAABAlVYsBwQUuFVe3dQgplNcQC4BOyctTc2o

Engagement activities

- Workshop held with Natural Hazards Commission on direction for volcanic risk research on August 13, 2024.
- Josh Hayes presented at the joint NHC/NHRP Workshop: The Future of Natural Hazard Risk Modelling in New Zealand: Volcano Vulnerability Models.
- Niamh Stratton presented at Geosciences New Zealand Thursday 25th November 2025: Evaluating the Reliability of Post-Volcanic Eruption Building Damage Data Collection Methods.

References

Allen NE (2021) Multi-volcanic hazard impact assessment for residential buildings in the Auckland Volcanic Field, New Zealand. University of Canterbury

Asadzadeh A, Kötter T, Salehi P, Birkmann J (2017) Operationalizing a concept: The systematic review of composite indicator building for measuring community disaster resilience. Int J Disaster Risk Reduct 25:147–162. https://doi.org/10.1016/j.ijdrr.2017.09.015

Ba Y, Mancenido MV, Chiou EK, Pan R (2025) Data quality in crowdsourcing and spamming behavior detection. Behav Res Methods 57:251. https://doi.org/10.3758/s13428-025-02757-5

Barker SJ, Van Eaton AR, Mastin LG, et al (2019) Modeling Ash Dispersal From Future Eruptions of Taupo Supervolcano. Geochem Geophys Geosystems 20:3375–3401. https://doi.org/10.1029/2018GC008152

Baxter PJ, Boyle R, Cole P, et al (2005) The impacts of pyroclastic surges on buildings at the eruption of the Soufri�re Hills volcano, Montserrat. Bull Volcanol 67:292–313. https://doi.org/10.1007/s00445-004-0365-7

Biass S, Jenkins S, Lallemant D, et al (2021) Remote sensing of volcanic impacts. In: Forecasting and Planning for Volcanic Hazards, Risks, and Disasters. Elsevier, pp 473–491

Biass S, Todde A, Cioni R, et al (2017) Potential impacts of tephra fallout from a large-scale explosive eruption at Sakurajima volcano, Japan. Bull Volcanol 79:73. https://doi.org/10.1007/s00445-017-1153-5

Blake DM, Wilson G, Stewart C, et al (2015) The 2014 eruption of Kelud volcano, Indonesia: impacts on infrastructure, utilities, agriculture and health. GNS Science, Lower Hutt, N.Z.

Blong R (2003) Building damage in Rabaul, Papua New Guinea, 1994. Bull Volcanol 65:43–54. https://doi.org/10.1007/s00445-002-0238-x

Cai L, Zhu Y (2015) The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Sci J 14:2. https://doi.org/10.5334/dsj-2015-002

Calantropio A, Chiabrando F, Codastefano M, Bourke E (2021) DEEP LEARNING FOR AUTOMATIC BUILDING DAMAGE ASSESSMENT: APPLICATION IN POST-DISASTER SCENARIOS USING UAV DATA. ISPRS Ann Photogramm Remote Sens Spat Inf Sci V-1–2021:113–120. https://doi.org/10.5194/isprs-annals-V-1-2021-113-2021

Chen H, Hailey D, Wang N, Yu P (2014) A Review of Data Quality Assessment Methods for Public Health Information Systems. Int J Environ Res Public Health 11:5170–5207. https://doi.org/10.3390/ijerph110505170

Chen Z, Lim S (2021) Social media data-based typhoon disaster assessment. Int J Disaster Risk Reduct 64:102482. https://doi.org/10.1016/j.ijdrr.2021.102482

Corbane C, De Groeve T, Ehrlich D, Poljansek K (2015) A EUROPEAN FRAMEWORK FOR RECORDING AND SHARING DISASTER DAMAGE AND LOSS DATA. Int Arch Photogramm Remote Sens Spat Inf Sci XL-3/W3:277–283. https://doi.org/10.5194/isprsarchives-XL-3-W3-277-2015

de Bruijn JA, Daniell JE, Pomonis A, et al (2022) Using rapid damage observations for Bayesian updating of hurricane vulnerability functions: A case study of Hurricane Dorian using social media. Int J Disaster Risk Reduct 72:102839. https://doi.org/10.1016/j.ijdrr.2022.102839

Deligne NI, Fitzgerald RH, Blake DM, et al (2017) Investigating the consequences of urban volcanism using a scenario approach I: Development and application of a hypothetical eruption in the Auckland Volcanic Field, New Zealand. J Volcanol Geotherm Res 336:192–208. https://doi.org/10.1016/j.jvolgeores.2017.02.023

Deligne NI, Jenkins SF, Meredith ES, et al (2022) From anecdotes to quantification: advances in characterizing volcanic eruption impacts on the built environment. Bull Volcanol

Eckhardt D, Leiras A, Thomé AMT (2019) Systematic literature review of methodologies for assessing the costs of disasters. Int J Disaster Risk Reduct 33:398–416. https://doi.org/10.1016/j.ijdrr.2018.10.010

Fitzgerald RH, Wilson TM, Hayes JL, et al (2023) A stocktake of global volcanic vulnerability models to inform future volcanic risk research in Aotearoa New Zealand. GNS Science

Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) emcee: The MCMC Hammer. Publications of the Astronomical Society of the Pacific 125:925. https://doi.org/10.1086/670067

Gunasekera R, Daniell JE, Pomonis A, et al (2018) Methodology Note on the Global RApid post-disaster Damage Estimation (GRADE) approach. Global Facility for Disaster Reduction and Recovery, Washington D.C.

Hao H, Wang Y (2021) Assessing Disaster Impact in Real Time: Data-Driven System Integrating Humans, Hazards, and the Built Environment. J Comput Civ Eng 35:04021010. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000970

Hayes JL, Calderón B R, Deligne NI, et al (2019) Timber-framed building damage from tephra fall and lahar: 2015 Calbuco eruption, Chile. J Volcanol Geotherm Res 374:142–159. https://doi.org/10.1016/j.jvolgeores.2019.02.017

Hayes JL, Fitzgerald RH, Wilson TM, et al (2024) Linking hazard intensity to impact severity: mini review of vulnerability models for volcanic impact and risk assessment. Front Earth Sci 11:1278283. https://doi.org/10.3389/feart.2023.1278283

Jenkins SF, McSporran A, Wilson TM, et al (2024) Tephra fall impacts to buildings: the 2017–2018 Manaro Voui eruption, Vanuatu. Front Earth Sci 12:. https://doi.org/10.3389/feart.2024.1392098

Ladds M, Keating A, Handmer J, Magee L (2017) How much do disasters cost? A comparison of disaster cost estimates in Australia. Int J Disaster Risk Reduct 21:419–429. https://doi.org/10.1016/j.ijdrr.2017.01.004

Magill C, Wilson T, Okada T (2013) Observations of tephra fall impacts from the 2011 Shinmoedake eruption, Japan. Earth Planets Space 65:677–698. https://doi.org/10.5047/eps.2013.05.010

Maqsood T, Wehner M, Ryu H, et al (2014) GAR15 Regional Vulnerability Functions: Reporting on the UNISDR/GA SE Asian Regional Workshop on Structural Vulnerability Models for the GAR Global Risk Assessment, 11-14 November, 2013, Geoscience Australia, Canberra, Australia. Geoscience Australia

Mashoufi M, Ayatollahi H, Khorasani-Zavareh D (2018) A Review of Data Quality Assessment in Emergency Medical Services. Open Med Inform J 12:19–32. https://doi.org/10.2174/1874431101812010019

Meredith ES, Jenkins SF, Hayes JL, et al (2022) Damage assessment for the 2018 lower East Rift Zone lava flows of Kīlauea volcano, Hawaiʻi. Bull Volcanol 84:65. https://doi.org/10.1007/s00445-022-01568-2

Moriyama K, Sasaki D, Ono Y, International Research Institute of Disaster Science (IRIDeS), Tohoku University 468-1-S302 Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-0845, Japan (2018) Comparison of Global Databases for Disaster Loss and Damage Data. J Disaster Res 13:1007–1014. https://doi.org/10.20965/jdr.2018.p1007

Osman S, Rossi E, Bonadonna C, et al (2019) Exposure-based risk assessment and emergency management associated with the fallout of large clasts at Mount Etna. Nat Hazards Earth Syst Sci 19:589–610. https://doi.org/10.5194/nhess-19-589-2019

Osman S, Thomas M, Crummy J, et al (2023) Laboratory tests to understand tephra sliding behaviour on roofs. J Appl Volcanol 12:11. https://doi.org/10.1186/s13617-023-00137-2

Paulik R, Horspool N, Woods R, et al (2023) RiskScape: a flexible multi-hazard risk modelling engine. Nat Hazards 119:1073–1090. https://doi.org/10.1007/s11069-022-05593-4

Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM 45:211–218. https://doi.org/10.1145/505248.506010

Pita G, Pinelli J-P, Gurley K, Mitrani-Reiser J (2015) State of the Art of Hurricane Vulnerability Estimation Methods: A Review. Nat Hazards Rev 16:04014022. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000153

Pita GL, Pinelli J-P, Gurley KR, Hamid S (2013) Hurricane vulnerability modeling: Development and future trends. J Wind Eng Ind Aerodyn 114:96–105. https://doi.org/10.1016/j.jweia.2012.12.004

Romão X, Paupério E (2016) A framework to assess quality and uncertainty in disaster loss data. Nat Hazards 83:1077–1102. https://doi.org/10.1007/s11069-016-2364-3

Scaini C, Biass S, Galderisi A, et al (2014) A multi-scale risk assessment for tephra fallout and airborne concentration from multiple Icelandic volcanoes – Part 2: Vulnerability and impact. Nat Hazards Earth Syst Sci 14:2289–2312. https://doi.org/10.5194/nhess-14-2289-2014

Scheele FR, Syed Y, Hayes JL, et al (2023) Building inventory and vulnerability functions for risk modelling in New Zealand. https://doi.org/10.21420/G34N-V958

Silva V, Akkar S, Baker J, et al (2019) Current Challenges and Future Trends in Analytical Fragility and Vulnerability Modeling. Earthq Spectra 35:1927–1952. https://doi.org/10.1193/042418EQS101O

Spence RJS, Pomonis A, Baxter PJ, et al (1996) Buillding damage caused by the Mount Pinatubo Eruption of 15 June 1991. In: Newhall CG, Punongbayan RS (eds) Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines. University of Washington Press, London, UK., pp 1055–1061

Stirling MW, Wilson CJN (2002) Development of a volcanic hazard model for New Zealand. Bull N Z Soc Earthq Eng 35:266–277. https://doi.org/10.5459/bnzsee.35.4.266-277

Tennant E, Jenkins SF, Miller V, et al (2024) Automating tephra fall building damage assessment using deep learning. Nat Hazards Earth Syst Sci 24:4585–4608. https://doi.org/10.5194/nhess-24-4585-2024

Thomas K-L, Woods RJ, Garlick R, et al (2020) User requirements of RiskScape 2.0 software and opportunities for disaster risk research in Aotearoa-New Zealand. https://doi.org/10.21420/RVDT-8R62

Weir AM, Williams JH, Wilson TM, et al (2024) Rapid remote volcanic ashfall impact assessment for the 2022 eruption of Hunga volcano, Tonga: a bespoke approach and lessons identified. Bull Volcanol 86:88. https://doi.org/10.1007/s00445-024-01776-y

Williams GT, Jenkins SF, Biass S, et al (2020) Remotely assessing tephra fall building damage and vulnerability: Kelud Volcano, Indonesia. J Appl Volcanol 9:10. https://doi.org/10.1186/s13617-020-00100-5

Wilson KM, Baldwin JW, Young RM (2022) Estimating Tropical Cyclone Vulnerability: A Review of Different Open-Source Approaches. In: Collins JM, Done JM (eds) Hurricane Risk in a Changing Climate. Springer International Publishing, Cham, pp 255–281

Wilson TM, Cole JW, Stewart C, et al (2011) Ash storms: impacts of wind-remobilised volcanic ash on rural communities and agriculture following the 1991 Hudson eruption, southern Patagonia, Chile. Bull Volcanol 73:223–239. https://doi.org/10.1007/s00445-010-0396-1

Wilson TM, Stewart C, Sword-Daniels V, et al (2012) Volcanic ash impacts on critical infrastructure. Phys Chem Earth 45–46:5–23. https://doi.org/10.1016/j.pce.2011.06.006

Xing Z, Zhang X, Zan X, et al (2021) Crowdsourced social media and mobile phone signaling data for disaster impact assessment: A case study of the 8.8 Jiuzhaigou earthquake. Int J Disaster Risk Reduct 58:102200. https://doi.org/10.1016/j.ijdrr.2021.102200

Yepes-Estrada C, Silva V, Rossetto T, et al (2016) The Global Earthquake Model Physical Vulnerability Database. Earthq Spectra 32:2567–2585. https://doi.org/10.1193/011816EQS015DP

Zhang G, Zhu A-X (2018) The representativeness and spatial bias of volunteered geographic information: a review. Ann GIS 24:151–162. https://doi.org/10.1080/19475683.2018.1501607

Zuccaro G, Leone MF, Del Cogliano D, Sgroi A (2013) Economic impact of explosive volcanic eruptions: A simulation-based assessment model applied to Campania region volcanoes. J Volcanol Geotherm Res 266:1–15. https://doi.org/10.1016/j.jvolgeores.2013.09.002