https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Open Access Earth System Science Data Discussions

The 2024 Noto Peninsula earthquake building damage dataset: Multi-source visual assessment

Ruben Vescovo[1], Bruno Adriano[2], Sesa Wiguna[1], Chia Yee Ho[1], Jorge Morales[1], Xuanyan Dong[1], Shin Ishii[1], Kazuki Wako[1], Yudai Ezaki[3], Ayumu Mizutani[2], Erick Mas[2], Satoshi Tanaka[4], and Shunichi Koshimura[2]

[1]Department of Civil and Environmental Engineering, Tohoku University, Aoba 468–1, Aramaki, Aoba-ku, Sendai, 980–8572, Japan
[2]International Research Institute of Disaster Science (IRIDeS), Tohoku University, Aoba 468–1, Aramaki, Aoba-ku, Sendai, 980–8572, Japan
[3]Department of Civil Engineering and Architecture, School of Engineering, Aoba–6–6–06 Aramaki, Aoba Ward, Sendai, Miyagi 980–8572
[4]Faculty of Social and Environmental Studies, Department of Social and Environmental Studies, Tokoha University, Yayoi-cho 6–1, Suruga-ku, Shizuoka city, 422–8581, Japan

Correspondence: Shunichi Koshimura (shunichi.koshimura.a3@tohoku.ac.jp)

Abstract. We present a building damage dataset following the 2024 Noto Peninsula Earthquake. The database was compiled from freely available, multi-source, remote sensing data, verified through opt-in crowd-sourced information. The dataset consists of geo-referenced vector polygons representing the pre-event building footprints of 140,208 structures. Each building was classified through visual inspection using pre-disaster and post disaster vertical, oblique, survey, and verifiable news reporting imagery. Entries were validated using voluntary-submission data sourced through a web-API hosting a live version of the database. We calculate classification metrics for a subset of the database where ground survey photographs were provided by independent surveyors. An average F_1-score of 0.94 suggests that the proposed assessment is consistent and high quality. We aim to inform future disaster research such as disaster dynamics models; statistical and machine learning damage models; logistics and evacuation studies. The present work describes the data collection process, damage assessment methodology, and rationale; including limitations encountered, the crowd sourcing validation process, and the dataset structure.

1 Introduction

At 16:10 January 1st 2024, shallow reverse faulting produced a $_w$ earthquake (United States Geological Survey, 2024) that propagated from the north-most point of Suzu City, Ishikawa Prefecture (Figure: 1). The disaster poses unique challenges for the disaster geophysics community due to its context and outfall. The intraplate faulting occurred on the relatively inert western coast of Honshu Island, Japan, following a 3 year long earthquake swarm (Ishikawa and Bai, 2024). Preliminary works attribute the prolonged seismic activity, and unique violence of the fault mechanism, to upwelling of fluid through an intensely cracked fault line. This fluid buildup is purported to have exacerbated seismic activity due to the reverse nature of the stress field; these constraints caused pressure to diffuse horizontally through the crust rather than escaping vertically, ultimately causing

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Open Access Earth System Science Data Discussions

the swarm to affect a larger than predicted area (Kato, 2024; Ishikawa and Bai, 2024; Shelly, 2024). The main shock originated north of Suzu City followed by more than 1500 strong aftershocks that continued into the next several days. Affected areas were the Suzu, Noto, Wajima, Nanao, Shika and Anamizu municipalities (NASA, 2024; Japan Meteorological Agency, 2024; Japanese Red Cross Society, 2024). The Prime Minister's Office of Japan has provided transcripts for several press conferences and emergency meetings reporting actions taken to address monitoring and relief operations. Initial reporting informed of near instant tsunami impacts around the main shock's epicenter (north Suzu), quickly followed by a comprehensive tsunami warning along the entire peninsula's coast. Subsequent statements confirmed the presence of catastrophic damage affecting infrastructure throughout the peninsula including ground shaking, land deformation, liquefaction, and landslides causing varied damage to buildings, interrupting roads, originating a fire. Consequent impacts to critical services such as water supply, sewage system, power outages, and telecommunication service disruptions (Egawa et al., 2024; Prime Minister's Office of Japan, 2024, a, b, c; British Broadcasting Corporation, 2024). The disaster and following outfall ultimately resulted in injuries and human casualties, the prevention of which represents an overarching focus of disaster research (preventable disaster deaths) (Egawa et al., 2024). The spatial distribution of infrastructure impacts is of particular importance to disaster research. Such data can inform emergency response studies (physical dynamics simulation, damage detection, damage estimation, evacuation simulation, etc.), long term recovery studies (socioeconomic studies, disaster epidemiology, disaster prevention, probabilistic hazard, etc.), and ultimately the development of more informed codes and regulation. Disaster damage visual assessments are critical to develop a comprehensive corpus of disaster impacts to infrastructure and to inform studies such as the aforementioned. Current global disaster datasets such as xBD (Gupta et al., 2019) have not included disaster data from Japan, in this capacity we hope to contribute with local coverage to such initiatives. Visual assessments can be carried out by an on-the-ground survey team ensuring the highest degree of fidelity and granularity. However such an investigation is often resource intensive, carries inherent risk of harm, and may be highly invasive. Alternative methods generally employ remote sensing data and have historically been carried out by experts (Gokon and Koshimura, 2012) and institutions (CEMS, 2017). Human visual assessments have informed several studies that have contributed to a deeper understanding of seismic and tsunami building damage Chua et al. (2021) for example conducted a limited scope visual assessment of the 2011 Great East Earthquake and Tsunami using multi-source imagery (including obliques) to generate fragility curves of port structures in Ishinomaki City. More recently automated methods employing pre-trained machine learning models have been explored (Deng and Wang, 2022; Miura et al., 2020; Wiguna et al., 2024a). Such methods generally leverage vertical imagery as input to a machine learning model to perform automatic classification of building damage. These automated image based assessments carry inherent limitations such as image resolution, observation angle, weather conditions, canopy obstruction, availability of baseline data, human error (within the training data), among others. Automated methods may suffer from additional limitations such as domain bias, model bias, etc. that make these methods particularly susceptible to error propagation and uncertainty on unseen domains: In Wiguna et al. (2024b) the authors explore mechanisms to address some of these technical limitations by fine tuning a pre-trained model to a new target domain using a semi-supervised framework. A preliminary investigation of the building footprint inventory revealed large discrepancies with pre-event imagery (Figure 8). Moreover, the variable aerial survey periods, image capture quality, meteorological conditions (Table 4), and different mechanisms driving building failure (Such as fire, tsunami, earthquake, etc.;

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

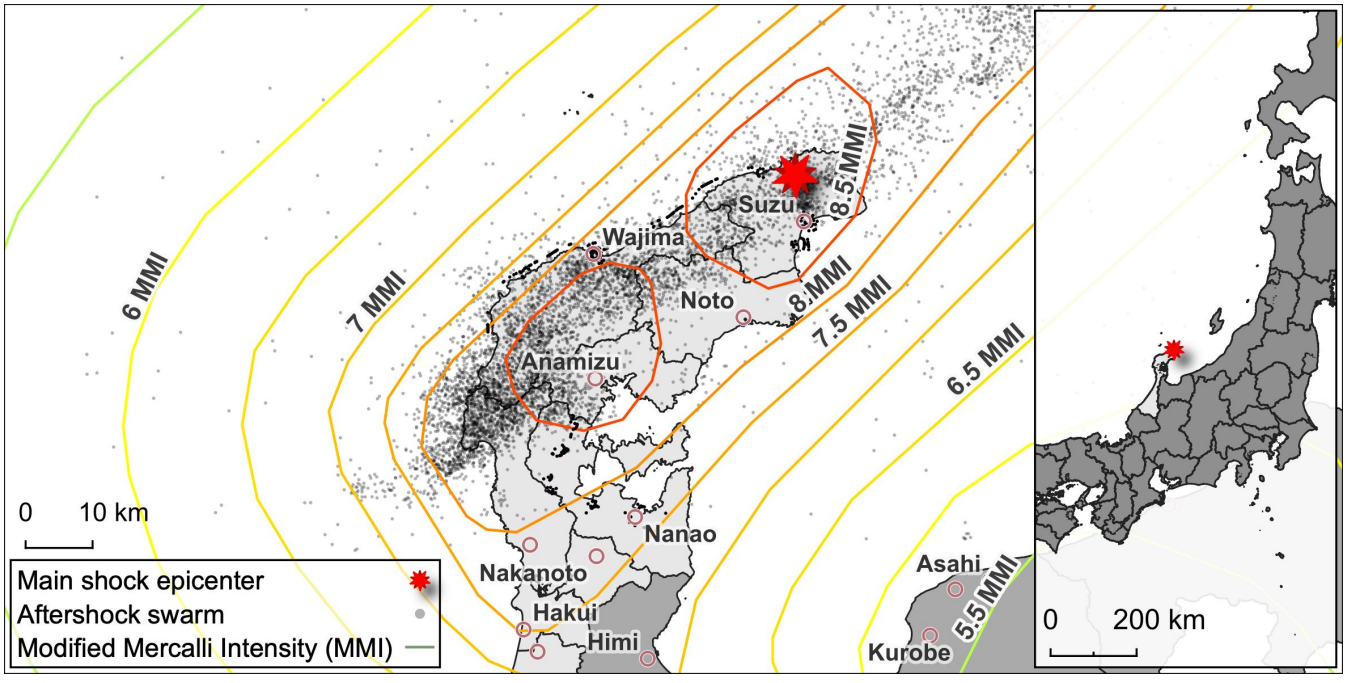


Figure 1. Seismic context of the 2024 Noto Peninsula Earthquake showing the distribution of the earthquake swarm following the aftershocks (United States Geological Survey, 2024; GSI, 2024).

Figure 5) contributed to a visually fragmented and inconsistent domain. Despite these challenges, the unprecedented availability of open-source & multi-source data provided us with a unique opportunity for a rapid visual damage assessment. With the above considerations in mind, we opted for a manual approach to curate a baseline dataset. We hope that this dataset will serve as a reference for future studies and as a benchmark for automated methods. The primary sources for the investigation were post-disaster vertical imagery captured by the Geospatial Information Authority of Japan (GSI) and made available online. In addition oblique imagery of select portions of Noto Peninsula were made available by Kokusai Kogyo (KKC) through the free version of their proprietary aerial survey database. The post-disaster imagery data informed the classification of the public GSI building footprint inventory vector data. Our criteria was developed iteratively in response to limitations presented by the data. Following an initially limited-scope investigation, the assessment was made available to the public for a progressive appraisal at the online portal: https://experience.arcgis.com/experience/70aae9964dc54e4190b6b360dcbb3759/. End users may request corrections regarding potential misclassifications. Requests must include proof to substantiate the amendment, usually in the form of a photo of the target building. Finally, two limited-scope onsite surveys by independent research teams informed a secondary round of corrections.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

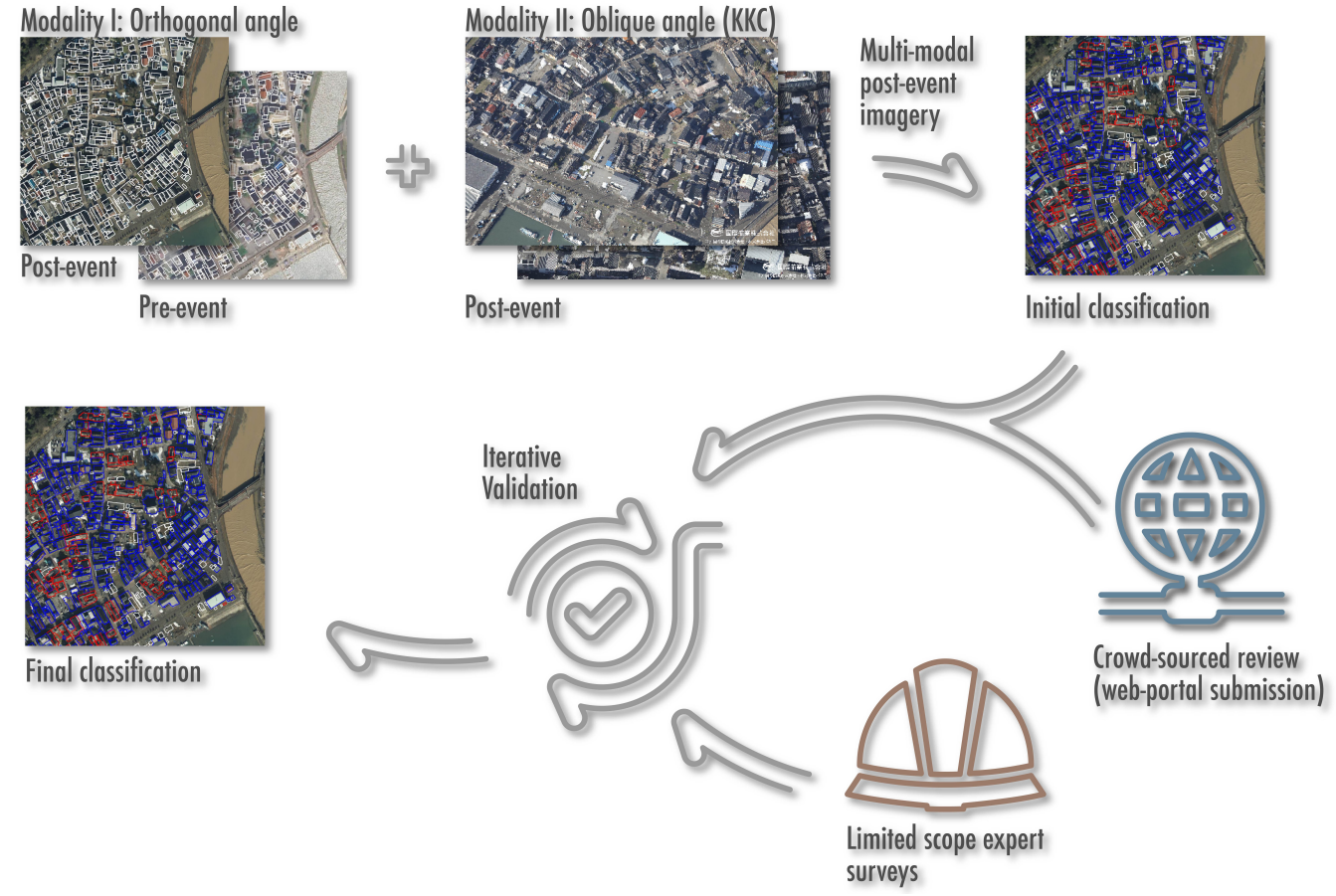


Figure 2. Building damage visual assessment workflow illustrating the working group's approach to multi-source & multi-modal data, each stage of inclusion and expert-feedback-driven iterative validation process. Pre-event orthophoto by © Google 2024, Post event aerial by GSI (2024), All oblique imagery courtesy of © KKC (2024).

2 Methods

Herein, we relate the methodology used to generate the dataset, including: considerations, challenges, and limitations encountered during the creation of the dataset (Figure 2).

2.1 Data sourcing

The initial review consisted of a general overview of available data from official sources. The Government of Japan provides basic geographic information through the GSI (https://fgd.gsi.go.jp/download/menu.php, Japanese). An inventory of building footprints, pre-event aerial imagery, and digital elevation data, provided a general level of clarity for the feasibility of a visual assessment. Moreover, GSI hosts an index of information pertinent to the Noto Peninsula Earthquake on a dedicated page (GSI,

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

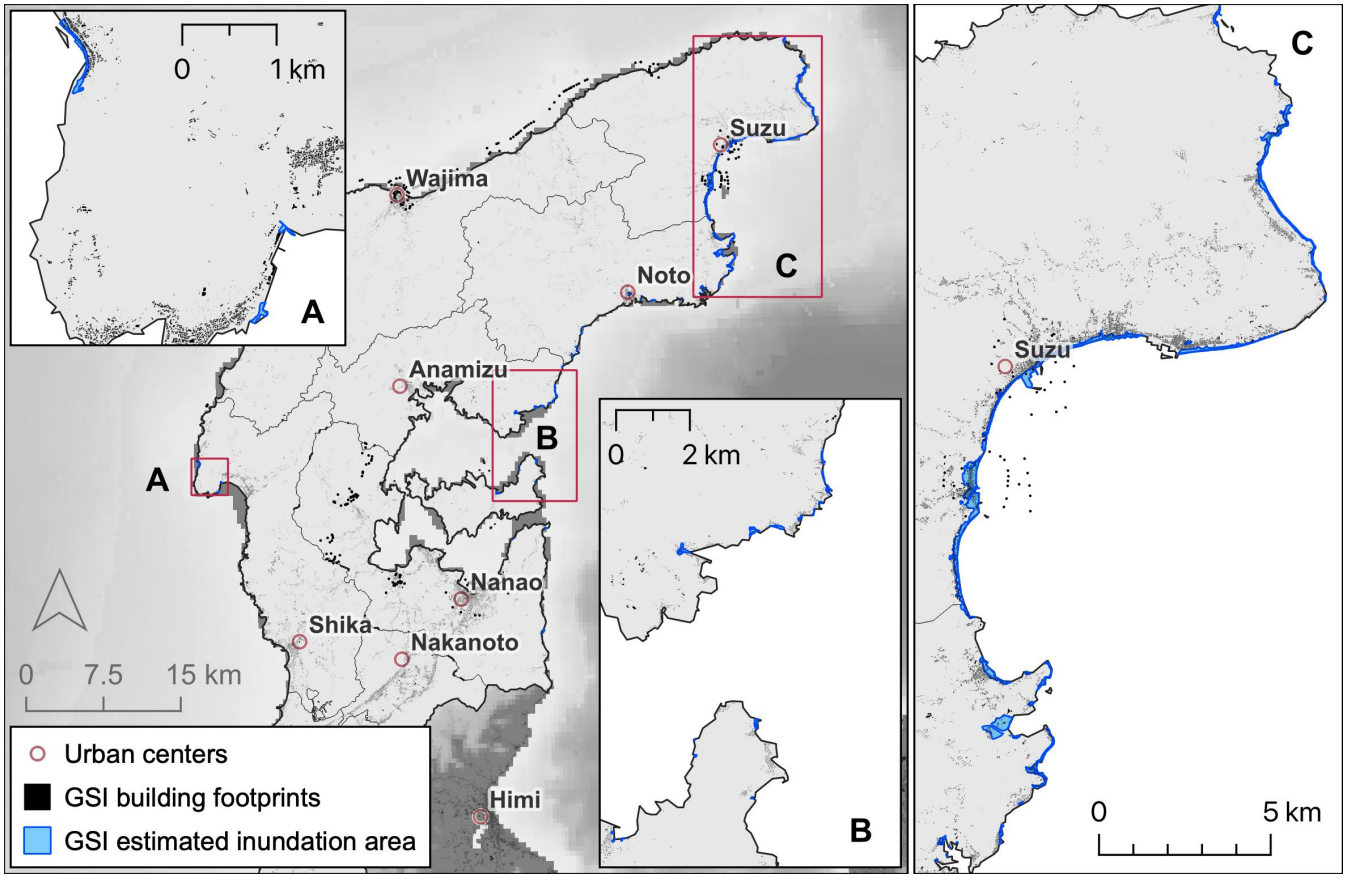


Figure 3. Estimated inundation area provided by GSI; backdrop: GEBCO Bathymetric Compilation Group 2023 (2023).

2024) (https://www.gsi.go.jp/BOUSAI/20240101_noto_earthquake.html, Japanese) at time of writing. Available data includes: ground subsidence and slope failure extents, post-event aerial vertical imagery, tsunami inundation extent estimates, and crustal deformation estimates. From this portal we obtained post-processed vertical imagery xyz tiles for the post disaster period: GSI conducted photographic missions on January 2nd, 5th, 11th, and 17th, covering the whole Noto Peninsula with a significant degree of redundancy to minimize visual obstruction due to atmospheric and environmental effects such a cloud coverage, smoke, sunshade, and snow. Similarly, Kokusai Kogyo (KKC), a for-profit consulting agency specializing in geospatial technology, has been issuing special investigative products free of charge through a Noto Peninsula Earthquake information page (KKC, 2024). Available data includes geotagged high-resolution oblique imagery of specific high-profile areas that were affected catastrophically by the disaster (KKC, 2024); oblique imagery missions were conducted on January 2nd; Oblique images are available through Kokusaki Kogyo's proprietary BOIS portal https://bois-free.bousai.genavis.jp/diarsweb (Japanese). The assessment was supplemented by news sources for select areas of Nanao City, due to visibility issues with the vertical and oblique images (Table 4)(Minami, 2024; Nikkei xTECH, 2024).

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

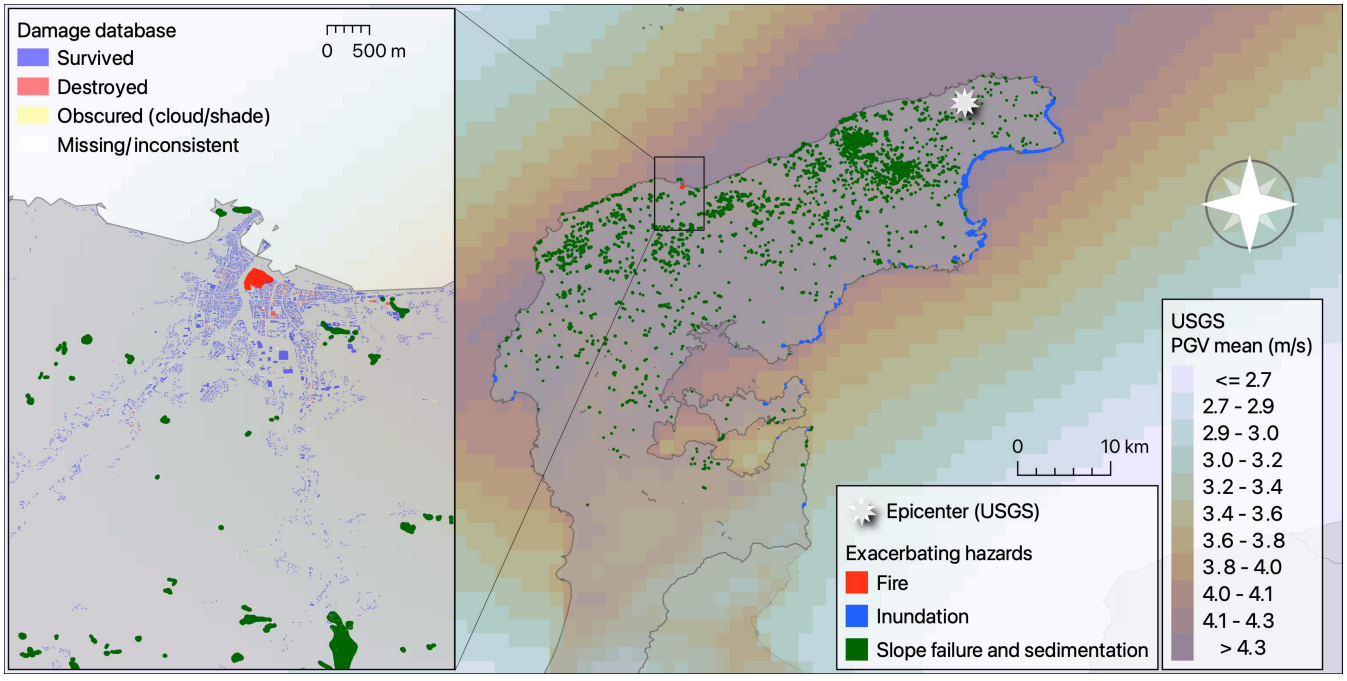


Figure 4. Composite of exacerbating hazards (in addition to seismic impact). Inundation, fire area, and slope failure/sedimentation extents are provided by the GSI (2024). Peak ground acceleration estimates provided by KKE (2024) in collaboration with NIED (2019).

2.2 Basis for the assessment

An initial criteria was drawn up for the tsunami-affected areas by computing the intersection between the GSI footprints and the GSI inundation extent. The building inventory for this portion of the study area originally accounted for 3,261 buildings. Based on the limited scope of the investigation, we conceived an initial "binary+" classification schema that was eventually formalized into the final classification system with minimal adjustment - these classes are defined as reported in Table 1:

Where possible, the visual assessment is supported by oblique imagery, which proved invaluable in many instances. This was especially true for edge cases such as areas with poor visibility, densely packed areas wherein buildings collapsed vertically ("pancake" collapse), or overcast areas. During the initial investigation we posited a multi-class damage assessment, for areas mere multi-source data was available, due to the high reliability of the oblique imagery. However, the limited scope of the obliques and poor weather conditions disallowed a comprehensive assessment (Figure 6) - We provide an approximate equivalence table between classification methods in Table 5. This decision is in part supported by previous findings by Huynh et al. (2014) who mention that crowd-sourcing may yield bias towards "No Damage" and "Destroyed" against middle classes. Since only ∼ % of buildings lie within the viewing angle of oblique imagery a multi-class assessment was deemed less viable. Miura et al. (2020) make a case for the inclusion of blue-tarp covered buildings as a separate class in their deep learning classification framework: they notice that presence of blue-tarp covered structures correlated with moderate-heavy building

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Open Access Earth System Science Data Discussions

Class label	Database value	Criterion
0	Survived	Damage does not appear to affect the bearing mode of the structure - Includes: • Partial damage of the roof requiring replacement or repair. • Buildings in the vicinity of structurally unsound buildings but appear structurally sound. • Undamaged buildings.
1	Destroyed	Structurally unsound based on visual interpretation - Includes: • Partially or completely washed away buildings. • Partially, completely collapsed, or severely inclined buildings. • Partially or completely buried buildings. • Buildings burned to the degree that they are structurally unsound.
9	Obstructed view	Building is marked by a footprint according to the GSI registry but is visually obstructed - Includes: • Buildings under cloud cover. • Buildings under sunshade such that they are indistinguishable from their surroundings. • Buildings under canopy cover such that structural features are indistinguishable.
99	Missing or inconsistent	Buildings that whose GSI registry footprint is significantly inconsistent relative to available imagery - Includes: • Building footprints that do not match an existing building across pre-event and post-event imagery even when allowing for a degree of vertical shift. • Building footprints that demarcate a non-existing building across pre-event and post-event imagery.

Table 1. Criteria used for binary classification of the entire Noto Peninsula building damage visual assessment.

damage classes. Although this was initially considered, there exist mismatches between the GSI provided vertical images. This is observable for segments where overlapping orthophotos, such as Wajima City, are available (A schedule of vertical imagery acquisition missions conducted by GSI is provided in Table 4). Figure 7 highlights the variability in tarp presence between mis-
105 sion dates. In other instances, such as Anamizu-machi, spotty cloud cover makes the identification of tarp-covered buildings particularly challenging. Ultimately, a conservative approach was deemed preferable.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

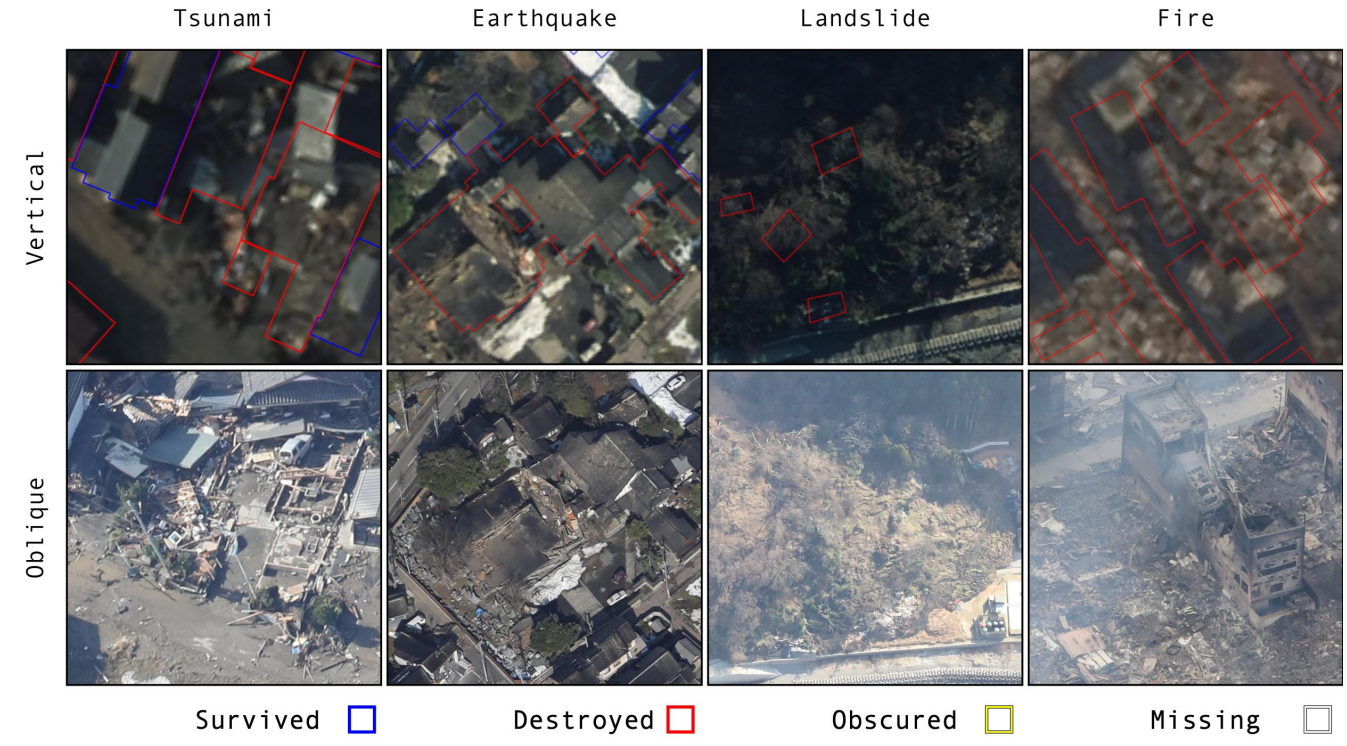


Figure 5. Different modalities of damage consequent to different impacts. Vertical aerial imagery by GSI (2024), Oblique imagery by © KKC (2024)

2.3 Tsunami damage assessment

The tsunami that impacted the eastern coast of the Noto Peninsula was purportedly generated in part by the rupturing of several offshore active faults. In addition, seismic activity may have aggravated submarine landslides in southern Toyama Bay, leading to subsequent tsunami amplification that was ultimately responsible for much of the damage experienced in along the eastern coast of the Noto Peninsula (Masuda et al., 2024). The estimated tsunami inundation extent lies almost entirely along the northeastern coast of the Noto Peninsula area and stretches from the northern most point of the Suzu municipality, to the Nanao municipality in the south. Yuhi et al. (2024a, b) conducted several surveys of the tsunami inundation area and provided comprehensive information on the inundation and run-up heights of the tsunami. On the western coast, only a small extent on the northern portion of the Shika municipality was indicated as inundated (Figure 3). The intersection between the estimated tsunami inundation and the GSI building inventory was the first portion of the damage assessment to be carried out as a preliminary measure. 3,261 were originally included, however significant mismatch exists between the GSI footprints and the orthophoto base map - particularly in non urban areas. Mismatches have been handled as noted in Table 2. There exist cases where buildings were removed in the interim period between the pre-event base map and the post-event GSI aerial; in disaster affected areas it

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

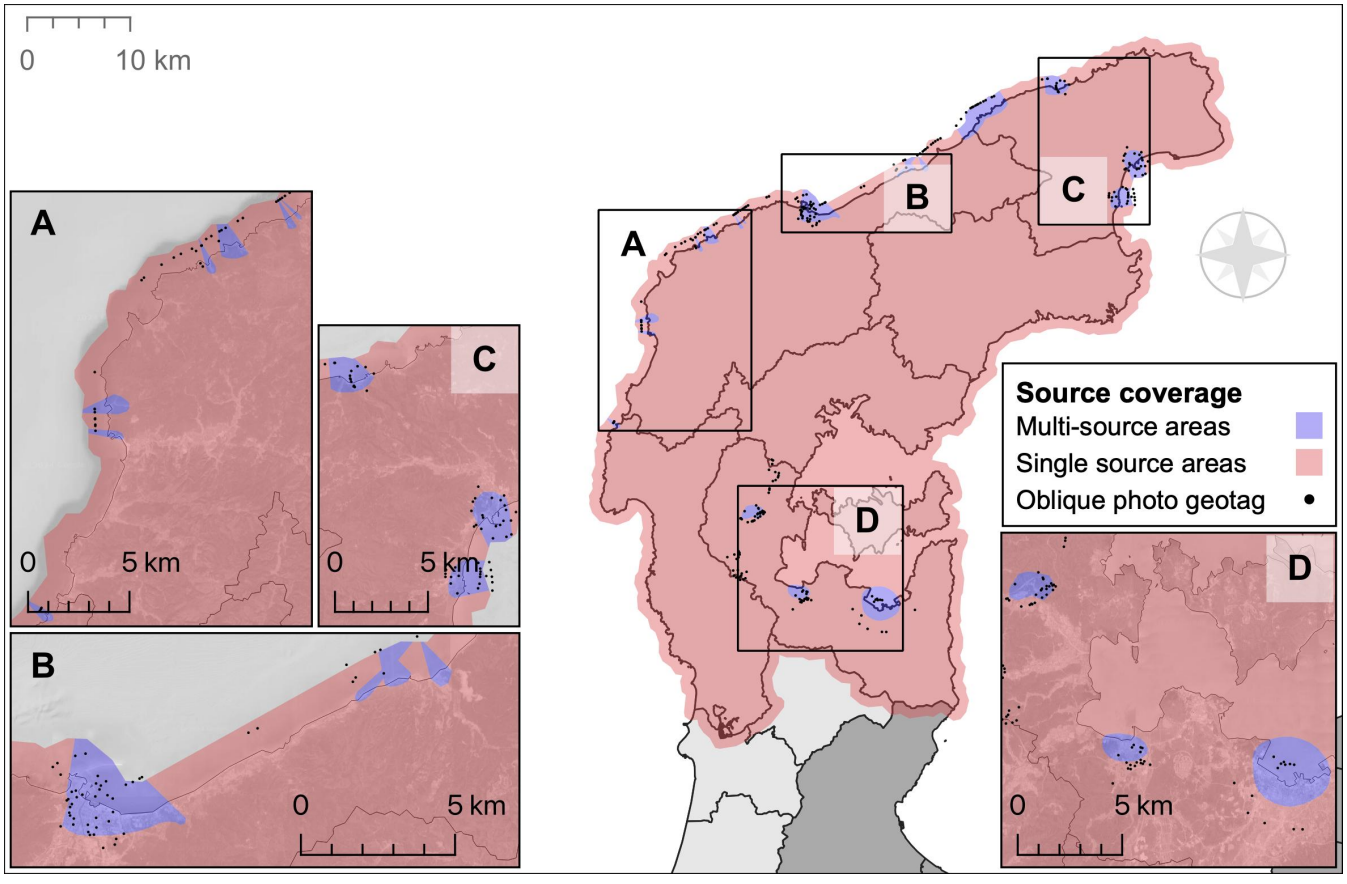


Figure 6. Oblique coverage by Kokusai Kogyo (KKC, 2024) and inherent confidence of visual assessment. Inset basemap © Google 2024

can be hard to distinguish these removed buildings from destroyed or washed away buildings due to debris. The final size of the

Case	Action
Polygon does not reflect the shape in the orthophoto	Adjust (add, split, merge)
Polygon does not appear to correspond to a pre-event or post-event building	Mark building as 99
Polygon does not exist, building is evident on pre-disaster orthophoto	Polygon is manually added

Table 2. Approach to mismatches between footprint polygons and the orthophoto baseline.

120

building inventory for the tsunami affected area stands at 3,400. 139 buildings were digitized based on pre-event orthophotos and subsequently classified. Missing buildings (marked 99) are kept in the inventory for posterity.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Figure 7. Examples of change in blue tarp coverage in Wajima City between aerial imagery capture missions. The figure highlights challenges faced through potential issues in coverage, atmospheric conditions, and source mismatch. Aerial imagery courtesy of GSI (2024).

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

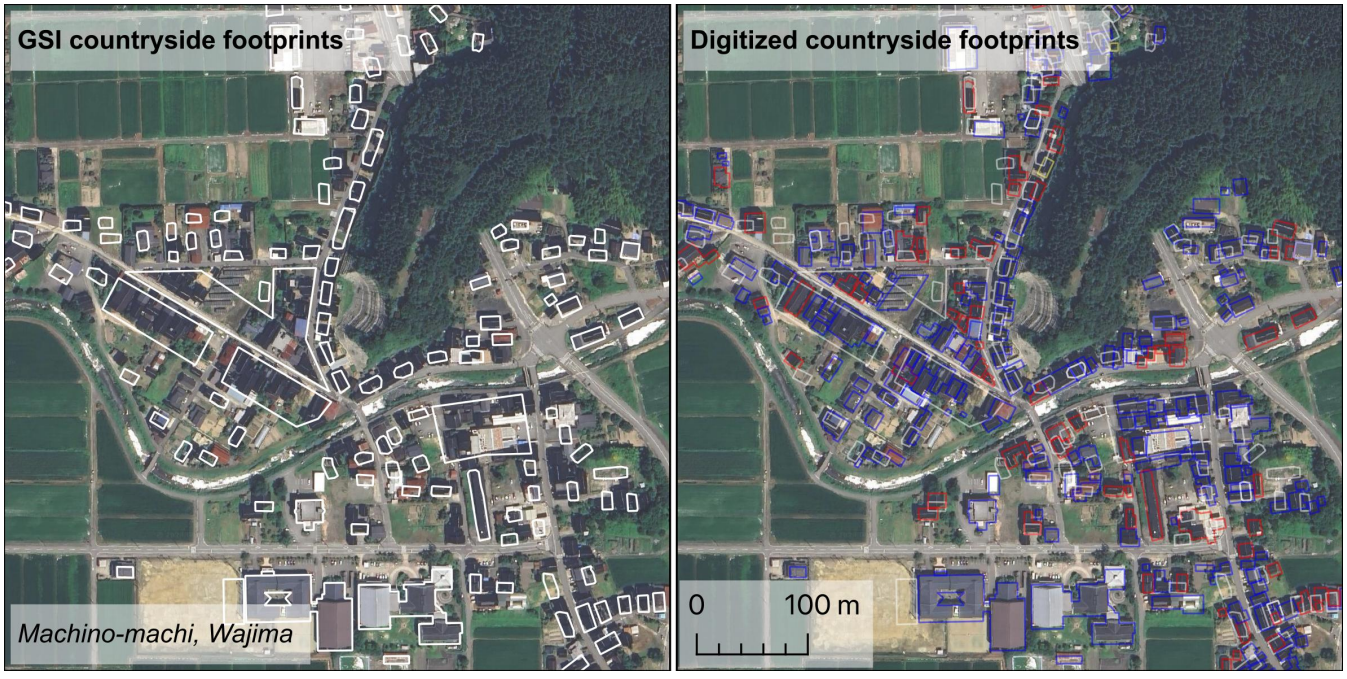


Figure 8. Mismatch between original GSI sourced building footprints and orthophoto imagery prevalent in the countryside. Missing or mismatched buildings were digitized or edited manually before being classified (Basemap attribution: © Google, 2024 Airbus, CNES/Airbus, Japan Hydrographic Association, Landsat/Copernicus, Maxar Technologies 2024).

2.4 Earthquake damage assessment

The scope of the visual assessment was expanded upon completion of the tsunami assessment. The criteria was adjusted to include modes of damage exogenous to tsunami induced failure: including considerations for landslide displacement & burial, fire damage. Moreover, concessions were made for sunshade and buildings under canopy; conditions that seldom affect urbanized coastal settlements where tree cover is diminished. The final damage inventory for the whole domain resolves to 140,208 buildings of which 25,685 were digitized manually. The large proportional disparity in digitized buildings between the tsunami affected areas (4.1%) and the entire domain (18.3%) is largely due to vast portions of building footprints in the countryside being mismatched (Figure 8).

2.5 Crowd sourced feedback

An initial version of the database was made public for viewing on **February 11, 2024** https://experience.arcgis.com/experience/70aae9964dc54e4190b6b360dcbb3759/. The working group encouraged specialist opinions to validate potential errors in the data. The initiative takes inspiration from previous efforts to democratize this process such as Tomnod (DigitalGlobe and formerly GEO-CAN) for cases like Haiti (Zhai et al., 2012) and Christchurch (Barrington et al., 2012; Ghosh et al., 2011).

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

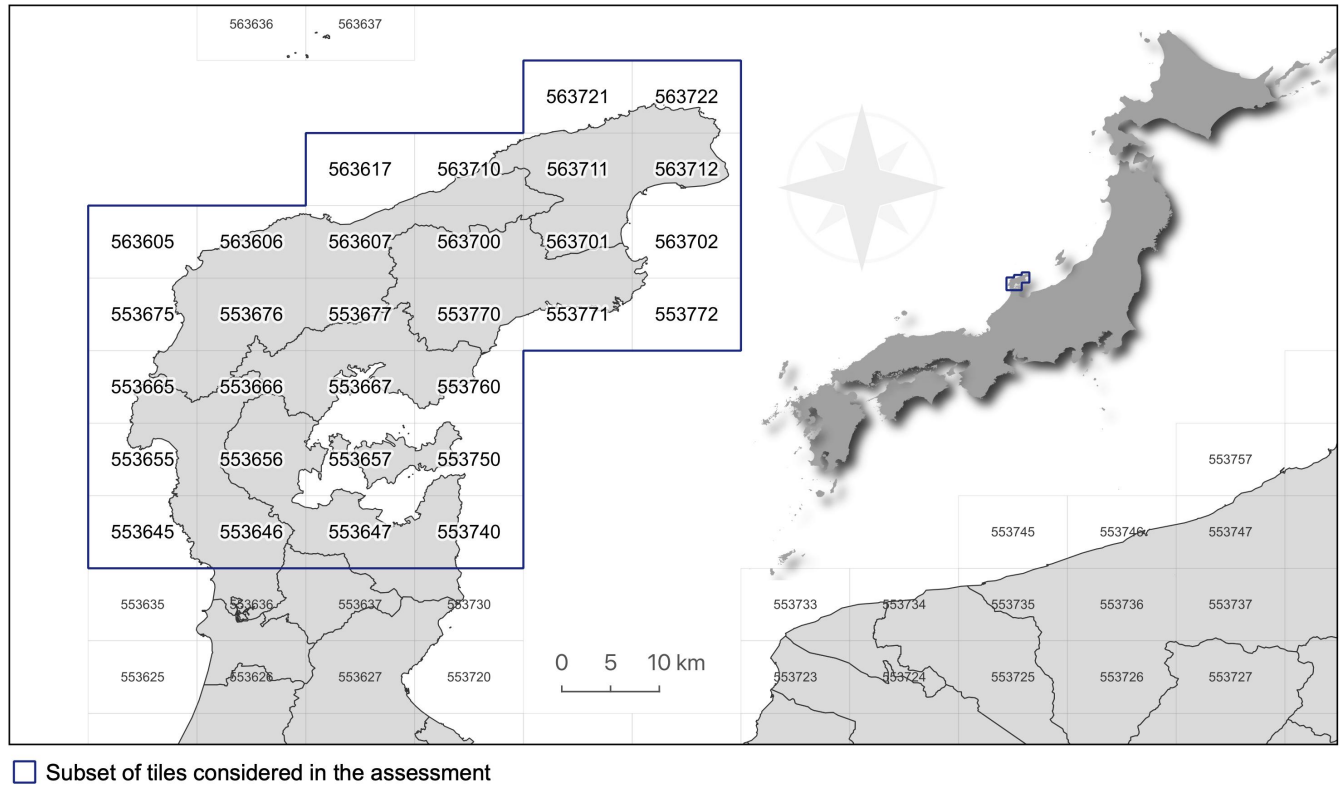


Figure 9. GSI mesh tiles considered for the assessment (GSI, 2024), available at: https://fgd.gsi.go.jp/download/.

Corrections are submitted directly through the website and must include photos for the review process to be formalized. Building damage in Nanao City was particularly challenging due to the a combination of poor exposure, vertical image desaturation, and densely packed houses. In this case we relied on news public reporting (Nikkei xTECH, 2024; Minami, 2024) that included images and location descriptions to identify damaged buildings.

A second set of review information was made available by limited-scope surveys (conducted by research teams) who provided photo evidence to assist the technical validation process. This data informed a quantitative statistical analysis of error margins for the assessment.

3 Data Description

In this section we describe the structure of the dataset, technical notes, attributes, and secondary sources. The database is stored as a GeoPackage (Yutzler, 2024) `Noto_Peninsula_Damage_X_Y.gpkg` (where X and Y are version values). A single layer (vX.Y) with table entries for contents (features) and geometries (`MultiPolygon`) is used to store the building footprints. Details regarding each feature are given in Table 3. A total of **140,208** entries (features + geometries) are included in the dataset. The

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

basis of the analysis was conducted on top of GSI (2024)'s "basic map information" publicly available at https://fgd.gsi.go.jp/download/. The raw data is organized in tiles comprising the standard national mesh defined in JIS X 0410:2002 (Japanese Industrial Standard Committee, 2021). Each tile archive (Provided as: `FG-GML-nnnnnn-ALL-YYYYMMDD.zip`, where `nnnnnn` is the mesh tile number and `YYYYMMDD` is the date of the last update). In our assessment we only consider the `FG-GML-nnnnnn-BldA-YYYYMMDD-0001.xml` files which contain the building polygons. We retain only the `geometry` and the `s_fid` attribute from the original dataset Tiles relevant to the present assessment are given in Figure 9, however we aggregate them into a single table in preprocessing. The dataset uses coordinate reference system (CRS) `EPSG:4326` (WGS 84).

Attribute	Type | Length	Valid entries	Description
`fid`	`Int64`	[1 — 140 208]	Unique identifier for the building (original).
`s_fid`	`String|80`	GSI serialization standard or 'manual'	Serial feature identifier from the original xml file GSI (2024); `manual` when manually added.
`damage`	`Int8`	[0, 1, 9, 99]	Damage class attributed as part of this assessment, as per Table 1.
`damage_val`	`Int8`	[0, 1, 9, 99]	Damage class after technical validation (Section 4) attributed as part of this assessment.
`source`	`String|30`	array or NULL	Oblique image source number from KKC inventory (KKC, 2024) (where available); see Section 6 for access to the KKC image repository.
`municipality`	`String|20`	*Prefecture-City-Town* (Japanese)	Municipality name from e-Stat[a] (Ministry of Internal Affairs and Communications, 2024).
`conf`	`String|10`	[single, multi]	Confidence level of the assessment as per figure 6 based on oblique coverage.
`geometry`	`MultiPolygon`	`MultiPolygon[Polygon(…)…]`	Vector geometry of the building footprint (GSI, 2024).

[a] Available at: https://www.e-stat.go.jp/gis.
Sitemap (JP): トップページ / 統計地理情報システム / 境界データダウンロード
Query tags (JP): 小地域, 国勢調査, 2020年, 小地域(基本単位区)(JGD2011), 世界測地系緯度経度・Shapefile,石川県

Table 3. Details regarding table attributes contained in the GeoPackage dataset.

4 Technical Validation

To guarantee a high degree of consistency across all working members, our classes needed to be as clear-cut as manageable. A potentially useful third class would necessarily need to split the "Survived" class, analogous to the Copernicus EMS

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

scale (CEMS, 2017). Roof damage and displaced rubble are generally the only visible signs of a damage spectrum between ideal "no-damage" and "destroyed" classes, hence may be significant indicators of a potential new class.

Regrettably, timing and weather conditions severely limited the return period of sufficiently clear or redundant vertical imagery (Table 4). Major seasonal pressure systems accompanied modest seasonal snow across the peninsula over the first 3 weeks of the year. The cloud cover and snow buildup effectively impeded the classification of 9,456 buildings (Criteria details are given in Table 1). Notable among the mosaics described in Table 4 is the `Nanao_2024-01-05` mosaic which is particularly faded and unsaturated.

The winter season poses particular challenges to image classification – much of the natural environment tends towards darker, less saturated colors due to a combination of factors: Houses in the Noto Peninsula generally feature traditional black roof tiles. Overcast weather can decrease color saturation by reducing available light (hence reflection). General darkening can reduce the contrast between dark roofs and the environment both cities (concrete & asphalt gray) and the countryside (deciduous vegetation will tend to browns and greys rather than greens). Many of these challenges can hamper the visibility of roof cracks, missing tiles, exposed roof beams, and scattered rubble that may used to distinguish between classification grades.

Ultimately, a binary classification was deemed preferable – a breakdown of how our assessment compares to popular reference scales is given in Table 5. The database was split into working subsections to annotate using single-source or multi-source remote-sensing imagery (Figure 6) – we refer to this process as human annotation (Xia et al., 2023). Once completely classified, each subsection was reviewed by a different team member and integrated into the live database. Our Iterative validation was twofold: Through our open web API, we collected voluntary requests for correction, each submission requiring photographic evidence (Figure 2). Each building for which a correction was submitted was given a new validated damage class (Table 3) with the new classification provided that the submitted evidence conformed to our criteria (Table 1).

Data provided by two independent on-site photographic surveys, respectively Tokoha University and Tohoku University, was used to validate portions of the database similar to how crowd-sourced data was handled. The surveys provide coherent coverage of 4 major settlements: Wajima City, Suzu City, Anamizu, and Monzenmachi (Wajima City); as well as scattered inland rural settlements (Figure 10). Since the data is unbiased with respect to the database (i.e., all damaged buildings were documented along the survey path irrespective of the damage assessment class), the coverage was used to statistically impute the accuracy of human annotation: Each photo was taken from ground level and geo-tagged, forming a dense set of nodes. An approximate path was generated using a range-limited nearest neighbor algorithm. Finally, the intersection between the building database and a 40m buffer (reasonable field of view, assumed from photo inspection) around the paths was taken as the surveyed extent. Our initial labels (human annotation) are taken as the estimates y and measured against the surveyed (corrected) ground truth y , we report standard classification metrics in Table 6. The harmonic F_1-score between survived and destroyed classes is 0.939, suggesting high confidence in the assessment. A spatial representation of the survey coverage is given in Figure 10, notably a large portion of the surveyed areas are outside of multi-source coverage, suggesting that despite the limitations described above, the proposed visual assessment framework is robust. We hope that this exercise in crowd-sourced and survey validation will permit further statistical investigations into the features and limitations of manual image-based rapid building damage visual assessments.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

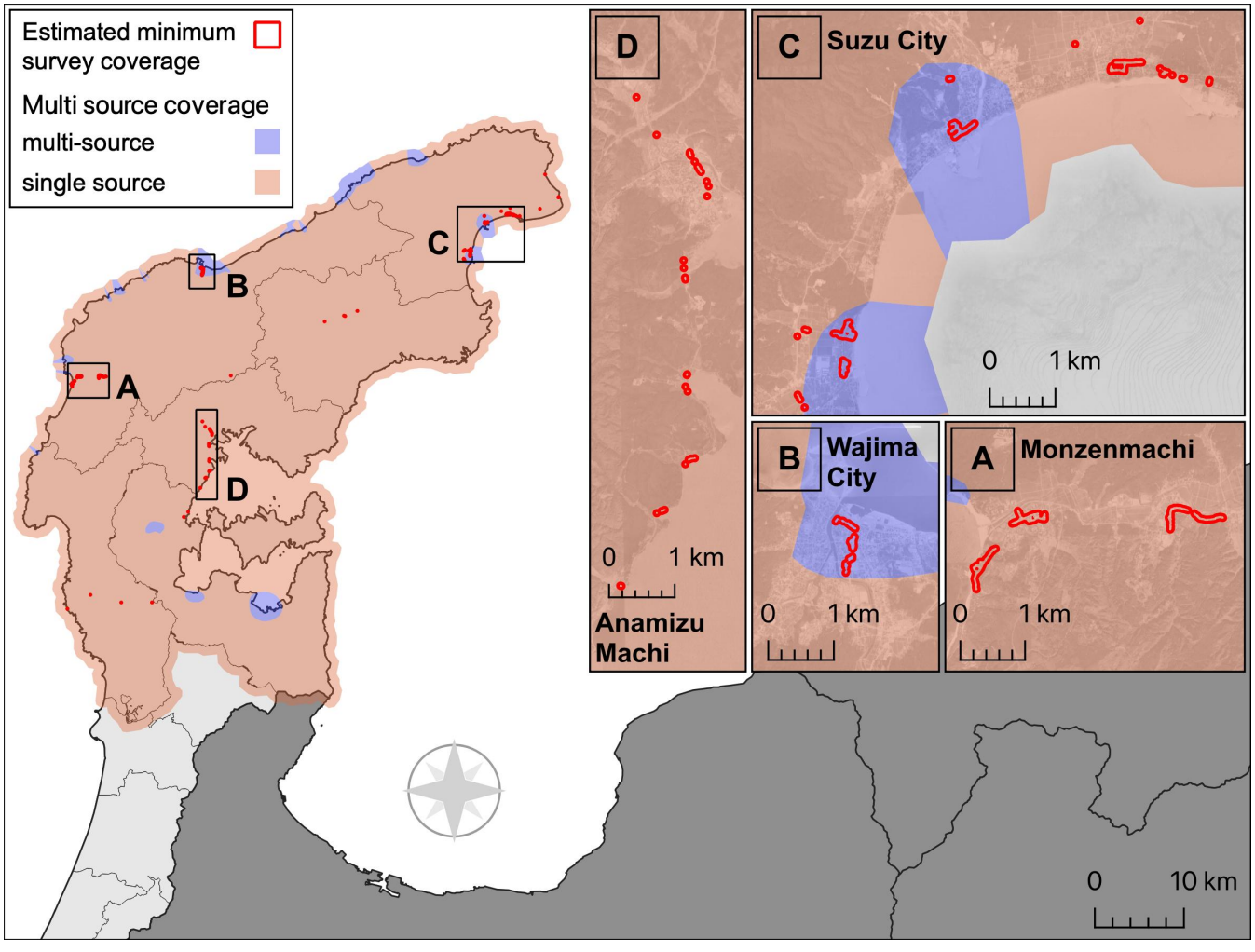


Figure 10. Dataset validation areas are estimated from imagery provided by an independent survey team. A path was fit through the location metadata of each photo. We assume a 40 m range buffer around the path as a reasonable visible area for the survey team, judging by the photographic evidence provided. Inset basemap courtesy of © Google 2024.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Date taken	GSI mosaic name	Notes
2024–01–02	Suzu	modest overcast (east), inland snow buildup (mild)
	Wajima-Naka	mostly overcast, inland snow buildup (modest)
	Wajima-Higashi	minimal overcast (east), inland snow buildup (mild)
2024–01–05	Suzu	mild overcast (west), otherwise clear
	Nanao	minimal overcast (southwest), generally clear, heavy desaturation
	Anamizu	major overcast, central coast clear
2024–01–11	Wajima-Naka	minimal overcast (southeast), snow buildup (modest)
	Anamizu	minimal overcast (center), inland snow buildup (modest)
	Wajima-Nishi	mild overcast (south), snow buildup (modest)
2024–01–17	Nanao	clear, snow buildup (heavy), slight desaturation
	Wajima-Nishi	clear, inland snow buildup (modest), coast snow buildup (mild)
	Anamizu	clear, snow buildup (heavy)

Table 4. Characteristics and date of each GSI vertical mosaic

5 Discussion

The unique nature of the disaster is reflected in its varied impacts on buildings, such as: ground shaking, subsidence, uplifting, tsunami surge, soil liquefaction, landslide, fire, among others (Figure 5). The dataset provides a comprehensive visual assessment of building damage across the Noto Peninsula, including all the aforementioned impacts. With this contribution, we aim to provide a reference for future studies and a benchmark for automated methods.

5.1 On multi-hazard failure modes

The dataset can inform studies that aim to understand the different multi-hazard failure modes given the different impacts listed above — Valentijn et al. (2020) explore multi-hazard damage detection models, but focus on aggregating each hazard discretely by type. However, as figure 13 illustrates, multi-hazard failures not only occur within the same domain, but can present in contiguous sections of the same town. In the figure, we show how earthquake damage is often compounded by fire, landslide, or tsunami damage; in cases of more populated areas, multiple hazards are present at once, as can be seen in Wajima City where fire, landslide, and earthquake damage are all present.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Open Access Earth System Science Data Discussions

Damage Grade (Gruenthal et al., 1998)	Damage Index (Okada and Takai, 1999)	Damage Index acronyms	Copernicus EMS (CEMS, 2017)	Present Study
D0	0.0	Nd0	Possibly damaged	Survived
D1	0.1∼0.2	Md1	Damaged	
D2	0.2∼0.4	Md2		
D3	0.4∼0.6	Ud3, Gd3, Ed3, Rd3, Sd3		
D4	0.6∼0.8	Ud4, Gd4, Ed4, Sd4	Destroyed	Destroyed
D5	0.8∼0.9	Ud5-, Ud5+, Gd5-, Gd5+, Sd5		
	0.9∼1.0	Cd5+		

Table 5. Comparison of popular reference damage scales for building damage visual assessment with approximate relative equivalences.

Class	Precision	Recall	F_1-score	Samples
Survived	0.95	0.99	0.97	1666
Destroyed	0.99	0.84	0.91	559

Table 6. Classification statistics for independently surveyed areas, showing the approximate accuracy of the visual assessment against comprehensive ground documentation.

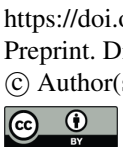
5.2 Machine learning applications

In the field of disaster geo-informatics, our dataset can serve as training data for machine learning tasks. In its current form, the dataset can be used to test pre-trained models such as those proposed by Miura et al. (2020); Deng and Wang (2022); Wiguna et al. (2024a). In this context our dataset offers a new, valuable out-of-domain test set (Wiguna et al., 2024a). A speculative framework, specifically focused on the multi-hazard nature of the Noto Earthquake disaster discussed above, is illustrated in Figure 11. Combined with population data, our database can enable more granular quantitative research into injury and mortality.

5.3 Statistical approaches and baseline model

To stoke the research community's engagement, we provide a statistical baseline of the damage across the non-inundated portion of the Noto Peninsula dataset (Figure 12). We propose an aggregated seismic empirical fragility function relative to

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025



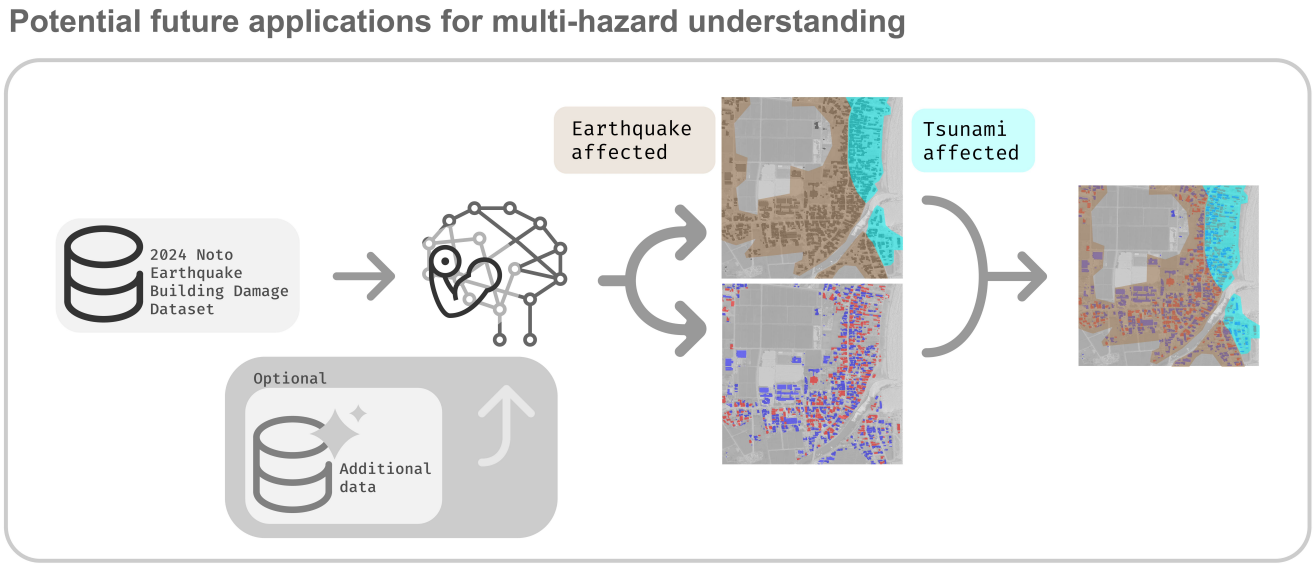


Figure 11. A speculative framework that might be used to investigate multi-hazard failure modes (for illustrative purposes). Basemaps courtesy of © Google 2024.

the Peak Ground Velocity (PGV) registered during the event. Importantly, this fragility function is built on the subset of data that was not affected by aggravating hazards (inundation, fire, or landslide) illustrated in Figure 4. Hence we assume that the damage is solely due to seismic shaking. We fit the aggregated data using a lognormal distribution (Equation 1) and estimate the parameters using ordinary least squares (0 .

$$\Phi \quad \ln \quad \Big) \tag{1}$$

As a frame of reference, we report two fragility functions proposed by Torisawa et al. (2022) for new wooden buildings affected by the Kumamoto Earthquake in 2016. Our baseline fragility function suggests that buildings in the Noto Peninsula were similarly vulnerable to wood buildings built between 2001 and 2016 and destroyed in the Kumamoto Earthquake.

6 Data Availability

The database is provided as a standard GeoPackage (Yutzler, 2024) containing a single vector layer accessible through any software implementing the Geospatial Data Abstraction Library (GDAL/OGR) such as QGIS or ArcGIS. Each entry is represented by a building footprint with 7 attributes summarized in Table 3.

1. The database is available in its most updated version at our public repository at (Vescovo et al., 2024) https://doi.org/10.5281/zenodo.11055712.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

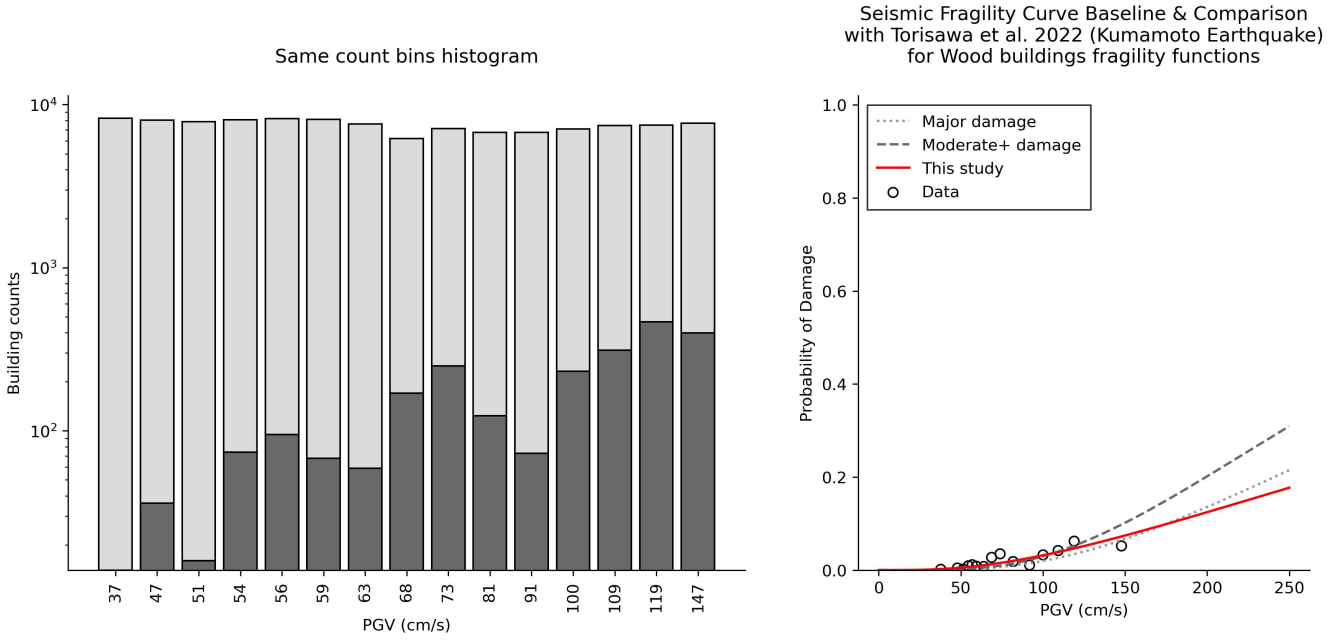


Figure 12. Left: Histogram of aggregated building damage. Right: Empirical fragility function (red solid line) for earthquake-affected buildings relative to PGV. We provide two wood buildings fragility functions for "Major" and "Moderate+" damage classes (respectively dotted line and dashed line) proposed by Torisawa et al. (2022) for buildings built between 2001 and 2016 and affected by the 2016 Kumamoto Earthquake.

2. Epicenter and intensity contours are available at the USGS event page (https://earthquake.usgs.gov/earthquakes/eventpage/us6000m0xl).

3. Earthquake swarm data is available through the Japan meteorological Agency (JMA)'s website (https://www.data.jma.go.jp/eqdb/data/shindo/).

4. Post event raster orthophotography, inundation, fire, and slope failure vector extents are available through the GSI's dedicated Noto Peninsula Earthquake page (GSI, 2024) (https://www.gsi.go.jp/BOUSAI/20240101_noto_earthquake.html, Japanese).

5. Oblique imagery is available through KKC's proprietary BOIS portal (KKC, 2024) (https://bois-free.bousai.genavis.jp/diarsweb, Japanese).

7 Conclusions

We present a comprehensive building damage database for the Noto Peninsula Earthquake of 2024, developed through a multi-source, multi-modal visual assessment of building damage. Our framework allows the dataset to have a much higher degree of

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

accuracy relative a single source approach. The added margin of confidence and singular circumstances of the Noto Earthquake place our dataset in a uniquely valuable position, offering the opportunity to study impacts of multi-hazard disasters on building damage. Figure 13 illustrates how different hazards manifested across contiguous areas of the Noto Peninsula. Understanding the different impacts may provide valuable insights to disaster response and recovery planning. Future studies may leverage our dataset to develop novel multi-hazard models that can predict building damage across different impacts (A speculative framework is shown in Figure 11). With this contribution we hope to enrich the global corpus of disaster building damage datasets. We provide the hand curated building inventory as a GeoPackage through the public repository at https://doi.org/10.5281/zenodo.11055712 (Vescovo et al., 2024). Each building was classified into 4 classes: Survived, Destroyed, Obstructed view through human inspection, and Missing or inconsistent. Limited scope validation was conducted through crowd-sourced community feedback through our online portal and independent survey data conducted by experts in the field. In its immediate form, the dataset may be used to:

- train site specific statistical and machine learning models for building damage assessment.
- test domain adaptation frameworks for building damage assessment by testing pre-trained models on our new out-of-domain dataset as illustrated by Wiguna et al. (2024a).
- fine-tune pre-trained models on our dataset to improve performance across datasets as shown in Wiguna et al. (2024b).
- develop novel multi-hazard models that can predict building damage across different impacts.

In combination with additional data source, such as population data, and post disaster information, our dataset can inform further investigation into disaster logistics, evacuation, injury, and mortality. We hope that this dataset will serve as a reference for future studies on building damage assessment, disaster response, and recovery planning.

Author contributions

R.V. Writing — original draft preparation, Validation, Visualization.
R.V. and S.W. Methodology, Software, Formal analysis, Data curation, Investigation.
CY.H., J.M., X.D., S.I., K.W., and Y.E. Data curation.
B.A., E.M., and A.M.: Conceptualization, Investigation, Supervision, Field Survey, Writing — review & editing.
S.K.: Conceptualization, Supervision, Field Survey, Writing — review & editing, Funding acquisition.

Competing interests. The authors declare that they have no conflict of interest.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

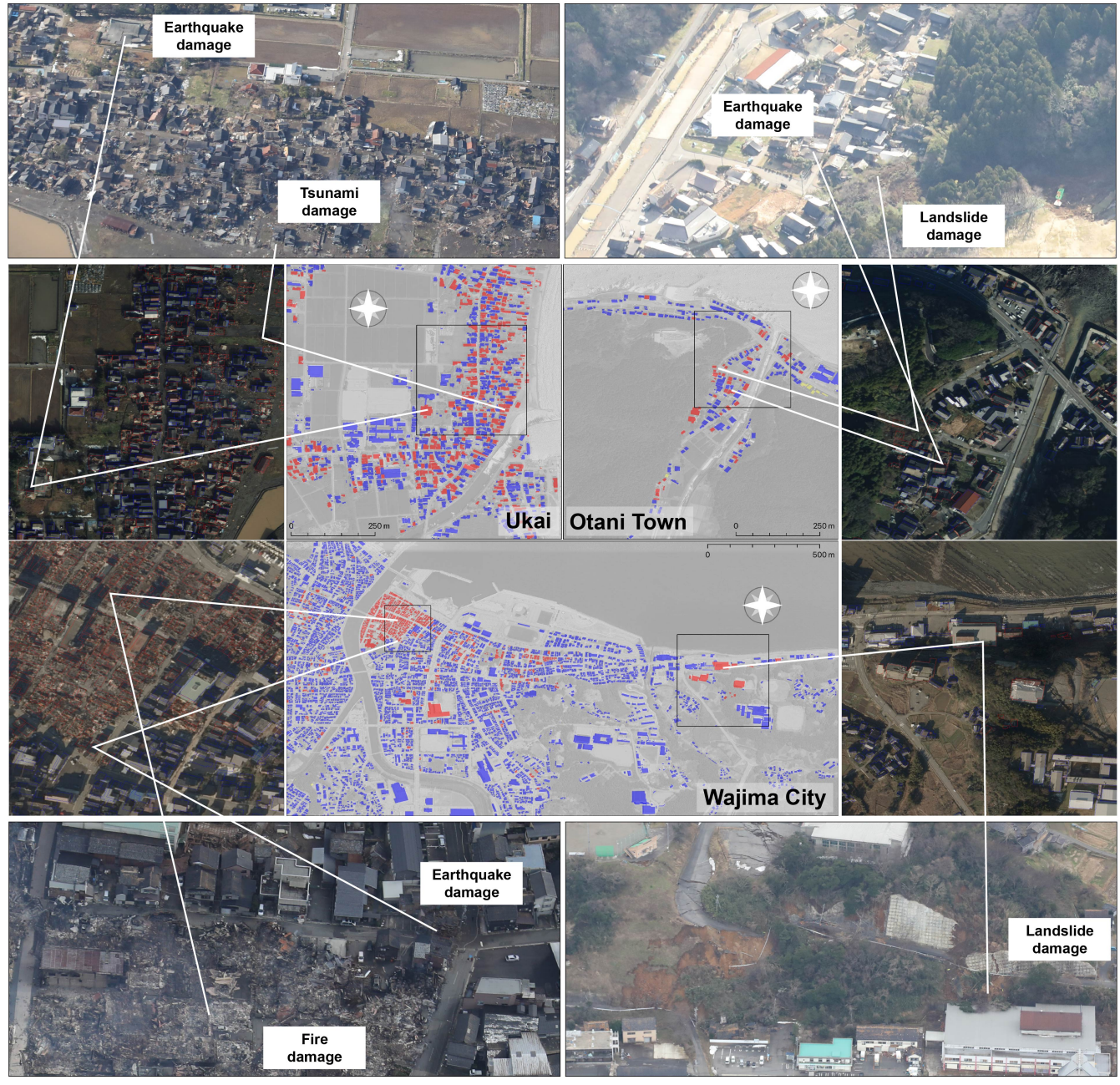


Figure 13. Different impacts across contiguous areas of the Noto Peninsula illustrate how multiple hazards may manifest across a single event with extreme proximity. Basemap courtesy of © Google 2024, Obliques by © KKC (2024).

Acknowledgements. This study was partly supported by JSPS KAKENHI (Grants-in-Aid for Scientific Research, 21H05001, 22K21372, and 22H01741) and the Cross-Ministerial Strategic Innovation Promotion Program (Grant Number JPJ012289). The use of building footprint data is approved by Geospatial Information Authority of Japan (GSI) (R5JHs641).

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

References

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., Yu-Min, A., and Huyck, C.: Crowdsourcing earthquake damage assessment using remote sensing imagery, Annals of Geophysics, 54, https://doi.org/10.4401/ag-5324, 2012.

British Broadcasting Corporation: Japan earthquake: Fires hit quake zone as rescuers race to reach survivors, BBC News, https://www.bbc.com/news/world-asia-67865502, 2024.

CEMS: Copernicus EMS - Rapid Mapping (EMSR) Building damage scale based on Vertical Imagery, Online, https://emergency.copernicus.eu/mapping/book/export/html/138313, 2017.

Chua, C. T., Switzer, A. D., Suppasri, A., Li, L., Pakoksung, K., Lallemant, D., Jenkins, S. F., Charvet, I., Chua, T., Cheong, A., and Winspear, N.: Tsunami damage to ports: cataloguing damage to create fragility functions from the 2011 Tohoku event, Natural Hazards and Earth System Sciences, 21, 1887–1908, https://doi.org/10.5194/nhess-21-1887-2021, 2021.

Deng, L. and Wang, Y.: Post-disaster building damage assessment based on improved U-Net, Scientific Reports, 12, 15 862, https://doi.org/10.1038/s41598-022-20114-w, 2022.

Egawa, S., Ishii, T., Furukawa, H., Fujita, M., Abe, Y., Sakamoto, A., Inaba, Y., Ono, K., Harigae, H., Tsuboi, M., Kuriyama, S., and Sasaki, H.: The 2024 Noto Peninsula Earthquake and the Strategy of Medical Assistance from the Tohoku University Hospital, The Tohoku Journal of Experimental Medicine, 262, 45–49, https://doi.org/10.1620/tjem.2024.J010, 2024.

GEBCO Bathymetric Compilation Group 2023: The GEBCO 2023 Grid - a continuous terrain model of the global ocean and land, https://doi.org/10.5285/f98b053b-0cbc-6c23-e053-6c86abc0af7b, 2023.

Ghosh, S., Huyck, C. K., Greene, M., Gill, S. P., Bevington, J., Svekla, W., DesRoches, R., and Eguchi, R. T.: Crowdsourcing for Rapid Damage Assessment: The Global Earth Observation Catastrophe Assessment Network (GEO-CAN), Earthquake Spectra, 27, 179–198, https://doi.org/10.1193/1.3636416, 2011.

Gokon, H. and Koshimura, S.: Mapping of Building Damage of the 2011 Tohoku Earthquake Tsunami in Miyagi Prefecture, Coastal Engineering Journal, 54, 1250 006–1–1250 006–12, https://doi.org/10.1142/S0578563412500064, 2012.

Gruenthal, G., Musson, R. M. W., Schwarz, J., and Stucchi, M.: European Microseismic Scale 1998. EMS-98, Cahiers du Centre Europeen de Geodynamique et de Seismologie ; 15, https://doi.org/10.2312/EMS-98.full.en, 1998.

GSI: Information on the Noto Peninsula Eartquake, Webpage, Geospatial Information Authority of Japan, https://www.gsi.go.jp/BOUSAI/20240101_noto_earthquake.html, 2024.

Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., and Gaston, M.: Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.

Huynh, A., Eguchi, M., Lin, A. Y.-M., and Eguchi, R.: Limitations of crowdsourcing using the EMS-98 scale in remote disaster sensing, in: 2014 IEEE Aerospace Conference, pp. 1–7, ISSN 1095-323X, https://doi.org/10.1109/AERO.2014.6836457, iSSN: 1095-323X, 2014.

Ishikawa, Y. and Bai, L.: The 2024 Mj7.6 Noto Peninsula, Japan earthquake caused by the fluid flow in the crust, Earthquake Research Advances, p. 100292, https://doi.org/10.1016/j.eqrea.2024.100292, 2024.

Japan Meteorological Agency: Reiwa 6 Noto Peninsula Earthquake Assessment (January 2nd), Tech. rep., Japan Meteorological Agency, https://www.static.jishin.go.jp/resource/monthly/2024/20240101_noto_1.pdf, 2024.

Japanese Industrial Standard Committee: JIS X0410:2002 Regional mesh code, 2021.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Japanese Red Cross Society: Operation Update No.30 : 2024 Noto Peninsula Earthquake: The Japanese Red Cross Society's Response | Emergency Relief, https://www.jrc.or.jp/english/relief/2024NotoPeninsulaEarthquake.html, 2024.

Kato, A.: Implications of Fault-Valve Behavior From Immediate Aftershocks Following the 2023 Mj6.5 Earthquake Beneath the Noto Peninsula, Central Japan, Geophysical Research Letters, 51, e2023GL106 444, https://doi.org/10.1029/2023GL106444, 2024.

KKC: Reiwa 6 Noto Peninsula Earthquake, Tech. rep., Kokusai Kōgyō, https://www.kkc.co.jp/disaster/2024/01/%E4%BB%A4%E5%92%8C%EF%BC%96%E5%B9%B4%E8%83%BD%E7%99%BB%E5%8D%8A%E5%B3%B6%E5%9C%B0%E9%9C%87/, 2024.

KKE: QUIET+, Tech. rep., Kōzō Keikaku Engineering Inc., https://site.quietplus.kke.co.jp/home/, 2024.

Masuda, H., Sugawara, D., Cheng, A.-C., Suppasri, A., Shigihara, Y., Kure, S., and Imamura, F.: Modeling the 2024 Noto Peninsula earthquake tsunami: implications for tsunami sources in the eastern margin of the Japan Sea, Geoscience Letters, 11, 29, https://doi.org/10.1186/s40562-024-00344-8, 2024.

Minami, F.: (Expert Opinion) Noto Peninsula earthquake hits family home in Nanao, Ishikawa Prefecture, Japan Shelters were set up by the residents themselves, https://news.yahoo.co.jp/expert/articles/8ab093df29bbadd1bf5c18e9d64fc581e677904d, 2024.

Ministry of Internal Affairs and Communications: E-Stat: Portal Site of Official Statistics of Japan, Online, https://www.e-stat.go.jp/, 2024.

Miura, H., Aridome, T., and Matsuoka, M.: Deep Learning-Based Identification of Collapsed, Non-Collapsed and Blue Tarp-Covered Buildings from Post-Disaster Aerial Images, Remote Sensing, 12, 1924, https://doi.org/10.3390/rs12121924, 2020.

NASA: Earthquake Lifts the Noto Peninsula, https://earthobservatory.nasa.gov/images/152350/earthquake-lifts-the-noto-peninsula, 2024.

NIED: Nied K-Net, Kik-Net, Tech. rep., National Research Institute for Earth Science and Disaster Resilience, https://doi.org/10.17598/NIED.0004, 2019.

Nikkei xTECH: Wooden houses collapse in rapid succession amid frequent liquefaction, Nanao City, https://xtech.nikkei.com/atcl/nxt/column/18/02706/011100038/, 2024.

Okada, S. and Takai, N.: Classifications of Structural Types and Damage Patterns of Buildings for Earthquake Field Investigation, Journal of Structural and Construction Engineering (Transactions of AIJ), 64, 65–72, https://doi.org/10.3130/aijs.64.65_5, 1999.

Prime Minister's Office of Japan: Press Conference by the Prime Minister on the 2024 Noto Peninsula Earthquake, January 1, 2024, https://japan.kantei.go.jp/101_kishida/statement/202401/01kaiken.html, a.

Prime Minister's Office of Japan: Press Conference by the Prime Minister on the 2024 Noto Peninsula Earthquake, January 2, 2024, https://japan.kantei.go.jp/101_kishida/statement/202401/02kaiken.html, b.

Prime Minister's Office of Japan: Press Conference by the Prime Minister on the 2024 Noto Peninsula Earthquake, January 3, 2024, https://japan.kantei.go.jp/101_kishida/statement/202401/03kaiken.html, c.

Prime Minister's Office of Japan: To those affected by the 2024 Noto Peninsula Earthquake — Information to assist disaster victims, https://japan.kantei.go.jp/ongoingtopics/notoearthquake20240101.html, 2024.

Shelly, D. R.: Examining the Connections Between Earthquake Swarms, Crustal Fluids, and Large Earthquakes in the Context of the 2020-2024 Noto Peninsula, Japan, Earthquake Sequence, Geophysical Research Letters, 51, e2023GL107 897, https://doi.org/10.1029/2023GL107897, 2024.

Torisawa, K., Matsuoka, M., Horie, K., Inoguchi, M., and Yamazaki, F.: Development of Fragility Curves for Japanese Buildings Based on Integrated Damage Data from the 2016 Kumamoto Earthquake, Journal of Disaster Research, 17, 464–474, https://doi.org/10.20965/jdr.2022.p0464, 2022.

United States Geological Survey: M 7.5 - 2024 Noto Peninsula, Japan Earthquake, online, https://earthquake.usgs.gov/earthquakes/eventpage/us6000m0xl/executive, 2024.

https://doi.org/10.5194/essd-2024-363
Preprint. Discussion started: 5 March 2025

Valentijn, T., Margutti, J., van den Homberg, M., and Laaksonen, J.: Multi-hazard and spatial transferability of a cnn for automated building damage assessment, Remote Sensing, 12, 2839, https://doi.org/10.3390/rs12172839, 2020.

Vescovo, R., Adriano, B., Mas, E., Wiguna, S., Mizutani, A., Ho, C. Y., Morales, J., Dong, X., Ishii, S., Ezaki, Y., Wako, K., Tanaka, S., and Koshimura, S.: 2024 Noto Peninsula Earthquake Building Damage Visual Assesment, Tech. rep., Tohoku University, https://doi.org/10.5281/zenodo.11055712, 2024.

Wiguna, S., Adriano, B., Mas, E., and Koshimura, S.: Evaluation of Deep Learning Models for Building Damage Mapping in Emergency Response Settings, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 1–17, https://doi.org/10.1109/JSTARS.2024.3367853, 2024a.

Wiguna, S., Adriano, B., Vescovo, R., Mas, E., Mizutani, A., and Koshimura, S.: Building Damage Mapping of the 2024 Noto Peninsula Earthquake, Japan, Using Semi-Supervised Learning and VHR Optical Imagery, IEEE Geoscience and Remote Sensing Letters, 21, 1–5, https://doi.org/10.1109/LGRS.2024.3407725, 2024b.

Xia, J., Yokoya, N., Adriano, B., and Broni-Bediako, C.: OpenEarthMap: A Benchmark Dataset for Global High-Resolution Land Cover Mapping, in: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 6243–6253, IEEE Computer Society, ISBN 9781665493468, https://doi.org/10.1109/WACV56688.2023.00619, 2023.

Yuhi, M., Umeda, S., Arita, M., Ninomiya, J., Gokon, H., Arikawa, T., Baba, T., Imamura, F., Kawai, A., Kumagai, K., Kure, S., Miyashita, T., Suppasri, A., Nobuoka, H., Shibayama, T., Koshimura, S., and Mori, N.: Post-event survey of the 2024 Noto Peninsula earthquake tsunami in Japan, Coastal Engineering Journal, 0, 1–14, https://doi.org/10.1080/21664250.2024.2368955, 2024a.

Yuhi, M., Umeda, S., Arita, M., Ninomiya, J., Gokon, H., Arikawa, T., Baba, T., Imamura, F., Kumagai, K., Kure, S., Miyashita, T., Suppasri, A., Kawai, A., Nobuoka, H., Shibayama, T., Koshimura, S., and Mori, N.: Dataset of Post-Event Survey of the 2024 Noto Peninsula Earthquake Tsunami in Japan, Scientific Data, 11, https://doi.org/10.1038/s41597-024-03619-z, 2024b.

Yutzler, J.: OGC GeoPackage Encoding Standard, https://www.geopackage.org/spec140/index.html, 2024.

Zhai, Z., Kijewski-Correa, T., Hachen, D., and Madey, G.: Haiti earthquake photo tagging: Lessons on crowdsourcing in-depth image classifications, in: Seventh International Conference on Digital Information Management (ICDIM 2012), IEEE, https://doi.org/10.1109/ICDIM.2012.6360130, 2012.