



**HAL**  
open science

# Data-driven-based models for seismic damage assessment of buildings : using machine learning at a regional scale and in-situ structural monitoring at building scale

Subash Ghimire

## ► To cite this version:

Subash Ghimire. Data-driven-based models for seismic damage assessment of buildings : using machine learning at a regional scale and in-situ structural monitoring at building scale. Earth Sciences. Université Grenoble Alpes [2020-..], 2023. English. NNT : 2023GRALU021 . tel-04370250

**HAL Id: tel-04370250**

**<https://theses.hal.science/tel-04370250>**

Submitted on 3 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : STEP - Sciences de la Terre de l'Environnement et des Planètes

Spécialité : Sciences de la Terre et de l'Environnement

Unité de recherche : Institut des Sciences de la Terre

**Modèles basés sur des données pour l'évaluation des dommages sismiques des bâtiments : utilisation de l'apprentissage automatique à l'échelle régionale et de la surveillance structurelle in situ à l'échelle du bâtiment.**

**Data-driven-based models for seismic damage assessment of buildings: using machine learning at a regional scale and in-situ structural monitoring at building scale.**

Présentée par :

**Subash GHIMIRE**

#### Direction de thèse :

**Philippe GUEGUEN**

DIRECTEUR DE RECHERCHE, Université Gustave Eiffel

Directeur de thèse

**Danijel Schorlemmer**

GFZ German Research Centre for Geosciences

Co-directeur de thèse

#### Rapporteurs :

**Andrea PENNA**

PROFESSEUR, University of Pavia

**Emily SO**

PROFESSEUR, Department of Architecture, University of Cambridge

#### Thèse soutenue publiquement le **5 juin 2023**, devant le jury composé de :

**Philippe GUEGUEN**

DIRECTEUR DE RECHERCHE, Université Gustave Eiffel

Directeur de thèse

**Danijel SCHOERLEMMER**

CHARGE DE RECHERCHE, German Research Center for Geosciences

Co-directeur de thèse

**Andrea PENNA**

PROFESSEUR, University of Pavia

Rapporteur

**Emily SO**

PROFESSEUR, Department of Architecture, University of Cambridge

Rapporteuse

**Vitor SILVA**

PROFESSEUR ASSOCIE, Global Earthquake Model (GEM) Foundation

Examineur

**Jocelyn CHANUSSOT**

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Examineur





## Résumé

Les tremblements de terre sont moins fréquents, mais leurs conséquences représentent un cinquième des catastrophes naturelles. Par conséquent, pour une gestion et une atténuation efficaces des risques sismiques, il est essentiel d'évaluer les dommages potentiels causés à un bâtiment (au niveau du bâtiment) et leur étendue dans un environnement urbain (à grande échelle) avant qu'un tremblement de terre ne se produise. Récemment, des données collectées sur les bâtiments sont devenues librement accessibles, notamment des bases de données d'enquêtes sur les dommages après les tremblements de terre et des enregistrements de données sur les mouvements violents. Ces ensembles de données permettent de tester les nouveaux modèles d'évaluation des dommages sismiques : évaluation rapide des pertes à grande échelle et prévision opérationnelle des pertes dues aux tremblements de terre au niveau des bâtiments sur la base d'une évaluation de la vulnérabilité variable dans le temps. Par conséquent, dans cette étude, nous avons utilisé ces données sur les tremblements de terre des bâtiments pour développer des idées basées sur les données dans l'évaluation des dommages sismiques en se concentrant sur l'évaluation rapide des dommages à grande échelle et l'évaluation des dommages au niveau des bâtiments. Cette thèse est divisée en deux parties. Premièrement, en se concentrant sur l'évaluation rapide des dommages à grande échelle, nous présentons nos conclusions sur les tests des modèles d'apprentissage automatique pour la prédiction des dommages à l'aide d'un ensemble de données de dommages aux bâtiments après le tremblement de terre. Nous avons évalué différentes questions liées aux modèles d'apprentissage automatique pour l'évaluation des dommages à grande échelle, y compris la sélection du modèle d'apprentissage automatique, la relation entre les caractéristiques des bâtiments et le niveau des dommages, l'efficacité des caractéristiques facilement disponibles (telles que l'âge, le nombre d'étages, la surface au sol, la hauteur) dans l'évaluation des dommages à grande échelle, l'efficacité de l'utilisation d'un modèle d'apprentissage automatique formé sur les données des dommages des bâtiments après un tremblement de terre pour prédire les dommages potentiels lors de futurs tremblements de terre, l'efficacité de la méthode d'apprentissage automatique par rapport aux méthodes classiques d'évaluation des dommages, et la transférabilité des modèles d'apprentissage automatique développés dans une région pour évaluer les dommages dans une autre région. Deuxièmement, en nous concentrant sur l'évaluation des dommages au niveau des bâtiments, nous présentons nos conclusions sur l'évaluation de la réponse cosismique des bâtiments pendant la charge sismique à l'aide de l'enregistrement des données sur les tremblements de terre à fort mouvement à partir des bâtiments. Plus précisément, nous avons examiné la relation entre la réponse cosismique d'un bâtiment en fonction des mesures d'intensité du mouvement du sol et les variabilités associées liées à la variation temporelle de l'état structurel due aux dommages induits par le tremblement de terre dans un bâtiment, à la magnitude du tremblement de terre et à la distance, en regroupant les bâtiments par classe et par région. En outre, nous avons évalué la capacité de résistance aux charges latérales d'un bâtiment et sa variation due aux dommages induits par le tremblement de terre et leur accumulation

dans le temps dans un bâtiment spécifique, parmi les bâtiments de la même classe, et les ajustements nécessaires pour tenir compte de l'état structurel réel au cours de la modélisation des dommages variables dans le temps. Enfin, des conclusions générales sont présentées, ainsi que des perspectives de travail.

**Mots-clés:** Évaluation des dommages sismiques dans les bâtiments, apprentissage automatique, efficacité et suffisance de l'IM, courbes de capacité

## **Abstract**

Earthquakes are less frequent, but their consequences represent one-fifth of natural catastrophes. Therefore, for efficient seismic risk management and mitigation, it is crucial to assess potential damage to a building (building-level) and its extent within an urban environment (large-scale) before the occurrence of an earthquake. Recently, data collected from buildings have become openly accessible, including post-earthquake damage survey databases and strong-motion data recordings. These datasets offer an opportunity to test the new models in seismic damage assessment: rapid loss assessment at large-scale and operational earthquake loss forecasting in building-level based on time-varying vulnerability assessment. Therefore, in this study, we utilized these earthquake data from buildings to develop data-driven-based insights into seismic damage assessment focusing in rapid damage assessment at large scale and building-level damage assessment. This thesis is divided into twofold. First, focusing on rapid damage assessment at a large scale, we present our findings on testing machine learning models for damage prediction using post-earthquake building damage dataset. We evaluated different issues related to machine learning models for large-scale damage assessments, including the machine learning model selection, the relationship between building features and damage level, the effectiveness of readily available features (such as age, number of stories, floor area, height) in large-scale damage assessment, the effectiveness of using machine learning model trained on past earthquake building damage data to predict potential damage during future earthquakes, the effectiveness of machine learning method compared to classical damage assessment methods, and the transferability of machine learning models developed in a region to assess damage in another region. Second, focusing on building-level damage assessment, we present our findings on evaluating the co-seismic building response during earthquake loading using the strong motion earthquake data recording from buildings. More precisely, we examined the relationship between the co-seismic response of a building as a function of ground motion intensity measures and the associated variabilities related to the time-variation of structural state due to induced earthquake damage in a building, earthquake magnitude and distance, grouping buildings by class and region. In addition, we evaluated the lateral load-resisting capacity of a building and its variation due to the earthquake-induced damage and its accumulation over time in a specific building, among buildings with the same class, and the adjustments needed to account for the actual structural health required for time-varying damage modeling. Finally, some general conclusions are presented, together with some perspective work.

**Keywords:** Seismic damage assessment in buildings, Machine learning, Efficiency and sufficiency of IM, Capacity curves



## **Acknowledgement**

I would like to express my sincere gratitude to my supervisors Philippe Guéguen and Danijel Schorlemmer. Philippe, thank you for providing me with valuable guidance, feedback, support, and care throughout this exciting journey. Danijel, your advice and support have been instrumental in helping me achieve my research goals.

I would like to express my gratitude to Jocelyn Chanussot, Sophie Giffard-Roisin, and Adrien Pothon for their invaluable advice during the annual thesis evaluation. Thank you for taking the time to provide me with constructive feedback, which has helped me to improve and refine my research.

I am also grateful to the URBASIS-EU project for their financial support and the invaluable experiences gained from my involvement organized during the project.

I extend my heartfelt thanks to the reviewers for their time and effort in reviewing my thesis.

To my family, thank you for your unwavering support and encouragement throughout my academic journey.

Lastly, I would like to thank all my friends for the memorable moments we shared. Your companionship has been a source of strength and inspiration for me.

## **Dedication**

I would like to dedicate this work to my late co-mother Bhawani Ghimire who lost their life in the 2015 Nepal earthquake. My personal experiences during the 2015 Nepal earthquake have been my primary motivation for working in this domain.

# Table of Contents

<b>INTRODUCTION.....</b>	<b>1</b>
<b>1. TESTING MACHINE LEARNING MODELS FOR SEISMIC DAMAGE PREDICTION AT REGIONAL SCALE USING BUILDING-DAMAGE DATASET COMPILED AFTER 2015 GORKHA NEPAL EARTHQUAKE.....</b>	<b>8</b>
1.1 INTRODUCTION .....	9
1.2 DATA .....	11
1.3 METHOD.....	16
1.4 RESULT.....	18
1.4.1 <i>Model selection</i> .....	18
1.4.2 <i>Management of data imbalance</i> .....	22
1.4.3 <i>Feature importance</i> .....	23
1.4.4 <i>Damage prediction using test dataset</i> .....	25
1.4.5 <i>Testing the proportionating of the dataset</i> .....	28
1.5 DISCUSSION .....	29
1.6 CONCLUSION.....	30
<b>2 TESTING MACHINE LEARNING MODELS FOR HEURISTIC BUILDING DAMAGE ASSESSMENT APPLIED TO THE ITALIAN DATABASE OF OBSERVED DAMAGE (DADO) .....</b>	<b>32</b>
2.1 INTRODUCTION .....	33
2.2 DATA .....	34
2.3 METHOD.....	39
2.4 RESULT.....	43
2.4.1 <i>First stage: model selection</i> .....	43
2.4.2 <i>Second stage: issues related to machine learning</i> .....	46
2.4.3 <i>Third stage: application to the whole DaDO portfolio and comparison with Risk-UE</i> ..	52
2.5 DISCUSSION .....	58
2.6 CONCLUSION.....	60
<b>3 TESTING HOST-TO-TARGET TRANSFER OF MACHINE LEARNING MODELS FOR REGIONAL-SCALE DAMAGE ASSESSMENT .....</b>	<b>62</b>
3.1 INTRODUCTION .....	62
3.2 DATA .....	64
3.3 METHOD.....	69
3.4 RESULTS.....	72
3.4.1 <i>Testing efficacy at building level</i> .....	72
3.4.2 <i>Testing efficacy by DG aggregation</i> .....	75
3.5 TESTING HOST-TO-TARGET MACHINE LEARNING MODELS .....	78
3.5.1 <i>Efficacy at building-level between the datasets</i> .....	78
3.5.2 <i>Efficacy at building-class level between the datasets</i> .....	79
3.5.3 <i>Damage probabilities between the datasets</i> .....	81
3.6 CONCLUSION.....	82
<b>4 ANALYSIS OF THE EFFICIENCY OF INTENSITY MEASURES FROM REAL EARTHQUAKE DATA RECORDED IN BUILDINGS.....</b>	<b>84</b>
4.1 INTRODUCTION .....	84

4.2	DESCRIPTION OF IMs AND EDPS .....	87
4.3	DATA AND METHOD .....	89
4.4	RESULTS ON EDP IM .....	93
4.4.1	<i>General trends - <math>\sigma</math></i> .....	93
4.4.2	<i>Variability associated with the tectonic context - <math>\sigma_R</math></i> .....	95
4.4.3	<i>Epistemic uncertainties related to building typology - <math>\sigma_T</math></i> .....	96
4.4.4	<i>Within-building variability associated with earthquake magnitude-distance - <math>\sigma_{MR}</math></i> .....	98
4.4.5	<i>Within-building variability associated with aging - <math>\sigma_A</math></i> .....	100
4.5	SUMMARY .....	101
4.5.1	<i>Building frequency variation and the average response spectral value as an IM</i> .....	103
4.6	CONCLUSIONS.....	104
<b>5</b>	<b>THE VARIABILITY OF EXPERIMENTAL CAPACITY CURVE MODELS FOR EXISTING BUILDINGS AND ITS WITHIN-BUILDING AND BETWEEN-BUILDING COMPONENTS.....</b>	<b>107</b>
5.1	INTRODUCTION .....	107
5.2	DATA .....	110
5.3	METHOD.....	112
5.4	RESULTS.....	114
5.4.1	<i>Within-building variability <math>\phi</math></i> .....	114
5.4.2	<i>Between-building variability <math>\tau</math></i> .....	118
5.4.3	<i>Adjustments to the generic capacity curves</i> .....	120
5.5	DISCUSSION .....	121
5.6	CONCLUSION.....	123
	<b>GENERAL CONCLUSION .....</b>	<b>125</b>
	<b>PERSPECTIVES .....</b>	<b>127</b>
	<b>REFERENCES.....</b>	<b>129</b>
	<b>APPENDIX.....</b>	<b>138</b>



## **Introduction**

Earthquakes are less frequent among natural catastrophes; however, they contribute significantly to the physical and social consequences of natural hazards such as human casualties, economic losses, psychological trauma, social disruption, and environmental destruction.

The global population is increasing, and the proportion of the population living in urban areas is also remarkably increasing. For example, 55% of the global population was living in urban space at the end of 2020, which is anticipated to be 68% by 2050. As a consequence, urban infrastructures are developing at an accelerated rate (UN, 2018). Globally, 16.8% of urban areas are exposed to earthquake-prone areas. Out of 94 earthquake-prone countries, the United States, Iran, Turkey, Mexico, and Japan are the top 5 countries with earthquake-prone urban areas; and Pakistan, China, Iran, India, Afghanistan, Nepal, and Bangladesh are the top 7 countries expected with higher fatalities (He et al., 2021).

Earthquakes' consequences represent one-fifth of annual economic losses, with an annual average of approximately US\$34.7 billion and more than 20 thousand deaths per year between 1900 and 2018 (Daniell et al., 2017; Silva et al., 2019). Out of all the earthquakes that occur worldwide, more than 56% of them result in casualties, and more than 91% of earthquake casualties are due to building collapses (He et al., 2021).

Some of the deadliest earthquakes that occurred in the last 15 years include (Munich Re, IMF, World Bank, EM-DAT disaster database): the 2010 Haiti earthquake, which resulted in an estimated fatalities more than 250,000, left more than 2 million people injured and 1.5 million people homeless, with an estimated economic loss more than US\$ 7.5 billion; the 2011 Tohoku earthquake, which resulted an estimated fatalities more than 20,000, left more than 6,000 people injured and 1.05 million people homeless, with an economic loss more than US\$ 250 billion; the 2015 Nepal earthquake, which resulted in an estimated fatalities more than 8000, left more than 22,000 people injured and 3.0 million people homeless, with an economic loss more than US\$ 1.9 billion; the 2021 Haiti earthquake, which caused an estimated fatalities more than 2500, left more than 12,000 people injured and 30,000 people homeless, with an economic loss more than US\$ 1.6 billion; the latest 2023 Turkey-Syria earthquake, which resulted in an estimated fatalities more than 56,000, left more than 121,000 people injured and 2.7 million people homeless, with an economic loss more than US\$ 109.1 billion, as of 01/04/2023.

A significant number of buildings were damaged during these earthquakes. The building's poor performance against the seismic loading, observed in recent earthquakes, highlights the need to enhance our state-of-art knowledge of the complex response of buildings during seismic loading and to

understand better the occurrence of damages, characterize their nature, and develop strategies to minimize them.

### **(i) Progress in collecting earthquake data from buildings**

The availability of earthquake data, such as accelerometric recordings, parametric values, post-earthquake damage surveys, collected from buildings is the key to understanding the earthquake's impact and necessary mitigation strategies (So, 2022). In recent years, communities worldwide are collaborating to collect earthquake data from buildings in the form of (a) post-earthquake building damage datasets (Dolce et al., 2019; MINVU, 2021; MTPTC, 2010; NPC, 2015; Omoya et al., 2022; Stojadinović et al., 2021), and (b) earthquake data recordings from buildings (Astorga et al. 2020), and share them with open access.

This post-earthquake building damage dataset includes information on the damage sustained by buildings, collected through visual screening, and building parameters describing the structural and built-up environment. Damage surveys were carried out on many buildings covering a wide spatial range. For example, the post-earthquake building damage database of the 2015 Nepal earthquake (NPC, 2015) and 2010 Haiti earthquake (MTPTC, 2010) has more than 700,000 and 300,000 buildings collected from several districts of Nepal and Haiti, respectively. In addition, the database of observed damage in Italy (DaDO), a platform of the Civil Protection Department developed by the Eucentre Foundation, has more than 100,000 buildings collected after several earthquakes of different magnitudes from different regions (Dolce et al., 2019).

Similarly, earthquake recordings in buildings (NDE1.0) released by Astorga et al. (2020) contains 8,250 records of strong motion data collected over the years from 2,737 events and 108 buildings from different designs and regions. The database includes structural engineering demand parameters (EDP) such as drift ratios (relative displacement between top and bottom floors normalized by their height), peak top values of acceleration, velocity, displacement, pre- and co-seismic fundamental frequencies, the ground motion intensity measures (IM) such as peak ground acceleration, velocity, displacement, spectral acceleration, velocity, displacement. In addition, this database contains information on buildings and earthquakes characteristics such as building design, location, magnitude, and epicentral distance.

### **(ii) Methods for assessing damage in buildings**

Understanding potential damage to buildings and its extent within an urban environment during an earthquake event is crucial for decision-makers, emergency planners, insurers, and reinsurers for effective seismic risk management and mitigation in earthquake-prone areas (Bommer and Crowley, 2006; Earle et al., 2010; Riedel et al., 2015, 2018).

A building damage assessment usually involves the definition of the hazards affecting the region of interest (i.e., expected frequency and intensity of the ground shaking), the definition of the vulnerability of the assets exposed to the hazards (i.e., likelihood to suffer damage or loss), and a classification of the buildings in the region as exposure (Silva et al., 2022). The extent of damage to buildings is then calculated as a function of hazard, exposure, and vulnerability (Crowley et al., 2021).

Several methods have been developed for assessing damage in buildings due to earthquakes depending on the level of information, methods, tools required for the assessment, and the objectives to be achieved (Erdik et al., 2011; Kassem et al., 2020). Some of the most advanced methods include:

**The machine learning method**, first introduced by (Riedel et al., 2015), is an emerging robust method capable of handling a large amount of data, superior computational efficacy, ability to handle complex problems, and incorporation of uncertainties for building damage assessment (Chi et al., 2020; Hegde and Rokseth, 2020; Karmenova et al., 2020; Rodriguez-Galiano et al., 2019; Sajedi and Liang, 2020; Salehi and Burgueño, 2018; Sun et al., 2021; Zhang et al., 2018; Zhao et al., 2020; Mangalathu et al., 2020a; Harirchian et al., 2021; Roeslin et al., 2020; Stojadinović et al., 2021). In this method, the input features defining a building portfolio are initially mapped to the damage level (target) using supervised machine learning algorithms to create a damage prediction model. Subsequently, this model is utilized in other building portfolios with the same input features to assess the potential damage.

**The Performance-based earthquake engineering (PBEE) method** is a probabilistic framework designed to improve seismic risk decision-making (Porter, 2003). PBEE works in four stages: hazard analysis, structural analysis, damage analysis, and consequence analysis. In hazard analysis, intensity measures (IM) and their annual frequency of exceedance are defined by probabilistic seismic hazard assessment (PSHA). In structural analysis, the structure's response to a given IM is modeled and expressed in terms of engineering demand parameters (EDPs) such as structural drift, maximal top acceleration, etc. In damage analysis, damage measure (DM) is calculated based on EDP values. Finally, the earthquake's consequences, in terms of repair costs, operability of the structure, and potential economic or human losses for a given DM, is calculated and expressed as decision variables (DVs) on which stakeholders base their decisions. In this framework, the relationship between the EDP and IM is utilized to assess damage in the building when compared with the reference EDP values at the damage-state threshold. Two natures of IM is defined: IM that results in the smallest variability in EDP|IM is defined as efficient IM, and IM that results in EDP|IM conditionally independent of earthquake magnitude and distance is defined as sufficient IM (Luco, 2002).

**The vulnerability modeling method** estimates the potential damage through the crossing of hazard, exposure, and vulnerability/fragility models (Crowley et al., 2021). The hazard model defines the ground-shaking potential in terms of expected frequency and intensity in a region of interest and can be calculated in three different ways: scenario-based, intensity-based, and frequency-based approach (Crowley et al., 2021).

The exposure model defines the built environment in terms of vulnerability class, function, reconstruction cost, and spatial distribution. Exposure modeling involves grouping buildings into a class based on a set of attributes that characterize their vulnerability rather than classifying each building individually (Silva et al., 2022).

The vulnerability and fragility model relates hazards with exposure to determine the likelihood of potential damage. The vulnerability model provides a more comprehensive overall estimate of damage using statistical functions. These functions are derived from post-earthquake damage observations from similar building classes experiencing a similar seismic load or based on experts' opinions (Kassem et al., 2020; Rossetto et al., 2014).

The fragility model estimates the probability of a building class reaching or exceeding a specific damage state as a function of ground motion intensity measures (Martins and Silva, 2021). Fragility models are defined in three different ways: based-on expert opinion (heuristic), post-earthquake damage evaluation data (empirical), structural modeling and response simulation (analytical) (Lallemant et al., 2015).

### **(iii) Recent insights for assessing damage in buildings**

In a recent European Project (RISE, Wiemer, 2018), the **dynamic damage assessment (DDA) framework** was summarized as providing real-time assessment of seismic damage by addressing the two major limitations of the existing vulnerability modeling method: (a) the assumption of constant exposure and (b) the inability of structural vulnerability models to account for damage accumulation in buildings (Wiemer, 2018). DDA method aims to consider the dynamic urban environment by using a time-varying exposure model and time-varying vulnerability models that can capture damage accumulation in buildings during different earthquakes over time (Wiemer, 2018; Nievas; et al., 2023). This framework was not new but compiled several ideas proposed by other authors for time-varying vulnerability and exposure.

First, the global dynamic exposure model is developing as a time-varying exposure model for DDA. It is a novel OpenStreetMap-based exposure model designed to capture the real-time spatial distribution of buildings, as well as their operational function, structural characteristics, occupancy, and valuation in real-time (Schorlemmer et al., 2020). The information related to building is collected explicitly (for example, location, number of floors, function), implicitly (for example, building shapes and positions), and semantically derived data based on expert judgment; and is growing at the rate of ~ 150,000

building data per day (consists of more than 375 million building footprints at the moment) (Schorlemmer et al., 2020).

The time-varying vulnerability modeling both at large-scale (urban or regional) and building-level is achieved through the use of the rapid loss assessment (RLA) and Operational Earthquake Loss Forecasting (OELF) methods, respectively (Wiemer, 2018; Nievas; et al., 2023).

The RLA method provides the first-order quantitative estimation of the damage induced in the buildings within a few hours after an earthquake at a larger scale through a time-varying exposure and vulnerability model (e.g., using machine learning methods as proposed initially by Riedel et al. (2014, 2015, 2018), updating existing vulnerability model with the Bayesian method).

The OELF method calculates the probable damage knowing the residual lateral load resisting capacity of a building after seismic loading. The residual capacity can be accessed through structural health monitoring, which involves continuous tracking of a building's structural state (for example, age, level of degradation, maintenance history, etc.) by measuring the dynamic parameters describing the building (e.g., structural period, (Trevlopoulos and Guéguen., 2018;, Trevlopoulos et al., 2020)).

The **Host-to-target transfer** method is an alternative approach for damage assessment when there is limited data available or when it is too costly due to time and resource-requirement to develop a region-specific damage assessment method or for a specific building. In this case, the model developed in one condition (host) is transferred to another condition (target) by applying certain adjustments. The adjustment is applied from the generic host condition to the generic target conditions and/or target-specific conditions: for example, transferring the damage assessment models from high-seismic areas to moderate-seismic areas with adjustments (e.g., Guéguen et al., 2007; Roca et al., 2006) or in regions with buildings of different seismic performance (e.g., Lagomarsino and Giovinazzi, 2006).

#### **(iv) Key issues**

During emergency situations, it may not be feasible to collect the necessary data for creating and updating the exposure model in relation to the vulnerability model for the RLA method due to time and resource constraints, as this information may not be readily available (Riedel et al., 2015). In this context, machine learning methods can offer a paradigm shift by reasonably assessing damage by relying on readily available data in a cost-effective way (Mangalathu et al., 2020a; Harirchian et al., 2021; Roeslin et al., 2020; Stojadinović et al., 2021). In addition, the machine learning method allows the development of relationships between readily available building features and damage level bypassing the vulnerability modeling step. It can handle a large amount of data so it could easily incorporate the global dynamic exposure model. However, the machine learning method is still in its early stages and requires rigorous testing across various issues such as method, data, effectiveness, and

transferability before it can be implemented in the RLA framework for large-scale damage assessment (Mangalathu et al., 2020a; Harirchian et al., 2021; Roeslin et al., 2020; Stojadinović et al., 2021). Specifically, the following issues need to be addressed:

**Which machine learning algorithm and damage assessment framework is robust for RLA application? What are the most crucial features, and how are they related to damage level? How about the host-to-target transfer, i.e., can machine models developed in one region be transferred between regions, and what parameters influence the model performance? The data are not clean and uniform, so how does that impact the performance of machine learning models? Do they outperform conventional damage assessment methods?**

In PBEE and OELF methods, building response to ground motion intensity measures is typically developed through structural modeling and response simulation of a representative building in a given class, using simple (e.g., single-degree-of-freedom oscillators) to most advanced (e.g., finite element-based methods) modeling methods (Iervolino, 2017; Martins and Silva, 2021; Nievas; et al., 2023). Structural modeling involves significant assumptions about the building's attributes (e.g., materials, design, geometry, foundation, period, damping, etc.). The response simulation uses different natural or synthetic accelerograms from different tectonic regions with or without scaling to obtain the desired building response values (Iervolino, 2017; Martins and Silva, 2021). Such modeling assumptions can have a significant impact on the nonlinear seismic response of a building, potentially limiting our understanding of the true response of structures, and its variability from building to building and from earthquake to earthquake. This can also contribute to epistemic uncertainty in damage assessment (Astorga et al., 2018; Kalakonas et al., 2020; Lestuzzi et al., 2016; Perrault et al., 2013, 2020; Perrault and Guéguen, 2015; Spence et al., 2003). Mainly, it raises the following questions:

**Which ground motion intensity measures (IM) can capture the best building response (EDP) in real-world problems? What about the experimental efficiency and sufficiency of IMs? How does the building response change over time after experiencing damage? Do all buildings with the same class respond similarly to a given ground motion? What features of the building contribute the most in response modeling uncertainties?**

In addition, to advance the OELF method as a reliable tool for DDA, it is crucial to enhance our understanding of earthquake-induced damage and residual capacity of the building, including its variation from building to building and earthquake to earthquake (Wiemer, 2018; Nievas; et al., 2023). Specifically, the following questions need to be answered:

**How does the building's co-seismic capacity change after experiencing damage? How do they vary from building to building? How can we consider the actual structural health required for the OELF method?**

## (v) Organization of the Thesis

This thesis is organized as follows:

**Chapter One** presents a comparative study on the efficacy of different machine learning models for predicting damage using the 2015 Nepal earthquake building damage dataset. This chapter also discusses crucial features for damage prediction, methods to address the imbalanced distribution of target features (damage grade) during model training, and the minimum number of data points required to develop a reasonable damage prediction model. The findings of this study have been published in the Earthquake Spectra journal (Ghimire et al., 2022).

**Chapter Two** presents the findings on testing machine learning models for heuristic building damage assessment at a regional scale using the database of observed damage (DaDO) in Italy. We evaluate the efficacy of machine learning models trained on past earthquake building damage datasets to predict potential damage in future earthquakes. This chapter also explores the contribution of building features to the damage level and compares the damage prediction efficacy of the machine learning model with the Risk-UE method. The finding of this study is under review in the Natural Hazards and Earth System Science journal (Ghimire et al., 2023).

**Chapter Three** compares the damage prediction efficacy of machine learning models developed in three different regions (Nepal, Italy, and Haiti) to test their host-to-target adjustment for damage assessment. The chapter also discusses the impact of a building's characteristic features, such as typology, construction practice, seismic regulations, etc., on transferring machine learning models for damage prediction from host to target dataset.

**Chapter Four** presents the findings of the structural response (EDP) analysis to a given set of intensity measures (IM) using the NDE1.0. This chapter examines the efficiency and sufficiency of IMs, as well as the associated variability related to different factors such as structural health, typology, earthquake magnitude and distance. This chapter also discusses the variation of building frequency in response to earthquake loading. Moreover, a building damage prediction equation is provided for different building typologies. The finding from this study have been published in the Soil Dynamics and Earthquake Engineering journal (Ghimire et al., 2021a).

**Chapter Five** presents the findings on the variability in the co-seismic capacity response of buildings under seismic loading using the NDE1.0. This chapter investigates the variation in the co-seismic capacity of a building before and after the first damage occurs, as well as the accumulation of damage over time and the variation between similar buildings. Additionally, this chapter provides the necessary adjustments to account for the actual structural health required for the OELF method. The finding of this study is currently under review at Bulletin of Earthquake Engineering for possible publication.

Finally, our conclusions are presented, along with perspectives for future work.

# 1 Testing machine learning models for seismic damage prediction at regional scale using building-damage dataset compiled after 2015 Gorkha Nepal earthquake

---

This chapter evaluates the effectiveness of machine learning models for a rapid damage assessment at a large scale using the 2015 Nepal earthquake building damage dataset. We analyze the efficacy of different machine learning models, methods for handling imbalanced target features, input feature's importance, the effectiveness of readily available building features for large-scale damage assessment, and minimum data points required for a reasonable prediction model. Our study's findings are published in Earthquake Spectra.

*Ghimire, S., Guéguen, P., Giffard-Roisin, S., and Schorlemmer, D. (2022). Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake. Earthquake Spectra, 38(4), 2970-2993.*

---

## Abstract

Assessing post-seismic damage on an urban/regional scale remains relatively difficult owing to the significant amount of time and resources required to acquire information and conduct a building-by-building seismic damage assessment. However, the application of new methods based on artificial intelligence, combined with the increasingly systematic availability of field surveys of post-seismic damage, has provided new perspectives for urban/regional seismic damage assessment. This study analyzes the effectiveness and relevance of a number of machine learning techniques for analyzing spatially distributed seismic damage after an earthquake at the regional scale. The basic structural parameters of a portfolio of buildings and the post-earthquake damage surveyed after the Nepal 2015 earthquake are analyzed and combined with macro-seismic intensity values provided by the United States Geological Survey ShakeMap tool. Among the methods considered, the random forest regression model provides the best damage predictions for specified ground motion intensity values and structural parameters. For traffic-light-based damage classification (three classes: green-, amber-, and red-tagged buildings based on post-earthquake damage grade), a mean accuracy of 0.68 is obtained. This study shows that restricting learning to basic features of buildings (i.e., number of stories, height, plinth area, and age), which could be readily available from authoritative databases (e.g., national census) or field-surveyed databases, yields a reliable prediction of building damage (4 features/3 damage grade accuracy: 0.64).

**Keywords:** Machine learning, 2015 Gorkha Nepal earthquake, post-earthquake building-damage data, building damage prediction, regional and urban scale risk assessment

## 1.1 Introduction

Earthquakes may not occur frequently; however, they contribute significantly to the physical and social consequences of natural hazards. From 1990 to 2017, the consequences of earthquakes represent an annual average of approximately USD 34.7 billion (OECD, 2018). Information regarding the estimated extent and spatial distribution of potential seismic damage within a built environment is crucial for decision makers, emergency planners, insurers, and reinsurers (e.g., Bommer and Crowley 2006; Ranf et al. 2007; Earle et al. 2010; Riedel and Guéguen 2018). The level of detail, methods, and tools required for the assessment are conditioned by the objectives to be achieved (Erdik et al., 2011). The estimation of potential damage could be obtained through fragility or vulnerability modelling. Seismic fragility or vulnerability modelling is the crossing of hazard, exposure, and fragility/vulnerability component (Crowley et al., 2019). The exposure-related component provides information regarding the built environment, including vulnerability class, function, reconstruction cost, and spatial distribution. The hazard-related component defines the ground-shaking potential in a region of interest based on a probabilistic seismic hazard assessment or from a deterministic rupture scenario. Finally, the vulnerability/fragility-related component relates hazards with exposure to determine the likelihood of potential damage (Silva et al., 2018). Many advanced empirical methods have been developed for seismic vulnerability assessment (e.g., FEMA 2003; Milutinovic and Trendafiloski 2003; Lagomarsino and Giovinazzi 2006; Guéguen et al. 2007; Hancilar et al. 2010, 2012; Silva et al. 2014), based on field surveys of building parameters in relation to standardized building typologies, which are associated with the vulnerability or fragility functions for a specified seismic intensity measurement. The application of these damage assessment methods for rapid damage assessment in regional or urban scale is still challenging because the acquisition of building features and the application of classical methods is time- and resources-consuming on urban or regional scale.

Over the last decade, substantial progress has been realized in the field of machine learning tools and their applications in various domains. Within the scope of this study, Riedel et al. (2015) demonstrated the capacity of the support vector machine in providing seismic vulnerability assessments at regional and national scales. As the application of conventional vulnerability assessment methods on a large scale requires information that is not readily available, they proposed assessing the ability of available data from national census for a region to estimate building vulnerabilities and modeling seismic damage for specified seismic intensities. More recent studies have demonstrated the efficiency of using machine learning techniques in seismic-risk engineering to solve the aforementioned time and resource issues (Chi et al., 2020; Hegde and Rokseth, 2020; Karmenova et al., 2020; Mangalathu et al., 2020a; Sajedi

and Liang, 2020; Salehi and Burgueño, 2018; Sun, 2019; Zhang and Burton, 2019; Zhao et al., 2020). Xie et al. (2020) summarized the ongoing research on the application of machine learning methods in earthquake engineering, they concluded that the implementation of machine learning in earthquake engineering is still in its early stage and needs further investigations. Limited number of studies are available concerning the post-earthquake damage classification using machine learning (Mangalathu et al., 2020b; Roeslin et al., 2020; Harirchian et al., 2021; Stojadinović et al., 2021). In particular, Mangalathu et al., (2020b) tested the performance of machine learning techniques in a building damage survey performed after the 2014 South Napa earthquake and demonstrated the effectiveness of such tools in interpreting the patterns of seismic damage observed in buildings. Harirchian et al., (2021) studied the performance of different machine learning methods for damage prediction in RC buildings. They observed reasonable damage prediction by machine learning models and suggested further investigation by considering a large number of buildings with more features to define the building vulnerability. Roeslin et al., (2020) explored the performance of different machine learning methods using 2017 Puebla-Morelos earthquake building damage data and suggested further investigation using a large number of buildings in each damage typology.

In fact, the added value of post-earthquake studies based on building-damage observations has long been recognized for its contribution in improving our understanding and validating state-of-the-art seismic risk assessment models (Colombi et al., 2008; Spence et al., 2009; Eleftheriadou and Karabinis, 2011, 2012; Guðmundsson, 2012; Karababa and Pomonis, 2011; Del Gaudio et al., 2017, 2020). In parallel to these machine learning developments, open access to a significant amount of information describing real-estate portfolios has improved (Crowley et al., 2020), and post-earthquake building-damage surveys are now available online (2015 Nepal Earthquake: Open Data Portal, 2021; Dolce et al., 2019; Loos et al., 2020). For example, the National Planning Commission of Nepal shared a massive data survey of damaged buildings after the  $M_w$  7.8 Nepal earthquake in 2015 (NPC, 2015a).

To prepare for the increased use of machine learning for earthquake and seismological engineering applications, combined with the increasing number of open-source data, independent studies on independent datasets and with different machine learning methods are essential to advance science in this emerging field. Unlike previous studies (e.g., Mangalathu et al., 2020b; Roeslin et al., 2020), this study herein compares the damage prediction performance of various machine learning models to explain post-earthquake seismic damage, considering two different classifications of damage and two different sets of building features, applied to the 2015 Nepal earthquake building-damage portfolio (NBDP). This dataset is first described in the data section (1.2). In the method section (1.3), we present a summary of the selected machine learning tools applied to the dataset. In the results section (1.4), we present and discuss the results, considering damage assessment as the target value of this study. Finally, we provide our conclusions and perspectives in the last section (1.5) and (1.6).

## 1.2 Data

On April 25, 2015, a devastating  $M_w$  7.8 earthquake struck central Nepal, with an epicentral distance of approximately 80 km NW of Kathmandu, a hypocentral depth of 8.2 km, and a 120 km long rupture toward the east. Level VIII epicentral intensity was estimated in Nepal based on the observed damage (Martin et al., 2015). Thousands of residential buildings were damaged, resulting in 8,790 fatalities and 22,300 injuries (NPC, 2015). Additionally, 31 among 75 administrative districts of Nepal were affected, with 14 districts being severely affected and declared as crisis-hit areas, and an estimated loss of around seven billion US dollars was reported (NPC, 2015). The government of Nepal conducted a massive post-earthquake survey in the 11 most severely affected districts (Fig. 1.1), excluding the Kathmandu Valley. The survey comprised a visual screening of damaged buildings by experts to map the damage in each district and to develop the NBDP database. The NBDP database comprised information regarding 762,106 buildings, each characterized by socio-economic status, engineering properties, and damage. The damage grades were classified into five levels (5DG) that are generally used in post-earthquake surveys following the EMS-98 damage classification system.

**Damage Grade 1 (DG1):** Thin cracks in walls and falling of plaster or loose stones from the upper part of the building, few architectural repairs required.

**Damage Grade 2 (DG2):** Cracks, falling of plaster or stones in many sections, damage to non-structural parts such as chimneys, and projecting cornices, with no significant reduction in the load-bearing capacity of the building.

**Damage Grade 3 (DG3):** Large, extensive cracks and collapse of a small portion of non-load-bearing walls. Detachment of roof tiles, tilting or falling of chimneys, failure of individual non-structural elements such as partitions/gable walls, and delamination of stone/adobe walls. Partial reduction in the load-bearing capacity of structural members, significant repairs required.

**Damage Grade 4 (DG4):** Large gaps or collapse of walls and partial structural failure of floors/roofs, resulting in the building classified as dangerous.

**Damage Grade 5 (DG5):** Complete or near collapse.

The color scale in Fig. 1.1 highlights the proportion of buildings tagged with DG5 in each district. Additionally, a set of building parameters related to structure and environment, frequently used in empirical seismic vulnerability assessment, was assigned to each screened building for the NBDP database:

**Numbers of stories:** total numbers of floors above the ground surface.

**Height of building:** total height of the building above the ground surface measured in feet.

**Building age:** calculated from the date of construction to the date of the earthquake.

**Plinth area:** total area occupied by the building at ground floor level in square feet.

**Ground slope condition:** ground surface topography at the building location, considering three types of ground slope conditions, i.e., flat, mild slope, and steep slope.

**Roof type:** three types of roofs are considered based on the material used, i.e., light timber/bamboo roof, heavy timber/bamboo roof, and reinforced concrete roof.

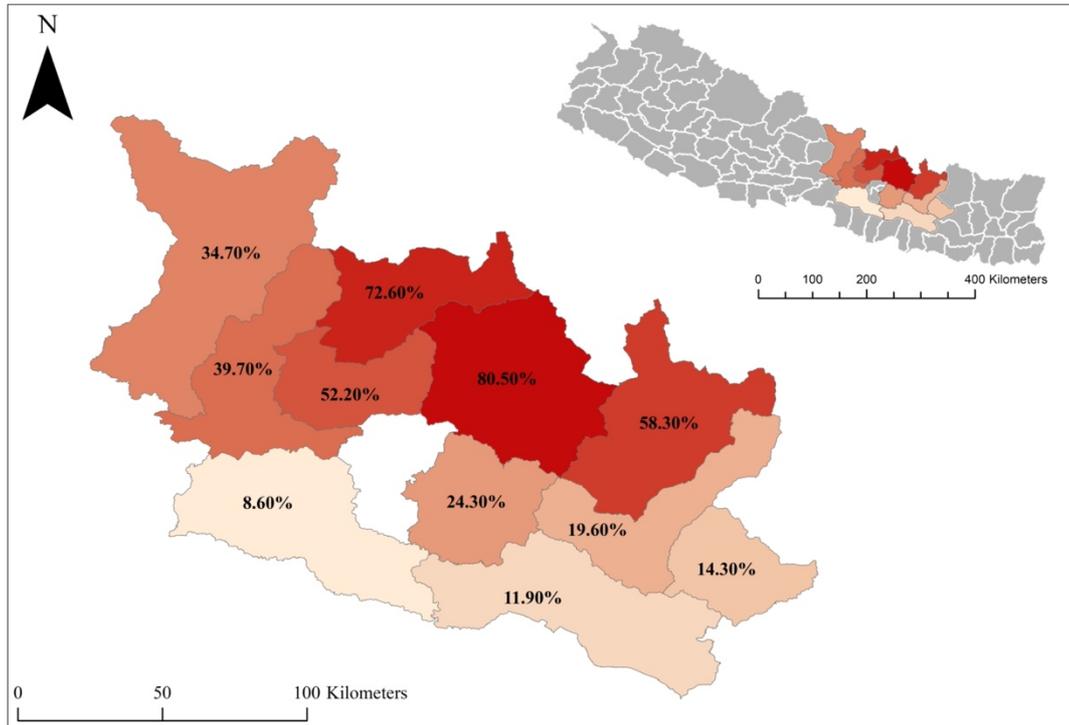
**Position of building:** indications of the building's location relative to other buildings. Four types of positions are cited, i.e., stand-alone, attached on one side, attached on two sides, and attached on three sides.

**Construction material:** Eleven types of construction materials used in the superstructure are considered: adobe, stone flange, mud-mortar stone, mud-mortar brick, cement-mortar stone, cement-mortar brick, reinforced concrete non-engineered, reinforced concrete engineered, timber, bamboo, and others.

Furthermore, in the NBDP database the geographic location of each building was assigned to a ward that belonged to a district. The ward was an elementary administrative cell. The total number of wards in 11 districts was 949. Considering building-by-building locations and the associated soil conditions, which may vary within a ward, can improve the damage prediction performance of a machine learning model (Mangalathu et al., 2020b; Roeslin et al., 2020; Stojadinović et al., 2021). However, having as objective the large-scale classification of damages (and not the classification and representation per building), and considering the macro-seismic intensities, the attribution of each building within a ward is consistent with our approach.

In a first case, all the building features contained in the NBDP database are considered in the learning phase, without any preconceived idea on their significance or possible cross-correlation. Their importance score will then be evaluated in order to know their contribution in the learning phase. In a second case, a subset of input features assumed easy to collect without significant expertise or available via authoritative national or global open databases are considered to test their damage prediction performance, without taking into account their importance.

No specific data cleaning methods were applied to the NBDP database. In the NBDP database, 12 buildings had missing damage information so they were removed from the database and 5,731 buildings had missing age value so they were replaced by the average age value.



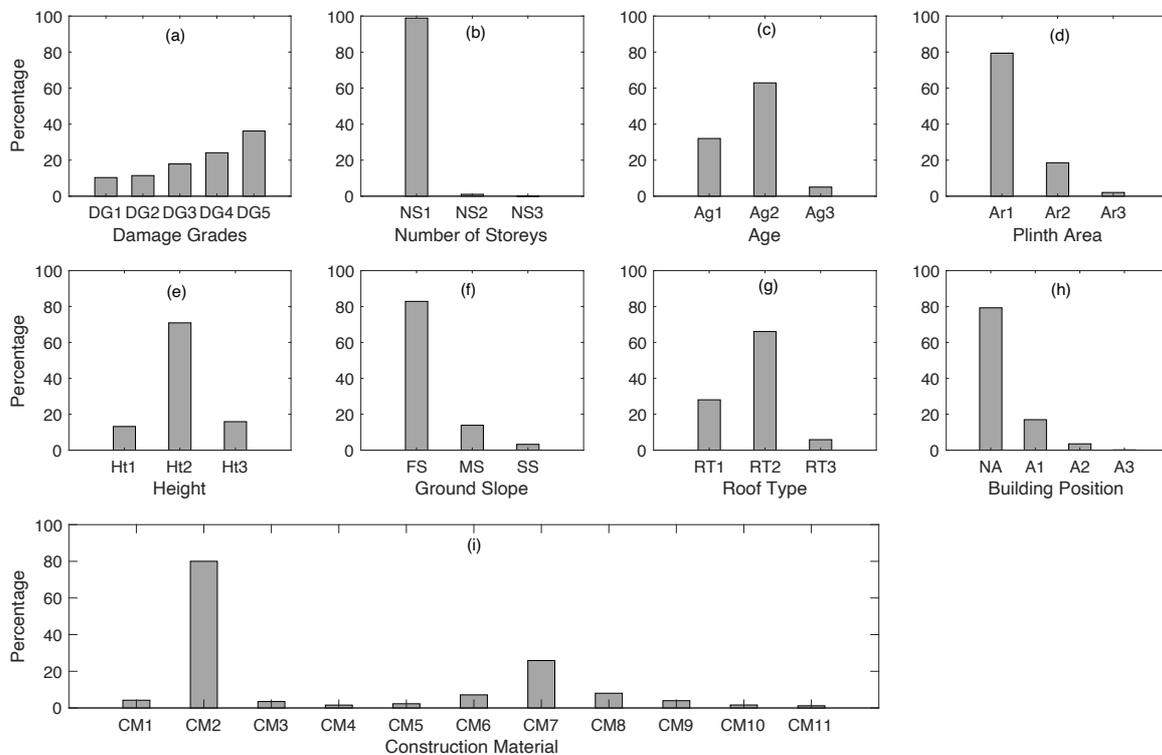
**Figure 1.1** Location of the studied area and the proportion of DG5 tagged buildings (percentage indicated and highlighted by the color scale) in the 11 districts surveyed by the Nepalese authorities (NPC 2015a).

Table 1.1 summarizes the distribution of the building parameters in the database. In the NBDP database, the distribution of the buildings per damage grade was as follows: 10.34% for DG1, 11.45% for DG2, 17.90% for DG3, 24.12% for DG4, and 36.19% for DG5. The number of stories ranged from one to nine floors, the building age ranged from 1 to 200 years, the building plinth area ranged from 70 to 5,000 square feet, and the building height was between 6 and 97 feet (approximately 2 to 30 m). The distribution of buildings based on the material used to construct the superstructure was as follows: 4.24% adobe, 80.02% mud-mortar stone, 3.51% stone flange, 1.58% cement mortar stone, 2.29% mud-mortar brick, 7.15% cement-mortar brick, 25.87% timber, 8.04% bamboo, 3.97% reinforced concrete (RC) non-engineered, 1.63% RC engineered, and 1.20% other materials.

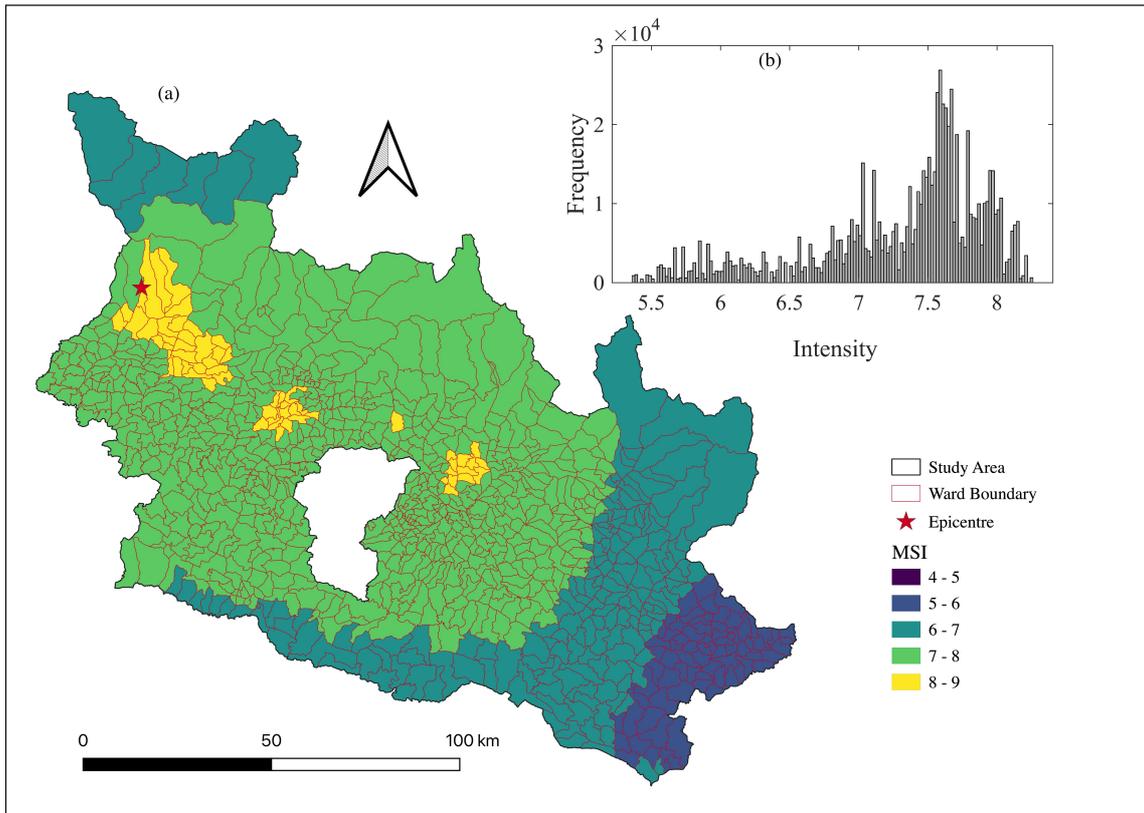
**Table 1.1** Distribution of different building parameters present in the NBDP database.

S.N.	Parameters	Data type	Distribution (%)	Remarks	
1	Damage Grades (DG)	Categorical	DG1	10.34	Fig. 1.2a
			DG2	11.45	
			DG3	17.90	
			DG4	24.12	
			DG5	36.19	
2	Number of stories	Numerical	0–3 (NS1)	98.9	Fig. 1.2b
			3–5 (NS2)	1.04	
			> 5 (NS3)	0.06	
3	Age (years)	Numerical	0–10 (Ag1)	32.03	Fig. 1.2c
			10–50 (Ag2)	62.92	
			> 50 (Ag3)	5.05	
4	Plinth area (square feet)	Numerical	0–500 (Ar1)	79.39	Fig. 1.2d
			500–1000 (Ar2)	18.54	
			> 1000 (Ar3)	2.07	
5	Height (feet)	Numerical	0–10 (Ht1)	13.19	Fig. 1.2e
			10–20 (Ht2)	70.93	
			> 20 (Ht3)	15.88	
6	Ground slope at building's location	Categorical	flat slope (FS)	82.89	Fig. 1.2f
			mild slope (MS)	13.87	
			steep slope (SS)	3.24	
7	Roof type	Categorical	bamboo/timber heavy roof (RT1)	28.05	Fig. 1.2g
			bamboo/timber light roof (RT2)	66.10	
			reinforce concrete (RT3)	5.85	
8	Building position	Categorical	stand-alone (NA)	79.32	Fig. 1.2h
			attached to one side (A1)	16.98	
			attached to two side (A2)	3.53	
			attached to three side (A3)	0.17	
9	Construction material	Categorical	adobe (CM1)	4.24	Fig. 1.2i
			mud-mortar stone (CM2)	80.02	
			stone-flange (CM3)	3.51	
			cement-mortar stone (CM4)	1.58	
			mud-mortar brick (CM5)	2.29	
			cement-mortar brick (CM6)	7.15	
			timber (CM7)	25.87	
			bamboo (CM8)	8.04	
			non-engineered reinforced concrete (CM9)	3.97	
			engineered reinforced concrete (CM10)	1.63	
			others (CM11)	1.20	

Additional information provided by the ShakeMap tool from the United States Geological Survey (USGS) supplied the macro-seismic intensity (MSI) field and completed the database (Wald et al., 2005). ShakeMap provides an estimate of the spatial distribution of earthquake ground shaking intensity by combining recorded ground motions, modeled ground motions, Did You Feel It? information, and slope- $v_s30$ -based site conditions. In recent studies requiring spatially distributed information as input ground motion, ShakeMap intensities have been successfully used to compensate for insufficient instrumental data (Jaiswal and Wald, 2010; Mak and Schorlemmer, 2016; Silva and Horspool, 2019; Del Gaudio et al., 2020; Pothon et al., 2020). The USGS ShakeMap provides the MSI value in modified Mercalli intensity scale (Fig. 1.3a). Because the geo-localization of the buildings is available at the ward level, we computed the mean MSI value for each ward and assigned it to all the buildings located within the same ward. The MSI values ranged from 5.30 to 8.30 (Fig. 1.3b).



**Figure 1.2.** Distribution of building parameters in the whole NBDP dataset. The y-axis shows percentage and the x-axis is (a) damage grade, (b) number of stories (NS1: 0 - 3, NS2: 3 - 5, NS3: > 5), (c) age (Ag1: 0 - 10, Ag2: 10 - 50, Ag3: >50), (d) plinth area (Ar1: 0 - 500, Ar2: 500 - 1000, Ar3: > 1000) (e) building height (Ht1: 0 - 10, Ht2: 10 - 20, Ht3: > 20, in ft as contained in the original database), (f) ground slope condition at building location (FS: flat slope, MS: mild slope, SS: steep slope), (g) roof construction material (RT1: heavy bamboo/timber roof, RT2: light bamboo/timber roof, RT3: reinforced concrete), (h) position of building with respect to other buildings (A1: attached on one side, A2: attached on two sides, A3: attached on three sides, NA: stand-alone building), and (i) superstructure construction material (CM1: adobe, CM2: mud-mortar stone, CM3: stone-flange, CM4: cement-mortar stone, CM5: mud-mortar brick, CM6: cement-mortar-brick, CM7: timber, CM8: bamboo, CM9: RC non-engineered, CM10: RC engineered, CM11: other).



**Figure 1.3.** (a) Spatial distribution of the macroseismic intensity for the  $M_w$  7.8 earthquake in 2015 obtained from the USGS ShakeMap tool and (b) distribution of the macro-seismic intensity in the database, the x-axis is the intensity value and the y-axis is the frequency.

### 1.3 Method

In this study, damage prediction was considered first as a classification problem. Machine learning classification involves assigning a label (or class) to categorical response variables. Damage grades (from 1 to 5) are considered categorical variables; hence, the application of classification models is recommended. In this study, we focused on two typically used classification machine learning methods, i.e., random forest classification (RFC) (Breiman, 2001) and gradient boosting classification (GBC) (Friedman, 1999). However, the damage grades were ordered and can be regarded as a continuous variable to minimize misclassification errors. In this study, regression models were considered simultaneously with the damage grade as a continuous variable ranging between 1 (DG1) and 5 (DG5). Because the regression model outputs a real value between 1 and 5 and not an integer, we rounded the output (real number) to the nearest integer to plot the confusion matrix. However, the error matrices were computed without rounding the model outputs to the nearest integer. Four regression machine learning models were considered, i.e., linear regression (LR) (Hastie et al., 2009), support vector

regression (SVR) (Cortes and Vapnik, 1995), gradient boosting regression (GBR) (Friedman, 1999), and random forest regression (RFR) (Breiman, 2001). Using classification methods, every misclassification yields the same penalty loss, whereas regression methods penalize the relative penalty loss (i.e., difference error) by minimizing the squared loss between the true and predicted damage grades. In this study, we tested both approaches (regression and classification), and their strengths and weaknesses in terms of damage prediction were compared.

RFC/RFR and GBC/GBR are based on a set of several decision trees as base learners; this enables them to achieve greater efficiency while integrating the complexity and nonlinear interaction of the features in the dataset. LR has simple analytical and computational properties; additionally, it can easily illustrate the interpretable description of the effect of the input on the output, i.e., the contribution of each feature in the model. SVR is effective in high-dimensional spaces and is extremely versatile in terms of application (i.e., it can be used for linear and polynomial models).

Hereinafter, the damage grade is defined as the target variable, and the building parameters (i.e., number of stories, height, age, plinth area, ground-slope condition, building position, roof type, and construction material) and the ground-motion intensity (MSI) are defined as the input features. A one-hot encoding technique was used to transform categorical values (i.e., construction material, position, ground slope condition, and roof type) into nominal values (Pedregosa et al., 2011). The one-hot encoding technique generates a unique binary value 1 or 0 for each categorical feature by creating a sparse matrix, which resulted in 26 input features for the model (Table 1.1).

The building damage dataset was randomly partitioned into three subsets. The training dataset constituted 60% of the entire dataset and was used to train the machine learning model. Meanwhile, the validation dataset constituted 20% of the entire dataset and was used to select the best model by comparing the strengths and weaknesses of the different machine learning models for damage prediction (Section 1.4.1). Model optimization was performed by investigating data imbalance arising from the unequal distribution of the target features in the training dataset (Section 1.4.2) and the importance of each feature in the model (Section 1.4.3). The test dataset constituted the remaining 20% of the entire dataset; it was kept hidden from the model until the final optimized model was developed. Once the optimized model was developed, the test dataset was used to test the performance of damage prediction of the optimized model (Section 1.4.4). The percentage distribution of each feature in the training, validation, and test dataset follows the same percentage distribution as in the entire NBDP database (Fig. 1.2).

In this study, we used the models provided in Python using the Scikit-learn package. The performance of each model was quantified by measuring the accuracy score (percentage of correctly predicted labels) (Pedregosa et al., 2011). A high accuracy score (approximately 1) indicates the high efficacy (i.e., ability to perform a task to a satisfactory or expected degree) of the machine learning model in damage

prediction. In addition, we computed the following statistical error indicators to compare the performance of each machine learning model: the mean of the absolute error (MAE) and the mean squared error (MSE). The scores of these indicators are referred to as the performance scores hereinafter. The smaller the MAE (approximately 0) and MSE (approximately 0), the higher is the efficacy of the models. We considered different metrics because we observed that one metric could not fully explain the predictive performance among classification and regression methods. For each machine learning model, the predictive performance is extremely sensitive to the model input parameters, which are known as hyperparameters. The hyperparameters of each machine learning model were tuned using a grid-search technique on the training set. Finally, computational time (time taken by each model in our personal computer, Macbook Pro 2019) is considered as one additional parameter to facilitate comparison among machine learning models.

The performance of each machine learning technique for damage prediction is presented graphically based on the distribution of errors as well as in a confusion matrix (see Fig. 1.4). A confusion matrix presents a comparison between the true and predicted damage grades. In the confusion matrix, the value of the predicted damage grades (cell value along the row of the confusion matrix) is normalized by the total number of true damage grades. The values along the main diagonal elements of the confusion matrix indicate the recall (i.e., the number of predicted values equal to the true values) of the machine learning model. A perfect model would have ones along the main diagonal of the confusion matrix, and zero everywhere else. In the confusion matrix, the accuracy score represents the number of correctly predicted instance (sum of the diagonal terms) over the total number of instances (sum of the full matrix). Because the regression model outputs a real value between 1 and 5 and not a label, we rounded the output (real number) to the nearest integer to plot the confusion matrix. But the error matrices were computed without rounding the model outputs to the nearest whole integer.

## **1.4 Result**

### **1.4.1 Model selection**

This section summarizes the performance of damage prediction provided by the machine learning models on the validation set. The input variables, performance scores, and relative computation time for each optimized model are listed in Table 1.2 and Appendix A. Within the classification model group, the GBC (MAE = 0.60) provided the best performance as compared with the RFC (MAE = 0.67). The GBC and RFC yielded MSE scores of 1.05 and 1.25, respectively. The GBC (Fig. 1.5a) and RFC (Fig. 1.5b) provided better classification recall for the two extreme damage grades (DG1 and DG5) than for the intermediate damage grades (DG2, DG3, and DG4). For these models, the intermediate damage grades were associated with a significant misclassification, independent of the classification model. However, this result shows that the GBC and RFC models did not yield high efficacies in damage

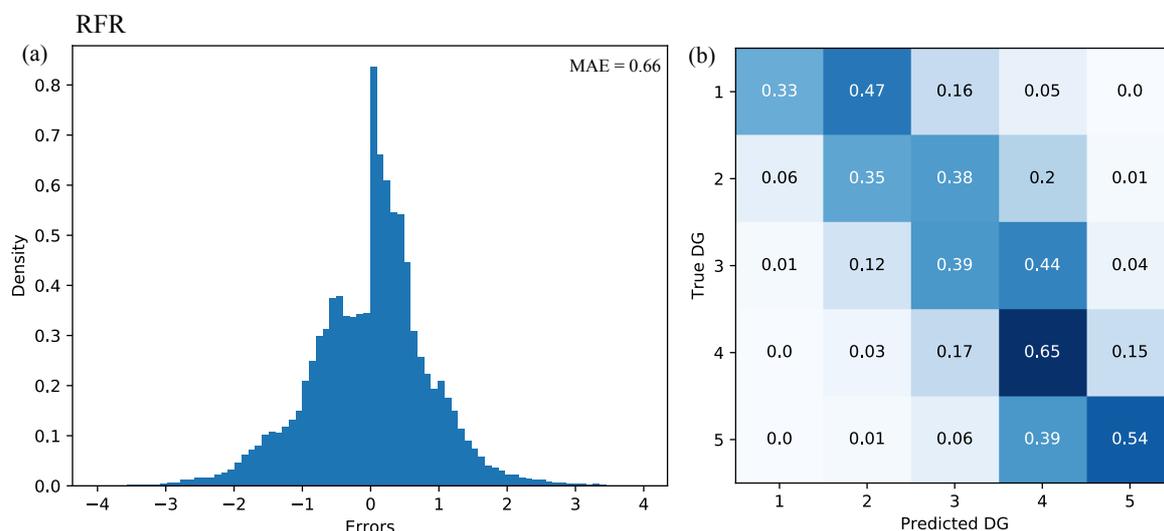
prediction. This may reflect the fact that, unlike DG1 (slight damage) and DG5 (near or complete collapse), the intermediate damage grades are more subjective as well as more difficult to classify and distinguish in the field.

The regression models, GBR (MAE = 0.64) and RFR (MAE = 0.66), demonstrated similar performance results, with the smallest values of MSE (GBR = 0.72; RFR = 0.76). For the RFR (Fig. 1.4b) and GBR (Fig. 1.5c), the predicted damage grades were similar to the ground-truth damage grades in the confusion matrix. High percentage values were observed on the diagonal of the matrix or offset by one level of damage, and they constituted 80% to 95% of the total prediction in each damage grade, as shown in Figs. 1.4b and 1.5c. As expected, the regression models penalized the large errors more efficiently than the GBC (Fig. 1.5a) and RFC (Fig. 1.5b). In addition, these methods are associated with low computational time (RFR: 563 s; GBR: 875 s, Tab. 1.2). It is noteworthy that the LR (MAE = 0.86) and SVR (MAE = 0.77) demonstrated worse performances, with a significant misclassification (LR: Fig. 1.5d and SVR: Fig. 1.5e) that might be due to the complexity and nonlinear feature interactions in the dataset.

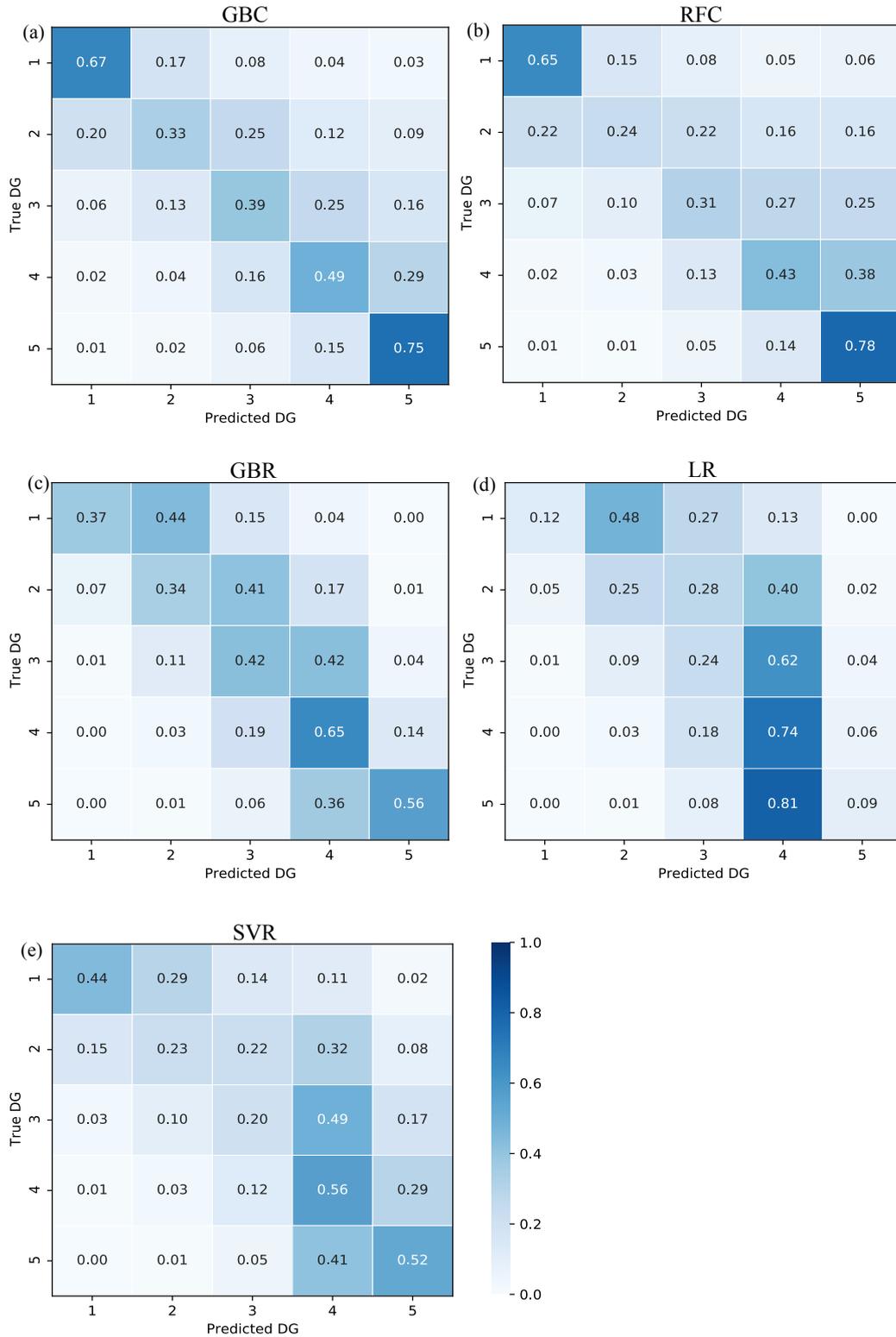
In conclusion, considering the damage-prediction task as a regression problem and using GBR/RFR machine learning models provides the best damage-prediction efficacy (see Fig. 1.4, 1.5 and Table 1.2). For the GBR and RFR models, the damage prediction efficacy, and computational time were extremely sensitive to the hyperparameters. For example, the GBR model required a careful tuning of a significant number of hyperparameters, which increased the computational time compared with the RFR model. In addition, the GBR model might be less generalizable to new data owing to its possible overfitting issue and is more difficult to implement compared with the RFR model. Hence, the RFR model was used in the remainder of this study to test the damage prediction.

**Table 1.2.** Summary of the optimized input parameters and the performance scores of different machine learning models observed on the validation dataset. The values of the input parameters are the hyperparameters for the machine learning methods (the other hyperparameters not mentioned here are the default ones in the Scikit-learn documentation (Pedregosa et al., 2011)). The best model for each measure is indicated in bold.

Machine learning model	Input parameters	Score			Computational time (sec)
		Classification Accuracy	Statistical MAE	MSE	
LR	Default	0.30	0.86	1.14	<b>3</b>
SVR	C = 1 Kernel = rbf	0.43	0.77	1.09	80900
GBR	learning rate = 0.1	0.51	0.64	<b>0.72</b>	875
	no_estimators = 500 max_depth = 10				
RFR	no_estimators = 1000	0.49	0.66	0.76	563
	max_depth = 20				
GBC	learning rate = 0.1	0.57	0.60	1.05	8584
	no_estimators = 1000 max_depth = 10				
RFC	no_estimators = 1000 max_depth = 30	0.53	0.67	1.25	450



**Figure 1.4.** Graphical representation of the predictive performance of the random forest regression (RFR) on the validation dataset. (a) Distribution of errors between the true and predicted damage grades (DG) and corresponding MAE value. (b) DG normalized confusion matrix for the RFR model. The model outputs were rounded to the nearest whole integer to plot the confusion matrix, which is not the case to compute the error matrices.

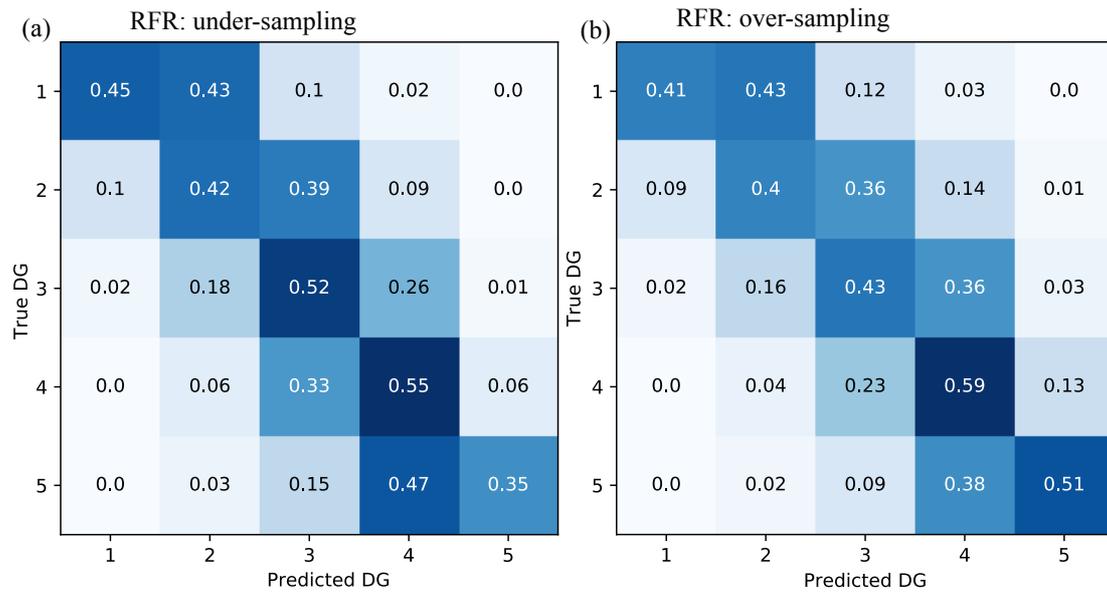


**Figure 1.5.** Damage grade (DG) normalized confusion matrix for (a) gradient boosting classification (GBC) model, (b) random forest classification (RFC) model, (c) gradient boosting regression (GBR) model, (d) linear regression (LR) model, and (e) support vector regression (SVR) model on the validation dataset. A color bar shows the normalized cell value for each confusion matrix.

## 1.4.2 Management of data imbalance

The distribution of labels in the training dataset affects the performance of machine learning models (e.g., Estabrooks and Japkowicz 2001; Japkowicz and Stephen 2002; Branco et al. 2017). The NBDP dataset shows an unequal distribution of the damage grades, i.e., the highest damage grade represents the largest fraction of the dataset (Fig. 1.2a). This data imbalance may have affected the predictive performance of the RFR model (Fig. 1.4). In this study, we considered some extensively adopted techniques to manage an imbalanced dataset: over and undersampling the target features (e.g., Estabrooks and Japkowicz 2001; Japkowicz and Stephen 2002). Undersampling is achieved by selecting an equal amount of data through random selection of a minimum number of values in each damage grade from the training dataset, whereas oversampling is achieved by replicating the minority damage grades (i.e., DG1 and DG2 in our case).

The RFR model was trained using the resampled dataset to observe its performance on the validation dataset. Fig. 1.6 shows the performance of the RFR model after (a) undersampling and (b) oversampling. Compared with the previous model trained using an imbalanced dataset (Fig. 1.4b), the undersampling method degraded the efficacy of the RFR model with an increase in the MAE by 9%. By oversampling, the performance yielded was similar to that of the original method (1% increase in the MAE). However, undersampling increased the recall value for the lowest damage grades (Fig. 1.6a) i.e., the recall value increased by 20%–25% for DG1 to DG3 and reduced by less than 15% for DG4 and DG5. Meanwhile, oversampling improved the recall value moderately for the lower grades without significantly affecting the recall value of the higher damage grades, e.g., the DG1/DG5 recall value changed from 0.33/0.54 (Fig. 1.4b) to 0.41/0.51 (Fig. 1.6b), respectively. Therefore, for further studies, we plan to use the oversampling technique to address data imbalance issues such that a better estimation of the lower grades can be obtained.



**Figure 1.6.** Data imbalance testing applied to the RFR model. Damage grade (DG) normalized confusion matrix for (a) undersampling and (b) oversampling techniques.

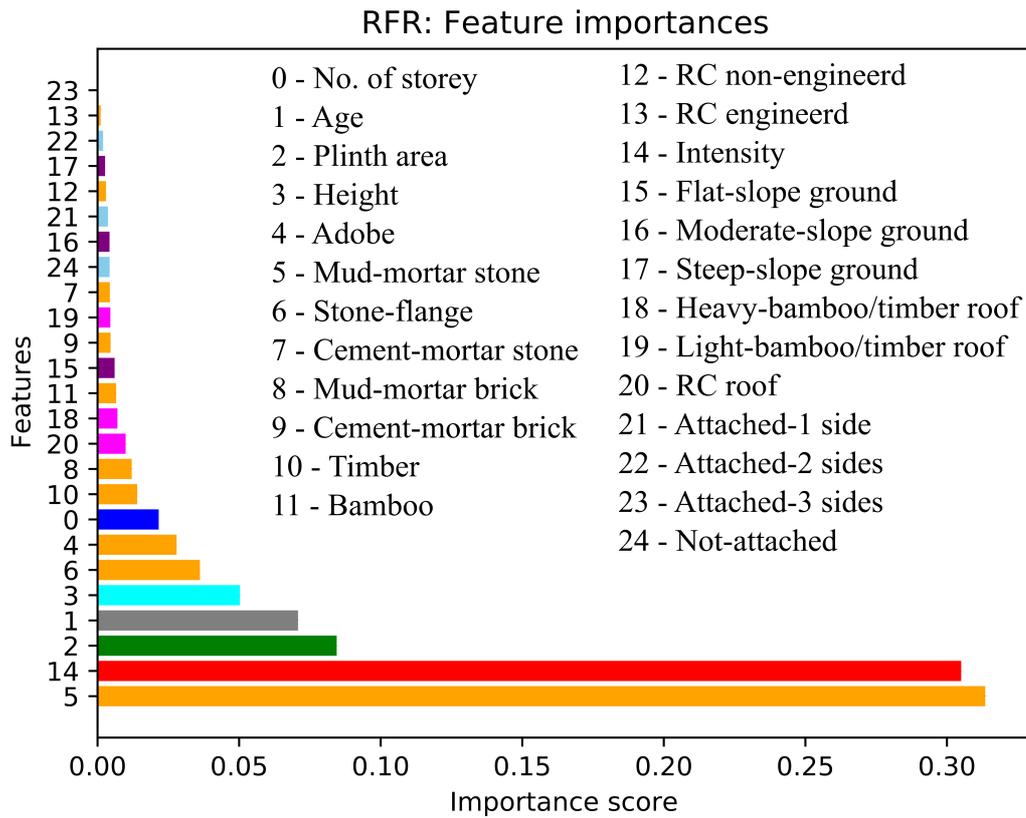
### 1.4.3 Feature importance

The importance of each input feature in the RFR damage-prediction model is analyzed in this section. In decision trees, each node is a condition for dividing the values into single characteristics such that similar values of the dependent variable appear in the same set after the division. The condition is based on impurity, which is the variance in the case of regression trees. In other words, each feature's contribution can be measured by the average decrease in impurity using all the trees in the forest, without considering the linear separability of the data (Pedregosa et al., 2011). The significance of each feature in the model is measured in terms of the feature importance scores. The feature importance score is a value assigned to each feature in the model while a model is developed. Note also that RFR uses the bagging algorithm (randomly splitting data into smaller subsets) while developing the trees, so the correlated features may or may not be used for a particular tree. Thus, correlated features may not affect the overall predictive performance, however, it may impact the importance ranking between two correlated features, removing one of the correlated features may lower the damage predictive performance of the RFR model.

Fig. 1.7 shows the importance score associated with each input feature considered in this study. The highest importance score (32% in Fig. 1.7) was associated with the mud-mortar stone material. Buildings constructed using mud-mortar-stone constituted the highest proportion in our dataset (Fig. 1.2i), and this building class was damaged severely during the Nepal earthquake, independent of the MSI. In fact, this type of building is generally associated with the most vulnerable class in most

vulnerability assessment methods. The fact that the RFR identified the mud-mortar-stone feature as the most important feature in damage prediction is consistent with previous studies (Maheri et al., 2005; Sayin et al., 2013; Webster and Tolles, 1999). Additionally, Fig. 1.7 shows that, as expected, the MSIs that characterized the ground motion were one of the most significant input features for earthquake damage prediction (31%), i.e., damage was first conditioned by the MSI regardless of the building parameters. Other building-related input features, such as plinth area (8%), construction age (7%), and height (5%) contributed significantly to earthquake damage prediction model; meanwhile, the other features contributed only marginally to the damage prediction model (RFR).

As mentioned earlier (Tab. 1.2), a higher damage-prediction accuracy (0.49) was observed when all features were considered (MAE = 0.66). However, information regarding building construction materials is typically not easily accessible. The four basic building parameters (i.e., number of stories, age, height, and plinth area) could be easily accessible from the institutional databases (e.g., national census, national housing database) (Crowley et al., 2020; Riedel et al., 2015), from filed survey, and partially available in open-source platform (e.g., OpenStreetMap (OSM) (Bennett, 2010), OpenBuildingMap (OBM) (Schorlemmer et al., 2017)). Thus, the performance of RFR model considering only the basic parameters of the building (i.e., number of stories, age, height, and plinth area) in addition to the ground-motion intensities is explored. The RFR model provided a similar level of accuracy (0.46) in damage prediction (MAE = 0.72). This result suggests that basic building parameters enable the development of satisfactory RFR models for predicting damage, and therefore earthquake losses, for the given value of ground-motion intensity (as reported by Riedel et al. (2015) and Guettiche et al. (2017) for seismic vulnerability classification using a support vector machine).

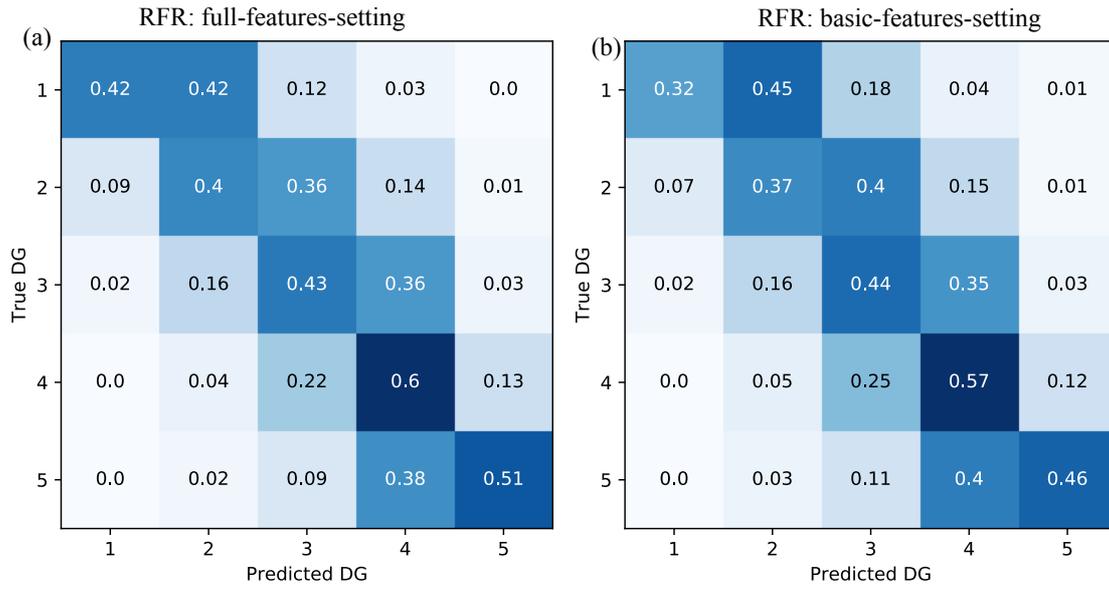


**Figure 1.7.** Graphical representation of the importance scores associated with the different input features considered for the RFR model. Categorical features are transformed to binary features. The features (same as in Fig. 1.2) considered in this study are on the y-axis and the x-axis is the importance score. In the figure the building features are represented by a set of colors: no. of storey (blue), age (grey), plinth area (green), height (cyan), construction material (orange), MSI (red), ground slope (purple), roof type (magenta), and building’s position (sky blue).

#### 1.4.4 Damage prediction using test dataset

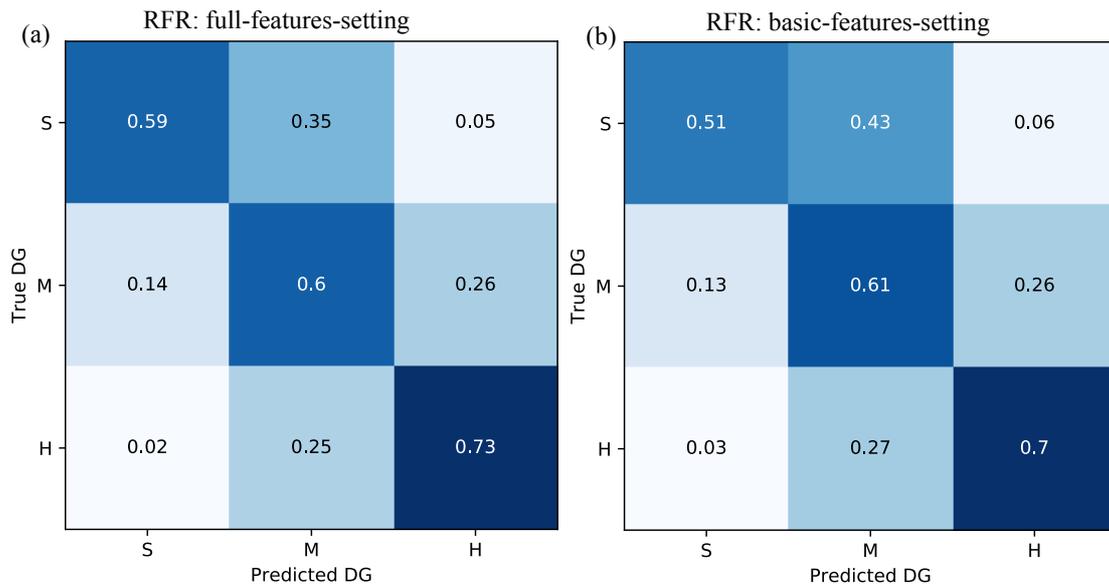
This section presents the predictive performance of the RFR model for the test dataset. The RFR model was developed for two sets of features, i.e., all building parameters (i.e., full-feature setting) and the basic parameters (i.e., number of stories, age, height, plinth area, and ground-motion intensity) (basic-feature setting).

The full-feature/basic-feature settings yielded MSE scores of 0.81/0.93. Confusion matrices for the full- and basic-feature settings are shown in Figs. 1.8a and 1.8b, respectively. The main diagonals and their adjacent values are associated with a higher recall value, illustrating higher efficacy in damage prediction and a lower misclassification rate. As shown in Figs. 1.4b and 1.6b, the RFR model applied to the test dataset predicted the highest damage grades (DG4/DG5) with a higher recall value, corresponding to 60%/51% and 57%/46% in the full- and basic-feature settings, respectively.



**Figure 1.8.** Damage grade (DG) normalized confusion matrix for (a) the full-features setting and (b) the basic-features setting observed in the test dataset.

During field observations for damage classification, it is often difficult to assign a damage level to buildings falling between two damage grades. Moreover, it is more convenient to classify damage into three categories, i.e., in the same manner as using the traffic-light-system (TLS)-based classification system (green, yellow, and red) in post-earthquake emergency surveys (ATC, 2005; Bazzurro et al., 2004). Therefore, we classified the five damage grades into three categories by considering the severity of damage based on the damage grade definition (Section 1.2), as follows: S (slight: DG1+DG2), M (medium: DG3), and H (heavy: DG4+DG5). The MSE scores for the full-feature/basic-feature settings were 0.34/0.39. Fig. 1.9 shows the confusion matrices. The damage prediction accuracy increased significantly when the TLS-based approach was used. For example, the recall values for the S/M/H damage grades were 59%/60%/73% and 51%/61%/70% for the full- and basic-feature settings, respectively.



**Figure 1.9.** Damage grade (DG) normalized confusion matrix for (a) the full-features setting and (b) the basic-features setting in the traffic-light (TLS) -based classification approach, grouping the five damage grades (DG) into three classes, small, medium, and heavy damage grades (S, M, H).

Table 1.3 summarizes the MAE, and MSE scores observed with the full- and basic-feature settings. For the basic-feature setting, the values of these scores (e.g., MAE = 0.73/0.45), considering that the 5DG/TLS damage scales were similar to those of the full-feature setting (e.g., MAE = 0.69/0.41). Section 4.3 shows that using the RFR model, a reasonable damage prediction was possible by merely considering the basic features of the buildings.

**Table 1.3.** Summary of the performance scores for RFR model considering the test datasets for the full-features setting and basic-features setting. The performance scores are associated to the 5DG and TLS-based damage classification systems.

Features	Damage grade	Accuracy score	MAE	MSE
Full-features setting	5DG	0.49	0.69	0.81
	TLS	0.68	0.41	0.34
Basic-features setting	5DG	0.46	0.73	0.93
	TLS	0.64	0.45	0.39

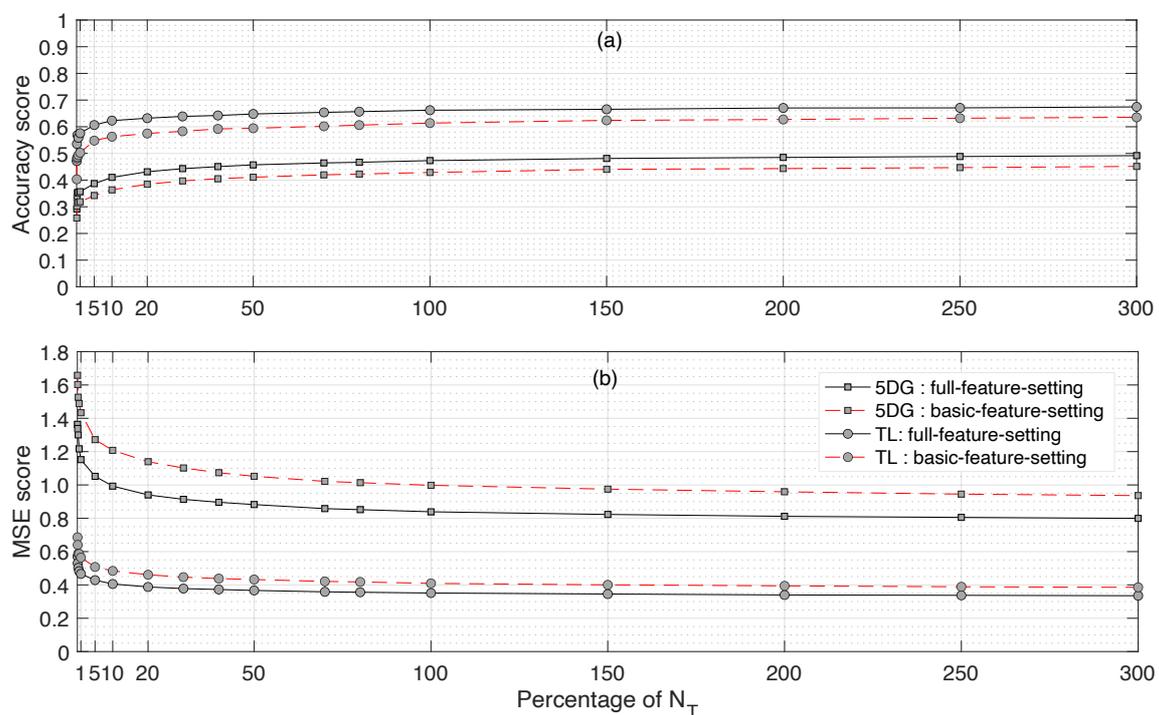
The results yielded by the RFR model confirmed that considering basic building features resulted in a relatively similar damage prediction efficacy as compared with considering all features. The possible reason behind this is maybe the combination of these basic features is able to indirectly capture some key parameters of buildings needed to define the vulnerability, such as the building typology. The RFR model developed by considering the basic features thereby may overcome the challenges of data

acquisition for vulnerability assessments as well as earthquake damage and loss prediction on the urban or regional scale.

### 1.4.5 Testing the proportionating of the dataset

We explored the damage prediction efficacy of the RFR model as a function of the amount of data in the training dataset. Fig. 1.10 shows the performance of the method (accuracy and MSE scores) for several percentage of data in the training phase according to the test dataset (% of  $N_T$ ). In this case 100% of the original training set correspond to 300% of  $N_T$ . For 10% of  $N_T$ , the accuracy score is 0.63/0.57 and 0.42/0.37 for TLS- and 5DG- based damage classifications considering full-features/basic-features-setting, respectively. These scores are very close to the performance scores noted in Table 1.3. Major improvement in the damage prediction efficacy of the RFR model is achieved up to 50% of  $N_T$ . After 50%, a minor improvement is observed.

Moreover, the damage prediction efficacy of the RFR model was further improved by taking into account the geographic location of the building (ward-id in this case). For example, at 10% of  $N_T$ , the accuracy score was 0.66/0.63 and 0.47/0.44 for TLS and 5DG based damage classification at full-features/basic-features-setting, respectively.



**Figure 1.10.** Variation of the accuracy (a) and MSE (b) of the RFR model as a function of the percentage of the whole test dataset,  $N_T$ , used for the training phase. The black-solid/red-dashed lines with the grey circle corresponds to TLS-based damage classification for the full-features/basic-features-setting, respectively. The black-solid/red-dash line with the grey square corresponds to 5DG-based damage classification for the full-features/basic-features-setting, respectively.

## 1.5 Discussion

In this study, we investigated the efficiency of various machine learning techniques for post-earthquake seismic damage assessment using the NBDP database compiled after the 2015  $M_w$  7.8 earthquake in Nepal. Machine learning models were trained based on a number of building characteristics and some basic parameters that contributed significantly to damage prediction performance. We tested the predictive efficacies of LR, SVR, GBC/GBR, and RFR/RFC models. For classification methods significant misclassification is observed for intermediate damage grades whereas for regression methods the predicted values were similar to the ground truth (see. Fig 1.4, 1.5 and Table 1.2). This could be because for classification methods every mis-classification has the same penalty loss whereas regression methods penalize the relative penalty loss (i.e. difference error) by minimizing the squared loss between true and predicted damage grades. Among these models, RFR was the most relevant model for damage prediction when applied to our dataset; it provided the best cost/benefit ratio in terms of performance and computing time. For the baseline comparison, the result obtained from the RFR model is compared with a random uniform baseline (randomly assigning damage grades following an uniform distribution) and a random stratified baseline (following the distribution of damage grades in the training set) to every input. The RFR model resulted largely higher accuracy score (0.49) as compared to the random uniform baseline (0.20) and the random stratified baseline (0.25) to every input for 5DG classification. Moreover, in this case study, the RFR model demonstrated superior performance in predicting higher damage grades, which is ultimately the most pursued information for seismic damage assessment and earthquake loss reduction.

Additionally, we achieved a moderate improvement in damage prediction by addressing data imbalance issues via data resampling using the RFR model. In this study, we considered an unseen test dataset to ensure the overfitting issues of the model. The error values of validation and test datasets are very similar (same 0.49 accuracy on both cases), indicating that the model did not over-fit on the validation set.

However, the results herein show that considering a traffic-light damage classification and a limited number of building features, the results reach very satisfactory scores according to the objectives set: damage prediction accuracy 0.64 for basic-features setting, to be compared to Mangalathu et al. (2020b), Roslin et al., (2020), and Harirchian et al., (2021) studies that reported accuracy values of 0.66, 0.67, and, 0.65, respectively also with the random forest method. The 3-class classification model lowers the number of targets and then lowers the confusion as is often observed during damage surveys for the intermediary damage grade (e.g., between DG3 and DG4). In addition, so-called basic building features (i.e., number of stories, age, height, and plinth area) associated with ground-motion intensity provided a relevant estimate of the damage grade with a significant level of accuracy. Similar observations have also been reported by Mangalathu et al. (2020b) and Stojadinović et al. (2021) in

their studies using the 2014 South Napa and 2010 Kraljevo, Serbian earthquake building damage dataset, respectively. Moreover, the RFR model trained on a relatively small amount of dataset (5% - 20% of the test dataset) resulted in a reasonable estimate of damage; similar observation has been reported by Stojadinović et al., (2021).

The input ground motion used was the USGS ShakeMap intensity of the mainshock, whereas the overall quality of the NBDP database results were based on the cumulative effects of the mainshock and aftershock events, which might have affected the prediction efficacy. In addition, missing building-by-building information in the NBDP database relative to their localization and their associated site condition reduce the damage prediction efficacy of the machine learning model (Mangalathu et al., 2020b; Roeslin et al., 2020; Stojadinović et al., 2021). However, we also observed that by considering the geographic locations of buildings (ward-id in our case) slightly improved the damage prediction efficacy of the RFR model.

Finally, in Nepal, a majority of the building typology inventory collected after the 2015 earthquake holds a good match with the building typology inventory as observed in the 2011 national census (Chaulagain et al., 2015). Thus, the findings from this study seems promising to the Nepal's exposure context. Further investigations should be carried out to strengthen these findings: after the 2015 Gorkha Nepal earthquake, national and international communities are collaborating to develop an exposure model for Nepal (Jordan, 2019).

## **1.6 Conclusion**

To summarize, a framework for combining building parameters and earthquake building damage information collected after an earthquake event with machine learning techniques for seismic-damage assessment was presented herein. This study shows a possibility of using machine learning methods for immediate damage assessment once ground-motion information is published via operational tools, such as the USGS ShakeMap. This study shows that the building's features (number of stories, age, floor area, height) could result in reasonable damage prediction. These basic building parameters could be available in existing institutional databases (e.g., national census, national housing database), thereby resolving data acquisition issues associated with seismic-damage assessments at the urban or regional scale.

Without anticipating how a city planner can use these results, machine learning developed on a building portfolio makes it possible to explain post-earthquake damage. One may assume that the model herein for a specific portfolio might be used to predict and represent seismic damage in another region with the same portfolio (for example, Riedel et al., 2014) or for another characterization of the seismic hazard (Riedel et al., 2015), in an emergency situation (immediately after an earthquake if the exposure model is known) or in a planning process. This will of course have to be confirmed on another dataset, e.g.

consisting of a series of earthquakes affecting the same region, i.e. characterized by the same portfolio of buildings. In addition, the results herein show that using a traffic-light damage classification and a limited number of building features, the results reach very satisfactory scores according to the objectives set (damage prediction accuracy 0.64 for 4 basic features/3 damage grade settings, to be compared to the studies of Mangalathu et al. (2020b), Roslin et al. (2020), and Harirchian et al., (2021) that reported accuracy values of 0.66, 0.67, and, 0.65, respectively with the random forest method).

Additionally, the collection of the building exposure data is the key challenge faced by the damage assessment communities. The observed performance of machine learning considering only basic parameters, easy to collect and without requiring much technical expertise, may suggest the interest of exploring the potential of crowd-source databases (e.g. OSM/OBM) as input parameters. This approach will require validation before being able to assert it definitively, but it would solve the difficulty of considering the evolution of the building in the definition of the exposure models.

## 2 Testing machine learning models for heuristic building damage assessment applied to the Italian Database of Observed Damage (DaDO)

---

In chapter 1, we compared the damage prediction performance of various machine learning models to explain post-earthquake seismic damage, considering two different classifications of damage and two different sets of building features, applied to the 2015 Nepal earthquake building-damage portfolio (NBDP).

In this chapter, we evaluate machine learning models for heuristic building damage assessment at a regional scale using the database of observed damage (DaDO) in Italy. We analyzed the efficacy of machine learning models trained on past earthquake building damage datasets to predict potential damage in future earthquakes, the relation between building features and damage level, and methods to handle imbalance distribution of target features. We also compared the damage prediction efficacy of the machine learning model with the Risk-UE method. The content of this chapter is under review in the Natural Hazards and Earth System Science journal.

*Ghimire, S., Guéguen, P., Pothon, A., and Schorlemmer, D. (2023). Testing machine learning models for heuristic building damage assessment applied to the Italian Database of Observed Damage (DaDO). Natural Hazards and Earth System Sciences Discussions, 1-29*

---

### Abstract

Assessing or forecasting seismic damage to buildings is an essential issue for earthquake disaster management. In this study, we explore the efficacy of several machine learning models for damage characterization, trained and tested on the database of damage observed after Italian earthquakes (DaDO). Six models were considered: regression- and classification-based machine learning models, each using random forest, gradient boosting and extreme gradient boosting. The structural features considered were divided into two groups: all structural features provided by DaDO or only those considered to be the most reliable and easiest to collect (age, number of storeys, floor area, building height). Macroseismic intensity was also included as an input feature. The seismic damage per building was determined according to the EMS-98 scale observed after seven significant earthquakes occurring in several Italian regions. The results showed that extreme gradient boosting classification is statistically the most efficient method, particularly when considering the basic structural features and grouping the damage according to the traffic-light based system used, for example, during the post-disaster period (green, yellow and red), 68% buildings were correctly classified. The results obtained by the machine

learning-based heuristic model for damage assessment are of the same order of accuracy (error values were less than 17%) as those obtained by the traditional Risk-UE method. Finally, the machine learning analysis found that the importance of structural features with respect to damage was conditioned by the level of damage considered.

### **Keywords**

Earthquake building-damage, DaDO building damage database, Machine learning, RISK-UE, Seismic vulnerability of buildings, Italy.

## **2.1 Introduction**

Population growth worldwide increases exposure to natural hazards, increasing consequences in terms of global economic and human losses. For example, between 1985 and 2014, the world's population increased by 50% and average annual losses due to natural disasters increased from US\$14 billion to over US\$140 billion (Silva et al., 2019). Among other natural hazards, earthquakes represent one-fifth of total annual economic losses and cause more than 20 thousand deaths per year (Daniell et al., 2017; Silva et al., 2019). To develop effective seismic risk reduction policies, decision-makers and stakeholders rely on a representation of consequences when earthquakes affect the built environment. Two main risk metrics generally considered at the global scale are associated with building damage: direct economic losses due to costs of repair/replacement and loss of life of inhabitants due to building damage. The damage is estimated by combining the seismic hazard, exposure models and vulnerability/fragility functions (Silva et al., 2019).

For scenario-based risk assessment, damage and related consequences are computed for a single earthquake defined in terms of magnitude, location, and other seismological features. Many methods have been developed to characterize the urban environment for exposure models. In particular, damage assessment requires vulnerability/fragility functions for all types of existing buildings, defined according to their design characteristics (shape, position, materials, height, etc.) and grouped in a building taxonomy (e.g. among other conventional methods FEMA, 2003; Grünthal, 1998; Guéguen et al., 2007; Lagomarsino and Giovinazzi, 2006; Mouroux and Le Brun, 2006; Silva et al., 2014). At the regional/country scale, damage assessment is therefore confronted with the difficulty of accurately characterizing exposure according to the required criteria and assigning appropriate vulnerability/fragility functions to building features. Unfortunately, the necessary information is often sparse and incomplete, and exposure model is suffering from economic and time constraints.

Over the past decade, there has been growing interest in artificial intelligence methods for seismic risk assessment, due to its superior computational efficiency, easy handling of complex problems, and the incorporation of uncertainties (e.g., Riedel et al., 2014, 2015; Azimi et al., 2020; Ghimire et al., 2022;

Hegde and Rokseth, 2020; Kim et al., 2020; Mangalathu and Jeon, 2020; Morfidis and Kostinakis, 2018; Salehi and Burgueño, 2018; Seo et al., 2012; Sun et al., 2021; Wang et al., 2021; Xie et al., 2020; Y. Xu et al., 2020; Z. Xu et al., 2020). In particular, several studies have tested the effectiveness of machine learning methods in associating damage degrees with basic building features and spatially-distributed seismic demand with acceptable accuracy compared with conventional methods or tested with post-earthquake observations (e.g., Riedel et al., 2014, 2015; Guettiche et al., 2017; Harirchian et al., 2021; Mangalathu et al., 2020; Roeslin et al., 2020; Stojadinović et al., 2021; Ghimire et al., 2022). In parallel, significant efforts have been made to collect post-earthquake building damage observations after damaging earthquakes (Dolce et al., 2019; MINVU, 2010; MTPTC, 2010; NPC, 2015). With more than 10,000 samples compiled, the Database of Observed Damage (DaDO) in Italy, a platform of the Civil Protection Department, developed by the Eucentre Foundation (Dolce et al., 2019), allows exploration of the value of heuristic vulnerability functions calibrated on observations (Lagomarsino et al., 2021), as well as the training of heuristic functions using machine learning models (Ghimire et al., 2022) and considering sparse and incomplete building features.

The main objective of this study is to investigate the effectiveness of several machine learning models trained and tested on information from the DaDO to develop a heuristic model for damage assessment. The model may be classified as heuristic because it applies a problem-solving approach in which a calculated guess based on previous experience is considered for damage assessment (as opposed to applying algorithms that effectively eliminate the approximation). The damage is thus estimated in a non-rigorous way defined during the training phase and the results must be validated and then tested against observed damage. By analogy with psychology, this procedure can reduce the cognitive load associated with uncertainties when making decisions based on damage assessment, by explicitly considering the uncertainties in the assessment, being aware about the incompleteness of the information and the accuracy level to make a decision. The dataset and methods are described in the data (2.2) and method (2.3) sections, respectively. The 2.4 section presents the results of damage prediction produced by machine learning models compared with conventional methods, followed by a conclusion section.

## **2.2 Data**

The Database of Observed Damage (DaDO, Dolce et al., 2019) is accessible through a web-GIS platform and is designed to collect and share information about building features, seismic ground motions and observed damage following major earthquakes in Italy from 1976 to 2019. A framework was adopted to homogenize the different forms of information collected and to translate the damage information into the EMS-98 scale (Grunthal et al., 1998) using the method proposed by Dolce et al.

(2019). For this study, we selected building damage data from seven earthquakes summarized in Table 2.1 and presented in Fig. 2.1.

**Table 2.1.** Building-damage data from the DaDO for the seven earthquakes considered in this study. ‘Ref’ is the reference to the earthquake. ‘DL’ is the number of the damage grade available in DaDO. ‘NB’ is the number of buildings considered in this study. AeDES is the post-earthquake damage survey form, first introduced in 1997 and become the official operational tool recognized by the Italian Civil Protection in 2002.

Ref	Earthquake	Event date	Mag.	Epicentre		Damage survey form	DL	NB
				Lat.	Long.			
E1	Irpinia-1980	23/11/1980	6.9	40.91	15.37	Irpinia-1980	8	37,828
E2	Pollino-1998	09/09/1998	5.6	40.04	15.98	AeDES-1998	4	9,485
E3	Molise-Puglia-2002	31/10/2002	5.9	41.79	14.87	AeDES-2000	4	6,396
E4	Emilia-Romagna-2003	14/09/2003	5.3	44.33	11.45	AeDES-2000	4	239
E5	L'Aquila-2009	06/04/2009	6.3	42.34	13.34	AeDES-2008	4	37,999
E6	Emilia-Romagna-2012	20/05/2012	6.1	44.89	11.23	AeDES-2008	4	10,581
E7	Garfagnana-Lunigiana-2013	21/06/2013	5.3	44.15	10.14	AeDES-2008	4	1,474

The converted EMS-98 damage grade (DG) ranges from damage grade DG0 (no damage) to DG5 (total collapse). The building features are available for each individual building and relate to the shape and design of the building and the built-up environment (Tab. 2.2, Fig. 2.2), as follows:

**Building location** - the location of each building is defined by its latitude and longitude, assigned using either the exact address of the building if available or the address of the local administrative centre (Dolce et al., 2019).

**Numbers of storeys** - total number of floors above the surface of the ground.

**Age of building** - time difference between the date of the earthquake and the date of building construction/renovation.

**Height of building** - total height of the building above the surface of the ground, in m.

**Floor area** – average of the storey surface area, in m<sup>2</sup>.

**Ground slope condition** - four types of ground slope conditions are defined (flat, mild slope, steep slope, and ridge).

**Roof type** – four types of roofs are defined (thrusting heavy roof, non-thrusting heavy roof, thrusting light roof, and non-thrusting light roof).

**Position of building** - indication of the building's position in the block: isolated, extreme, corner, and intermediate.

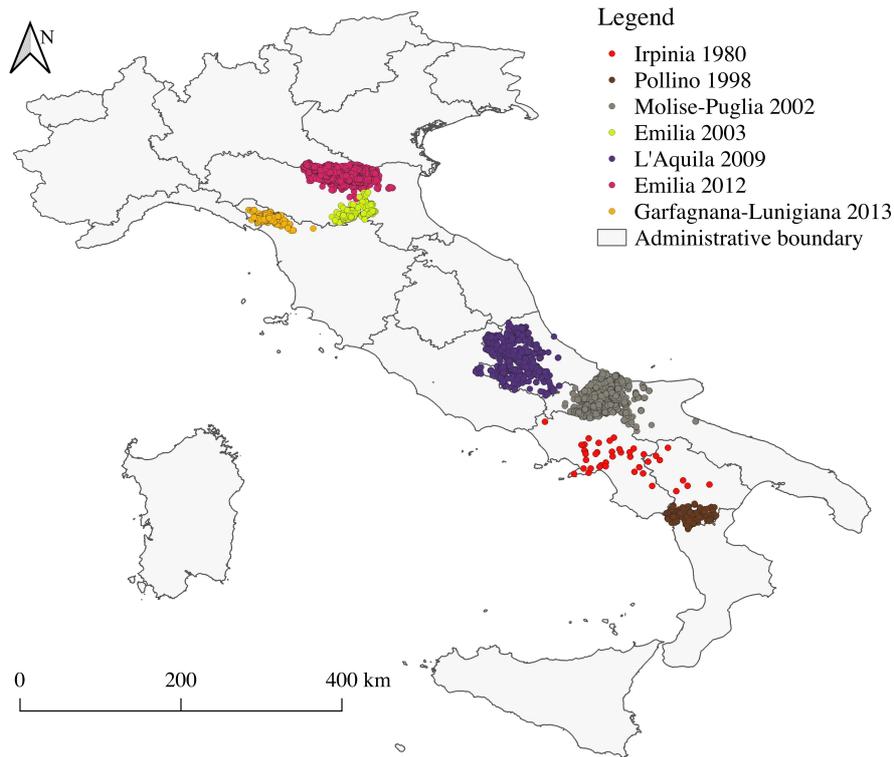
**Regularity:** building regularity in terms of plan and elevation, classified as either irregular or regular.

**Construction material:** vertical elements: good and poor-quality masonry, good and poor quality mixed frame masonry, reinforced concrete frame and wall, steel frame, and other.

For features defined as value ranges (e.g., date of construction/renovation, floor area, and building height), the average value was used. Furthermore, the Irpinia-1980 building damage portfolio (E1) was constructed using the specific Irpinia-1980 damage survey form, while the AeDES damage survey form was used for the others. The Irpinia-1980 dataset will therefore be analysed separately.

Building damage data from earthquake surveys other than the Irpinia-1980 earthquake damage survey primarily include damaged buildings. This is because the data was collected based on requests for damage assessments after the earthquake event (Dolce et al. 2019). The damage information in the DaDO database is still relevant for testing the machine learning models for heuristic damage assessment. Mixing these datasets to train machine learning models can lead to biased outcomes. Therefore, the machine learning models were developed on the other earthquake dataset excluding the Irpinia dataset, and the Irpinia earthquake dataset was used only in the testing phase.

The distribution of the samples is very imbalanced (Fig. 2.2): for example, there is a small proportion of buildings in DG4+DG5 (7.59%), and a large majority of masonry (65.47%) compared to reinforced concrete frame (21.31%) buildings. This imbalance should be taken into account when defining the machine learning models.



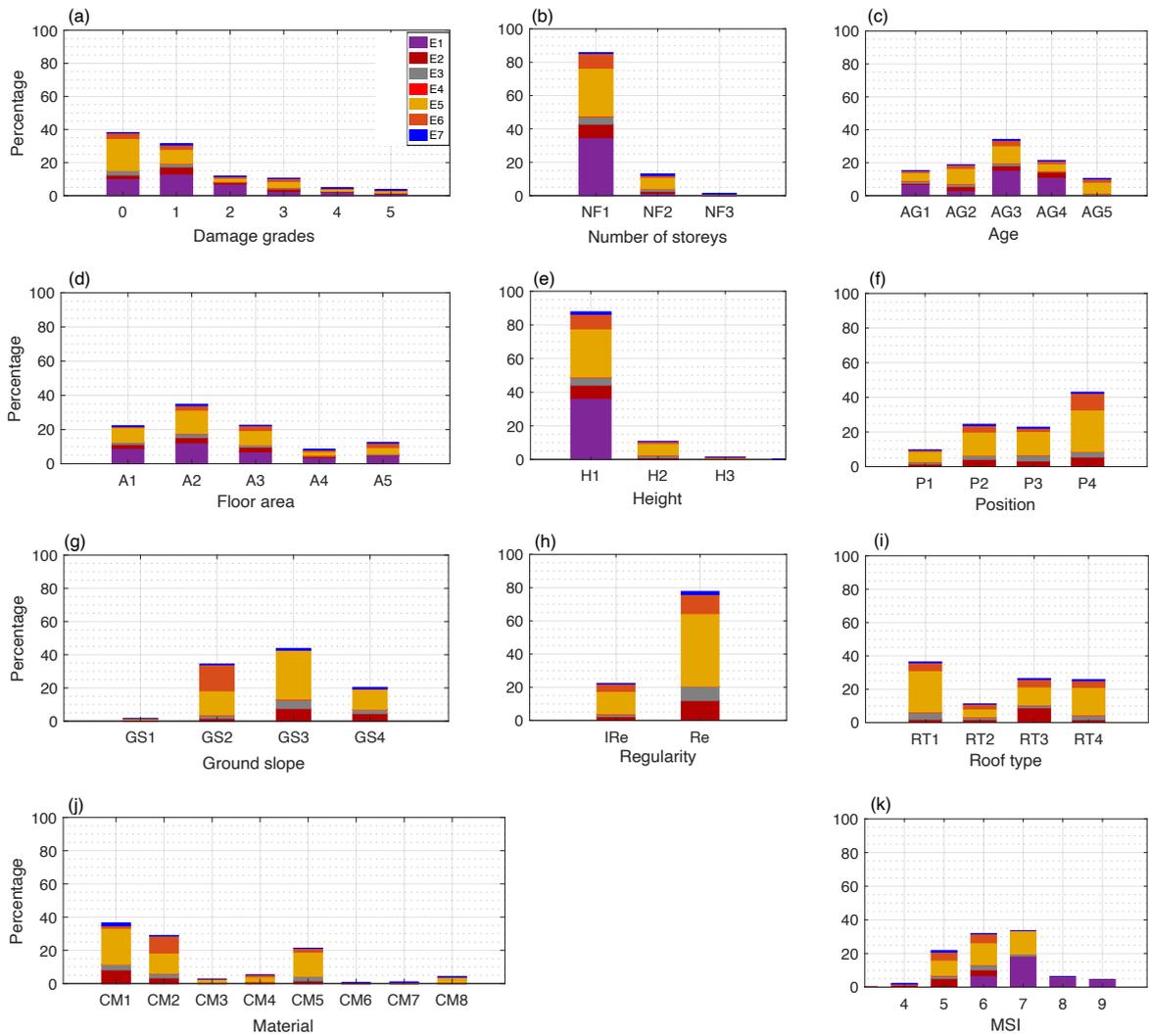
**Figure 2.1.** Geographic location of the buildings considered in this study.

To consider spatially-distributed ground motion, the original DaDO data are supplemented with the main event macroseismic intensities (MSI) provided by the United States Geological Survey (USGS) ShakeMap tool (Wald et al., 2005). Macroseismic intensities (MSI) given in terms of modified Mercalli intensities are considered and assigned to buildings based on their location. The distribution of MSI values in the database is shown in Fig. 2.2k.

**Table 2.2.** Distribution of the different features used in this study.

No.	Parameters	Data type	Distribution (%)	Remarks	
1	Damage grades (DG)	No damage	DG0	43.63	Fig. 2.2a
		Slight damage	DG1	28.90	
		Moderate damage	DG2	7.41	
		Substantial damage	DG3	12.48	
		Very heavy damage	DG4	3.94	
		Total collapse	DG5	3.65	
2	Number of storeys	0-3	NF1	85.81	Fig. 2.2b
		3-5	NF2	13.01	
		> 5	NF3	1.19	
3	Age (years)	0-20	AG1	15.22	Fig. 2.2c
		21-40	AG2	18.81	
		41-60	AG3	34.15	
		61-80	AG4	21.34	
		>80	AG5	10.49	
4	0-50	A1	Numerical	22.16	Fig. 2.2d

	Floor area (square metres)	50-100	A2		34.73	
		100-150	A3		22.53	
		150-200	A4		8.32	
		> 200	A5		12.26	
5	Height (metres)	0-10	H1	Numerical	87.78	Fig. 2.2e
		10-15	H2		10.69	
		>15	H3		1.50	
6	Position	Corner	P1	Categorical	9.71	Fig. 2.2f
		Extreme	P2		24.47	
		Internal	P3		22.80	
		Isolated	P4		43.02	
7	Ground slope	Ridge	GS1	Categorical	2.62	Fig. 2.2g
		Plain	GS2		34.25	
		Moderate slope	GS3		43.74	
		Steep Slope	GS4		20.39	
8	Regularity	Irregular in plan and elevation	IR	Categorical	22.28	Fig. 2.2h
		Regular in plan and elevation	Re		77.72	
9	Roof type	Heavy no thrust	R1	Categorical	36.43	Fig. 2.2i
		Heavy thrust	R2		11.25	
		Light thrust	R3		26.48	
		Light no thrust	R4		25.83	
10	Material	Masonry poor quality	CM1	Categorical	36.51	Fig. 2.2j
		Masonry good quality	CM2		28.96	
		Mixed frame masonry poor quality	CM3		2.64	
		Mixed frame masonry good quality	CM4		5.21	
		Reinforced concrete frame	CM5		21.31	
		Reinforced concrete wall	CM6		0.42	
		Steel frame	CM7		0.09	
		Other	CM8		4.10	



**Figure 2.2.** Distribution of the different features in the database. E1, E2, E3, E4, E5, E6, and E7, representing Irpinia-1980, Pollino-1998, Molise-Puglia-2002, Emilia-Romagna-2003, L'Aquila-2009, Emilia-Romagna-2012, and Garfagnana-Lunigiana-2013 building damage portfolios, respectively. The y-axis is the percentage distribution and the x-axis is (a) Damage grade, (b) Number of storeys (NF1: 0-3, NF2: 3-5, NF3: >5), (c) Building age (AG1: 0-20, AG2: 21-40, AG3: 41-60, AG4: 61-80, AG5: >80), (d) Floor area (A1: 0-50, A2: 51-100, A3: 101-150, A4: 151-200, A5: >200), (e) Height (H1: 0-10, H2: 10-15, H3: >15), (f) Building position (P1: corner, P2: extreme, P3: internal, P4: isolated), (g) Ground slope condition (GS1: ridge, GS2: plain, GS3: moderate slope, GS4: steep slope), (h) Regularity in plan and elevation (IRe: irregular, Re: Regular), (i) Roof type (RT1: heavy no thrust, RT2: heavy thrust, RT3: light no thrust, RT4: light thrust), (j) Construction material (CM1: poor-quality masonry, CM2: good-quality masonry, CM3: poor-quality mixed frame masonry, CM4: good-quality mixed frame masonry, CM5: reinforced concrete frame, CM6: reinforced concrete wall, CM7: steel frames, CM8: other), and (k) macro-seismic intensity.

## 2.3 Method

### Machine learning models

In chapter 1, we applied classification- and regression-based machine learning models to the damage observed after the 2015 Gorkha Nepal earthquake (NPC, 2015). The main concepts for method selection, the definition of the dataset for training and testing, and the representation of model performance are presented here.

To develop the heuristic damage assessment model, the damage grades are considered as the target feature. The damage grades are discrete labels, from DG0 to DG5. Three most advanced classification and regression machine learning algorithms were selected: random forest (RFC) and regression (RFR) (Breiman, 2001), gradient boosting classification (GBC) and regression (GBR) (Friedman, 1999), and extreme gradient boosting classification (XGBC) and regression (XGBR) (Chen and Guestrin, 2016). A label (or class) was thus assigned to the categorical response variables (DG) for the classification-based machine learning models. For the regression-based machine learning models, DG is converted into a continuous variable to minimize misclassifications (Ghimire et al., 2022).

Building features and macroseismic intensities were considered as input features. A one-hot encoding technique was used to convert the categorical features (i.e., ground slope condition, building position, roof type, construction material) into binary values (1 or 0), resulting in 28 input variables (Tab. 2.2). No input features were removed from the dataset: some building features (e.g., number of storeys and height) may be correlated but we assumed that the presence of correlated features does not impact the overall performance of these machine learning methods (Ghimire et al., 2022). No specific data cleaning methods were applied to the DaDO database.

The machine learning algorithms from the Scikit-learn package developed in Python (Pedregosa et al., 2011) were applied. The machine learning models were trained and tested on the randomly selected training (60% of the dataset) and testing (40% of the dataset) subsets of data, considering a single earthquake dataset or the whole DaDO dataset. The testing subset was kept hidden from the model during the training phase.

### **Machine learning model efficacy**

The efficacy of the heuristic damage assessment model (i.e., its ability to predict damage to a satisfactory or expected degree) was analysed in three stages: comparison of the efficacy of the machine learning models using metrics; analysis of specific issues related to machine learning using the selected models; and application of the heuristic model to the whole DaDO dataset.

#### **First stage: model selection**

In the first stage, only the L'Aquila-2009 portfolio was considered for the training and testing phases. This is the largest dataset in terms of the number of buildings and was obtained using the AeDES survey format (Baggio et al., 2007; Dolce et al., 2019). Model efficacy was provided by the confusion matrix, which represents model prediction compared with the so-called “ground truth” value. Accuracy was then represented on the confusion matrix by the ratio of the number of correctly predicted DGs to the total number of observed values per DG ( $A_{DG}$ ).

Total accuracy ( $A_T$ ) was computed as the ratio of the number of correctly predicted DGs to the total number of observed values.  $A_T$  and  $A_{DG}$  values close to 1 indicate high efficacy. Moreover, the quantitative statistical error was also calculated as the mean of the absolute value of errors (MAE) and the mean squared error (MSE) (MAE and MSE values close to 0 indicate high efficacy). For classification-based machine learning models, the ordinal value of the DG was used to calculate the MAE and MSE scores directly. For the regression-based machine learning models, the output DG values were rounded to the nearest integer for the accuracy scores plotted for the confusion matrix, but not for the MAE and MSE value calculations.

### **Second stage: machine learning related issues**

In the second stage, the best heuristic model for damage assessment was selected based on the highest efficacy, and used to analyse and test specific issues related to machine learning: (1) the imbalance distribution of DGs in the DaDO, (2) the performance of the selected model when only some basic, but accurately assessed, building features are considered (i.e., number of storeys, location, age, floor area), and (3) the simplification of the heuristic model, in the sense that DGs are grouped into a traffic-light-based classification (i.e., green, yellow and red, corresponding to DG0+DG1, DG2+DG3 and DG4+DG5, respectively). In the second stage, the issues related to machine learning were first analysed using the L'Aquila-2009 portfolio. The whole DaDO dataset was then used.

### **Third stage: application to the whole DaDO portfolio and comparison with Risk-UE**

In the third stage, several learning and testing sequences were considered, with the idea of moving to an operational configuration in which past information is used to predict damage from future earthquakes: either learning based on a portfolio of damage caused by one earthquake and tested on another portfolio, or learning based on a series of damage portfolios and tested on the portfolio of damage caused by an earthquake placed in the chronological continuity of the earthquake sequence considered. In this stage, the efficacy of the heuristic damage assessment model was analysed by comparing the prediction values with the so-called “ground truth” values through the error distribution, as follows:

$$\varepsilon_d(\%) = \left(\frac{n_e}{N}\right) * 100 \quad (2.1)$$

where  $n_e$  is the total number of buildings at a given error level (difference between observed and predicted DGs),  $N$  is the total number of buildings in the damage portfolio.

In this stage, the efficacy of the heuristic damage assessment model was compared with the conventional damage prediction framework proposed by the RISK-UE method (Milutinovic and Trendafiloski, 2003). The RISK-UE method assigns a vulnerability index (IV) to a building, based on its construction material and structural properties (e.g., height, building age, position, regularities, geographic location). For a given level of seismic demand (MSI), the mean damage ( $\mu_d$ ) and the probability,  $p_k$ , of observing a given damage level  $k$  ( $k = 0$  to  $5$ ) are given by:

$$\mu_d = 2.5 \left[ 1 + \tanh \left( \frac{MSI + 6.25IV - 13.1}{2.3} \right) \right] \quad (2.2)$$

$$p_k = \frac{5!}{k!(5-k)!} \left( \frac{\mu_d}{5} \right)^k \left( 1 - \frac{\mu_d}{5} \right)^{5-k} \quad (2.3)$$

Herein, comparing the heuristic model and the RISK-UE method amounts to considering the following steps, based on the equations given by RISK-UE:

**Step 1** - The buildings in the training and testing datasets are grouped into different classes according to construction material.

**Step 2** - For a given building class in the training dataset, computation of

**Step 2.1** - mean damage ( $\mu_d$ ) using the observed damage distribution at a given MSI value by:

$$\mu_d = \sum_{k=0}^5 p_k k \quad (2.4)$$

**Step 2.2** - vulnerability index (IV) with the  $\mu_d$  obtained in step 2.1 by:

$$IV = \frac{1}{6.25} \left[ 13.1 - MSI + 2.3 \left( \tanh^{-1} \left( \frac{\mu_d}{2.5} - 1 \right) \right) \right] \quad (2.5)$$

**Step 3** - For the same building class in the test dataset, calculation of

**Step 3.1** - mean damage ( $\mu_d$ ) Eq. 2.2 for a given MSI value with the value of IV obtained in step 2.2;

**Step 3.2** - damage probability ( $p_k$ ) Eq. 2.3 with the value of  $\mu_d$  obtained in step 3.1;

**Step 3.3** - distribution of buildings in each damage grade within a range of MSI values observed in the test dataset as follows:

$$N_{pred,k} = \sum_{MSI} p_k n_{obs,MSI} \quad (2.6)$$

where  $n_{\text{obs,MSI}}$  is the total number of buildings observed in the test set for a given MSI value;

**Step 3.4** –absolute error ( $\varepsilon_k$ ) in each damage level  $k$ , given by:

$$\varepsilon_k = \left| \frac{N_{\text{obs},k} - N_{\text{pred},k}}{N} \right| \quad (2.7)$$

where,  $N_{\text{obs},k}$  is the total number of buildings observed in the given damage grade  $k$ .

Similarly, the heuristic damage assessment model was also compared with the mean damage relationship (Eq. 2.4) applied to the test set. Thus, for each building class in the test set, the error value (Eq. 2.7) for each DG was computed from the  $\mu_d$  on the observed damage using Eq. (2.4), the probability  $p_k$  of obtaining a given DG  $k$  ( $k= 0$  to  $5$ ) using Eq. (2.3), and the distribution of buildings in each DG  $N_{\text{pred},k}$  for a given MSI value using Eq. (2.6).

## 2.4 Result

### 2.4.1 First stage: model selection

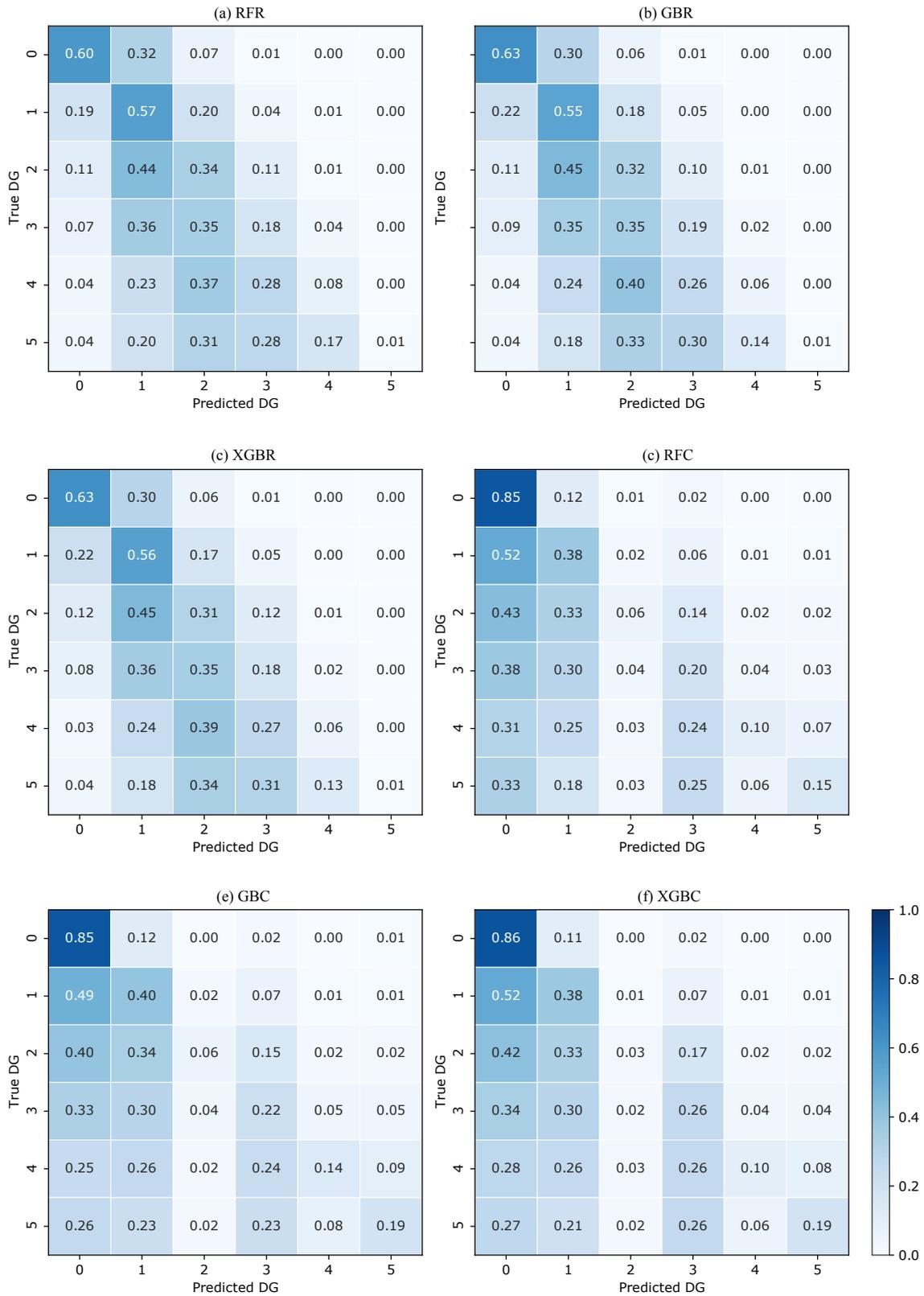
The efficacy of the regression (RFR, GBR, XGBR) and classification (RFC, GBC, XGBC) machine learning models trained and tested on the randomly selected 60% (training set) and 40% (test set) of the 2009 -L'Aquila earthquake building damage portfolio is summarized in Table 2.3. The hyperparameters indicated in Tab. 2.3 were chosen after tests performed by Ghimire et al. (2021). The regression-based machine learning models RFR, GBR and XGBR yielded similar MSE scores (1.22, 1.22 and 1.21) and accuracy scores ( $A_T = 0.49, 0.50$  and  $0.50$ ), considering the five DGs of the EMS-98 scale. In the confusion matrix (Fig. 2.3a: RFR, Fig. 2.3b: GBR, and Fig. 2.3c: XGBR), the accuracy  $A_{DG}$  values show that the efficacy of these models is higher for the lower DGs (around 60% for DG0 and 55% for DG1) and lower for the higher DGs (6% and 1% of the buildings are correctly classified in DG4 and DG5, respectively).

For the classification-based machine learning models, the XGBC model ( $[MSE, A_T] = [1.78, 0.59]$ ) was more effective than the RFC ( $[MSE, A_T] = [1.86, 0.57]$ ) and GBC ( $[MSE, A_T] = [1.80, 0.58]$ ) models, considering the EMS-98 scale. In the confusion matrix (Fig. 2.3d: RFC, Fig. 2.3e: GBC, and Fig. 2.3f: XGBC), the accuracy  $A_{DG}$  values also show higher model efficacy for the lower DGs (86% for DG0 and 39% for DG1) and lower efficacy for the higher DGs (5%, 23%, 12% and 17% buildings correctly classified in DG2, DG3, DG4 and DG5, respectively).

**Table 2.3.** Summary of optimized hyperparameters parameters, accuracy  $A_T$  and quantitative statistical error values for the regression-based and classification-based machine learning methods. The parameters are the hyperparameters chosen for the machine learning models (the other hyperparameters not mentioned here are the default parameters in the Scikit-learn documentation (Pedregosa et al., 2011)). The best accuracy and error values are indicated in bold.

Method	Parameters	Accuracy $A_T$	MSE	MAE
RFR	n_estimators = 1000 max_depth = 25	0.49	1.22	0.77
GBR	n_estimators = 1000 max_depth = 10 learning_rate = 0.01	0.50	1.22	0.77
XGBR	n_estimators = 1000 max_depth = 10 learning_rate = 0.01	0.50	<b>1.21</b>	0.76
RFC	no_estimators = 1000 max_depth = 25	0.57	1.86	0.77
GBC	no_estimators = 1000 max_depth = 10 learning_rate = 0.01	0.58	1.80	0.77
XGBC	n_estimators = 1000 max_depth = 10 learning_rate = 0.01	<b>0.59</b>	1.78	<b>0.74</b>

The classification-based machine learning models thus yielded slightly better predictive efficacy, but still lower than recent studies applied to other datasets (Ghimire et al., 2022; Harirchian et al., 2021; Mangalathu et al., 2020; Roeslin et al., 2020; Stojadinović et al., 2021). The high classification error in the higher DGs could be related to the characteristics of the building portfolio and the imbalance of DG distribution. Among the classification methods, the XGBC model showed slightly higher classification efficacy; the XGBC model was therefore selected for the next stages 2.4.2 and 2.4.3.



**Figure 2.3.** Normalized confusion matrix between predicted and observed DGs. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . All values are also represented by the colour scale.

## 2.4.2 Second stage: issues related to machine learning

### 2.4.2.1 Imbalance distribution of the DGs in the DaDO

The efficacy of the heuristic damage assessment model depends on the distribution of target features in the training dataset. This can lead to low prediction efficacy, especially for minority classes (Estabrooks and Japkowicz 2001; Japkowicz and Stephen 2002; Branco et al. 2017; Ghimire et al., 2022). The previous section reports significant misclassification associated with the highest DGs for all classification- and regression-based models (Fig. 2.3), i.e., for the DGs with the lowest number of buildings (Fig. 2.2a). The efficacy of the XGBC model is analysed below, addressing the class-imbalance issue with data resampling techniques applied to the training phase and considering the L'Aquila-2009 portfolio.

Four strategies to solve the class imbalance issue were tested:

- (a) random undersampling: randomly selecting the number of data entries in each class equal to the number of data entries in the minority class (DG4 in our case);
- (b) random oversampling: randomly replacing the number of data entries in each class equal to the number of data entries in the majority class (DG0 in our case);
- (c) Synthetic Minority Oversampling Technique (SMOTE): creating an equal number of data entries in each class by generating synthetic samples by interpolating the neighbouring data in the minority class;
- (d) a combination of oversampling and undersampling methods: oversampling of the minority class using the SMOTE method, followed by the Edited Nearest Neighbours (ENN) undersampling method to eliminate data that is misclassified by its three nearest neighbours (SMOTE-ENN).

Fig. 2.4 shows the confusion matrices of the four strategies considered for the class imbalance issue. Compared with Fig. 2.3f (i.e., XGBC), the effects of addressing the issue of imbalance were as follows:

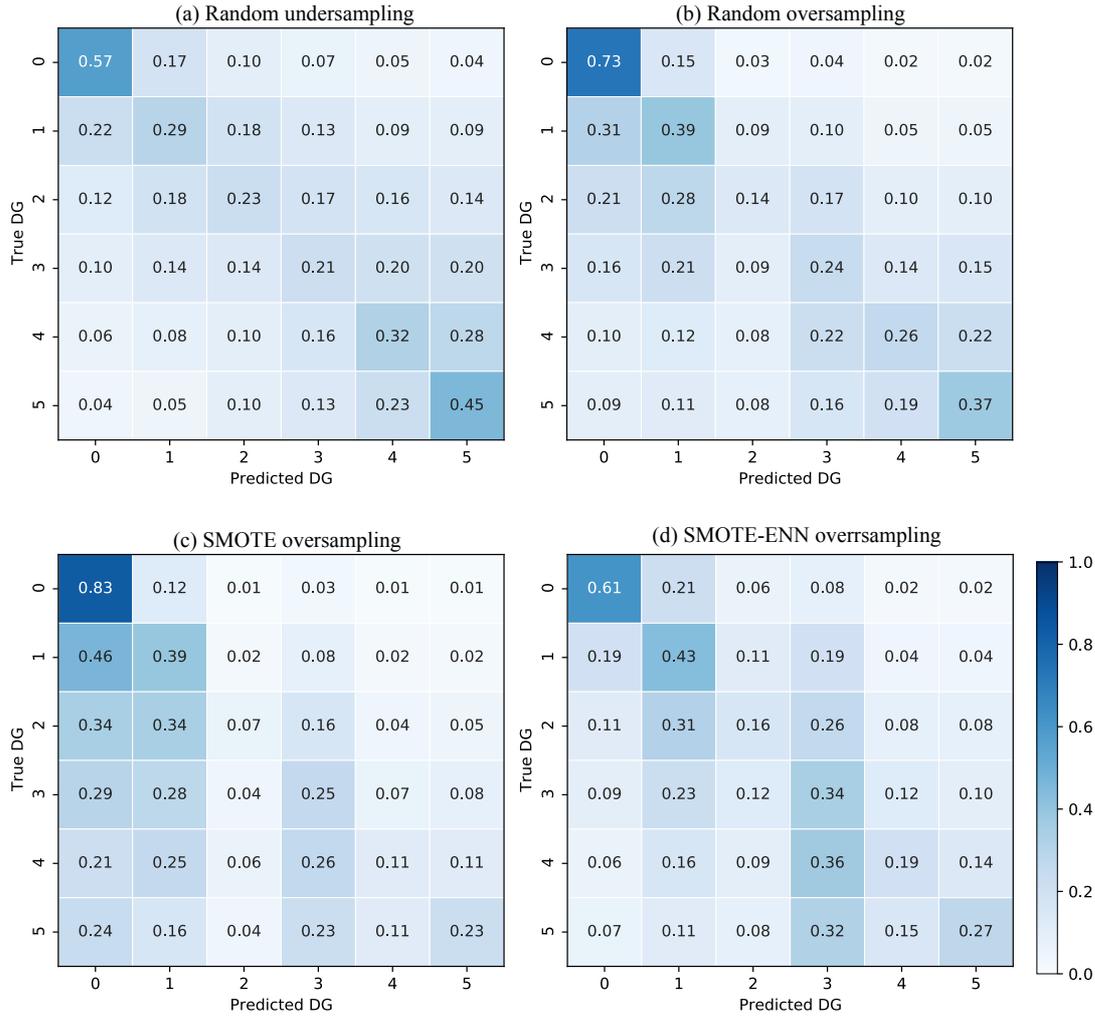
- (a) undersampling (Fig. 2.4a):  $A_{DG}$  value increased by 20/22/26% for DG2/DG4/DG5 and decreased by 29% for DG0.
- (b) oversampling (Fig. 2.4b):  $A_{DG}$  value increased by 11/16/18% for DG2/DG4/DG5 and decreased by 13% for DG0
- (c) SMOTE (Fig. 2.4c):  $A_{DG}$  value increased by 4/1/4% for DG2/DG4/DG5 and decreased by 3% for DG0
- (d) SMOTE-ENN (Fig. 2.4d):  $A_{DG}$  value increased by 13/9/8% for DG2/DG4/DG5 and decreased by 25% for DG0.

The  $A_T$ , MAE and MSE scores are given in Table 2.4 with the associated effects.

**Table 2.4** – Scores of the accuracy  $A_T$ , MSE and MAE metrics considering the imbalance issue and their variation  $\Delta$  compared with values without consideration of the imbalance.

Method	Accuracy $A_T$		MSE		MAE	
	Scores	$\Delta$	Score	$\Delta$	Score	$\Delta$
Undersampling	0.26	-0.33	1.24	-0.34	1.20	0.46
Oversampling	0.53	-0.06	2.13	0.35	0.86	0.12
SMOTE	0.57	-0.02	1.87	0.09	0.77	0.03
SMOTE-ENN	0.49	-0.10	2.28	0.50	0.93	0.19

In conclusion, the random oversampling method improves prediction in the minority class without significantly decreasing prediction in the majority class. The random oversampling method was therefore applied in this study.



**Figure 2.4.** Confusion matrices for the four methods to solve the DG imbalance issue in the DaDO. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . All values are also represented by the colour scale.

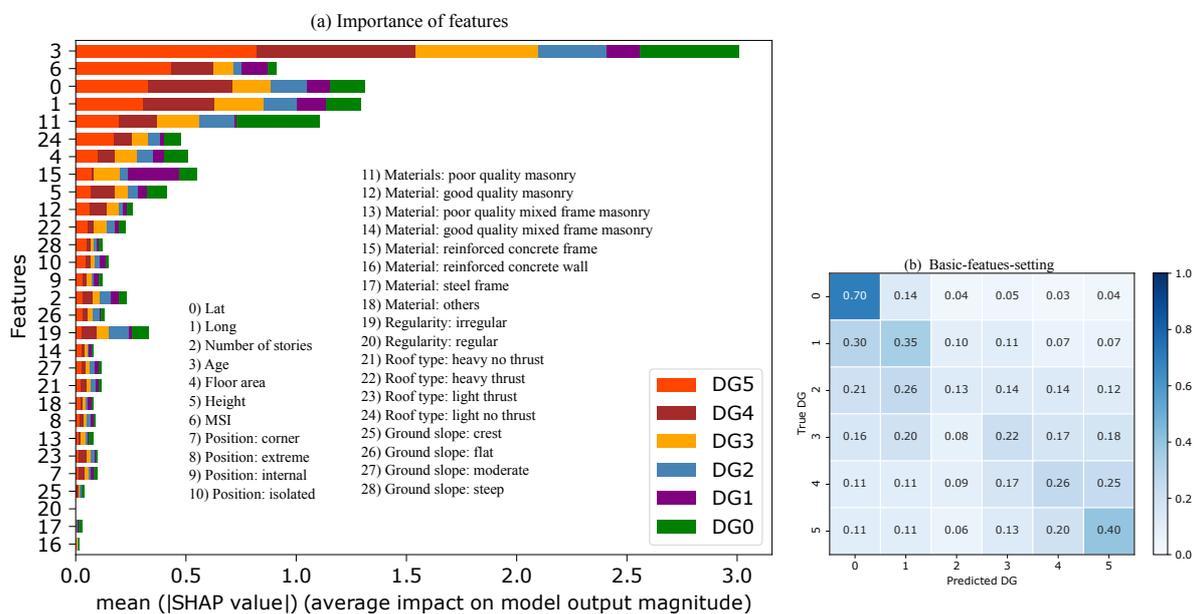
### 2.4.2.2 Testing the XGBC model with basic features

This section begins by exploring the importance of each feature in the heuristic damage assessment model applied to the L'Aquila-2009 portfolio. We used the Shapely Additive Explanations (SHAP) method developed by Lundberg and Lee (2017). The SHAP method compares the efficacy of the model with and without considering each input feature to measure its average impact, provided in terms of mean absolute SHAP values.

Figure 2.5a shows the average SHAP value associated with each feature considered in this study as a function of DG. The most weighted features are building age, location (latitude and longitude), material (poor quality masonry, RC frame), MSI, roof type, floor area, and height. Interestingly, the mean SHAP values are dependent on the DG, i.e., the weight of the feature is not linear depending on the DG considered; this is never taken into account in vulnerability methods. For example, Scala et al. (2022)

and Del Gaudio et al. (2021) observed a decrease in the vulnerability of structures as construction year increases, without distinguishing the DG considered, which is not the case herein. Note also that the importance score associated with the location feature can indirectly capture variations in local geological properties and the spatially distributed vulnerability associated with the built-up area of the L'Aquila-2009 portfolio (e.g., the distinction between the historic town and more modern urban areas). Furthermore, the average SHAP value obtained for poor quality masonry buildings for DG3/DG4/DG5 confirms the same high vulnerability of this typology as in the EMS-98 scale (Grünthal, 1998), regardless of DG.

Some basic features of the building (e.g., location, age, floor area, number of storeys, height) are observed with a high mean SHAP value (Fig. 2.5a). Compared with others, these five basic features can be easily collected from the field or provided by national census databases, for example. Fig. 2.5b shows the efficacy of the heuristic damage assessment model using XGBC trained with a set of easily accessible building features (i.e., basic-features-setting: geographic location, floor area, number of stories, height, age, MSI), after addressing the class-imbalance issue using the random oversampling method. Compared with Fig. 2.4b (considering all features and named as the full-features-setting), the XGBC model with the basic-features-setting (Fig. 2.5b) gives almost the same efficacy with only a 6% average reduction in the accuracy scores.

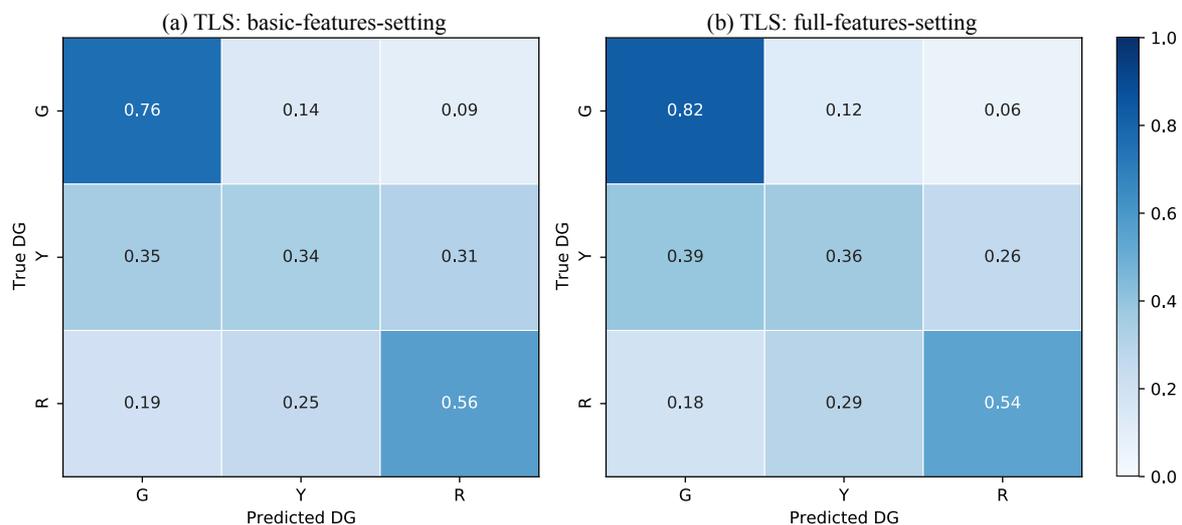


**Figure 2.5.** (a) Graphic representation of the importance scores associated with the different input features considered for the XGBC model. The features (the same as in Fig. 2.2) considered in this study are on the y-axis, and the x-axis is the mean SHAP score according to DG. (b) Confusion matrices considering the basic-features-setting. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . All values are also represented by the colour scale.

### 2.4.2.3 Testing the XBGC model with the traffic-light system for damage grades

In this section, a simplified version of the DG scale was used, in the sense that the DGs are classified according to a traffic-light system (TLS) (i.e., green G, yellow Y and red R classes, corresponding to DG0+DG1, DG2+DG3 and DG4+DG5, respectively), as monitored during post-earthquake emergency situations (Mangalathu et al., 2020a; Riedel et al., 2015; ATC, 2005; Bazzurro et al., 2004). For the TLS-based damage classification, the XBGC model (after oversampling to compensate of the imbalance issue) with the basic-features-setting applied to the L'Aquila-2009 portfolio (Fig. 2.6a) gives almost the same efficacy compared to the full-features-setting (Fig. 2.6b). For example, accuracy values  $A_{DG}$  using the basic-features-setting and the full-features-setting were 0.76/0.34/0.56 and 0.82/0.36/0.54 for G/Y/R classes, with the accuracy score  $A_T$  of 0.68 and 0.72, respectively. Mangalathu et al. (2020), Roslin et al., (2020), and Harirchian et al., (2021) reported similar damage grade classification accuracy values of 0.66, 0.67, and 0.65 respectively.

The efficacy of the heuristic damage assessment model using TLS-based damage classification indicates that classifying damage into three classes is much easier for the machine learning model compared with the six-class classification system (EMS-98 damage classification). This is also observed during damage surveys in the field, which sometimes find it hard to distinguish the intermediate damage grades, such as DG2 and DG3, or DG3 and DG4. Similar observations have been reported in previous studies by Guettiche et al., (2017); Harirchian et al., (2021); Riedel et al., (2015); Roeslin et al., (2020) and Stojadinović et al., (2021).



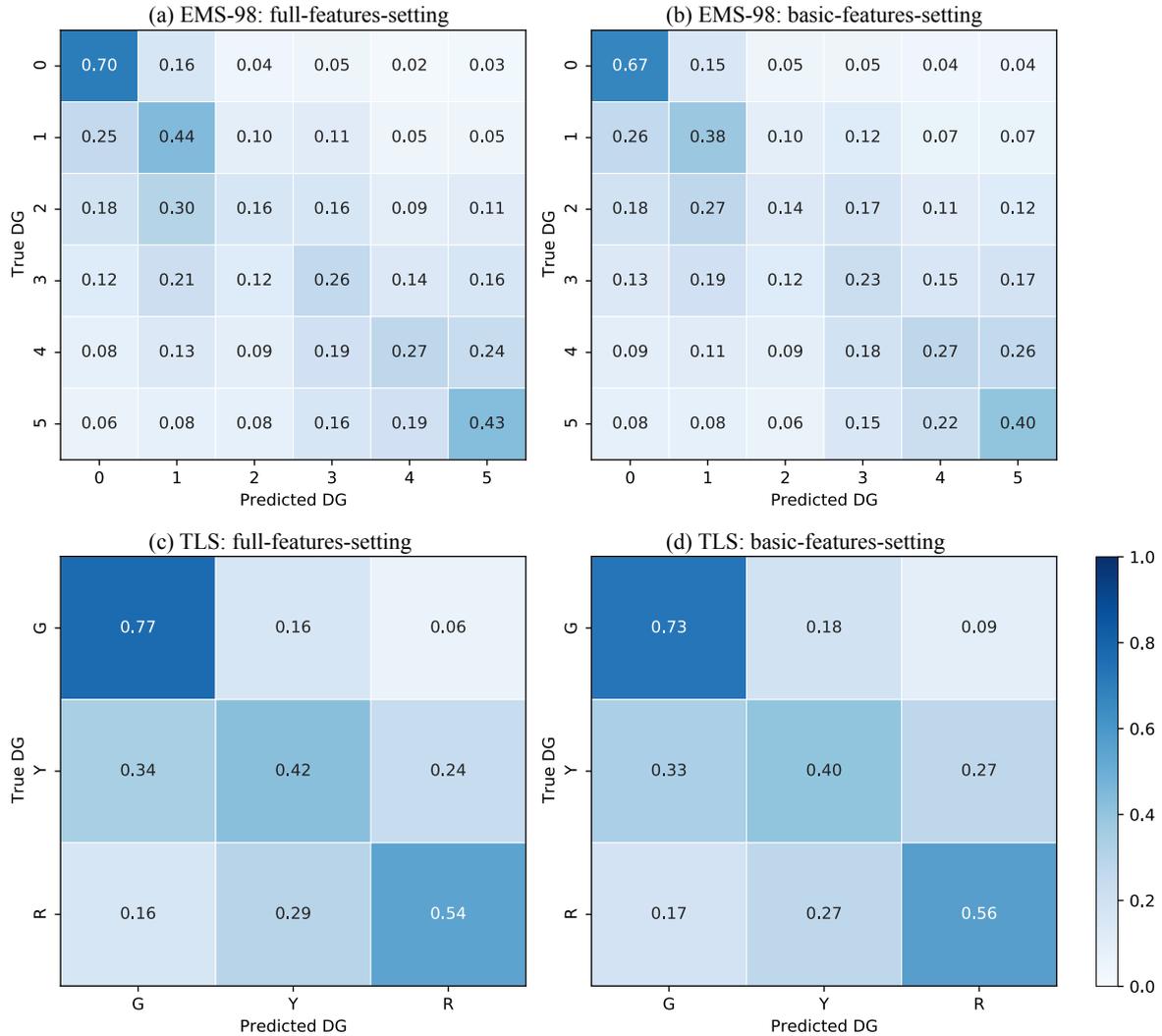
**Figure 2.6.** Confusion matrices for (a) the basic-features-setting and (b) the full-features-setting using the traffic-light (TLS)-based classification, grouping the EMS-98 damage grades (DG) into three classes (green for no or slight damage; yellow for moderate damage; and red for heavy damage). The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . All values are also represented by the colour scale.

#### 2.4.2.4 Testing the XGBC model with the whole dataset

The efficacy of the XGBC model was tested using a dataset with six building damage portfolios, excluding the 1980-Irpinia building damage portfolio. The XGBC model was trained and tested on the randomly selected 60% (training set) and 40% (test set) of the dataset for EMS-98/TLS damage classification, with two sets of features (full-features-setting and basic-features-setting), applying the random oversampling method to compensate for class-imbalance issues. Fig. 2.7 shows the associated confusion matrix.

The basic-features-setting resulted in a similar level of damage prediction compared with the full-features setting for both EMS-98 and TLS-based damage classification systems. For EMS-98 damage classification (Fig. 2.7a, b), the accuracy  $A_{DG}$  scores indicated in the confusion matrices are almost the same for the basic-features-setting and the full-features-setting. Furthermore, the accuracy  $A_T$  and MAE scores are also almost the same (0.45 and 1.08 for the basic-features-setting and 0.48 and 0.95 for the full-features-setting).

Likewise, for TLS-based damage classification (Fig. 2.7c, d), the accuracy values  $A_{DG}$  for the basic-features-setting and the full-features-setting are almost the same, with similar accuracy  $A_T$ / MAE scores (0.63/0.45 and 0.67/0.39, respectively).



**Figure 2.7.** Confusion matrices for EMS-98 (a, b) and TLS (c, d) damage classification systems using the basic- and full-features-settings (green for no or slight damage; yellow for moderate damage; red for heavy damage) with (c) the full-features-setting and (d) the basic-features-setting. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . All values are also represented by the colour scale.

### 2.4.3 Third stage: application to the whole DaDO portfolio and comparison with Risk-UE

In this section, the efficacy of the heuristic damage assessment model was considered for building damage predictions, without respecting the time frame of the earthquakes. Two scenarios were considered: (1) a single building damage portfolio was used for training and the model was then tested on the others (named single-single), in situations using a single portfolio to predict future damage; and (2) some building damage portfolios were used for training but testing was performed on a single portfolio (named aggregate-single), i.e. a larger number of damage portfolios were used as a training

set to predict the damage caused by the next earthquake. The model XGBC was applied with the basic-features-setting (number of storeys, building age, floor area, height, MSI for EMS-98) and EMS-98- and TLS-based damage classification.

#### **2.4.3.1 Single-single scenario**

First, a series of building damage portfolios, concerning earthquakes occurring in northern or southern Italy and of different magnitudes, was used for training and testing:

Training set: E3 – test set: E1, E5, E7.

Training set: E5 – test set: E1, E3, E7.

Training set: E7 – test set: E1, E3, E5.

Figure 2.8 shows the distribution of correct DG classification (i.e.,  $1 - \varepsilon_d$  in % given by Eq. 2.1) observed for each building for the EMS-98 damage grade (2.8a) and the TLS (2.8b) systems. The x-axis represents the incremental error in the damage grade (e.g., 1 corresponds to the delta of damage grade between observation and prediction, regardless of the DG considered).

For the EMS-98 damage scale, correct classification (x-value centred on 0) in the range of 31% to 48% was found, depending on the training/test data sets. The error distribution is quite wide with incorrect predictions of +/-1 DG in the range of +/- 13-35%. Remarkably, when considering the E1 portfolio (Irpinia-1980), for which the post-earthquake inventory was based on another form, as the test set, the error is larger. The predictions at +/-1 DG (i.e., the sum of the x-values Fig. 2.8a between -1 and +1) were 70.5%, 69.9% and 72.8% with portfolios E3, E5 and E7 as the test set, respectively, for an average of 71%. For the other portfolios, the average of the predictions at +/- 1 DG was 77%, 78% and 77%, respectively, for portfolios E5, E3 and E7 as the test set. This tendency was also observed for the TLS damage system (Fig. 2.8b). In this case, the classification of the E1 portfolio was correct on average (average of x-values centred on 0) at 63% and equal to 72%, 73%, and 70.5% for the test on portfolios E5, E3, and E7. For both damage scales, the distributions were skewed, with a larger number of predictions being underestimated (positive x-values), as certainly a consequence of the choice of machine learning models, their implementation (including imbalance issues), the distribution of input and target features considered, or all. The interest of machine learning model is also to have a relevant representation of the errors and limits of these methods.

#### **2.4.3.2 Aggregate-single scenario**

Secondly, several aggregated building damage portfolio scenarios were considered to predict a single earthquake, thus testing whether the prediction was improved by increasing the number of post-earthquake damage observations. Three scenarios were tested. They are represented in Fig. 2.9 applying the EMS-98 damage grade (9a) and the TLS (9b):

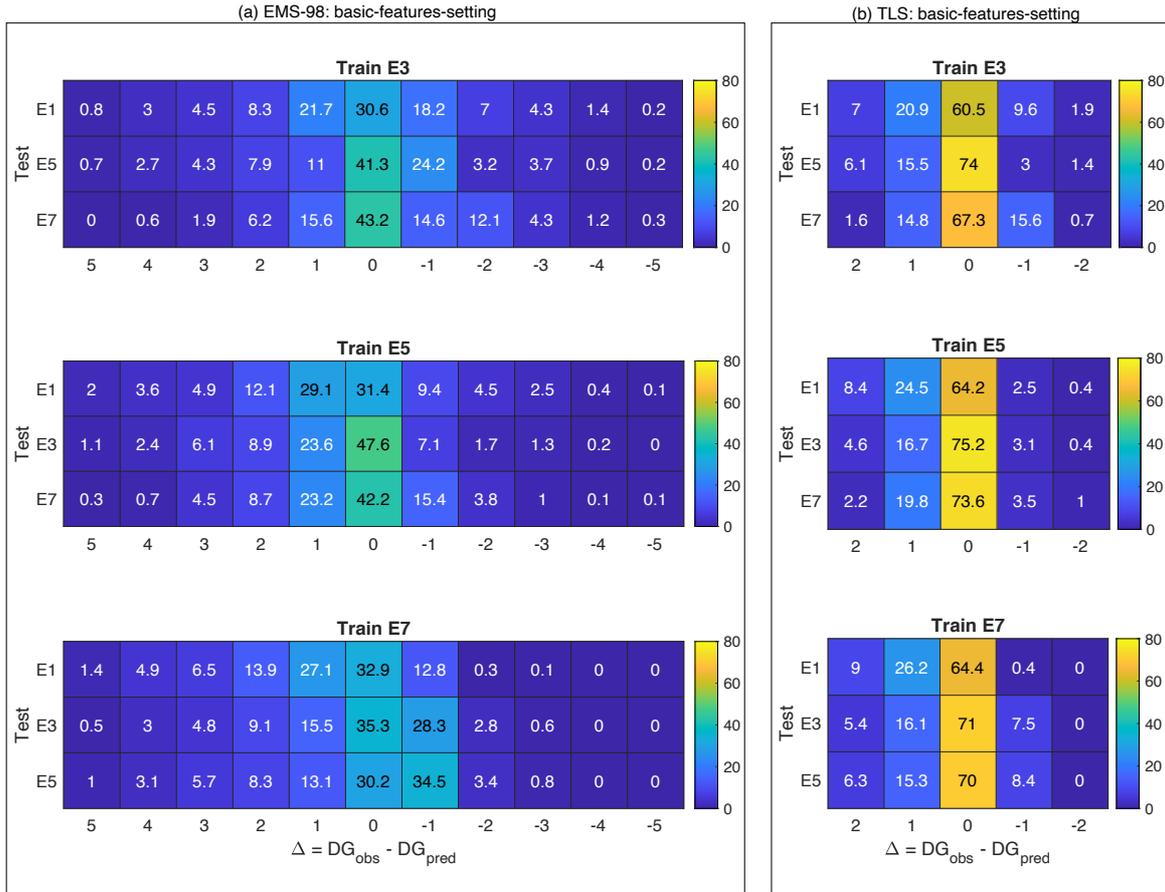
Training set: E2+E3+E4+E6 (shown as E2346) – test set: E1, E5 and E7.

Training set: E2+E4+E5+E6 (shown as E2456) – test set: E1, E3 and E7.

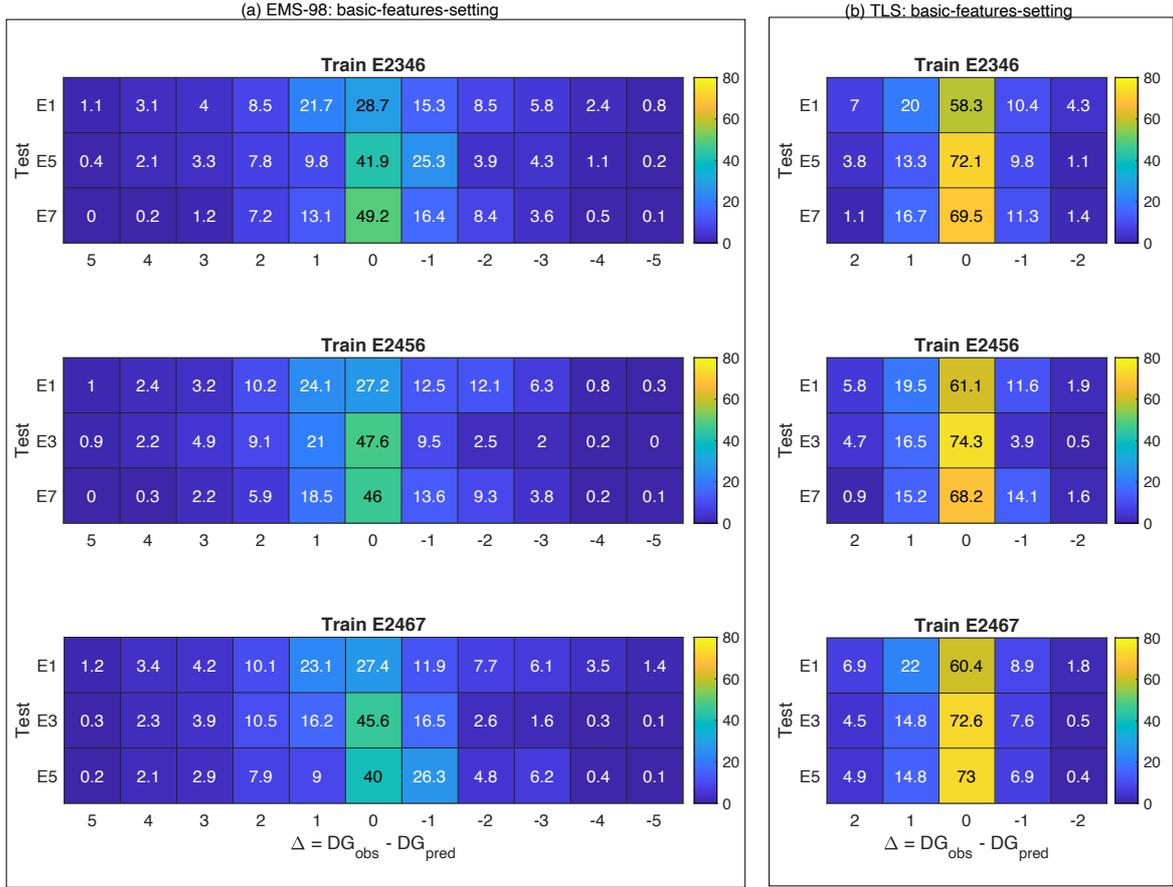
Training set: E2+E4+E6+E7 (shown as E2467) – test set: E1, E3 and E5.

For the EMS-98 damage scale, correct classification (x-value centred on 0) in the range of 27% to 49% was found, depending on the training/test datasets. As in Fig. 2.8, using the E1 (Irpinia-1980) earthquake for testing score is lower regardless of the portfolio used for training (28.7%, 27.2% and 27.4% prediction accuracy). With E1 as the test set, the predictions at +/-1 DG (i.e., the sum of the x-values on Fig. 2.9a between -1 and +1) were 65.7%, 63.8% and 62.4% considering the E2346, E2456 and E2467 portfolios as the training set, respectively, for an average of 64% (compared with the 70% score for the single portfolio scenario, Fig. 2.8a). Other scenarios were also tested by aggregating the building damage portfolios differently (not presented herein), leading to the two main conclusions: (1) the quality and homogeneity of the input data (i.e., building features) affect the efficacy of the heuristic model and (2) this efficacy is limited and not improved by increasing the number of building damage observations, with a score (excluding E1) between 40% and 49% (x-value centred on 0), and up to 78% (average of the two scenarios, Fig. 2.8a and Fig. 2.9a) at +/-1 DG. Considering the TLS damage scale (Fig. 2.9b), a damage prediction efficacy of about 72% was obtained (compared with 72% in Fig. 2.8b), i.e., but no significant improvement was observed when the number of damaged buildings in the training portfolio was increased. For EMS-98 and TLS, the distributions were skewed, with a larger number of predictions being underestimated (positive x-values).

Finally, in conclusion, the heuristic damage assessment model based on the XGBC model gives a better score for TLS damage assessment than for the EMS-98 damage scale. The TLS system also allows for quick assessment of damage on a large scale such as a city or region from an operational point of view.



**Figure 2.8.** Distribution of the classification value ( $1 - \varepsilon_d$  in % given by Eq. 2.1) for (a) EMS-98- and (b) TLS-based damage classification using XGBC machine learning models and considering a single damage portfolio to predict a single portfolio (single-single scenario). The colour bar indicates the associated value in each cell. The x-values are the difference between the DG observed and the DG predicted, regardless of the DG considered.



**Figure 2.9.** Distribution of the classification value ( $1 - \varepsilon_d$  in % given by Eq. 2.1) for (a) EMS-98- and (b) TLS-based damage classification using XGBC machine learning models and considering an aggregate damage portfolio to predict a single portfolio (aggregate-single scenario). The colour bar indicates the associated value in each cell. The x-values are the difference between the DG observed and the DG predicted, regardless of the DG considered.

### 2.4.3.3 Comparing efficacy with the Risk-UE model

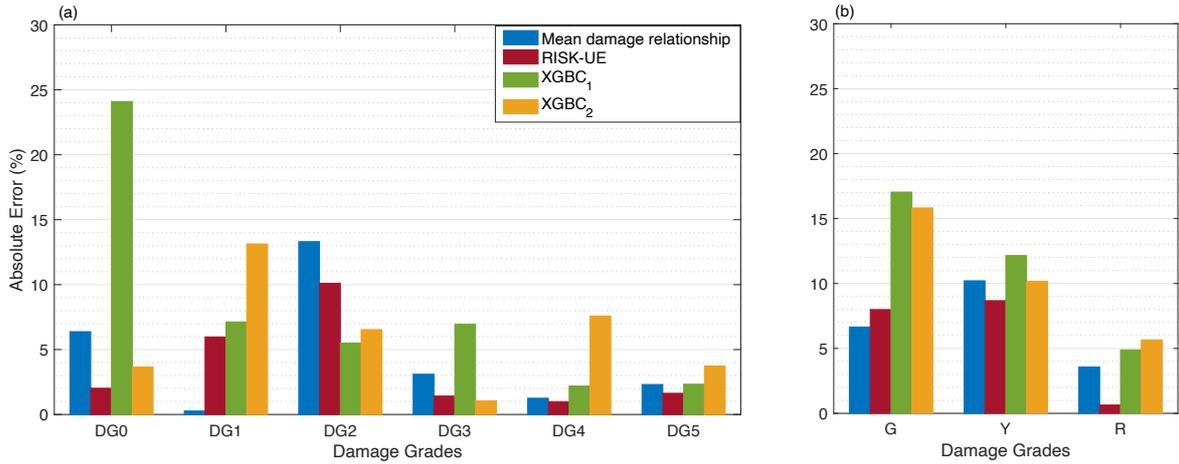
The efficacy of the heuristic damage assessment model was then compared with conventional damage prediction methods, i.e., RISK-UE and mean damage relationship (Eq. 2.2 to 2.7), considering the basic-features-settings. For RISK-UE, mean damage  $\mu_d$  (Eq. 2.4) was computed using the training set and the vulnerability index  $IV$  for each building (Eq. 2.5). A vulnerability index was then attributed to all the buildings in each class defined according to building features. The vulnerability indexes were then attributed to every building in the test set, mean damage ( $\mu_d$ ) was computed with Eq. 2.2 and then DG distribution with Eq. 2.3, before being compared with the damage portfolio used for testing. Finally, the distribution of the mean damage observed (Eq. 2.4) was compared with the distribution of damage directly on the test set, using Eq. 2.3.

Fig. 2.10 shows the distribution of absolute errors associated with the RISK-UE, mean damage relationship, and XGBC methods (with and without compensation for the class-imbalance issue) trained on earthquake building damage portfolio E5 and tested on E3. For EMS-98 damage classification (Fig. 2.10a), the XGBC model (without compensation for class-imbalance issues) resulted in a level of absolute errors similar to that of the RISK-UE and/or mean damage relationship, except for DG0 (24%). Random oversampling to compensate for the class-imbalance issues improved the distribution of errors for the XGBC model (errors less than 8%, except for DG1: 13%).

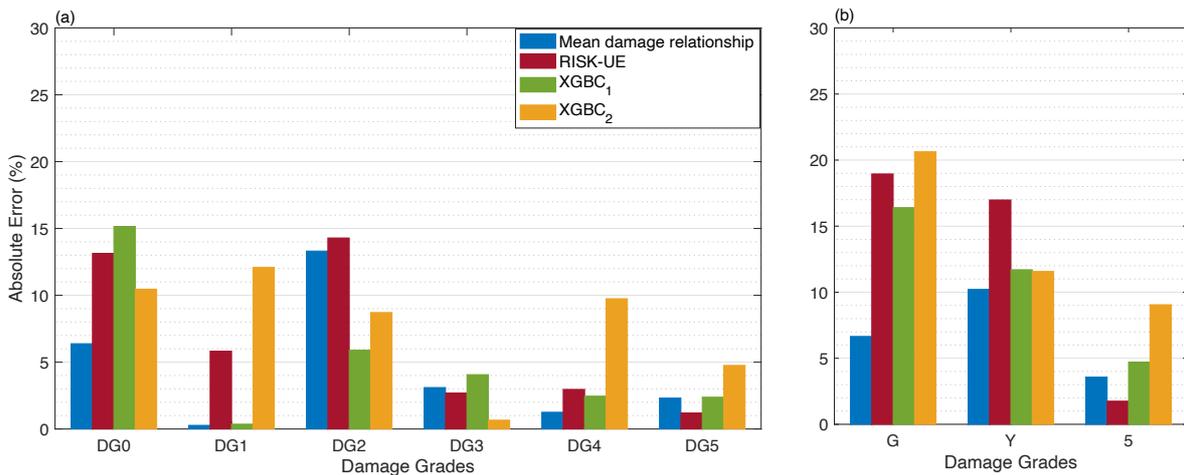
For TLS-based damage classification, the XGBC model also resulted in a similar level of errors compared with the mean damage relationship and/or RISK-UE methods (Fig. 2.10b), except for the green class (no or slight damage, 17.04%). Compensation for class-imbalance issues slightly improved the distribution of errors for the XGBC model with a 2% drop in errors for green (no/slight damage) and yellow (moderate damage) classes.

Figure 2.11 shows the distribution of absolute errors trained using the E2456 portfolio and tested on the E3 portfolio. For EMS-98 damage classification (Fig. 2.11a), the XGBC model (without compensation for class-imbalance issues) resulted in a level of errors similar to that of the RISK-UE and/or mean damage relationship; errors were highest for DG0 with 15.15%. With compensation for the class-imbalance issues, the XGBC model achieved a slightly lower error distribution for DG0 (5%) and DG3 (4%); however, for other damage grades, the error value increased significantly (DG1: 11%, DG2: 12%, DG4: 7%, DG5: 2%). For TLS-based damage classification, the distribution of absolute errors was similar for both the XGBC model and the mean damage relationship and/or RISK-UE methods (Fig. 2.11b). The highest absolute error value was associated with the green (no or slight damage) class of buildings (16.40%). Compensation for the class-imbalance issues slightly increased the error distribution for the XGBC model with nearly 5% for buildings in the green (no or slight) and red (heavy) classes.

These results show that the heuristic building damage model based on the XGBC model, trained using building damage portfolios with the basic-features-setting, provides a reasonable estimation of potential damage, particularly with TLS-based damage classification.



**Figure 2.10.** Comparison of the efficacy of the heuristic model with the conventional model considering the DaDO portfolio (training set: E5; test set: E3) for (a) EMS-98- and (b) TLS-based damage classification. The x-axis is the damage grade and the y-axis is the percentage of absolute error ( $\epsilon_k$  in % given by Eq. 2.7). The blue bar corresponds to the mean damage relationship, the red bar corresponds to the RISK-UE method, the green and orange bars correspond to the heuristic model without (XGBC<sub>1</sub>) and with (XGBC<sub>2</sub>) compensation for the class-imbalance issues, respectively.



**Figure 2.11.** Comparison of the efficacy of the heuristic model with the conventional model considering the DaDO portfolio (training set: E2456; test set: E3) for (a) EMS-98- and (b) TLS-based damage classification. The x-axis is the damage grade and the y-axis is the percentage of absolute error ( $\epsilon_k$  in % given by Eq. 2.7). The blue bar corresponds to the mean damage relationship, the red bar corresponds to the RISK-UE method, the green and orange bars correspond to the heuristic model without (XGBC<sub>1</sub>) and with (XGBC<sub>2</sub>) compensation for the class-imbalance issues, respectively.

## 2.5 Discussion

Previous studies have aimed to test a machine learning framework for seismic building damage assessment (e.g., Mangalathu et al., 2020; Roeslin et al., 2020; Harirchan et al., 2021; Ghimire et al., 2022). They evaluated various machine learning and data balancing methods to classify earthquake damage to buildings. However, these studies (Mangalathu et al., 2020, Roeslin et al., 2020, Harirchan

et al., 2021) had limitations such as limited data samples, damage classes, and building characteristics limited to a spatial coverage and range of seismic demand values. Ghimire et al. (2022) also used a larger building damage database, but did not investigate the importance of input features as a function of damage levels and did not compare machine learning with conventional damage assessment methods.

Our study aims to go beyond previous studies by testing advanced machine learning methods and data resampling techniques using the unique DaDO dataset collected from several major earthquakes in Italy. This database covers a wide range of seismic damage and seismic demands of a specific region, including undamaged buildings. Most importantly, this study highlights the importance of input features according to the degrees of damage and finally compares the machine learning models with a classical damage prediction model (Risk-UE). The machine learning models achieved comparable accuracy to the Risk-UE method. In addition, TLS-based damage classification, using red for heavily damaged, yellow for moderate damage, and green for no to slight damage, could be appropriate when the information for undamaged buildings is unavailable during model training.

Indeed, it is worth noting that the importance of the input features used in the learning process changes with the degree of damage: this indicates that each feature may have a contribution to the damage that changes with the damage level. Thus, the weight of each feature does not depend linearly on the degree of damage, which is not considered in conventional vulnerability methods.

The prediction of seismic damage by machine learning remains until now tested on geographically limited data. The damage distribution is strongly influenced by region-specific factors such as construction quality and regional typologies, implementation of seismic regulations and hazard level. Therefore, machine learning-based models can only work well in regions with comparable characteristics and a host-to-target transfer of these models should be studied. In addition, the distribution of damage is often imbalanced, impacting the performance of machine learning models by assigning higher weights to the features of the majority class. However, data balancing methods like random oversampling can reduce bias caused by imbalanced data during the training phase, but they may also introduce overfitting issues depending on the distribution of input and target features. Thus, integrating data from a wider range of input features and earthquake damage from different regions, relying on a host-to-target strategy, could help achieve a more natural balance of data sets and lead to less biased results. Moreover, the machine learning methods trained only on the data available in the learning phase, reflects the building portfolio in the study area. The importance of the features contributing to the damage could thus be modulated, and would require a host-to-target adjustment for the application of the model to another urban zone/seismic region.

However, the machine learning models trained and tested on the DaDO dataset resulted in similar damage prediction accuracy values reported in existing literature using different models and datasets with different combinations of input features. This might suggest that the uncertainty related to building

vulnerability in damage classification may be smaller than the primary source of uncertainty related to the hazard component (such as ground motion, fault rupture, slip duration, etc.).

In recent years, there has been a proliferation of open building data, such as the OpenStreetMap-based dynamic global exposure model (Schorlemmer et al., 2020) and building damage dataset after an earthquake (such as DaDO). We must therefore continue this paradigm shift initiated by Riedel et al. (2014, 2015) which consisted in identifying the exposure data available and as certain as possible, and in finding the most effective relationships for estimating the damage, unlike conventional approaches which proposed established and robust methods but relying on data not available and therefore difficult to collect. The global dynamic exposure model will make it possible to meet the challenge of modelling exposure on a larger scale on available data, using a tool capable of integrating this large volume of data. Machine learning methods are one such rapidly growing tool that can aid in exposure classification and damage prediction by leveraging readily available information. It is therefore necessary to continue in this direction in order to evaluate the performance of the methods and their pros and cons for maximum efficacy of the prediction of damage.

Future works will therefore have to address several key issues that have been discussed here but that need to be further investigated. For example, the weight of the input features varies according to the level of damage, but one can question the systematization of this observation whatever the dataset and the feature considered. The efficiency of the selected models and the management of imbalance data remain to be explored, in particular by verifying regional independence. Taking advantage of the increasing abundance of exposure data and post-seismic observations, the imbalanced feature distribution and observed damage levels could be solved by aggregating datasets independent of the exposure and hazard contexts of the regions, once the host-to-target transfer of the models has been resolved. Finally, key input features (still not yet identified) describing hazard or vulnerability may be unexplored, and incorporating them into the models may improve the accuracy of damage classification.

## **2.6 Conclusion**

In this study, we explored the efficacy of machine learning models trained using DaDO post-earthquake building damage portfolios. We compared six machine learning models: RFC, GBC, XGBC, RFR, GBR, and XGBR. These models were trained on numbers of building features (location, number of storeys, age, floor area, height, position, construction material, regularity, roof type, ground slope condition) and ground motion intensity defined in terms of macro-seismic intensity. The classification models performed slightly better than the regression methods and the XGBC model was ultimately found to be the most efficient model for this dataset. To solve the imbalance issue concerning observed damage, the random oversampling method was applied to the training dataset to improve the efficacy of the heuristic damage assessment model by rectifying the skewed distribution of the target features (DGs).

Surprisingly, we found that the weight of the most important building feature evolves according to DG, i.e., the weight of the feature for damage prediction changes depending on the DG considered, which is not taken into account in conventional methods.

The basic-features-setting (i.e., considering number of storeys, age, floor area, height and macroseismic intensity, which are accurately evaluated for the existing building portfolio) gave the same accuracy (0.68) as the full-features-settings (0.72) with the TLS-based damage classification method. For training and testing, the homogeneity of the information in the portfolios is a key issue for the definition of a highly effective machine learning model, as shown by the data from the E1 earthquake (Irpinia-1990). However, the efficacy of the model reaches a limit which is not improved by increasing the number of damaged buildings in the portfolio used as the training set, for example. For damage prediction, this type of heuristic model results in approximately 75% correct classification. Other authors (e.g., Riedel et al., 2014, 2015; Ghimire et al. 2022) have already reached this same conclusion by increasing the percentage of the training set compared with the test set.

Despite this limit threshold, the level of accuracy achieved remains similar to that attained by conventional methods, such as Risk-UE and the mean damage relationship, for the basic-features-settings and TLS-based damage classification (error values less than 17 %). Machine learning models trained on post-earthquake building damage portfolios could provide a reasonable estimation of damage for a different region with similar building portfolios, after host-to-target adjustment.

Some variability may have been introduced into the damage prediction model due to the framework defined to translate the original damage scale to the EMS-98 damage scale and because in the DaDO database, the year of construction and the floor area of each building are provided as interval values, and missing locations of buildings were replaced with the location of local administrative centres. The latter can lead to a smoothing of the macro-seismic intensities to be considered for each structure and also affect the distance to the earthquake. Similarly, the building damage surveys were carried out after the seismic sequence, which includes aftershocks as well as the mainshock, whereas the MSI input corresponds to the mainshock from the USGS ShakeMap. All these issues may reduce the efficacy of the heuristic model and its limit threshold. Addressing these issues could improve the damage prediction performance of machine learning models.

# 3 Testing host-to-target transfer of machine learning models for regional-scale damage assessment

---

In chapter 2, we analyzed the efficacy of machine learning models trained on past earthquake building damage datasets to predict potential damage in future earthquakes in regions with similar building portfolios. This chapter focuses on the influence of a building's characteristic features, such as typology, construction practice, seismic regulations, etc., on transferring machine learning models for damage prediction from host-to-target dataset. In this chapter, we compared the efficacy of machine learning models developed in Nepal, Italy, and Haiti for large-scale damage assessment.

---

## Abstract

This study evaluates the effectiveness of machine learning models for classifying seismic damage, using post-earthquake observation of the 2010 Haiti earthquake, the 2015 Nepal earthquake, and the database of observed damage from Italy. Two regression and two classification methods, random forest and extreme gradient boosting, were compared. The result showed that the extreme gradient boosting classification (XGBC) achieved higher accuracy (0.57 for three-class classification) used in post-earthquake damage surveys. Furthermore, the XGBC model estimated building damage using readily accessible features such as the building's age, number of storeys, and MSI, with accuracy score of 0.45 (building-by-building comparison) but error value below 10% when analyzed by aggregating buildings according to the damage grades in the test dataset, indicating its suitability for large-scale portfolio level rapid damage classification. A comparative study on damage prediction using XGBC models trained and tested with age, number of storeys, and MSI from building damage datasets in Haiti, Nepal, and Italy revealed varying efficacy and damage distribution, emphasizing the influence of factors such as building typologies, construction practices, seismic regulations, etc.

**Keywords:** Machine learning, 2010 Haiti earthquake building damage dataset, host-to-target transfer, large-scale damage assessment.

## 3.1 Introduction

The main objective of seismic risk mitigation is to ensure people's safety and protect their sources of sustenance by minimizing earthquake threats (Bommer, 2022). To achieve this objective, accurate risk assessment requires a thorough examination of seismic damage as a crucial component of the process. Advanced methods have been developed to estimate seismic damage (among other, FEMA, 2003;

Milutinovic and Trendafiloski, 2003; Porter, 2003; Vamvatsikos and Cornell, 2004; Silva et al., 2014; Guéguen et al., 2007; Hancilar et al., 2010; Lagomarsino and Giovinazzi, 2006). These methods provide damage as a function of seismic hazard, exposure models describing the portfolio of buildings, and the vulnerability/fragility functions describing the damage probability for building typologies. However, for large-scale damage assessment, the information necessary for these methods is often sparse, incomplete, or low resolution, making regional and national-scale damage assessment very challenging and time-consuming (Riedel et al., 2015).

Communities worldwide are collaborating to build exposure models and fragility/vulnerability functions for large-scale seismic damage assessment by aggregating available information from existing databases (Crowley et al. 2020). Similarly, efforts have been made to test novel methods that use satellite and ground-based remote sensing for exposure modeling and damage classification (Pittore and Wieland, 2013; Wieland et al., 2015; Stone et al., 2018; Rao et al., 2022). In parallel, platforms outlining basic building features suitable for damage assessment have increased in number (e.g., OpenBuildingMaps (Schorlemmer et al., 2017); the National Institute of Statistics and Economic Studies in France (INSEE) ([www.insee.fr](http://www.insee.fr)); global exposure database (Gamba et al., 2012); the World Housing Encyclopaedia (<http://db.world-housing.net/>); e-Stat Japan (<https://www.e-stat.go.jp/>)). These platforms offer access to the low-resolution, high-density building information, which opens an opportunity to be used for seismic damage assessment (Riedel et al., 2015). In this context, Riedel et al. (2014, 2015) investigated machine learning methods to estimate seismic damage on a regional and national level using readily available data for many buildings in the INSEE database.

On the other hand, numerous studies have investigated the effectiveness of machine learning methods in accurately connecting building features and spatially-distributed ground motion to different levels of damage, with satisfactory levels of accuracy when compared to traditional methods or tested using post-earthquake observations (e.g., Riedel et al. 2015; Harirchian et al., 2021; Mangalathu et al., 2020; Roeslin et al., 2020; Stojadinović et al., 2021; Ghimire et al., 2022, 2023). In parallel, significant efforts have been made to collect and share post-earthquake building damage observations after damaging earthquakes (MTPTC, 2010; NPA, 2021; MINUV, 2021; Dolce et al., 2019; Omoya et al., 2022; Stojadinović et al., 2021). By taking advantage of the convergence of situations, i.e. rapidly developing learning methods and the increasing amount of available building open-data, it is time to evaluate the efficiency conditions of the damage classification methods, in particular for different post-earthquake data according to the building criteria and damage classification, but also for the construction of a building damage model through host-to-target region adjustments, i.e. developing a machine from one specific region in terms of seismicity and building design to be applied to another region.

The main objective of this study is to perform a comparative study on damage prediction across different regions using machine learning models trained on the readily available building features from the post-earthquake damage database. Therefore, this study compares the damage prediction effectiveness of

machine learning models trained on the age, number of storeys, and macroseismic intensity using building damage datasets from Haiti, Nepal, and Italy. The dataset and methods are described in the 3.2 and 3.3 sections. The 3.4 and 3.5 section presents the results of the comparative study of damage prediction, followed by a conclusion in 3.6.

## 3.2 Data

### 2010-Haiti earthquake building damage dataset (HBDP)

A Mw 7 earthquake struck Haiti on January 12, 2010, with the epicentre located 25 km SW of Port-au-Prince, with a hypocentre depth of 13 km and a 15-40 km long rupture, followed by three large aftershocks 6.0 Mw, 5.7 Mw, and 5.9 Mw, causing over 300,000 casualties, leaving over 1.3 million people homeless, and resulting an estimated losses of US\$ 7 to US\$ 14 billion, exceeding the Haiti gross domestic product (DesRoches et al., 2011). The government of Haiti conducted a massive post-earthquake damage survey with the help of more than 300 trained engineers, alongside a third-party structural engineers, and developed a database of observed damage (MTPTC, 2010).

During the field survey of damage, the ATC-20 damage classification methodology (ATC, 2005) was adopted for damage classification (MTPTC, 2010).

The damage was grouped into seven discrete classes based on visual observation.

**Damage grades (DG):** none for no damage (DG0) slight for 0-1% damage (DG1), light for 1-10% damage (DG2), moderate for 10-30% damage (DG3), heavy for 30-60% damage (DG4), major for 60-100% damage (DG5) and destroyed for 100% damage (DG6), respectively.

In the same survey, building features related to structural typology were also collected in the field:

**Number of storeys-** total number of floors above the ground surface.

**Age of building-** time difference between the date of the earthquake and the date of building construction/renovation; buildings were grouped into four categories according to age: 0-10 years, 11-25 years, 26-50 years, and older than 50 years.

**Floor plan-** geometric shape defining the building plan; eight types of plans were defined: E-shape, H-shape, L-shape, O-shape, Rectangular-shape, T-shape, U-shape, and Other-shape.

**Wall type-** materials used in vertical wall; seven wall types were defined: block-masonry with reinforcement, block-masonry without reinforcement, brick-masonry, reinforced concrete, stone-masonry, wood-masonry, and others.

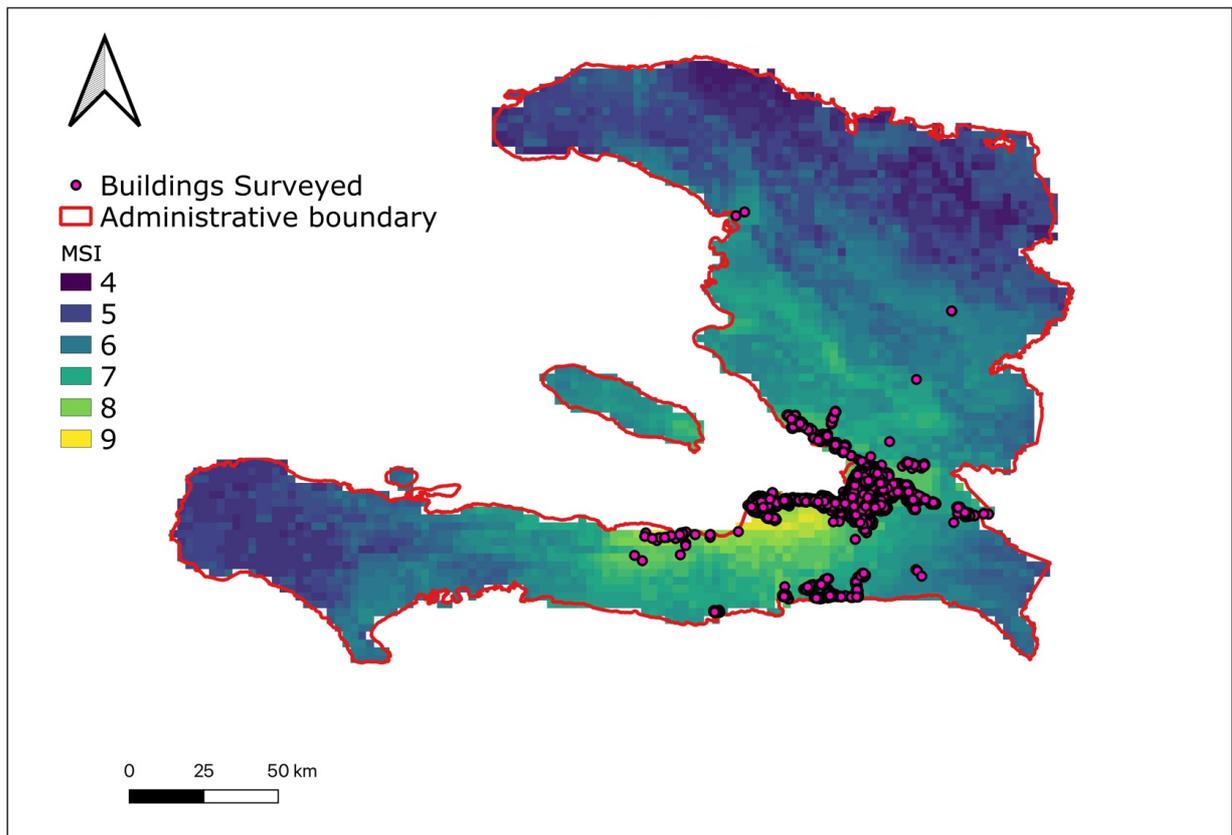
**Structure type-** material used in the vertical structure; four structure types were defined: reinforced concrete structures, load-bearing wall structures, steel sheet-metal structures, and wood sheet-metal structures.

**Floor-type-** materials used for flooring system in a building: three floor types were defined: reinforced concrete floor, concrete floor, and wooden floor.

For each building, the **geographic location** was provided in terms of latitude and longitude.

The 2010 Haiti earthquake building damage portfolio (HBDP) contained 351,819 buildings with complete information on the above-defined features. Fig. 3.1 shows their location, and Fig. 3.2 and Tab 3.1. show the distribution of building features in the HBDP database. The distribution of the samples was imbalanced: for example, there was a small portion of buildings with higher damage grades DG5+DG6 (16.01%) as compared with DG0+DG1 (42.23%), a large majority of buildings were 1-storey (83.05%), reinforced concrete structure (82.11%), and unreinforced block masonry wall (87.81%).

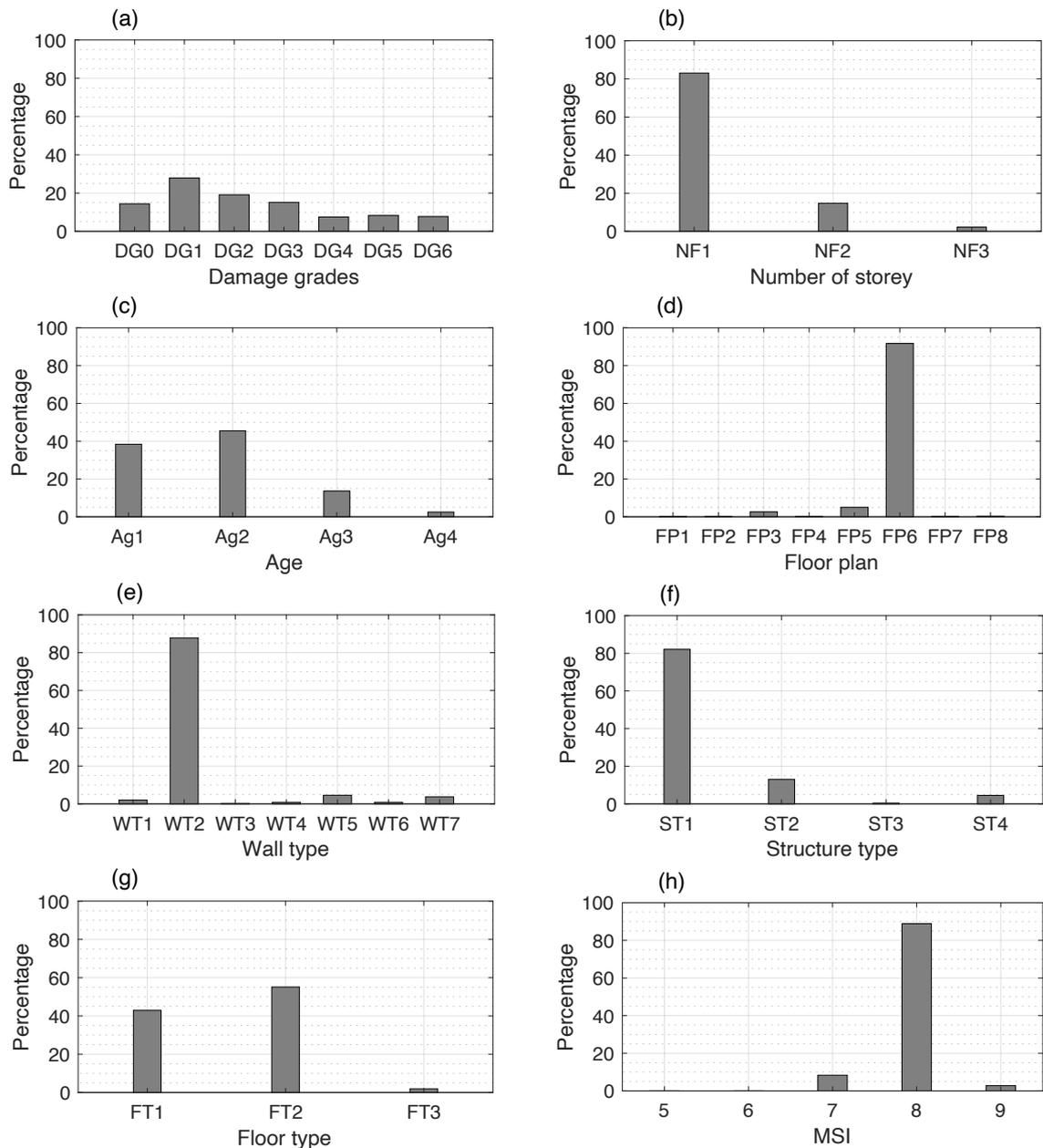
For this study, the HBDP was completed with the spatially distributed ground motion information in terms of macroseismic intensities (MSI) values defined in terms of modified Mercalli intensities provided by the United States Geological Survey (USGS) ShakeMap tool (Wald et al., 2005). The MSI values corresponding to the 2010 mainshock were added to each building information using their geographic location (Fig. 3.1).



**Figure 3.1.** The geographic location of buildings used in this study and the distribution of macroseismic intensity of mainshock obtained from the USGS ShakeMap.

**Table 3.1.** Distribution of building features in the HBDP and their label reported in Fig. 3.2.

No.	Parameters		Data type	Distribution (%)	Remarks
1	Damage grades	none	DG0	14.37	Fig. 3.2a
		slight	DG1	27.86	
		light	DG2	19.10	
		moderate	DG3	15.14	
		heavy	DG4	7.52	
		major	DG5	8.28	
		destroyed	DG6	7.73	
2	Number of storeys	1	NF1	83.05	Fig. 3.2b
		2	NF2	14.75	
		3-10	NF3	2.20	
3	Age (years)	1-10	Ag1	38.38	Fig. 3.2c
		11-25	Ag2	45.49	
		26-50	Ag3	13.64	
		> 50	Ag4	2.49	
4	Floor plan	E-shape	FP1	0.02	Fig. 3.2d
		H-shape	FP2	0.08	
		L-shape	FP3	2.61	
		O-shape	FP4	0.09	
		Other-shape	FP5	4.99	
		Rectangle-shape	FP6	91.73	
		T-shape	FP7	0.12	
		U-shape	FP8	0.36	
5	Wall type	Block-masonry with reinforced	WT1	2.00	Fig 3.2e
		Block-masonry without reinforced	WT2	87.81	
		Brick-masonry	WT3	0.26	
		Reinforced concrete	WT4	0.83	
		Stone-masonry	WT5	4.54	
		Other	WT6	0.83	
		Wood-masonry	WT7	3.73	
6	Structure type	Reinforced concrete	ST1	82.11	Fig. 3.2f
		Load-bearing walls	ST2	13.02	
		Steel sheet-metal	ST3	0.40	
		Wood sheet-metal	ST4	4.47	
7	Floor type	Reinforced concrete floor	FT1	42.92	Fig. 3.2g
		Concrete floor	FT2	55.14	



**Figure 3.2.** Distribution of different features in the Haiti building damaged database. The y-axis is the percentage distribution, and the x-axis is (a) Damage grades, (b) Number of storeys (NF1: 1-storey, NF2: 2-storeys, NF3: 3-10 storeys), (c) Building age (Ag1: 0-10 years; Ag2: 11-25 years; Ag3: 26-50 years; and Ag4: > 50 years), (d) Floor plan (FP1: E-shape; FP2: H-shape; FP3: L-shape; FP4: O-shape; FP5: other-shape; FP6: rectangle-shape; FP7: T-shape; FP8: U-shape), (e) Wall type (WT1: block-masonry with reinforcement; WT2: block-masonry without reinforcement; WT3: brick-masonry; WT4: reinforced concrete; WT5: stone-masonry; WT6: other; WT7: wood-masonry), (f) Structure type (ST1: reinforced concrete; ST2: load-bearing wall; ST3: steel sheet-metal; ST4: wood sheet-metal), (g) Floor type (FT1: reinforced concrete floor; FT2: concrete floor; FT3: wooden floor), and (h) macroseismic intensity.

### 2015-Nepal earthquake building damage database (NBDP)

Following the 7.8 Mw earthquake on 25 April 2015 in Nepal, the government of Nepal conducted a post-earthquake damage survey in the 11 most affected districts and released a database of the observed damage. Tab. 3.2 summarizes the distribution of building features considered for this study and the full description of the database, including the building features, the damage classification and the seismic hazard classification according to USGS macroseismic shake-map is described in Ghimire et al. (2022) (Chapter 1). In this database, information are given by districts (corresponding to an administrative unit).

**Table 3.2.** Distribution of building features in the NBDP.

No.	Parameters	Data type	Distribution (%)	
1	Damage grades	Categorical	DG1	10.34
			DG2 + DG3	29.35
			DG4 + DG5	60.31
2	Number storeys of	Numerical	1-3	98.94
			4-6	1.05
			7-10	0.01
3	Age (years)	Numerical	0-20	62.9
			21-40	27.63
			41-60	6.51
			> 60	2.95
4	MSI	Numerical	5	0.54
			6	11.90
			7	35.90
			8	51.66

### Italian earthquakes building damage database (DaDO)

The Database of Observed Damaged (DaDO) provides information on damaged buildings surveyed after several earthquakes in Italy. The full description of the data used for this study is given in Ghimire et al. (2023) (Chapter 2). Tab. 3.3 summarizes the categories of the features from DaDO and the number of building features used in this study, corresponding to seven Italian post-earthquake surveys.

**Table 3.3.** Distribution of building features in the DaDO.

No.	Parameters	Data type	Distribution (%)
1	Damage grades	DG0+DG1	69.60
		DG2+DG3	22.26
		DG4+DG5	8.15
2	Number storeys of	1-3	85.82
		4-6	13.76
		7-10	0.43
3	Age (years)	0-20	15.22
		21-40	18.80
		41-60	34.16
		> 60	31.82
4	MSI	4	2.23
		5	21.75
		6	31.78
		7	33.57
		8	6.28
		9	4.39

### 3.3 Method

A detailed description on the development of machine learning models is presented in Ghimire et al. (2022, 2023), and a short summary is presented here.

The damage grades were considered as target features. Damage grades were considered as categorical response variables from DG0 to DG6 to test the classification machine learning models. Two advanced classification-based machine learning methods were selected: random forest classification (RFC) (Breiman, 2001) and extreme gradient boosting classification (XGBC) (Chen and Guestrin, 2016). Similarly, the damage grades were ordinal variables, and they were converted as continuous variables from 0 (DG0) to 6 (DG6). Two advanced regression models were selected: random forest regression (RFR) (Breiman, 2001) and extreme gradient boosting regression (XGBR) (Chen and Guestrin, 2016).

Building features and MSI were considered as input features. A one-hot encoding technique was used to convert the categorical features (i.e., wall type, structure type, floor plan, floor type) into binary values (1 or 0), resulting in 30 input variables (Tab. 3.2). No specific data cleaning methods were applied to the HBDP database.

The machine learning algorithms were applied from the Scikit-learn package developed in Python (Pedregosa et al., 2011). The HBDP was randomly divided into training (60% of the whole HBDP) and test (40% of the whole HBDP) dataset with the identical distribution of target features in the training and the test set. The test dataset was kept hidden from the model during the training phase and used for model testing. The efficacy of the machine learning damage prediction model (i.e., its ability to predict damage to a satisfactory or expected degree) was analyzed in three different ways: at building level, at building-class level (defined according to the combination of three input features, i.e. number of storeys, age, and MSI), and at portfolio level.

## **Testing efficacy at building level**

### **Model selection**

At the building level, model efficacy was evaluated by the confusion matrix, representing model prediction compared with the observed (so-called “ground truth”) value. Accuracy was then represented on the confusion matrix by the ratio of the number of correctly predicted DG to the total number of observed values per DG ( $A_{DG}$ ). Total accuracy ( $A_T$ ) was computed in a similar manner as the ratio of the number of correctly predicted DGs to the total number of observed buildings in the whole test dataset.  $A_T$  and  $A_{DG}$  values close to 1 indicate high efficacy. Moreover, the quantitative statistical error was also calculated as the average of the absolute value of errors (MAE): value close to 0 indicate high efficacy.

For classification-based machine learning models, the ordinal value of the DG was used to calculate the MAE score directly. For the regression-based machine learning models, the output DG values were rounded to the nearest integer for the accuracy scores plotted in the confusion matrix but not to calculate the MAE score.

### **Machine learning related issues**

The best model for damage assessment was selected based on the highest efficacy and used to test specific issues related to (a) the imbalance distribution of DGs in the HBDP and (b) the simplification in the damage classification. In this case, DGs are grouped into a traffic-light based damage classification, i.e., green, yellow and red, corresponding to DG0+DG1, DG2+DG3, and DG4+DG5+DG6, respectively.

## Testing efficacy by DG aggregation

### DG aggregation in the whole test dataset

The buildings in the test dataset were aggregated according to the DG values and the error values was computed using Eq. 3.1:

$$\varepsilon_p(\%) = \left| \frac{\Delta n_{DG}}{N} \right| * 100 \quad (3.1)$$

where  $\Delta n_{DG}$  is the difference between the number of buildings observed and predicted at each DG level, and  $N$  is the total number of buildings in the test dataset.

### DG aggregation in the building-class

The buildings in the test dataset were grouped into different building-class defined according to the combination of three input features (number of storeys, age, and MSI). The buildings in each building-class were aggregated according to the DG values and the error values were computed using Eq. 3.2:

$$\varepsilon_{c1}(\%) = \left| \frac{\Delta n_{DG/c}}{N} \right| * 100 \quad (3.2)$$

where  $\Delta n_{DG/c}$  is the difference between the number of buildings observed and predicted at each DG level for a given building-class, and  $N$  is the total number of buildings in the test dataset.

## Comparison between the datasets

The efficacy of the damage prediction model was explored in the HBDP, NBDP, and DaDO datasets. These datasets were randomly divided into training (60%) and test (40%) datasets. The machine learning model was trained on the training set, and its efficacy was explored in the test dataset.

### Efficacy at building-level between the datasets

The efficacy at the building level was compared through the error distribution computed using Eq. 3.3:

$$\varepsilon_d(\%) = \left| \frac{n_e}{N_b} \right| * 100 \quad (3.3)$$

where  $n_e$  is the total number of buildings at a given error level (difference between observed and predicted DG), and  $N_b$  is the total number of buildings in each test datasets.

### **Efficacy at building-class level between the datasets**

The efficacy was compared at building-class level through the error distribution computed using Eq. 3.4:

$$\varepsilon_{c2} (\%) = \left| \frac{\Delta n_{DG/c}}{N_c} \right| * 100 \quad (3.4)$$

where  $\Delta n_{DG/c}$  is the difference between the number of buildings observed and predicted at each DG in each building class,  $N_c$  is the total number of buildings in each building class.

### **Damage probabilities between the datasets**

Finally, the damage probabilities were computed at the building-class level using Eq. 3.5 and compared between HBDP, NBDP, and DaDO datasets.

$$P_{dg} (\%) = \left| \frac{n_{DG/c}}{N_c} \right| * 100 \quad (3.5)$$

where  $n_{DG/c}$  is the difference between the number of buildings observed and predicted at each DG in each building class,  $N_c$  is the total number of buildings in each building class.

## **3.4 Results**

### **3.4.1 Testing efficacy at building level**

#### **3.4.1.1 Model selection**

The efficacy of the regression (RFR, XGBR) and classification (RFC, XGBC) models trained and tested on the randomly selected 60% (training set) and 40% (test set) of the HBDP is summarized in Tab. 3.4.

For the ATC-20 damage classification, the regression-based machine learning models RFR and XGBR yielded similar MAE scores (1.14, 1.18) and accuracy scores ( $A_T = 0.30, 0.27$ ). The classification-based machine learning models XGBC and RFC provide very similar scores ( $[MAE, A_T] = [1.21, 0.38]$  and  $[MAE, A_T] = [1.21, 0.37]$ , respectively). The efficacy of XGBC model is shown in Fig. 3.3a. In the confusion matrix (Fig. 3.3a), the accuracy  $A_{DG}$  values show that the efficacy of the XGBC model is higher for DG1 (0.75) and DG6 (0.41) as compared to DG4 (0.03) and DG5 (0.18). The smaller  $A_{DG}$  values in Fig. 3.3(a) may be due to the imbalance distribution of the target feature as reported for Nepal and Italian datasets in Chapter 1 and 2 (Ghimire et al. 2022, 2023).

In conclusion, the classification models showed slightly better efficacy as compared to the regression models and finally, the XGBC classification model is selected for this study.

**Table 3.4.** Summary of optimized input parameters, accuracy  $A_T$ , and quantitative statistical error values for the regression-based and classification-based machine learning methods for ATC-20 damage classification (seven-class classification). The parameters are the hyperparameters chosen for the machine learning models (the other modal parameters not mentioned here are the default parameters in the Scikit-learn documentation (Pedregosa et al., 2011)). The best accuracy and error values are indicated in bold.

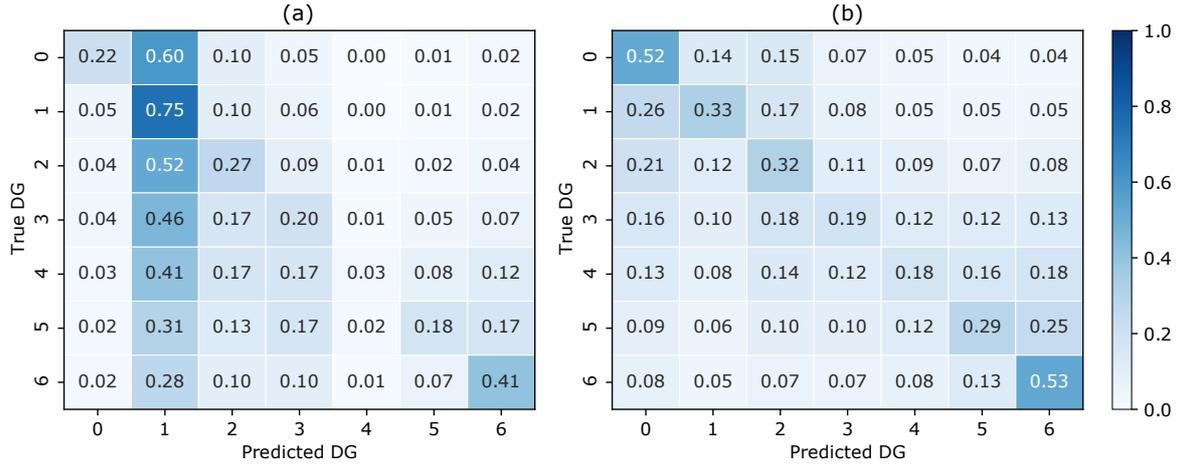
Method	Parameters	Accuracy $A_T$	MAE
RFR	n_estimators = 1000 max_depth = 20	0.30	<b>1.14</b>
XGBR	n_estimators = 1000 max_depth = 10 learning_rate = 0.01	0.27	1.18
RFC	n_estimators = 1000 max_depth = 25	0.37	1.21
XGBC	n_estimators = 1000 max_depth = 10 learning_rate = 0.01	<b>0.38</b>	1.21

### 3.4.1.2 Machine learning related issues

#### 3.4.1.2.1 Handling of class-imbalance issue

The imbalance distribution of target features in the training dataset, known as class-imbalance issues, significantly impact the efficacy of the machine learning models (Estabrooks and Japkowicz 2001; Japkowicz and Stephen 2002; Branco et al. 2017; Ghimire et al. 2022). Resampling the training data using the random oversampling method effectively address the class-imbalance issues (Ghimire et al. 2022, 2023).

In this study, the random oversampling method was then applied in the training set before developing the XGBC model. The damage prediction efficacy on the test set is shown in Fig. 3.3(b). Compared with the result in Fig. 3.3(a), the XGBC model with random oversampling method yielded slightly better efficacy. For DG0/DG2/DG4/DG5/DG6, the  $A_{DG}$  value increased by 30/5/15/11/12%; however, the  $A_{DG}$  value for DG1 decreased by 42%, with a slightly higher MAE score (1.39) and lower  $A_T$  score (0.33). The lower damage prediction efficacy could be originated from the confusion for the machine learning algorithms in multi-class classification depending on the resolution and distribution of input features (Ghimire et al., 2022; Harirchian et al., 2021; Roeslin et al., 2018). This is also observed during damage surveys in the field, which sometimes find it hard to distinguish the intermediate damage grades, such as DG2 and DG3, or DG3 and DG4.



**Figure 3.3.** Normalized confusion matrix between predicted and observed damage grades (DG) (a) without and (b) with handling the DG imbalance issue in HBDP. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . The color scale represents all values.

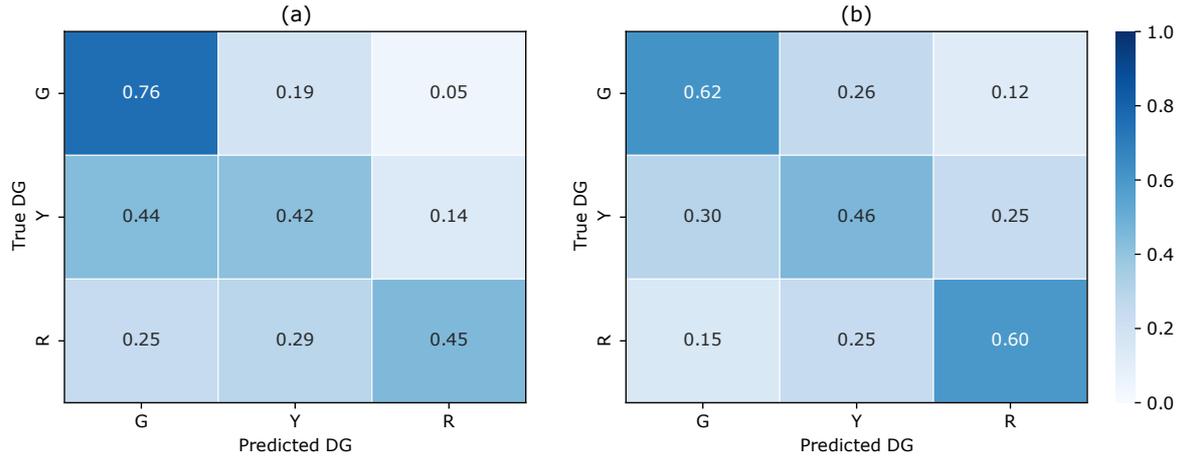
### 3.4.1.2.2 Testing traffic-light based damage classification

The ATC-20 damage classification was reframed into three classes according to the Traffic-Light based classification System (named TLS) (i.e., green G, yellow Y, red R classes, corresponding to  $DG_0+DG_1$ ,  $DG_2+DG_3$ ,  $DG_4+DG_5+DG_6$ , respectively), as followed during post-earthquake emergencies (Mangalathu et al., 2020; Riedel et al., 2015; ATC, 2005; Bazzurro et al., 2004; Ghimire et al., 2022, 2023).

XGBC model was trained on the training dataset with the TLS-based damage classification, and its efficacy on the test dataset is shown in Fig. 3.4 without (Fig. 3.4a) and with (Fig. 3.4b) handling the class-imbalance issues using the random oversampling method. Without addressing the class-imbalance issue, the XGBC model resulted in  $A_{DG}$  values of 0.76/0.42/0.45 for G/Y/R, respectively, with the MAE score of 0.51 and  $A_T$  score of 0.57. After addressing the class imbalance issues, efficacy improved slightly, with  $A_{DG}$  values of 0.62/0.46/0.60 for G/Y/R, respectively, and an MAE score of 0.52 and  $A_T$  score of 0.57.

The  $A_T$  score obtained here (0.57) is slightly lower than those reported by other similar studies: 0.66 by Mangalathu et al. (2020), 0.67 by Roeslin et al. (2020), 0.65 by Harirchian et al. (2021), and 0.68 and 0.72 by Ghimire et al. (2022 and 2023) using Nepal and Italian datasets.

The XGBC model is observed to have higher damage prediction efficacy for TLS-based damage classification, indicating that it is easier for the machine to classify damage into three-class than seven-class (ATC-20 classification system) with the given level resolution of input features. This is consistent with previous studies by (Ghimire et al., 2023 and 2023; Harirchian et al., 2021; Riedel et al., 2015; Roeslin et al., 2020).



**Figure 3.4.** Normalized confusion matrix between predicted and observed damage grades (DG) in HBDP using the traffic-light (TLS)-based classification, grouping the ATC-20 damage classification into three classes (green for no or slight damage; yellow for moderate damage; and red for heavy damage) (a) without and (b) with handling the DGs imbalance issue. The values given in each main diagonal cell are the accuracy scores  $A_{DG}$ . The color scale represents all values.

### 3.4.2 Testing efficacy by DG aggregation

#### 3.4.2.1 DG aggregation in the whole test dataset

Five XGBC models were developed using different combinations of input features and random oversampling method in HBDP dataset:

- (a) XGBC<sub>1</sub> with all input features and without handling class-imbalance issues;
- (b) XGBC<sub>2</sub> with all features and random oversampling;
- (c) XGBC<sub>3</sub> with the number of storeys, age, floor type, wall type, structure type, floor plan, MSI, and random oversampling;
- (d) XGBC<sub>4</sub> with location, number of storeys, age, MSI, and random oversampling;
- (e) XGBC<sub>5</sub> with the number of storeys, age, MSI, and random oversampling.

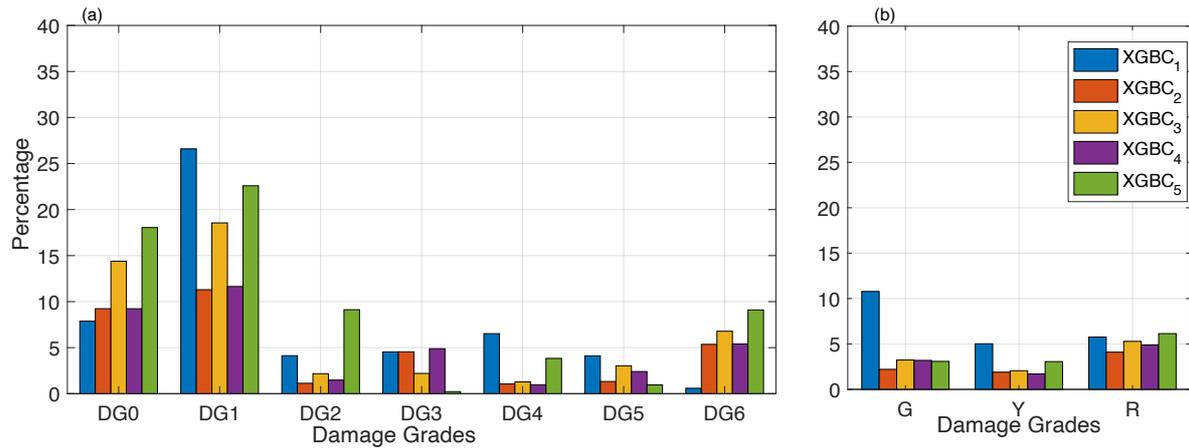
The damage prediction efficacy of these XGBC models was tested by DG aggregation in the HBDP test dataset. The distribution of error values ( $\epsilon_p$  from Eq. 3.1) is shown in Fig. 3.5.

The XGBC<sub>2</sub> and XGBC<sub>4</sub> models show a better damage prediction efficacy for ATC-20 classification (Fig. 3.5a), with error values less than 6% for all DGs except DG0 (9%) and DG1 (11%). Other models give error values of less than 10% for all DGs except DG0 (7-18%) and DG1(18-27%).

For TLS-based damage classification (Fig. 3.5b), XGBC<sub>1</sub> shows higher error values (5-11%) among the DGs. Other models have smaller error values (less than 6%). Among them, XGBC<sub>4</sub> and XGBC<sub>5</sub>

show higher damage prediction efficacy with the smallest error values (2-4% and 3-6%, respectively) for the DGs prediction.

In conclusion, considering DGs aggregation through TLS, easily accessible building features (number of stories, age, location, and MSI) and random oversampling for class-imbalance issues provide reasonable damage estimates for ATC-20 and for TLS-based damage classification, highlighting their broad applicability for larger-scale damage classification. Therefore, the XGBC<sub>5</sub> model for TLS-based damage classification is selected for further investigation in this study.



**Figure 3.5.** Distribution of error values ( $\epsilon_p$  from Eq. 3.1) for (a) ATC-20 and (b) TLS-based damage classification. The x-axis is the damage grades. The y-axis is the error values in percentage. The blue, orange, yellow, purple, and green bar corresponds to the XGBC<sub>1</sub>, XGBC<sub>2</sub>, XGBC<sub>3</sub>, XGBC<sub>4</sub>, and XGBC<sub>5</sub> models, respectively.

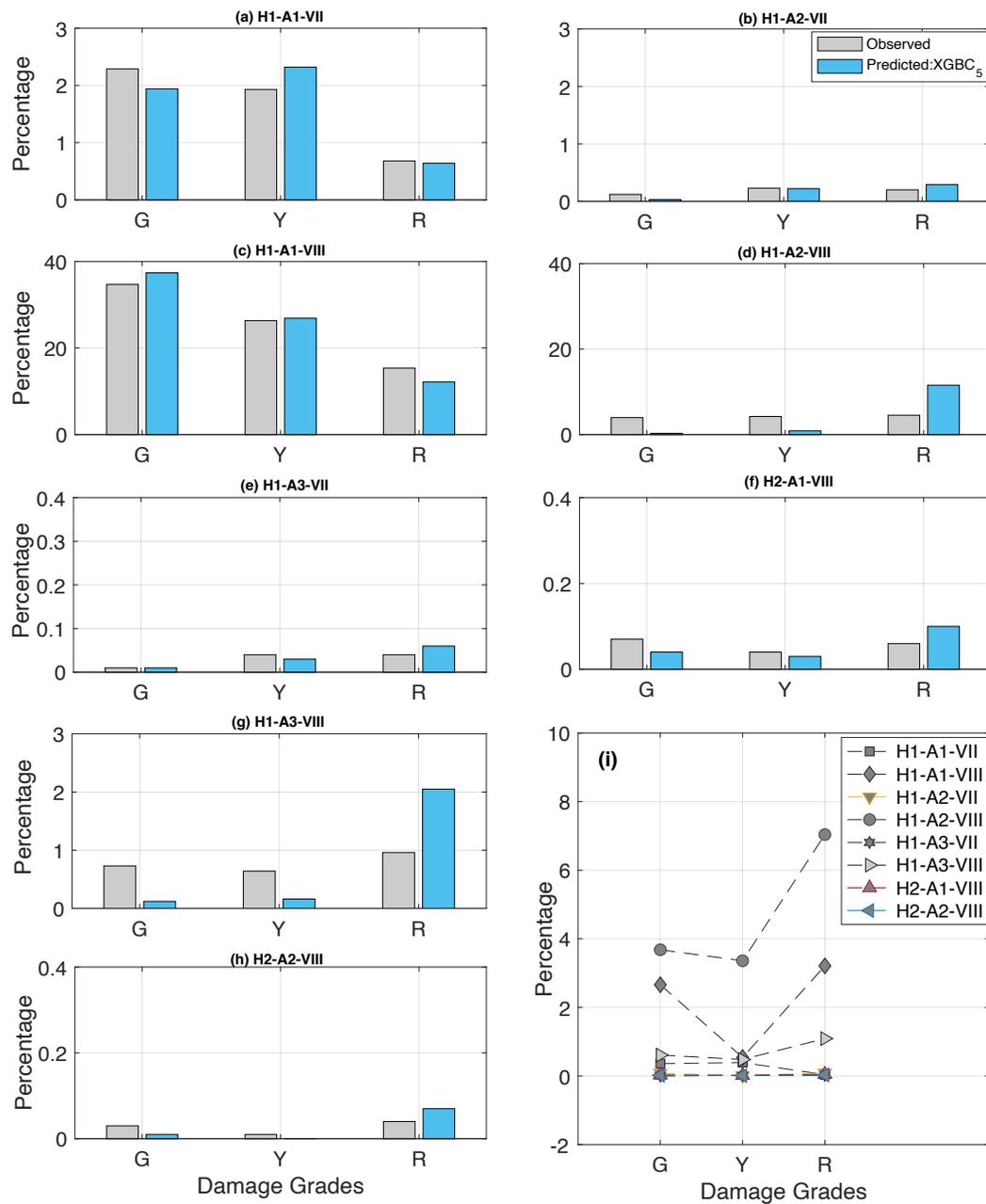
### 3.4.2.2 DG aggregation in the building-class using XGBC<sub>5</sub> model

Building classes were defined based on the combination of number of stories, age, and MSI (H1= 1-3 stories, H2=4-6 stories, H3=7+stories. A1=0-20 years, A2=21-40 years, A3=41-60 years, A4=60+ years. MSI values between 4-9 as IV-XI). Only eight building classes had enough samples for analysis in the test dataset: (a) H1-A1-VII, (b) H1-A1-VIII, (c) H1-A2-VII, (d) H1-A2-VIII, (e)H1-A3-VII, (f) H1-A3-VIII, (g) H2-A1-VIII, and (h) H2-A2-VIII.

The damage prediction efficacy of the XGBC<sub>5</sub> model for TLS-based damage classification was tested by aggregating observed and predicted DG values in each building class. The distribution of buildings in each class and the associated error values ( $\epsilon_{c1}$  from Eq. 3.2) is shown in Fig. 3.6. Most buildings in the test set were observed in H1-A1-VIII and H1-A2-VIII building classes (Fig. 3.6c and 6d), while other classes had smaller number of buildings.

The XGBC<sub>5</sub> model resulted higher error values (Fig. 3.6i) for H1-A2-VIII (3-7%) and H1-A1-VIII (0.5-3.5%) but smaller than 1% for other classes. The XGBC<sub>5</sub> model showed reasonable damage prediction

efficacy in TLS-based damage classification for large-scale classification at the building-class level considering age and number of storeys.



**Figure 3.6.** Distribution of (a) - (h) the observed and predicted DG values and (i) the distribution of error values ( $\epsilon_{c1}$  from Eq. 3.2) in each building class level for the TLS-based damage classification XGBC<sub>5</sub> model. The x-axis is the damage grades. The y-axis for (a) - (h) is building density and (i) error values in percentage. In figure (a) - (g), the grey and blue bar corresponds to the observed and predicted DG.

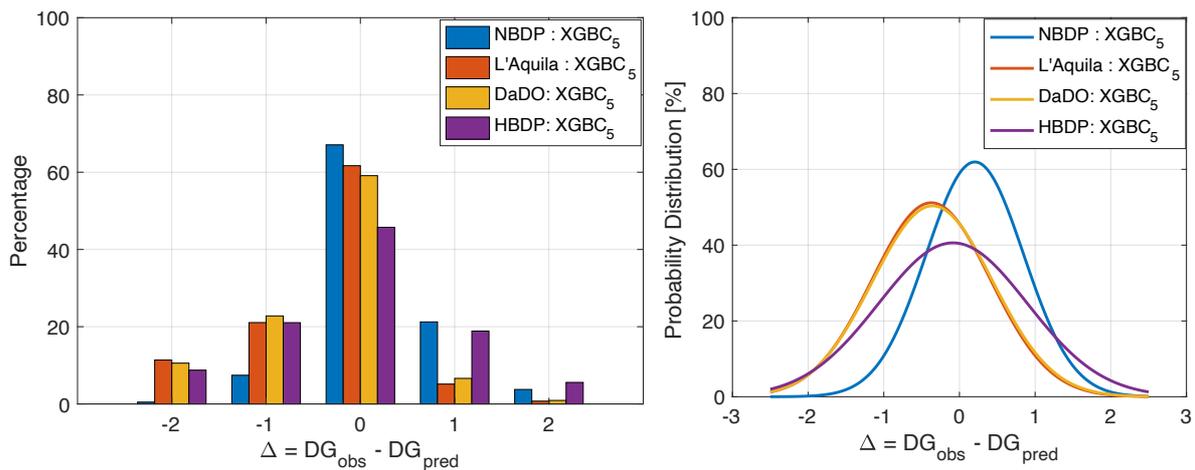
### 3.5 Testing host-to-target machine learning models

The damage prediction efficacy was compared among three datasets: HBDP, NBDP, and DaDO. These datasets were randomly divided into training (60%) and test (40%) datasets. The XGBC<sub>5</sub> model considering TLS-based damage classification was trained on the training set, and its efficacy was explored in the test dataset. The damage prediction efficacy was first compared at the building level and then at the building-class level.

#### 3.5.1 Efficacy at building-level between the datasets

The distribution of error values in the DG ( $\epsilon_d$  from Eq. 3.3) (e.g., 1 corresponds to the delta of 1 damage grade between observed and prediction, regardless of the DG considered) is shown in Fig. 3.7. The percentage of buildings correctly classified (x-value centred on 0 in Fig. 3.7a) in HBDP, DaDO, L'Aquila (a single earthquake of the DaDO dataset) and NBDP is 45%, 59%, 61% and 67%, respectively. The error values were concentrated in +/-1 DG in the range of 7-21% in NBDP, 5-23% in DaDO and L'Aquila, and 18-22% in HBDP. The error values in +/-2 DG was higher for HBDP (5-9%) and DaDO (1-11%) as compared to NBDP (below 4%).

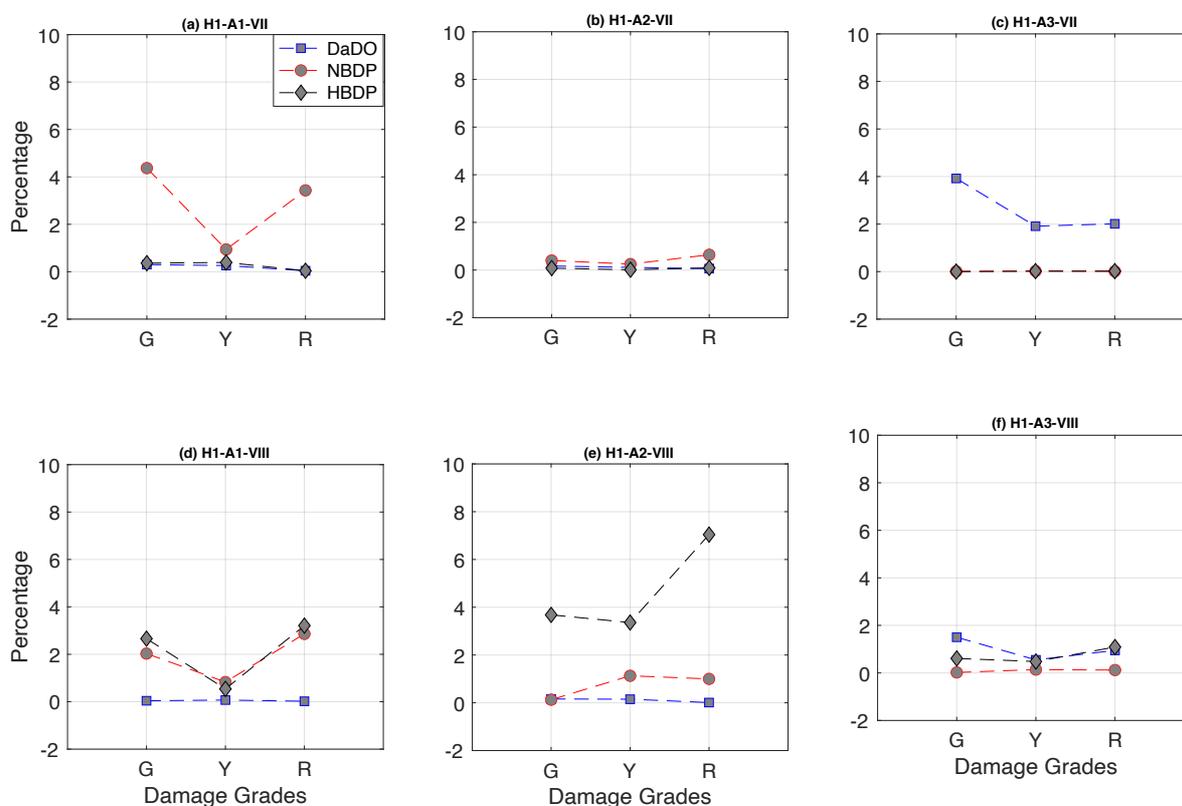
In the probability density function (Fig. 3.7b), skewed error distribution is observed, with a large number of predictions being underestimated in NBDP (positive x-value at peak: 0.21) and overestimated significantly in DaDO (negative x-value at the peak: -0.28) and slightly in HBDP (negative x-value at the peak: -0.08).



**Figure 3.7.** The distribution of (a) error values ( $\epsilon_d$  from Eq. 3.3) and (b) probability distribution function for TLS-based damage classification using the XGBC<sub>5</sub> model. The x-axis is the difference between the observed and predicted DG. The y-axis is the error values in percentage. The blue, orange, yellow, and purple colour corresponds to the NBDP, L'Aquila, DaDO, and the HBDP database, respectively.

### 3.5.2 Efficacy at building-class level between the datasets

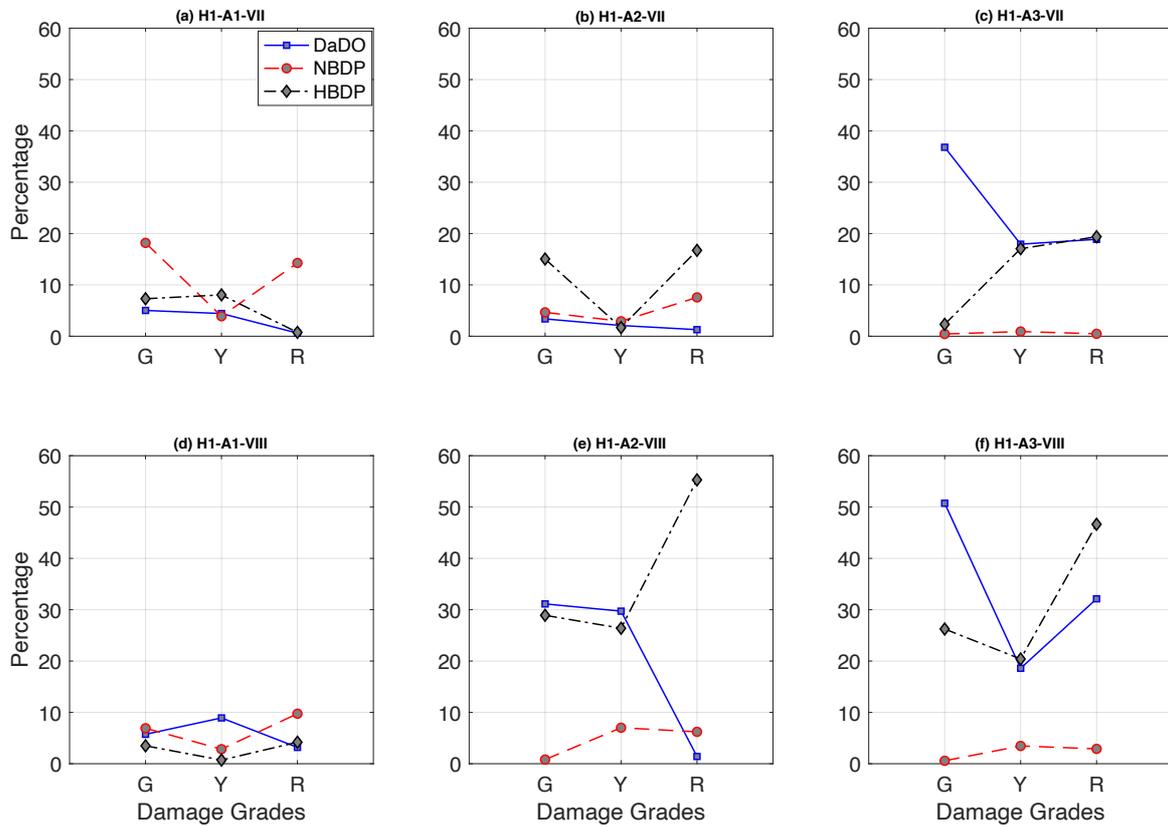
The HBDP, NBDP, and DaDO dataset have six common building classes according to the nomenclature considered for the Haiti dataset. Fig. 3.8 shows the distribution of the error values ( $\epsilon_{c1}$  from Eq. 3.2) by building classes for each dataset. For HBDP, error values are below 1% except for class H1-A2-VIII (3-7%) and H1-A1-VIII (below 4%). For NBDP, error values are below 3% except for class H1-A1-VII (between 1-5%). For DaDO, error values are below 2% except for class H1-A3-VII (between 2-4%)



**Figure 3.8.** Distribution of the error values ( $\epsilon_{c1}$  from Eq. 3.2) for TLS-based damage classification using the XGBC5. The x-axis is the damage grades. The y-axis is the error values in percentage. The blue, red, and black lines with square, circle, and diamond markers correspond to the DaDO, NBDP, and HBDP databases.

Fig. 3.9 shows the distribution of the error values ( $\epsilon_{c2}$  from Eq. 3.4) by building classes for each dataset. For NBDP, error values were below 10%, except for class H1-A1-VII (4-18%). For DaDO, error values are below 9% for three building classes (H1-A1-VII, H1-A2-VII, and H1-A1-VIII), between 19-36% in H1-A3-VII, 1-31% in H1-A2-VIII, and 18-50% in H1-A3-VIII. For HBDP, error values are below 8% for H1-A1-VII and H1-A1-VIII, and between 1-19% for H1-A2-VII and H1-A3-VII. For H1-A2-

VIII and H1-A3-VIII, the error are larger than 20% for H1-A2-VIII and H1-A3-VIII for DaDO and HBDP. Among the datasets, the lowest error values (below 10%) was observed in H1-A1-VIII, followed by H1-A2-VII (below 17%) and H1-A1-VII (below 18%) building classes.



**Figure 3.9.** Distribution of the error values ( $\epsilon_{c2}$  from Eq. 3.4) for TLS-based damage classification using the XGBC<sub>5</sub>. The x-axis is the damage grades. The y-axis is the error values in percentage. The blue, red, and black lines with square, circle, and diamond markers correspond to the DaDO, NBDP, and HBDP databases.

The XGBC model showed smaller efficacy in the HBDP when compared with NBDP and DaDO for TLS-based damage classification. The difference in damage prediction efficacy across datasets may be attributed by the varying degrees of feature variability. For instance, HBDP and DaDO consider interval age values whereas NBDP considers discrete values. In addition, damage information was collected adopting different damage survey form and the damage survey covered the seismic sequence, including aftershocks, while the MSI input only includes the mainshock from USGS ShakeMap.

Interestingly, low-resolution features resulted in reasonable efficacy for large-scale damage classification at the building-class level or higher. The easily accessible building's features (age and the number of stories) and MSI can then provide a reasonable estimate of damage in TLS-based damage classification systems. This method is suitable for rapid damage classification at the regional and national scales, as Riedel et al. (2015) noted.

### 3.5.3 Damage probabilities between the datasets

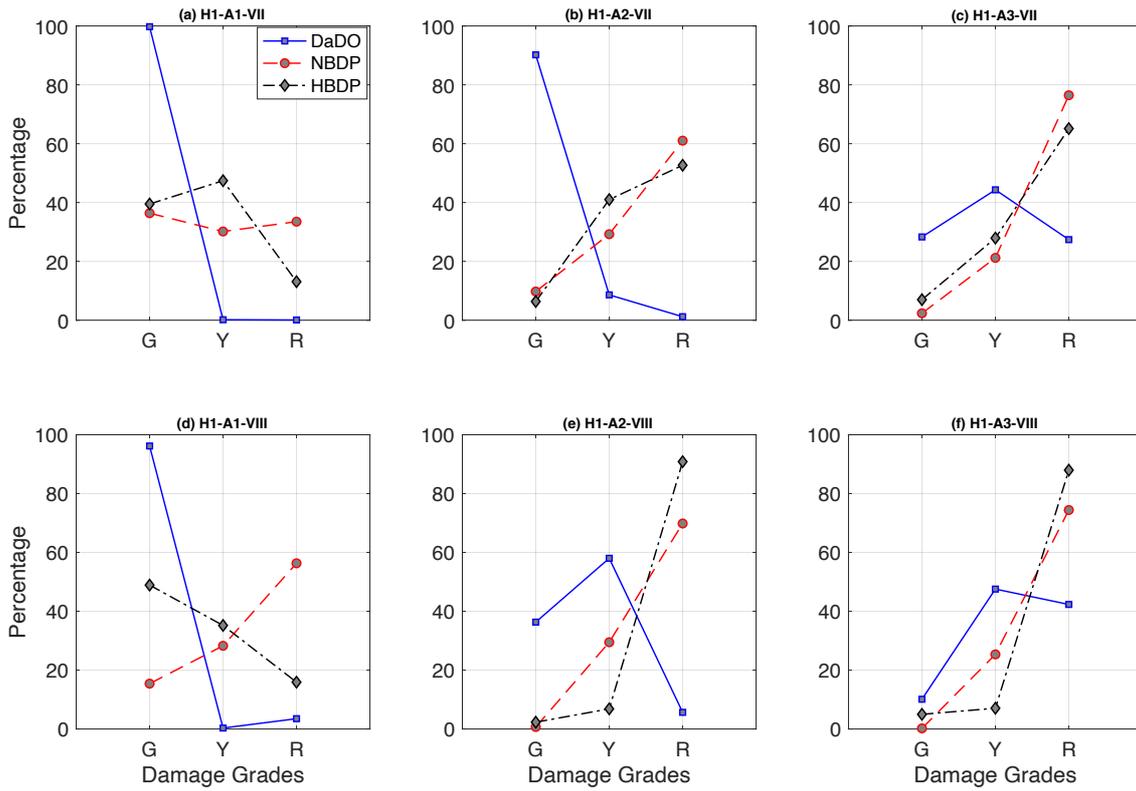
We observed that the smallest error values for the H1-A1-VIII class, followed by H1-A1-VII, and higher error values for H1-A2-VIII and H1-A3-VIII class. Fig. 3.10 shows the damage probabilities ( $P_{dg}$ ) for a building in these classes using Eq. 3.5.

For a building in the H1-A1-VIII class, the probability of being in no-to-slight (G), moderate (Y) and heavy (R) damage grade is 96%, 1% and 3% in Italy, 49%, 35% and 16% in Haiti, and 16%, 28% and 56% in Nepal, respectively. For the building class H1-A2-VII, the probability is 90%, 9% and 2% in Italy, 6%, 41% and 53% in Haiti, and 10%, 29% and 61% in Nepal, respectively. For H1-A1-VII class, the probability is 99%, 0.5% and 0.5% in Italy, 40%, 47% and 13% in Haiti, and 36%, 30% and 34% in Nepal, respectively.

For H1-A3-VII class, the probability of G, Y and R is 28%, 45% and 27% for Italy, 7%, 28% and 65% in Haiti, and 2%, 21% and 77% in Nepal, respectively.

Finally, for the H1-A3-VIII class, the probability of G, Y and R is 10%, 47% and 43% in Italy, 5%, 7% and 88% in Haiti, and 0.5%, 25.5% and 74% in Nepal, respectively.

Thus, a given building class with a given MSI value has different probability to be in each DG, as a consequence of the regional structural typologies and seismic regulations. For example, Nepal and Haiti provide very similar results whatever the class of building, compared to Italian dataset. Therefore, based on these three datasets, raw machine-learning based models for damage prediction are only valid for the hosting region. Thus, empirical adjustments for host-to-target region design differences using regional observation of damage are unlikely to be robust unless the buildings class are similar. In essence, these questions the relevance of generic vulnerability models developed for a host region and which are very tempting to apply to a target region. Testing empirical adjustments of machine learning model for regional characteristics is needed, isolating each host-to-target difference to analyse the contributions of the regional building design differences to the epistemic uncertainty of the model and the damage grades related to macroseismic intensity.



**Figure 3.10.** Distribution of the damage probabilities values ( $P_{dg}$  from Eq. 3.5) using the  $XGBC_5$  model. The x-axis is the damage grades. The y-axis is the probability values in percentage. The blue, red, and black lines with square, circle, and diamond markers correspond to the DaDO, NBDP, and HBDP databases, respectively.

### 3.6 Conclusion

In this study, the damage classification efficacy of machine learning models was explored by using the 2010 Haiti earthquake building damage portfolio (HBDP), 2015-Nepal earthquake building damage portfolio (NBDP) and the observed damage for several earthquakes in Italy (DaDO). Compared to building-level damage classification, the XGBC model showed higher damage prediction efficacy when analysed by aggregating buildings according to DG values in building classes or in the whole test dataset.

When tested with the readily available features (age, number of storeys, and MSI) with TLS-based damage classification, the XGBC model show smaller efficacy in the HBDP when compared with NBDP and DaDO. This could be due to the difference in the resolution of features in these datasets. For instance, HBDP and DaDO has interval age values whereas NBDP has discrete values, and damage information was collected adopting different damage survey form.

Similarly, the XGBC model trained and tested on HBDP, NBDP, and DaDO showed different damage probabilities among the DGs for a given building class. It suggests that when transferring damage prediction models (e.g., Ge et al., 2023) or vulnerability models from active seismic regions to moderate or less active regions, careful consideration on building typologies and their distribution, construction practices and seismic regulations are required.

Higher-resolution features can result in higher efficacy in damage prediction. Interestingly, the low-resolution features (age, the number of stories, MSI) and ground truth damage information can also provide in reasonable efficacy for large-scale damage classification and enable a rapid large-scale damage classification, as also noted in previous studies (Riedel et al. 2014, 2015; Ghimire et al. 2022, 2023). This study opens an opportunity to test building data from the institutional database (e.g., national census, national housing database) for large-scale seismic risk assessment, as already discussed by Riedel et al. (2014, 2015). It can also benefit and encourage the exposure and vulnerability modeling communities (e.g., Jordan, 2019; Schorlemmer et al., 2017). Further investigation should be carried out by collecting building damage portfolios from different regions and building information from different sources (e.g., OpenBuildingMap, European exposure model, INSEE database) to develop this emerging field.

## 4 Analysis of the efficiency of Intensity Measures from real earthquake data recorded in buildings

---

Chapter 1-3 is focused on seismic damage prediction in buildings at regional scale using post-earthquake damage survey data, whereas, this chapter is focused on seismic response of building at building-level using earthquake data recordings from buildings. In this chapter, we investigate the relationship between engineering demand parameters (EDP) and intensity measures (IM). We analyze the efficiency and sufficiency of IMs, sources of uncertainties in the EDP|IM relationship, building frequency variation under earthquake loading, and develop an empirical building damage prediction equation for different building typologies. The study findings have been published in the Soil Dynamics and Earthquake Engineering journal.

*Ghimire, S., Guéguen, P., and Astorga, A. (2021). Analysis of the efficiency of intensity measures from real earthquake data recorded in buildings. Soil Dynamics and Earthquake Engineering, 147, 106751.*

---

### Abstract

In this chapter, a number of spectral and ordinary ground motion intensity measures (IMs) are tested for use in structural performance assessment. Real strong motion values recorded at the top and the bottom of US, Japanese and Romanian buildings are analyzed in order to identify the source of uncertainties in the prediction of engineering demand parameters (i.e., structural drift) for given IMs (i.e.  $\sigma_{EDP|IM}$ ). The efficiency and sufficiency of each IM from a large set of building and earthquake motion data are tested for different criteria characterizing the seismic source (magnitude and source-to-site distance), and considering several building classes and a specific single-building analysis including aging due to cumulative earthquake damage over time. The spectral values at co-seismic resonance frequency was found to be the most efficient IMs for the range of buildings and earthquakes investigated, particularly for velocity with a reduction of approximately 50% of the  $\sigma_{EDP|IM}$  value. Conversely, most IMs are relatively insufficient.

### Keywords

Structural performance assessment, strong motion recordings in buildings, building frequency, efficiency and sufficiency of IM

### 4.1 Introduction

Performance-based earthquake engineering (PBEE) refers to the probabilistic framework in which earthquake consequences are expressed by a set of performance objectives, based on a comprehensive scientific foundation (Porter, 2003). Depending on the application, these performance objectives can

help stakeholders to make decisions with regard to crisis management and structural capacity, such as immediate occupancy or near-collapse levels, by predicting human or economic losses. In the framework proposed by PEER (Porter, 2003), PBEE works in four stages, starting with the hazard itself through to the consequence analysis. In hazard analysis, intensity measures (IM) and their annual frequency of exceedance ( $\lambda_{IM}$ ) are defined by probabilistic seismic hazard assessment (PSHA). In structural analysis, the response of the structure to a given IM can be modeled and expressed in terms of engineering demand parameters (EDPs), such as structural drift, maximal top acceleration, etc. In damage analysis, damage measurement (DM) is calculated based on EDP values and models of structure capacity or fragility. Finally, the earthquake's consequences, in terms of repair costs, operability of the structure and potential economic or human losses for a given DM, can be calculated and expressed as decision variables (DVs) on which stakeholders can base their decisions in view of the expected performance levels.

The four steps of the underlying probabilistic framework of PBEE estimate the frequency of failure of a performance level over a given period of time; this involves uncertainties. For example, the annual frequency of exceeding a given EDP value ( $\lambda_{EDP}$ ) is expressed by:

$$\lambda_{EDP} = \int_{im} P[EDP/IM = im] d\lambda_{im} \quad (4.1)$$

where  $P(EDP|IM=im)$  is the conditional probability of occurrence of each EDP value, taking into account the value of the IM, and  $d\lambda_{im}$  is the annual rate of exceeding an IM value, derived from the hazard curves.  $P(EDP|IM=im)$  is usually obtained by considering a series of nonlinear dynamic analyses of the structure. Baker and Cornell (2008a) provide a detailed description of approaches to characterize and propagate uncertainties at each step. Current research on PBEE is mainly focused on identifying the origins of uncertainties, distinguishing between epistemic and random uncertainties, in order to boost scientific efforts on the reducible elements that contribute most to performance uncertainty (e.g., Iervolino, 2017). In practice,  $P(EDP|IM=im)$  satisfies a chosen model of EDP distribution for a given IM and is obtained by regression of EDP values for IM values. Luco (2002), Luco and Cornell (2007), and Baker and Cornell (2008b) considered an IM to be sufficient if the prediction of EDP given IM is statistically independent of earthquake magnitude and epicentral distance values. Furthermore, the efficiency of IMs is assessed by measuring the variability of values of EDP (given IM) around the regression on the IM values. In general, Peak Ground Acceleration (PGA) or the acceleration spectral value at the resonance period  $T_1$  of the structure  $S_a(T_1)$  (with 5% damping) are considered as outputs of the seismic hazard curves.

Structure response and the associated uncertainties are conditioned by time-history seismic excitation, considering the IM at which the EDP value is exceeded. The efficiency and sufficiency of  $S_a(T_1)$  compared with PGA were investigated by Shome and Cornell (1999) according to the type of building and the contribution of the higher modes to the total response was also considered (Shome and Cornell,

1999). Other ground motion parameters have also been investigated in terms of sufficiency and efficiency, such as peak values in velocity, duration or energy, or spectral values considering different resonant periods (Buratti, 2012; Ebrahimian et al., 2015), spectral values in acceleration (Luco and Cornell, 2007; Shome and Cornell, 1999; Bianchini et al., 2009; Lin et al., 2011; Eads et al., 2015) or velocity (Jayaram et al., 2011; Mollaioli et al., 2011), or by combining IMs based on vector-valued approach (Baker and Cornell, 2008b; Luco et al., 2005; Vamvatsikos and Cornell, 2005). All these studies are based on the numerical modeling of structures considering different ground motion datasets, mostly using the Incremental Dynamic Analysis approach (IDA). In structural analysis, the selection or generation of natural or synthetic accelerograms from different tectonic areas, the scaling applied to obtain the desired structural response values, the selection of physical modal parameters (e.g., structural period and damping) and their co-seismic variations, as well as other modeling assumptions related to component fragility functions, affect the overall uncertainty of the performance estimate. Furthermore, a typical assumption in the assessment of  $P(\text{EDP}|\text{IM})$  is that the building response variability for a class of buildings is the same as the response variability for a given building in this class (this assumption is an ergodic assumption affecting fragility curves).

According to several authors (e.g., Guéguen et al., 2016; Trifunac et al., 2010), considering that data from full-scale observations in real buildings are much more representative than even the most sophisticated laboratory or numerical experiments, one way of improving engineering science to understand the physical behavior of structures is to use a complete database of earthquake recordings in real structures. For example, Perrault and Guéguen (2015) analyzed the variability of EDP versus IM using accelerometric data recorded in Californian buildings, taking structural drift as the EDP, and derived a single-building damage prediction equation (BDPE) with its associated uncertainties. Astorga et al. (2018, 2019, 2020) completed the analysis, confirming the added value of physical data in understanding the seismic response of Japanese buildings in terms of co-seismic demand parameters related to modal (i.e. resonance frequency) parameter variations, especially during repetitive earthquake sequences.

In this study, the efficiency and sufficiency of several IMs for  $P(\text{EDP}|\text{IM})$  from a large number of experimental datasets are analyzed using the regression model of EDP values for IM values. In the 4.2 section, the IMs and EDPs are described based on the study by Astorga et al. (2020). The 4.3 section describes the datasets and the methodology used. Then, the results in terms of efficiency and sufficiency are discussed in the 4.4 section, completed in the 4.5 by a specific analysis of the co-seismic frequency value versus EDP. Finally, the conclusion develops a simple empirical BDPE using available experimental data.

## 4.2 Description of IMs and EDPs

This study uses accelerometric data recorded in several sets of buildings and processed by Astorga et al. (2020) for NDE1.0. Herein, we propose a brief description of the building information and earthquake data; more detailed information is available in NDE1.0 (Astorga et al. 2020).

Six ordinary IM values are considered in this study, computed from the data recorded at the bottom floor of each building:

- Peak ground acceleration (PGA) (Fig. 4.1a), velocity (PGV) (Fig. 4.1b), and displacement (PGD) (Fig. 4.1c), corresponding to the absolute values of maximum acceleration, velocity and displacement time histories, respectively.
- Arias intensity (AI), destructive potential (DP) and cumulative absolute velocity (CAV). Arias intensity (Arias, 1970) includes both the amplitude and duration of seismic shaking, computed as follows:

$$AI = \frac{\pi}{2g} \int_0^{t_f} a^2(t) dt \quad (4.2)$$

where  $g$  is the acceleration due to gravity,  $a(t)$  is the acceleration recorded at time  $t$ , and  $t_f$  is the total duration of the recording. AI is an energy-based parameter that considers amplitude and duration of the ground motion, but it is unable to capture the frequency characteristics of ground motions. To overcome this, Araya and Saragoni (1984) define DP, as follows:

$$DP = \frac{AI}{v_0^2} \quad (4.3)$$

where  $v_0^2$  is the intensity of zero crossings, calculated over the entire duration of ground motion, as defined in its original version. Araya and Saragoni (1984) have shown a strong correlation between DP and observed real damage. Actually,  $v_0^2$  provides a measure of the dominant frequency content of the seismic ground motion.

Finally, Cumulative absolute velocity (EPRI, 1988) is computed as follows:

$$CAV = \int_0^{t_f} |a(t)| dt \quad (4.4)$$

where  $|a(t)|$  is the absolute value of acceleration at time  $t$ .

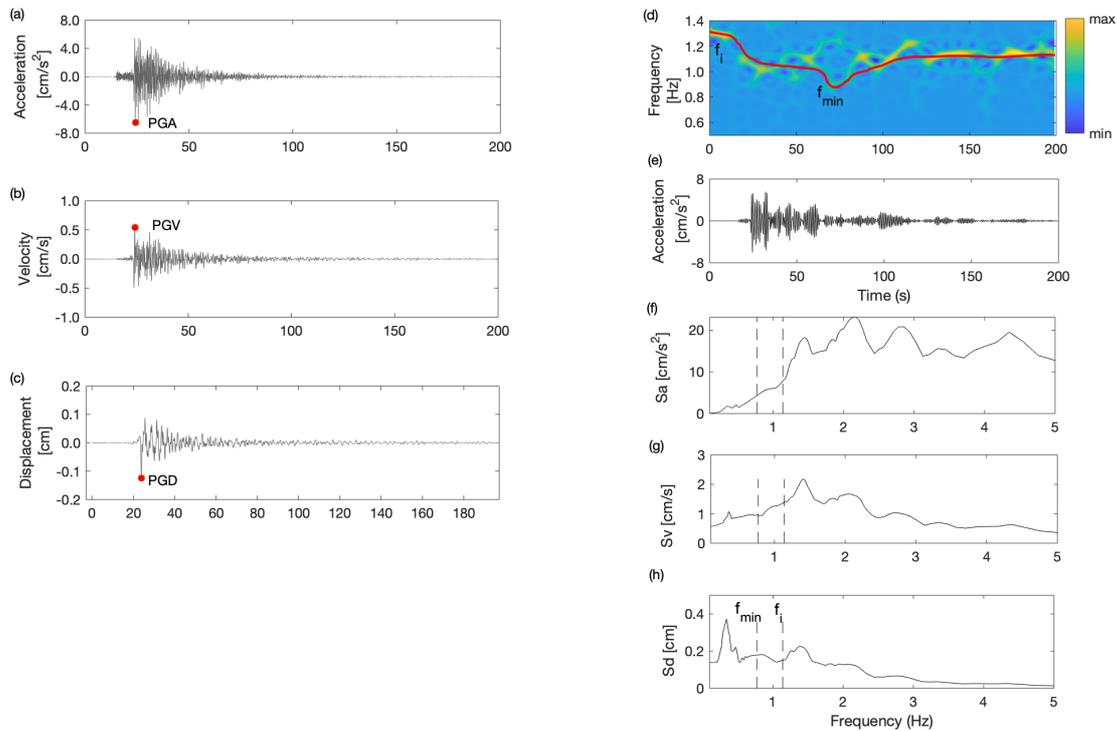
Six spectral IM values are also considered: spectral acceleration (5% damping) ( $SA_i$  and  $SA_{min}$ ) (Fig. 4.1f), velocity ( $SV_i$  and  $SV_{min}$ ) (Fig. 4.1g) and displacement ( $SD_i$  and  $SD_{min}$ ) (Fig. 4.1h) for two specific frequency values ( $i$  and  $min$ ) impacting seismic demand. Index  $i$  corresponds to the spectral value computed at the elastic frequency of the system  $f_i$ , i.e., the frequency obtained by Fourier analysis of the pre-event noise window before the earthquake. Index  $min$  corresponds to the minimal co-seismic value of the resonance frequency  $f_{min}$  during the strongest loading and obtained by applying a time-frequency distribution to the seismic recording (Fig. 4.1d), using the reassigned smoothed Wigner-Ville

distribution (Auger and Flandrin, 1995) .The time-frequency distribution is applied to the top accelerometric time history (Fig. 4.1e). The time and frequency smoothing windows are Hamming windows, with N/10 and N/4 points, respectively. Then, a 3<sup>rd</sup> order Savitzky-Golay filter is applied to the maximum values of energy window. Fig 4.1(e) shows an example of the time-frequency process applied to the data. The  $f_{\min}$  corresponds to the average value of  $\pm 10$  samples around the minimum value observed in the smoothing function (Astorga et al. 2018).

Finally, in this study, drift ratio (DR) is considered as an EDP to describe the building response (Astorga et al. 2020). DR corresponds to the peak of the transient drift observed during the ground motion, obtained by computing the maximum relative displacement, as follows:

$$DR = \frac{\Delta_{\text{top}} - \Delta_{\text{bottom}}}{h} \quad (4.5)$$

where  $\Delta_{\text{top}}$  and  $\Delta_{\text{bottom}}$  are the horizontal displacements recorded at the top and bottom floors of each building, respectively, and  $h$  is the height between the top and bottom floors.



**Figure 4.1.** Illustrations on the left side is the time history of (a) acceleration, (b) velocity, (c) displacement, recorded at the bottom floor sensor of the building. The red dots (a-c) correspond to the peak values. Similarly, the figures on the right are (d) the pseudo-Wigner-Ville time-frequency distribution, the red-line is the smoothing of the maximum of Wigner-Ville, (e) the acceleration recorded at the top floor sensors. Response spectrum of (f) acceleration, (g) velocity, and (h) displacement, the dashed vertical line represents the position of elastic frequency ( $f_i$ ) and co-seismic resonance frequency ( $f_{\min}$ ).

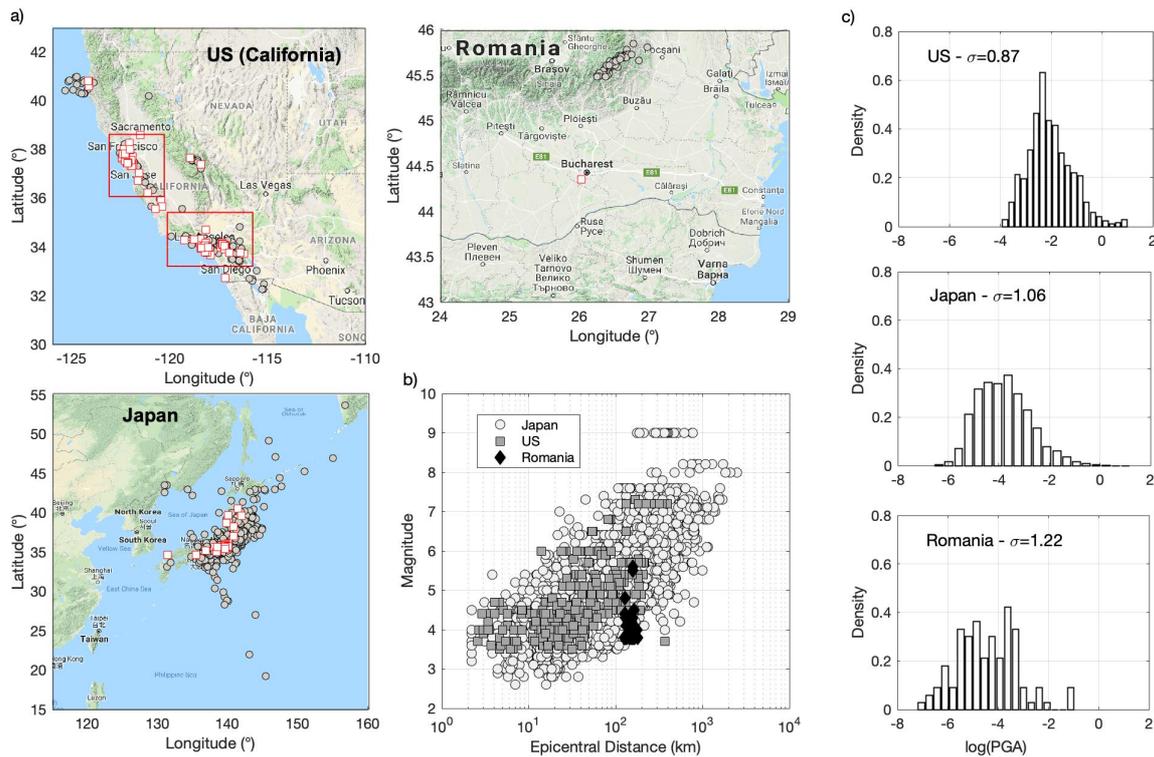
### 4.3 Data and Method

NDE1.0 accelerometric data (Astorga et al., 2020) from the US and Japan plus additional data collected from Romania, processed according the same method as that defined in NDE1.0, were used in this study (Fig. 4.2). Information such as magnitude and epicentral distance is available for each earthquake but there is no description of the source parameters. A brief description of the buildings, earthquakes and datasets is given hereafter.

US data - Data from 84 US buildings provided by Center for Engineering Strong Motion Data (CESMD) (<https://strongmotioncenter.org/>) were considered (Fig. 4.2a). The distribution of the buildings according to construction material is as follows: 27% reinforced concrete (US-RC), 57% steel (US-ST), 11% masonry (US-MA), and 5% wood (US-WO). 684 accelerometric recordings were collected; among them, 225/302/134/24 recordings were collected from concrete/steel/masonry/wooden buildings. Moment magnitude ( $M_w$ ) varies from 3.5 to 7.3 and epicentral distance varies from 2.6 to 331 km (Fig. 4.2b). The dataset includes strong earthquakes, such as the 7.2  $M_w$  Landers event in 1992 and the 7.3  $M_w$  Baja California event in 2010. Two subsets of Californian data are considered to assess uncertainties related to the tectonic context. These subsets are named specific tectonic source STS1 and STS2. The latitude and longitude boundaries of STS1 and STS2 are 33 to 35 and 35 to 39 degrees; 116 to 120 and 120 to 123 degrees, respectively (Fig. 4.2a).

RO data - A ten-story reinforced concrete building monitored by the National Center for Seismic Risk Reduction (NCSRR) of Romania is also considered. This building has been monitored since December 2013. 108 accelerometric records were collected, most of them corresponding to earthquakes located in the Vrancea seismic zone to the north of Bucharest (Fig. 4.2a). Epicentral distance thus varies slightly, between 127 and 178.3 km for  $M_w$  ranging from 3.8 to 5.6 (Fig. 4.2b). The largest earthquakes  $M_w$  5.6 and 5.4 in 2016 and 2014, respectively, are included.

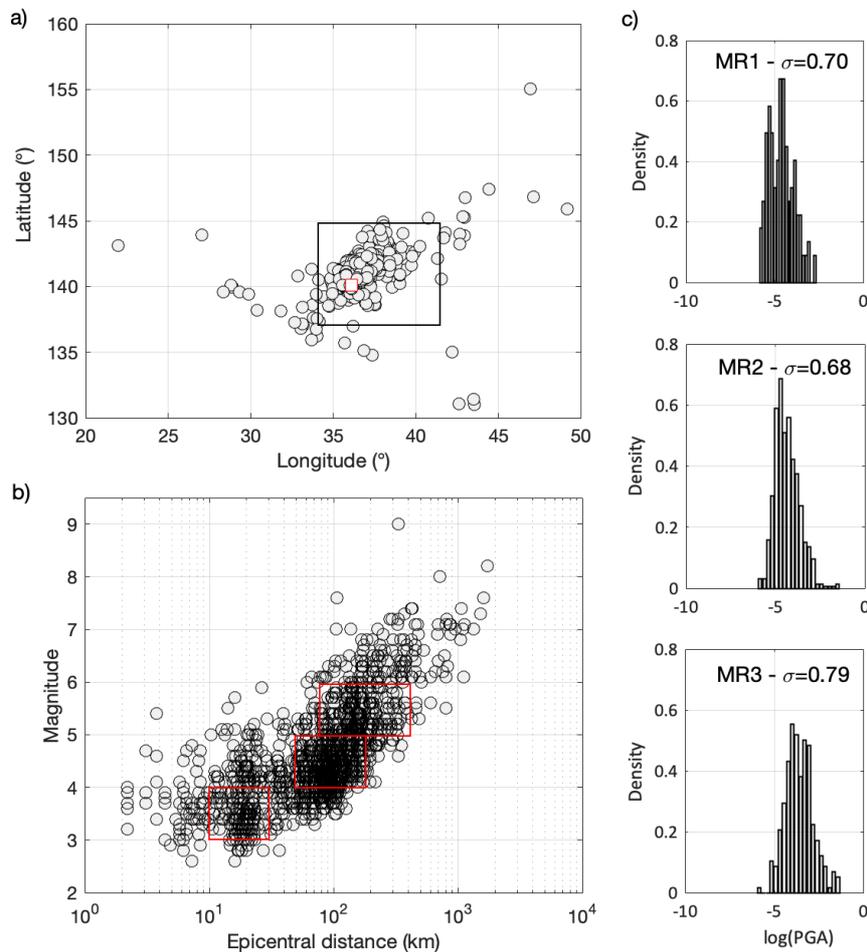
JPN data – 11,763 accelerometric recordings from 32 high/mid-rise Japanese buildings were collected from the BRI strong motion network (<https://smo.kenken.go.jp/>) (Fig. 4.2a). The building distribution according to elementary typology is as follows: 24% steel (JPN-ST), 40% reinforced concrete (JPN-RC), and 36% steel-reinforced concrete (JPN-SRC). The main shock and aftershock sequences of the strongest earthquake, 2011 Tohoku, are included. Magnitude varies from 2.6 to 9.1 (JMA magnitude) and epicentral distance varies from 2.2 to 2,394 km (Fig. 4.2b).



**Figure 4.2.** View of the whole dataset used in this study. (a) Positions of epicenters (gray circles) and buildings (red squares) in the US (California), Romania and Japan. For the US dataset, the two red rectangular boxes define the area of the two specific regions discussed in the manuscript. (b) Magnitude versus epicentral distance distribution of the whole dataset including Japan (open circles), the US (gray squares) and Romania (solid diamonds). (c) Distribution of natural  $\log(\text{PGA})$  for American, Japanese and Romanian datasets, respectively.  $\sigma$  is the standard deviation of the distribution.

ANX – One of the Japanese buildings, ANX, is a building that has been studied extensively by the BRI strong motion network. A detailed description of ANX is available in Astorga et al. (2018). ANX is an 8-story, steel-reinforced concrete building located approximately 60 km northwest of Tokyo, in Tsukuba (Japan) (Fig. 4.3a). ANX has one basement floor resting on spread foundations (8.2 m deep) lying on soft soil made up of alternating layers of clay and sandy-clay to a depth of 40 m. A description of the instrumentation is provided by Kashima (2004, 2014). The ANX dataset is the largest of our datasets, comprising 1,630 recordings in both horizontal directions, made over a period of 20 years, starting immediately after building completion in March 1998 and including the main shock and aftershocks of the 2011 Tohoku earthquake. Magnitude varies from 2.6 to 9.1 and epicentral distance varies from 2.2 to 1,730 km. (Fig. 4.3b). A specific subset of data for ANX is considered, with a geographical boundary of 34 to 41.5 degrees (latitude) and 137 to 145 degrees (longitude). Furthermore, three data subsets are defined based on the distribution of magnitude-distance criteria considered to have an adequate number of data in each dataset (Fig. 4.3b): MR1 corresponding to 166 entries with  $R = 20 \pm 50\%$  and  $M = 3.5 \pm 0.5$ ; MR2 corresponding to 575 entries with  $R = 120 \pm 60\%$  and  $M = 4.5 \pm 0.5$

and MR3 corresponding to 274 entries with  $R = 250 \pm 70\%$  and  $M = 5.5 \pm 0.5$ ). The distribution of PGA for MR1/MR2/MR3 is shown in Fig. 4.3(c). Astorga et al. (2018, 2019) analyzed the time variation of the resonance frequency of the ANX building since 1998. They defined four time periods corresponding to changes in its behavior. During the first period (T1), the fundamental frequency starts to decrease immediately after the completion of construction work, from 1998 to 2005. Frequency stabilizes during period T2 (2006-2011/02/30) until the Tohoku earthquake sequence in 2011. During period T3, the fundamental frequency drops significantly and a slow recovery of the resonance frequency is observed directly after the Tohoku earthquake during the immediate aftershock sequence between 2011/03/01 and 2011/09/30. Finally, T4 corresponds to the period between 2011/10/01 and 2018/05/15. T1/T2/T3/T4 comprise a total of 366/313/402/468 data, respectively. Four further subsets of data within the magnitude distance criteria MR2 are considered according to the period criteria: T1-MR2 (118 data), T2-MR2 (119 data), T3-MR2 (193 data), and T4-MR2 (121 data).



**Figure 4.3.** Dataset for the specific Annex (ANX) building in Japan. (a) Location of the ANX building (red square) and related earthquake epicenters (gray circles). The black square represents the specific subset of data considered. (b) Magnitude versus epicentral distance distribution. The red rectangles define the boundaries of the three magnitude-distance criteria (MR1, MR2 and MR3) described herein. (c) Distribution of  $\log(\text{PGA})$  for specific magnitude-distance criteria MR1, MR2 and MR3.  $\sigma$  is the standard deviation of the distribution.

One method of obtaining  $P[EDP|IM]$  is to perform a series of non-linear dynamic analyses for a given structure and for a given series of earthquakes (Luco, 2002). Another method is to perform regression between EDP and IM, knowing the probability distribution (Shome and Cornell, 1999; Cornell and Luco, 1999) for variability analysis. The degree of scattering around the fitted model represents the uncertainty of the EDP|IM model. To analyze EDP|IM uncertainty and testing efficiency and sufficiency of IMs, one-parameter log-log (log: natural logarithm) linear regression of EDP on IM (Luco, 2002) is used, defined as follows:

$$\log(EDP) = a + b \cdot \log(IM) + \varepsilon \quad (4.6)$$

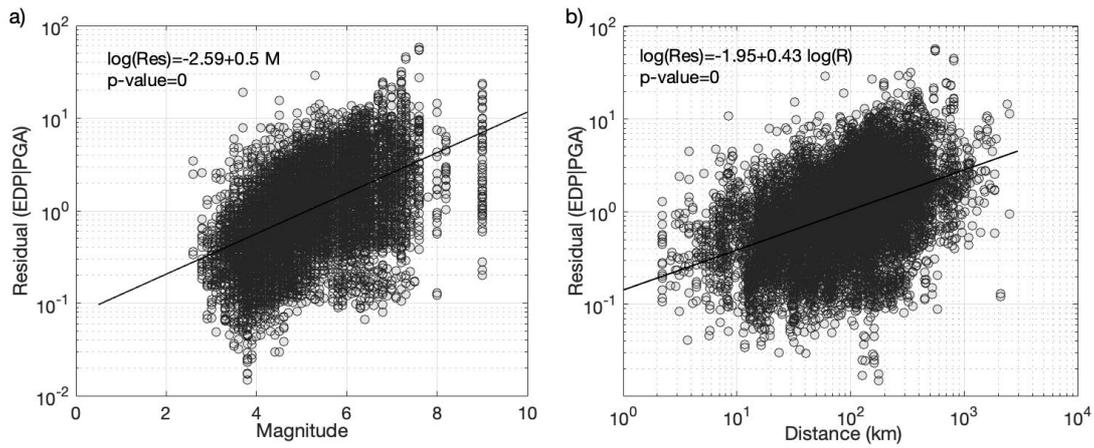
where  $a$  and  $b$  are the estimated regression coefficients and  $\varepsilon$  is the standard error.

The variability associated with IMs and EDP is represented hereafter as  $\sigma_{IM}$ , and  $\sigma_{EDP}$ , respectively, i.e. the standard deviation of the log of IM and EDP values, normalized by their mean value.

The efficiency of IMs is defined simply as the IM that results in a small variability of EDP given IM (Luco and Cornell, 2007). The variability associated with EDP for a given IM is measured by calculating the standard deviation of the residuals of the fitted regression model between EDP and IM (Eq. 4.6), represented hereafter as  $\sigma_{EDP|IM}$ . However, an efficient IM reduces the record-to-record variability between building responses. For practical purposes, this can then reduce the number of non-linear time history analyses for IDA (Shome and Cornell, 1999) with the necessary degree of precision.

The sufficiency of IMs is defined as the IM that makes EDP conditionally independent on earthquake parameters such as magnitude ( $M$ ) and source-to-site distance ( $R$ ). Sufficiency is estimated by computing the linear regression between EDP and IM regression residuals ( $\varepsilon|IM$ ) of Eq. 4.6 and the corresponding value of  $M$  or  $\log(R)$  (Luco and Cornell, 2007).

Fig. 4.4 shows the distribution of residual values EDP|PGA for the whole dataset, versus  $M$  or  $\log(R)$ . As expected,  $\sigma_{EDP|PGA}$  varies significantly, reflecting several sources of uncertainties, which will be explored in the following section.



**Figure 4.4.** Distribution of residual values (EDP|PGA) as function of (a)  $M$  and (b)  $\log(R)$ , considering the whole dataset. The lines represent the fitted linear model between  $\log(R)/M$  and the residuals.

## 4.4 Results on EDP|IM

### 4.4.1 General trends - $\sigma$

For several candidate IMs, global and regional ground motion prediction equation (GMPE) models are continuously evolving (Douglas, 2020). For e.g., Dhakal et al. (2008), Podili and Raghukanth (2019), and Zhao et al. (2016) developed GMPE models for Japan, the value of sigma ( $\sigma_{\text{GMPE}}$ ) reported in these models were 0.86 for PGA, 0.70 for PGV, 0.82 for PGD, 1.46 for AI, 0.69 for CAV, and 0.64-0.88 for acceleration response spectra, and 0.28-0.39 for velocity response spectra with a structural period ranging from 0.01-5 sec. Atkinson (2015) developed GMPE models for the US, the value of sigma reported for these models were 0.37 for PGA, 0.33 for PGV, and 0.31-0.41 for acceleration response spectra with a structural period ranging between 0.03-5 sec.

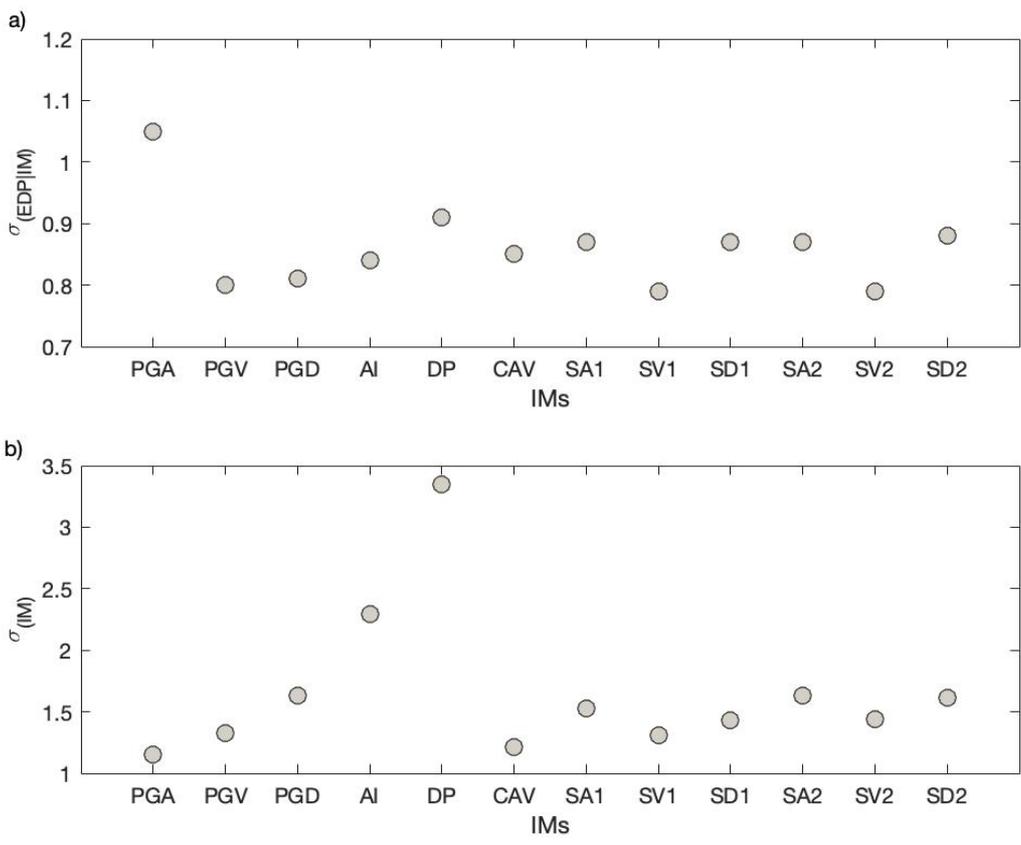
This section compares the efficiency of these candidates IMs in structural response prediction using our experimental data. Fig. 4.5 shows the standard deviation of the residuals of the fitted standard log-linear regression model between EDP and IM (Eq. 4.6) for ALL (JPN+US+RO) datasets. For each IM parameter,  $\sigma_{\text{IM}}$  is given in Fig. 4.5b. All the  $\sigma_{\text{EDP}}$ ,  $\sigma_{\text{IM}}$  and  $\sigma_{\text{EDP|IM}}$  values are provided in Appendix B. In the main body of the manuscript, only the most relevant results are discussed.

The first observation is that the JPN dataset ( $\sigma_{\text{EDP}}=1.44$ ), largest in terms of numbers, is predominant in our global dataset (1.48) compared with the US dataset (1.32). Fig. 4.5a shows that the value of  $\sigma_{\text{EDP|IM}}$  is oscillating between 0.8 and 1.1, with an average value of 0.9. The highest value of  $\sigma_{\text{EDP|IM}}$  corresponds to the spectral and peak ground acceleration values and DP intensity measures. Fig. 4.5b shows that PGA corresponds to the smallest value of  $\sigma_{\text{IM}}$  (1.15), whereas spectral acceleration corresponds to the

highest value of  $\sigma_{IM}$  (in average around 1.60) after DP (3.33) and AI (2.29). Here we can see that the velocity related IMs (PGV,  $SV_i$  and  $SV_{min}$ ) corresponds to the relatively smaller value of  $\sigma_{IM}$ . However, for AI and DP,  $\sigma_{EDP|IM}$  remains similar to the other values. These IMs are also associated with the relatively higher value of  $\sigma_{GMPE}$  in the above mentioned GMPE models. This indicates that these parameters are not good indicators of the natural variability of ground motion and do not enable a high degree of certainty for predicting the response of structures for a given IM.

The velocity related IMs are observed with a smaller value of  $\sigma_{EDP|IM}$  (e.g., 0.80 for PGV, 0.79 for  $SV_i$ , and 0.79 for  $SV_{min}$ ). The efficiency of velocity IMs has already been reported for US data by Perrault and Guéguen (2015) and is confirmed herein, regardless of the dataset considered. It is interesting to note that these IMs are associated with the relatively smallest value of  $\sigma_{GMPE}$  in the above mentioned GMPE models, thus, the velocity related IMs are the most efficient IMs i.e. PGV,  $SV_i$ , and  $SV_{min}$ .

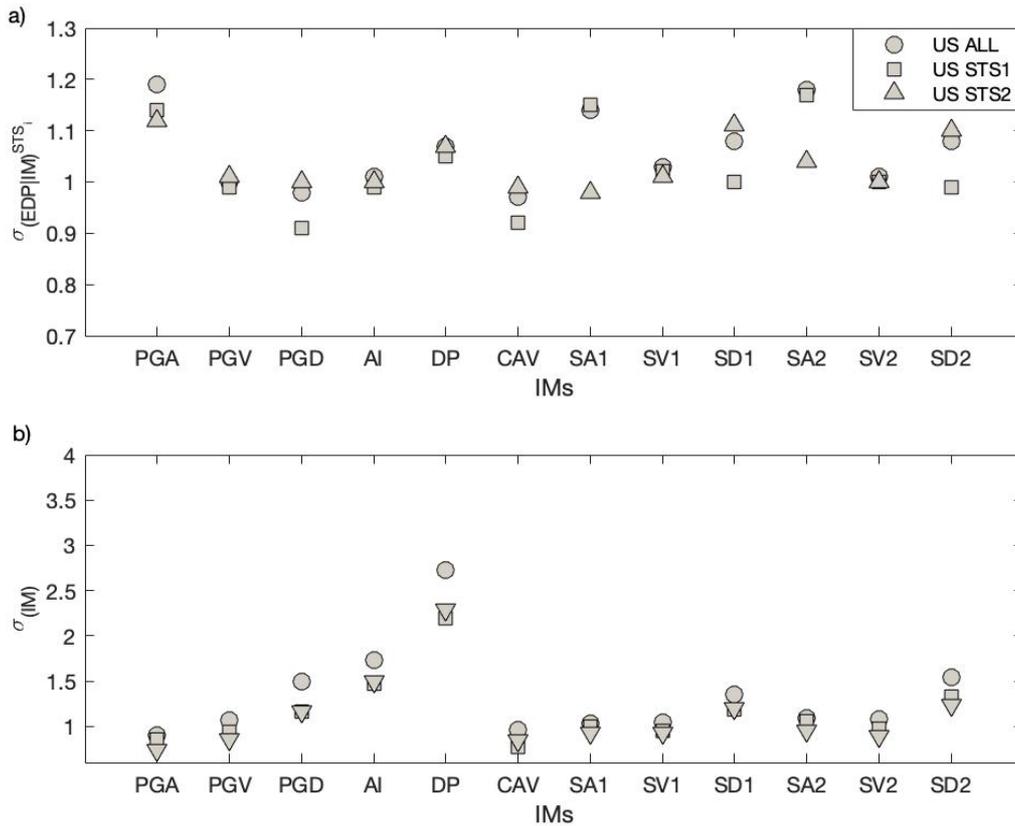
In order to capture the origins of the uncertainties in building response prediction, several relationships are tested in the following sections, according to tectonic region, building typology and ageing effect. The paucity of the data for specific analysis in some datasets makes it necessary to separate the studies; sub-datasets are therefore presented.



**Figure 4.5.** (a)  $\sigma_{EDP|IM}$  values for the IMs concerned, computed for the whole dataset. (b)  $\sigma_{IM}$  values.

#### 4.4.2 Variability associated with the tectonic context - $\sigma_R$

Atkinson and Morisson (2009) demonstrated that seismic ground-motion amplitudes in northern and southern California were significantly different for the same magnitude/distance pair of earthquakes, without identifying the origin of this difference, but related to different tectonic regions. In this study, the effect of the region on building response is examined for all building classes. Fig. 4.6 shows the effect of the tectonic region on the US dataset. The whole US dataset (US ALL) and both STS1 and STS2 subsets are tested, considering all of the previously mentioned US building typologies. Note that for STS1 and STS2, the  $\sigma_{EDP}$  values are the same (1.24 and 1.20 respectively, Appendix B), and  $\sigma_{IM}$  differs only marginally. In Fig. 4.6, the effect of considering the data by specific region barely minimizes the  $\sigma_{EDP|IM}$  values, for the same values of  $\sigma_{IM}$  (Fig. 4.6b). Some exceptions should be noted, the most remarkable being displacement and acceleration. Firstly, the figure shows that for velocity IMs (i.e. PGV,  $SV_i$  and  $SV_{min}$ ), the  $\sigma_{EDP|IM}$  values are similar, being around 1 for ALL, STS1 and STS2 (values in Appendix B). On the other hand, the  $\sigma_{EDP|IM}$  values for STS1 and STS2, respectively, correspond to 1.15 and 0.97 for  $SA_i$ , 1.15 and 1.04 for  $SA_{min}$ , 1.03 and 1.10 for  $SD_i$  and 1.03 and 1.10 for  $SD_{min}$ . Thus, a trend inversion (the smallest values for STS2 or STS1) is observed depending on whether acceleration or displacement IM values are considered. Although the origin of this inversion has not been confirmed, the class of the buildings concerned in these two geographical areas is likely to be cause, since some buildings are more sensitive to acceleration than others, depending on their period of resonance (i.e. stiff or flexible buildings). The following sections will therefore focus mainly on velocity IMs, testing the variability observed in relation to the class of structure in particular, and assuming an insignificant effect of the tectonic context.



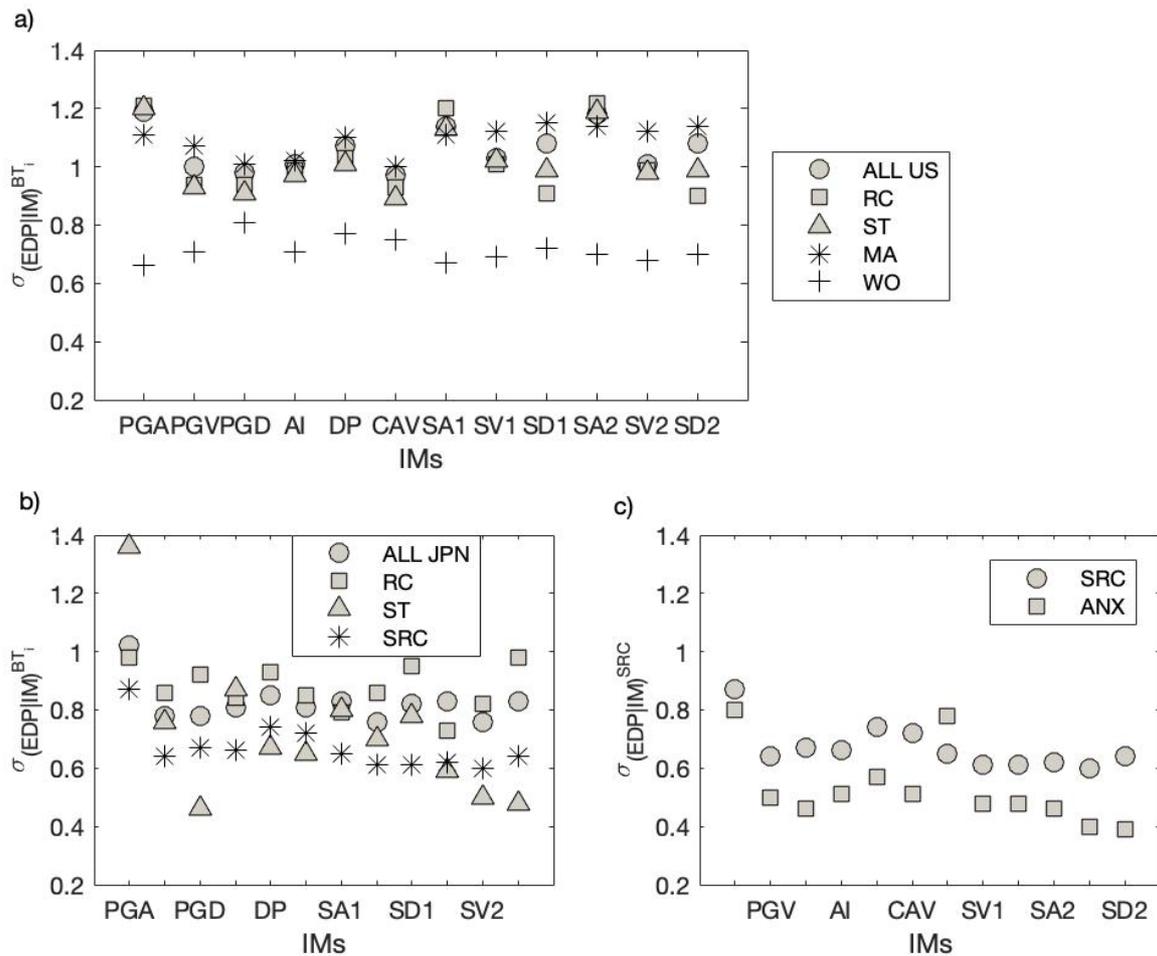
**Figure 4.6.** (a) Values of  $\sigma_{EDP|IM}$  for the IMs concerned computed for US buildings by tectonic region (US STS1 and US STS2) and for all US buildings (US ALL). (b)  $\sigma_{IM}$  values associated with each IM.

#### 4.4.3 Epistemic uncertainties related to building typology - $\sigma_{\tau}$

The seismic performance of buildings depends on their design and characteristics (construction material, height, plan area, regularity etc.). The uncertainties affecting vulnerability assessments are mainly epistemic because, according to Spence et al. (2003), they are due to the classification of buildings into typologies and the attribution of a single generic model to a whole class of buildings. Furthermore, when evaluating seismic capacity, we suppose that many buildings of the same typology have the same ergodic epistemic uncertainties, implying that the values of the epistemic uncertainties do not change between buildings. In this section, the variability associated with different classes of buildings is explored using the US and JPN datasets. Only a basic description of the buildings, based on material, is available in our database. A more detailed classification according to international standards (e.g., HAZUS typology, GEM taxonomy) could be considered in a more comprehensive analysis.

Fig. 4.7(a) shows the variability observed for different classes of buildings in the US dataset. The trends are the same as those observed previously between the velocity IMs and the other IMs. Two typologies stand out: MA and WO. For these two typologies, the small amount of data in our dataset does not

allow a more detailed analysis nor a definitive conclusion as to the effectiveness of certain IMs for EDP prediction. However, for MA buildings, the velocity IMs give higher values of  $\sigma_{EDP|IM}$ , and IMs in acceleration and displacement seem more and less efficient, respectively, than for the other typologies, due to the greater stiffness (i.e. smaller resonance period) generally observed in such buildings. For WO, all the  $\sigma_{EDP|IM}$  values are well below those of the other typologies, but the small number of buildings involved ultimately reduces the epistemic uncertainty related to structural differences within each building class.



**Figure 4.7.** Variability of  $\sigma_{EDP|IM}$  values as a function of the class of buildings (a) US dataset, (b) JPN dataset, (c) ANX single building dataset.

For the US classes (Fig. 4.7a), the type of structure only has a slight influence on the  $\sigma_{EDP|IM}$  values for the velocity IMs (i.e. 0.94 and 0.93 for PGV, 1.01 and 1.02 for  $SV_i$ , and 0.99 and 0.98 for  $SV_{min}$  for RC and ST, respectively, Appendix B). On the other hand, a notable difference exists between US ST and US RC buildings in particular, the latter having a lower  $\sigma_{EDP|IM}$  value for displacement IMs (i.e. 0.94 and 0.91 for PGD, 0.91 and 0.99 for  $SD_i$  for RC and ST buildings, respectively).

There are significant differences between the JPN data (Fig. 4.7b) and the US data. First of all, the velocity IMs give different  $\sigma_{EDP|IM}$  values for different classes of buildings. For JPN ST, the  $\sigma_{EDP|IM}$  values are all lower, reflecting lower epistemic uncertainty related to the diversity of buildings within this class; this contrasts with JPN RC buildings (e.g. 0.86 and 0.76 for PGV, 0.86 and 0.70 for  $SV_i$  and 0.82 and 0.50 for  $SV_{min}$  for JPN RC and JPN ST buildings, respectively). This epistemic uncertainty is confirmed in Fig. 4.7c, where the  $\sigma_{EDP|IM}$  values for one specific single building (ANX building) are compared with those of its building class. There is a significant contribution to the specific single building  $\sigma_{EDP|IM}$  values, with significantly reduced  $\sigma_{EDP|IM}$  values (e.g., 0.64 to 0.50 for PGV, 0.61 to 0.48 for  $SV_i$ , and 0.60 to 0.40 for  $SV_{min}$ ), particularly for parameters other than acceleration. It is also interesting to note an evident contribution of the response spectra calculated by taking into account the co-seismic response values to significantly reduce  $\sigma_{EDP|IM}$  for the SRC buildings (e.g., 0.78/0.48/0.48 for  $SA_i/SV_i/SD_i$  and 0.46/0.40/0.39 for  $SA_{min}/SV_{min}/SD_{min}$  for the ANX building, Fig. 4.7c). Co-seismic resonance frequency, which modifies co-seismic demand, is known to vary for this building Astorga et al. (2018, 2019), as well as for a specific US ST building reported by Guéguen et al., (2016). A similar variation is observed for JPN ST buildings (Fig. 4.7b), significantly reducing the  $\sigma_{EDP|IM}$  values (e.g., 0.80/0.70/0.78 for  $SA_i/SV_i/SD_i$  and 0.59/0.50/0.48 for  $SA_{min}/SV_{min}/SD_{min}$ ). This point will be analyzed more specifically in the last section of this manuscript.

#### 4.4.4 Within-building variability associated with earthquake magnitude-distance - $\sigma_{MR}$

Fig. 4.8a shows the effect of M/R pairs on the variability of the ANX building response. The M/R criteria are described in Fig. 4.3 (section 4.3). Firstly, there is a significant effect on  $\sigma_{EDP|IM}$  values compared with the totality of the ANX data, regardless of the IMs considered, except for the displacement values of the IMs for the MR3 data subset ( $R = 250 \pm 70\%$  and  $M = 5.5 \pm 0.5$ ). These events generated longer periods of ground motion, to which the ANX building, with its resonance period of around 1Hz (Astorga et al., 2018), is more sensitive. For MR1 and MR2, the  $\sigma_{EDP|IM}$  values are lower than the values of the ANX dataset, particularly for the velocity IMs (values for PGV/ $SV_i$ / $SV_{min}$  are 0.50/0.48/0.40 for all ANX data compared with 0.42/0.40/0.34 for MR1 and 0.41/0.41/0.37 for MR2) and for the displacement IMs (values for PGD/ $SD_i$ / $SD_{min}$  are 0.46/0.48/0.39 for all ANX data compared with 0.38/0.37/0.33 for MR1 and 0.35/0.42/0.29 for MR2).

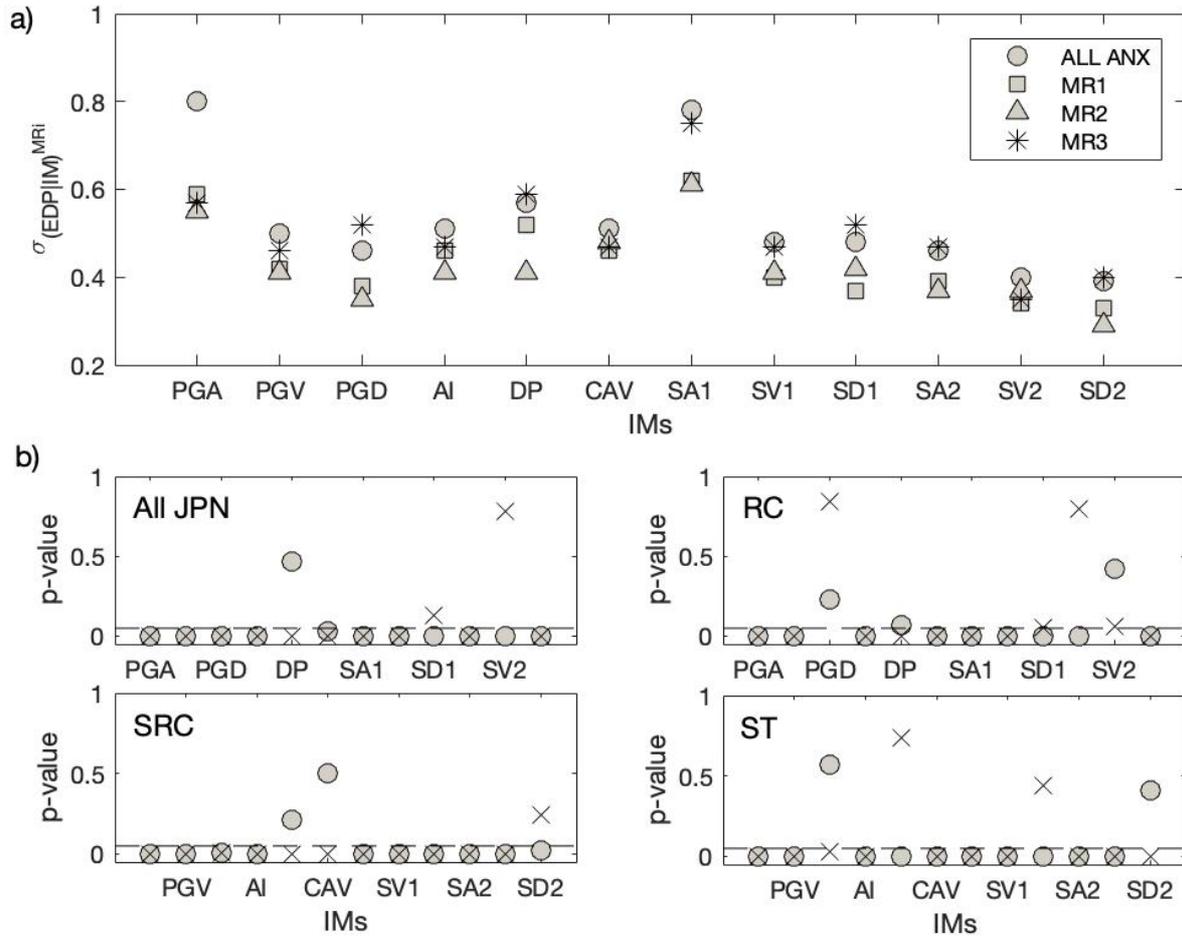
IM sufficiency is tested by considering the JPN building class dataset. The statistical significance of the coefficient obtained from the standard linear regression for M and  $\log(R)$  is assessed based on the p-value (i.e. the probability of obtaining an estimated value of the coefficient at least as large as the actual value, the actual value of the coefficient being zero) (Benjamin and Cornell, 1970). If the p-value observed is greater than or equal to 0.05, the estimated coefficient of M or  $\log(R)$  is statistically insignificant and the IM is considered sufficient (Luco and Cornell, 2007). For the JPN building class, Fig. 4.8b shows the p-value considering all the IM parameters, summarized in Tab. 4.1.

Based on Fig. 4.8b, it appears difficult to conclude on the sufficiency of the IMs tested on our dataset, i.e. EDP is not conditionally independent of magnitude and distance. For the ALL-JPN dataset, the most sufficient IM with respect to magnitude is DP (p-value=0.47) and with respect to distance  $SD_i$  (0.13) and  $SV_{min}$  (0.78), the latter value (the highest) allowing us to assume that the prediction of EDP knowing  $SV_{min}$  is statistically independent of distance. However, for individual building classes, the p-values differ between IMs. It seems that displacement IMs are the most sufficient in distance, such as PGD (p-value=0.57) and  $SD_{min}$  (0.41) for the ST class, as these buildings are the most slender - long-period buildings, i.e. more sensitive to ground displacements. For RC buildings, the most sufficient IMs (in distance) are  $SA_{min}$  (p-value=0.80) and PGD (0.84) and, to a lesser extent,  $SD_i$  (0.05) and  $SV_{min}$  (0.06). For the same class of buildings, the sufficient IMs in magnitude are  $SV_{min}$  (p-value=0.42) and, to a lesser extent, PGD (0.23) and DP (0.07). Finally, for SRC buildings, only two IMs (DP and CAV) are sufficient in magnitude, with p-values of 0.21 and 0.50, respectively.

Since sufficiency differs according to building class, these results suggest that particular attention should be paid when selecting the accelerometric time histories used to perform non-linear time history analysis or PBEE assessment. Furthermore, if insufficient IMs are considered, site-specific ground motion data must be provided to avoid inaccurate estimation of the damage levels or failure rates used in PBEE if the ground motion characteristics do not match the source and site requirements (Kazantzi and Vamvatsikos, 2015).

**Table 4.1** – p-values for the whole JPN dataset, and by JPN building class. Values in bold are greater than 0.05, i.e. the threshold for evaluating IM sufficiency.

		PGA	PGV	PGD	AI	DP	CAV	$SA_i$	$SV_i$	$SD_i$	$SA_{min}$	$SV_{min}$	$SD_{min}$
All	M	0.00	0.00	0.00	0.00	<b>0.47</b>	0.03	0.00	0.00	0.00	0.00	0.00	0.00
JPN	R	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.13</b>	0.00	<b>0.78</b>	0.00
RC	M	0.00	0.00	<b>0.23</b>	0.00	<b>0.07</b>	0.00	0.00	0.00	0.00	0.00	<b>0.42</b>	0.00
	R	0.00	0.00	<b>0.84</b>	0.00	0.00	0.00	0.00	0.00	<b>0.05</b>	<b>0.80</b>	<b>0.06</b>	0.00
SRC	M	0.00	0.00	0.01	0.00	<b>0.21</b>	<b>0.50</b>	0.00	0.00	0.00	0.00	0.00	0.02
	R	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.24</b>
ST	M	0.00	0.00	<b>0.57</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	<b>0.41</b>
	R	0.00	0.00	0.03	0.00	<b>0.74</b>	0.00	0.00	0.00	<b>0.44</b>	0.00	0.00	0.00



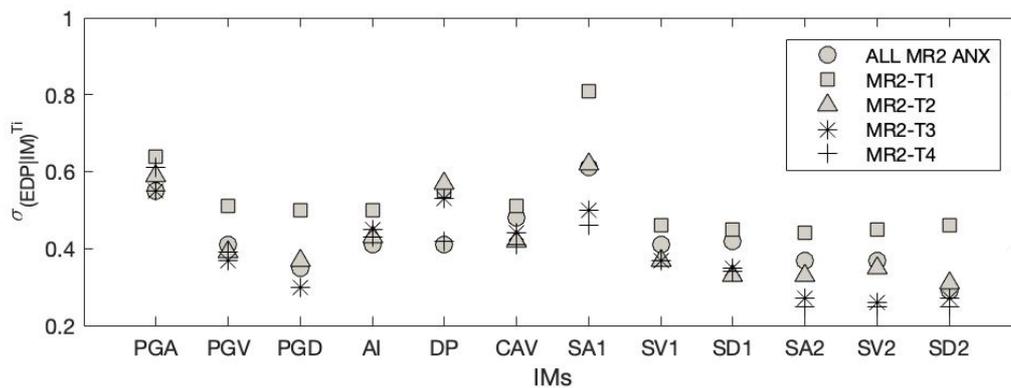
**Figure 4.8.** (a) Variability of  $\sigma_{EDP|IM}$  values as a function of the IMs concerned for different magnitude and earthquake-to-building distance criteria, considering the ANX single building dataset. Magnitude/distance criteria are  $R = 20 \pm 50\%$  and  $M = 3.5 \pm 0.5$  for MR1;  $R = 120 \pm 60\%$  and  $M = 4.5 \pm 0.5$  for MR2;  $R = 250 \pm 70\%$  and  $M = 5.5 \pm 0.5$  for MR3. (b) Sufficiency analysis with respect to magnitude (o) and distance (x) observed for the JPN building class datasets. The dashed line corresponds to a p-value of 0.05.

#### 4.4.5 Within-building variability associated with aging - $\sigma_A$

Karapetrou et al. (2016) discussed the effect of aging over time on the seismic vulnerability of buildings. Moreover, Astorga et al. (2018, 2019) demonstrated the time-dependent response of the ANX building to cumulative events during a long sequence of moderate to strong earthquakes in experimental conditions. Incorporating the real state of a structure may therefore help to reduce variability, yielding more reliable results for PBEE analysis. Fig. 4.9 shows the  $\sigma_{EDP|IM}$  variations as a function of the age of the ANX building. Astorga et al. (2018, 2019) distinguished four specific periods (T1 to T4) during which the frequency of the building changed over time, depending on structural health related to the cumulative damage in the structure. Astorga et al. (2018, 2019) also showed that ANX's response

between seismic events stabilized with its degradation, expressed as a function of the dispersion of structural drift values.

These four periods are considered here, focusing only on the M/R dataset corresponding to MR2 (i.e. the dataset with the most data). A progressive reduction of  $\sigma_{EDP|IM}$  values is observed between T1 and T4, the last two periods being the most efficient (Appendix B), which confirms the results previously reported in Astorga et al. (2018) concerning the stability of the building response with degradation. It therefore appears that taking into account the ageing or actual state of a structure in performance analysis will help to modify the efficiency of the IMs, particularly as even moderate seismic shaking may change the building response (Perrault et al., 2020). For example, for  $PGV/SV_i/SV_{min}$ , the  $\sigma_{EDP|IM}$  values correspond to 0.39/0.37/0.35 for the MR2-T2 ANX dataset and 0.37/0.37/0.26 for the MR2-T3 ANX dataset, compared with 0.41/0.41/0.37 for the MR2 ANX dataset as a whole. This results in a reduction of the performance prediction uncertainties as required during aftershock sequences for the short-time operative assessment of time-dependent building capacity assessment, based on resonance period shift (Trevlopoulos et al., 2020; Trevlopoulos and Guéguen, 2016).



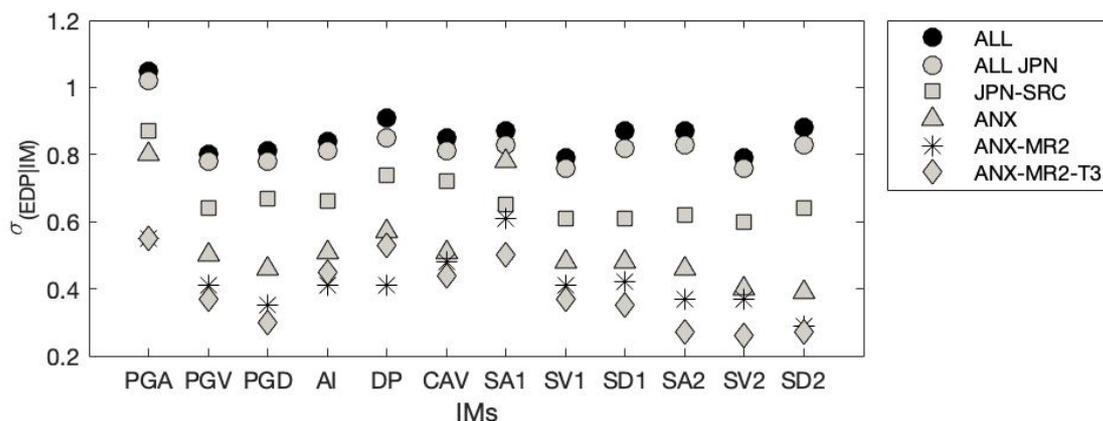
**Figure 4.9.** Variation of  $\sigma_{EDP|IM}$  values as a function of the IMs for different periods, considering seismic cumulative damage in the ANX building. Magnitude/distance criteria (MR2) are  $R = 120 \pm 60\%$  and  $M = 4.5 \pm 0.5$  and periods T1 to T4 are described in the manuscript.

## 4.5 Summary

Fig. 4.10 is a summary of the identification of building response prediction uncertainties for the different IMs considered, based on available data and metadata. In general, for the specific case of the ANX building, the  $\sigma_{EDP|IM}$  values are considerably lower than the model evaluated on the dataset to take into account structural characteristics such as construction type or ageing. For example, for  $PGV/SV_i/SV_{min}$ , the  $\sigma_{EDP|IM}$  values decrease from 0.41/0.41/0.37 to 0.37/0.37/0.26. Table 4.2 summarizes the contribution of each component to the epistemic uncertainties of Eq. 4.6. While the

regional distinction ( $\sigma_R$ ) does not bring any significant gain (4% on average), distinction by type of construction ( $\sigma_T$ ) and specific building ( $\sigma_B$ ) contributes significantly (19% and 21% respectively). Concerning the IMs that make EDP conditionally independent from magnitude  $M$  and source-to-site distance ( $\sigma_{MR}$ ), the figure shows that all the IMs are globally non-sufficient, with a reduction of  $\sigma_{EDP|IM}$  values of approximately 19% for the specific case of the ANX building. When ageing ( $\sigma_A$ ), i.e. the actual health of the structure, is taken into account, the  $\sigma_{EDP|IM}$  values are reduced by 8%.

In total, the spectral IMs benefit most from these successive components. For example, the  $SA_{min}/SV_{min}/SD_{min}$  value reductions are equal to 69%/67%/69% for  $\sigma_{EDP|IM}$  values corresponding to 0.27/0.26/0.27. These results concern one specific building, with a resonance frequency of approximately 1Hz. For longer- or shorter-resonance period buildings, the results would be different, particularly for acceleration or displacement IM values. However, spectral values integrating the co-seismic increase of the resonant period (index 2) allow a reduction of the  $\sigma_{EDP|IM}$  values of approximately 10%.



**Figure 4.10.** Summary of the variation of  $\sigma_{EDP|IM}$  values as a function of the IMs concerned, considering different components of the uncertainties in prediction models.

**Table 4.2** – Summary of the  $\sigma_{EDP|IM}$  values and their reduction (in %) applied to the specific ANX building. The Avg column is the mean value of all IMs.

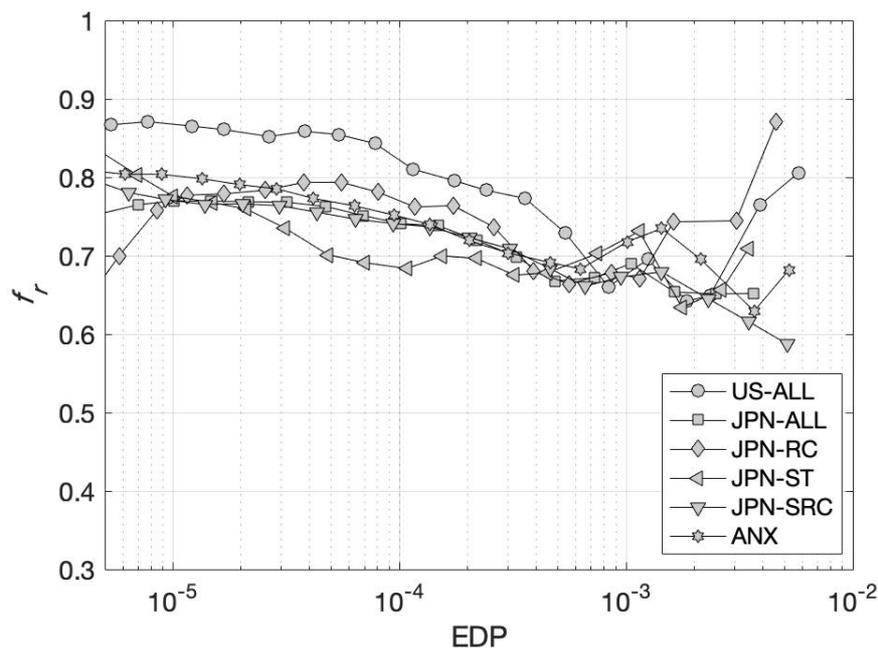
	PGA	PGV	PGD	AI	DP	CAV	SA <sub>i</sub>	SV <sub>i</sub>	SD <sub>i</sub>	SA <sub>min</sub>	SV <sub>min</sub>	SD <sub>min</sub>	Avg.
$\sigma$	1.05	0.80	0.81	0.84	0.91	0.85	0.87	0.79	0.87	0.87	0.79	0.88	<b>0.86</b>
$\sigma_R$	1.02	0.78	0.78	0.81	0.85	0.81	0.83	0.76	0.82	0.83	0.76	0.83	<b>0.82</b>
$\sigma_R / \sigma$	3 %	3 %	4 %	4 %	7 %	5 %	5 %	4 %	6 %	5 %	4 %	6 %	<b>4 %</b>
$\sigma_T$	0.87	0.64	0.67	0.66	0.74	0.72	0.65	0.61	0.61	0.62	0.60	0.64	<b>0.67</b>
$\sigma_T / \sigma_R$	15 %	18 %	14 %	19 %	13 %	11 %	22 %	20 %	26 %	25 %	21 %	23 %	<b>19 %</b>
$\sigma_B$	0.80	0.50	0.46	0.51	0.57	0.51	0.78	0.48	0.48	0.46	0.40	0.39	<b>0.53</b>
$\sigma_B / \sigma_T$	8 %	22 %	31 %	23 %	23 %	29 %	-20 %	21 %	21 %	26 %	33 %	39 %	<b>21 %</b>
$\sigma_{MR}$	0.55	0.41	0.35	0.41	0.41	0.48	0.61	0.41	0.42	0.37	0.37	0.29	<b>0.42</b>
$\sigma_{MR} / \sigma_B$	31 %	18 %	24 %	20 %	28 %	6 %	22 %	15 %	13 %	20 %	8 %	26 %	<b>19 %</b>
$\sigma_A$	0.55	0.37	0.30	0.45	0.53	0.44	0.50	0.37	0.35	0.27	0.26	0.27	<b>0.39</b>
$\sigma_A / \sigma_{MR}$	0 %	10 %	14 %	-10 %	-29 %	8 %	18 %	10 %	17 %	27 %	30 %	7 %	<b>8 %</b>
$\sigma_A / \sigma$	48 %	54 %	63 %	46 %	42 %	48 %	43 %	53 %	60 %	69 %	67 %	69 %	<b>55 %</b>

#### 4.5.1 Building frequency variation and the average response spectral value as an IM

Many previous studies (Astorga et al., 2018, 2019; Calvi et al., 2006; Clinton et al., 2006; Guéguen et al., 2016; Kashima, 2014; Masi and Vona, 2010; Michel and Gueguen, 2010; Mucciarelli et al., 2004) have observed the co-seismic shifting of the resonance frequency for different building typologies. Using a US dataset, Perrault and Guéguen (2015) showed that mean spectral values computed between the pre- and co-seismic periods provided the most effective IM for the EDP|IM model. On the other hand, several authors (Ebrahimian et al., 2015; Bianchini et al., 2009; Eads et al., 2015; Kazantzi and Vamvatsikos, 2015; Bommer et al., 2004; Adam et al., 2017; Kohrangi et al., 2016, 2017) investigated the average spectral values computed between two periods to take into account the co-seismic nonlinear response of building in structural analysis. All these studies were carried using numerical modeling. Boomer et al. (2004) considered the correlation of damage measures and average spectral IMs computed between T1 and F\*T1, T1 being the elastic period of structure, for F values from 1.7 to 3. Bianchini et al. (2009) found the most efficient and sufficient spectral values computed from 0.2\*T1 to 2\*T1 and 0.2\*T1 to 3\*T1. Eads et al. (2015) concluded that average spectral acceleration computed between 0.2\*T1 and 3\*T1 yields lower variability in terms of structural response. Ebrahimian et al. (2015) also considered three average spectral values between 0.2\*T1 and 1.5\*T1, T1 and 2\*T1, and 0.2\*T1 and 2\*T1. Adam et al. (2017) used average spectral acceleration values computed from T1 to 1.6\*T1 and from 0.2\*T1 to 1.6\*T1 and observed a reduction of dispersion in the collapse capacity relationship. Finally, Kohrangi et al. (2017, 2016) observed that the average spectral acceleration value from 0.2\*T1

to  $1.5 \cdot T_1$  yields better prediction of structural response. We found no clear consensus on the values to be considered to reproduce co-seismic demand in structure analysis.

In this study, the frequency variation with respect to EDP was observed for the JPN dataset by building class. Fig. 4.11 summarizes the variation of the frequency ratio  $R_f = f_{\min}/f_i$  between pre-seismic frequency ( $f_i$ ) and co-seismic frequency  $f_{\min}$  with respect to EDP for different JPN building classes and the whole US database, with EDP ranging from  $5 \cdot 10^{-6}$  to  $10^{-2}$ . The variation of  $R_f$  confirms that, regardless of building class, the frequency shift between the pre- and the co-seismic period increases with EDP, which means large frequency drops occur for the strongest earthquakes. Significant variation of  $R_f$  is observed even at the lower end of the EDP range from 0.9 to 0.65 (below the slight damage threshold=0.0025). For EDP values between  $10^{-5}$  and  $10^{-3}$ , a relatively similar trend is observed regardless of building class, with  $R_f$  values decreasing from 0.78 to 0.65. There are no stronger earthquakes, which prevent us from extending this result to a higher level of drift, but this first experimental evaluation suggests the need to collect a large amount of earthquake data in buildings in order to refine our performance prediction models.



**Figure 4.11** – Variation of the frequency ratio ( $R_f=f_{\min}/f_i$ ) for different datasets.

## 4.6 Conclusions

Experimental data is very useful in helping us to understand the complex physical processes at work in civil engineering structures to be able to integrate them into our models to reduce the epistemic

uncertainty of these complex process. Earthquake data collected from buildings under long-term monitoring in Japan, the US and Romania were used to attempt to identify the components of the uncertainties associated with EDP|IM. Region-to-region, building-to-building and within-building uncertainties associated with earthquake magnitude-distance and ageing were explored.

Compared with the conventional IMs based on peak values or conventional spectral value ( $SA_i$ ,  $SV_i$  or  $SD_i$ ), the ground motion intensity measure, denoted  $SA_{min}$ ,  $SV_{min}$  and  $SD_{min}$ , which considers inelastic period lengthening, was found to be the most efficient IM for estimating EDP, taken as structural drift herein. In terms of sufficiency, generally speaking, it appears that no IMs are sufficient due to a significant conditional dependence of EDP on  $R$  (i.e. earthquake source-to-building distance) and  $M$  (i.e. magnitude). Some exceptions are pointed out in Fig. 4.8 for specific building classes and IMs. In fact, depending on the type of building and, in particular, its period, displacement and acceleration IMs might be more efficient or sufficient; this could be confirmed with additional data and a specific analysis of the building characteristics, which is not considered by this study. Nevertheless, all our results indicate that velocity IMs are those that provide the lowest variability for predicting EDP given IM.

Based on the ANX building results, the components that make the largest contribution to overall uncertainties are building class and specific building associated with the  $M/R$  condition (Fig. 4.10, Tab. 4.2). When analyzing specific buildings using long-term monitoring data, the real structural state also appears to make a significant contribution to the uncertainties, reflecting the real co-seismic demand in EDP prediction. The underlying key issue is related to the variation of frequency, which is strongly dependent on EDP. Note that regardless of building class, this frequency variation follows the same trend for all the drift values in our dataset.

Several numerical studies have been carried out to investigate the efficiency and sufficiency of IMs considering one building or a group of buildings and using suits of input ground motions (Buratti, 2012; Ebrahimian et al., 2015; Bianchini et al., 2009; Lin et al., 2011; Eads et al., 2015; Jayaram et al., 2011; Mollaioli et al., 2011; Luco et al., 2005; Vamvatsikos and Cornell, 2005). In contrast, this study compares the efficiency and sufficiency of IMs and the variation of sigma in the EDP|IM relationship using real experimental data. We are able to capture the real structural response from different building typologies during several earthquakes, which is the main advantage of this study. This study shows that the most commonly used IMs i.e., PGA and  $SA_i$  are not efficient and sufficient to predict the structural response within the range of our dataset. The velocity related IMs i.e., PGV,  $SV_i$ , and  $SV_{min}$  are observed to be the most efficient in building response prediction. Our findings strongly agree with previous studies carried out on real data by Perrault and Guéguen (2015). It is very interesting note here that the IMs which are found efficient in our studies are also associated with relatively smaller variability in the available GMPE models.

This study highlights the importance of real experimental dataset. Having more information on the earthquakes and descriptions of the building characteristics would help to improve the prediction of structural response for analyzing seismic vulnerability or loss assessment. Although the amount of data contained in our dataset provides relevant results, the paucity of data concerning specific classes of buildings or components of uncertainties limits the strength of the conclusions that can be drawn. To resolve outstanding issues, we must continue our international collaborative efforts and motivate building owners to share their data, which would increase their interest in this type of study. In particular, having more specific data would enable verification of the aforementioned conclusions.

Moreover, building response prediction models can be developed considering several parameters related to earthquakes and buildings, such as ground motion IM, magnitude, distance, building typology, height, structural properties, etc. (Perrault and Guéguen, 2015; FEMA, 1999; Hancock et al., 2008). In conclusion to this study, an empirical building damage prediction model is proposed (Tab. 4.3) based on the entire dataset (US, Japan and Romania) according to building class and considering the most efficient IMs ( $SV_{\min}$  and PGV) using the functional form given in Eq. 4.6:

$$\log(\text{EDP}) = a + b \cdot \log(\text{IM}) + \varepsilon$$

**Table 4.3-** Empirical building damage prediction model according to building class.

IM	Parameter	BT-ALL	BT-RC	BT-SRC	BT-ST	BT-MA	BT-WO
$SV_{\min}$	a	-10.22	-10.65	-10.00	-9.81	-9.17	-9.26
	b	0.87	0.83	0.89	1.08	0.55	0.58
	$\sigma$	0.79	0.84	0.60	0.60	1.10	0.68
PGV	a	-9.41	-9.78	-9.21	-9.17	-8.80	-9.02
	b	0.94	0.86	0.95	1.16	0.60	0.55
	$\sigma$	0.80	0.87	0.64	0.80	1.07	0.71

## 5 The variability of experimental capacity curve models for existing buildings and its within-building and between-building components.

---

In this study, we developed an experimental co-seismic capacity curve based on strong motion data recordings from buildings. Using the experimental capacity curves, we analyzed the evolution of building co-seismic capacity during a seismic event, including the variation before and after initial damage, accumulation of damage over time, and variation in a building class, and proposed adjustments to account for the actual structural health required for the OELF method.

---

### Abstract

In this study, accelerometric data from seven Japanese buildings under long-term monitoring were analysed to explore the variability of the buildings' co-seismic response over time and its within- and between-building components, using co-seismic capacity curves developed in acceleration-displacement-response-spectrum format. The data include the 2011 Tohoku Mw9.1 earthquake, which caused building damage of different levels of severity, and the time-varying actual capacity curves were analysed considering earthquakes before and after 2011. We observed that the initial slope of the capacity curves reflects the amount of damage. The between-building and within-building components of the variability are discussed by comparing a single building and several buildings in the same class for several earthquakes. Finally, the epistemic uncertainty of seismic risk assessment studies is discussed in relation to the selection of a generic capacity model for all buildings in a single class.

**Keywords:** experimental capacity curves, earthquake data recordings from buildings, building's co-seismic capacity variation after damage, operational earthquake loss forecasting, structural health monitoring.

### 5.1 Introduction

Earthquake engineering is currently focused on improving the seismic resistance of structures through earthquake capacity design. Capacity design refers to the design of a building to ensure controlled ductile behavior to avoid collapse in a design-level earthquake specified by stakeholders. The main objective is to minimize the direct and indirect losses generally correlated with any level of structural damage that affect the assets and people exposed and, indirectly, to limit the downtime of vital functions and essential facilities in urban areas following an earthquake. Herein, damage refers to any undesirable change in

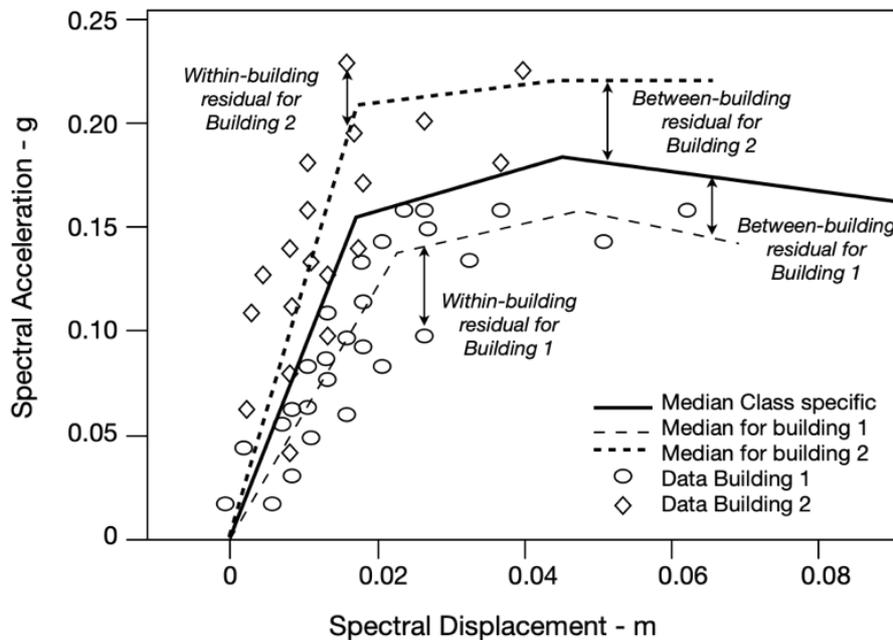
structural properties that significantly affects the intended functions of a structure over its expected lifespan (Farrar & Worden, 2007). Anticipating seismic consequences requires prior information on the damage that structures might be expected to suffer in a given earthquake, i.e., capacity models for the structures exposed.

In the recent European Seismic Risk Model (Crowley et al., 2021), three main approaches are considered to assess seismic risk: scenario-based, intensity-based and frequency-based. The main difference lies in the definition of seismic hazard, but once the latter has been defined, risk is calculated as the convolution of the hazard with the vulnerability or fragility functions. This results in a loss estimate for a given class of building, which is then aggregated to the geographical unit of the area concerned. The fragility functions are developed based on the capacity model of the building or building class, which describes the lateral strength and deformation capacity in acceleration-displacement-response-spectrum format (ADRS). Capacity curves are then developed, considering various structural attributes (e.g., materials, lateral load resistance system, number of storeys, etc.) to cover a wide range of building classes in the exposed area. In urban-specific seismic risk studies, it is common practice to start by assessing the exposure model at the site concerned and then to attribute class-capacity curves to each building (Crowley et al., 2021). This approach offers the potential advantage of taking into account a large set of buildings, but specific attributes within a class can cause significant deviations from the generic model. For example, Martins and Silva (2021) released a databank of capacity curves that has been used to represent the vulnerability classes of current European buildings (Crowley et al., 2021). For each class, the backbone capacity curves were compiled from simulated design-based research studies, i.e., push-over analysis (Fajfar, 2000) or incremental dynamic analysis (Vamvatsikos and Cornell, 2002). The method returned an average capacity curve for a generic building model without considering the effects of any specific building features. The application of generic backbone models to large sets of buildings in a single class introduces epistemic uncertainties into damage assessment (Ghimire et al., 2021; Perrault & Guéguen, 2015; Spence et al., 2003). Among other attributes, specific features may control the nonlinear seismic response of a building (Lagomarsino & Giovinazzi, 2006; Martins & Silva, 2021), thus bringing significant epistemic uncertainty to the damage estimation (Lestuzzi et al., 2016). The most advanced numerical models (e.g., finite element-based models, Mazzoni et al., 2006) explicitly take into account all the specific attributes of the design and geometry of an individual building. However, these more complex models may also have to cope with additional sources of uncertainties: (i) the description of the design and materials, including foundation systems, and (ii) the assessment of the actual condition of the structures. These uncertainties are all the more critical in seismic prone regions because of the possible effects of cumulative seismic damage (e.g., Perrault et al., 2020).

To resolve the afore-mentioned issues, other approaches can be used to implement a so-called “host-to-target” adjustment to eliminate the adverse effects of considering average “generic” building class models. This adjustment can be based on a modal analysis of existing buildings to define their elastic properties related to design and actual condition, or the processing of earthquake data recorded in

buildings (e.g., Mucciarelli et al. 2004; Dunand et al. 2004; Clinton et al. 2006; Masi and Vona 2010; Michel et al. 2012; Vidal et al. 2014; Perrault et al., 2013; Astorga et al. 2018, 2019; Astorga and Guéguen 2020) to take into account the nonlinear processes activated in the buildings during earthquakes and estimate their residual capacity (Dowgala & Irfanoglu, 2016; Freeman et al., 1999; Pan et al., 2019). Using these data, within-building variability  $\phi$  (i.e., the misfit between a single structure response for a given earthquake and the structure-specific median response model, which is defined as the median prediction of the model for a given structure plus the variability between earthquakes) and between-building variability  $\tau$  (i.e., the average shift of the observed structure-specific median model from the class-specific median model of the same class) can be examined to quantify, understand, and process the variability of the seismic building response and to address the origins of uncertainties through residual analysis (Fig. 5. 1). Ultimately, the total variability  $\sigma$  for a single building is given by:

$$\sigma = \sqrt{\tau^2 + \phi^2} \quad (5.1)$$



**Figure 5.1.** Between-building and within-building components of building response variability modelled in acceleration-displacement-response-spectrum format.

The main objective of this study is to develop a chain of processing to obtain experimental capacity curves using strong motion data recorded in buildings and to quantify and process the variability of the building response, taking into account the building class adjustment to capacity curves provided by Martins and Silva (2021). The dataset is described first, the selected buildings are presented in the second section, and the methods used to adjust the capacity curves from experimental data are discussed in the third section. The fourth section presents the results, and the paper ends with a discussion and conclusion.

## 5.2 Data

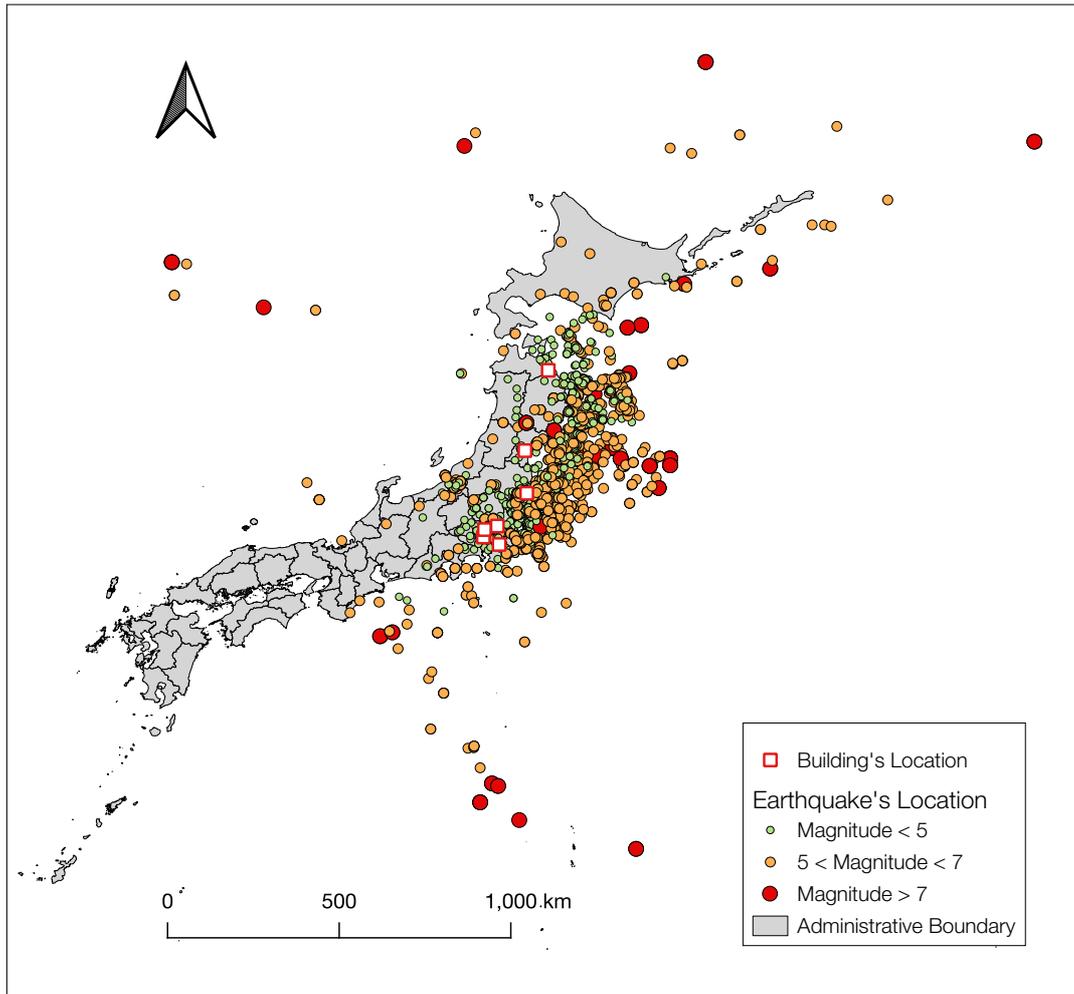
Seven Japanese buildings monitored by the Building Research Institute (BRI) with slight-to-extensive damage reported by Kashima (2014) after the 2011 Tohoku Mw9.1 earthquake were selected from NDE1.0 (Astorga et al., 2020, Fig. 5.2). Accelerometric data were recorded at the top and bottom of each building in both horizontal directions and processed (double integration for displacement and time-frequency analysis for fundamental frequency assessment) by Astorga et al. (2020).

(1) The Annex building (ANX) is an eight-storey steel reinforced concrete (SRC) structure. Since its completion in 1998, a seismic monitoring system has been installed and 1,630 earthquakes have been recorded over a 20-year period. During the 2011 Tohoku earthquake, slight-to-moderate damage was reported.

(2) The Tohoku University building (THU) is a nine-storey SRC structure constructed in 1969. Monitoring started in 1994 and comprises 203 earthquake recordings, including a series of very strong ground motion events (e.g., 2005 Miyagi-ken earthquake, 2008 Iwate–Miyagi earthquake, 2008 Iwate earthquake). Moderate-to-extensive damage was reported after the Tohoku earthquake, during which multi-storey shear walls suffered serious bending failures.

(3, 4, and 5) The Iwaki city hall building (IWK), Toda city hall building (TDS) and Chiba government office building (CHB) are eight-storey SRC buildings, with monitoring periods of 1993 to 2014, 1998 to 2013 and 2000 to 2012, respectively, corresponding to 415, 337 and 29 earthquakes. None-to-slight damage was reported in the IWK building.

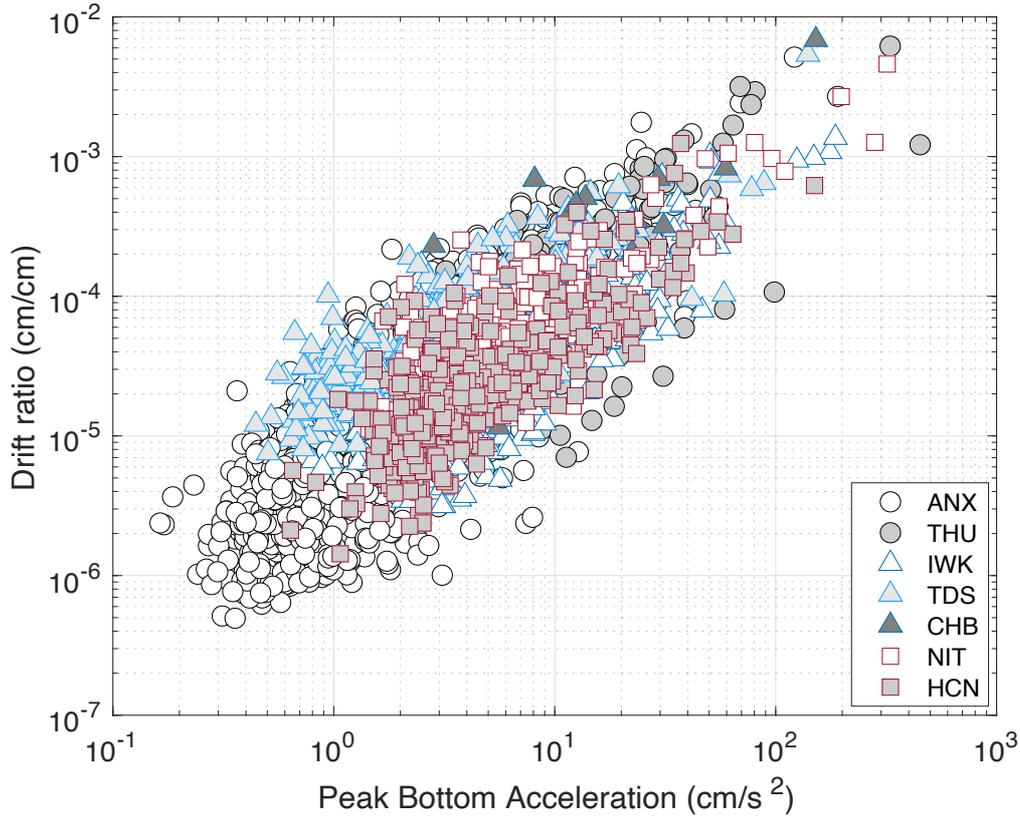
(6 and 7) The Hachinohe city hall building (HCN) and the Nippon Institute of Technology building (NIT) are six-storey reinforced concrete (RC) buildings, with recordings from 428 and 213 earthquakes over 1998-2014 and 1994-2014, respectively.



**Figure 5.2.** Building locations (open red squares) and earthquake epicentres (solid colour circles) for this study.

Structural drift, computed as the relative displacement between top and bottom and normalized by the distance between sensors, is considered as the engineering demand parameter (EDP). The intensity measure (IM) considered hereafter is the peak acceleration recorded at the bottom floor (PBA). The distribution of EDP versus IM is shown in Fig. 5.3. Structural drift ranges from about  $10^{-7}$  to  $10^{-2}$ , corresponding to PBA from approximately  $10^{-1}$  to  $500 \text{ cm/s}^2$ .

In addition to IM and EDP, resonance frequency variations were obtained by applying a Wiegner-Ville time-frequency distribution (see Astorga et al., 2020 for details) to the top recordings. Analysis of the ante-, co- and post-seismic resonance frequency of buildings identifies the nonlinear processes activated during the earthquakes due to the opening/closing of cracks (Astorga et al., 2018, 2019; Clinton et al., 2006) and improves the efficiency analysis  $P(\text{EDP}|\text{IM})$  (Ghimire et al., 2021). In this study,  $f_i$  corresponds to the frequency obtained by analyzing the noise window before the earthquake, and  $f_{\min}$  is the minimal value of the co-seismic frequency observed during strong loading.



**Figure 5.3.** Distribution of maximum acceleration at the bottom of the building versus the drift ratios (relative roof displacement normalized by height) for each building.

### 5.3 Method

Empirical capacity curves are typically represented in acceleration-displacement-response-spectrum (ADRS) format (Freeman et al., 1999). This study uses the mass normalized capacity curve in the Sa-Sd spectrum proposed by Freeman et al. (1999). The following four-step procedure is used to obtain the empirical capacity curve.

**Step 1.** Calculation of the relative displacement time histories ( $\delta_{top}$ ) as follows:

$$\delta_{top} = u_T(t) - u_B(t) \quad (5.2)$$

where  $u_T(t)$  and  $u_B(t)$  are the top and bottom floor displacements (Fig. 5.4b) obtained from the accelerometric data (Fig. 5.4a).

**Step 2.** Calculation of the time history of the fundamental frequency from  $\delta_{top}$  by applying a 4<sup>th</sup> order Butterworth filter between  $f_i$  and  $f_{min}$  (Fig. 5.4c).

**Step 3.** Calculation of Sa and Sd to develop the building response hysteresis (Fig. 5.4d) as follows:

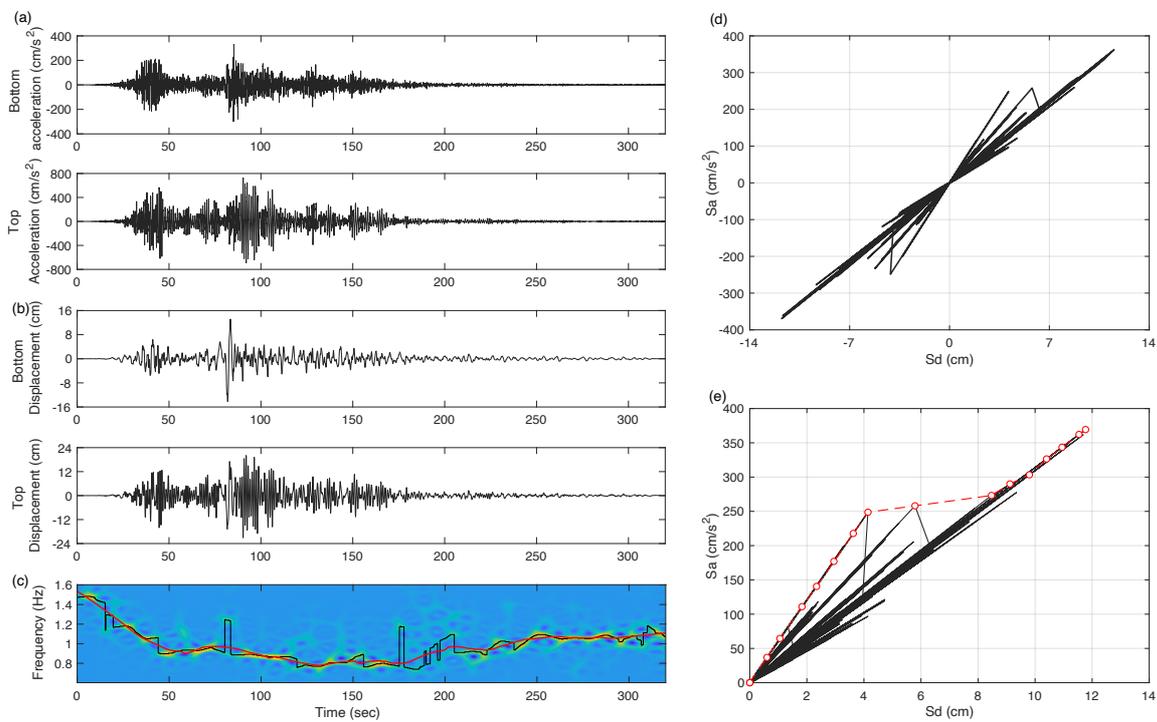
$$S_d = \frac{\delta_{top}}{\Gamma} \quad (5.3)$$

$$S_a = S_d \left( \frac{2\pi}{T} \right)^2 \quad (5.4)$$

where  $\Gamma$  is the fundamental mode performance factor ranging from 1.3 to 1.4 for a multi-degree-of-freedom system (Martins and Silva, 2021), and  $T$  corresponds to the fundamental period given by the time-frequency distribution (Fig. 5.4c).

**Step 4.** Finally, the empirical capacity curve is the envelope of the absolute value of the ADRS spectrum (Fig. 5.4e).

The mass-normalized co-seismic stiffness (i.e., mean  $\pm$  standard deviation), measured in  $s^{-2}$ , is approximated by the slopes of the peaks in the ADRS map (Fig. 5.4d).



**Figure 5.4.** Four-step procedure used to obtain the empirical capacity curve illustrated with the 2011 Tohoku earthquake recorded in the THU building. (a) Bottom and (b) Top floor recordings in acceleration and displacement time histories, (c) Time-frequency distribution of the fundamental frequency obtained from the Wigner-Ville distribution applied to the top recording with before (black) and after (red) Savitsky-Golay smoothing functions, (d) Fundamental mode building response hysteresis in the  $S_a$ - $S_d$  spectrum, (e) The  $S_a$ - $S_d$  spectrum in absolute value (black) with empirical ADRS curve (red dashes).

In this study, a selection of data from NDE1.0 is used to analyze the variation of the experimentally derived capacity considering two different contexts (before and after the 2011 Tohoku earthquake) presented hereafter. In this manuscript, the ADRS curves are referenced according to the NDE1.0 format, i.e., date of the earthquake, building code, and direction of the component considered.

One-parameter linear regression on  $S_a$  and  $S_d$  is used to analyze ADRS curve variability, as follows:

$$S_a = a + b.S_d + \varepsilon \quad (5.5)$$

where  $a$  and  $b$  are the estimated regression coefficients and  $\varepsilon$  is the standard error. The variability associated with ADRS curves is measured by calculating the standard deviation of the residuals of the fitted regression model between  $S_a$  and  $S_d$  (Eq. 5.5), representing within-building variability  $\phi$ .

First, we considered the within-building residual (Fig. 5.1) as the misfit between the capacity curves of a given structure for a single earthquake (i.e., single ADRS model) and the median response to several earthquakes (i.e., building-specific ADRS model). Within-building variability accounts for (1) the time-varying co-seismic dynamic properties related to the cumulative effect of earthquakes and the related structural ageing (Astorga et al. 2018, 2019; Trevlopoulos et al. 2020) and (2) IM efficiency obtained by measuring the variability of the building response (i.e., EDP given IM) around the regression on the IM values. Post-seismic visual inspections of ANX, THU and IWK reported slight to moderate structural damage after the 2011 Tohoku earthquake. We explored the capacity variation of these buildings over time (within-building variation) using a set of seismic events recorded before and after the 2011 earthquake.

Secondly, we considered the between-building variability (Fig. 5.1), which refers to the average shift of the model of a single structure in a specific class of structures (i.e., building-specific ADRS model), from the median model defined for this class (i.e., class-specific ADRS model). This variability can be attributed to several components, including within-building components related to ground motion, variability of structural design, site conditions, and ageing of the building. Here, we explored the between-building variability of the capacity response for the two building classes that include the 8-storey SRC buildings (i.e., IWK, TDS, and CHB) and the 6-storey RC buildings (i.e., NIT and HCN).

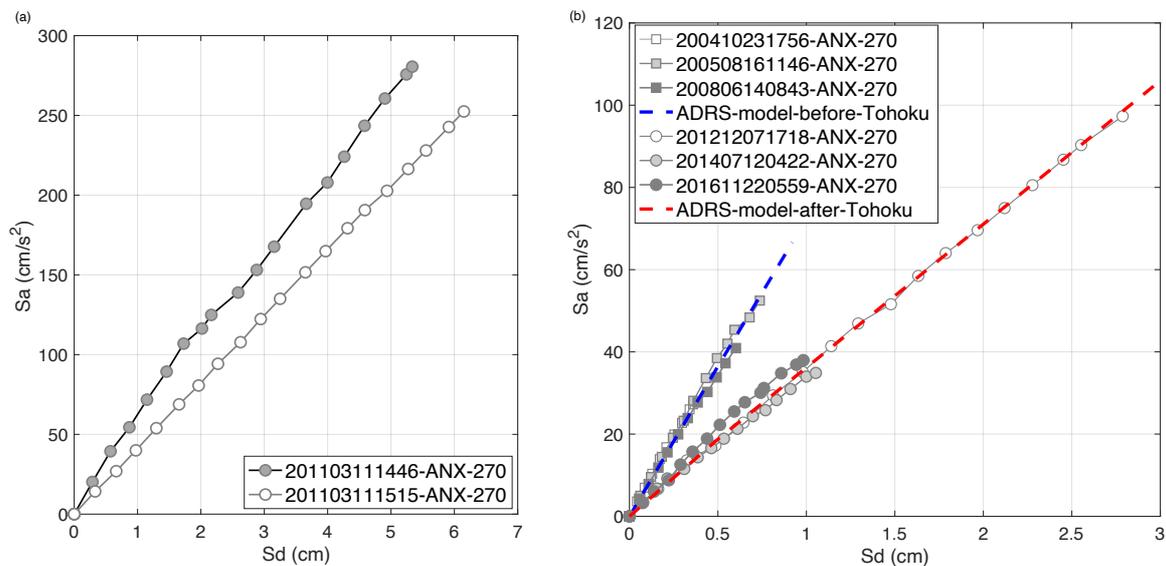
Finally, we compared the experimental co-seismic capacity curves of the ANX and THU buildings with the corresponding GEM-taxonomy capacity curves provided by Martins and Silva (2021). The comparison is made by measuring the difference in the initial slope of the ADRS curves as follows:

$$\Delta (\%) = \frac{\frac{S_a}{S_d^{\text{generic}}} - \frac{S_a}{S_d^{\text{experimental}}}}{\frac{S_a}{S_d^{\text{generic}}}} \times 100 \quad (5.6)$$

## 5.4 Results

### 5.4.1 Within-building variability $\phi$

Figure 5.5 shows the variation of the co-seismic capacity of the ANX building along one horizontal direction (N270) during a selection of earthquakes from 2004 to 2016. Fig. 5.5a corresponds to the co-seismic ADRS curve of the 2011 Tohoku earthquake (main shock file 201103111446-ANX and one immediate aftershock file 201103111515-ANX), which produced the largest displacement and acceleration values at the top of the building. Fig. 5.5b shows the ADRS curves for a selection of moderate earthquakes before and after the 2011 sequence, which clearly indicate the impact of the Tohoku earthquake (and the damage induced) on the capacity curve. The seismic damage caused an increase in variability of the ADRS response (about 39%), as indicated by the higher  $\phi$  value (1.72) after the Tohoku event compared with the  $\phi$  value (1.24) from before.

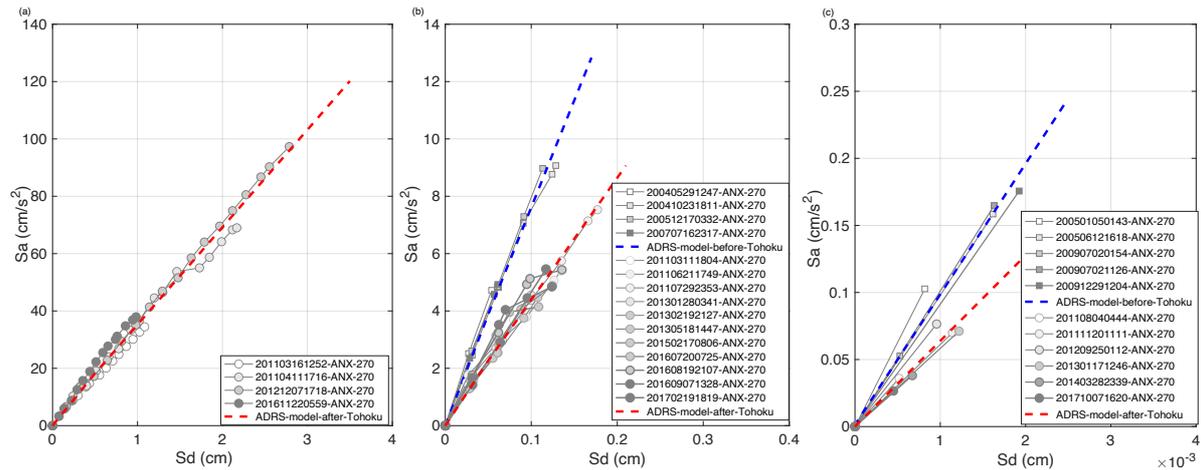


**Figure 5.5.** ADRS curve of the ANX building along the lateral direction (N270 component) before and after the Tohoku 2011 earthquake. (a) Co-seismic ADRS curves corresponding to the main shock (file 201103111446-ANX-270) and one immediate aftershock (file 201103111515-ANX-270). (b) Co-seismic ADRS curves before (squares) and after (circles) the Tohoku event. The dashed-blue and dashed-red lines represent the mean ADRS models before and after 2011.

Figure 5.6 shows the within-building variation of the capacity curves before and after the Tohoku earthquake in the ANX building for three ranges of shaking amplitude: strong, corresponding to a drift ratio of  $10^{-3}$  or above (Fig. 5.6a), moderate for a drift ratio between  $10^{-4}$  and  $1.1 \times 10^{-4}$  (Fig. 5.6b), and slight for a ratio between  $1.25 \times 10^{-6}$  and  $1.35 \times 10^{-6}$  (Fig. 5.6c). The impact of the 2011 Tohoku earthquake is clear even for moderate and slight amplitudes, as indicated by the change in slope of the before and after ADRS curves. The variation in the ADRS response is captured by the differences in the  $\phi$  values before (0.31 for moderate and 0.007 for slight shaking) and after (0.32 for moderate and 0.005 for slight shaking) the Tohoku earthquake. The impact of the deterioration of structural health on response variability is more pronounced in the ADRS response to the earthquakes that caused strong shaking

(Fig. 5.6a), also indicated by the higher  $\phi$  value (2.50) after the Tohoku earthquake, compared with the events that caused only moderate or slight shaking.

The different  $\phi$  values for earthquakes causing strong, moderate and slight shaking illustrate the direct relationship between variability and degree of shaking, which has already been reported by other studies (e.g., Masi and Vona 2010; Trifunac et al. 2010; Michel et al. 2011; Vidal et al. 2014; Astorga et al. 2018; Astorga and Guéguen 2020; Ghimire et al. 2021).



**Figure 5.6.** Co-seismic ADRS curves for the ANX building along the lateral direction, illustrating the within-building variability during various earthquakes before (grey lines with square markers) and after (grey lines with circular markers) the 2011 Tohoku earthquake, considering events inducing (a) strong shaking (drift ratio  $\geq 10^{-3}$ ), (b) moderate shaking ( $10^{-4} \leq$  drift ratio  $\leq 1.1 \times 10^{-4}$ ), and (c) slight shaking ( $1.25 \times 10^{-6} \leq$  drift ratio  $\leq 1.35 \times 10^{-6}$ ), respectively. The dashed-blue and dashed-red lines represent the curves of the mean ADRS model before and after 2011.

The same plots are given in Fig. 5.7b and 5.7c for the THU and IWK buildings, respectively, and Fig. 5.7a includes ANX for the main shock. The three buildings belong to the same class of buildings (Steel Reinforced Concrete buildings) but have different numbers of storeys. During the Tohoku earthquake, THU, ANX, and IWK experienced severe, slight-to-moderate, and none-to-slight structural damage, respectively (Kashima, 2014). In Fig. 5.7a, the slope break in the ADRS curve illustrates the co-seismic response of these buildings during the 2011 mainshock, with the threshold Sd values corresponding to the onset of damage: a significant increase in Sd values without a corresponding increase in Sa values indicates severe damage. In the building-specific ADRS model for THU (Fig. 5.7b) and IWK (5.7c), the same trend as for ANX (Fig. 5.5 and Fig. 5.6) is observed, i.e., the slopes of the ADRS model change over time as a consequence of the damage caused by the 2011 earthquake. The THU and IWK models have  $\phi$  values of 7.98 and 1.46 before Tohoku and 2.54 and 3.80 afterwards, respectively.

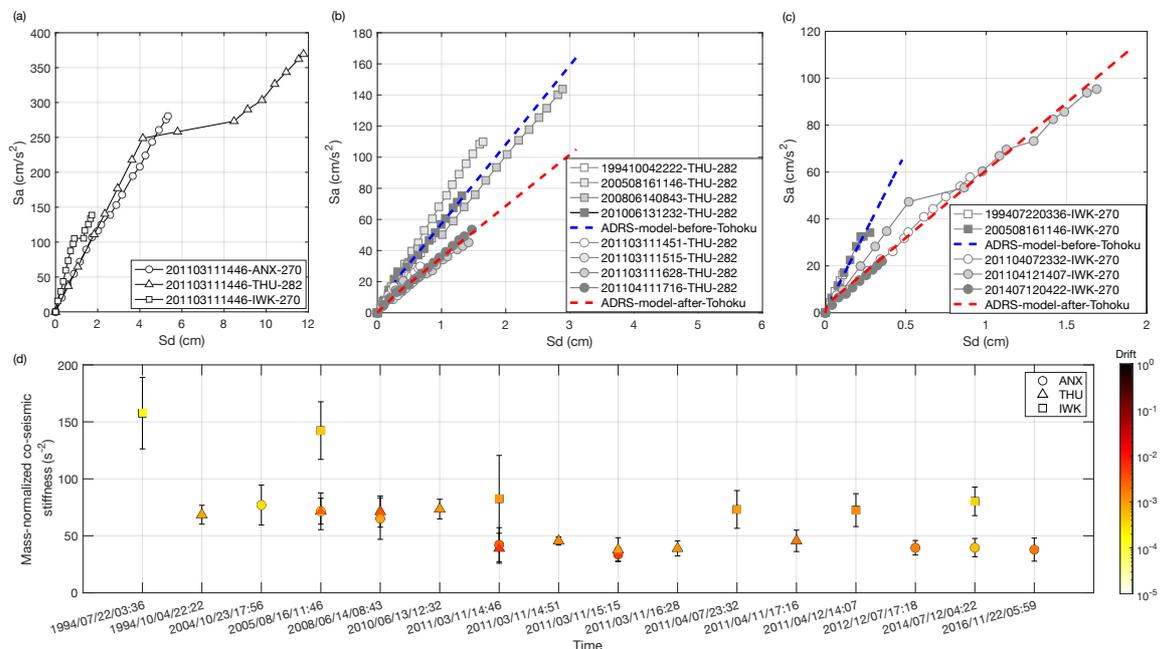
Fig. 5.7d shows the time variation of the ADRS slopes through mass-normalised co-seismic stiffness (mean  $\pm$  standard deviation) for ANX, THU, and IWK. For ANX during Tohoku (2011/03/11/14:46),

the mass-normalised co-seismic stiffness value is  $42.43 \pm 15.65 \text{ (s}^{-2}\text{)}$  with a drift ratio of  $2.7 \times 10^{-3}$ . The lowest co-seismic stiffness value ( $33.82 \pm 7.54 \text{ s}^{-2}$ ) is observed during the strongest aftershock (2011/03/11/15:15) with a drift ratio of  $5.2 \times 10^{-3}$ . No further reduction of mass-normalised co-seismic stiffness is observed during the subsequent earthquakes with a similar loading amplitude. After 2011, the mass-normalised co-seismic stiffness of the ANX building dropped by an average of 45.35%.

THU and IWK exhibit a similar trend in mass-normalised co-seismic stiffness following the Tohoku earthquake. After this event, the mass-normalised co-seismic stiffness of THU and IWK dropped by 41.39% and 24.83%, respectively. These results show that the extent of the decline in co-seismic rigidity is proportional to the amplitude of loading and the level of residual damage in the building.

Furthermore, the initial slopes of the ADRS response (Fig. 5.7a) and the mass-normalised co-seismic stiffness (Fig. 5.7d) are similar for the ANX and THU buildings. The damage also reduces the variability of the structure response, as seen here with the rigidity variation (Fig. 5.7d). This reflects the decreasing variability of the within-building capacity curve based on experimental data, as previously reported by Astorga et al. (2018) for ANX. Astorga et al. (2018, 2019), Astorga and Guéguen (2020), and Abeele et al. (2000) related the variability of the building's response to the seismic activation of pre-existing cracks in the structure.

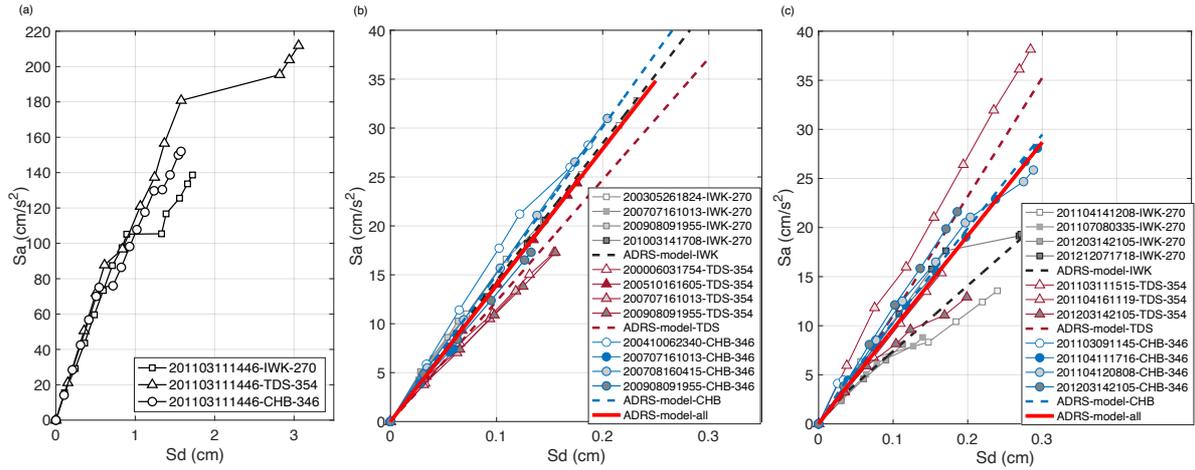
These examples illustrate how long-term monitoring enables the development of building-specific SHM-based solutions for updating actual capacity and reducing the within-building variability of the vulnerability function.



**Figure 5.7.** Co-seismic ADRS curves for the Steel Reinforced Concrete class of buildings. (a) ADRS curves for ANX, THU and IWK during the 2011 Tohoku earthquake. (b) ADRS curves before and after the 2011 earthquake for THU. (c) Same as (b) for IWK. (d) Mass-normalised co-seismic stiffness variation (mean  $\pm$  standard deviation) for ANX, THU and IWK. The colour bar indicates the range of drift ratios observed.

#### 5.4.2 Between-building variability $\tau$

In this section, the between-building variation of co-seismic capacity is discussed by considering buildings from the same class.  $\tau$  is computed as the standard deviation of the residual between the experimental building-specific ADRS model and the mean experimental class-specific ADRS model for this class. Fig. 5.8 shows the co-seismic ADRS curves for the eight-storey SRC buildings (IWK, TDS and CHB) during (Fig. 5.8a), before (Fig. 5.8b), and after (Fig. 5.8c) the 2011 Tohoku earthquake, the building-specific and the mean class-specific ADRS model. In Figures 5.8a and 5.8b, the same earthquakes are considered for all buildings. For the first part of the co-seismic ADRS curves in Fig. 5.8a, the response of the structures is very similar, characterised by the same slope. During the seismic loading caused by the Tohoku event, the responses evolve differently because of the shaking induced in each building (related to earthquake-to-building distance, for example), illustrating the damage caused with the broken slope at a given  $S_d$  value. For example, the CHB building suffered the least seismic loading, characterised by a co-seismic ADRS curve that remains mostly linear during the Tohoku earthquake, and also illustrated by the within-building variability  $\phi$  which remains similar before (1.22) and after (1.25) the 2011 earthquake. The ADRS curves for TDS and IWK show clear breaks in the slope, probably due to the strong and moderate damage suffered, respectively. The level of damage is also confirmed by the  $\phi$  values before (1.2 for TDS and 0.48 for IWK) and after (3.92 for TDS and 2.02 for IWK) the Tohoku earthquake. As a result, a significant deviation is observed between the mean ADRS curve for the class and the building-specific ADRS curves both before (Fig. 5.8b) and after (Fig. 5.8c). The changes are illustrated by the different between-building variability and total variability (Eq. 5.1) values for each building  $[\tau, \sigma]$ . The before Tohoku values are [0.53, 0.71], [2.43, 2.71], and [1.72, 2.11] for IWK, TDS, and CHB, respectively, which evolve to [4.66, 5.08], [3.69, 5.39], and [0.72, 1.44] after the earthquake. Finally, total variability for this class (corresponding to the mean class-specific ADRS model) before (1.50) and after (3.44) the Tohoku earthquake is observed to differ from the total variability for each building-specific ADRS model belonging to the same class.

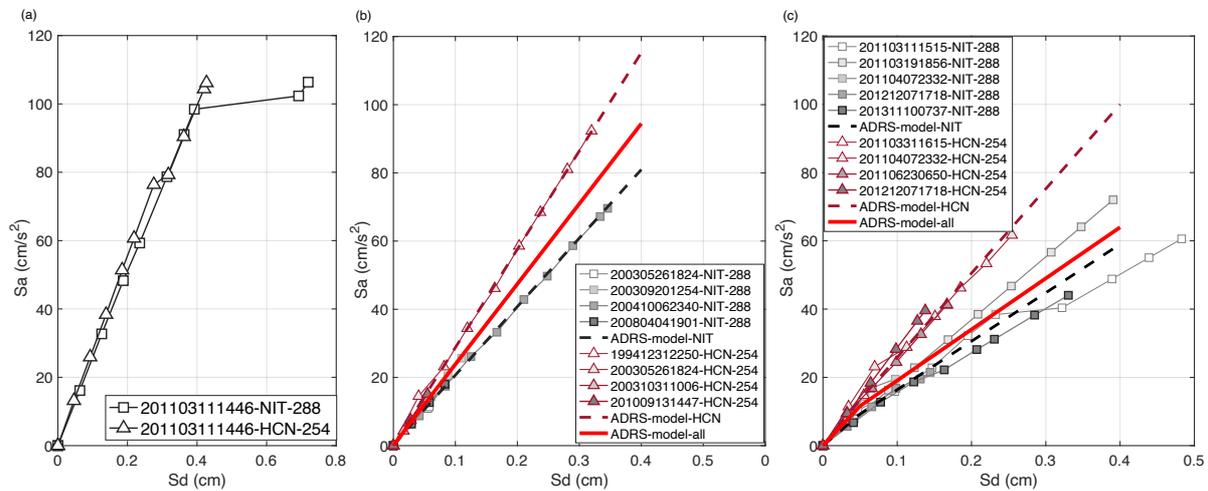


**Figure 5.8.** Co-seismic ADRS curves for 8-storey Steel Reinforced Concrete (SRC) buildings. (a) ADRS curves for TDS, CHB and IWK during the 2011 Tohoku earthquake. (b) ADRS curves before the 2011 Tohoku earthquake for IWK, TDS and CHB. (c) Same as (b) after the 2011 earthquake. The dashed-black, dashed-red, dashed-blue, and solid-red lines represent the mean ADRS curves for IWK, CHB, TDS, and for the whole class, respectively.

Fig. 5.9 gives the between-event variability analysis for the 6-storey reinforced concrete buildings (NIT and HCN). As reported by Kashima (2014), the NIT building suffered more damage than HCN in the 2011 Tohoku earthquake. As a consequence, for  $S_d < 0.4$  cm (Fig. 5.9a), the ADRS curves are quite similar for both buildings, despite a minor slope difference due to the slight variation in the response of the buildings in the same class. For  $S_d > 0.4$  cm, the NIT ADRS curve shows a broken slope, indicating damage. Also note that during the 2011 Tohoku earthquake, HCN suffered slight co-seismic degradation around  $S_d = 0.28$ cm. Furthermore, the difference in damage levels in both buildings is also illustrated by comparing the  $\phi$  value before (0.97 for NIT and 0.76 for HCN) and after (5.22 for NIT and 1.91 for HCN) the 2011 earthquake (Fig. 5.9b and Fig. 5.9c), confirming that the strongest 2011 earthquake caused damage to the NIT building. The before Tohoku values of  $[\tau, \sigma]$  are [6.45, 6.53] and [9.93, 9.96] for NIT and HCN, respectively, which become [3.46, 6.26] and [17.96, 18.06] after the earthquake. Similarly, the total variability of the class before (5.67) and after (7.29) the Tohoku earthquake differs from the total variability for each specific building in the same class.

This shows that the variability  $[\phi, \tau, \sigma]$  associated with a building class is time-dependent, being related to the actual condition of the individual buildings in the class. Furthermore, the ergodic assumption, which supposes that class-specific variability is the same as that of a single building within that class, is not validated. This is confirmed by the total variability of the class-specific ADRS model compared with that of the building-specific ADRS model for that class, both before and after the Tohoku earthquake for 8-storey SRC and 6-storey RC buildings. This is a key observation for large-scale building damage assessment methods, for which a generic class-specific capacity curve is considered

for all buildings in the same class. This can introduce significant epistemic uncertainty, as reported previously by other studies (Ghimire et al., 2021; Spence et al., 2003). Furthermore, an updated capacity curve based on experimental data enables the assessment of damage level and the residual capacity of monitored buildings after a mainshock.



**Figure 5.9.** Same as Fig. 5.8 for the 6-story Reinforced Concrete (RC) buildings, NIT and HCN.

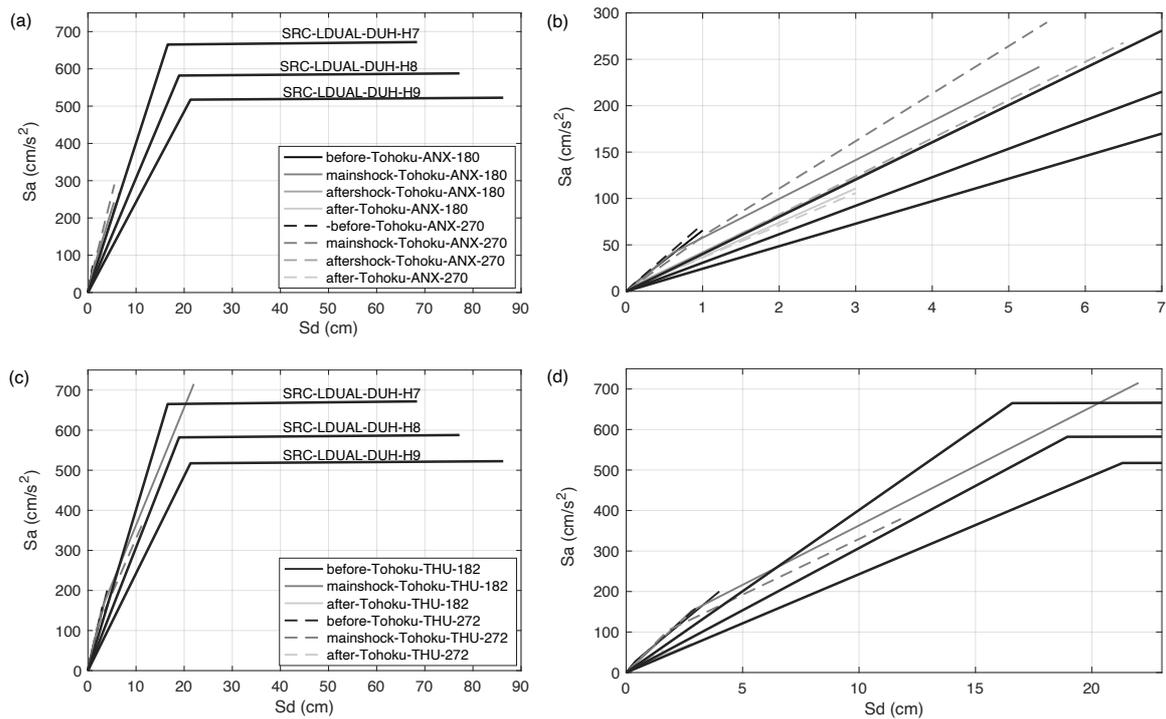
### 5.4.3 Adjustments to the generic capacity curves

Fig. 5.10 compares the experimental co-seismic capacity curves for the ANX and THU buildings in both horizontal directions with the corresponding curves of the GEM-taxonomy generic class-specific ADRS model provided by Martin and Silva (2021). The GEM-taxonomy reference for these buildings is SRC-LDUAL-DUH-Hx, where x corresponds to the number of storeys. The experimental ADRS model curves are observed to be similar in both lateral and transversal directions, with only a slight variation observed during the 2011 Tohoku earthquake.

For the ANX building (Fig. 5.10a, zoomed in 5.10b), significant differences in the initial slope are observed between the generic and experimental ADRS curves ( $\Delta$  calculated using Eq. 5.6). Over time, ANX experienced slight to moderate damage and the  $\Delta$  values fall by 114% and 20% compared with the generic capacity model values for this 8-storey SRC building before and after the Tohoku earthquake, respectively. For THU (9-storey SRC class), the  $\Delta$  values are -63% and +8% before and after the Tohoku earthquake, respectively.

As reported by Kashima (2014), the THU building (Fig. 5.10c, zoomed in 5.10d) was severely damaged by the Tohoku earthquake, and its  $\Delta$  values are high. For example, the initial slope of the generic capacity model for the nine-storey building has a  $\Delta$  value of -107% before the Tohoku earthquake and -4% after, while the generic model for the eight-storey building shows a  $\Delta$  value of -63% before and +18% after.

The difference observed between the experimental and generic capacity models in ANX and THU quantifies the adjustment needed to make the generic model more relevant to these buildings, and thus reduce epistemic uncertainty in the seismic damage assessment model. For example, for the ANX building, increasing the initial slope of the eight-storey model by 20% helps to account for the actual structural condition of ANX building. This significantly reduces the level of uncertainty in building vulnerability models applied to existing structures and contributes to the time-updating of seismic vulnerability, thus improving seismic risk assessment (Ghimire et al., 2021; Guéguen et al., 2014; Hannewald et al., 2020; Michel et al., 2010, 2012).



**Figure 5.10.** Comparison between the building-specific ADRS model and the generic GEM-Taxonomy model for the THU and ANX class. (a) ANX building, zoomed in (b), (c) THU building, zoomed in (d). The generic capacity model curves are represented by the thick black lines, the thin, solid, and dashed lines represent the experimental capacity curves before, during, and after the Tohoku earthquake along the transversal and lateral directions.

## 5.5 Discussion

Previous studies have attempted to develop capacity models describing the lateral strength and deformation capacity of buildings or building classes (Martins and Silva, 2021). These capacity models are defined based on simulated design-based research studies that characterise building attributes (e.g., materials, lateral load resisting system, number of storeys, etc.) using simple single- or multi-degree-of-freedom systems. Such developments are based on significant modelling assumptions, which may prevent full understanding of the actual response of structures, as well as increasing epistemic

uncertainty in risk assessment. Some studies have attempted to develop experimental capacity curves using strong motion data recorded in buildings (Dowgala & Irfanoglu, 2016; Freeman et al., 1999; Pan et al., 2019). Few of these studies have analysed the variability of the co-seismic response of buildings using experimental capacity curves.

In this study, we explore the variability of building response for specific buildings with time-varying conditions and different earthquakes, and between buildings with the same typology using experimental capacity curves developed from strong motion data recorded in buildings. Inspired by Al Atik et al. (2010), who analysed the variability and variability components of empirical ground motion models, we examined the within-building variation of co-seismic capacity by analysing separately the ANX, THU and IWK buildings, which suffered slight to moderate structural damage during the 2011 Tohoku earthquake (Kashima, 2014). Our analysis of the experimental capacity curves revealed clear evidence of damage in these buildings resulting from the Tohoku earthquake, indicated by a shift in the initial slope and higher variability in the ADRS models. We also found that the degree of variability is linked to the degree of damage in the building. For instance, after the Tohoku earthquake, the ANX, THU, and IWK buildings show within-building variability of 1.72, 2.54, and 3.80, respectively. We also show that the variability of the ADRS curve depends on the amplitude of seismic loading. For example, for ANX, variability is 2.50, 0.32, and 0.01 for strong, moderate, and slight shaking, respectively, which we interpret as being due to the activation of existing cracks in the building (Astorga et al. 2018, 2019; Astorga and Guéguen, 2020; Abeele et al., 2000). Thus, the time-varying condition of the buildings, due to the cumulative effects of earthquakes, produces a change in the variability of the building-specific ADRS model for the same class. For example, IWK, TDS, and CHB (6-storey SRC class) show between-building variability of 0.53, 2.43, and 1.72 before the Tohoku earthquake and 4.66, 3.69, and 0.72 after the earthquake, respectively. The differences in response variability among buildings of the same class may be related to the degree of heterogeneity in the residual crack systems and the nature of the crack activation mechanisms during seismic activity.

Comparison of the experimental capacity curves of buildings with long-term monitoring (ANX and THU) and their GEM-taxonomy generic capacity models shows significant differences between the experimental (building-specific) and generic (class-specific) models. For example, following the Tohoku earthquake, a 20% difference was observed for the ANX building, indicating that the initial slope of the generic capacity model should be adjusted by 20% to account for the actual structural condition of the building. This updated ADRS model enables the use of time-varying co-seismic capacity for seismic risk assessment studies or operational loss forecasting, and at the same reducing epistemic uncertainty (Trevlopoulos and Guéguen 2016, Trevlopoulos et al. 2020, Ghimire et al. 2021, Perrault and Gueguen 2015).

Our study found that, during the Tohoku earthquake, the ANX building experienced a 45.35% co-seismic stiffness drop with slight-to-moderate damage, while the THU building experienced a 41.39% drop with moderate-to-extensive damage. Furthermore, when we analysed the THU building for a

different range of shaking amplitudes, we observed different values of co-seismic stiffness degradation: 36.43% for strong shaking (drift ratio of  $1 \times 10^{-3}$  or more), 21.71% for moderate shaking (drift ratio between  $1 \times 10^{-4}$  and  $1.4 \times 10^{-4}$ ), and 9.34% for slight shaking (drift ratio between  $2 \times 10^{-5}$  and  $3 \times 10^{-5}$ ). These findings suggest that the change in co-seismic stiffness is proportional to the shaking amplitude and the level of residual damage in the building before the earthquake under consideration. Previous studies (e.g., Dunand et al. 2004; Masi and Vona 2010; Trifunac et al. 2010; Michel et al. 2011; Vidal et al. 2014) have defined different thresholds of co-seismic stiffness/period elongation for detecting damage and estimating building condition. However, our results highlight the need for caution when implementing co-seismic stiffness-degradation-threshold-based approaches for structural health monitoring, as highlighted by Astorga et al. (2018).

Experimental capacity curves also reflect earthquake-induced damage in buildings, as illustrated by the broken slope of the ADRS model. The plateau of the break is proportional to the amount of damage in these buildings, as reported by Kashima (2014). These curves highlight the fact that actual structural condition is required to evaluate capacity immediately after an earthquake. The experimental capacity curves developed in this study only consider the fundamental mode, and additional parameters (high modes, damping, etc.) could be used to assess the impact of such parameters on total variability.

## 5.6 Conclusion

Experimental data from buildings are crucial for quantifying, understanding and processing the variability of the seismic building response and for identifying the origin of uncertainties through residual analysis. Experimental data offers the advantage of including the complex physical processes that affect buildings during seismic loading, and consideration of the time-varying capacity due to the actual condition of the building. In this study, we used strong motion data collected from seven Japanese buildings equipped with permanent instrumentation and post-processed in the NDE1.0 database (Astorga et al., 2020) to investigate variability in the co-seismic capacity curves. We developed an ADRS-based framework, considering the method proposed by Freeman et al. (1999). We explored the within-building variability related to the building-specific response for a given earthquake, including its change in condition, and the between-building variability related to the drift of building-specific and class-specific models, which are derived from experimental data or selected from the GEM-Taxonomy models.

Analysis of the ANX building's co-seismic capacity before and after the 2011 Tohoku earthquake shows that the damage helped to reduce ANX's co-seismic stiffness by 45% and increase the variability of the co-seismic capacity response by 39%. Furthermore, we observed that strong shaking earthquakes resulted in greater variability in the co-seismic capacity response compared with moderate to slight earthquakes. The IWK and THU buildings, which experienced none-to-slight and moderate-to-severe damage during the Tohoku earthquake, also showed the same trend. Similarly, the other buildings in the

same class (IWK, CHB, TDS, and NIT, HCN) confirmed that cumulative seismic loading of varying amplitudes over time resulted in significant changes in the co-seismic capacity curves. This, in turn, led to changing levels of variability among the buildings in the same class, as indicated by different between-building variability values before and after the 2011 Tohoku earthquake. Consequently, the response variability of a given building in a given class differs from that of the class as a whole, which contradicts the ergodic assumption that the response variability for a class of buildings is the same as the response variability for any given building in that class. Time-varying experimental capacity models are useful for calculating time-varying seismic risk, especially after strong earthquakes or during a mainshock-aftershock sequence. Furthermore, the difference observed between experimental curves and the generic capacity curves obtained from the global databank of capacity curves can quantify the adjustments required to account for a building's actual structural condition.

Using the model developed in this study, we derived experimental capacity curves that took into account the damage-related condition and the necessary adjustment of the generic capacity models by reducing epistemic uncertainty. These experimental capacity curves can be helpful for characterising the actual seismic integrity of structures, for both time-variant seismic risk assessment and operational loss forecasting. These empirical capacity curves were developed by considering the fundamental mode period, and do not reflect the entire capacity of the building, i.e., considering the contribution of the higher modes in the building co-seismic response and the associated variability. Finally, there is a huge need for building testing to develop and calibrate actual models, which are superior to even the most sophisticated numerical models. The most important aspect is possibly not only related to the median values of engineering demand parameters, but their variability according to the physical processes in the structure that are activated by shaking.

## General conclusion

Recently, earthquake data collected from buildings have become openly accessible, including post-earthquake damage survey databases and strong-motion data recordings. These data represent the actual physical processes involved in buildings during earthquake loading more than even the most sophisticated numerical studies. In this study, we used the post-earthquake building damage survey database and strong-motion data recordings from buildings to develop new insights into the large-scale rapid damage assessment and building-level seismic damage assessment, respectively.

In the framework of the rapid damage assessment at a large scale, we evaluated the effectiveness of machine learning models by training and testing on the post-earthquake building damage dataset. Machine learning models could reasonably estimate damage in a cost-effective way. Readily available building features such as the number of stories, age, floor area, and height can result in a reasonable assessment of damage at a large scale, particularly when using a traffic-light-based (green, yellow, and red) damage classification framework. Machine learning models trained on these readily available building features showed efficacy similar to conventional methods like Risk-UE. In addition, these parameters can be easily obtained from an institutional database, such as the national census or national housing database, thereby resolving data acquisition issues associated with seismic-damage assessments at the urban or regional scale.

We observed that the distribution of target features (damage grades) in the training set could significantly affect the damage prediction efficacy of the machine learning model. Random oversampling method or integrating naturally balanced datasets (selecting data from a wider range of input features and earthquake damage from different regions) while training the machine learning model can penalize the skewed distribution of the target features and improve the damage prediction efficacy of the machine learning model. The efficacy of a machine learning model is influenced by the quality and diversity of the training dataset, i.e., the efficacy of a machine learning model depends more on the quality of the dataset's resolution than on the quantity of dataset used in the training phase. Thus, increasing the number of damaged buildings in the training set may not improve the model's efficacy beyond a certain limit, but improving the dataset's resolution can enhance the prediction accuracy of the damage.

We also observed that the contribution of the building's feature changes depending on the damage level considered. This behaviour has not been discussed previously. The machine learning models trained on past earthquake building damage portfolios can reasonably estimate damage during the future earthquake for a different region with similar building portfolios. However, the damage distribution is strongly influenced by region-specific factors such as construction quality and regional typologies, implementation of seismic regulations, and hazard level. Thus, machine learning models trained with

building damage datasets from one region may not directly apply to another region without first accounting for host-to-target adjustments.

Focusing on the building-level damage assessment, we evaluated the co-seismic response of a building during seismic loading using the earthquake data recordings from buildings. We found that using ground motion intensity measures that accurately capture a building's co-seismic response (for example, as spectral velocity at the co-seismic period) by considering the building's structural state, along with the earthquake's magnitude and distance, can significantly decrease the level of uncertainty in damage assessment.

We also observed that the damage caused by an earthquake and the damage accumulation over time could alter the co-seismic response of a building. The seismic loading of different amplitude can impact buildings with similar typologies differently. As a result, the co-seismic response of buildings with the same typologies can be significantly different over time. Assuming a generic co-seismic response for a building throughout its lifetime or grouping buildings into a class based on taxonomy and assuming a generic co-seismic response to all buildings during seismic loading (such as in fragility modeling or building response modeling in the performance-based earthquake engineering framework) can lead to significant uncertainty in damage assessment.

Co-seismic capacity curves can accurately capture a building's lateral load-resisting capacity during earthquake loading. The extent of damage caused by the first occurrence and the accumulation of damage over time can be easily reflected by a shift in the co-seismic capacity curves proportional to the degree of damage. The capacity curves help to quantify the necessary adjustments to generic capacity models and can help to reduce epistemic uncertainties in damage assessment. A time-varying capacity model is necessary for calculating time-varying seismic risk. The experimental capacity curves can be an efficient tool for characterizing a structure's actual seismic integrity for both time-variant seismic risk assessment and operational loss forecasting, particularly after strong earthquakes or during a mainshock aftershock sequence.

This study further highlighted the importance of real earthquake datasets to enhance our understanding of the seismic damage response of buildings and develop a tool for seismic damage assessment. More information on the earthquakes and descriptions of the building characteristics would help cover diverse aspects of the seismic damage or loss assessment framework. To further develop our insights into the seismic response of buildings and strengthen the conclusions mentioned above through data-driven studies, we must continue our global collaboration to collect and share earthquake data collected from buildings.

## Perspectives

Developing the dynamic damage assessment method requires rigorous testing of the damage response of buildings during seismic loading. Several issues need to be studied to fully develop the dynamic damage assessment method as a tool for seismic risk decision-making. Potential issues related to this study need to be addressed in further studies and are presented here as follows.

In the framework of the rapid damage assessment at a large scale:

- (a) In Chapter 3, we observed that the accuracy of machine learning-based damage assessment at large-scale is heavily influenced by region-specific factors such as construction quality, regional typologies, implementation of seismic regulations, and hazard level; host-to-target adjustments are required to apply machine learning models trained in one region to another. Further studies should focus on quantifying these adjustments based on the region-specific factors mentioned above. Similarly, the accuracy of damage assessment must also be compared between machine learning models developed in different regions with and without focusing on these specific features (e.g., typology and age).
- (b) In Chapters 1-3, we observed that the readily available basic building features (age, number of stories, floor area, height) could result in a reasonable building damage assessment. The global dynamic exposure model incorporates a significant amount of low-resolution but high-density open-building data describing some basic features of the building. Further studies can be focused on evaluating the value of this data; and developing a system that integrates the machine learning-based damage prediction models and the global dynamic exposure model to facilitate real-time building damage assessment. To start, we can develop a database that includes damaged buildings surveyed in the field which are also included in the global dynamic exposure model. Using this database, we can develop a machine learning model and conduct a series of comparative analysis to assess the significance of the open building data in the global dynamic exposure model for large-scale damage assessment in real-time.
- (c) Future studies could also focus on gathering earthquake loss data and evaluating machine learning models directly for loss assessment. These models will map building features with the loss value, bypassing vulnerability modeling and translating damage to loss.

In the framework of real-time building-level damage assessment:

- (d) In Chapter 4, we observed that the engineering demand parameter (EDP) is not conditional independent with earthquake magnitude and distance. In reality, buildings near fault rupture were observed to have significant damage (Bessason and Bjarnason, 2016). Previous study suggest that, the pulse-like signals of near-fault ground motions can induce fewer cycles in structures than far-field ground motions but with higher energy content (Mollaioli et al., 2011).

As a result, buildings may experience sudden demands for energy and displacement, leading to a larger concentration of drift demand (Mollaioli et al., 2011; Baker and Cornell, 2008b). Additionally, higher modes may affect high-rise buildings near the fault area and the maximum displacement may not necessarily be at the top-story level (Mollaioli et al., 2011). In NDE1.0, we have earthquake data recordings in buildings located near to the epicentre. Further studies can be carried out to explore the building response for given intensity measures (EDP|IM) near the epicentre using the earthquake data recordings from the building available in NDE1.0.

- (e) Chapter 5 presents co-seismic capacity curves, which enable real-time tracking of the residual load-resisting capacity of buildings that are equipped with strong motion sensors. In reality, not all buildings are currently being monitored by a strong motion sensors network. Various analytical methods have been developed to facilitate the real-time evaluation of potential damage to non-instrumented buildings based on structural health monitoring applications to assess the actual building's condition (e.g., by measuring certain dynamic parameters, such as the fundamental frequency) (e.g., Cremen and Baker, 2018; Goulet, Michel and Kiureghian, 2015; Hwang and Lignos, 2018; Porter et al., 2006; Reuland et al., 2019; Tubaldi et al., 2022; Uma, 2007). However, a paradigm shift from model-driven to data-driven methods is needed to test existing methods and to provide data-driven solutions improving our understanding in this domain (Ozer et al., 2022). The NDE1.0 database contains a substantial amount of earthquake data that can be utilized to develop and test solutions for the structural health monitoring-based operational earthquake loss forecasting (OELF) framework required to facilitate real-time damage assessment for non-instrumented buildings. Further studies can be carried out in this domain.
- (f) The earthquake data recordings in the NDE1.0 database can be used to test the feasibility of integrating machine learning-based solutions for the OELF framework (for e.g., Iaccarino et al., 2021). Using earthquake data recordings from buildings available in NDE1.0, further studies can be carried out to develop a framework to integrate the machine learning methods with the OELF method for building level damage assessment.

## References

- Abeele, V. Den, Visscher, K. De, and Joëlle: Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques, *Cem. Concr. Res.*, 30, 1453–1464, [https://doi.org/10.1016/S0008-8846\(00\)00329-X](https://doi.org/10.1016/S0008-8846(00)00329-X), 2000.
- Adam, C., Kampenhuber, D., Ibarra, L. F., and Tsantaki, S.: Optimal Spectral Acceleration-based Intensity Measure for Seismic Collapse Assessment of P-Delta Vulnerable Frame Structures, *J. Earthq. Eng.*, 21, 1189–1195, <https://doi.org/10.1080/13632469.2016.1210059>, 2017.
- Akhlaghi, M. M., Bose, S., Mohammadi, M. E., Moaveni, B., Stavridis, A., and Wood, R. L.: Post-earthquake damage identification of an RC school building in Nepal using ambient vibration and point cloud data, *Eng. Struct.*, 227, <https://doi.org/10.1016/j.engstruct.2020.111413>, 2021.
- Araya, R. and Saragoni, G. R.: Earthquake accelerogram destructiveness potential factor, in: Paper presented at the Proc. 8th World Conference on Earthquake Engineering, 7, 1984.
- Arias, A.: MEASURE OF EARTHQUAKE INTENSITY, 1970.
- Astorga, A. and Guéguen, P.: Influence of seismic strain rates on the co- and post-seismic response of civil engineering buildings, *Earthq. Eng. Struct. Dyn.*, 49, 1758–1764, <https://doi.org/10.1002/EQE.3328>, 2020a.
- Astorga, A. and Guéguen, P.: Structural health building response induced by earthquakes: Material softening and recovery, *Eng. Reports*, 2, e12228, <https://doi.org/10.1002/ENG2.12228>, 2020b.
- Astorga, A., Guéguen, P., and Kashima, T.: Nonlinear elasticity observed in buildings during a long sequence of earthquakes, *Bull. Seismol. Soc. Am.*, 108, 1185–1198, <https://doi.org/10.1785/0120170289>, 2018.
- Astorga, A., Guéguen, P., Ghimire, S., and Kashima, T.: NDE1.0: a new database of earthquake data recordings from buildings for engineering applications, *Bull. Earthq. Eng.*, 18, 1321–1344, <https://doi.org/10.1007/s10518-019-00746-6>, 2020.
- Astorga, A. L., Guéguen, P., Rivière, J., Kashima, T., and Johnson, P. A.: Recovery of the resonance frequency of buildings following strong seismic deformation as a proxy for structural health, *Struct. Heal. Monit.*, 18, 1966–1981, <https://doi.org/10.1177/1475921718820770>, 2019.
- ATC: ATC-20-1, Field Manual: Postearthquake Safety Evaluation of Buildings Second Edition,, Applied Technology Council, Redwood City, California., 2005.
- Atkinson, G. M.: Ground-motion prediction equation for small-to-moderate events at short hypocentral distances, with application to induced-seismicity hazards, *Bull. Seismol. Soc. Am.*, 105, 981–992, <https://doi.org/10.1785/0120140142>, 2015.
- Atkinson, G. M. and Morrison, M.: Observations on regional variability in ground-motion amplitudes for small-to-moderate earthquakes in North America, *Bull. Seismol. Soc. Am.*, 99, 2393–2409, <https://doi.org/10.1785/0120080223>, 2009.
- Auger, F. and Flandrin, P.: Improving the Readability of T-F and T-S by Reassignment Method.pdf, *Ieee Trans. Signal Process.* Vol. 43, No. 5, May 1995, 43, 1068–1089, 1995.
- Azimi, M., Eslamlou, A. D., and Pekcan, G.: Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review, <https://doi.org/10.3390/s20102778>, 2020.
- Baggio, C., Bernardini, A., Colozza, R., Pinto, A. V., and Taucer, F.: Field Manual for post-earthquake damage and safety assessment and short term countermeasures (AeDES) Translation from Italian: Maria ROTA and Agostino GORETTI, 2007.
- Baker, J. W. and Cornell, C. A.: Uncertainty propagation in probabilistic seismic loss estimation, *Struct. Saf.*, 30, 236–252, <https://doi.org/10.1016/j.strusafe.2006.11.003>, 2008a.
- Baker, J. W. and Cornell, C. A.: Vector-valued intensity measures for pulse-like near-fault ground motions, *Eng. Struct.*, 30, 1048–1057, <https://doi.org/10.1016/j.engstruct.2007.07.009>, 2008b.
- Bazzurro, P., Cornell, C. A., Menun, C., and Motahari, M.: GUIDELINES FOR SEISMIC ASSESSMENT OF DAMAGED BUILDINGS, in: 13th World Conference on Earthquake Engineering, Vancouver, B.C. m Canada, 74–76, <https://doi.org/10.5459/bnzsee.38.1.41-49>, 2004.
- Benjamin, J. R. and Cornell, C. A.: Solutions Manual to Accompany Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill, 1970.
- Bennett, J.: OpenStreetMap, 1st ed., Packt Publishing Ltd., 252 pp., 2010.
- Bessonon, B. and Bjarnason, J. Ö.: Seismic vulnerability of low-rise residential buildings based on damage data from three earthquakes (Mw6.5, 6.5 and 6.3), *Eng. Struct.*, 111, 64–79, <https://doi.org/10.1016/j.engstruct.2015.12.008>, 2016.
- Bianchini, M., Diotallevi, P. P., and Baker, J. W.: Prediction of Inelastic Structural Response Using an Average of Spectral Accelerations, 8, 2009.
- Bommer, J. J.: Earthquake hazard and risk analysis for natural and induced seismicity: towards objective assessments in the face of uncertainty, *Bull. Earthq. Eng.* 2022 206, 20, 2825–3069, <https://doi.org/10.1007/S10518-022-01357-4>, 2022.
- Bommer, J. J. and Crowley, H.: The influence of ground-motion variability in earthquake loss modelling, *Bull. Earthq. Eng.*, 4, 231–248, <https://doi.org/10.1007/s10518-006-9008-z>, 2006.

- Bommer, J. J., Magenes, G., Hancock, J., and Penazzo, P.: The influence of strong-motion duration on the seismic response of masonry structures, *Bull. Earthq. Eng.*, 2, 1–26, <https://doi.org/10.1023/B:BEEE.0000038948.95616.bf>, 2004.
- Borzi, B., Pinho, R., and Crowley, H.: Simplified pushover-based vulnerability analysis for large-scale assessment of RC buildings, *Eng. Struct.*, 30, 804–820, <https://doi.org/10.1016/j.engstruct.2007.05.021>, 2008.
- Branco, P., Ribeiro, R. P., Torgo, L., Krawczyk, B., and Moniz, N.: SMOGN: a Pre-processing Approach for Imbalanced Regression, *Proc. Mach. Learn. Res.*, 74, 36–50, 2017.
- Breiman, L.: Random Forests, *Mach. Learn.*, 5–32, 2001.
- Buratti, N.: A comparison of the performances of various ground-motion intensity measures, in: *In 15th World Conference on Earthquake Engineering*, Lisbon, Portugal, 77, 2012.
- Calvi, G. G. M., Pinho, R., and Crowley, H.: State-of-the-knowledge on the period elongation of RC buildings during strong ground shaking, *First Eur. Conf. Earthq. Eng. Seismol.*, 3–8, 2006.
- Chaulagain, H., Rodrigues, H., Silva, V., Spacone, E., and Varum, H.: Seismic risk assessment and hazard mapping in Nepal, *Nat. Hazards*, 78, 583–602, <https://doi.org/10.1007/s11069-015-1734-6>, 2015.
- Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: *22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794, <https://doi.org/10.1145/2939672.2939785>, 2016.
- Chi, N. W., Wang, J. P., Liao, J. H., Cheng, W. C., and Chen, C. S.: Machine learning-based seismic capability evaluation for school buildings, *Autom. Constr.*, 118, 103274, <https://doi.org/10.1016/j.autcon.2020.103274>, 2020.
- Clinton, J. F., Bradford, S. C., Heaton, T. H., and Favela, J.: The observed wander of the natural frequencies in a structure, *Bull. Seismol. Soc. Am.*, 96, 237–257, <https://doi.org/10.1785/0120050052>, 2006.
- Colombi, M., Borzi, B., Crowley, H., Onida, M., Meroni, F., and Pinho, R.: Deriving vulnerability curves using Italian earthquake damage data, *Bull. Earthq. Eng.*, 6, 485–504, <https://doi.org/10.1007/s10518-008-9073-6>, 2008.
- Cornell, C. A. and Luco, N.: *The Effect of Connection Fractures on Steel Moment Resisting Frame Seismic Demands and Safety*, 1999.
- Cortes, C. and Vapnik, V.: Support-vector networks, *Mach. Learn.*, 273–297, 1995.
- Cremen, G. and Baker, J. W.: Quantifying the benefits of building instruments to FEMA P-58 rapid post-earthquake damage and loss predictions, *Eng. Struct.*, 176, 243–253, <https://doi.org/10.1016/j.engstruct.2018.08.017>, 2018.
- Crowley, H., Rodrigues, D., Crowley, H., and Rodrigues, D.: *The European Seismic Risk Model 2020 ( ESRM 2020 )*, Iconhic 2019, 2020, 2019.
- Crowley, H., Despotaki, V., Rodrigues, D., Silva, V., Toma-Danila, D., Riga, E., Karatzetou, A., Fotopoulou, S., Zugic, Z., Sousa, L., Ozcebe, S., and Gamba, P.: Exposure model for European seismic risk assessment, *Earthq. Spectra*, 36, 252–273, <https://doi.org/10.1177/8755293020919429>, 2020.
- Crowley, H., Dabbeek, J., Despotaki, V., Rodrigues, D., Martins, L., Silva, V., Romao, X., Pereira, N., Weatherill, G., and Danciu, L.: *European Seismic Risk Model (ESRM20)*, 2021.
- Daniell, J. E., Schaefer, A. M., Wenzel, F., and Tsang, H. H.: The global role of earthquake fatalities in decision-making: earthquakes versus other causes of fatalities, *Proc. Sixt. world Conf. Earthq. Eng. Santiago, Chile*, 9–13, 2017.
- DesRoches, R., Comerio, M., Eberhard, M., Mooney, W., and Rix, G. J.: Overview of the 2010 Haiti Earthquake, <https://doi.org/10.1193/1.3630129>, 27, <https://doi.org/10.1193/1.3630129>, 2011.
- Dhakal, Y. P., Takai, N., and Sasatani, T.: Empirical analysis of path effects on prediction equations of pseudo-velocity response spectra in northern Japan, *Earthq. Eng. Struct. Dyn.*, 39, 443–461, <https://doi.org/10.1002/eqe.952>, 2008.
- Dolce, M., Speranza, E., Giordano, F., Borzi, B., Bocchi, F., Conte, C., Meo, A. Di, Faravelli, M., and Pascale, V.: Observed damage database of past Italian earthquakes: The da.D.O. WebGIS, *Boll. di Geofis. Teor. ed Appl.*, 60, 141–164, <https://doi.org/10.4430/bgta0254>, 2019.
- Douglas, J.: *Ground Motion Prediction Equations 1964-2020*, <http://www.gmpe.org.uk>, 2020.
- Dowgala, J. and Irfanoglu, A.: A Method for Extracting Building Empirical Capacity Curves from Earthquake Response Data, *Earthq. Spectra*, <https://doi.org/doi:10.1193/122714EQS219M>, 2016.
- Dunand, F., AIT MEZIANE, Y., MACHANE, D., MEZOUEUR, N., NOUR, A., OUBAICHE, E.-H., REMAS, A., GUEGUEN, P., CHATELAIN, J.-L., GUILLIER, B., BEN SALEM, R., HADID, M., HELLEL, M., KIBOUA, A., and LAOUAMI, N.: Utilisation du bruit de fond pour l’analyse des dommages des bâtiments de Boumerdès suite au séisme du 21 Mai 2003, *Mémoires du Serv. géologique l’Algérie*, 177–191, 2004.
- Eads, L., Miranda, E., and Lignos, D. : Average spectral acceleration as an intensity measure for collapse risk assessment, *Earthq. Eng. Struct. Dyn.*, 44, 2057–2073, <https://doi.org/10.1002/eqe>, 2015.
- Earle, P. S., Wald, D. J., Jaiswal, K. S., Allen, T. I., Hearne, M. G., Marano, K. D., Hotovec, A. J., and Fee, J. M.: Prompt assessment of global earthquakes for response (pager): A system for rapidly determining the impact

- of earthquakes worldwide, *Earthq. Res. Backgr. Sel. Reports*, 31–46, 2010.
- Ebrahimian, H., Jalayer, F., Lucchini, A., Mollaioli, F., and Manfredi, G.: Preliminary ranking of alternative scalar and vector intensity measures of ground shaking, *Bull. Earthq. Eng.*, 13, 2805–2840, <https://doi.org/10.1007/s10518-015-9755-9>, 2015.
- Eleftheriadou, A. K. and Karabinis, A. I.: Development of damage probability matrices based on Greek earthquake damage data, *Earthq. Eng. Eng. Vib.*, 10, 129–141, <https://doi.org/10.1007/s11803-011-0052-6>, 2011.
- Eleftheriadou, A. K. and Karabinis, A. I.: Seismic vulnerability assessment of buildings based on damage data after a near field earthquake (7 September 1999 Athens - Greece), *Earthq. Struct.*, 3, 117–140, <https://doi.org/10.12989/eas.2012.3.2.117>, 2012.
- EPRI: A Criterion for Determining Expedience of the Operating Basis Earthquake, 1988.
- Erdik, M., Şeşetyan, K., Demircioğlu, M. B., Hancilar, U., and Zülfikar, C.: Rapid earthquake loss assessment after damaging earthquakes, *Soil Dyn. Earthq. Eng.*, 31, 247–266, <https://doi.org/10.1016/j.soildyn.2010.03.009>, 2011.
- Estabrooks, A. and Japkowicz, N.: A mixture-of-experts framework for learning from imbalanced data sets, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2189, 34–43, [https://doi.org/10.1007/3-540-44816-0\\_4](https://doi.org/10.1007/3-540-44816-0_4), 2001.
- Fajfar, P.: A Nonlinear Analysis Method for Performance-Based Seismic Design, *Earthq. Spectra*, 16, 573–592, <https://doi.org/10.1193/1.1586128>, 2000.
- Farrar, C. R. and Worden, K.: An introduction to structural health monitoring, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 365, 303–315, <https://doi.org/10.1098/RSTA.2006.1928>, 2006.
- FEMA: Hazus –MH 2.1 Multi-hazard Loss Estimation Methodology Earthquake, 2003.
- FEMA, F.: Case Studies-An Assessment of the NEHRP Guidelines for the Seismic Rehabilitation of Buildings., 343 pp., 1999.
- Freeman, S. A., Gilmartin, U. M., and Searer, G. R.: Using strong motion recordings to construct pushover curves, 263–268, 1999.
- Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine, 1999.
- Gamba, P., Cavalca, D., Jaiswal, K. S., Huyck, C., and Crowley, H.: The GED4GEM project: development of a Global Exposure Database for the Global Earthquake Model Initiative, *Proc. 15th World Conf. Earthq. Eng.*, 1–13, 2012.
- Del Gaudio, C., De Martino, G., Di Ludovico, M., Manfredi, G., Prota, A., Ricci, P., and Verderame, G. M.: Empirical fragility curves from damage data on RC buildings after the 2009 L’Aquila earthquake, *Bull. Earthq. Eng.*, 15, 1425–1450, <https://doi.org/10.1007/s10518-016-0026-1>, 2017.
- Del Gaudio, C., Di Ludovico, M., Polese, M., Manfredi, G., Prota, A., Ricci, P., and Verderame, G. M.: Seismic fragility for Italian RC buildings based on damage data of the last 50 years, *Springer Netherlands*, 2023–2059 pp., <https://doi.org/10.1007/s10518-019-00762-6>, 2020.
- Del Gaudio, C., Scala, S. A., Ricci, P., and Verderame, G. M.: Evolution of the seismic vulnerability of masonry buildings based on the damage data from L’Aquila 2009 event, *Bull. Earthq. Eng.*, 19, 4435–4470, 2021.
- Ghimire;Subash;, Philippe, G., Adrien, P., Schorlemmer, D., and Eiffel, G.: Testing machine learning models for heuristic building damage assessment applied to the Italian Database of Observed Damage (DaDO), *Nat. Hazards Earth Syst. Sci. Discuss.*, 1–29, <https://doi.org/10.5194/NHESS-2023-7>, 2023.
- Ghimire, S., Guéguen, P., and Astorga, A.: Analysis of the efficiency of intensity measures from real earthquake data recorded in buildings, *Soil Dyn. Earthq. Eng.*, 147, <https://doi.org/10.1016/j.soildyn.2021.106751>, 2021a.
- Ghimire, S., Guéguen, P., Giffard-Roisin, S., and Schorlemmer, D.: Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake, *Earthq. Spectra*, <https://doi.org/10.1177/87552930221106495>, 2022.
- Goulet, J. A., Michel, C., and A. Der, K.: Data-driven post-earthquake rapid structural safety assessment, *Earthq. Eng. Struct. Dyn.*, 549–462, <https://doi.org/10.1002/eqe.2541>, 2015a.
- Goulet, J. A., Michel, C., and Kiureghian, A. Der: Data-driven post-earthquake rapid structural safety assessment, *Earthq. Eng. Struct. Dyn.*, 44, 549–562, <https://doi.org/10.1002/eqe.2541>, 2015b.
- Grünthal, G.: Escala Macro Sísmica Europea EMS - 98, 101 pp., 1998.
- Guðmundsson, A.: Comprehensive damage analysis of buildings affected by the 2008 South Iceland earthquake, *15th World Conf. Earthq. Eng.*, 5, 2012.
- Guéguen, P., Michel, C., and Lecorre, L.: A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: Application to Grenoble (France), *Bull. Earthq. Eng.*, 5, 467–490, <https://doi.org/10.1007/s10518-007-9036-3>, 2007.
- Guéguen, P., Gallipoli, M. R., Navarro, M., Michel, C., Guillier, B., Karakostas, C., Lekidis, V., Mucciarelli, M., Ponzio, F., and Spina, D.: Testing Buildings Using Ambient Vibrations for Earthquake Engineering: a European Review, *Second Eur. Conf. Earthq. Eng. and Seismol.*, 1–12, 2014.
- Guéguen, P., Johnson, P., and Roux, P.: Nonlinear dynamics induced in a structure by seismic and environmental

- loading, *J. Acoust. Soc. Am.*, 140, 582–590, <https://doi.org/10.1121/1.4958990>, 2016.
- Guettiche, A., Guéguen, P., and Mimoune, M.: Seismic vulnerability assessment using association rule learning: application to the city of Constantine, Algeria, *Nat. Hazards*, 86, 1223–1245, <https://doi.org/10.1007/s11069-016-2739-5>, 2017.
- Guillier, B., Chatelain, J. L., Tavera, H., Perfettini, H., Ochoa, A., and Herrera, B.: Establishing Empirical Period Formula for RC Buildings in Lima, Peru: Evidence for the Impact of Both the 1974 Lima Earthquake and the Application of the Peruvian Seismic Code on High-Rise Buildings, *Seismol. Res. Lett.*, 85, 1308–1315, <https://doi.org/10.1785/0220140078>, 2014.
- Hancilar, U., Tuzun, C., Yenidogan, C., and Erdik, M.: ELER software - A new tool for urban earthquake loss assessment, *Nat. Hazards Earth Syst. Sci.*, 10, 2677–2696, <https://doi.org/10.5194/nhess-10-2677-2010>, 2010.
- Hancilar, U., Taucer, F., and Corbane, C.: Empirical fragility assessment after the january 12, 2010 haiti earthquake, *WIT Trans. Inf. Commun. Technol.*, 44, 353–365, <https://doi.org/10.2495/RISK120301>, 2012.
- Hancock, J., Bommer, J. J., and Stafford, P. J.: Numbers of scaled and matched accelerograms required for inelastic dynamic analyses, *Earthq. Eng. Struct. Dyn.*, 14, 1585–1607, 2008.
- Hannewald, P., Michel, C., Lestuzzi, P., Crowley, H., Pinguet, J., and Fäh, D.: Development and validation of simplified mechanics-based capacity curves for scenario-based risk assessment of school buildings in Basel, *Eng. Struct.*, 209, 110290, <https://doi.org/10.1016/j.engstruct.2020.110290>, 2020.
- Harirchian, E., Kumari, V., Jadhav, K., Rasulzade, S., Lahmer, T., and Das, R. R.: A synthesized study based on machine learning approaches for rapid classifying earthquake damage grades to rc buildings, *Appl. Sci.*, 11, <https://doi.org/10.3390/app11167540>, 2021.
- Hastie, T., Tibshirani, R., and Friedman, J.: Springer Series in Statistics The Elements of Statistical Learning, *Math. Intell.*, 27, 83–85, <https://doi.org/10.1007/b94608>, 2009.
- He, C., Huang, Q., Bai, X., Robinson, D. T., Shi, P., Dou, Y., Zhao, B., Yan, J., Zhang, Q., Xu, F., and Daniell, J.: A Global Analysis of the Relationship Between Urbanization and Fatalities in Earthquake-Prone Areas, *Int. J. Disaster Risk Sci.*, 12, 805–820, <https://doi.org/10.1007/s13753-021-00385-z>, 2021.
- Hegde, J. and Rokseth, B.: Applications of machine learning methods for engineering risk assessment – A review, *Saf. Sci.*, 122, 104492, <https://doi.org/10.1016/j.ssci.2019.09.015>, 2020.
- Hwang, S. H. and Lignos, D. G.: Nonmodel-based framework for rapid seismic risk and loss assessment of instrumented steel buildings, *Eng. Struct.*, 156, 417–432, <https://doi.org/10.1016/j.engstruct.2017.11.045>, 2018.
- Iaccarino, A. G., Gueguen, P., Picozzi, M., and Ghimire, S.: Earthquake Early Warning System for Structural Drift Prediction Using Machine Learning and Linear Regressors, *Front. Earth Sci.*, 9, 545, <https://doi.org/10.3389/feart.2021.666444>, 2021.
- Iervolino, I.: Assessing uncertainty in estimation of seismic response for PBEE, *Earthq. Eng. Struct. Dyn.*, 46, 1711–1723, <https://doi.org/10.1002/eqe.2883>, 2017.
- Jaiswal, K. S. and Wald, D. J.: Development of a semi-empirical loss model within the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system, 9th US Natl. 10th Can. Conf. Earthq. Eng. 2010, Incl. Pap. from 4th Int. Tsunami Symp., 5, 3720–3731, 2010.
- Japkowicz, N. and Stephen, S.: The class imbalance problem A systematic study [fulltext.pdf](https://arxiv.org/abs/2002.05202), 6, 429–449, 2002.
- Jayaram, N., Lin, T., and Baker, J. W.: A Computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance, *Earthq. Spectra*, 27, 797–815, <https://doi.org/10.1193/1.3608002>, 2011.
- Jordan, C.: METEOR: Modelling Exposure Through Earth Observation Routines to Aid Sustainable Development, in: *Geophysical Research Abstracts.*, 17990, 2019.
- Kalakonas, P., Silva, V., Mouyiannou, A., and Rao, A.: Exploring the impact of epistemic uncertainty on a regional probabilistic seismic risk assessment model, <https://doi.org/10.1007/s11069-020-04201-7>, 2020.
- Karababa, F. S. and Pomonis, A.: Damage data analysis and vulnerability estimation following the August 14, 2003 Lefkada Island, Greece, *Earthquake, Bull. Earthq. Eng.*, 9, 1015–1046, <https://doi.org/10.1007/s10518-010-9231-5>, 2011.
- Karapetrou, S., Manakou, M., Bindi, D., Petrovic, B., and Pitilakis, K.: “Time-building specific” seismic vulnerability assessment of a hospital RC building using field monitoring data, *Eng. Struct.*, 112, 114–132, <https://doi.org/10.1016/j.engstruct.2016.01.009>, 2016.
- Karmenova, M., Nugumanova, A., Tlebalidina, A., Beldeubaev, A., Popova, G., and Sedchenko, A.: Seismic Assessment of Urban Buildings Using Data Mining Methods, 154–159, <https://doi.org/10.1145/3397125.3397152>, 2020.
- Kashima, T.: Dynamic Behavior of an Eight-Storey SRC Building Examined from Strong Motion Records, in: Paper presented at the 13th World Conference on Earthquake Engineering., 2004.
- Kashima, T.: Dynamic behaviour of SRC buildings damaged by the 2011 great east japan earthquake based on strong motion records, in: In Second European Conference on Earthquake Engineering and Seismology., 1–

11, 2014.

- Kassem, M. M., Mohamed Nazri, F., and Noroozinejad Farsangi, E.: The seismic vulnerability assessment methodologies: A state-of-the-art review, <https://doi.org/10.1016/j.asej.2020.04.001>, 1 December 2020.
- Kazantzi, A. K. and Vamvatsikos, D.: Intensity measure selection for vulnerability studies of building classes, *Earthq. Eng. Struct. Dyn.*, 44, 2677–2694, <https://doi.org/10.1002/eqe.2603>, 2015.
- Kim, T., Song, J., and Kwon, O. S.: Pre- and post-earthquake regional loss assessment using deep learning, *Earthq. Eng. Struct. Dyn.*, 49, 657–678, <https://doi.org/10.1002/eqe.3258>, 2020.
- Kohrangi, M., Bazzurro, P., and Vamvatsikos, D.: Vector and scalar IMs in structural response estimation, Part II: Building demand assessment, *Earthq. Spectra*, 32, 1525–1543, <https://doi.org/10.1193/053115EQS081M>, 2016.
- Kohrangi, M., Bazzurro, P., Vamvatsikos, D., and Spillatura, A.: Conditional spectrum-based ground motion record selection using average spectral acceleration, *Earthq. Eng. Struct. Dyn.*, 46(10), 16, 2017.
- Lagomarsino, S. and Giovinazzi, S.: Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings, *Bull. Earthq. Eng.*, 4, 415–443, <https://doi.org/10.1007/s10518-006-9024-z>, 2006.
- Lagomarsino, S., Cattari, S., and Ottonelli, D.: The heuristic vulnerability model: fragility curves for masonry buildings, Springer Netherlands, 3129–3163 pp., <https://doi.org/10.1007/s10518-021-01063-7>, 2021.
- Lallemant, D., Kiremidjian, A., and Burton, H.: Statistical procedures for developing earthquake damage fragility curves, *Earthq. Eng. Struct. Dyn.*, 44, 1373–1389, <https://doi.org/10.1002/EQE.2522>, 2015.
- Lestuzzi, P., Podestà, S., Luchini, C., Garofano, A., Kazantzidou-Firtinidou, D., Bozzano, C., Bischof, P., Haffter, A., and Rouiller, J. D.: Seismic vulnerability assessment at urban scale for two typical Swiss cities using Risk-UE methodology, *Nat. Hazards*, 84, 249–269, <https://doi.org/10.1007/s11069-016-2420-z>, 2016.
- Lin, L., Naumoski, N., Saatcioglu, M., and Foo, S.: Improved intensity measures for probabilistic seismic demand analysis. part 1: Development of improved intensity measures, *Can. J. Civ. Eng.*, 38, 79–88, <https://doi.org/10.1139/L10-110>, 2011.
- Loos, S., Lallemant, D., Baker, J., McCaughey, J., Yun, S. H., Budhathoki, N., Khan, F., and Singh, R.: G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage, *Earthq. Spectra*, 36, 1695–1718, <https://doi.org/10.1177/8755293020926190>, 2020.
- Luco, N.: Probabilistic seismic demand analysis, SMRF connection fractures, and near-source effects, *Dep. Civ. Environ.*, 260 pp., 2002.
- Luco, N. and Cornell, C. A.: Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions, *Earthq. Spectra*, 23, 357–392, 2007.
- Luco, N., Manuel, L., Baldava, S., and Bazzurro, P.: Correlation of damage of steel moment-resisting frames to a vector-valued set of ground motion parameters, *Proc. 9th Int. Conf. Struct. Saf. Reliab.*, 2719–2726, 2005.
- Luna, B. N. and Irfanoglu, A.: Estimating base shear versus roof drift curves using earthquake-response data, 9th US Natl. 10th Can. Conf. Earthq. Eng. 2010, Incl. Pap. from 4th Int. Tsunami Symp., 2, 890–899, 2010.
- Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: 31st Conference on Neural Information Processing Systems, 2017.
- Maheri, M. R., Naeim, F., and Mehrain, M.: Performance of adobe residential buildings in the 2003 Bam, Iran, earthquake, *Earthq. Spectra*, 21, 337–344, <https://doi.org/10.1193/1.2098861>, 2005.
- Mak, S. and Schorlemmer, D.: A comparison between the forecast by the United States national seismic hazard maps with recent ground-motion records, *Bull. Seismol. Soc. Am.*, 106, 1817–1831, <https://doi.org/10.1785/0120150323>, 2016.
- Mangalathu, S. and Jeon, J.-S.: Regional Seismic Risk Assessment of Infrastructure Systems through Machine Learning: Active Learning Approach, *J. Struct. Eng.*, 146, 04020269, [https://doi.org/10.1061/\(asce\)st.1943-541x.0002831](https://doi.org/10.1061/(asce)st.1943-541x.0002831), 2020.
- Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., and Burton, H. V.: Classifying earthquake damage to buildings using machine learning, *Earthq. Spectra*, 36, 183–208, <https://doi.org/10.1177/8755293019878137>, 2020a.
- Mangalathu, S., Jang, H., Hwang, S. H., and Jeon, J. S.: Data-driven machine-learning-based seismic failure mode identification of reinforced concrete shear walls, *Eng. Struct.*, 208, 110331, <https://doi.org/10.1016/j.engstruct.2020.110331>, 2020b.
- Martin, S. S., Hough, S. E., and Hung, C.: Ground motions from the 2015 Mw 7.8 Gorkha, Nepal, Earthquake constrained by a detailed assessment of macroseismic data, *Seismol. Res. Lett.*, 86, 1524–1532, <https://doi.org/10.1785/0220150138>, 2015.
- Martins, L. and Silva, V.: Development of a fragility and vulnerability model for global seismic risk analyses, *Bull. Earthq. Eng.*, 19, 6719–6745, <https://doi.org/10.1007/s10518-020-00885-1>, 2021.
- Masi, A. and Vona, M.: Experimental and numerical evaluation of the fundamental period of undamaged and damaged RC framed buildings, *Bull. Earthq. Eng.*, 8, 643–656, <https://doi.org/10.1007/s10518-009-9136-3>, 2010.
- Mazzoni, S. ., McKenna, F. ., Scott, H. M. ., Fenves, L. G., and Al., E.: OpenSees Command Language Manual,

- Pacific Earthq. Eng. Res. Cent., 246 (1), 137–158, <https://doi.org/10.1080/15732479.2021.1876106>, 2006.
- Michel, C. and Guéguen, P.: Time-frequency analysis of small frequency variations in civil engineering structures under weak and strong motions using a reassignment method, *Struct. Heal. Monit.*, 9, 159–171, <https://doi.org/10.1177/1475921709352146>, 2010.
- Michel, C. and Guéguen, P.: Seismic vulnerability assessment based on vibration data at the city-scale, 15th World Conf. Earthq. Eng., 2–9, 2012.
- Michel, C., Guéguen, P., Lestuzzi, P., and Bard, P. Y.: Comparison between seismic vulnerability models and experimental dynamic properties of existing buildings in France, *Bull. Earthq. Eng.*, 8, 1295–1307, <https://doi.org/10.1007/s10518-010-9185-7>, 2010.
- Michel, C., Zapico, B., Lestuzzi, P., Molina, F. J., and Weber, F.: Quantification of fundamental frequency drop for unreinforced masonry buildings from dynamic tests, *Earthq. Eng. Struct. Dyn.*, 40, 1283–1296, <https://doi.org/DOI:10.1002/eqe.1088> Quantification, 2011.
- Michel, C., Guéguen, P., and Causse, M.: Seismic vulnerability assessment to slight damage based on experimental modal parameters, *Earthq. Eng. Struct. Dyn.*, 41, 81–98, <https://doi.org/DOI:10.1002/eqe.1119> Seismic, 2012.
- Milutinovic, Z. and Trendafiloski, G.: Risk-UE An advanced approach to earthquake risk scenarios with applications to different European towns, Rep. to WP4 vulnerability Curr. Build., 1–83, 2003.
- Ministry of Housing and Urbanism of Chile, Terremoto y Tsunami 27F 2010:
- Mollaioli, F., Silvia, B., Luis, D., and Rodolfo, S.: Correlations between energy and displacement demands for performance-based seismic engineering, *Pure Appl. Geophys.*, 168, 237–259, <https://doi.org/10.1007/s00024-010-0118-9>, 2011.
- Morfidis, K. and Kostinakis, K.: Approaches to the rapid seismic damage prediction of r/c buildings using artificial neural networks, *Eng. Struct.*, 165, 120–141, <https://doi.org/10.1016/j.engstruct.2018.03.028>, 2018.
- Mouroux, P. and Le Brun, B.: Presentation of RISK-UE Project, *Bull. Earthq. Eng.* 2006 44, 4, 323–339, <https://doi.org/10.1007/S10518-006-9020-3>, 2006.
- Ministere des Travaux Publics, Transports et Communications: Evaluation des Bâtiments: [https://www.mtpc.gouv.ht/accueil/recherche/article\\_7.html](https://www.mtpc.gouv.ht/accueil/recherche/article_7.html).
- Mucciarelli, M., Masi, A., Gallipoli, M. R., Harabaglia, P., Vona, M., Ponzio, F., and Dolce, M.: Analysis of RC building dynamic response and soil-building resonance based on data recorded during a damaging earthquake (Molise, Italy, 2002), *Bull. Seismol. Soc. Am.*, 94, 1943–1953, <https://doi.org/10.1785/012003186>, 2004.
- Nievas, C. I., Crowley, H., Reuland, Y., Weatherill, G., Baltzopoulos, G., Bayliss, K., Chatzi, E., Chioccarelli, E., Guéguen, P., Iervolino, I., Marzocchi, W., Naylor, M., Orlacchio, M., Pejovic, J., Popovic, N., Serafini, F., and Serdar, N.: Deliverable 6.1 D6.1 Integration of RISE Innovations in the Fields of OELF, RLA and SHM Deliverable information, <https://doi.org/https://doi.org/10.5281/zenodo.7785133>, 2023.
- NPA: Police Countermeasures and Damage Situation associated with 2011 Tohoku district - off the Pacific Ocean Earthquake Total burn down Inundated below floor level Partially damaged Property damages Damaged roads Partial burn down March 10, 2021, [https://doi.org/https://www.npa.go.jp/news/other/earthquake2011/pdf/higaijokyo\\_e.pdf](https://doi.org/https://www.npa.go.jp/news/other/earthquake2011/pdf/higaijokyo_e.pdf) (last access: 22 March 2021), 2021.
- 2015 Nepal Earthquake: Open Data Portal: /eq2015.npc.gov.np/#/, last access: 22 March 2021.
- NPC: Post Disaster Needs Assessment, 2015.
- OECD: Financial Management of Earthquake Risk, 108, 2018.
- Omoya, M., Ero, I., Zaker Esteghamati, M., Burton, H. V., Brandenburg, S., Sun, H., Yi, Z., Kang, H., and Nweke, C. C.: A relational database to support post-earthquake building damage and recovery assessment, <https://doi.org/10.1177/87552930211061167>, 38, 1549–1569, <https://doi.org/10.1177/87552930211061167>, 2022.
- Ostrovsky, L. A. and Johnson, P. A.: Dynamic nonlinear elasticity in geomaterials, *Riv. DEL NUOVO Cim.*, 24, 2001.
- Ozer, E., Özcebe, A. G., Negulescu, C., Kharazian, A., Borzi, B., Bozzoni, F., Molina, S., Peloso, S., and Tubaldi, E.: Vibration-Based and Near Real-Time Seismic Damage Assessment Adaptive to Building Knowledge Level, *Buildings*, 12, 416, <https://doi.org/10.3390/buildings12040416>, 2022.
- Pan, H., Kusunoki, K., and Hattori, Y.: Capacity-curve-based damage evaluation approach for reinforced concrete buildings using seismic response data, *Eng. Struct.*, 197, 109386, <https://doi.org/10.1016/j.engstruct.2019.109386>, 2019.
- Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., and Mueller, A.: Scikit-learn, *GetMobile Mob. Comput. Commun.*, 19, 29–33, <https://doi.org/10.1145/2786984.2786995>, 2011.
- Perrault, M. and Guéguen, P.: Correlation between ground motion and building response using California earthquake records, *Earthq. Spectra*, 31, 2027–2046, <https://doi.org/10.1193/062413EQS168M>, 2015.

- Perrault, M., Gueguen, P., Aldea, A., and Demetriu, S.: Using experimental data to reduce the single-building sigma of fragility curves: Case study of the BRD tower in Bucharest, Romania, *Earthq. Eng. Eng. Vib.*, 12, 643–658, <https://doi.org/10.1007/s11803-013-0203-z>, 2013.
- Perrault, M., Guéguen, P., Parra, G., and Sarango, J.: Modification of the data-driven period/height relationship for buildings located in seismic-prone regions such as Quito (Ecuador), *Bull. Earthq. Eng.*, 18, 3545–3562, <https://doi.org/10.1007/s10518-020-00840-0>, 2020.
- Pittore, M. and Wieland, M.: Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing, *Nat. Hazards*, 68, 115–145, <https://doi.org/10.1007/s11069-012-0475-z>, 2013.
- Podili, B. and Raghukanth, S. T. G.: Ground motion prediction equations for higher order parameters, *Soil Dyn. Earthq. Eng.*, 118, 98–110, <https://doi.org/10.1016/j.soildyn.2018.11.027>, 2019.
- Porter, K.: An overview of PEER’s performance-based earthquake engineering methodology., in: *Proceedings of ninth international conference on applications of statistics and probability in civil engineering (ICASP9)*, San Francisco (US), 2003.
- Porter, K., Mitrani-Reiser, J., and Beck, J. L.: Near-real-time loss estimation for instrumented buildings, *Struct. Des. Tall Spec. Build.*, 15, 3–20, <https://doi.org/10.1002/tal.340>, 2006.
- Pothon, A., Gueguen, P., Buisine, S., and Bard, P.: Comparing Probabilistic Seismic Hazard Maps with ShakeMap Footprints for Indonesia, *Seismol. Res. Lett.*, <https://doi.org/10.1785/0220190171>, 2020.
- Ranf, R. T., Eberhard, M. O., and Malone, S.: Post earthquake prioritization of bridge inspections, *Earthq. Spectra*, 23, 131–146, <https://doi.org/10.1193/1.2428313>, 2007.
- Rao, A., Jung, J., Silva, V., Molinario, G., and Yun, S.-H.: Earthquake Building Damage Detection based on Synthetic Aperture Radar Imagery and Machine Learning, *Nat. Hazards Earth Syst. Sci.*, <https://doi.org/10.5194/nhess-2022-125>, 2022.
- Régnier, J., Michel, C., Bertrand, E., and Guéguen, P.: Contribution of ambient vibration recordings (free-field and buildings) for post-seismic analysis: The case of the Mw 7.3 Martinique (French Lesser Antilles) earthquake, November 29, 2007, *Soil Dyn. Earthq. Eng.*, 50, 162–167, <https://doi.org/10.1016/j.soildyn.2013.03.007>, 2013.
- Reuland, Y., Lestuzzi, P., and Smith, I. F. C.: Measurement-based support for post-earthquake assessment of buildings, *Struct. Infrastruct. Eng.*, 15, 647–662, <https://doi.org/10.1080/15732479.2019.1569071>, 2019.
- Riedel, I. and Guéguen, P.: Modeling of damage-related earthquake losses in a moderate seismic-prone country and cost–benefit evaluation of retrofit investments: application to France, *Nat. Hazards*, 90, 639–662, <https://doi.org/10.1007/s11069-017-3061-6>, 2018.
- Riedel, I., Guéguen, P., Dunand, F., and Cottaz, S.: Macroscale vulnerability assessment of cities using association rule learning, *Seismol. Res. Lett.*, 85, 295–305, <https://doi.org/10.1785/0220130148>, 2014.
- Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., and Chanussot, J.: Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine methods, *Nat. Hazards*, 76, 1111–1141, <https://doi.org/10.1007/s11069-014-1538-0>, 2015.
- Roca, A., Goula, X., Susagna, T., Ch’avez, J., Ch’avez, C., Gonz’alez, M., Gonz’alez, G., and Reinoso, E.: A Simplified Method for Vulnerability Assessment of Dwelling Buildings and Estimation of Damage Scenarios in Catalonia, Spain, <https://doi.org/10.1007/s10518-006-9003-4>, 2006.
- Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., Ribeiro, L., Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., Friedman, J. H., Mangalathu, S., Jong-Su, J., Jeon, J. S., Sun, H., Nweke, C. C., Yi, Z., Burton, H. V., Saad, S. E., Yang, J., Sebastian, R., Vahid, M., Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., and Mueller, A.: Classifying earthquake damage to buildings using machine learning, *GetMobile Mob. Comput. Commun.*, 476–477, 1–12, <https://doi.org/10.1177/8755293019878137>, 2019.
- Roeslin, S., Ma, Q. T. M., and García, H. J.: Damage assessment on buildings following the 19th september 2017 Puebla, Mexico earthquake, *Front. Built Environ.*, 4, 1–18, <https://doi.org/10.3389/fbuil.2018.00072>, 2018.
- Roeslin, S., Ma, Q., Juárez-García, H., Gómez-Bernal, A., Wicker, J., and Wotherspoon, L.: A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake, *Earthq. Spectra*, 36, 314–339, <https://doi.org/10.1177/8755293020936714>, 2020.
- Rossetto, T., Ioannou, I., Grant, D., and Maqsood, T.: Guidelines for the empirical vulnerability assessment, *GEM Tech. Rep.*, 08, 140, <https://doi.org/10.13117/GEM.VULN-MOD.TR2014.11>, 2014.
- Sajedi, S. O. and Liang, X.: A data-driven framework for near real-time and robust damage diagnosis of building structures, *Struct. Control Heal. Monit.*, 27, 1–22, <https://doi.org/10.1002/stc.2488>, 2020.
- Salehi, H. and Burgueño, R.: Emerging artificial intelligence methods in structural engineering, *Eng. Struct.*, 171, 170–189, <https://doi.org/10.1016/j.engstruct.2018.05.084>, 2018.
- Sayin, E., Yön, B., Calayir, Y., and Karaton, M.: Failures of masonry and adobe buildings during the June 23, 2011 Maden-(Elaziğ) earthquake in Turkey, *Eng. Fail. Anal.*, 34, 779–791, <https://doi.org/10.1016/j.engfailanal.2012.10.016>, 2013.

- Scala, S. A., Del Gaudio, C., and Verderame, G. M.: Influence of construction age on seismic vulnerability of masonry buildings damaged after 2009 L'Aquila earthquake, *Soil Dyn. Earthq. Eng.*, 157, 107199, <https://doi.org/10.1016/J.SOILDYN.2022.107199>, 2022.
- Schorlemmer, D., Beutin, T., Hirata, N., Hao, K., Wyss, M., Cotton, F., and Prehn, K.: Global Dynamic Exposure and the OpenBuildingMap - Communicating Risk and Involving Communities, *EGU Gen. Assem. Conf. Abstr.*, 19, 7060, 2017.
- Schorlemmer, D., Beutin, T., Cotton, F., Ospina, N. G., Hirata, N., Ma, K.-F., Nievas, C., Prehn, K., and Wyss, M.: Global Dynamic Exposure and the OpenBuildingMap - A Big-Data and Crowd-Sourcing Approach to Exposure Modeling, *EGU2020*, <https://doi.org/10.5194/EGUSPHERE-EGU2020-18920>, 2020.
- Seo, J., Dueñas-Osorio, L., Craig, J. I., and Goodno, B. J.: Metamodel-based regional vulnerability estimate of irregular steel moment-frame structures subjected to earthquake events, *Eng. Struct.*, 45, 585–597, <https://doi.org/10.1016/j.engstruct.2012.07.003>, 2012.
- Shome, N. and Cornell, C. A.: *Probabilistic Seismic Demand Analysis of Nonlinear Structures*, 1999.
- Silva, V. and Horspool, N.: Combining USGS ShakeMaps and the OpenQuake-engine for damage and loss assessment, *Earthq. Eng. Struct. Dyn.*, 48, 634–652, <https://doi.org/10.1002/eqe.3154>, 2019.
- Silva, V., Crowley, H., Pagani, M., Monelli, D., and Pinho, R.: Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment, *Nat. Hazards*, 72, 1409–1427, <https://doi.org/10.1007/s11069-013-0618-x>, 2014.
- Silva, V., Crowley, H., Jaiswal, K., Acevedo, A. B., Pittore, M., and Journey, M.: Developing a Global Earthquake Risk Model, *16th Eur. Conf. Earthq. Eng.*, 1–11, 2018.
- Silva, V., Pagani, M., Schneider, J., and Henshaw, P.: Assessing Seismic Hazard and Risk Globally for an Earthquake Resilient World, *Contrib. Pap. to GAR 2019*, 24 p., 2019.
- Silva, V., Brzev, S., Scawthorn, C., Yepes, C., Dabbeek, J., and Crowley, H.: A Building Classification System for Multi-hazard Risk Assessment, *Int. J. Disaster Risk Sci.*, 13, 161–177, <https://doi.org/10.1007/s13753-022-00400-x>, 2022.
- So, E.: Data and its role in reducing the risk of disasters in the built environment, *Nat. Hazards*, <https://doi.org/10.1007/s11069-022-05590-7>, 2022.
- Spence, R., Bommer, J., Del Re, D., Bird, J., Aydinoglu, N., and Tabuchi, S.: Comparing loss estimation with observed damage: A study of the 1999 Kocaeli earthquake in Turkey, *Bull. Earthq. Eng.*, 1, 83–113, <https://doi.org/10.1023/A:1024857427292>, 2003.
- Spence, R. J. S., Coburn, A. W., and Ruffle, S. J.: Estimation of vulnerability functions based on a global earthquake damage database, *Assembly*, 11, 12387, 2009.
- Stojadinović, Z., Kovačević, M., Marinković, D., and Stojadinović, B.: Rapid earthquake loss assessment based on machine learning and representative sampling, *Earthq. Spectra*, 38, 152–177, <https://doi.org/10.1177/87552930211042393>, 2021.
- Stone, H., Putrino, V., and D'Ayala, D.: Earthquake damage data collection using omnidirectional imagery, *Front. Built Environ.*, 4, 51, <https://doi.org/10.3389/FBUIL.2018.00051/BIBTEX>, 2018.
- Sun, H.: A Data-driven Building Seismic Response Prediction Framework: from Simulation and Recordings to Statistical Learning, 1–212, 2019.
- Sun, H., Burton, H. V., and Huang, H.: Machine learning applications for building structural design and performance assessment: State-of-the-art review, *J. Build. Eng.*, 33, 101816, <https://doi.org/10.1016/j.jobe.2020.101816>, 2021.
- Todorovska, M. I. and Trifunac, M. D.: Earthquake damage detection in the Imperial County Services Building I: The data and time-frequency analysis, *Soil Dyn. Earthq. Eng.*, 27, 564–576, <https://doi.org/10.1016/j.soildyn.2006.10.005>, 2007.
- Trevlopoulos, K. and Guéguen, P.: Period elongation-based framework for operative assessment of the variation of seismic vulnerability of reinforced concrete buildings during aftershock sequences, *Soil Dyn. Earthq. Eng.*, 84, 224–237, <https://doi.org/10.1016/j.soildyn.2016.02.009>, 2016.
- Trevlopoulos, K., Guéguen, P., Helmstetter, A., and Cotton, F.: Earthquake risk in reinforced concrete buildings during aftershock sequences based on period elongation and operational earthquake forecasting, *Struct. Saf.*, 84, 101922, <https://doi.org/10.1016/j.strusafe.2020.101922>, 2020.
- Trifunac, M. D., Todorovska, M. I., Manić, M. I., and Bulajić, B. Đ.: Variability of the fixed-base and soil–structure system frequencies of a building—The case of Borik-2 building, *Struct. Control Heal. Monit. Off. J. Int. Assoc. Struct. Control Monit. Eur. Assoc. Control Struct.*, 2, 120–15, <https://doi.org/10.1002/stc>, 2010.
- Tubaldi, E., Ozer, E., Douglas, J., and Gehl, P.: Examining the contribution of near real-time data for rapid seismic loss assessment of structures, *Struct. Heal. Monit.*, 21, 118–137, <https://doi.org/10.1177/1475921721996218>, 2022.
- Uma, S. R.: Seismic Instrumentation of Buildings - A Promising Step for Performance Based Design in New Zealand, *New Zeal. Soc. Earthq. Eng. Conf.*, 1–9, 2007.
- UN: United Nations, Department of Economic and Social Affairs, Population Division (2019). World

- Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420), 197–236 pp., 2018.
- Vamvatsikos, D. and Cornell, C. A.: Incremental dynamic analysis, *Earthq. Eng. Struct. Dyn.*, 31, 491–514, <https://doi.org/10.1002/eqe.141>, 2002.
- Vamvatsikos, D. and Cornell, C. A.: Applied incremental dynamic analysis, *Earthq. Spectra*, 20, 523–553, <https://doi.org/10.1193/1.1737737>, 2004.
- Vamvatsikos, D. and Cornell, C. A.: Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information, *Earthq. Eng. Struct. Dyn.*, 34, 1573–1600, <https://doi.org/10.1002/eqe.496>, 2005.
- Vidal, F., Navarro, M., Aranda, C., and Enomoto, T.: Changes in dynamic characteristics of Lorca RC buildings from pre- and post-earthquake ambient vibration data, *Bull. Earthq. Eng.*, 12, 2095–2110, <https://doi.org/10.1007/s10518-013-9489-5>, 2014a.
- Vidal, F., Navarro, M., Aranda, C., and Enomoto, T.: Changes in dynamic characteristics of Lorca RC buildings from pre-and post-earthquake ambient vibration data, *Bull. Earthq. Eng.*, 12, 2095–2110, <https://doi.org/10.1007/s10518-013-9489-5>, 2014b.
- Wald, D. J., Worden, B. C., Quitariano, V., and Pankow, K. L.: ShakeMap manual: technical manual, user's guide, and software guide, *Techniques and Methods*, <https://doi.org/10.3133/tm12A1>, 2005.
- Wang, C., Yu, Q., Law, K. H., McKenna, F., Yu, S. X., Taciroglu, E., Zsarnóczay, A., Elhaddad, W., and Cetiner, B.: Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management, *Autom. Constr.*, 122, <https://doi.org/10.1016/j.autcon.2020.103474>, 2021.
- Webster, F. A. and Tolles, E. L.: Earthquake Damage To Historic and Older Adobe Buildings During the 1994 Northridge , California Earthquake, Program, 1, 1–8, 1999.
- Wieland, M., Pittore, M., Parolai, S., Begaliev, U., Yasunov, P., Tyagunov, S., Moldobekov, B., Saidiy, S., Ilyasov, I., and Abakanov, T.: A multiscale exposure model for seismic risk assessment in Central Asia, *Seismol. Res. Lett.*, 86, 210–222, <https://doi.org/10.1785/0220140130>, 2015.
- Wiemer, S.: Real-time earthquake risk reduction for a Resilient Europe ( RISE ), 2018.
- Xie, Y., Ebad Sichani, M., Padgett, J. E., and DesRoches, R.: The promise of implementing machine learning in earthquake engineering: A state-of-the-art review, *Earthq. Spectra*, 36, 1769–1801, <https://doi.org/10.1177/8755293020919419>, 2020.
- Xu, Y., Lu, X., Cetiner, B., and Taciroglu, E.: Real-time regional seismic damage assessment framework based on long short-term memory neural network, *Comput. Civ. Infrastruct. Eng.*, 1–18, <https://doi.org/10.1111/mice.12628>, 2020a.
- Xu, Z., Wu, Y., Qi, M. zhu, Zheng, M., Xiong, C., and Lu, X.: Prediction of structural type for city-scale seismic damage simulation based on machine learning, *Appl. Sci.*, 10, <https://doi.org/10.3390/app10051795>, 2020b.
- Zhang, Y. and Burton, H. V.: Pattern recognition approach to assess the residual structural capacity of damaged tall buildings, *Struct. Saf.*, 78, 12–22, <https://doi.org/10.1016/j.strusafe.2018.12.004>, 2019.
- Zhang, Y., Burton, H. V., Sun, H., and Shokrabadi, M.: A machine learning framework for assessing post-earthquake structural safety, *Struct. Saf.*, 72, 1–16, <https://doi.org/10.1016/j.strusafe.2017.12.001>, 2018.
- Zhao, J. X., Jiang, F., Shi, P., Xing, H., Huang, H., Hou, R., Zhang, Y., Yu, P., Lan, X., Rhoades, D. A., Somerville, P. G., Irikura, K., and Fukushima, Y.: Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions, *Bull. Seismol. Soc. Am.*, 106, 1535–1551, <https://doi.org/10.1785/0120150056>, 2016.
- Zhao, X., Lovreglio, R., and Nilsson, D.: Modelling and interpreting pre-evacuation decision-making using machine learning, *Autom. Constr.*, 113, 103140, <https://doi.org/10.1016/j.autcon.2020.103140>, 2020.

## Appendix

**Table A.** Distribution of precision and recall observed for different machine learning models developed in this study. RFC represents random forest classification, GBC represents gradient boosting classification, LR represents linear regression, SVR represents support vector regression, GBR represents gradient boosting regression, RFR represents random forest regression, US represents undersampling, OS represents oversampling, FFS represents full-feature-setting, BFS represents basic-features-setting, Re represents recall, and Pr represent precision (ratio of correctly predicted instances over the total instances predicted in a particular class).

Dataset	Method	Score	5DG					TLS			Overall accuracy		
			DG1	DG2	DG3	DG4	DG5	S	M	H	5DG	TLS	
Validation	RFC	Re	0.65	0.24	0.31	0.43	0.78	-	-	-	0.53	-	
		Pr	0.59	0.38	0.40	0.46	0.64	-	-	-	-	-	
	GBC	Re	0.67	0.33	0.39	0.49	0.75	-	-	-	0.57	-	
		Pr	0.61	0.40	0.42	0.50	0.71	-	-	-	-	-	
	LR	Re	0.12	0.25	0.24	0.74	0.09	-	-	-	0.30	-	
		Pr	0.60	0.27	0.24	0.28	0.59	-	-	-	-	-	
	SVR	Re	0.44	0.23	0.20	0.56	0.52	-	-	-	0.43	-	
		Pr	0.64	0.31	0.30	0.32	0.63	-	-	-	-	-	
	GBR	Re	0.37	0.34	0.42	0.65	0.56	-	-	-	0.51	-	
		Pr	0.79	0.34	0.36	0.40	0.83	-	-	-	-	-	
	RFR	Re	0.33	0.35	0.39	0.65	0.54	-	-	-	0.49	-	
		Pr	0.80	0.33	0.36	0.39	0.81	-	-	-	-	-	
	RFR	US	Re	0.45	0.42	0.52	0.55	0.35	-	-	-	0.45	-
			Pr	0.73	0.32	0.33	0.37	0.88	-	-	-	-	-
		OS	Re	0.41	0.40	0.43	0.59	0.51	-	-	-	0.49	-
			Pr	0.75	0.33	0.35	0.39	0.82	-	-	-	-	-
Test	RFR	FFS	Re	0.42	0.40	0.43	0.60	0.51	0.59	0.60	0.73	0.49	0.68
			Pr	0.75	0.34	0.35	0.39	0.82	0.77	0.32	0.88	-	-
		BFS	Re	0.32	0.36	0.45	0.56	0.46	0.51	0.61	0.70	0.46	0.64
			Pr	0.71	0.30	0.32	0.38	0.82	0.73	0.30	0.88	-	-

**Table B.**  $\sigma_{IM}$ ,  $\sigma_{EDP}$ , and  $\sigma_{EDP|IM}$  values considering different sub datasets discussed in the manuscript

			PGA	PGV	PGD	AI	DP	CAV	SA <sub>i</sub>	SV <sub>i</sub>	SD <sub>i</sub>	SA <sub>min</sub>	SV <sub>min</sub>	SD <sub>min</sub>	
US+JPN +RO	ALL	IM	1.15	1.33	1.63	2.29	3.35	1.21	1.53	1.31	1.43	1.63	1.44	1.61	
		EDP	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48
		EDP IM	1.05	0.80	0.81	0.84	0.91	0.85	0.87	0.79	0.87	0.87	0.87	0.79	0.88
US data	ALL	IM	0.90	1.07	1.49	1.73	2.73	0.96	1.03	1.04	1.35	1.09	1.08	1.54	
		EDP	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32
		EDP IM	1.19	1.00	0.98	1.01	1.07	0.97	1.14	1.03	1.08	1.18	1.01	1.08	
	STS1	IM	0.86	0.94	1.15	1.48	2.20	0.77	1.00	1.06	1.20	1.03	1.08	1.31	
		EDP	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24
		EDP IM	1.14	1.00	0.93	1.00	1.07	0.93	1.15	1.03	1.03	1.15	1.01	1.03	
	STS2	IM	0.74	0.85	1.15	1.49	2.29	0.84	0.92	0.95	1.20	0.95	0.92	1.29	
		EDP	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
		EDP IM	1.12	1.01	0.99	1.00	1.06	0.98	0.97	1.01	1.10	1.04	1.01	1.10	
	RC	IM	0.84	0.98	1.47	1.58	2.63	0.91	1.03	1.03	1.25	1.07	1.04	1.31	
		EDP	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32	1.32
		EDP IM	1.21	0.94	0.94	0.99	1.03	0.93	1.20	1.01	0.91	1.22	0.99	0.90	
ST	IM	0.84	1.01	1.46	1.55	2.78	0.87	0.91	1.02	1.32	0.99	1.08	1.64		
	EDP	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.29	1.28	1.28	1.29	1.28	
	EDP IM	1.20	0.93	0.91	0.97	1.01	0.89	1.13	1.02	0.99	1.19	0.98	0.99		
MA	IM	0.91	1.08	1.43	1.83	2.35	1.02	0.94	1.27	1.71	1.00	1.18	1.46		
	EDP	1.23	1.23	1.23	1.23	1.23	1.23	1.23	1.21	1.21	1.23	1.23	1.23		
	EDP IM	1.11	1.07	1.01	1.02	1.10	1.00	1.11	1.12	1.15	1.14	1.12	1.14		
WO	IM	1.02	1.36	1.83	2.20	3.21	1.21	1.07	1.32	1.60	1.06	1.32	1.65		
	EDP	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02		
	EDP IM	0.66	0.71	0.81	0.71	0.77	0.75	0.67	0.69	0.72	0.70	0.68	0.70		
Japanese	ALL	IM	1.11	1.31	1.61	2.26	3.34	1.22	1.50	1.30	1.42	1.61	1.43	1.61	
		EDP	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	
		EDP IM	1.02	0.78	0.78	0.81	0.85	0.81	0.83	0.76	0.82	0.83	0.76	0.83	
	RC	IM	0.98	1.13	1.50	2.02	2.98	1.12	1.25	1.09	1.22	1.37	1.20	1.35	
		EDP	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	1.28	
		EDP IM	0.98	0.86	0.92	0.84	0.93	0.85	0.79	0.86	0.95	0.73	0.82	0.98	
	SRC	IM	1.14	1.40	1.66	2.36	3.53	1.24	1.48	1.37	1.46	1.56	1.52	1.68	
		EDP	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	
		EDP IM	0.87	0.64	0.67	0.66	0.74	0.72	0.65	0.61	0.61	0.62	0.60	0.64	
	ST	IM	1.11	1.24	1.61	2.18	3.24	1.21	1.72	1.37	1.56	1.65	1.43	1.70	
		EDP	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	
		EDP IM	1.36	0.76	0.46	0.87	0.67	0.65	0.80	0.70	0.78	0.59	0.50	0.48	
ALL	IM	1.00	1.27	1.51	2.12	3.25	1.13	1.54	1.31	1.42	1.44	1.41	1.59		
	EDP	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42	1.42		
	EDP IM	0.80	0.50	0.46	0.51	0.57	0.51	0.78	0.48	0.48	0.46	0.40	0.39		
ANX building	STS	IM	1.01	1.28	1.50	2.13	3.25	1.13	1.55	1.32	1.43	1.45	1.42	1.60	
		EDP	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	
		EDP IM	0.80	0.50	0.45	0.50	0.56	0.50	0.78	0.48	0.47	0.45	0.39	0.39	
MR1	IM	0.69	0.80	0.90	1.30	2.10	0.65	1.06	0.85	0.91	0.89	0.88	0.97		
	EDP	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86		
	EDP IM	0.59	0.42	0.38	0.46	0.52	0.46	0.62	0.40	0.37	0.39	0.34	0.33		
MR2	IM	0.71	0.75	0.78	1.16	1.82	0.57	1.21	0.79	0.89	0.89	0.79	0.86		

	EDP	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
	EDP IM	0.55	0.41	0.35	0.41	0.41	0.48	0.61	0.41	0.42	0.37	0.37	0.29
MR3	IM	0.79	0.86	0.96	1.47	2.17	0.73	1.10	0.85	0.93	0.88	0.92	1.03
	EDP	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
	EDP IM	0.57	0.46	0.52	0.47	0.59	0.47	0.75	0.47	0.52	0.47	0.35	0.40
T1-MR2	IM	0.83	0.95	1.02	1.58	2.19	0.81	1.18	0.92	0.99	0.98	0.98	1.05
	EDP	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	EDP IM	0.64	0.51	0.50	0.50	0.55	0.51	0.81	0.46	0.45	0.44	0.45	0.46
T2-MR2	IM	0.66	0.77	0.89	1.24	1.93	0.61	1.17	0.86	0.94	0.86	0.84	0.90
	EDP	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
	EDP IM	0.59	0.39	0.37	0.43	0.57	0.42	0.62	0.37	0.33	0.33	0.35	0.31
T3-MR2	IM	0.60	0.71	0.80	1.14	1.98	0.57	0.91	0.78	0.82	0.79	0.80	0.90
	EDP	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
	EDP IM	0.55	0.37	0.30	0.45	0.53	0.44	0.50	0.37	0.35	0.27	0.26	0.27
T4-MR2	IM	0.63	0.75	0.90	1.22	2.13	0.61	0.98	0.83	0.87	0.91	0.89	1.01
	EDP	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
	EDP IM	0.61	0.39	0.30	0.43	0.42	0.41	0.46	0.37	0.34	0.25	0.25	0.25