Universität Potsdam

Institute of Geosciences

GFZ Helmholtz Centre Potsdam

Enhanced computational approaches for data-driven characterization of earthquake ground motion and rapid earthquake impact assessment

Henning Lilienkamp

Cumulative dissertation

to obtain the degree:
"doctor rerum naturalium" (Dr. rer. nat.)
in the scientific discipline Seismology.

Submitted to:
the Faculty of Mathematics and Natural Sciences
at the University of Potsdam, Germany.

Prepared at:
Section 2.6 (Seismic Hazard and Risk Dynamics)
Department of Geophysics
Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences.

This work is protected by copyright and/or related rights. You are free to use this work in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s).
https://rightsstatements.org/page/InC/1.0/?language=en

Principal supervisor: Prof. Dr. Fabrice Cotton
University of Potsdam
Germany

Secondary supervisor: Prof. Giuseppe Caire, Ph.D.
Technical University Berlin
Germany

Reviewers: Prof. Dr. Fabrice Cotton
Prof. Dr. Gregory Beroza
Prof. Dr. Martin Mai

Examination committee: Prof. Dr. Eva Eibl
Prof. Dr. Fabrice Cotton
Prof. Dr. Gregory Beroza
Prof. Dr. Martin Mai
Prof. Giuseppe Caire, Ph.D.
Prof. Dr. Edward Sobel

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-63195
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-631954

Statement of Originality

I confirm that this assignment is my own work, is not copied from any other person's work (published or unpublished), and has not previously been submitted for assessment either at the University of Potsdam or elsewhere. I hereby certify that I have prepared this thesis without the help of third parties and without the use of sources and aids other than those indicated. The passages taken verbatim or in terms of content from the sources used are identified as such.

Potsdam, April 5, 2024

Henning Lilienkamp

"The most elementary and valuable statement in science
– the beginning of wisdom – is: I do not know."

– Lt. Commander Data

Acknowledgements

The past five years that I spent on my doctoral studies are among the most formative, challenging, but also rewarding periods of my life. I owe the successful completion of this endeavor largely to the support of my colleagues, friends, and my family, whom I would like to thank at this point.

First of all, I want to thank all the colleagues and friends at GFZ sections 2.6, as well as the members and coordinators of the HEIBRiDS graduate school for providing both a comfortable and stimulating working environment to conduct my doctoral studies.

In particular, I want to thank my principal supervisor Prof. Fabrice Cotton. Throughout my time at the GFZ, he has been the most supportive supervisor one could wish for. Fabrice encouraged me to find my own style of thinking, writing, and doing research, and to communicate it confidently to the outside world. Thanks to Fabrice's commitment, I never had to seriously worry about questions of financing, which made it possible for me to always focus completely on my research.

I would like to thank my secondary supervisor Prof. Giuseppe Caire for the fruitful scientific cooperation that led to the first publication of this dissertation. Maybe even more importantly, Giuseppe not only gave me the freedom, but actively encouraged me to follow my own research ideas, even when they moved far away from his own fields of interest and expertise in the second half of my doctoral studies.

Although Dr. Graeme Weatherill was not an official supervisor of my thesis, our detailed scientific discussions, the extensive and valuable reviews of my manuscripts and this thesis, as well as the close cooperation on two of its chapters, significantly contributed to the successful completion of my Ph.D. project. I highly appreciate Graeme's willingness to help, even when his own calendar is bursting at the seams.

I want to thank Dr. Sebastian Specht for his great scientific support from which I was able to benefit over the entire course of my time at the GFZ. This is especially true for the fruitful cooperation at the Academia Sinica in Taiwan

Furthemore I want to thank my Mentor Dr. Matthias Ohrnberger, who has provided me with reliable scientific and moral support throughout my doctoral studies.

The time at GFZ and in HEIBRiDS would not have been the same without my friends and colleagues Philipp, Tallte, Sissi, Ming-Hsuan, Karina, Smallte, Pouria, Shubham, Gesa, Peter, Pablo, Lorenzo, Reza, Zahra, Lukas, Jannes, Elif, Camilo, Laurens, Tara, Chen-Ray, Ugur, Sreeram, Annabel, Cecilia, Dino, Riccardo, Marco, Christian, Dorina, Solveig, Uwe [...]. Thanks to you, my doctorate was a time I will always remember with pleasure.

I want to thank also my dear flatmates Camille, Cathi, Bianca, and Nils for ensuring a healthy work-life-balance over the course of my studies.

Finally I would like to thank my parents Doris & Edmund, my sister Annika, and my brother Thomas for their unconditional moral support. Without them, this dissertation would not have come about.

Abstract

Rapidly growing seismic and macroseismic databases and simplified access to advanced machine learning methods have in recent years opened up vast opportunities to address challenges in engineering and strong motion seismology from novel, datacentric perspectives. In this thesis, I explore the opportunities of such perspectives for the tasks of ground motion modeling and rapid earthquake impact assessment, tasks with major implications for long-term earthquake disaster mitigation.

In my first study, I utilize the rich strong motion database from the Kanto basin, Japan, and apply the U-Net artificial neural network architecture to develop a deep learning based ground motion model. The operational prototype provides statistical estimates of expected ground shaking, given descriptions of a specific earthquake source, wave propagation paths, and geophysical site conditions. The U-Net interprets ground motion data in its spatial context, potentially taking into account, for example, the geological properties in the vicinity of observation sites. Predictions of ground motion intensity are thereby calibrated to individual observation sites and earthquake locations.

The second study addresses the explicit incorporation of rupture forward directivity into ground motion modeling. Incorporation of this phenomenon, causing strong, pulse like ground shaking in the vicinity of earthquake sources, is usually associated with an intolerable increase in computational demand during probabilistic seismic hazard analysis (PSHA) calculations. I suggest an approach in which I utilize an artificial neural network to efficiently approximate the average, directivity-related adjustment to ground motion predictions for earthquake ruptures from the 2022 New Zealand National Seismic Hazard Model. The practical implementation in an actual PSHA calculation demonstrates the efficiency and operational readiness of my model. In a follow-up study, I present a proof of concept for an alternative strategy in which I target the generalizing applicability to ruptures other than those from the New Zealand National Seismic Hazard Model.

In the third study, I address the usability of pseudo-intensity reports obtained from macroseismic observations by non-expert citizens for rapid impact assessment. I demonstrate that the statistical properties of pseudo-intensity collections describing the intensity of shaking are correlated with the societal impact of earthquakes. In a second step, I develop a probabilistic model that, within minutes of an event, quantifies the probability of an earthquake to cause considerable societal impact. Under certain conditions, such a quick and preliminary method might be useful to support decision makers in their efforts to organize auxiliary measures for earthquake disaster response while results from more elaborate impact assessment frameworks are not yet available.

The application of machine learning methods to datasets that only partially reveal characteristics of Big Data, qualify the majority of results obtained in this thesis as explorative insights rather than ready-to-use solutions to real world problems. The practical usefulness of this work will be better assessed in the future by applying the approaches developed to growing and increasingly complex data sets.

Zusammenfassung

Das rapide Wachstum seismischer und makroseismischer Datenbanken und der vereinfachte Zugang zu fortschrittlichen Methoden aus dem Bereich des maschinellen Lernens haben in den letzen Jahren die datenfokussierte Betrachtung von Fragestellungen in der Seismologie ermöglicht. In dieser Arbeit erforsche ich das Potenzial solcher Betrachtungsweisen im Hinblick auf die Modellierung erdbebenbedingter Bodenerschütterungen und der raschen Einschätzung von gesellschaftlichen Erdbebenauswirkungen, Disziplinen von erheblicher Bedeutung für den langfristigen Erdbebenkatastrophenschutz in seismisch aktiven Regionen.

In meiner ersten Studie nutze ich die Vielzahl an Bodenbewegungsdaten aus der Kanto Region in Japan, sowie eine spezielle neuronale Netzwerkarchitektur (U-Net) um ein Bodenbewegungsmodell zu entwickeln. Der einsatzbereite Prototyp liefert auf Basis der Charakterisierung von Erdbebenherden, Wellenausbreitungspfaden und Bodenbeschaffenheiten statistische Schätzungen der zu erwartenden Bodenerschütterungen. Das U-Net interpretiert Bodenbewegungsdaten im räumlichen Kontext, sodass etwa die geologischen Beschaffenheiten in der Umgebung von Messstationen mit einbezogen werden können. Auch die absoluten Koordinaten von Erdbebenherden und Messstationen werden berücksichtigt.

Die zweite Studie behandelt die explizite Berücksichtigung richtungsabhängiger Verstärkungseffekte in der Bodenbewegungsmodellierung. Obwohl solche Effekte starke, impulsartige Erschütterungen in der Nähe von Erdbebenherden erzeugen, die eine erhebliche seismische Beanspruchung von Gebäuden darstellen, wird deren explizite Modellierung in der seismischen Gefährdungsabschätzung aufgrund des nicht vertretbaren Rechenaufwandes ausgelassen. Mit meinem, auf einem neuronalen Netzwerk basierenden, Ansatz schlage ich eine Methode vor, um dieses Vorhaben effizient für Erdbebenszenarien aus dem neuseeländischen seismischen Gefährdungsmodell für 2022 (NSHM) umzusetzen. Die Implementierung in einer seismichen Gefährdungsrechnung unterstreicht die Praktikabilität meines Modells. In einer anschließenden Machbarkeitsstudie untersuche ich einen alternativen Ansatz der auf die Anwendbarkeit auf beliebige Erdbebeszenarien abzielt.

Die abschließende dritte Studie befasst sich mit dem potenziellen Nutzen der von makroseismischen Beobachtungen abgeleiteten pseudo-Erschütterungsintensitäten für die rasche Abschätzung von gesellschaftlichen Erdbebenauswirkungen. Ich zeige, dass sich aus den Merkmalen solcher Daten Schlussfolgerungen über die gesellschaftlichen Folgen eines Erdbebens ableiten lassen. Basierend darauf formuliere ich ein statistisches Modell, welches innerhalb weniger Minuten nach einem Erdbeben die Wahrscheinlichkeit für das Auftreten beachtlicher gesellschaftlicher Auswirkungen liefert. Ich komme zu dem Schluss, dass ein solches Modell, unter bestimmten Bedingungen, hilfreich sein könnte, um EntscheidungsträgerInnen in ihren Bestrebungen Hilfsmaßnahmen zu organisieren zu unterstützen.

Die Anwendung von Methoden des maschinellen Lernens auf Datensätze die sich nur begrenzt als Big Data charakterisieren lassen, qualifizieren die Mehrheit der Ergebnisse dieser Arbeit als explorative Einblicke und weniger als einsatzbereite Lösungen für praktische Fragestellungen. Der praktische Nutzen dieser Arbeit wird sich in erst in Zukunft an der Anwendung der erarbeiteten Ansätze auf wachsende und zunehmend komplexe Datensätze final abschätzen lassen.

Contents

Chapter 1

Introduction

1.1 Earthquakes – a perpetual threat throughout human history

The terrifying and at times disastrous consequences of large earthquakes have troubled mankind since earliest history. An impressive historic example for the awe-inspiring intimidation that these natural events posed to early societies can be found in the bamboo annals (see Fig. 1.1), a chronicle of ancient China, that mentions the shaking caused by a major earthquake that occurred presumably in the 16th century B.C. close to Mount Tai in today's Shandong province in China (Legge, 1865).

Figure 1.1: Excerpt from the bamboo annals which is commonly considered the first report of an earthquake in human history. The event presumably occurred at Mount Tai, China, in the 16th century B.C. and is apparently related to the death of the then emperor Fah. Figure modified from Legge (1865).

After this transcription, which is today considered the first ever qualitative description of an earthquake, it took another 2,000 years until the first seismoscope, a rudimentary instrument to detect ground shaking, was developed by the chinese philosopher Chan Heng in A.D. 132

(Dewey and Byerly, 1969). Ever since, the observation of seismic waves and the examination of the underlying processes has progressed, not least motivated by the disastrous consequences of events such as the 1556 earthquake in the Shaanxi province, China, the deadliest earthquake to date, causing approximately 830,000 fatalities (Kuo, 1957).

With the emergence of the elastic rebound theory (Reid, 1910) in the early 20th century and the broad acceptance of the theory of plate tectonics in the 1960s, the theoretical foundation for the description of earthquake formation was laid. At the center of these theories is the concept that the Earth's lithosphere, is divided into multiple tectonic plates that float on the underlying, more viscous asthenosphere. Variability in the direction and speed of movement of tectonic plates causes the accumulation of mechanical stress along the locked plate boundaries, which eventually releases in a sudden rupture, which is accompanied by an energy release in the form of elastic waves. With the installation of the first global network of seismic instruments, also in the 1960s (Lay and Wallace, 1995), modern day Seismology was finally established.

Observations of seismic waves with networks like the Global Seismographic Network (GSN), comprising 152 instruments distributed around the globe, and additional regional networks such as KiK-net and K-net (National Research Institute for Earth Science and Disaster Resilience, 2019) covering over 1,000 observation sites in Japan for example, have made Seismology a remarkably data rich science in recent years. In addition, both the United States Geological Survey (USGS) and the European Mediterranean Seismological Centre (EMSC) have launched programs to systematically collect observations of earthquake related phenomena by non-expert citizens (Bossu et al., 2017; Wald et al., 1999), basically a return to the very same methodology that gave us the first earthquake description in the Bamboo annals over 3,500 years ago.

The subsequent abundance of earthquake related data that we are facing today offers a wide range of possibilities to gather new insights into the processes of the Earth's interior system and the resulting risks they potentially pose to modern societies. The development and proper application of sophisticated data analysis frameworks, including also advanced technologies from the field of machine learning, is an essential, complementary element in order to maximize knowledge gain from large data sets.

Despite all the technological progress and the knowledge gained in the past centuries, earthquakes have never lost their terror. Disasters such as the recent earthquakes in February 2023 affecting Turkey and Syria, which claimed more than 50,000 lives (National Geophysical Data Center / World Data Service, 2023), are shaking our modern society today just as they did over 400 years ago in the Shaanxi province in China. In the light of a rapidly growing world population, the actual number of at-risk societies is expected to increase substantially in the future, such that epicentral hits of megacities may cause up to 1 million fatalities (Bilham, 2009). Thus, the need for seismic risk mitigation has never been greater than today.

The purpose of this thesis is to contribute to this daunting task by exploring how the abundance of available data and the access to advanced methodologies such as deep learning can be best utilized in concert with proven standards to mitigate earthquake consequences in the future.

1.2 Preventive earthquake disasater mitigation – Probabilistic seismic hazard and risk assessment

Despite the considerable research efforts in the last decades regarding the precise forecasting of future earthquakes in terms of location, magnitude, and timing, it is the prevailing opinion among the seismological community that the occurrence of individual earthquakes is largely unpredictable (Jordan et al., 2011). Instead, probabilistic seismic hazard assessment (PSHA) has established as the preferred methodology to achieve long-term preparedness for earthquake-induced ground shaking. Rather than aiming at the precise prediction of individual earthquakes, PSHA aims to identify all earthquake ruptures that could possibly occur in the vicinity of a site of interest. In addition, also the probability of occurence for each of these earthquake scenarios within a given period of time (typically decades) is quantified. In a final step, the intensity of ground shaking that would be caused by such earthquake scenarios is estimated, yielding a probability distribution of ground shaking intensity that is to be expected in coming decades. Such a result can subsequently be compared against the resistance of engineered structures at the site of interest in a seismic risk assessment, and reinforcement measures can be initiated if insufficient ability to withstand the expected shaking is attested.

Empirical ground motion models (GMMs) that predict a probability distribution of a ground motion intensity measure (IM) given a description of an earthquake scenario (e.g., the earthquake's magnitude, its distance to the site of interest, and a geophysical characterization of the site) are an essential component in PSHA. In their classical formulation, GMMs are equation based models, the functional forms of which represent fundamental physical principles of earthquake source scaling, wave propagation, and site amplification (Kramer, 1996b), but whose coefficients are calibrated empirically upon observed ground motion data. Although purely physics-based ground motion simulation is possible both from a theoretical and methodological point of view (e.g. Graves and Wald, 2001; L. Zhao et al., 2006), limited knowledge about the structure and the geophysical properties of the Earth's subsurface require the empirical calibration of GMM coefficients with actual observations in most regions on the planet. When the first GMMs were developed in the 1960s, ground motions observations were sparse, and the equations consisted of only a hand full of free coefficients (e.g. Esteva and Rosenblueth, 1964). As data became abundant and variable over the last decades, GMMs evolved into more and more complex statistical instruments, a development that in the end could no longer achieve any significant progress as far as model uncertainties are concerned (Douglas and Edwards, 2016; Strasser et al., 2009), indicating that the ability of classical ground motion modeling frameworks to describe ground motion data are increasingly being exhausted.

As a consequence, the ground motion modeling community has started to explore alternative modeling strategies, first and foremost the relaxation of the ergodic assumption. The ergodic assumption states that samples of ground shaking recorded over time at one specific location follow the same distribution as samples collected at the same point in time across multiple sites (Anderson and Brune, 1999). In practice, this assumption allows a modeler to calibrate a GMM with a finite number of point observations sampled within a single, or across multiple tectonic regions, and subsequently utilize it for ground motion prediction in regions tectonically

analogous to those represented in the sampled data. In fact, however, ground motion datasets may contain region or site specific characteristics due to geophysical properties that are either unknown, or difficult to obtain in practice, such as the average stress drop of earthquake ruptures within a specific tectonic setting, or the geological conditions of the subsurface underneath a site of interest. Therefore, calibrating a GMM under the ergodic assumption with a joint dataset containing different region- and site-specific characteristics may lead to inflation of aleatory uncertainty that could be reduced if these characteristics were taken into account (Stafford, 2014). Such region- or even site-specific calibration of GMMs is a first step towards relaxation of the ergodic assumption, and can considerably reduce the aleatory uncertainty provided that a sufficiently large and dense dataset from a specific region is available.

Besides the relaxation of the ergodic assumption, also the rise of novel machine learning techniques, especially deep learning (Lecun et al., 2015), has brought about new perspectives to ground motion modeling in recent years. Deep learning is an increasingly popular machine learning method that provides the opportunity to implement predictive models without actually defining the relation between a set of predictive parameters (e.g. earthquake magnitude and the source-to-site distance) and target parameters (e.g. a ground motion intensity measure) in an equational form. Instead, example pairs of values of predictive and target parameters are fed to an artificial neural network (ANN), a modeling framework designed to mimic biological information processing (Bishop, 2006), which then autonomously adjusts its internal coefficients (neurons) in order to map the given predictive parameters to an output that closely resembles the desired value of the target parameter. Typical deep ANNs comprise hundreds or even millions of neurons and are therefore prone to overfitting if no appropriate countermeasures are taken. Therefore, a huge amount of data samples is required to ensure a stable calibration and predictive power of the final model. If such a dataset is available, ANNs are capable of adopting very closely subtle, non-linear nuances in datasets that might not be evident to the modeler from classical data analysis or theoretical a priori knowledge. While the calibration of ANNs goes along with considerable computational effort and, therefore, the necessity for advanced computing resources, inference is highly efficient and easy to implement.

In the context of ground motion modelling, ANNs offer the opportunity to explore ground motion data in a fully data-driven way. GMMs based on simple feed-forward ANN architectures have been implemented successfully in the past (e.g. Derras et al., 2014; Dhanya and Raghukanth, 2018; Pozos-Estrada et al., 2014), however, advanced architectures such as convolutional neural networks, designed to operate on array shaped data such as images (e.g. LeCun et al., 1989), have received little attention in the ground motion modeling community. Yet such architectures offer interesting properties: While equation-based GMMs and simple feed-forward ANNs consider ground motion observations as spatially independent point measurements, representing a regional ground motion dataset in its spatial context as maps and operating on it using a convolutional neural network potentially allows the modeler to natively incorporate the spatial context into the final GMM. The resulting model may therefore adapt to region- or site-specific properties and can, therefore, be considered as nonergodic. Consequently, convolutional neural networks pose a unique opportunity to join the so far independently pursued approaches of fully data-driven modeling via ANNs on the one hand, and nonergodic modeling on the other hand.

Another aspect in current ground motion modeling and PSHA practices that would potentially benefit from refinement is the implicit way of incorporating directivity effects. Rupture forward directivity is a kind of doppler effect that occurs when the rupture propagation along a tectonic fault is exceptionally fast (Somerville et al., 1997). As a consequence, sites that are located in the vicinity of tectonic faults may experience extraordinarily intense, pulse-like ground shaking, that poses a larger seismic demand to buildings than non pulse-like shaking. Ground motion measurements that are affected by directivity effects are a small but integral component of most ground motion databases and are difficult to explain with GMMs that do not explicitly take directivity into account. Consequently, the predictive uncertainty of such a GMM will be underestimated at sites close to tectonic faults, where directivity effects are most prominent, and overestimated at larger distances, where directivity effects are vanishing (Weatherill, 2022).

While explicit incorporation of directivity effects in ground motion modeling is realizable via the use of directivity models such as the one by Bayless et al. (2020), extension of the PSHA framework in this regard, although theoretically straightforward, goes along with a several hundredfold increase in computational demand that would prohibit the implementation of seismic hazard studies at regional or national scale in many cases. With the introduction of the modifier of moments approach, e.g., Donahue et al. (2019) and Watson-Lamprey (2018) suggest to circumvent the simulation of hundreds of individual directivity scenarios that could possibly occur on a specific rupture in a PSHA computation on the fly, and instead calculate the average amplification pattern and its variability in the vicinity of the considered ruptures (the moment modifiers) beforehand, as to remove this step from the actual PSHA procedure.

Because such a lookup table of amplification patterns would be impractically large to handle in PSHA studies at regional or national scale, simplified empirical equations are fit to reproduce the moment modifiers based on the geometric and seismic properties of ruptures. While Watson-Lamprey (2018) demonstrated the suitability of this approach for simple rupture geometries, accurate performance on highly complex ruptures, that are encountered more and more often in advanced seismic source models such as the Uniform California Earthquake Rupture Forecast v3.0 (Field et al., 2014) or the New Zealand Community Fault Model (Seebeck et al., 2023), is questionable.

Yet again, deep learning offers promising alternatives for the implementation of the modifier of moments approach that could potentially extend its applicability towards more complex ruptures. First of all, the ability of ANNs to transfer given information into abstract features in an extremely efficient way is an appealing feature in the light of the impractical size of moment modifier lookup tables. Considering also the highly efficient inference, this raises the idea of utilizing an ANN as an efficiently compressed and easily accessible lookup table for implementation of the modifier of moments approach in PSHA. Furthermore, the ability of ANNs to identify complex and subtle relations in large datasets may enable a deep learning-based model to find meaningful relations between complicated rupture geometries and the spatial patterns of the corresponding moment modifiers and reproduce them with greater accuracy than is achievable via empirical equations.

1.3 Post-event earthquake disaster mitigation – Rapid earthquake impact assessment and response to earthquakes

Although PSHA provides a comprehensive quantification of earthquake hazard upon which further long-term disaster mitigation strategies can be developed, this approach cannot prevent the occurence of individual catastrophes such as the disastrous consequences of the recent Kahmaranmaras Earthquakes in Turkey in the near term. Therefore, also effective rapid response to earthquakes (RRE) strategies, i.e., efficient organization and navigation of auxiliary measures in the immediate aftermath of disastrous earthquakes, are an indispensable component in the overall concept of earthquake disaster mitigation. Rapid earthquake impact assessment (RIA) via frameworks such as the United States Geological Survey's (USGS) service PAGER (Wald et al., 2010) typically forms the basis on which authorities build their decisions in the immediate aftermath of significant earthquakes. The PAGER methodology is a multi-step procedure where the first step involves the analysis of seismic recordings in order to characterize the earthquake source. Subsequently, empirical GMMs are utilized together with actual observations to estimate the ground shaking across the affected area via a ShakeMap (Wald et al., 2005). Finally, information about population density and the engineering properties of buildings are taken into account to estimate the amount and spatial distribution of casualties and economic losses. While the actual processing time of PAGER strongly depends on the available infrastructure in the affected region, the average global delay between an earthquake origin time and the first available loss estimate amounts to 30 min (Wald et al., 2010). Although this efficiency might be sufficient for most practical purposes, it still leaves decision makers in an information vacuum for a considerable period of time in which any kind of information, even if it is rudimentary and uncertain, might be valuable.

Pseudo-intensity felt reports collected via the European Mediterranean Seismological Centre's service LastQuake (Bossu et al., 2018b) are available rapidly after the occurence of major earthquakes (average reporting time of 10 min), and therefore present a promising option to fill the information gap in the immediate aftermath. Pseudo-intensity felt reports are based on macroseismic observations, i.e., observations of earthquake induced phenomena such as shattering windows, swinging ceiling lights, or damage to buildings, by non-expert citizens, who report their experience via a smartphone application. In a second step, these observations are converted into a pseudo-intensity value between 1 (not felt) and 12 (completely devastating) according to the EMS-98 macroseismic scale (Grünthal, 1998). Compared to instrumental recordings of ground shaking, uncertainties accociated with pseudo-intensities are large for multiple reasons. First of all, the lack of expertise and potential emotional bias of earthquake witnesses may lead to an exaggeration of the personal experience, leading to pseudo-intensity values being biased high (Bossu et al., 2017). Furthermore, individual macroseismic observations are not considered to be representative for the actual macroseismic intensity of an area, for which averaging over multiple observations is crucial (Grünthal, 1998).

Despite these limitations, the general value of a rapid information source that pseudo-intensity reports pose is widely recognized, although little benefit is derived from such in current RRE practices. Further investigation of this so far unused potential is therefore an important

task in the course of improving the resilience of at-risk societies to earthquakes.

1.4 The emerging relevance of Big Data in probabilistic seismic hazard assessment and rapid response to earthquakes

According to Arrowsmith et al. (2022), Big Data Seismology is characterized by the application of advanced machine learning (ML) algorithms for the analysis of large and complex datasets, using specialized and high-performance computational resources. Multiple studies in recent years have demonstrated that appropriate application of ML, especially deep learning (DL), to Big Data can lead to scientific breakthroughs. In seismology, major breakthroughs were achieved for example in the DL-based detection of earthquake signals in continuous, raw waveform data via PhaseNet (Zhu and Beroza, 2019) and GPD (Ross et al., 2018). A common foundation of the aforementioned studies, that actually brought about results of considerable practical relevance, is the availability of data sets that can be largely characterized as Big Data in the sense of the "5 Vs of Big Data" (*Volume*, *Velocity*, *Variability*, *Variety*, *Value*) (L'Heureux et al., 2017). However, application of ML algorithms to Small Data may also grant valuable insights into the underlying relations and structures, although the resulting models may not be readily applicable in practice.

The previously described developments in PSHA and RRE reveal a considerable overlap with the field of Big Data Seismology. Making a profit of the growing amounts of ground motion records and macroseismic observations is increasingly becoming the focus of current research and the application of DL on specialized computing hardware to further develop PSHA practices has established in recent years. However, while the raw seismic data available at centers such as the IRIS DMC waveform archive IRIS, 2023 (about 900 TB) is indeed entering the realm of Big Data, individual studies are usually conducted only on subsets of the same or on derived, dimensionally reduced [1] data products. Furthermore, the *Volume* of data is only one among commonly five considered criteria that describe Big Data. Therefore, careful assessment of the actually used datasets with respect to Big Data is crucial in order to judge whether ML based results pose experimental insights or practical solutions to real world problems.

1.5 Purpose and structure of this thesis

1.5.1 Research questions

The overarching concept of my doctoral studies was to explore novel, data focused perspectives in the fields of probabilistic seismic hazard assessment and rapid response to earthquakes, with a strong emphasis on the evaluation of deep learning methods. Rather than advancing and challenging existing methodologies and strategies, it was my intention to open alternative perspectives on well-researched topics and to explore novel modeling approaches that may complement traditional methods to mutual advantage. In particular, I addressed the following research questions:

[1] Often only a handful of features obtained from ground motion time series are utilized, rather than the entire time series itself. Compare section 2.1.

Research question 1 "What are the opportunities and limitations of utilizing advanced artificial neural network architectures for ground motion modeling in the light of rapidly growing datasets obtained from spatially dense instrumental networks?"

Contrary to the early days of strong motion seismology when observations were rare and the design of ground motion models heavily relied on theoretical a priori knowledge, the abundance of observations we are facing today allows us to relax such restrictions and instead learn from actual evidence. Fully data-driven training of artificial neural networks (ANNs) is the antithesis to hand designed physics-inspired equations that are commonly utilized as GMMs today, because ANN based models represent a fully impartial interpretation of ground motion data that is free of any a priori assumptions. Breakthrough successes of basic deep learning models in modeling complex datasets in general, and advances in ergodic modeling of strong motion data in particular, sparked my motivation to investigate an advanced ANN architecture, the U-Net, with respect to its potential for fully data-driven and nonergodic ground motion modeling. In chapter 3, I present my aspirations to implement the prototype of such a model for the Kanto region in Japan in order to learn more about the opportunities and limitations that such an approach brings. A continuing discussion of this study in a broader scientific context is presented in section 7.1.

Research question 2 "How can deep learning assist the implementation of explicit incorporation of directivity related amplification effects in probabilistic seismic hazard calculations?"

While explicit incorporation of directivity effects in PSHA calculations via extension of the hazard integral (full hypocenter randomization) is theoretically a straightforward endeavor, the associated computational demand makes the actual application to PSHA studies at regional or national scale infeasible. The modifier of moments approach offers an efficient alternative to the full hypocenter randomization, one in which the mean amplification and variability in the vicinity of an earthquake rupture due to varying hypocenter locations is described via a simplified, hand-designed, and empirically calibrated equation. However, the applicability of such simplified equations to complex rupture geometries is a matter of debate. Compared to a simple empirical equation, ANNs offer increased flexibility that could potentially lead to more accurate representations of empirically learned spatial amplification patterns, also for complex rupture geometries. Therefore, in chapter 4 I investigate the opportunity to store a synthetically generated database of moment modifiers

for a fix set of ruptures in the hidden layers of an ANN, such that it can be retrieved efficiently during seismic hazard calculations. In the subsequent chapter 5, I follow a different idea in which I explore an actual deep learning approach, i.e., one in which an ANN is trained to relate earthquake rupture properties to the spatial patterns of the moment modifiers in its vicinity. An extended discussion of this research question is presented in section 7.2.

Research question 3 "How can we take optimal advantage of the rapid availability of pseudo-intensity reports with regard to rapid impact assessment?"

Pseudo-intensity reports collected globally via the LastQuake service are among the first available information sources in the immediate aftermath of an earthquake. While macroseismic intensity values obtained from averaging over multiple macroseismic observations are commonly used as an additional constraint in the ShakeMap methodology, utilization of pseudo-intensity values has been widely neglected so far. The fact that pseudo-intensity values are obtained from macroseismic observations, suggests that the statistical and geographic properties of collections of pseudo-intensities should be correlated with the final impact that the corresponding earthquake has on the affected society. Chapter 6 describes an exploratory study in which I present my attempts to firstly verify this assumption, and secondly to utilize this correlation to formulate a statistical model that allows rapid, probabilistic classification of earthquakes to cause substantial societal impact or not. An evaluation of the model using earthquakes from the February 2023 Kahmaranmaras sequence is included as part of the continuing discussion in section 7.3.

Research question 4 "To what extent can the studies presented within the scope of this thesis be associated with the field of Big Data Seismology, and what are the implications for the presented results?"

Both the application of machine learning methods on specialized hardware and the optimal usage of rapidly growing databases are key aspects of my doctoral studies. Subsequently, a certain proximity to the advent field of Big Data Seismology cannot be denied. However, incautious association of my studies to this field might lead to misconception of the presented results and conclusions, since many thought processes also contradict typical Big Data principles. Therefore, as a part of the overarching discussion of this thesis in section 7.4, I want to take a closer look and clarify in which aspects the presented studies exhibit characteristics of Big Data, and in which they do not.

1.5.2 Structure of this thesis

In the following chapter 2, I provide a brief introduction of seismological datasets and selected foundations and concepts from data science and earthquake disaster mitigation to facilitate the entry into the core chapters 3 – 6. The organization of the core chapters with respect to the scientific context is summarized in Fig. 1.2. In chapter 3 I present my first published manuscript regarding the development of an ANN-based ground motion model for the Kanto

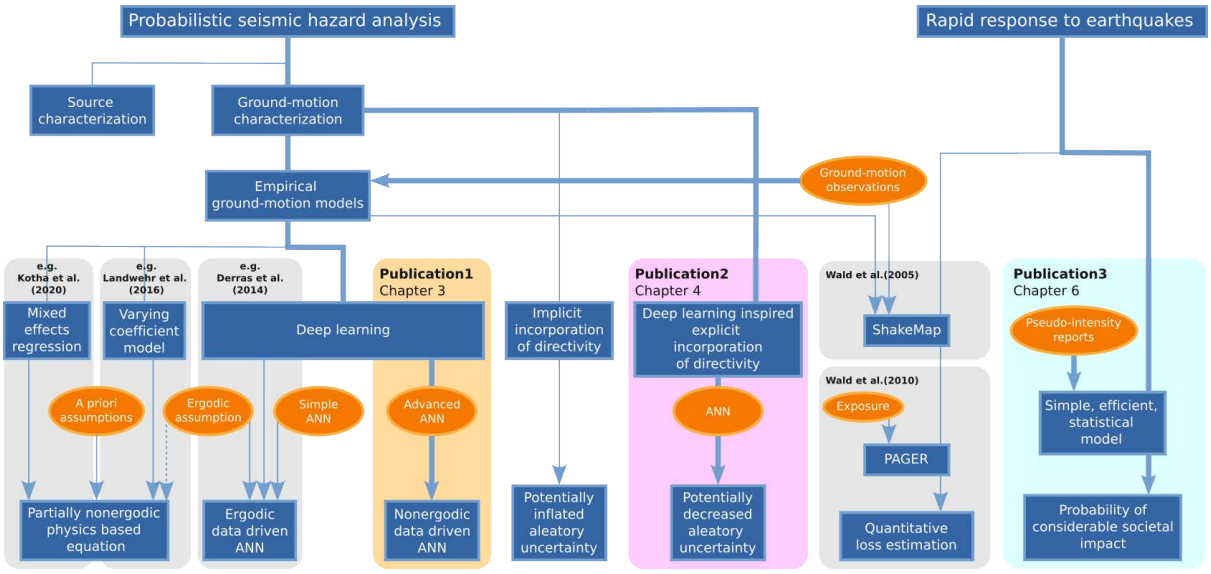

Figure 1.2: Organization of the core chapters 3, 4, and 6 of this thesis within the scientific context.

basin, Japan, following the ideas outlined in research question 1. Implementation of explicit incorporation of directivity effects in PSHA, as introduced in research question 2, is addressed in a manuscript that is currently under review presented in chapter 4. The subsequent chapter 5 contains unpublished material regarding an alternative approach to research question 2. My third key contribution regarding the utilization of pseudo-intensity reports for the assessment of earthquake impact is presented in the form of a published manuscript in chapter 6. Chapter 7 contains an overarching and continuing discussion with respect to the research questions 1–3 (sections 7.1 – 7.3) and research question 4 (section 7.4), which is not addressed explicitly in the core chapters. I close my thesis with a summary of the key conclusions in chapter 8.

1.6 Author's contributions to publications in this thesis

The chapters 3, 4, and 6 of this thesis contain manuscripts that are either published in peer-reviewed journals, or are currently under review of such. The content of these manuscripts represents the essential material that forms this cumulative dissertation. In the following I will provide a list of said publications and clarify my respective contributions.

- **Chapter 3:**
  - **Lilienkamp, H.**, S. von Specht, G. Weatherill, G. Caire, and F. Cotton (2022). "Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach". In: Bulletin of the Seismological Society of America 112 (3), pp. 1565–1582. doi: 10.1785/0120220008.
  - **Lilienkamp, H.**, S. von Specht, G. Weatherill, G. Caire, and F. Cotton (2022). "Erratum to Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach". In: Bulletin of the Seismological Society of America 112 (5), 2770–2771, doi: 10.1785/0120220131
    - H.L. participated in the conceptual design of the study. H.L. wrote python computer codes regarding data processing and modeling. H.L. chose and implemented the modeling framework. H.L. ***lead authored*** the manuscript. H.L. designed and prepared all figures in the manuscript. H.L. discovered post-publication a coding error. H.L. investigated the effects of the coding error on the main results in the published manuscript. H.L. ***lead authored*** the erratum to clarify this influence. S.v.S. contributed to the development of the statistical framework and writing the manuscript. G.W., G.C., and F.C. contributed to the design of the study and editing of the manuscript.

- **Chapter 4:**
  - Weatherill, G., and **H. Lilienkamp** (2023). "Implementing Directivity in Probabilistic Seismic Hazard Analysis for New Zealand". Submitted to: Bulletin of the Seismological Society of America. Accepted for publication with "minor revisions" suggested.
    - H.L. participated in the design of the synthetic dataset, described in section 4.5.2.4. H.L. introduced the idea of using an artificial neural network as a data storage for moment modifiers. H.L. participated in editing the manuscript. H.L. conducted deep learning experiments and wrote the related python computer code. H.L. designed and implemented the model evaluation presented in section 4.5.2.7. H.L. participated in writing the manuscript, in particular sections 4.5.2.2, 4.5.2.3, 4.5.2.5, 4.5.2.6, and 4.5.2.7. H.L. designed and prepared Figures 4.9–4.11. H.L. produced texts, figures, and tables shown in appendix 4.A. G.W. designed the study, conducted seismic hazard calculations, designed and produced figures, and lead-authored the manuscript.

- **Chapter 6:**
  - **Lilienkamp, H.**, R. Bossu, F. Cotton, F. Finazzi, M. Landès, G. Weatherill, S. von Specht (2023). "Utilization of Crowdsourced Felt Reports to Distinguish High-Impact from Low-Impact Earthquakes Globally within Minutes of an Event". In: The Seismic Record 3 (1): 29–36. doi: 10.1785/0320220039

* H.L. participated in the conceptual design of the study. H.L. chose and implemented the statistical modeling framework. H.L. conducted modeling experiments. H.L. wrote python computer codes regarding data processing and modeling. H.L. ***lead authored*** the manuscript. H.L. designed and prepared all figures in the manuscript. R.B. lead the design of the study. R.B. and M.L. granted access to data. F.C. organized the scientific cooperation. F.F. and S.v.S. contributed to the development of the methodological framework. R.B., F.C., F.F., M.L., G.W., and S.v.S. edited the manuscript.

1.7 Additional relevant publications

- Gómez Zapata, J.C., Pittore, M., Cotton, F., **Lilienkamp, H.** et al. "Epistemic uncertainty of probabilistic building exposure compositions in scenario-based earthquake loss models". In: Bull. Earthquake Eng. 20, 2401–2438 (2022). doi: 10.1007/s10518-021-01312-9

  – H.L. contributed to the implementation of a spatial cross-correlation model for ground motion intensity measures.

- **Lilienkamp, H.**, Lilienkamp, T. "Detecting spiral wave tips using deep learning". Sci. Rep. 11, 19767 (2021). doi: 10.1038/s41598-021-99069-3

  – H.L. contributed to hyperparameter optimization of artificial neural network models. H.L. was partially involved in writing the manuscript. H.L. edited the manuscript.

1.8 Popular scientific dissemination of findings

My studies regarding the utilization of LastQuake pseudo-intensity reports for the evaluation of the societal impact of earthquakes has attracted attention beyond the scientific community and was actively disseminated in the following popular scientific documents:

- Sidik, Saima May. Felt Reports Could Shake Up Earthquake Response. Eos, (2023). https://eos.org/articles/felt-reports-could-shake-up-earthquake-response

- Yavas, Civan. Earthquake emergency response: How can felt reports help?. Temblor (2023). https://temblor.net/temblor/earthquake-emergency-response-felt-reports-15354/

- Meldung der "empfundenen Stärke" von Erdbeben via App könnten im frühen Katastrophenmanagement helfen. idw - Informationsdienst Wissenschaft (2023). https://idw-online.de/de/news810359

Chapter 2

Scientific background

The purpose of this chapter is to establish the scientific context around my core studies presented in the chapters 3 – 6. In particular, I will give an introduction to the general data types and sets used over the course of my doctoral studies, as well as a brief explanation of the field of Big Data Seismology and its relevance for this thesis. Special emphasis is placed on introducing the field of deep learning, as it is used extensively throughout this thesis. Finally, I will establish an overview on the diverse strategies for earthquake disaster mitigation and explain the connection to my studies.

2.1 Data in seismology

Among the various data types, recordings of ground motion time series at the Earth's surface, i.e., seismograms or waveforms, are the primary source of information, whereby one distinguishes low amplitude ground motion records obtained at large distances from the earthquake source (weak motion) from large amplitude records obtained in the close vicinity of the earthquake rupture (strong motion). Strong motion records are of particular interest for earthquake disaster mitigation strategies, i.e., seismic hazard analysis, earthquake early warning, and rapid response to earthquakes (see section 2.3.1), because they provide the most information about the characteristics of ground motions that have the greatest potential to lead to damage.

The time series of strong motion records are seldom used for data analysis directly, but usually only a small amount of features such as the peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD) (maximum values of acceleration, velocity, and displacement, respectively, obtained from the time series of ground shaking) are extracted from the time series (e.g. Ancheta et al., 2013). Moreover, response spectral acceleration $SA(T)$, i.e., the peak absolute acceleration of single-degree-of-freedom oscillators with eigenperiods T that are exposed to the recorded ground motion is a widely utilized feature. The consequence of such processing is a substantial data reduction, that leads to a loss of information, as time series with tens of thousands of samples are summarized into sets of a couple of tens of strong motion features. Databases yielding such processed strong motion data are referred to as flat files. The strong motion flat file provided by Bahrampouri et al. (2021) obtained from the KiK-net accelerometer network in Japan is the data foundation for the study presented in chapter 3 of this thesis.

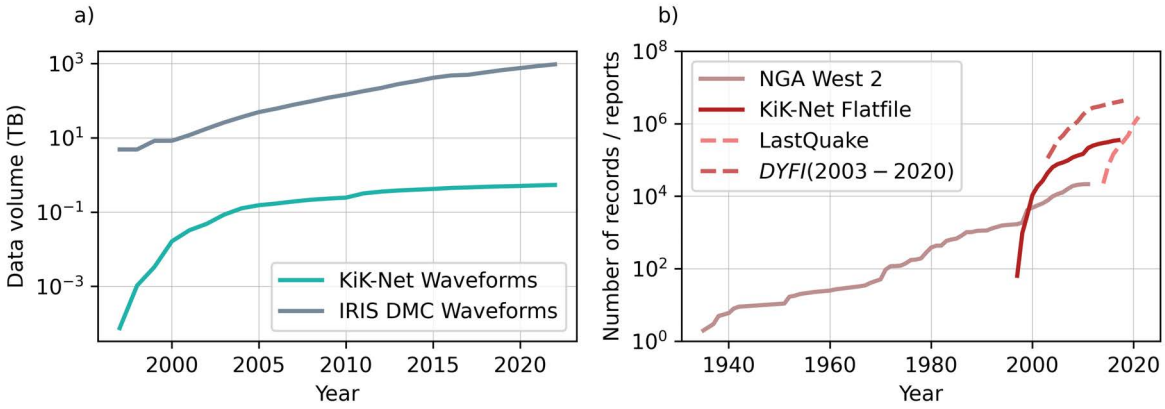

Figure 2.1: Temporal evolution of (a) seismic waveform databases and (b) strong motion flat files (solid lines) / macroseismic observation databases (dashed lines). The KiK-net flat file and the LastQuake database are utilized in chapters 3 and 6 of this thesis, respectively. Numbers for the NGA-West2 database are obtained from the corresponding flat file (Ancheta et al., 2014). The LastQuake database used for visualization is available via Bossu et al. (2023). Numbers regarding the "Did you feel it?" (DYFI) database development since 2003 are obtained graphically from Figure 6 in Quitoriano and Wald (2020). Note that DYFI reports before 2003 are neither considered in Quitoriano and Wald (2020) nor in this figure. Numbers regarding the KiK-net flat file are obtained from the Bahrampouri et al. (2021) database. KiK-net waveform data was downloaded from the database published by National Research Institute for Earth Science and Disaster Resilience (2019). Numbers regarding the IRIS DMC wavefrom archive were obtained graphically from the IRIS summary sheet available via IRIS (2023).

In addition to instrumental recordings, also macroseismic observations, i.e., observations of effects that are caused by ground shaking, such as perception by people and animals, swinging motion of hanging objects or damage to buildings, are an important source of information. The traditional approach of incorporating such data is to conduct post-earthquake surveys, in which experts collect a large amount of macroseismic observations within a limited geographical area, and subsequently quantify the average shaking intensity in the area via a macroseismic intensity scale such as EMS-98 (Grünthal, 1998). Nowadays, this procedure is also accomplished in a more rapid way via internet-based services such as "Did you feel it?" operated by the United States Geological Survey (USGS) (Wald et al., 1999).

Contrary to macroseismic intensities, pseudo-intensity values are obtained from individual macroseismic observations made by non-expert citizens evaluated against a macroseismic scale. Due to the lack of both expertise and spatial averaging of multiple observations, the uncertainties associated with pseudo-intensities are large, and usage of different macroseismic scales among different services, e.g., the EMS-98 scale in Europe or the modified mercalli scale (MMI, Wood and Neumann, 1931) which is popular in the US, limits the compatibility of datasets. However, collection of pseudo-intensity values in the aftermath of an earthquake via internet-based services is efficient and such databases are among the first available sources of information. A database of pseudo-intensities collected globally via the LastQuake service (Bossu et al., 2018b) operated by the European Mediterranean Seismological Center (EMSC) is the data foundation for the study presented in chapter 6 of this thesis.

The technological progress of the last decades has led to a dense instrumentation of the Earth's surface with seismometers and efficient collection of macroseismic intensities and pseudo-intensities via the internet, such that the volumes of the corresponding databases have been growing rapidly (see Fig. 2.1), and will continue to do so in the future. Due to global databases such as the IRIS DMC archive nowadays comprising several hundreds of terabytes of waveform data, the term Big Data is brought into play more and more often regarding data-driven studies in seismology. Ultimately, this has led to the advent of an entirely new sub-discipline referred to as Big Data Seismology.

2.2 Big Data Seismology

Arrowsmith et al. (2022) define Big Data Seismology as the field of study where large seismic datasets obtained from increasingly dense seismometer networks are analyzed with highly specialized algorithms that make use of advanced computational resources (compare Fig. 2.2). Since the key contributions to this thesis touch on various aspects of this discipline in one way or the other, I want to briefly elaborate on each of the three driving factors in the following. In doing so, I also want to lay the foundation for later discussion of research question 4 regarding the relation of my own studies to this rather new and advent discipline.

2.2.1 Big Data

A common scheme to describe the term Big Data is the definition of a dictionary containing its presumed key properties, such as "The 5 Vs of Big Data" (Volume, Velocity, Variety, Veracity, Value) (L'Heureux et al., 2017). The fact that the size of such dictionaries ranges from "The 3 Vs of Big Data" (Sagiroglu and Sinanc, 2013) all the way up to "The 10+ Vs of Big Data", indicates how vastly the perception of what Big Data actually is depends on the individual domain-specific data challenges that are faced in various scientific disciplines and economic sectors. In the context of this thesis, I consider "The 5 Vs of Big Data" to be an appropriate definition, and will describe the associated properties in the following.

2.2.1.1 Volume

The key term *Volume* refers to challenges in data processing or analyses due to the sheer amount of available data. One typical *Volume* related issue in Seismology is the identification of earthquake signals and the subsequent picking of seismic wave arrivals in huge amounts of continuous, noisy waveform data. While until the 1970s this task was conducted manually by experts, the interest in fully automated algorithms was sparked around that time (e.g. Allen, 1978), due to the rapid growth of waveform archives (Kong et al., 2019). Today, seismic phase picking on large data volumes is almost exclusively performed with advanced machine learning-based picking algorithms such as PhaseNet (Zhu and Beroza, 2019) or GPD (Ross et al., 2018).

2.2.1.2 Velocity

In the context of Big Data, the term *Velocity* refers to the challenge to instantaneously process rapidly generated data. Earthquake early warning (EEW, compare section 2.3.1) is among the most prominent disciplines where efficient management of rapidly generated data is a key prerequisite. Due to the physically limited warning times of a few seconds it is self-explanatory that any additional delay due to data processing and analysis should be avoided as a matter of urgency. Consequently, EEW might be considered a discipline that is facing a Big Data challenge in terms of *Velocity*. Another illustrative example is the field of rapid impact assessment (RIA, section 2.3.3), that focuses on the fast assessment of earthquake impact on the affected society to then organize appropriate auxiliary measures in time. Again, efficient collection and near real-time evaluation of rapidly generated waveform data and macroseismic observations are crucial to organize disaster management in the most efficient way.

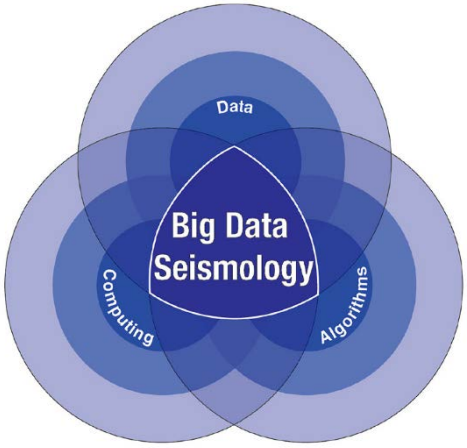

Figure 2.2: Visualiziation of the Big Data Seismology concept from Arrowsmith et al. (2022). The three driving factors are large, increasingly growing seismic datasets, the application of accordingly specialized algorithms for analyses, and the associated utilization of advanced computational resources.

2.2.1.3 Variety

Challenges in data processing and analysis may also arise due to the *Variety* of datasets that contain, e.g., various types of data from different locations, instrument types and formats (L'Heureux et al., 2017). The datasets analyzed within the scope of this thesis are rather homogeneous, and therefore exhibit a low level of *Variety*.

2.2.1.4 Veracity

The term *Veracity* refers to questions regarding the quality and reliability of data, such as the handling of noisy records and missing values (Arrowsmith et al., 2022), but also the inherent uncertainty of data sources itself (L'Heureux et al., 2017). For example, pseudo-intensity values collected from non-expert citizens naturally contain large uncertainties due to the lack of averaging over multiple observations from a region. On top of that, emotionally captured earthquake witnesses may exaggerate their actual experience, adding additional uncertainty to values at the upper end of macroseismic scales (Bossu et al., 2017). The handling of such reports is a challenge that requires consideration of the *Veracity* of a given dataset.

2.2.1.5 (Hidden) Value

According to L'Heureux et al. (2017), the extraction of *Value* from a dataset is more a desired outcome of Big Data processing rather than a property of the underlying dataset itself. I agree to this interpretation insofar as the general presence of *Value* in and the extraction of *Value* from data are not Big Data exclusive features. For example, in the light of a well established apparatus of physical theories, foremost the theory of elastic wave propagation, the *Value* of a small dataset of high-quality seismic observations for classical tasks, such as tomography, is obvious, and solving the corresponding inversion problem is not necessarily a Big Data challenge.

Following Arrowsmith et al. (2022), the term *Value* in a Big Data context does not refer to the presence of arbitrary, but of subtle or hidden *Value* in a dataset, such as weak earthquake signals hidden in high levels of ambient noise. In such cases, the reliability of a standard method such as LTA/STA (long term average, short term average, Allen (1978)) that is based on the theoretical consideration that earthquake signals cause short-term alterations of the power content in a recorded time series with respect to the long term average, is limited, whereas advanced deep learning-based models excel at retrieving such hidden *Value*. Adapting this perspective, one could argue that the commonly recognized *Value* of pseudo-intensity datasets is actually hidden, because there is no physical theory or standard methodology according to which *Value* could be extracted. Consequently, specialized algorithms are required to extract the hidden *Value* from such data and utilize it in a profitable way. I adapt this interpretation of Big Data related *Value* issues for the rest of this thesis.

2.2.2 Algorithms and resources – deep learning

Among the various algorithms designed for the processing and analysis of Big Data in general and Big Data Seismology in particular, deep learning methods, a specific type of machine learning algorithms, have experienced the greatest popularity, owing to the numerous scientific breakthroughs that they brought. One prime example for the success story of deep learning is an algorithm called AlphaFold (Jumper et al., 2021) that addresses the Big Data problem of protein folding, an open research problem in molecular biology with major implications for, e.g., drug development that has been addressed for over 50 years. The technical challenge in the protein folding problem is to determine the 3D structure of a protein from its sequence of amino-acids. While previous modeling attempts already managed to provide reasonable approximations of 3D structures, the introduction of AlphaFold in 2018 and its successor AlphaFold2 in 2020 abruptly and surprisingly improved the quality of predictions to the level where they are barely distinguishable from the experimentally determined structure (Callaway, 2020). Part of the huge

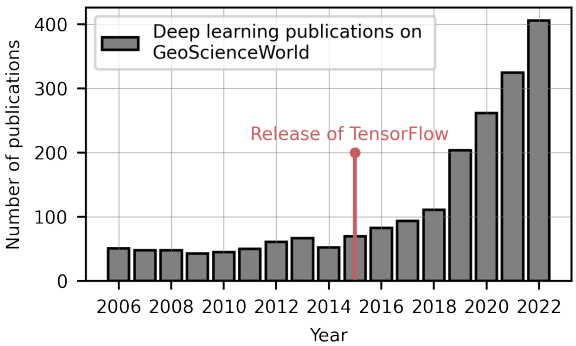

Figure 2.3: Evolution of deep learning related publications in the geosciences. Numbers refer to publications available on the portal GeoScienceWorld (2023) containing the terms "Deep learning" or "Neural Network".

success of AlphaFold can be attributed to the ability of deep learning to autonomously identify and utilize complex, non-linear relations in large datasets that may not be recognized by expert modelers. As a downside, however, traditional deep learning measures are black-box models that do not grant any insights into the learned relationships, i.e., the relations that AlphaFold utilizes to make its predictions are not evident to the modelers and therefore do not grant deeper insights into the mechanisms of protein folding.

While the original concept of artificial neural networks (ANNs), the methodological key component of deep learning, dates back to the 1940s (e.g. McCulloch and Pitts, 1943), and algorithms to efficiently calibrate such models evolved in the 1980s (Rumelhart et al., 1986), deep learning has been experiencing a renaissance in the last decade (Fig. 2.3). This large popularity can be attributed to few recent key developments, foremost the rapidly growing amount of scientific data (Mousavi and Beroza, 2022), but also the considerably eased access to computationally suitable hardware and user friendly software packages. Graphical processing units (GPUs) that were initially designed for visualization of video games turned out to be up to 20 times more efficient than regular Central computing units (CPUs) in terms of calibrating deep learning models (Lecun et al., 2015), and have nowadays evolved into the standard computational resource for deep learning. On the software side, the dissemination of easily accessible deep learning software packages like Tensorflow (Abadi et al., 2015) (Fig. 2.3), has enabled the utilization of deep learning also for users without a distinct background in mathematics or computer science.

Deep learning is utilized extensively over the course of this thesis and is therefore of key significance. Consequently, I want to give a brief introduction to the methodology in the following, using a simplified example of a ground motion model (GMM) (compare section 2.3.2.0.1). GMMs are semi-empirical models that describe the intensity of ground motion at a site, given, in the simplest case, the magnitude of the corresponding earthquake and the distance between the earthquake and the site. The following example largely follows the work of Derras et al. (2014) and paves the way to my own study presented in section 3.

2.2.2.1 Methodology

Artificial neural networks (ANNs) evolved from the attempt to find a mathematical representation for biological information processing in brains, and are the key tool in deep learning (Bishop, 2006). The structure of a simple feed-forward artificial neural network (ANN), consisting of an input layer x_i, a number of hidden units or neurons z_i, and an output unit $\hat{y}$, is depicted in Fig. 2.4. In the following I will interpret the nodes in the input layer x_i as the magnitude M of an earthquake and the distance R between the earthquake and the observation site, and the output node $\hat{y}$ as a prediction of the corresponding ground motion intensity measure PGA. In its final stage, according to e.g. Bishop (2006), the ANN is expected to forward-propagate a given input pair (M, R) as follows:

$$\begin{aligned} \hat{y} &= f\left(\sum_{j=1}^{5} W_j^{(2)} \cdot z_j\right) \\ z_j &= f\left(\sum_{i=1}^{2} W_{ij}^{(1)} x_i\right) \; j \in [1, ..., 5], \end{aligned} \tag{2.1}$$

to eventually estimate a value of the corresponding PGA. Here, $W_{ij}^{(1)}$ and $W_j^{(2)}$ denote the model coefficients, or weights, that determine the values of the hidden units z_i and the output unit $\hat{y}$, depending on the values in the previous layer, respectively. The function f denotes a non-linear activation function, such as the *tanh* or the Rectified Linear Unit ($ReLU$). Without these non-linear transformations the ANN would only be capable of learning linear relations. In its initial state, the ANN's coefficients $W_{ij}^{(1)}$ and $W_j^{(2)}$ are random values, such that for any given input x_i, the output $\hat{y}$ would just be a random number, too, instead of the desired PGA value. Consequently, the ANN must be trained, i.e., the values $W_{ij}^{(1)}$ and $W_j^{(2)}$ must be tweaked to provide a meaningful output. To this end, a training dataset is required that holds a large amount of PGA observations (labels) and corresponding values of the predictive parameters (M, R) such that the output $\hat{y}$ can be evaluated against the actual PGA value via a loss function such as the mean squared error (MSE). Subsequently, the derivative of the loss function with respect to $W_{ij}^{(1)}$ and $W_j^{(2)}$ can be calculated, and coefficients are updated in order to minimize the MSE. This procedure is called backpropagation (Rumelhart et al., 1986), and is the key step that enables the ANN to iteratively learn the relation between (M, R) and PGA, when applied repeatedly on a large training dataset. Modern, deep ANN architectures typically involve multiple hidden layers and millions of free coefficients which allow very close adaption of the model to subtle, non-linear relationships in the training dataset. This high flexibility comes at the price of a tendency to overfit to the data examples in the training dataset, since the number of coefficients in the model typically exceeds the dimensionality of the actual underlying problem by far. To this end, a second so-called validation dataset, one that is strictly separated from the training set, is utilized during the training procedure in order to verify the ANNs ability to generalize to examples outside the training set. In practice, if the ANN's ability to describe the training data increases over time, but the performance on the validation set becomes worse at some point, overfitting is detected and the training procedure is interrupted. At this stage the training of the ANN is assumed to be complete and it can henceforth be utilized as a ground motion model.

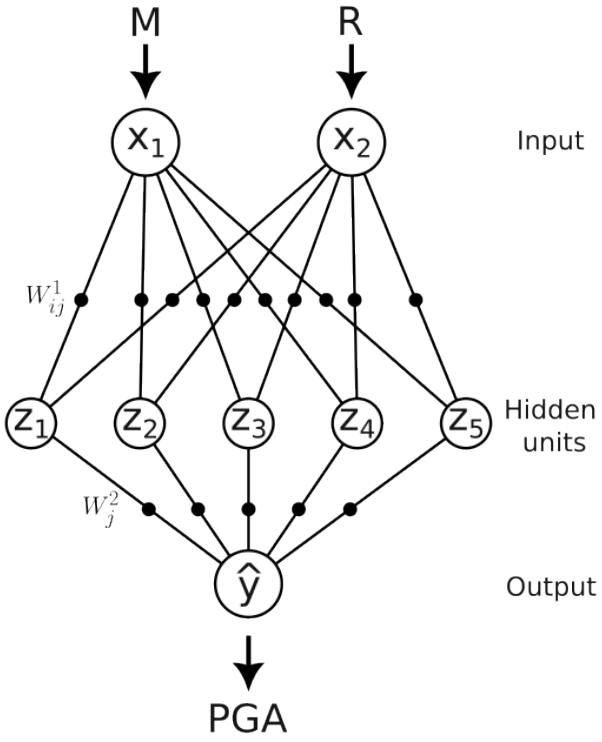

Figure 2.4: Example visualization of an ANN used for the prediction of a ground motion intensity measure (PGA) given the magnitude M of the corresponding earthquake and the distance R of the observation site from the earthquake source.

The example described here concerns a rather simple ANN architecture and refers to the specific case of supervised regression. However, the general concepts are equally applicable

to more elaborate architectures such as the U-Net (Ronneberger et al., 2015) introduced in chapter 3. Although deep learning also finds application for classification and unsupervised learning problems, such approaches are not discussed in further detail because they are not relevant within the scope of this thesis.

2.3 Seismological contributions to earthquake disaster mitigation

2.3.1 Different strategies for different time scales

Over the course of time, four different strategies have emerged to mitigate the consequences of earthquakes: Rapid response to earthquakes (RRE), earthquake early warning (EEW), operational earthquake forecasting (OEF), and probabilistic seismic hazard analysis (PSHA). These four approaches address the mitigation of earthquake losses on four different time scales, ranging from long term assessment of future earthquakes (PSHA), to rapid post-event assistance measures (RRE) (compare Fig. 2.5).

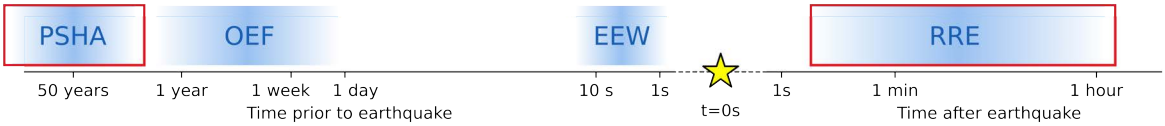

Figure 2.5: Earthquake impact mitigation strategies and their application time windows. PSHA: Probabilistic seismic hazard analysis; OEF: Operational earthquake forecasting; EEW: Earthquake early warning; RRE: Rapid response to earthquakes. The studies presented in this thesis are related to PSHA and RRE (highlighted in red).

A reliable short term earthquake rupture prediction or forecasting framework is one of the great dreams in seismology, because it would give authorities the necessary scope for timely evacuation of at-risk communities. Unfortunately, extensive and diverse research into identification of earthquake precursors, ranging from the analysis of animal behavior (e.g. Savage, 1982; Woith et al., 2018) to geochemical monitoring of gas emissions (e.g. Okabe, 1956), have to date not led to any results that would make a deterministic earthquake prediction framework seem achievable (Jordan et al., 2011).

The most promising candidates closest to this objective are time-dependent, probabilistic forecasting models such as ETAS (epidemic type aftershock sequence), which aim at utilizing the spatio-temporal statistics of a region's seismicity in the past to infer the occurence of earthquakes in the future (Ogata, 1998). Indeed, for several major earthquakes in the past, e.g., the L'Aquila earthquake 2009 in Italy, ETAS models indicated increased probabilities of rupture a day before the actual event (Marzocchi and Lombardi, 2009). However, despite such relative increase, the absolute probability of an earthquake occuring within short periods given by time-dependent earthquake forecasts typically remains low, making this type of model insufficiently reliable for actual initiation of evacuation (Jordan et al., 2011).

Earthquake early warning utilizes the delay between an earthquake rupture and the arrival of seismic waves at a nearby densely populated region to raise an alert shortly before the actual onset of ground shaking (e.g. Cremen and Galasso, 2020). With warning times on the order of a

few seconds, EEW does not provide the opportunity for sophisticated evacuation measures prior to earthquake shaking, but grants the alerted population the chance to seek shelter under, e.g., tables or to move outdoors, to slow down high speed trains, and to initiate shutdown/safety procedures for, e.g., nuclear power plants or medical surgery.

The studies presented within the scope of this thesis are essentially related to the fields of PSHA and RRE, which are elaborated in more detail in the following paragraphs.

2.3.2 Probabilistic seismic hazard analysis

Probabilistic seismic hazard analysis (PSHA) is designed to estimate the expected long-term seismic influence at a particular site of interest. The knowledge thus gained may considerably increase the earthquake resilience of a region way before a significant earthquake occurs, provided that the results are implemented accordingly in carefully respected regional building codes (Kramer, 1996a). The fundamental target of PSHA is to determine the seismic hazard, i.e., the probability of exceeding given levels of ground motion intensity within a given time window (typically decades), at a particular site of interest (Baker et al., 2021b). The PSHA framework comprises two fundamental stages: Source characterization and ground motion characterization. Source characterization deals with the identification of possible earthquake rupture scenarios that may cause significant ground motion at the site of interest. Thereby a complete source characterization comprises an estimation of size and location of future earthquake ruptures, as well as the corresponding probabilities of occurrence for each individual scenario. Modern seismogenic source models address this task via the use of earthquake rupture forecasts (ERFs, e.g., Field et al., 2014; Seebeck et al., 2023), i.e., inventories of possible earthquake ruptures across a regional tectonic fault system that are established prior to actual PSHA calculations. In ground motion characterization, ground motion models (GMMs) are employed to estimate the ground shaking intensity that is effected at a site of interest, given a particular earthquake scenario. Integration of all possible earthquake scenarios potentially relevant for a site of interest, corresponding occurrence rates, effected ground motion intensity, and all associated uncertainties, eventually yields the site-specific hazard curve, i.e. a model that describes the annual rate of exceedance of various levels of ground motion intensity. The hazard curve is the basis on which probabilistic seismic risk calculations can subsequently be performed in order to estimate the risk of damage/collapse of engineered structures and, ultimately, the risk of injury or loss of life for the affected population. Although uncertainties in PSHA frameworks can be large, the fully probabilistic description that it yields is the most expressive form of describing the seismic hazard and is today preferred by experts in earthquake engineering and insurance agencies (e.g. Paté-Cornell, 1994).

2.3.2.0.1 Ground motion modeling Ground motion models (GMMs) are the key tools for ground motion characterization in PSHA. The input to a ground motion model usually comprises a description of the earthquake source, the path between the earthquake and the observation site, and geophysical site properties that may amplify or de-amplify the amplitude of seismic waves. For example, a typical set of predictive parameters consists of the earthquake magnitude M (source), the source-to-site distance R (path), and the average shear-wave veloc-

ity in the uppermost 30 m of the crust V_{S30} (site). Provided with such information, a GMM predicts a probability distribution for a set of ground motion intensity measures (IMs) Y for the respective scenario. It is important to note here that while ground motion is initially recorded by seismometers as a time series, the derived IMs utilized in ground motion modeling, and for other engineering seismology purposes, are scalar features, such as the peak ground acceleration (PGA) or the response spectral acceleration $SA(T)$ (compare section 2.1). Following Kramer (1996b) a GMM can be described in its basic form as follows:

$$\overbrace{\ln Y}^{\text{log-scaled IM}} = \overbrace{C_1 + C_2 M + C_3 M^{C_4}}^{\text{Magnitude scaling}} + \overbrace{C_5 \ln[R}^{\text{Geometric spreading}} + \overbrace{C_6 \exp(C_7 M)]}^{\text{Rupture extent}} + \overbrace{C_8 R}^{\text{Anelastic attenuation}} + \underbrace{f(source)}_{\text{Earthquake source effects}} + \underbrace{f(site)}_{\text{Site specific amplification}} + \underbrace{\varepsilon \cdot \sigma_{\ln Y}}_{\text{Total variability}} , \tag{2.2}$$

where the functional form is designed according to physical principles such as considerations regarding the scaling of energy released at the earthquake source with the effected ground motion (magnitude scaling), geometric spreading, or anelastic attenuation. Here, ε is a standard gaussian random variable and $\sigma_{\ln Y}$ is the variability obtained from the misfit of the model mean and the actual strong-motion observations that are used to empirically determine the coefficients C_i. Because ground motion intensity values are commonly assumed to be log-normally distributed, GMMs are typically not formulated for the respective IM itself but for its logarithm. A fundamental issue with the empirical calibration of ground motion models is the fact, that observations close to the sources of large earthquakes are rare. On the one hand, this is due to the rarity of large earthquakes according to the Gutenberg-Richter magnitude-frequency distribution (Gutenberg and Richter, 1944), on the other hand it is for geometrical reasons that the observation of shaking at large distances is more likely than at short distances, given that sensor networks are often spatially evenly distributed. Subsequently, GMMs are usually least constrained where the predictions are of largest engineering interest (Baker et al., 2021a). It should be noted that the use of physics-based ground motion simulations to generate synthetic data for the calibration of empirical ground motion models, as suggested by, e.g., Paolucci et al. (2021) is a promising candidate to address this problem in the future, provided that the required high quality 3D velocity models are available (Bradley et al., 2017).

Since the first suggestions of a GMM with only three empirically determined coefficients (Esteva and Rosenblueth, 1964), modern GMMs (e.g. Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014; Idrissa, 2014) have evolved into complex statistical frameworks containing dozens of coefficients that are calibrated with large ground motion databases such as NGA-West2 (Ancheta et al., 2014) (compare Fig. 2.6 a and b). Despite this increase in model complexity, no significant reduction of aleatory uncertainty, i.e., reduction of misfit compared to actually observed data has been noted in recent years (Douglas and Edwards, 2016; Strasser et al., 2009) (Fig. 2.6 c), a circumstance that indicates that model uncertainties as defined in classical frameworks are nowadays well characterized. This exhaustion of traditional GMM modeling paradigms has motivated the research into alternative approaches,

that may allow a more detailed evaluation of what parts to the uncertainty are truly aleatory, and which are, in principle, reducible.

One such approach is the waiving of the ergodic assumption, a necessity for the development of early ground motion models when the amount of observations was small. The ergodic assumption states that samples of ground shaking recorded over time at one specific location follow the same distribution as samples collected at the same point in time across multiple sites (Anderson and Brune, 1999). This assumption finds expression in equation 2.2, insofar as the predicted ground motion intensity Y only depends on the distance R between the source and the observation site, and not on the absolute site location. Relaxing the ergodic assumption in ground motion modeling leads to GMMs that more closely resemble the ground motion characteristics of specific regions or even observation sites where observations of ground motion are dense and numerous (e.g. Landwehr et al., 2016). As a downside, nonergodic GMMs come with inherently large uncertainties in locations where observations are sparse. Due to the rapid growth of ground motion databases in recent years, various frameworks for partially nonergodic ground motion modeling have established, e.g., crossed and nested mixed effects regression (MER) as suggested by Stafford (2014) and adopted by Kotha et al. (2016, 2020), and the varying-coefficient model (VCM) as suggested by Bussas et al. (2015) and implemented by Landwehr et al. (2016).

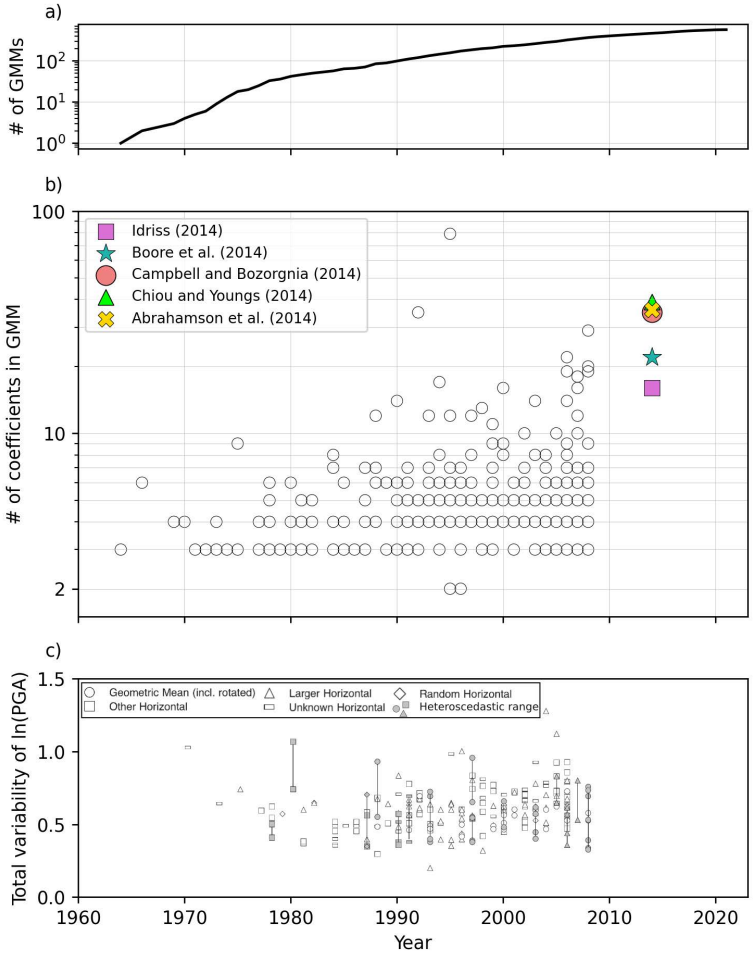

Figure 2.6: Temporal evolution of ground motion modeling. The number of published GMMs in (a) was derived from Douglas (2022). The number of model coefficients shown in (b) was derived from Fig. 2 in Bommer et al. (2010). The total variability of GMMs over time presented in (c) was obtained from Fig. 2 in Strasser et al. (2009). Here, the different symbols refer to the respective type of intensity measure calculation described in the legend. Dark grey markers indicate uncertainty ranges for models with heteroscedastic uncertainty quantification.

The observation by Douglas and Edwards (2016) and Strasser et al. (2009) regarding the stabilization of model uncertainties in GMM development is also one of the key motivations to

employ artificial neural networks (ANNs) to the task of ground motion modeling. As described in section 2.2.2, ANNs excel at identifying subtle and complex non-linear relations in datasets without any instructions regarding the corresponding equational form (Lecun et al., 2015). This property makes ANN-based GMMs interesting baseline candidates for the development of equation-based GMMs, because their high flexibility allows the assumption that the obtained misfit with respect to the ground motion observations is the best possible result that is achievable for a given set of input parameters.

Furthermore ANNs provide a convenient framework to test the predictive power of uncommon predictive parameters for which a relation to the target ground motion IM is presumed, but cannot be expressed in an equational form due to the lack of a physical model (Kong et al., 2019). For example, while Ameri et al. (2017) suggest a simple linear adjustment term to account for stress-drop in a GMM, Trugman and Shearer (2018) demonstrate how the correlation of the stress-drop parameter with ground motion intensity can be incorporated in a GMM, without assuming an equational form of the underlying physical relation.

One of the key insights from ANN-based GMM studies is that already basic ANN architectures with few coefficients yield GMMs that closely match observed ground motion data, and furthermore exhibit physically sound features such as magnitude scaling of distance dependency (e.g. Derras et al., 2014).

As a downside, simple ANN-based models are usually black box models that do not grant any insight into the relations obtained from data, such that only the validation of predictions for well-understood scenarios might be used as an evaluation criterion. Furthermore the aforementioned lack of ground motion observations of large events and at short distances make the corresponding predictions of ANN-based GMMs inherently untrustworthy if no further measures such as data augmentation or the introduction of additional physical constraints are conducted.

Recent advances in both ANN-based and nonergodic ground motion modeling sparked my motivation to merge these two paradigms into a single GMM. The subsequent development of a GMM prototype for the Kanto basin in Japan is presented in chapter 3 of this thesis.

2.3.2.1 Incorporation of directivity effects in ground motion modeling and PSHA

Rupture forward directivity is a phenomenon that occurs when the velocity at which an earthquake rupture progagates along a fault is close to that of the shear waves that are radiated continuously during the rupture process (Somerville et al., 1997). It can therefore be considered a doppler effect, as a result of which the radiated shear waves are compressed in space and time parallel to the rupture propagation, which can lead to extraordinarily intense, pulse-like ground motion at sites close to the earthquake source.

Such pulse-like signals have been observed occasionally as summarized in Yen et al. (2022), e.g., during the 2010 Darfield, New Zealand, earthquake (Bradley, 2012) and the recent Kahramanmaras events in February 2023 (Gabriel et al., 2023). In the classical formulation of GMMs (e.g. equation 2.2), amplification due to directivity effects is not considered explicitly. Consequently, and because directivity is primarily observed close to seismic sources, the predictive uncertainty of GMMs at short source-to-site distances should in principle be underestimated, whereas uncertainty should be inflated at longer distances. However, whether the explicit incor-

poration of directivity effects actually has a substantial practical influence on the variability of GMMs and PSHA results is subject of current research.

A number of directivity models summarized in Spudich et al. (2013) (e.g. Bayless et al., 2020) have been suggested to simulate spatial patterns of directivity related amplification in the near field of large earthquakes, where amplification is assumed to depend on the earthquake's magnitude, the rupture geometry, the relative location of an observation site with respect to the rupture, and the exact nucleation point of the rupture on the fault surface (the hypocenter location). Incorporating such directivity models for ground motion adjustment in PSHA calculations is theoretically straightforward, but would lead to an unaffordable increase in computational demand, because the sampling from the distribution of possible hypocenter locations would result in a several-hundredfold repetition of the actual hazard calculation. Therefore, e.g., Watson-Lamprey (2018) suggests to circumvent this additional computational load by estimating the mean and variability of the spatial amplification pattern of a rupture, obtained from averaging over all possible hypocenter locations, prior to the actual hazard calculation. The resulting so called moment modifiers may then be utilized efficiently in PSHA to modify the "directivity neutral" predicted ground motion and, thus, explicitly consider directivity related amplification.

The implementation of an advanced modifier of moments approach utilizing an artificial neural network, and subsequent efficient application to the 2022 New Zealand National Seismic Hazard Model, is a contribution to a study that was lead-managed by my colleague Dr. Graeme Weatherill, and is presented in chapter 4 of this thesis.

2.3.3 Rapid response to earthquakes

While the previously described strategies focus on the preparation for future earthquakes, the first key step in rapid response to earthquakes (RRE) is the assessment of the intensity and the spatial distribution of ground shaking as well as the resulting impact on the affected communities in near-realtime after an earthquake occurred (Fig. 2.5).

Such a well founded rapid impact assessment (RIA) is an indispensable source of information for the authorities that are responsible for disaster management, as it may considerably assist the decision-making process of how to organize auxiliary measures (e.g. Guérin-Marthe et al., 2021). Although the detailed characteristics of various RIA implementations may differ considerably, the wide majority of systems follow a similar multi-step framework, which is illustrated here briefly using the United States Geological Survey's (USGS) service PAGER (Wald et al., 2010) by example.

In a first step, seismic records are utilized to infer the fundamental earthquake characteristics, most importantly its magnitude and location. Subsequently, a ground motion model is applied to simulate a continuous field of ground motion for the affected region, which is then refined using macroseismic intensities (see section 2.1). The resulting product is referred to as ShakeMap (Wald et al., 2005), which is then fed to the actual PAGER loss assessment system. Together with information on the exposure, i.e., the number and spatial distribution of buildings and population, and vulnerability, i.e., the limited ability of an engineered structure to withstand ground shaking, estimates of casualties and economic losses can be provided within

approximately 30 minutes of any globally occuring earthquake (Wald et al., 2010).

RIA systems are most effective when operated quickly and efficiently after an earthquake, such that timely actions can be initiated to considerably mitigate the final impact. Thereby also rudimentary information with high uncertainty as obtained from, e.g., pseudo-intensity reports (see section 2.1) might prove insightful if it fills the information vacuum in the very first minutes after the shaking. Subsequently acquired information might then be incorporated successively as to receive an increasingly more precise picture of the situation.

In a study presented in chapter 6 in this thesis, I demonstrate how pseudo-intensity reports might be utilized exclusively to rapidly identify whether an ongoing earthquake will have a a substantial societal impact. It can therefore be considered a contribution to the fields of RIA and RRE.

Chapter 3

Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven and Nonergodic Approach

originally published as:

Lilienkamp, H., S. von Specht, G. Weatherill, G. Caire, and F. Cotton (2022). "Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach", Bull. Seismol. Soc. Am. 112 (3), 1565–1582, doi: 10.1785/0120220008

and:

Lilienkamp, H., S. von Specht, G. Weatherill, G. Caire, and F. Cotton (2022). "Erratum to Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach", Bull. Seismol. Soc. Am. 112 (5), 2770–2771, doi: 10.1785/0120220131

The supplemental material to this article is available online at:

https://gsw.silverchair-cdn.com/gsw/Content_public/Journal/bssa/112/3/10.1785_0120220008/2/bssa-2022008_supplement.pdf?Expires=1695636704&Signature=UdMa6LUdV7jsC~UevIx1MuIIE8NMUQh8csRIiRR-dX0Y0CGyLVj28COFa4HUUE99n0RLgvbYLzu3ikmVIcb-Waem8AplOVTejcCKjLizRjsHR7NkmhGvTr8PZRqO6A1RKeS2Un7TGUJxxF1OygyiPDGkg5H-~HnwKSvrwWU8tPcDY1IqegiyVQeJqBISvUVTIdC4EYkusposMHihbGwrXN0PudqGRWvT7LDlj-VHz5~WgmW6WWRbEaKF0zxNgzZltMsRCLvBpGPAJ2c4vNjDyXN0Fs2~JoVpHW1BDsQYnNxKVSyOAw1dN5DgQtIvTC8JHBBCKGG0sKIRta5nH3fg-A__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA

3.1 Abstract

We construct and examine the prototype of a deep learning-based ground motion model (GMM) that is both fully data-driven and nonergodic. We formulate ground motion modeling as an image processing task, in which a specific type of neural network, the U-Net, relates continuous, horizontal maps of earthquake predictive parameters to sparse observations of a ground motion intensity measure (IM). The processing of map-shaped data allows the natural incorporation of absolute earthquake source and observation site coordinates, and is, therefore, well suited to include site-, source-, and path-specific amplification effects in a nonergodic GMM. Data-driven interpolation of the IM between observation points is an inherent feature of the U-Net and requires no a priori assumptions. We evaluate our model using both a synthetic dataset and a subset of observations from the KiK-net strong motion network in the Kanto basin in Japan. We find that the U-Net model is capable of learning the magnitude-distance scaling, as well as site-, source-, and path-specific amplification effects from a strong motion dataset. The interpolation scheme is evaluated using a fivefold cross validation and is found to provide on average unbiased predictions. The magnitude-distance scaling as well as the site amplification of response spectral acceleration at a period of 1 s obtained for the Kanto basin are comparable to previous regional studies.

3.2 Introduction

A ground motion model (GMM) relates the probability distribution of ground motion intensity caused by an earthquake to a set of predictive parameters, for example, earthquake magnitude, distance at which the shaking is observed, and observation site characteristics that influence the local amplification. Commonly, GMMs are equations that are expert designed to represent physical processes causing observable relations in strong motion data, such as decreasing ground motion intensity with increasing distance from the earthquake source. The coefficients of these equations are calibrated via regression analysis using observations from previous earthquakes. The choice of the equation design, that is, which physical aspects are reflected and how, depends largely on the application purpose of the model, resulting in a vast number of proposed models over the last decades (Douglas, 2003). Although the amount of available strong motion observations and, therefore, also the complexity of GMMs have increased significantly since the early stages of ground motion modeling (e.g. Esteva and Rosenblueth, 1964; Trifunac, 1979), associated uncertainties have remained stable during the past 50 yr (Douglas and Edwards, 2016; Strasser et al., 2009), indicating that the development of GMMs might benefit from the complementary incorporation of novel conceptual and methodological modeling paradigms.

One of the new conceptual model paradigms in recent years is the waiving of the ergodic assumption, which was previously a fundamental feature of GMMs. The ergodic assumption states that the distribution of ground motion intensity values observed at a single location over time converges to the same distribution of values sampled across multiple locations (Anderson and Brune, 1999). This assumption was an indispensable feature of GMMs in the past when strong motion observations were scarce. However, in recent years, the amount of ground motion data and computational power increased to levels that allow for the development of more regional

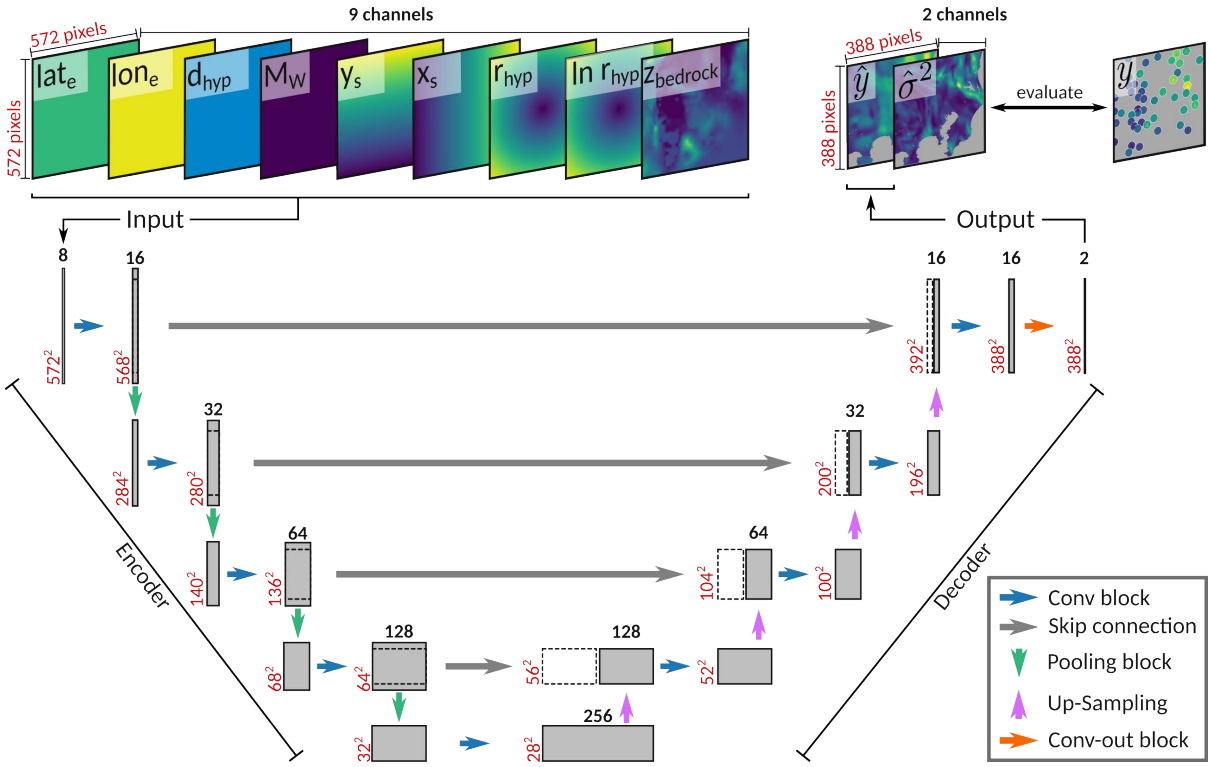

Figure 3.1: The U-Net architecture used throughout this study. The predictive parameters for a single earthquake are provided to the U-Net as input in the form of a stack of maps. The input is processed through the encoder and decoder branches of the U-Net, until eventually the mean $\hat{y}$ and variance $\hat{\sigma}^2$ estimators of an intensity measure (IM) are provided as output. In the training phase, the output is evaluated against sparse measurements y of the IM that are available for this specific earthquake. Bold black numbers indicate the number of features at the respective locations within the U-Net. Red numbers indicate the resolution of the features. The depicted configuration of the input corresponds to the setup used for training with data from the Kanto basin (see section 3.5). The latitude, longitude, and depth of the event hypocenter are denoted as lat_e, lon_e, d_{hyp} respectively, M_W denotes the moment magnitude, x_s and y_s denote the coordinates of each pixel in the input layer, r_{hyp} is the hypocentral distance and $z_{bedrock}$ denotes the depth to seismic bedrock. A more detailed explanation of this figure is provided in the appendix (section 3.A.1). This figure is based on Figure 1 in Ronneberger et al. (2015).

and (partially) nonergodic models. Landwehr et al. (2016), for example, suggest a model with spatially varying coefficients for California that takes into account the absolute locations of earthquake source and observation sites as model parameters, leading to a significant decrease in aleatory uncertainty compared to the previous ergodic models. Following Stafford (2014), a similar strategy is followed by Kotha et al. (2016, 2020), who consider region-specific properties of ground motion as a random effect in a mixed-effects regression and, consequently, provide partially nonergodic GMMs for Europe. Recently, the problem of modeling spatial correlation of ground motion intensity within a nonergodic framework was adressed by Kuehn and Abrahamson (2020).

On the methodological side, another trend has arisen across many fields in recent years: fully data-driven modeling through artificial neural networks (ANNs). ANNs have successfully been applied to ground motion modeling in numerous studies. Derras et al. (2014, 2012) developed

GMMs for shallow crustal earthquakes in Japan and Europe, Pozos-Estrada et al. (2014) modeled inslab and interplate earthquakes in Mexico, and Dhanya and Raghukanth (2018) derived a GMM from the Pacific Earthquake Engineering Research Center - Next Generation Attenuation - West2 Project database (Ancheta et al., 2014). The big advantage of ANNs over model-based approaches is that no a priori definition of a functional form is required. Instead, ANNs autonomously learn the relations between predictive parameters and the target ground motion intensity measures (IMs) from ground motion observations. Even though subjective decisions about the specific design of an ANN may have an impact on the results, the independence from a user-defined functional form opens up entirely new opportunities to reveal previously unknown relations and compare the predictive power of descriptive parameters objectively (e.g. Derras et al., 2012). Although ANNs have shown to provide excellent predictions within the data range that was used to train them, it is also clear that predictions outside this range are highly uncertain due to the lack of physical constraints that would allow such extrapolation. This becomes particularly relevant regarding the persistent lack of strong motion observations of large earthquakes at short distances (e.g. Kong et al., 2019).

Although considerable advances in both nonergodic and fully data-driven ground motion modeling have been made in recent years, there was no attempt to merge these two, so far isolated, concepts within the ground motion modeling community. This study aims at the assessment of opportunities and limitations of a specific ANN architecture, the U-Net (Ronneberger et al., 2015), for the task of ground motion modeling. The U-Net is a well-established architecture that bears the potential to develop GMMs that are both fully nonergodic and fully data-driven. The U-Net was initially developed for image processing tasks such as identifying and differentiating individual cells within tissue specimens, and has since been adopted to a vast range of applications, such as the prediction of the strength of wireless communication signals across maps (Levie et al., 2020), and pansharpening of satellite images (Yao et al., 2018). It is the inherent ability to process data in the form of 2D arrays (maps) that makes this architecture especially interesting for ground motion modeling too, because it offers the opportunity to natively operate on map data and therefore preserves the actual spatial distribution of ground motion observations and seemlessly links it to other geospatial information, for example, lithology and subsurface velocity structure.

We present our study as follows: First, we introduce the methodological adaption of the U-Net architecture to the task of ground motion prediction. Subsequently, we present a proof of concept using a synthetic dataset to demonstrate the functionality of our model. We then show an exemplary application of our method to a subset of the Kiban–Kyoshin (KiK-net National Research Institute for Earth Science and Disaster Resilience, 2019) strong motion dataset in the Kanto basin in Japan and discuss our findings in comparison to local studies.

3.3 Methodology

The U-Net neural network architecture (Ronneberger et al., 2015) is the key component of our proposed GMM. Similar to all supervised machine learning methods, it learns the relation between predictive parameters and target parameters from a large number of examples that

are provided to train the neural network. In ground motion modeling, the target parameter is a ground motion IM, which can be inferred from predictive parameters such as the moment magnitude M_W and the hypocentral distance r_{hyp}. For a more detailed overview on neural networks we refer to the review of Lecun et al. (2015). The U-Net architecture is of particular interest for the task of ground motion modeling, because it is designed to process data that comes in the shape of 2D numerical arrays, that is, maps. Representing ground motion data as maps naturally preserves the absolute locations of observations and their relative position to each other—information that is required to develop a fully nonergodic GMM. Apart from its technical suitability, the U-Net has already been approved in a similar application—the pathloss prediction of wireless telecommunication signals (Levie et al., 2020).

The functionality of a U-Net operating as a GMM is depicted in Fig. 3.1. The predictive parameters for a single earthquake event represent the input features to the U-Net and are provided in the form of a stack of maps covering a predefined area. The input is then processed through all the layers of the U-Net until eventually two output maps are generated. Repeated convolution of the input maps with filter masks is the main operation within the U-Net, causing boundary value loss. The output maps are, therefore, slightly smaller than the input maps and cover a smaller region. Consequently, we distinguish between the U-Net input area and output area (the latter being the actual area of interest for which IM predictions are obtained) from here on. More details on this circumstance and a more detailed description of the implemented U-Net architecture are given in the appendix (section 3.A.1).

If the U-Net is in an untrained state, outputs are just randomly generated maps, because the U-Net just consists of a number of randomly initialized coefficients and has not learned the relation between the predictive parameters and the target IM yet. However, we want the U-Net to provide estimates of the mean $\hat{y}$ and variance $\hat{\sigma}^2$ of the target IM, and therefore need to train the U-Net on a large training dataset that contains examples of event predictive parameters and corresponding IM observations. During training, the loss L_e between the U-Net outputs and the observations y_e for an event e with N_{obs} observations is evaluated as the negative log likelihood of the normal distribution:

$$L_e = \sum_{n=1}^{N_{obs}} \ln \hat{\sigma}_{e,n}^2 + \frac{(y_{e,n} - \hat{y}_{e,n})^2}{\hat{\sigma}_{e,n}^2}. \tag{3.1}$$

The U-Net learns iteratively how to relate the input predictive parameters to the IM observations, that is, the loss is iteratively minimized through the gradient descent method Adam (Kingma and Ba, 2015), with the gradient efficiently implemented by backpropagation (Rumelhart et al., 1986). By the choice of the loss function and the optimization routine, training of the U-Net is equivalent to nonlinear least-squares regression, as it is commonly used in the most mixed-effects models. After each epoch of training, that is, one episode of training during which the U-Net sees all data examples in the training set, its ability to generalize to previously unseen events in a second dataset, the validation set, is evaluated. After a number of epochs, the loss on the validation dataset does not decrease any further, so the training is assumed to be complete, and the U-Net can henceforth be used as a GMM. The exact technical configuration of the training procedure followed throughout this study is given in the supplemental

material S1, available to this article. Although the observations y are point-wise measurements, the U-Net predictions $\hat{y}$ and $\hat{\sigma}^2$ are continuous maps. This means first that the loss function is only evaluated at those locations where actual observations are available, and second that the U-Net automatically interpolates the learned attenuation relation from the observation locations across the U-Net output area. The quality of this interpolation is examined in more detail in the following sections.

A clean separation between training and validation data could be achieved by splitting a strong motion dataset strictly according to both events and sites with seismic stations. However, huge portions of the dataset would be lost, because records of "training events" on "validation stations" and vice versa would be discarded. To address this issue, we developed the following training strategy:

1. The data is separated into training and validation events.

2. Stations (sites where seismic stations are located) are randomly grouped into a number N_{chu} of "chunks".

3. One U-Net is trained per station chunk, in which the respective station chunk is excluded from the training procedure and the remaining four chunks are used for both training and validation on the training and validation events, respectively. The fifth chunk is then used to evaluate the capability to generalize to new locations after training.

4. Training of each U-Net is conducted N_{init} times to sample the variability that is caused by the random initialization of coefficients in the U-Net prior to training.

The final predictions for the mean and variance of the target IM ($\hat{Y}_e$ and $\hat{\Sigma^2}_e$, respectively) for the event e are subsequently obtained via ensemble averaging the mean and variance predictions of the separate U-Nets ($\hat{y}_e^{ij}$ and $\hat{\sigma}_e^{ij^2}$, respectively). The ensemble prediction is then a mixture distribution, the mean and variance of which are obtained based on the law of total expectation and the law of total variance, respectively (Blitzstein and Hwang, 2014):

$$\hat{Y}_e = \frac{1}{N_U} \cdot \sum_{i=1}^{N_{chu}} \sum_{j=1}^{N_{init}} \hat{y}_e^{ij} \tag{3.2}$$

$$\hat{\Sigma}_e^2 = \frac{1}{N_U} \cdot \sum_{i=1}^{N_{chu}} \sum_{j=1}^{N_{init}} [\hat{\sigma}_e^{ij^2} + \hat{y}_e^{ij^2}] - \hat{Y}_e^2, \tag{3.3}$$

in which $N_U = N_{chu}\ N_{init}$ equals the total number of U-Nets.

We note that the suggested procedure is not a classic example for the separation of training and validation data. However, we consider this adoption to the particular challenges of strong motion datasets an appropriate compromise between methodological accuracy and optimal data usage.

3.4 Proof of concept with synthetic data

Before we apply the U-Net methodology to derive a fully nonergodic GMM from real data, we demonstrate its abilities to:

- Learn the relationship between an IM and predictive parameters (such as M_W and r_{hyp})
- Learn site-, source-, and path-specific amplification from site and event coordinates
- Perform reasonable interpolation between observation sites

We will present the case of a synthetic dataset including measurements of an IM that depends on hypocentral distance, magnitude, and site-specific amplification. Similar studies for datasets containing source- and path-specific amplification are conducted alike and are provided in the supplemental material S2.

3.4.1 Dataset generation

According to the input and output resolutions of the U-Net architecture (see the appendix in section 3.A.1), we define two regular grids of sites for the U-Net input area (572 × 572 pixels) and the U-Net output area (388 × 388 pixels). We choose a grid spacing of 500 m, resulting in extents of $(286\,\text{km})^2$ and $(194\,\text{km})^2$ for the input and output areas, respectively (Fig. 3.2 a).

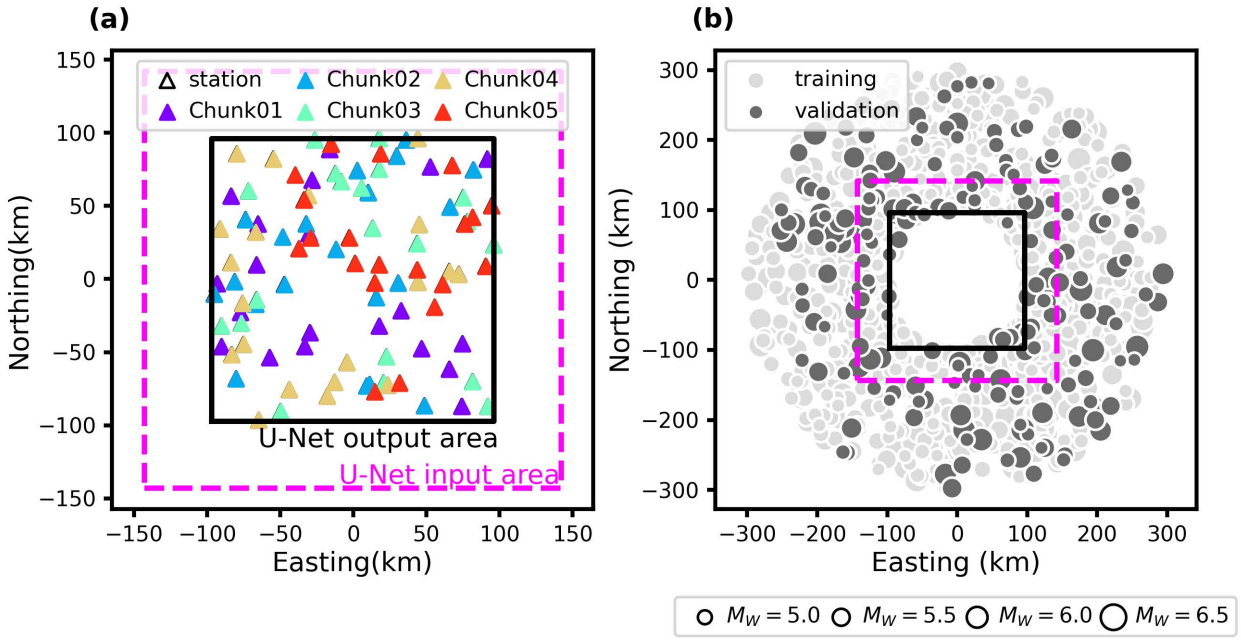

Figure 3.2: Data generation setup for the synthetic experiments. (a) The geometric setup of the area of interest and the locations of randomly sampled stations. (b) The spatial distribution of the synthetic earthquake catalog with respect to the area of interest.

We then randomly define $N_s = 100$ sites from the output area as observation locations with seismic stations. We simulate $N_e = 1000$ events with uniformly distributed moment magnitudes $5.0 \leq M_W \leq 6.5$. The spatial distribution of events with respect to the U-Net input and output

areas is presented in Fig. 3.2 b. We use the GMM by J. X. Zhao et al. (2016c) (from here on Zhao16ASC) as the basis to simulate values of the natural logarithm of the acceleration response spectrum at period $T = 1s$ ($\ln(Sa(T = 1s))$) for all pairs of events and sites on the regular grid within the U-Net output area. Thereby, we only consider the scalings with M_W and r_{hyp} and keep all other predictive parameters required by Zhao16ASC fixed. The synthetic dataset of ground motion observations u follows from

$$u = \bar{u}(M_W, r_{hyp}) + \delta_{S2S_s} + \varepsilon. \tag{3.4}$$

Here $\bar{u}$ describes the mean prediction of Zhao16ASC, δ_{S2S_s} is a predefined site-specific amplification function, and ε is a zero-mean Gaussian with a standard deviation of 0.1 that accounts for aleatory uncertainty. The term δ_{S2S_s} is defined as the portion of the site-specific amplification that is not captured by predictive parameters such as the shear wave velocity of the upper 30 m of the Earth's crust (V_{S30}) and can only be obtained from repeated measurements at a site (Al Atik et al., 2010). To model δ_{S2S_s} we first initialize an uncorrelated Gaussian random field and subsequently smooth it with an anisotropic Gaussian kernel with zero mean and covariance:

$$C = \begin{pmatrix} \sigma_E^2 & 0 \\ 0 & \sigma_N^2 \end{pmatrix}, \tag{3.5}$$

where the standard deviations σ_E and σ_N scale the correlation in east-west and north-south directions, respectively. The resulting δ_{S2S} field is rescaled to a standard deviation of 0.2 and presented in Fig. 3.3. The choice of a Gaussian kernel function is in agreement with various semivariogram analysis studies that investigate correlation of ground motion features (e.g. Jayaram and Baker, 2009; Loth and Baker, 2013; Markhvida et al., 2018). The choice of values of $\sigma_E = 26$ km and $\sigma_N = 8$ km are in agreement with the findings of Sgobba et al. (2021), who report an isotropic range of correlation (the distance at which only 5 % correlation remain, approximately twice the standard deviation) of 25 km for δ_{S2S}. Although radial isotropy, i.e. $\sigma_E = \sigma_N$, is commonly assumed in ground motion modeling for reasons of model simplicity, we choose anisotropic δ_{S2S} as it is more realistic and to test whether the U-Net can capture such features.

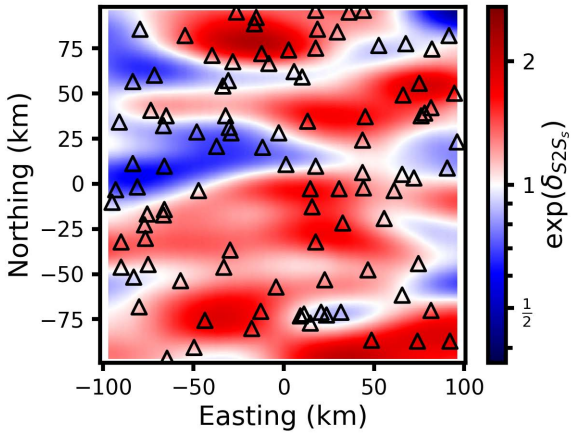

Figure 3.3: The predefined spatially correlated random field representing $\exp \delta_{S2S_s}$ used in the synthetic experiments. The black triangles indicate the locations of seismic stations.

The final dataset does not contain all simulated IM values on the regular grid. Instead, 25 observations are randomly selected from the 100 station locations. The synthetic dataset therefore consists of 25,000 records from 1000 events recorded at 100 stations.

3.4.2 U-Net training (synthetic)

We generally follow the training procedure described in section 3.3. The synthetic dataset is randomly split into N_e^{tr}=800 training and N_e^{val}=200 validation events. The 100 stations are separated into N_{chu}=5 chunks and we consider $N_{init} = 5$ different random initializations per chunk. The derived GMM:

$$\hat{U} = f(M_W, r_{hyp}, \ln(r_{hyp}), x_s, y_s) \tag{3.6}$$

is an ensemble of 5 chunks $\times$ 5 initializations = 25 U-Nets. Parameters x_s and y_s denote coordinates of each grid point in the U-Net input area. We emphasize that we use no predictive parameter related to δ_{S2S_s}, which can subsequently only be learned as a function of x_s and y_s.

3.4.3 Evaluation of U-Net predictions (synthetic)

In a first step, we need to verify that our GMM has learned the fundamental scaling relations of u with M_W and r_{hyp}. Because of the abstract representation of information inside the U-Net, we cannot disaggregate the individual scalings with the separate parameters. However, the learned scaling $\hat{U}_{basic}$ of u with M_W and r_{hyp} can be approximated by:

$$\hat{U}_{basic}(M_W, r_{hyp}) = \hat{U} - \hat{\delta}_{S2S_s}, \tag{3.7}$$

where $\hat{\delta}_{S2S_s}$ is the approximated site amplification

$$\hat{\delta}_{S2S_s} = \frac{1}{N_e^{val}} \sum_{e=1}^{N_e^{val}} \hat{U}_e - \bar{u}_e. \tag{3.8}$$

In Fig. 3.4 we present the comparison between the predefined relationship $\bar{u}(M_W, r_{hyp})$ and the estimated relationship $\hat{U}_{basic}(M_W, r_{hyp})$ approximated from predictions for validation events. We observe an overall good agreement indicating that the attenuation relation has been learned successfully. In a second step we compare the estimated site amplification $\hat{\delta}_{S2S_s}$ to the predefined δ_{S2S_s} to assess whether the site amplification was learned successfully at the station locations and interpolated with acceptable precision across the U-Net output area. Visual inspection of the learned site effect $\hat{\delta}_{S2S_s}$ (Fig. 3.5 a) and comparison to the predefined site effect (Fig. 3.3) indicate a high similarity. From the distribution of residuals between the predefined and the estimated site amplification depicted in Fig. 3.5 b, one can see that at those locations where observations were made, the

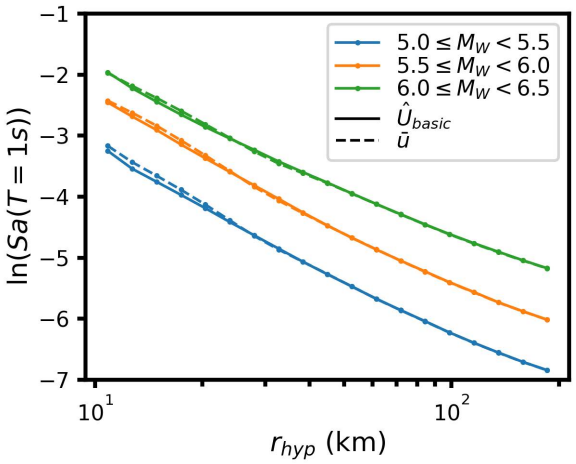

Figure 3.4: Comparison of the predefined magnitude-distance scaling $\bar{u}(M_W, r_{ryp})$ (dashed lines) and the estimated relation $\hat{U}_{basic}(M_W, r_{ryp})$ (solid lines). The lines are derived via binning and averaging $\ln(Sa(T = 1s))$ predefined and predicted values for the validation events according to r_{hyp} and M_W. For this comparison, the entire maps of $\hat{U}_{basic}$ and corresponding $\bar{u}$ values are used.

learned site amplification closely resembles the predefined one, whereas the estimations are less precise, though unbiased, in interpolated areas. For comparison purposes we also extract the predefined δ_{S2S_s} at station locations and perform bicubic spline interpolation to reconstruct the

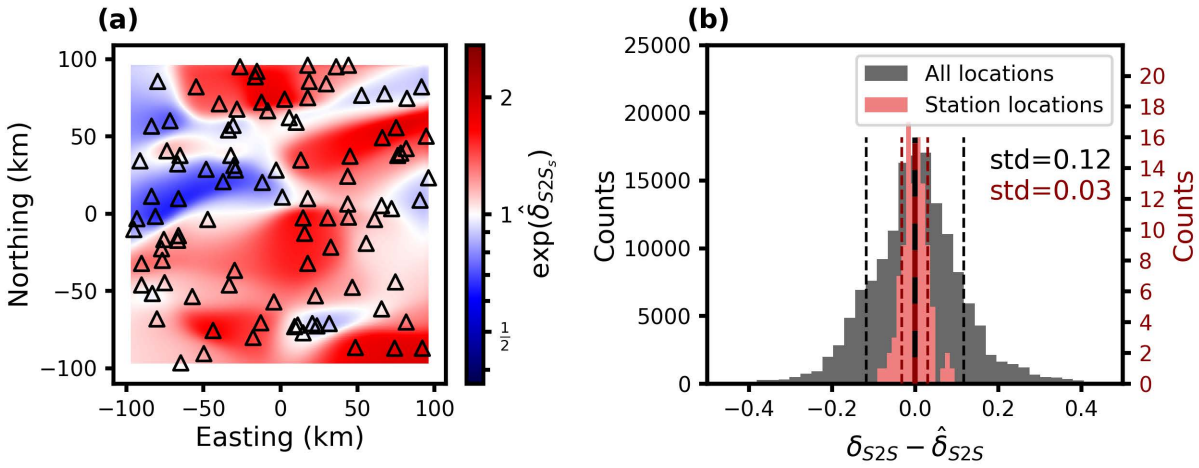

Figure 3.5: (a) Estimated site amplification map $\hat{\delta}_{S2S_s}$ and (b) distributions of residuals $\delta_{S2S_s} - \hat{\delta}_{S2S_s}$ with respect to the predefined site amplification. Solid and dashed vertical lines in (b) indicate distribution means and standard deviations (std), respectively. From the visual comparison of (a) to Fig. 3.3 and the distribution of residuals in (b) one can see that the site amplification is accurately retrieved at station locations. Predictions in interpolated areas are less accurate, but seem reasonable from visual inspection and appear to be unbiased.

site amplification field. The standard deviation between this reproduction and the predefined field amounts to 0.11, very similar to the value of 0.12 that is obtained from the U-Net reproduction $\hat{\delta}_{S2S_s}$. We note that although the spatial anisotropy introduced during generation of δ_{S2S_s} clearly emerges also in the interpolated site amplification $\hat{\delta}_{S2S_s}$, no advantage over the bicubic spline interpolation can be stated in terms of misfit to the predefined δ_{S2S_s}.

Similar studies concerning the recovery of source-location specific variations δ_{L2L_l} and path-specific amplification $\delta_{P2P_{sl}}$ have been conducted successfully and are presented in the supplemental material S2. In summary, we come to the conclusion that the proposed U-Net methodology is capable of extracting and interpolating the scaling of an IM with magnitude and distance, as well as site-, source-, and path-specific amplification from a strong motion dataset.

3.5 Application to the Kanto basin

In this section, we present an exemplary application of our U-Net GMM to the Kanto basin area, Japan. We will first describe the used dataset and the U-Net training procedure, followed by the discussion of an example prediction and a more general evaluation of the model performance on the entire dataset.

3.5.1 Kanto basin dataset

We chose the Kanto basin as our study area due to its high seismic activity and high station density of the regional KiK-net strong motion network. Operating for over 20 yr, KiK-net provides a substantial strong motion dataset necessary to constrain the U-Net methodology.

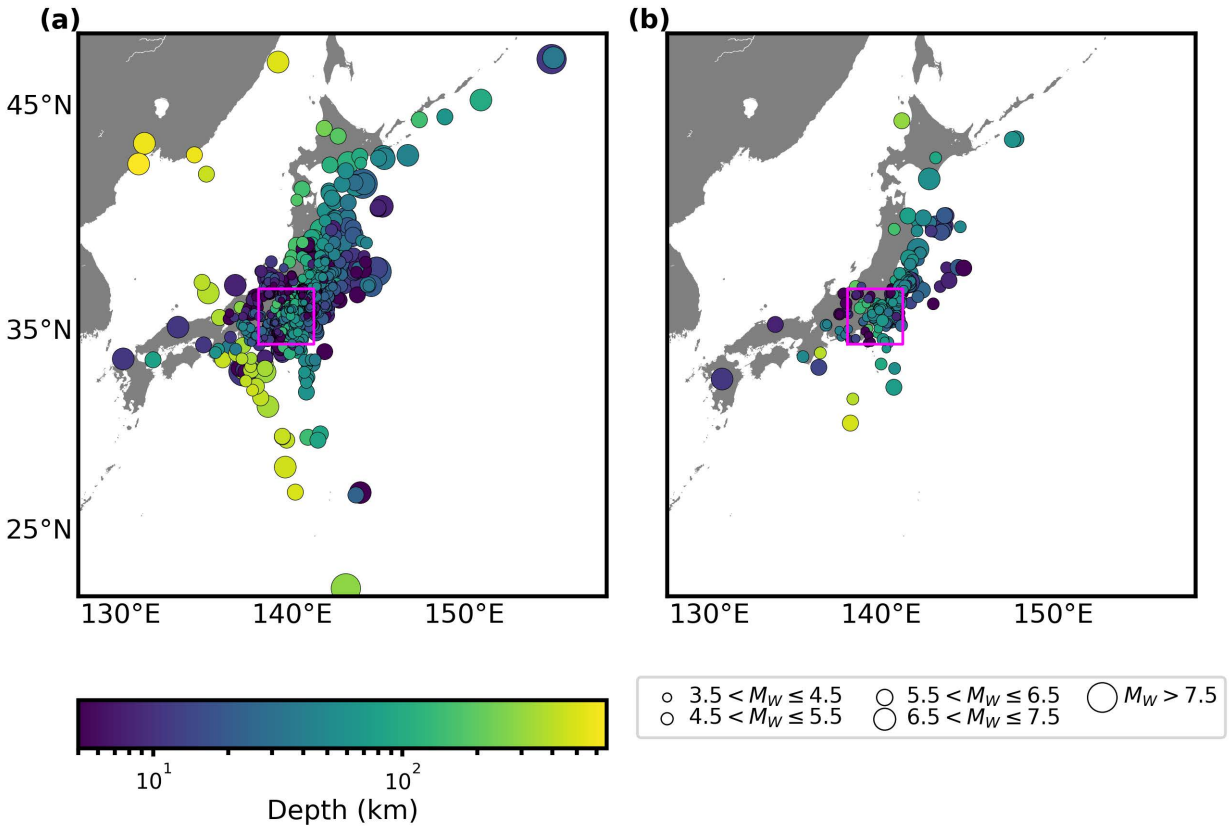

Figure 3.6: (a) Training and (b) validation events selected from the Bahrampouri et al. (2021) database. The magenta square indicates the U-Net input area.

We use the strong motion database by Bahrampouri et al. (2021) and select a subset of records from the Kanto basin, in total 46,191 records of $N_e = 2864$ events (Fig. 3.6) recorded at 65 stations (Fig. 3.7). Thereby, we use records from events of all tectonic region types. The average interstation distance of the selected KiK-net stations is about 94 km, with almost 20% of the interstation distances being less than 50 km. We select the geometric mean of the two horizontal components of $\ln(Sa(T = 1s))$ as our target IM due to its wide application in probabilistic seismic hazard analysis (PSHA). The distribution of the records with respect to M_W and r_{hyp} is presented in Fig. 3.8. The locations of the U-Net input (572×572 pixels) and output (388×388 pixels) areas in the Kanto basin are presented in Fig. 3.7. The corresponding regular grids have a spacing of 500 m and encompass regions of $(286\,\text{km})^2$ and $(194\,\text{km})^2$, respectively. We use a collection of nine predictive parameters: M_W, r_{hyp}, $\ln r_{hyp}$, event longitude lon_e, event latitude lat_e, hypocentral depth d_{hyp}, coordinates of each grid point x_s, and y_s, and the depth to seismic bedrock $z_{bedrock}$ (compare Fig. 3.1). We expect the U-Net to learn site amplification as a function of x_s, y_s, and $z_{bedrock}$, source-location-specific variability from lon_e and lat_e, and path-specific amplification from x_s, y_s, lon_e, and lat_e. The $z_{bedrock}$ data was downloaded from the J-SHIS

map web service (see section 3.8).

3.5.2 U-Net Training (Kanto basin)

We generally follow the training procedure described in section 3.3. We split the dataset by events, in which events before the year 2015 are used for training and after for validation. The 65 stations are divided into N_{chu}=5 chunks and we consider N_{init}=10 different random initializations per chunk. The arrangement of stations into chunks is illustrated in the supplementary material S3.

We derive 5 chunks × 10 initializations = 50 U-Net estimators:

$$\hat{y}^{ij} = f(M_W, r_{hyp}, \ln(r_{hyp}), lon_e, lat_e, d_{hyp}, x_s, y_s, z_{bedrock}) \quad \begin{matrix} i \in \{1, 2, \ldots, 5\} \\ j \in \{1, 2, \ldots, 10\}. \end{matrix} \tag{3.9}$$

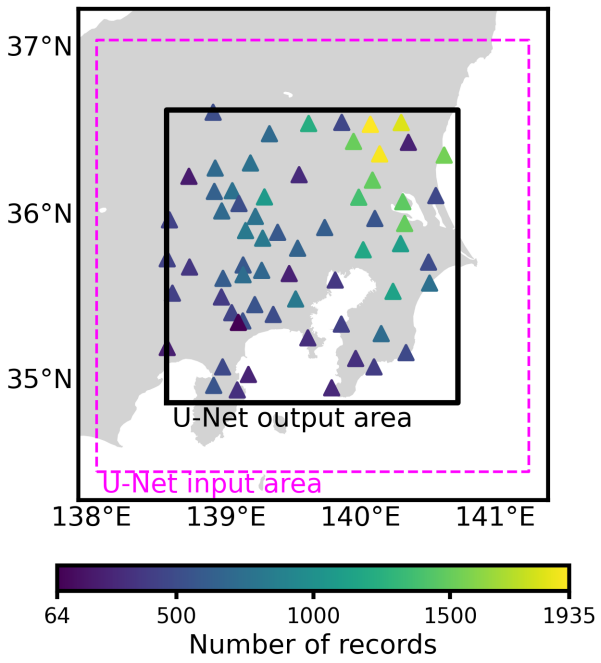

Figure 3.7: Station coverage of the KiK-net network in the Kanto basin. Triangles indicate the locations of seismic stations. The color indicates the number of available strong motion records for $\ln(Sa(T = 1s))$ in the Bahrampouri et al. (2021) database.

We perform ensemble averaging over all 50 U-Nets $\hat{y}^{ij}$ according to equations (3.2) and (3.3) to derive the full ensemble estimator $\hat{Y}$ (compare Fig. 3.9) as our final GMM for the Kanto basin. Because $\hat{Y}$ is trained on data from all stations in the Kanto basin, it is not suited for evaluation of the quality of interpolated values. Therefore, we also derive partial ensemble estimators $\hat{\mathcal{Y}}^i$, which only average over those respective 10 U-Nets for which the same ith station chunk was excluded from the training procedure and is used to evaluate the ability to interpolate spatially after training.

3.5.3 Evaluation of U-Net predictions (Kanto basin)

In Fig. 3.10 we present an example prediction of our full ensemble estimator for an event from the validation set. A comparison between the mean prediction $\hat{Y}$ and the available observations y for this event is shown in Fig. 3.10 a. From visual inspection, we notice an overall satisfying agreement. The corresponding predicted standard deviation $\hat{\Sigma}$ is given in Fig. 3.10 b. One should recall that $\hat{\Sigma}$ depends on the same set of input variables as the mean prediction $\hat{Y}$, and is therefore a spatially varying and event specific quantity. One can identify the general trend that $\hat{\Sigma}$ is small in regions where station density is high and spatial variability in the mean prediction is low. This distribution indicates that the individual U-Net estimators $\hat{y}^{ij}$ learn the same scatter of observations

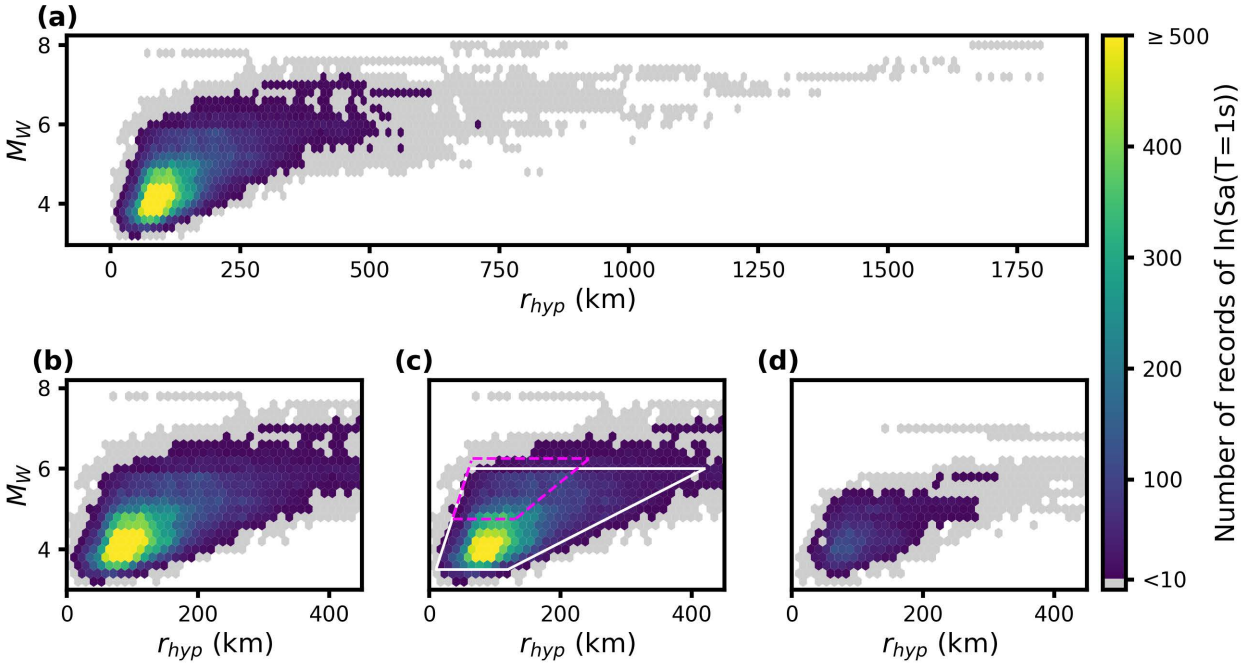

Figure 3.8: Distribution of all KiK-net records from the Kanto basin according to (a,b) M_W and r_{hyp} available in the Bahrampouri et al. (2021) database. The entire dataset is split into a set of (c) training events and (d) validation events. Although all available records were used for either training or validation, we assume that predictions from the ground motion model (GMM) derived in this study for M_W, r_{hyp} combinations with less than 10 records in the training set are not reliable. The white polygon in (c) indicates our suggested range for application. The magenta polygon indicates the range for which comparison to conventional GMMs is presented in the Discussion (section 3.6).

at station locations, but interpolate slightly differently due to their random initialization of coefficients. Consequently, $\hat{\Sigma}$ can, at least, be interpreted as a comprehensible measure for the relative confidence of the GMM in its prediction. More examples and comparisons to the predictions of the ergodic GMMs of J. X. Zhao et al. (2016a,b,c) are available in the supplementary material S4. The comparison of predictions $\hat{Y}$ with observations y from validation events is illustrated in Fig. 3.11 and is quantified using the root mean square error (rmse = 0.48 in natural log units). The use of the negative log-likelihood loss function given in equation (3.1) provides that $\hat{\Sigma}$ should be a proxy for the scatter in observations at station locations. To verify the success of this strategy, we first calculate residuals between observations from validation events and the corresponding mean predictions of our full ensemble GMM. We then standardize these residuals by dividing each individual residual by its predicted standard deviation:

$$\tilde{\Delta}_n = \frac{y_n - \hat{Y}_n}{\hat{\Sigma}_n} \qquad n \in \{1, \ldots, N_{obs,val}\}, \tag{3.10}$$

in which $N_{obs,val}$ denotes the number of observations from validation events. If $\hat{\Sigma}$ accurately described the scatter in observations, then the standardized residuals $\tilde{\Delta}_n$ should be standard normally distributed.

The actual distribution of $\tilde{\Delta}_n$ together with the targeted standard normal probability density function (PDF), is depicted in Fig. 3.12. We notice that the actual standard deviation of 0.85 is

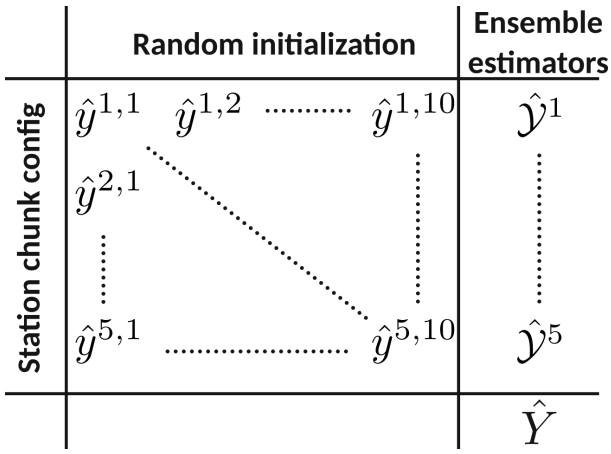

Figure 3.9: Schematic representation of ensemble averaging of mean predictions. The full ensemble estimator $\hat{Y}$ is derived via averaging over all individual U-Net estimators $\hat{y}^{ij}$. Partial ensemble estimators $\hat{\mathcal{Y}}^i$ are derived via averaging over those subsets of U-Nets for which the same ith station chunk was excluded from training and can thus be used to evaluate interpolated predictions. Partial ensemble estimators are required to evaluate predictions in interpolated regions of the area of interest. Ensemble averaging of individual standard deviation estimators $\hat{\sigma}^{ij}$ into partial ensembles $\hat{\varsigma}^i$ and the full ensemble $\hat{\Sigma}$ is performed accordingly.

$\approx$15 % smaller than the targeted value of 1. Because $\hat{\Sigma}_n$ represents the denominator in equation (3.10), this indicates an overestimation of the uncertainty in predictions at station locations.

In addition to the performance at station locations, we also verify the model's ability to interpolate between stations. For this purpose, we use the partial ensemble estimators $\hat{\mathcal{Y}}^i$, for which the ith station chunk was excluded from training and can thus be used to evaluate interpolated predictions. Each of the five $\hat{\mathcal{Y}}^i$ can be compared with the observed data in four different categories: 1) training events recorded on training stations, 2) validation events recorded on training stations, 3) training events recorded on validation stations, 4) validation events recorded on validation stations. The rmses between observations and predictions averaged over the five $\hat{\mathcal{Y}}^i$ are presented in Table 3.1. As expected the smallest rmse is obtained for those observations that were used during training. A similar value is obtained for records from validation events observed at training stations, indicating successful generalization to events outside the training dataset. The rmse increases significantly for validation stations, because these predictions include the additional error due to interpolation. The rmse at interpolated sites obtained from $\hat{\mathcal{Y}}^i$ can be interpreted as an upper bound estimate of the corresponding rmse of the full ensemble estimator $\hat{Y}$, because it is trained using all stations and is, therefore, better constrained due to the higher station density. In a final step, we evaluate the precision of the predicted standard deviations $\hat{\varsigma}^i$ from the partial ensemble estimators. As for the full ensemble estimator, we calculate standardized residuals between all observations from both training and validation events with respect to all partial ensemble estimators:

$$\tilde{\Lambda}_n^i = \frac{y_n - \hat{\mathcal{Y}}_n^i}{\hat{\varsigma}_n^i} \qquad n \in \{1, \ldots, N_{obs}\}, \tag{3.11}$$

in which N_{obs} denotes the number of observations. The comparison of the distribution of $\tilde{\Lambda}_n^i$ to the targeted standard normal PDF is depicted in Fig. 3.13. From Fig. 3.13 a, b, we conclude that the predictive uncertainty at the training stations closely resembles the data scatter and, therefore, represents a reliable uncertainty measure. At interpolated sites, $\tilde{\Lambda}_n^i$ is zero centered and well described by a normal distribution (Fig. 3.13 c, d). However, we observe that the standard deviation of $\tilde{\Lambda}_n^i$ is over 40 % larger than the targeted value of 1, indicating a significant

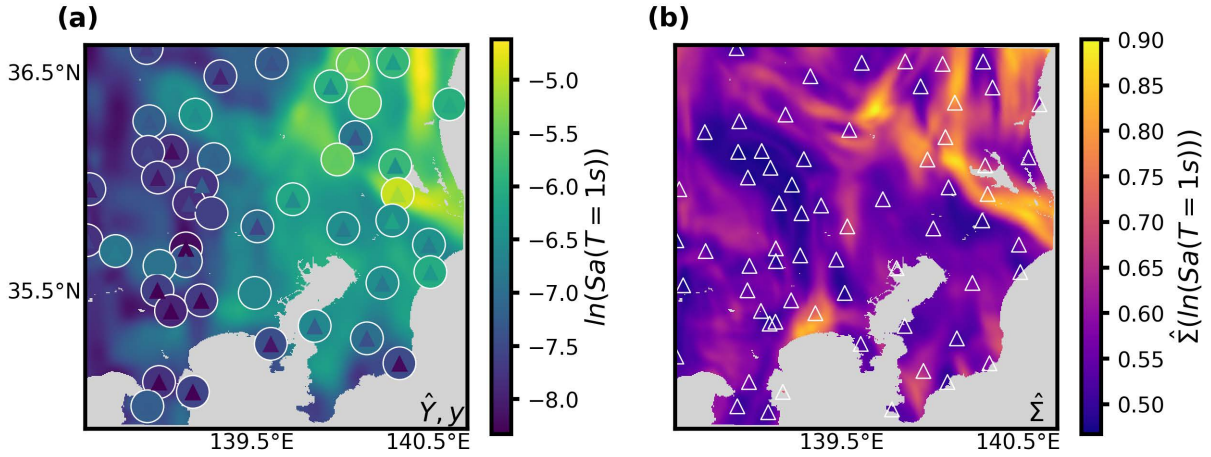

Figure 3.10: Example prediction of our GMM for a magnitude 5.2 validation event at longitude 140.98° and latitude 36.86°. The plotted raster data and the circles in (a) show the mean prediction $\hat{Y}$, in which the circles represent the values at exactly those locations, where observations y (triangles) are available. The root mean square error (rmse) amounts to 0.41 in natural log units. The raster data in (b) shows the predicted standard deviation $\hat{\Sigma}$. The white triangles indicate the locations of all stations in the dataset, including those that have no record for this specific event.

underestimation of the predicted standard deviation.This means that although the relative tendency to assign larger uncertainties to less constraint estimates is given, predictive uncertainty at interpolated sites is quantitatively only a lower bound uncertainty estimate. The precision in Fig. 3.13 b is actually higher than the one obtained for the full ensemble estimator presented in Fig. 3.12. This is due to the fact that when ensemble averaging over all 50 U-Nets at a specific station site, those 10 U-Nets for which this site was in the validation chunk contribute less well-calibrated predictions.

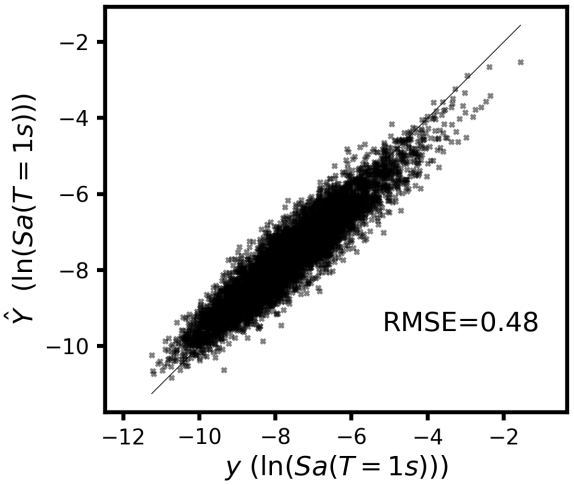

Figure 3.11: Full ensemble mean predictions $\hat{Y}$ versus observations y of $\ln(Sa(T = 1s))$ for all records from validation events.

3.6 Discussion

3.6.1 Comparison to ergodic GMMs

Because the presented U-Net methodology is a fully data-driven approach without any physical constraints on the model design, we want to validate that the fundamental functionality of our full ensemble estimator $\hat{Y}$ is comparable to conventional GMMs. We compare our prediction for the example event presented in Fig. 3.10, as well as the overall learned magnitude and distance

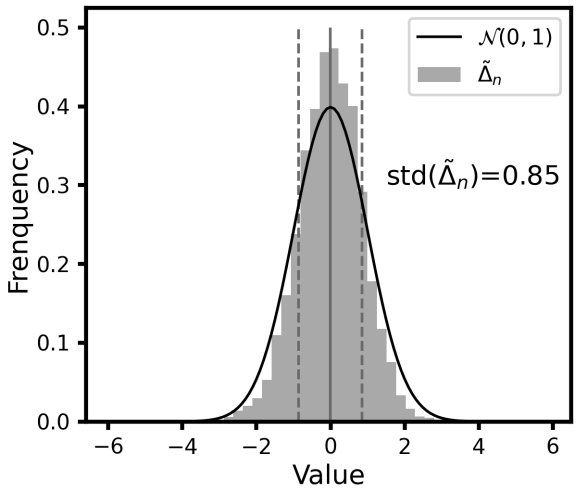

Figure 3.12: Comparison of the distribution of standardized residuals $\tilde{\Delta}_n$ with the targeted standard Gaussian. Solid and dashed vertical lines indicate the empirical mean and standard deviation (std) of standardized residuals $\tilde{\Delta}_n$, respectively.

scalings to the ergodic GMMs for Japan developed by J. X. Zhao et al. (2016a,b,c) (hereafter Zhao16 GMMs). We emphasize, that the purpose of these comparisons is not to identify superiority of one model over the other. Instead, we want to demonstrate that the basic functionality of our model is comparable to a conventional GMM that is derived from a well-established and mature methodology. A detailed, quantitative comparison would be inappropriate for a number of reasons: first of all, there is no standardized procedure to compare nonergodic to ergodic GMMs, because they make fundamentally different assumptions on the distribution of ground motion data. Furthermore, our model is trained on data from the Kanto basin only, whereas the Zhao16 models are calibrated using data from entire Japan. Another aspect is that in this study we do not exhaust the full power of the Zhao16 models, because we disregard its volcanic path terms and always use r_{hyp} instead of the suggested rupture plane distance r_{rup}, if 3D rupture models are available. Finally, we derive the site classes that are required as site parameters for the Zhao16 models indirectly via conversion of the V_{S30} map accessed via the J-SHIS web map service (see section 3.8), according to J. X. Zhao et al. (2015). However, we consider this use of Zhao16 models appropriate for the qualitative comparison that we attempt.

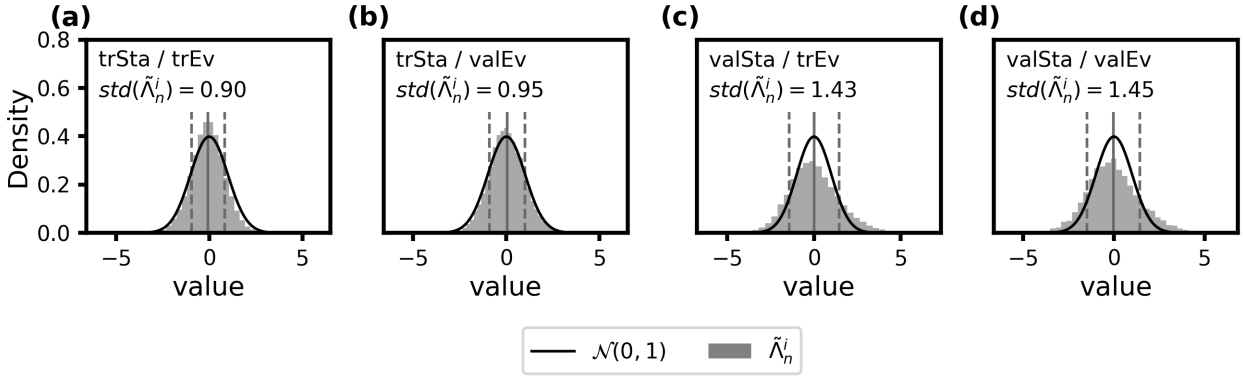

Figure 3.13: Standardized residuals $\tilde{\Lambda}_n^i$ with respect to partial ensemble estimators $\hat{\mathcal{Y}}^i$ at (a) training station locations for training events, (b) training stations for validation events, (c) validation stations for training events, and (d) validation stations for validation events. Solid and dashed vertical lines indicate the empirical mean and standard deviation (std) of standardized residuals $\tilde{\Lambda}_n^i$, respectively.

The prediction of the Zhao16 GMM for the event presented in Fig. 3.10 is given in Fig. 3.14. In comparison to the prediction $\hat{Y}$ derived in this study (Fig. 3.10 a), we note that the range of IM values and the large scale attenuation with distance are similar.

Table 3.1: Average root mean square errors ± 1 standard deviation between observations y and predictions of the five partial ensemble estimators $\hat{\mathcal{Y}}^i$.

	training stations	validation stations
training events	0.428 ± 0.005	0.809 ± 0.082
validation events	0.454 ± 0.003	0.821 ± 0.110

However, we also note that $\hat{Y}$ adapts better to the small-scale spatial variations in the observations of the IM for this specific event. The direct comparison between $\hat{Y}$ and the Zhao16 GMM is presented in Fig. 3.15. Although our model appears as a point cloud and the Zhao16 model as discrete lines, due to the use of continuous and discrete site parameters, respectively, we note that the attenuation with distance of both models are comparable, with our model providing a closer fit to the observations. In a next step, we compare the average magnitude-distance scaling learned by our model to that of the actual observations and Zhao16 GMMs. Because the individual scalings between input and target parameters cannot be disaggregated for a neural network, we need to approximate them from predictions. Only events and records within the data range that is valid for comparison in terms of magnitude, distance, and depth to top of rupture (z_{tor}) are selected for comparison. The suitable data range in terms of magnitude and distance is given in Fig. 3.8 c. According to J. X. Zhao et al. (2016a,b,c), limits of $z_{tor} < 25$ km for active shallow crust, $z_{tor} > 25$ km for upper mantle, $z_{tor} < 50$ km for subduction interface, and $z_{tor} < 200$ km for subduction intraslab events are selected. We average predictions of the selected training and validation events per tectonic region type along magnitude-distance bins. The results are depicted in Fig. 3.16. Because the accuracy of this procedure depends highly on the number of predictions within a bin, we only consider bins with more than 500 predicted values. Predictions of our model in water covered areas are excluded. The dots representing the observations are derived accordingly, using the same bins. For the Zhao16 predictions, we use average magnitude and rake values obtained per tectonic region type and magnitude bin and $V_{S30} = 410$ m/s, which corresponds to site class SCII (hard soil) according to J. X. Zhao et al. (2015). More details on the configuration of the Zhao16 models are given in the supplementary material S5. For Active Shallow Crust events, predictions of our model, Zhao16, and observations provide a consistent picture (Fig. 3.16 a). Although magnitude scalings are similar between our model and Zhao16 for upper mantle events, distance scalings significantly differ (Fig. 3.16 b). However, our model is closer to the observations than Zhao16, indicating that our model has

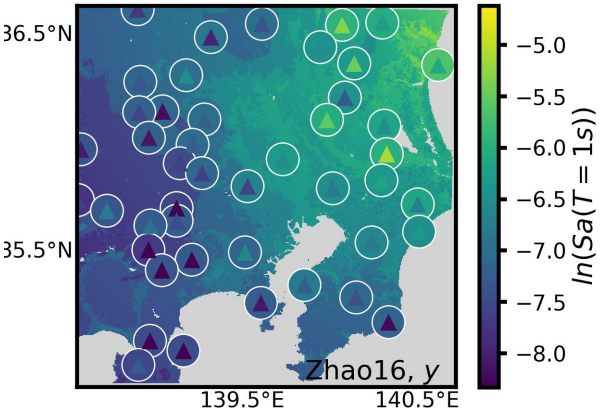

Figure 3.14: The raster data and the circles show the mean IM prediction of the Zhao16 GMM for the same event presented in Fig. 3.10. The circles represent the values at exactly those locations, where observations (triangles) are available. The rmse amounts to 0.65 in natural log units.

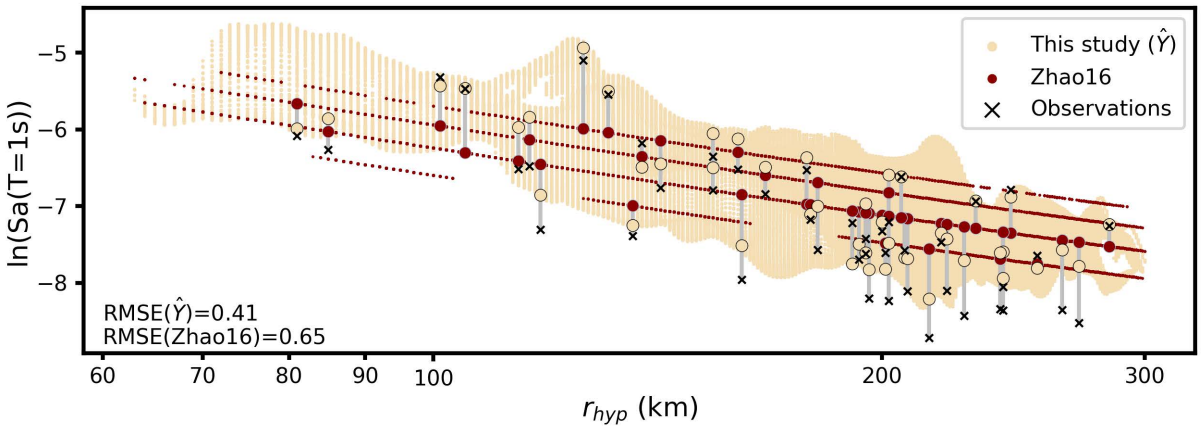

Figure 3.15: Direct comparison between our model (yellow point cloud), the Zhao16 GMM (red lines) and observations (black crosses) for the event introduced in Fig. 3.10, 3.14. Because of the continuous site parameters used in our model, corresponding predictions appear as a point cloud, contrary to the 4 lines that represent predictions of the Zhao16 GMM for four different site classes. Gray lines indicate the affiliation of observations with corresponding Zhao16 predictions (large red markers) and $\hat{Y}$ predictions (large yellow markers).

learned a Kanto basin specific distance scaling that differs from the average Japanese one. For subduction interface events, scalings are slightly different between Zhao16 and our model (Fig. 3.16 c). However, compared to the observations, our model provides reasonable results. An overall good agreement of the two models and observations can be stated for subduction intraslab events (Fig. 3.16 d). We emphasize at this stage that the predictions of our model adapt to the tectonic region type of an event, although it was not used as a predictive parameter during training. We, therefore, conclude that the U-Net extracted the relevant information from the location of the event hypocenters. This feature is fundamentally different from common GMMs that rely on the accurate classification of events into tectonic region types. Based on the presented comparison, we conclude that our U-Net based GMM provides predictions in terms of magnitude and distance comparable to Zhao16 — a set of models based on a well-established and mature methodology.

In a final comparison, we evaluate the performance of our model and the Zhao16 models on those observations from validation events that are within the validity range of both models (Fig. 3.17). The fact that the misfit of our model and Zhao16 are of the same order of magnitude indicates that the U-Net methodology is capable of providing reasonable GMMs.

3.6.2 Evaluation of learned site amplification

Our proof of concept demonstrated that our model is capable of providing fully nonergodic GMMs, including site-specific, as well as source-location and path-specific amplification. Although the latter two are difficult to evaluate on real data, because we cannot disaggregate the scaling relations within a neural network, we can approximate the average learned site amplification from the absolute site coordinates x_s, and y_s, and $z_{bedrock}$ by averaging predictions.

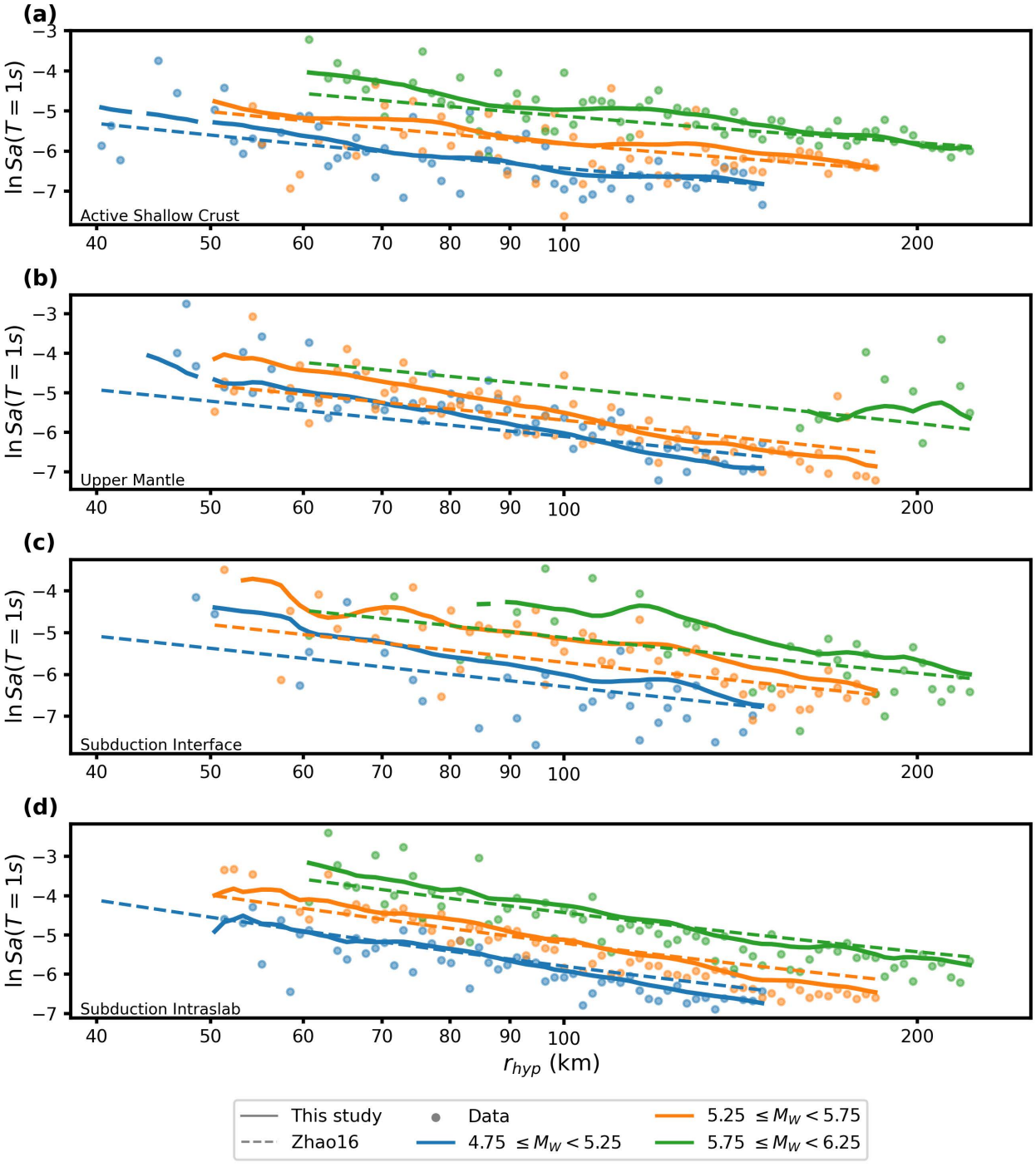

Figure 3.16: Comparison between the GMM developed in this study (solid lines), observations from the Bahrampouri et al. (2021) database (dots), and Zhao16 GMMs (dashed lines), for (a) active shallow crust events, (b) upper mantle events, (c) subduction interface events, and (d) subduction intraslab events. The dots depicting the data and the curves depicting our model are derived from averaging suitable records and predictions of $\ln Sa(T = 1s)$ along M_W and r_{hyp} bins for both training and validation events.

In a first step we train another GMM $\hat{F}$ identically to $\hat{Y}$ but without using x_s, y_s, and $z_{bedrock}$ as predictive parameters. We then approximate the site amplification $\hat{Amp}_s$ learned by $\hat{Y}$ with

$$\ln \hat{Amp}_s = \frac{1}{N_e} \sum_{1}^{N_e} \hat{Y}_e - \hat{F}_e. \tag{3.12}$$

The resulting site amplification map is depicted in Fig. 3.18. We compare this map to the depth to bedrock map used for training, the commonly used site parameter V_{S30}, and site amplification of pseudo-spectral velocity at a period of 1 s (PSV(T=1s)) derived by Nakano and Kawase (2021) (Fig. 3.19). The V_{S30} map was downloaded from the J-SHIS map web service (see section 3.8). The degrees of correlation are quantified using Spearman's ρ correlation coefficient. From the correlation of $\hat{Amp}_s$ with the depth to bedrock map (compare Fig. 3.19 a, d), one can see that the model successfully learned a relation between depth to bedrock and the target IM. Because the outline where $\hat{Amp}_s = 1$ systematically follows the outline of the Kanto basin (compare Fig. 3.19 a), we conclude that this relation is successfully transferred from the observation locations across the Kanto basin. Furthermore, one can see small-scale anomalies around seismic station locations, indicating that the U-Net learned a site-specific amplification from the absolute locations of observations. The approximated site amplification $\hat{Amp}_s$ shows moderate correlation with the commonly used site parameter V_{S30} (Fig. 3.19 b, e) and the site amplification of PSV(T=1s) (Fig. 3.19 c, f). Good agreement between $\hat{Amp}_s$ and amplification of PSV(T = 1 s) can be stated at KiK-net sites in the central and eastern parts of the Kanto basin (compare Fig. 3.18). However, major differences are observable west of the Kanto basin, where the PSV(T = 1 s) map consistently shows deamplification, and our map shows a more variable picture. Reasons for this difference might be the different data selection of Nakano and Kawase (2021). Furthermore, they used not only the KiK-net, but also K-net and Japan Meteorological Agency (JMA) strong motion networks, and therefore provide a more detailed map. Using the same data density in our model might provide a more insightful comparison but is beyond the scope of this study.

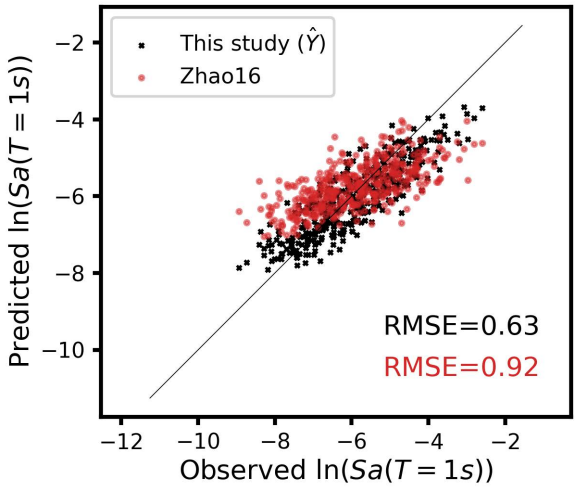

Figure 3.17: Performance of our model and Zhao16 models on the fraction of observations from validation events suitable for comparison (compare Fig. 3.8).

3.6.3 Concerning practical application

All calculations related to this study were performed on a desktop computer with an Intel(R) Core(TM) i9-7980XE CPU (977 GFLOPS), 128 GB of RAM and an NVIDIA RTX2080Ti GPU (13.45 TFLOPS). With this setup, training of a single U-Net takes approximately 1.5 hr. It

means that the presented work can be transferred to other datasets and areas of interest with reasonable effort. The major shortcoming of the U-Net GMM methodology at the current stage is its limited applicability to near source ground motion of large-magnitude events due to the lack of corresponding observations. Although all GMMs suffer from this problem, physical a priori information can be incorporated in equation-based GMMs to overcome this issue. An analog stabilization for the U-Net method could be the augmentation of strong motion datasets with predictions from ergodic GMMs. With such a strategy, the U-Net method would provide the same predictions as ergodic GMMs at the respective data ranges and provide nonergodic predictions where more observations are available. A similar solution is implemented for example by Landwehr et al. (2016), whose model converges towards an ergodic backbone model in regions where station density is low.

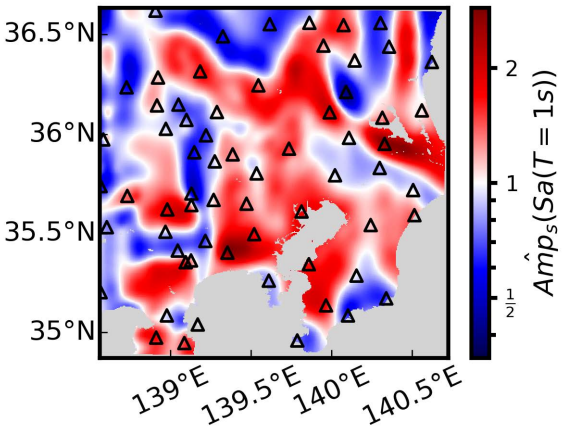

Figure 3.18: Estimated site amplification of $Sa(T = 1s)$ in the Kanto basin approximated from averaged mean predictions $\hat{Y}$.

Unlike other commonly used interpolation approaches, such as Gaussian Process Regression (as used in Landwehr et al., 2016), the U-Net methodology provides data-driven spatial interpolation without the need to define the type and parameters of an interpolation kernel function. Although this functionality is convenient and grants some degree of freedom of subjective decisions by the modeler, it is yet an open question how its performance compares against conventional interpolation techniques.

Because of the open questions discussed earlier, we currently do not consider our presented methodology sufficiently mature for immediate application in seismic hazard and risk studies, yet. However, potential future research paths toward answering these questions are in sight. We are optimistic that U-Net-based fully data-driven and nonegodic GMMs can become a permanent, complementary element of hazard and risk studies in data-rich regions such as Japan, Taiwan, and California in the foreseeable future.

3.7 Conclusions

In this study, we have presented the development of a nonergodic, fully data-driven GMM prototype based on the U-Net neural network architecture. As such, the model is free of any a priori choices concerning the model design itself or the interpolation scheme.

From experiments with synthetic data we found that the U-Net methodology is capable of learning observation site, earthquake location, and propagation path-specific amplification effects and is, thus, comparable to fully nonergodic approaches to ground motion modeling. However, the U-Net's fully data-driven, inherent interpolation scheme significantly eases the use of such a model to predict continuous ground motion maps, because no a priori assumption

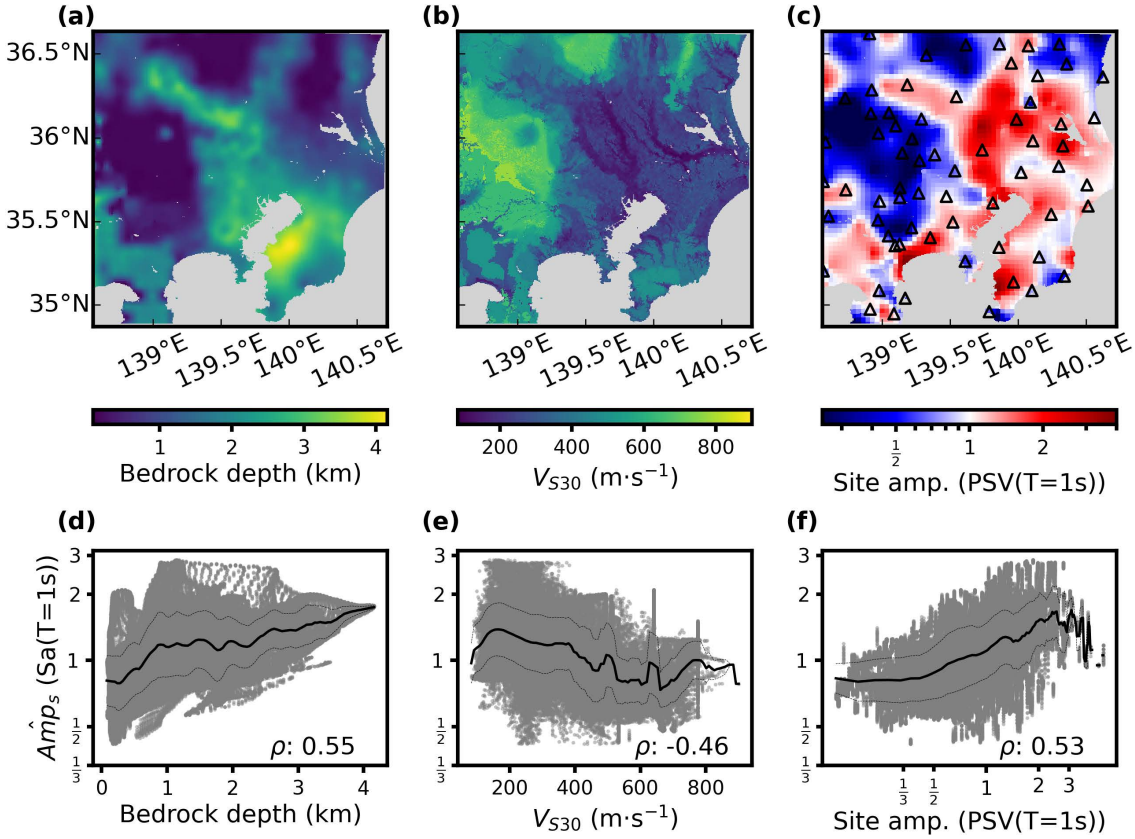

Figure 3.19: Maps of (a) bedrock depth, (b) V_{S30}, and (c) site amplification of pseudo spectral velocity at a period of 1 s (PSV(T=1s)) estimated from Nakano and Kawase (2021). The respective correlations with the approximated site amplification $\hat{Amp}_s(Sa(T = 1s))$ are depicted in (d–f), quantified using the Spearman's ρ correlation coefficient. The bold and thin black lines in (d–f) represent the average and ±1 standard deviation for the respective data bins.

about the interpolation have to be made. In the synthetic experiments, we demonstrated that the interpolation scheme provides unbiased predictions within acceptable error bounds, whereas from the results of the application to real data from the Kanto basin we can state that interpolated predictions correlate with commonly used site parameters. However, a quantitative comparison to conventional interpolation methods is yet an open task.

From the application of our model to real observations from the Kanto basin, we have learned that ground shaking of future earthquakes can be predicted, within acceptable errorbounds, at those sites where observations from previous earthquakes are available. We can furthermore state that the predictions of our model at observation sites and in interpolated regions agree on average with the predictions of equation-based, ergodic GMMs for Japan.

However, due to data scarcity, application of our model in the near field is currently not advised without augmenting the used strong motion dataset in this regards. Although associated predictive uncertainties were found to provide useful information about the relative confidence of the model in a prediction, we noticed that the absolute values of uncertainties are overestimated at observation sites and underestimated at interpolated sites, with respect to the data scatter.

The fact that our model can be derived on a consumer-level desktop computer with a decent GPU allows the transfer to an arbitrary area of interest with reasonable effort.

Summarizing the main findings from this study, we come to the conclusion that U-Net-based GMMs pose a worthwhile complementary tool to the already large and well-established pool of GMM methods. In our opinion, the presented results indicate that U-Net-based GMMs have the potential to become a permanent feature in seismic hazard and risk studies, complementary to conventional GMMs, in the foreseeable future.

3.8 Data and resources

The depth to seismic bedrock (V.3.2 ESRI shapefile "Subsurface Structure" layer 30; The Headquarters for Earthquake Research promotion, 2021), as well as V_{S30} (ESRI shapefile "Site amplification factors", parameter "AVS"; Fujimoto and Midorikawa, 2006; Senna et al., 2013, 2019; Wakamatsu and Matsuoka, 2013, 2020) maps for the kanto basin were downloaded from the J-SHIS web map service available at https://www.j-shis.bosai.go.jp/map/?lang=en (last accessed on November 26, 2021). The Bahrampouri et al. (2021) strong motion flat file was downloaded from https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2547 (last accessed on November 29, 2021). We used NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020). Figures were made with Matplotlib (Hunter, 2007) and Inkscape (Inkscape Project, 2020). Tensorflow (Abadi et al., 2015) was used for deep learning. We used QGis (QGIS Development Team, 2021) for data preparation. GMMs of J. X. Zhao et al. (2016a,b,c) were accessed via Openquake (Pagani et al., 2014). The supplementary material to this article encompasses the following content: S1, technical details on U-Net training; S2, documentation of complementary synthetic experiments; S3, a figure on the grouping of station chunks in the Kanto basin; S4, example predictions from the validation dataset and comparisons to Zhao16 models; and S5, technical details on the use of Zhao16 models.

3.9 Declaration of competing interests

The authors acknowledge there are no conflicts of interest recorded.

3.10 Acknowledgements

The first author acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS). We warmly thank Prof. Hiroshi Kawase and Dr. Kenichi Nakano for fruitful discussions and for willingly sharing their site amplification data.

3.A Appendix

3.A.1 U-Net architecture

In this appendix, we provide a detailed technical description of the U-Net architecture depicted in Fig. 3.1.

From the original U-Net, we adapt the resolution of 572 × 572 pixels for the input features, which ensures even-sized features throughout the U-Net, a requirement for a smooth processing

pipeline. The corresponding size of the input maps in terms of square kilometers is determined by the spacing between pixels that can be chosen by the user. A larger resolution of input features could be chosen to either investigate a larger area of interest or to increase the spatial resolution. The practical limitation for the resolution is the amount of available memory on the used graphics processing unit (GPU).

The first element of the U-Net is the encoder branch – a sequence of alternating Conv blocks and Pooling blocks through which the input is processed (compare Fig. 3.1). Each Conv block encompasses two convolutional layers, each followed by a tanh activation function. Features that are fed to a convolutional layer are convolved with a number of filtermasks of size (3×3), in which each filter mask generates one output feature. Thus, the number of features at a given point in the U-Net equals the number of filtermasks in the previously passed convolutional layer. We decided to reduce the number of filtermasks by a factor of 2 compared to the original U-Net in Ronneberger et al. (2015) due to GPU memory limitations. Subsequent passing through a nonlinear activation function (in our case the tanh) is required to learn nonlinear relations between the predictive parameters and the target parameter. Although the original U-Net uses the Rectified Linear Unit (ReLU) as default activation function, we observed that using the tanh leads to more stable convergence during training in our case. The purpose of the Max pooling blocks is to reduce the resolution of the features along the encoder branch. This is achieved via dividing each feature into subarrays of size 2×2 and picking the respective maximum values, thus, reducing its resolution by a factor of 2. This factor is adapted from the original U-Net. Along the encoder branch, the size of the features continuously decreases, whereas the number of features increases. It can thus be considered a feature extractor concentrating the information given in the input layer into a number of low resolution, highly abstract features.

The second element of the U-Net is the decoder branch – a sequence of alternating upsamplings, skip connections and Conv blocks. Upsampling operations increase the resolution of the features via bilinear interpolation, conversely with respect to the Pooling blocks in the encoder branch. The upsampled features are subsequently concatenated with the corresponding features from the encoder branch, which is called skip connection. Via skip connection, the abstract features derived in the feature extraction are combined with the less abstract features from the encoder branch that include more spatial context. In the original U-Net architecture, the upsampling operations are followed by convolutional layers before the skip connections. In our setup, these layers cause significant artifacts in the final results, and, therefore, we removed them from the architecture.

The spatial context information and the abstract features are then combined in the Conv blocks, which eventually leads to less abstract and more practical representations of the extracted features toward the output layer. The final Conv-out block consists of two convolutional layers that are organized in parallel, in which the first one is followed by a linear, and the second one by an exponential activation function. These two parallel layers provide two outputs that we interpret as the mean and the variance of the target intensity measure (IM), respectively. Because no padding is applied to the features during convolution operations, the resolution of the output features is smaller (388×388) than that of the input features (572×572).

Chapter 4

Capturing Directivity in Probabilistic Seismic Hazard Analysis for New Zealand: Challenges, Implications and a Machine Learning Approach for Implementation

originally submitted to *Bulletin of the Seismological Society of America* as:

Weatherill, G., **H. Lilienkamp** (2023). "Capturing Directivity in Probabilistic Seismic Hazard Analysis for New Zealand: Challenges, Implications and a Machine Learning Approach for Implementation", submitted to Bull. Seismol. Soc. Am. (accepted with minor revisions)

4.1 Abstract

The proximity of shallow, fast-slipping crustal faults to urban areas may result in pulse-like ground motions from rupture directivity contributing to increased levels of damage, even for engineered structures. Systematic modelling of directivity within probabilistic seismic hazard analysis (PSHA) remains challenging to implement at regional scale, despite the availability of directivity models in the literature. In the process of developing the 2022 National Seismic Hazard Model for New Zealand (2022 NSHM), we explored the feasibility and impact of modelling directivity for PSHA at national scale using the previous generation 2010 National Seismic Hazard Model. We identify a suitable state-of-the-art directivity amplification model for this purpose according to a set of proposed selection criteria. The results of this analysis allowed us to quantify the impact of directivity on the resulting seismic hazard maps for New Zealand, and gain insights into the factors that contribute to the expected increases (and decreases) in ground motion level. For the 2022 NSHM, earthquake rupture forecast (ERFs) seismogenic source models introduced enormous challenges for directivity modelling due to the abundance of large multi-segment/ multi-fault ruptures with complex geometries. To overcome these challenges, we apply a machine-learning based strategy to "overfit" an artificial neural network to capture the distributions of directivity amplification for each unique rupture in the earthquake rupture forecast. This produces a compact representation of the spatial fields of amplification that are computationally efficient to generate within a complete PSHA calculation for the 2022 NSHM. This flexible and reproducible framework facilitates implementation of directivity in PSHA at regional scale for complex ERF source models and opens the possibility of more complex characterization of epistemic uncertainties for near source ground motion in practice.

4.2 Introduction

Strong ground motions in the near-field region of large earthquake ruptures present a significant threat to the built environment in seismically active regions, even for buildings designed to modern seismic resistant code standards. Such motions are often not only large amplitude owing to their close proximity to the earthquake rupture, but they also contain pulse-like characteristics that can produce greater seismic demands for engineered structures than non pulse-like motions of similar amplitude (Baker, 2007; Bertero et al., 1978; Hall et al., 1995; Mavroeidis et al., 2004). One of the principal factors that contribute to these characteristics is rupture forward directivity, which occurs when a rupture propagates toward a site at a velocity approaching that of the local shearwave velocity in the crust, causing the wave to arrive predominantly as a large pulse shortly after the onset of shaking (Somerville et al., 1997). The conditions for the phenomenon to occur at a site during earthquake shaking depend on the position of the site with respect to the earthquake fault, the magnitude of the earthquake itself, and the speed and direction of rupture. Forward directivity has been observed in near-source records for many major earthquakes, such as the 1993 Landers, California, (M 7.3) (Wald and Heaton, 1994), 1999 Chi-Chi, Taiwan (M 7.6) (Xie, 2019), 2011 Darfield, New Zealand (Bradley et al., 2014) and 2023 Kahramanmaras, Türkiye, (M 7.8, M 7.6) earthquakes (Mai et al., 2023), among many

others.

4.2.1 Capturing directivity in PSHA

While near-source directivity effects have been well observed, and physical mechanisms controlling their generation largely understood, characterization of these effects in ground motion models used for probabilistic seismic hazard analysis (PSHA) remains limited. There are several reasons for this, beginning with the challenge of classifying forward directivity pulse-like ground motion records objectively (e.g. Baker, 2007; Shahi and Baker, 2014), quantifying the probability of a record displaying pulse-like motion given the source, path and site conditions (Iervolino and Cornell, 2008; Tothong et al., 2007), prediction of the change in amplitude of strong shaking due to forward (and reverse) directivity and the frequency range over which this occurs, i.e. the *directivity amplification models* (e.g. Abrahamson, 2000; Donahue et al., 2019; Somerville et al., 1997; Spudich et al., 2013), and, finally, characterization of the probability distributions that describe the aleatory uncertainty of the source and path properties relevant for directivity (e.g. Mai et al., 2005; Melgar and Hayes, 2019). Existing directivity amplification models require further parameterization of the earthquake source and path, often requiring additional source-to-site rupture distance metrics that are seldom computed by many PSHA software, and their predictions of both the scale and spatial pattern of amplification remain divergent from model-to-model (Spudich et al., 2013, 2014).

Recent developments in seismogenic fault source characterization in regions of active seismicity introduce a new dimension of complexity to directivity modelling in PSHA. The recent state-of-the-art Uniform California Earthquake Rupture Forecast v3 (UCERF v3) introduced a transformative approach to modelling earthquake recurrence on fault systems in PSHA (Field et al., 2014), wherein seismogenic sources are represented as an inventory of possible ruptures across a fault system and their associated probabilities of occurrence (the *earthquake rupture forecast*, or ERF hereafter). The long-term probabilities of occurrence of the ruptures are determined via a large-scale inversion with respect to seismic, geological, and geodetic observations in the region (Page et al., 2014). ERF source models permit a broader range of earthquake rupture characteristics than most existing fault sources in PSHA. Each rupture will comprise a set of microsegments from a fault (or several faults), permitting discontinuities in the fault surface such as offsets and step-overs, and many ruptures may include participation of microsegments from multiple faults with different styles-of-faulting. Such ruptures are substantially more geometrically complex than those for which even recent directivity amplification models have been designed and calibrated. Recent efforts in California to apply directivity to PSHA using UCERF version 3.0 have been undertaken by Al Atik et al. (2023) and Mazzoni et al. (2023) to produce seismic maps for California incorporating this phenomenon; the only such study of its kind to address the challenges posed by ERF models and apply directivity to PSHA at regional scale.

4.2.2 Relevance for seismic hazard in New Zealand

One region where the impact of directivity in seismic hazard assessment may be particularly relevant is New Zealand. As one of the world's most seismically active countries, New Zealand is host to two subduction zones at opposing ends of the country (the Hikurangi and Peysegur

subduction zones), which are connected by a fast-slipping transform fault system capable of producing large (M > 8) strike-slip ruptures, as shown in Fig. 4.1. Tectonic deformation is high across much of the country's South Island and eastern North Island, resulting in large earthquakes not only on the faults comprising that main Alpine fault system but also on shallow crustal faults away from this system, such as those of the 1855 Wairarapa (M_W 8.25), 1931 Hawke's Bay (M_W 7.6), the 2010 Darfield (M_W 7.0) and 2016 Kaikoura (M_W 7.8) events. These active shallow crustal faults present a clear threat to several major cities across the country, some of which can be found in the near-field region of the faults and would likely experience strong directivity effects in future earthquakes. Several records from the 2010 – 2011 Canterbury earthquake sequence showed pulse-like features in the Christchurch basin (Bradley et al., 2014), and may have been one of several contributing factors to the strong shaking and widespread damage observed during the sequence.

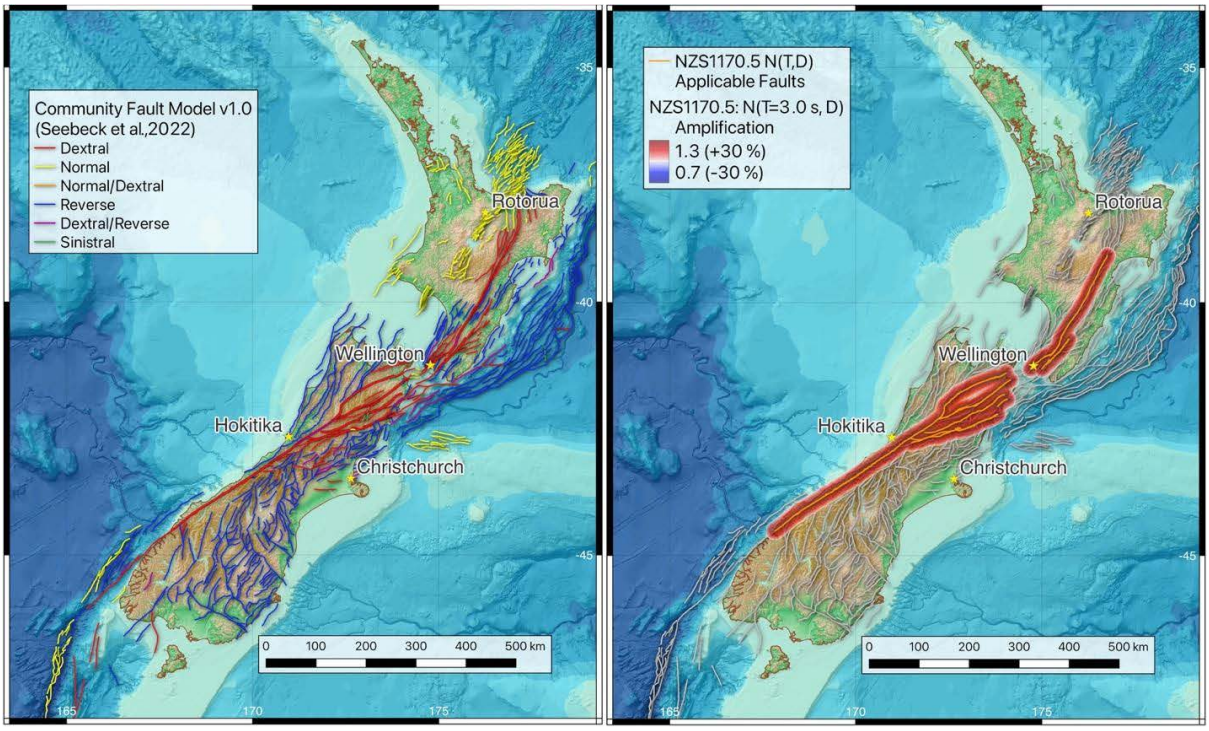

Figure 4.1: Active faults in and around New Zealand according to the 2022 New Zealand Community Fault Model – version 1.0 (Seebeck et al., 2023) (left), and region of application of the NZS1170.5 near-fault factor $N(T, D)$ with its corresponding amplification for $Sa(3.0\,\mathrm{s})$ (right).

The potential threat that directivity poses to engineered structures in New Zealand has been recognized by seismologists and engineers. The 2005 New Zealand Standard for seismic loads (NZS, 2004) is one of the few seismic design codes worldwide to introduce an explicit factor to increase the amplitude of the elastic design spectra at spectral periods $T \geq 1.5\,\mathrm{s}$ in order to account for directivity in the near-source region of major active faults. The elastic design spectrum for horizontal loads is multiplied by a period- and distance-dependent "near-fault" factor, $N(T, D)$, to be applied when the expected exceedance of the seismic design level ground motion is less than 1 / 250 yr^{-1}. $N(T, D) > 1$ for sites within 20 km of faults whose annual slip rates, $\dot{s}$, exceed 5 mm yr^{-1} and are capable of producing earthquakes of magnitude $M_W > 7$.

It tapers from a period-dependent peak when distance from the fault, D, is less than 2 km to $N(T, D) = 1$ when $D > 20$ km. Fig. 4.1 also shows the areas for which $N(T, D) > 1$ and the corresponding level of amplification for the $T = 3$ s period spectral acceleration, $Sa(3.0\,\text{s})$. This factor is not based on a probabilistic analysis but on a *reasonable* engineering basis, with N(T, D $\leq$ 2) calibrated considering an adverse rupture and site configuration for a M_W 7.5 scenario strike-slip earthquake (McVerry, 2003).

4.2.3 Aims of the study

Our overarching goals are to formulate explicit modelling of directivity in PSHA, to identify and address methodological and computational challenges, and to evaluate the relevance for the seismic hazard in New Zealand. This exploration has been undertaken within the scope of the 2022 update to the New Zealand National Seismic Hazard Model (2022 NSHM). We first aimed to implement directivity in PSHA for New Zealand and to understand its potential impact on seismic hazard prior to the completion of the 2022 NSHM. For this purpose, we adopted the fault source model from the existing 2010 NSHM (Stirling et al., 2012), which allowed us to gauge the impact of the directivity model and gain an understanding of the factors that influence its impact on the resulting seismic hazard model. The general framework for implementation of directivity in PSHA and considerations for model selection are found in the following section before presentation of the results from the exploratory study using the NSHM thereafter.

Following completion of the 2022 NSHM we then aimed to assess the feasibility of implementing directivity into PSHA at regional scale for a PSHA based on the new earthquake rupture forecast source model. The 2022 NSHM is the second such model, after UCERF v3, to adopt the fault system modelling approach that results in an earthquake rupture forecast as an input seismic source model (Gerstenberger et al., 2022a,b). As was the case for California (Mazzoni et al., 2023), we needed to address the issue of rupture complexity and feasibility of computation at regional scale. Here we introduce a novel strategy that built on components of the approach proposed by Al Atik et al. (2023) to address this problem in California but adapted them into a machine learning based framework for efficient implementation at scale. The development and the outcomes of this application to the 2022 NSHM are shown at the end of the paper, alongside general considerations for implementation of directivity for complete regional scale PSHA calculations in practice.

4.3 Modelling directivity in PSHA

4.3.1 Overview

The formulation of the probabilistic seismic hazard methodology adopted here based around the *earthquake rupture forecast* (ERF) (Field et al., 2003), defined previously. This formulation, rather than the original form from Cornell (1968) and McGuire (1976), is relevant in our case as it is through the ERF concept that fault system seismogenic source models such as UCERF3 or the 2022 NSHM are input into PSHA calculations. The ERF formulation for PSHA modelling is adopted within both OpenSHA (Field et al., 2003) and OpenQuake (Pagani et al., 2014),

the latter being the software utilized in the current work. The probability of ground motion intensity a at a site exceeding a level A in t years, $P(a \geq A|t)$, is given by:

$$P(a \geq A|t) = 1 - \prod_{i}^{N_{SRCS}} \prod_{j}^{N_{RUP,i}} \left(1 - P_{rup_{ij}}(n \geq 1|t)\right)^{P(a \geq A|rup_{ij})} \tag{4.1}$$

where rup_{ij} is the jth rupture of $N_{RUP,i}$ ruptures generated by the ith seismogenic source of N_{SRCS}, $P_{rup_{ij}}(n \geq 1|t)$ describes the probability of n=1 or more occurrences of this rupture in time t, and $P(a \geq A|rup_{ij})$ the probability of acceleration a at the target site exceeding A given the occurence of rup_{ij}. The latter term is determined from the ground motion model (GMM), which takes the form:

$$ln(Y|T) = \mu(m, r, \boldsymbol{\theta}, T) + \delta_Y = \mu(m, r, \boldsymbol{\theta}, T) + \varepsilon \cdot \sigma(T) \tag{4.2}$$

where the logarithm of ground motion spectral acceleration, Y, at period T is described by a normal distribution with expected value $\mu(m, r, \boldsymbol{\theta}, T)$ and total standard deviation $\sigma(T)$, i.e., $\ln(Y|T) = \mathcal{N}(\mu(m, r, \boldsymbol{\theta}, T), \sigma)$. Here m is the magnitude of rup_{ij}, r the distance from the rupture to the target site and $\boldsymbol{\theta}$ the set of other relevant source and site parameters to describe the expected distribution of ground motions (e.g., hypocentral depth h_D, style of faulting, shearwave velocity averaged over the upper 30 m of the crust V_{S30}, depth of soil sediment to a given shearwave velocity layer Z_h, etc.). A common misconception of PSHA is that it does not account for directivity in ground motion unless the GMM itself contains an explicit term for this phenomenon. This is not exactly the case, however, as calibration of both $\mu(m, r, \boldsymbol{\theta}, T)$ and $\sigma(T)$ for ground motions from large-magnitude earthquakes will be based on a set of observed or simulated ground motions that will contain a proportion of pulse-like records. Directivity can therefore be said to be captured implicitly in current PSHA practice owing to its presence in the near-fault ground motion variability $\sigma(T)$. However, this implicit approach may not always be adequate in application as $\sigma(T)$ is usually constrained from ground motion records over a large range of distances and azimuths, and thus represents a mixture of pulse-like and non pulse-like records. The predicted variability in ground motion in the near-fault region may be underestimated, as it is reduced by averaging with records from far-field events for which directivity effects are weaker or absent (Abrahamson, 2000). The expected variation in ground motion around an active fault will likely be more complex, changing with magnitude, distance and source-to-site azimuth, as well as with spectral period. Non-ergodic GMMs with spatially varying coefficients may capture part of this complexity via repeatable path effects, although with very few records in the near field of large magnitude earthquakes on a fault it is unlikely to be representative of the true variability. These limitations of the implicit approach are a key motivation for modelling directivity in PSHA using explicit predictive models of amplification that depend on the characteristics of the earthquake rupture and the configuration of the target sites for which ground motion is calculated (e.g. Donahue et al., 2019; Somerville et al., 1997; Spudich et al., 2013). We refer to this as the explicit approach to modelling directivity in PSHA.

4.3.2 Explicit directivity modelling for PSHA

In our nomenclature, *explicit* modelling of directivity requires modification of the ground motion model to incorporate a near-fault directivity amplification term:

$$lnY(T) = [\mu(m, r, \boldsymbol{\theta}, T) + \Delta\mu_{dir}(m, r, h, \boldsymbol{\theta}, \boldsymbol{\theta_{dir}}, T)] + \varepsilon \cdot \sigma_{dir}(m, r, h, \boldsymbol{\theta}, \boldsymbol{\theta_{dir}}, T) \tag{4.3}$$

where $\Delta\mu_{dir}(m, r, h, \boldsymbol{\theta}, \boldsymbol{\theta_{dir}}, T)$ is a scenario-dependent modifying factor describing the change in expected ground motion due to near-fault directivity, with respect to the median of the GMM ($\mu(m, r, \boldsymbol{\theta}, T)$) that represents a directivity neutral condition, and $\sigma_{dir}(m, r, h, \boldsymbol{\theta}, \boldsymbol{\theta_{dir}}, T)$ is the modified total variability of the GMM owing to the inclusion of the directivity amplification term. σ_{dir} should, at this stage, be reduced with respect to the original σ of the GMM, as inclusion of $\Delta\mu_{dir}$ into the expected near-source ground will reduce the corresponding aleatory variability. h is the position of the hypocenter within the rupture plane, which we will describe here as $h = \{X, Z\}$, where $X \in [0, 1]$ is the along-strike position of the hypocentre as a fraction of the total length L of the rupture plane and $Z \in [0, 1]$ the down-dip position of the hypocenter as a fraction of the total down-dip width W of the rupture plane. A distinction is made between $\boldsymbol{\theta}$, the set of source, path and site parameters (excluding magnitude and distance) required by a GMM to describe the expected ground motion at a site, and $\boldsymbol{\theta_{dir}}$ the additional set of parameters required to describe the directivity amplification, which are limited only to source and path parameters. Depending on the complexity of the GMM in question there may be some overlap between $\boldsymbol{\theta}$ and $\boldsymbol{\theta_{dir}}$ particularly regarding properties related to the dimension of the earthquake rupture, such as fault width (W), fault length (L) and the mechanism of the rupture represented by its strike (ϕ), dip (δ) and rake (ξ). Among $\boldsymbol{\theta_{dir}}$, however, there may also be parameters relating to the dynamic properties of a rupture, such as local direction of slip (Rowshandel, 2018), which need to be treated as aleatory variables in the context of PSHA.

The modified GMM in equation 4.3 introduces potentially two new aleatory variables compared to conventional PSHA: h and $\boldsymbol{\theta_{dir}}$. To incorporate these variables into PSHA we would need to adapt the formulations in equation 4.1 to include integrals over their respective probability distributions $f_H(h)$ $f_\gamma(\boldsymbol{\theta_{dir}})$. Equation 4.1 therefore becomes:

$$P(a \geq A|t) = 1 - \prod_{i}^{N_{SRCS}} \prod_{j}^{N_{RUP,i}} \prod_{k}^{N_h} \prod_{l}^{N_\gamma} \left(1 - P_h(h_k) \cdot P_\gamma(\boldsymbol{\theta_{dir,l}}) \cdot P_{rup_{ij}}(n \geq 1|t)\right)^{P(a \geq A|rup_{ij}, h_k, \boldsymbol{\theta_{dir,l}})}, \tag{4.4}$$

where $P_h(h_k)$ is the probability of the $h = \{X, Z\}_k$ hypocenter position on the rupture from N_h positions approximating $f_H(h)$ (such that $\sum_k^{N_h} P_h(h_k) \equiv 1$), and $P_\gamma(\boldsymbol{\theta_{dir,l}})$ the probability of the directivity predictor $\boldsymbol{\theta_{dir,l}}$ from N_γ values approximating $f_\gamma(\boldsymbol{\theta_{dir}})$ (also requiring that $\sum_{l=1}^{N_\gamma} P_\gamma(\boldsymbol{\theta_{dir,l}}) \equiv 1$). Following the terminology of (Donahue et al., 2019), we refer to this complete extension of the PSHA integral as the *full hypocenter randomization* approach (notwithstanding that the formulations above may randomize both the hypocenter and potentially other predictors of directivity). Implementation of this approach in the PSHA calculation results in a significant increase in computation, however, as the number of earthquake source

scenarios is increased by a factor of $N_h \cdot N_\gamma$. This is compounded by the additional cost of calculating $\Delta\mu_{dir}$, σ_{dir} and potentially complex parameters within $\boldsymbol{\theta_{dir}}$ itself. When attempting to run PSHA at regional scale, which considers potentially tens of thousands of target locations, the increased computation needed for the *full hypocenter randomization* approach is not just challenging, it may become entirely prohibitive.

An alternative strategy for explicit directivity modelling in PSHA is to modify the moments of the GMM to reflect the mean change in expected ground motion ($\overline{\Delta\mu}_{dir}$) and its corresponding aleatory variability (σ'_{dir} based on the additional distributions $f_H(h)$ and $f_\gamma(\boldsymbol{\theta_{dir}})$. (Donahue et al., 2019) term this approach modified moments from a randomized hypocenter (hereafter just modified moments approach), which effectively separates $f_H(h)$ and, if necessary,$f_\gamma(\boldsymbol{\theta_{dir}})$ from the PSHA integral and uses the mean and variance of the amplification from given directivity model to modify the GMM directly:

$$ln(Y|T) = [\mu(m, r, \boldsymbol{\theta}, T) + \overline{\Delta\mu}_{dir}(m, r, T|f_h, f_\gamma)] + \varepsilon \cdot \sigma'_{dir}(m, r, T|f_h, f_\gamma), \tag{4.5}$$

and

$${\sigma'_{dir}}^2(m, r, T|f_H, f_\gamma) = \sigma^2(T) - \sigma^2_{red}(T) + \sigma^2_{f_H f_\gamma}(m, r, T|f_h, f_\gamma) \tag{4.6}$$

where $\sigma^2(T)$ is the original variance of the unadjusted GMM, $\sigma^2_{red}(T)$ the factor by which the original GMM variance should be reduced on account of the explicit inclusion of directivity into the median of the GMM, and $\sigma^2_{f_H f_\gamma}(m, r, T|f_h, f_\gamma)$ the additional variance in ground motion at a given site emerging from the assumed aleatory variability in hypocenter and directivity predictor properties. The *moment modifiers* of the ground motion model, simply $\overline{\Delta\mu}_{dir}$ and σ_{f_H, f_γ} hereafter, can be fit using simpler parametric equations conditioned upon certain properties of the source, path and site, as presented by Watson-Lamprey (2018).

The obvious advantage of the *modified moments* approach is that the additional computational effort is undertaken outside of the PSHA calculation, removing the need to integrate over$f_H(h)$ and $f_\gamma(\boldsymbol{\theta_{dir}})$ per rupture within the execution. Depending on the complexity of the parametric model for the *moment modifiers* and the source, path and site parameters they may require, if any, the additional computational costs are minimal while the overall spatial distribution of the mean amplification and its impact on the total aleatory variability of the GMM can still be captured explicitly. The potential limitations to this approach are, firstly, the requirement to define separate coefficients of the model for each GMM and/or directivity model combination, and secondly the loss of accuracy in the predictions due to the simpler parametric approximation. The latter becomes more relevant as we consider complex earthquake ruptures for which the spatial patterns of directivity amplification may diverge more from those implied by the simpler parametric approximation. Ultimately, the choice of approach may depend on the context of application and could be guided by preliminary assessments of the likely impact of explicit integration of directivity into PSHA using the full hypocenter randomization approach on smaller scale hazard calculations as a point of reference.

4.3.3 Selecting directivity models for application in PSHA

Since the first general framework for inclusion of directivity in PSHA was proposed by Somerville et al. (1997), several studies have published parametric directivity models intended for practical application. These have evolved in the subsequent 25 years, first with an updated calibration proposed by Abrahamson (2000), then with the development of a new directivity predictor model based on isochrone theory (Spudich and Chiou, 2008). In addition to developing state-of-the-art GMMs for application to shallow crustal earthquakes, the Next Generation Attenuation (NGA) West2 project also produced a new generation of directivity models with a view to application in PSHA. These were summarized in a report by Spudich et al. (2013) and are listed as: Bayless and Somerville (2013)[BS13], Spudich and Chiou (2013)[SC13], Shahi and Baker (2013, 2014)[SB13/14], Rowshandel (2013) updated in Rowshandel (2018)[R13/18] and Chiou and Spudich (2013)[CS13]. For a full explanation and comparison of these models the reader is referred to Donahue et al. (2019) and Spudich et al. (2013, 2014). Among these models CS13 is noteworthy due to its explicit integration into the Chiou and Youngs (2014) NGA West 2 GMM, the only such GMM to date to include an explicit directivity term. Like its predecessor (Spudich and Chiou, 2008, 2013), CS13 is based on isochrone theory but instead introducing the Direct Point Parameter (DPP) as a predictor of directivity.

Following the NGA West2 Directivity project comparatively few new models have emerged for PSHA application, and usage has still been limited. Among the factors that may have contributed to this are the additional computational demands required for their deployment, the divergence between the models themselves, and the potential challenges and inconsistencies that emerge from applying these models to more complex fault rupture surfaces encountered in recent PSHA models. Two subsequent models have been developed that have taken these factors under consideration, which have been aided by the introduction of the Generalized Coordinate System 2 [GC2] (Spudich and Chiou, 2015) to facilitate their application to complex multi-segment ruptures (described in more detail in due course). Watson-Lamprey (2018) [WL18] developed the moment modifier model, which is calibrated upon the mean and standard deviation of the amplification predicted by Chiou and Spudich (2013)'s DPP and parameterized using an efficient functional form that adopts the GC2 framework. More recently, Bayless et al. (2020) [B20] provided a successor to BS13 that updates several of the modelling assumptions, is well suited to multi-segment ruptures using the GC2 framework and is efficient for use in PSHA.

The summary of the available directivity models in the scientific literature serves to illustrate that while there is currently a limited selection of models, each model has its own characteristics that may influence its suitability for application in PSHA. Following the analogy of defining selection criteria for GMMs (e.g. Bommer et al., 2010; Cotton et al., 2006), the choice of directivity models to use should also follow clearly defined criteria that can not only guide current selection but also help future directivity model developers construct models that meet such requirements. At the same time, with such a limited number of models one may run the risk that stringent criteria will eliminate *all* potential directivity models, and therefore aim to strike a practical balance between accuracy and utility. The criteria we propose for directivity model selection, which we based on the discussion provided in Donahue et al. (2019), are:

1) The directivity amplification model is derived from ground motion records appropriate to

the tectonic environment and for the definition of horizontal ground motion compatible with the GMMs to which it will be applied.

2) The directivity model spans a sufficient spectral period range for the application in question and covers *adequately* the range of scenarios relevant for seismic hazard at the site(s).

3) The model includes appropriate tapers of amplification with magnitude and distance.

4) The model is narrowband, with the peak pulse period, t_p, increasing with magnitude.

5) The model is appropriately calibrated such that the average value of the directivity predictor over all azimuths represents a directivity neutral condition, i.e., the model is *directivity centered*.

6) The model is applicable to multi-segment rupture surfaces and should allow for discontinuities between the surfaces without singularities or discontinuities in the resulting amplification field.

7) The model should not be superseded by that found in a subsequent publication.

In the analysis presented here we have limited our focus only to the model of Bayless et al. (2020). This model not only fulfills the stated selection criteria, but it also has several appealing features for our purpose. It uses GC2 to allow for application to complex multi-segment ruptures, has an efficient functional form using rupture parameters and distance metrics frequently computed in PSHA, proposes a reduction factor for the within-event variability of the GMM from which we can calculate σ_{red}. It is also empirically centered, based on the residual distributions of observed ground motions. This means that it can be applied as an *a posteriori* adjustment to the median and standard deviation of the GMM at the site in question after calculation of the nonlinear site response (Jeff Bayless, personal communication). This contrasts with CS13 and R13/18, which should be applied to the calculation of motion on bedrock prior to application of the nonlinear site response, as is the case when implemented in the GMM of Chiou and Youngs (2014). The latter may be preferable from a theoretic perspective as directivity is a source and path phenomenon, but this can be challenging to implement in PSHA software as it requires modification of each of the GMMs to which the directivity adjustment is being applied. Further work to explore the implications of this assumption is ongoing.

4.4 Application to PSHA in New Zealand using the 2010 NSHM

4.4.1 Seismogenic source model

The 2010 National Seismic Hazard Model of New Zealand is presented in Stirling et al. (2012) and was subsequently translated into OpenQuake by Abbott et al. (2020). The OpenQuake implementation is utilized for the current purposes, which is known to be in good agreement with the original version of the model. The 2010 NSHM source model comprises two types of seismogenic source: 1) three-dimensional fault sources with characteristic earthquake recurrence models, which are applied to shallow crustal faults as well as to ruptures on Hikurangi and

Fiordland subduction interfaces, 2) distributed (smoothed) seismicity with exponential recurrence models, which characterize the background shallow crustal seismicity and the subduction in-slab sources. Shallow crustal faults in the volcanic region of the North Island, which runs from Tongariro National Park in the south through to the Bay of Plenty, are classified as "volcanic" fault sources. Altogether a total of 541 faults are present in the model (shown in Fig. 4.2), of which 333 are classified as "active shallow crust", 11 as "subduction interface" and 196 as "volcanic".

The fault sources of the 2010 NSHM have some features that make them particularly well suited as a test case for directivity implementation in PSHA. In terms of the geometry, the fault surfaces are generally regular and topologically simple. Each fault is represented by either a single planar rupture surface (for the smaller faults) or a multi-segment surface comprising multiple planar segments. For the multi-segment surfaces no complexities such as offsets or step-overs are present, and the rake and dip are constant across all segments in the rupture. A characteristic earthquake magnitude frequency distribution is assumed for each fault such that a single annual rate of occurrence $\lambda_{char,i}$ is assumed for the fault characteristic magnitude M_{char} (the characteristic magnitude). These conditions make the ERF for the entire fault source model compact, with only 541 ruptures ($N_{rup,i} = 1$ for all i), and the sources themselves are assumed time-independent so that $P_{rup_{ij}}(n \geq 1|t)$ is determined using the Poissonian recurrence model:

$$P_{rup_{ij}}(n \geq 1|t) = 1 - e^{-\lambda_{char,i} \cdot t} \tag{4.7}$$

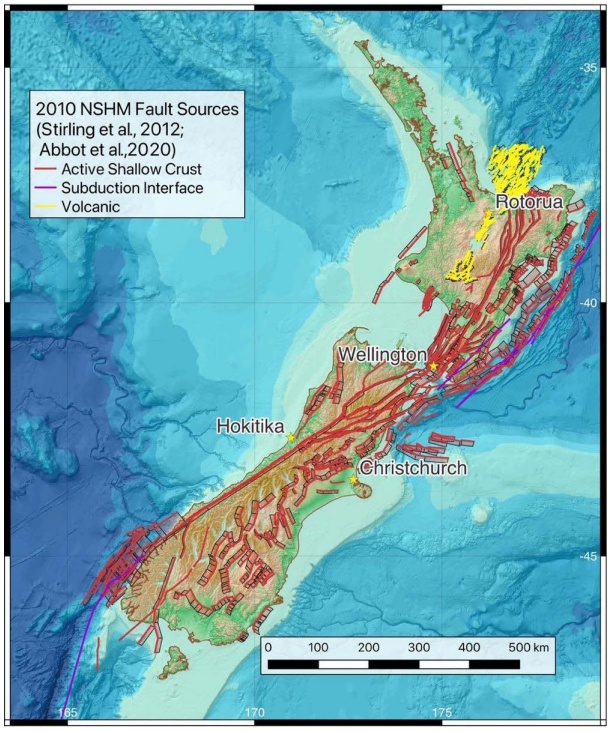

Figure 4.2: Fault Sources in the 2010 NSHM (Abbott et al., 2020; Stirling et al., 2012).

One important question to address is that of the sources to which the directivity model should be applied. The available models in the literature have been developed principally for application to shallow faults in the crust, and while there are examples of directivity observed

from subduction interface earthquakes (Abercrombie et al., 2017; Folesky et al., 2018), to the author's knowledge no models have been developed for such events that could be applied in a PSHA context. Furthermore, while models for crustal earthquakes could in practice be implemented for subduction interface sources in the current example, to do so would be to apply the existing models to contexts and rupture sizes well outside of the range for which the models were developed. We therefore do not apply directivity to subduction interface fault sources, but rather only to the active shallow crust and the volcanic faults.

4.4.2 Ground motion model (GMM)

The 2010 NSHM adopts exclusively the GMM of McVerry et al. (2006), which provides alternative parameterizations for the prediction of PGA and $Sa(T)$ in the period range $0.075 \leq T(s) \leq 3.0$ from earthquake shallow crustal, subduction interface and subduction in-slab regions. This model is calibrated on strong motions recorded within New Zealand from 51 earthquakes occurring prior to the end of 1995. However, as this data set contains few records from the near-field regions of large faults, additional records from other regions were added to supplement the database used for a calibration of the GMM for peak ground acceleration (PGA).

For this test case application of directivity, adoption of the McVerry et al. (2006) model, as originally implemented, would result in several inconsistencies between the GMM and the directivity model under consideration. The first is the definition of the horizontal component of ground motion, for which the 2010 NSHM adopted the larger of the two as-recorded components in the GMM, while the available directivity models have been calibrated assuming the RotD50 measure (Boore, 2010). To address this inconsistency in our comparisons McVerry et al. (2006) is adapted by converting from the larger-horizontal component of motion to the RotD50 using factors defined by Bradley and Baker (2015). We also adopted the magnitude-dependent weighting to the McVerry et al. (2006) model described in Gerstenberger et al. (2014). This weighting applies only to the short period motion ($T \leq 0.5$ s), however, which is below the period range to which our selected directivity model applies but relevant for the comparison of the GMM spectra and uniform hazard spectra shown subsequently. To reflect these adaptations the model will be referred to as McVerry et al. (2006)* hereafter.

For the McVerry et al. (2006)* GMM we do not know whether the expected motion in the near field region of the seismic source reflects a directivity neutral condition even after the above adjustments have been applied. Similarly, the spectral period range of the GMM is relatively limited, meaning that the period of the pulse in large magnitude events may not be captured in the variability of the GMM itself. Given these limitations, we also include in our comparison the two additional GMMs that have been developed for the 2022 NSHM of New Zealand: Stafford (2022) and Atkinson (2022), further details of which can be found in their original publications. These additional GMMs are better suited for application with the directivity model under consideration (here Bayless et al., 2020) as they are defined directly in terms of RotD50 and span a sufficiently broadband period range ($0.01 \leq T(s) \leq 10.0$).

Stafford (2022) can be considered a New Zealand specific calibration of the NGA West 2 model of Chiou and Youngs (2014) using the approach of Stafford et al. (2022). The calibrations applied by Stafford (2022) largely address the source stress parameter depth-scaling effects

and the anelastic attenuation, which are unlikely to unduly bias the median GMMs for large magnitude, short distance ruptures with respect to that of the original GMM. The assumption of directivity centering in the median ground motion can therefore be verified in the Stafford (2022) model owing to this inheritance from Chiou and Youngs (2014). In the original model the median ground motion is explicitly directivity centered when the directivity adjustment ($\overline{\Delta DPP}$) is set to zero. Scaling of Stafford (2022) and Chiou and Youngs (2014) with distance and period (Fig. 4.3) shows that the two models are virtually identical at large magnitude and short distances.

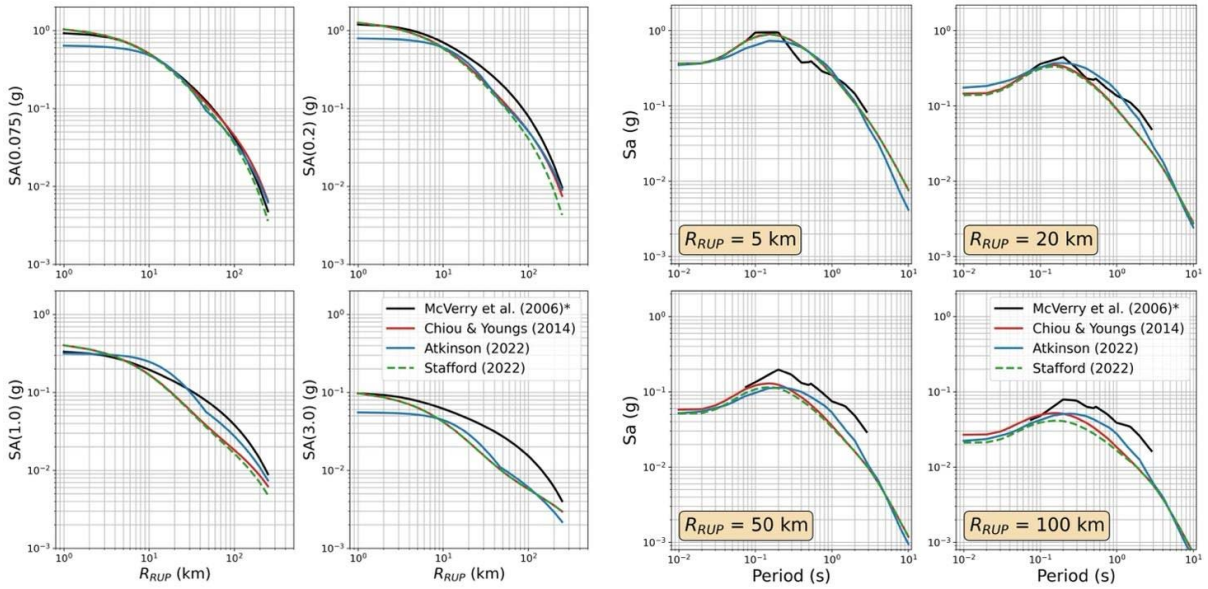

Figure 4.3: Comparison of median ground motion from GMMs considered for current application to New Zealand. Attenuation with rupture distance (R_{RUP}) for Sa at periods T = 0.075, 0.2, 1.0 and 3.0 s (left) and scaling of spectral acceleration with period for a M_W 7 strike-slip earthquake at distances of R_{RUP} = 5, 20, 50 and 100 km (assuming a site condition of V_{S30} 750 m/s) (right).

While we cannot necessarily directly verify that the Atkinson (2022) model is directivity centered as we could for Stafford (2022), the functional form of the crustal ground motion model is based on previous work of Yenier and Atkinson (2015a,b), which itself was calibrated using the NGA West 2 database. Both this database and the models derived from it have been shown to be directivity centered (Donahue et al., 2019), which we tentatively assume is reflected in the magnitude- and near-source scaling of the GMMs in the Yenier and Atkinson (2015a) model. Atkinson (2022)'s adaption of Yenier and Atkinson (2015a) GMM for application to New Zealand is undertaken using a database that is still limited in large-magnitude near-source records for crustal earthquakes (Hutchinson et al., 2022). The large-magnitude near-source records that are present in the New Zealand database are dominated by two events (the 2010 Darfield M7.1 event and 2016 M7.8 Kaikoura earthquake), for which directivity was observed and quite predominant in the records. Nevertheless, we will for our purposes assume that with so few records in this range the observed New Zealand data *may* be insufficient to de-center the model from a directivity neutral condition.

4.4.3 Hypocenter distribution within the rupture plane

The choice of $f_H(h)$ is generally an epistemic variable that in certain contexts may be appropriate to include within the logic tree of the seismic hazard model. Several models for along-strike and down-dip hypocenter distribution are available in the scientific literature (e.g. Mai et al., 2005; Melgar and Hayes, 2019). For implementation within PSHA, the definition of $f_H(h)$ should consider the shape of the distribution and the number of discrete points used in its approximation within the calculation. It is generally assumed that the distributions are symmetric, meaning that there is no preferred direction of rupture propagation; however, models may differ in terms of the relative weight assigned to predominantly bilateral rupture propagation and unilateral propagation. Watson-Lamprey (2018) explored several different distributions and suggest a preferred along-strike hypocenter position model (X) for strike slip ruptures that places 0.6 weight on bilateral propagation and 0.4 on unilateral propagation with no preference for direction of propagation. For dip-slip ruptures a more typical Gaussian distribution is assumed. We adopt this as the preferred model for along-strike position both for strike-slip and dip slip ruptures, which in their proposal is discretized into 10 bins ($X \approx [0, 0.1], [0.1, 0.2], \ldots, [0.9, 1.0]$) as shown in Fig. 4.4.

For down-dip hypocenter position Z, we use the model of Melgar and Hayes (2019), which describes Z as a Gaussian distribution with $\mathcal{N}(\mu = 0.55, \sigma = 0.2)$, which is truncated in the range [0, 1]. For both strike-slip and dip-slip events this distribution is discretized into bins using the Gaussian Quadrature approximation of Miller and Rice (1983). In the case of strike-slip ruptures a three-bin approximation is used, while for dip-slip ruptures five bins are preferred. These too are shown in Fig. 4.4. Altogether, for strike-slip faults $f_H(h)$ is approximated by 24 positions of $\{X, Z\}$, while for dip-slip earthquakes 50 are used. It is certainly arguable that a higher resolution discretization may improve accuracy, particularly in the case of the longer strike-slip ruptures. In a full randomization of hypocenter approach, however, this increases computational cost, which may be prohibitive for analysis at regional scale.

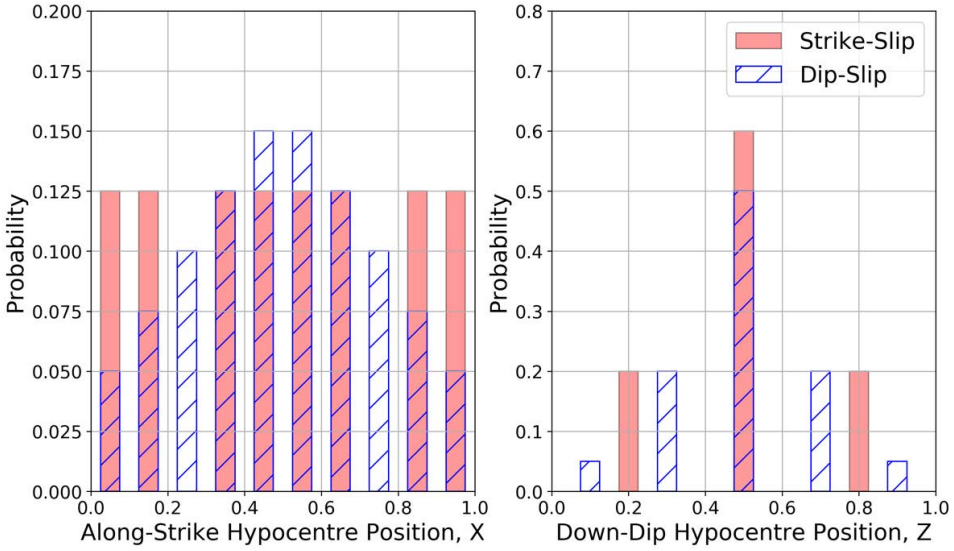

Figure 4.4: Probability mass functions for hypocenter position within the rupture for strike-slip and dip-slip earthquakes: Along-Strike (left) and Down-Dip (right).

4.4.4 Changes in seismic hazard with explicit directivity modelling

Full probabilistic seismic hazard calculations were run for New Zealand combining the Bayless et al. (2020) directivity model with three different GMMs for shallow crustal earthquakes: McVerry et al. (2006)*, Stafford (2022), and Atkinson (2022). In the first instance only the crustal fault sources are considered, meaning that background sources and subduction sources are neglected ("crustal faults only"), while in the second instance all sources are included ("all sources"). In the "all sources" case the McVerry et al. (2006)* GMM is run for the subduction sources, while the GMM for the shallow seismicity sources changes as described. Directivity is applied only to the crustal fault sources. Seismic hazard maps corresponding to the 475-year return period (10 % probability of exceedance in 50 years) and 2,475 year return period (2 % probability of exceedance in 50 years) are produced for New Zealand for multiple spectral periods in the range $0.075 \leq T(\text{s}) \leq 3.0$; however, we show only the results for $Sa(T = 3.0\,\text{s})$ here. Difference maps are given in terms of the % change in ground motion for the specified probability of exceedance when including directivity explicitly into the calculation, which can be seen in Fig. 4.5 and 4.6 for the 10 % and 2 % probability of exceedance in 50 years respectively. A reference rock site condition of V_{S30} 800 m/s is assumed throughout (with depth to the 1 km/s V_S layer for the site, $Z_{1.0}$, is set to 31 m).

Increases in seismic hazard can be seen clearly around the main Alpine fault system on the western side of the South Island, then continuing around the Hope and Clarence fault systems that cross the northern South Island. These are the most active strike slip fault systems in the 2010 NSHM, so the increase in seismic hazard is certainly expected and the change is on the order of around 5 – 15 % at the 475-year return period and 20 – 30 % for the 2,475-year return period for $Sa(3.0\,\text{s})$. We also see an increase at the very northern end of the Alpine fault, close to the intersection with the Wairau and Waimea faults in the northern South Island close to Nelson and the Tasman Bay. In the North Island there are two main regions of increased hazard: 1) the extended fault system that runs northwest from Wellington for several hundred kilometers and incorporates the Wellington, Wairapa, Mohaka faults, 2) the central North Island running from south of Lake Taupo north-northeast toward Rotorua. This latter region is within the Taupo Volcanic Zone and is characterized by dozens of smaller fault sources with characteristic magnitudes in the range $6.0 \leq M_{char} \leq 6.5$ and annual rates of occurrence on the order of 0.001 to 0.005 per year.

While largest increases in seismic hazard due to directivity can be seen close to the major active strike-slip fault systems, the overall change is modest. When considering only the crustal fault sources, much of the hazard in the North Island is reduced for having included explicit modelling of directivity, while other regions see changes only on the order of less than 2 – 3 %. These percentages decrease for shorter spectral periods, reducing to barely 1 – 2 % for $Sa(1.0\,\text{s})$ at the 475-year return period. When we include the subduction and background seismicity sources (lower rows of Fig. 4.5 and 4.6) the influence of directivity on the hazard is more "washed out" (particularly in the North Island), with the decreases in seismic hazard less visible owing to the relative increase in the contributions of the background and subduction sources to the probabilities of exceedance of ground motion at the return periods of interest here.

To view the influence of directivity across the full spectral period range, we show in Fig. 4.7

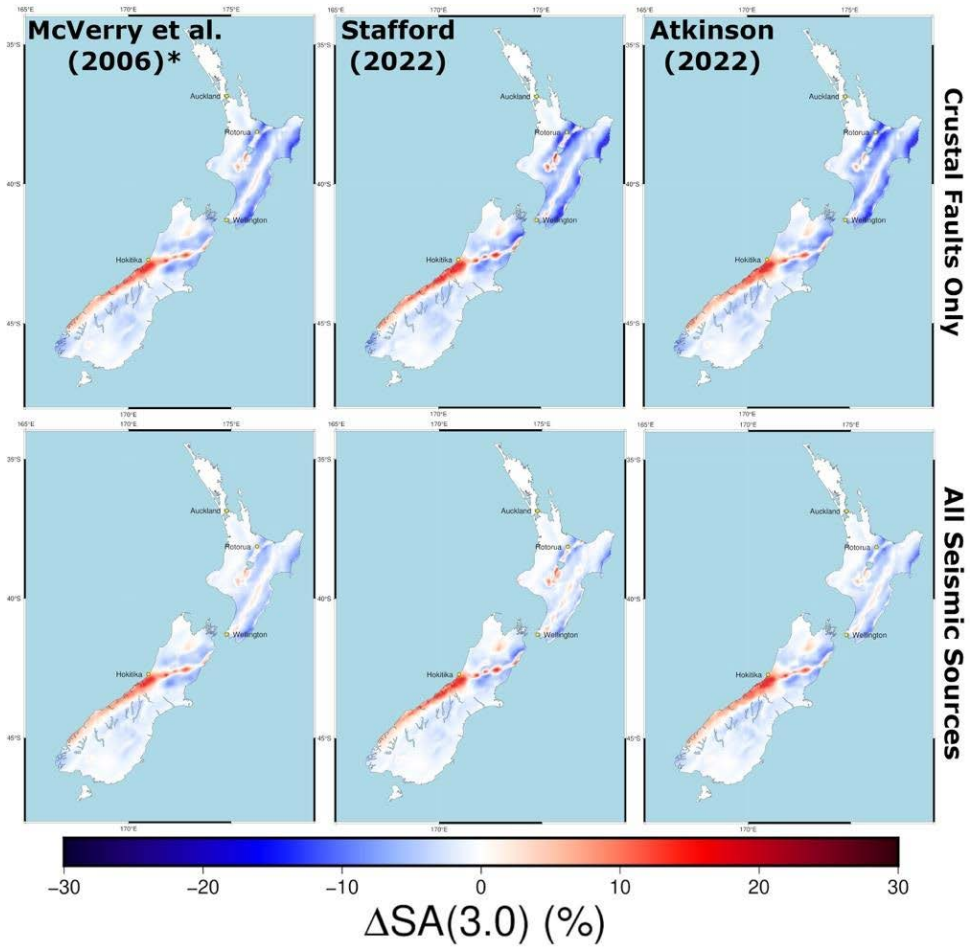

Figure 4.5: Change in $Sa(3.0\,\mathrm{s})$ with a 10 % Probability of Exceedance in 50 years when including explicit directivity modelling into PSHA, considering only the 2010 NSHM crustal faults sources (upper row) and considering all NSHM 2010 seismogenic sources (lower row).

the uniform hazard spectra (UHS) for both return periods (475 year and 2,475 year) for four selected cities: Wellington, Christchurch, Rotorua and Hokitika. Given its relevance as a long-period phenomenon we include the UHS for both spectral acceleration and spectral displacement ($Sd(T)$). While these comparisons serve to reinforce how modest the changes are overall, each city reveals a relevant facet of the influence of directivity.Wellington, for example, sits in the very near-field region of the active Wellington Fault, and in close proximity to several other active structures (e.g., Whitemans Fault, Wairarapa Fault, Ohariu Fault). The annual rates of recurrence of nearly all the characteristic ruptures are less than 1 / 475, however, meaning that they do not contribute to the hazard at the site for this return period. Only the reverse-slipping Needles fault 60 km south of the city, offshore in the Cook Straight, contributes to the hazard in Wellington at this return period, and there the directivity models slightly de-amplify shaking (with respect to a directivity neutral condition) in this position. At the 2,475-year return period, faults closer to the city begin to contribute to the hazard, but their contribution is complex. Wellington lies toward the middle of the Wellington fault, a region where directivity is likely to be neutral or de-amplifying under most of the hypocenter positions. Positive directivity amplification may be experienced at the city from ruptures farther to the north at distances of

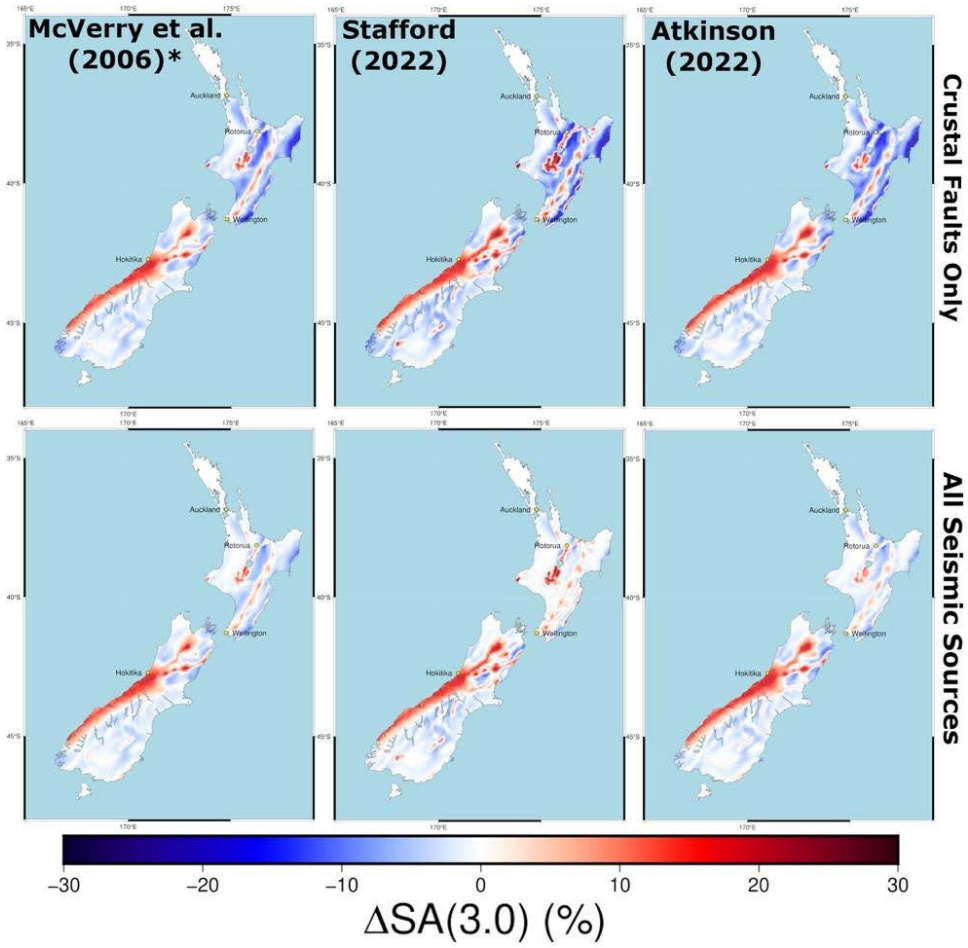

Figure 4.6: As Fig. 4.5, for $Sa(3.0\,\mathrm{s})$ and a 2 % Probability of Exceedance in 50 years.

20 – 30 km away, but then the reverse slipping Wharekauhau and Ohariu faults also contribute to the hazard and their influence would serve to de-amplify the hazard in the Wellington Central Business District itself. The result is a complex mix of ruptures that would individually increase or decrease hazard in the city, the net result being only a modest overall change.

The combined influence of several faults on seismic hazard, their relative contribution to the hazard at a site and their source-to-site configuration is illustrated well in the case of Christchurch. The 2010 NSHM does contain the Greendale fault to the west of the city, which was the source of the 2010 Darfield Earthquake (M_W7.1). During this earthquake directivity was well observed in the city (Bradley et al., 2014), and as the fault strikes in a near east-west direction then positive directivity amplification is expected at the city for all hypocenter positions. However, the rate of occurrence assigned to the M_W7.1 characteristic earthquake on the Greendale fault is 4.0×10^{-5}/yr, much lower than even the 2,475-year return period. Several reverse dipping faults are found to the north of the city in the Canterbury plains and offshore in Pegasus Bay, and these are associated with a higher rate of occurrence. This places Christchurch itself in a region of de-amplification with respect to a directivity neutral condition under most hypocenter scenarios. Once again, the net result is a reduction in hazard, which occurs due to the negative amplification predicted by the more active faults, combined with the slight reduction in σ resulting from the directivity amplification model.

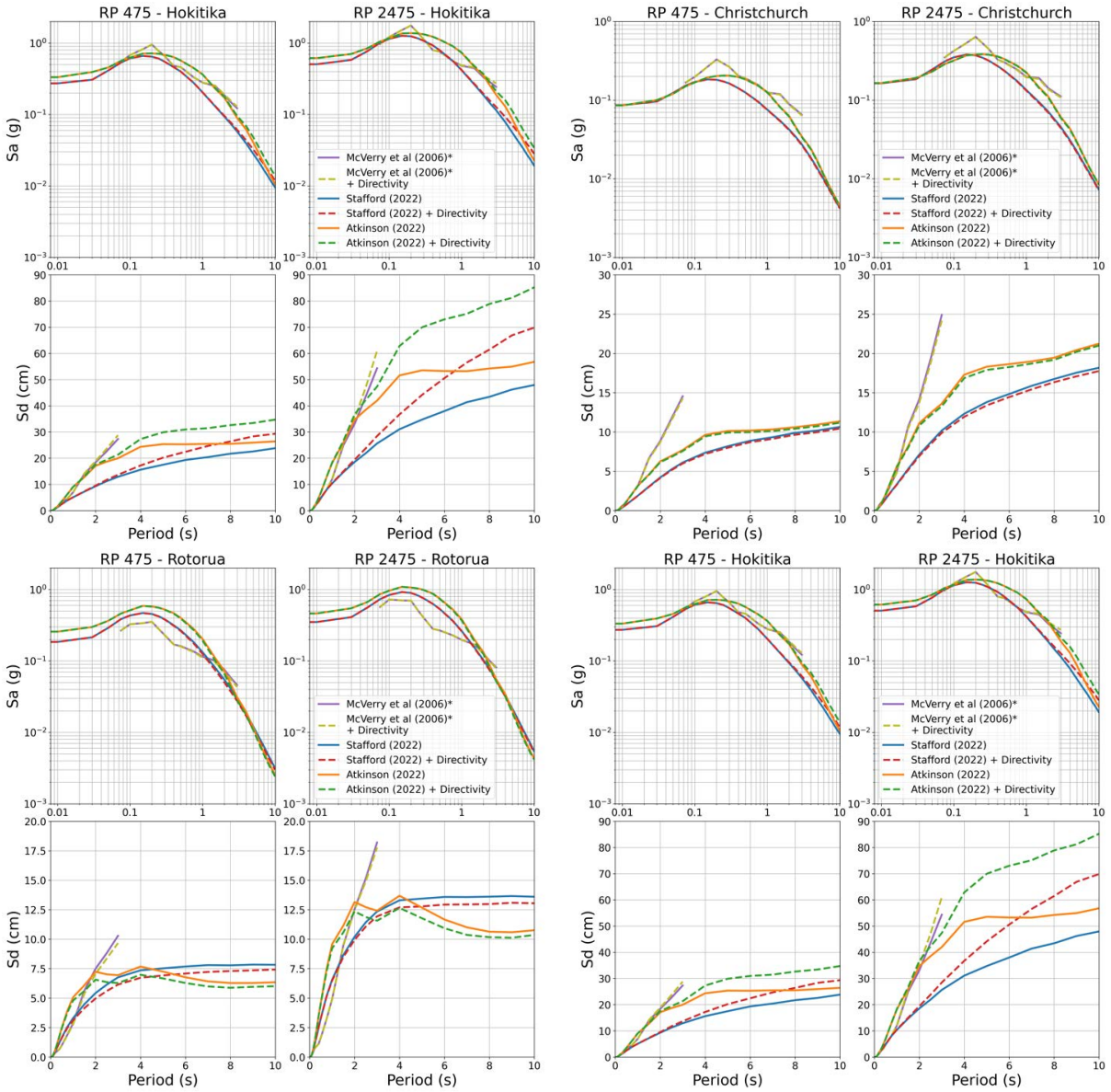

Figure 4.7: Comparison of UHS using the 2010 NSHM active fault sources for Wellingon (top left), Christchurch (top right), Rotorua (lower left) and Hokitika (lower right) for crustal faults only: with directivity (dashed line) and without directivity (solid line).

Hokitika and Rotorua are shown in these comparisons as they sit within regions of high positive and moderate negative amplification respectively. Hokitika sits close to the confluence of the Alpine, Clarence and Hope Faults and is thus optimally situated for positive directivity amplification from several ruptures with high rates of occurrence. Rotorua, meanwhile, is located close to a cluster of active faults in the northern Taupo region but sits several kilometers from the down-dip extent of the faults. Under most hypocenter positions it experiences de-amplification with respect to a directivity neutral condition, which yields a reduction in hazard.

The UHS also reveal another interesting feature of the directivity model, which relates to the period range affected. For Rotorua the greatest changes in the hazard are seen in the $2 \leq T(\mathrm{s}) \leq 4$ range, while for Hokitika the increase is seen at much longer ($T \geq 5\,\mathrm{s}$) periods. This reflects the influence of the narrow-band amplification of the directivity model and the magnitude-

dependence of t_p. This results in amplification at shorter periods for the $6.0 \leq M_{char} \leq 6.5$ faults in the northern Taupo region, while for the larger ruptures on the Alpine, Clarence and Hope faults the t_p is significantly longer, thus affecting the longer period accelerations to a greater degree but with a reduced impact for shorter period seismic hazard.

The results shown here illustrate several key points that are relevant in the application of directivity in PSHA and the interpretation of the changes in hazard. The first is that in many places the probability of exceedance of ground motion at a site may represent the net contribution from many active faults. In each case the degree to which directivity will increase or decrease the hazard will depend on the position of the site with respect to the rupture and the rate of occurrence of the rupture in question. The corollary to this point is that not only is the position of the fault important, so too is the activity rate and magnitude frequency distribution (MFD). In the case of the 2010 NSHM the use of characteristic ruptures with no uncertainty on magnitude results in a binary condition in which the fault either contributes to the change in hazard at a given annual rate of exceedance if the said rate is less than $\lambda_{M_{char}}$, or else the directivity will not influence the hazard (as was the Christchurch case) regardless of whether the position of the site is optimally situated for directivity amplification. Difference maps such as those in Fig. 4.5 and 4.6 may change significantly were the possibility of the fault to produce ruptures smaller than M_{char} to be considered or an entirely different MFD assumed.

The second major factor influencing PSHA with directivity is of course the uncertainty in the hypocenter position within the rupture. Whether one considers distributions assigning greater weight to bilateral or to unilateral rupture propagation, $f_H(h)$ is assumed to be symmetric with no specific preference for direction of propagation. Under the segmentation and characteristic earthquake assumptions present in the 2010 NSHM this focuses the regions of positive directivity amplification primarily to those areas just off the ends of the rupture (in the strike-slip case) or to areas close to the up-dip projection of the rupture to the surface (in the dipping fault case). This concentrates amplification at the ends of the characteristic rupture segments, while minimizing or even de-amplifying it toward the middle of the ruptures where ground motions are expected to represent a neutral or backward directivity condition under most hypocenter positions. When integrating over this uncertainty, the impact of directivity on the resulting PSHA is moderate compared to the more adverse deterministic scenarios.

Other factors that influence the resulting change in seismic hazard are the bandwidth of the pulse in the directivity model and the scaling of the pulse-period with magnitude. With most recent models being narrow-banded there is a stronger connection between the probability density of the magnitudes for a given fault source and the period range in which most amplification is to be expected. Fault source models pushing toward larger magnitude earthquakes will yield directivity changes in the resulting PSHA that may only be relevant for very long period motion. This may also have downstream consequences for seismic risk analysis, where fragility functions for most common structural types seldom adopt such long period intensity measures as predictors, or where the use of vector measures of ground motion such as Average Sa may smooth out the influence of directivity even for intermediate and long-period structures.

4.5 Applying directivity to the 2022 NSHM

4.5.1 Earthquake rupture forecast

The results of the calculations of the 2010 NSHM using directivity demonstrated that both the geometry and the magnitude frequency distributions of the fault ruptures play an important role in controlling the spatial trends in amplification. The radical change from the 2010 NSHM to the 2022 NSHM by way of adopting the grand inversion approach (Field et al., 2014; Page et al., 2014) will inevitably yield differences in the faults contributing to hazard at given return periods. For full details on the construction of the crustal fault source model via the grand inversion the reader is referred to Van Dissen et al. (2022). What is relevant for the application of directivity into PSHA are some of the key assumptions made within the grand inversion and how they change the crustal fault sources with respect to the 2010 NSHM.

The inversion fault model begins initially with the New Zealand Community Fault Model (NZCFM v1.0 Seebeck et al., 2023), which is transformed into sub-sections (or micro-segments) each with a length equal to half its down-dip width. This results in 2,325 sub-sections, each with a slip rate and rake value that is constrained for the section by both geological and geodetic observations. The complete set of sections and their styles-of-faulting are shown in Fig. 4.8. A minimum magnitude of M6.8 is applied to the fault model, on the assumption that smaller ruptures would have a higher likelihood of not breaking the surface; hence the paleoearthquake record would likely be incomplete. This change alone may have a significant impact on the spatial patterns of directivity in the Lake Taupo region, where previously the clusters of smaller, slower slipping extensional faults with $M_{char} < 6.8$ had been present. In the 2022 NSHM far fewer ruptures remain in this region and those that do represent connections of ruptures across several faults that have a lower probability of occurrence.

Given the sub-section geometries, their geological/geodetic constraints and the input modelling assumptions, initial rupture sets are created by connecting sub-sections into combinations that form plausible ruptures. These connections are subject to rules ("plausibility filters") that govern whether sections can rupture together, which include maximum jump ("gap") distance, cumulative change in rake, and coulomb stress change (Milner et al., 2022). No prior constraint on maximum magnitude is imposed, so it is the plausibility filters that control the degree of connectivity across sections. Application of these filters results in a rupture set of 232,906 feasible ruptures, ranging from ≈ 50 km in length to nearly 1,200 km. These ruptures sets are then input into the inversion process, where their corresponding rates of occurrence are determined so that they satisfy constraints of target magnitude frequency distribution and recurrence intervals of slip at measured sites. While more than 232,000 ruptures are input into the inversion, the outputs yield between 600 and 1,736 ruptures with non-zero occurrence rates per logic tree branch depending on different modelling assumptions and weighting of constraints. A total of 3,884 different ruptures are found across all 324 different logic tree branches.

The most critical feature of the ERF from the perspective of directivity modelling is the increased complexity of the constituent ruptures. Without a characteristic magnitude or maximum magnitude constraint, ruptures can accumulate sub-sections and grow significantly in size; often entraining sections from multiple faults in the process provided they comply with the plausibility

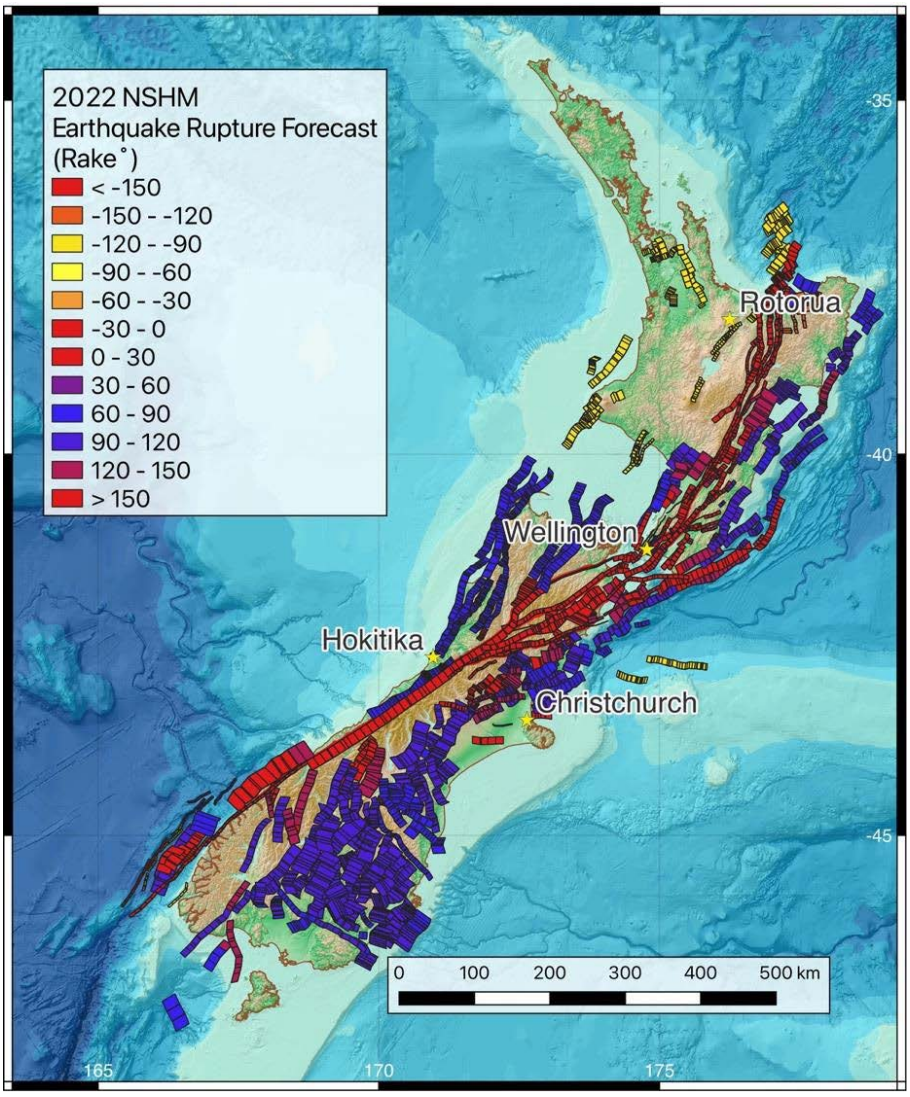

Figure 4.8: Complete rupture set from the 2022 NSHM earthquake rupture forecast with colors scaled according to rake.

constraints. The resulting ERFs therefore contain some large-magnitude ($M \geq 8.0$) earthquake sources with exceptionally long lengths, usually representing ruptures connecting several fault systems. The ruptures themselves may contain complex geometries, including offsets, step-overs, and significant changes of strike, dip and rake. With the potential for participation of segments from different fault systems we are likely to encounter ruptures that we would consider multi-fault ruptures, which are distinguished in the ERF by large gaps between consecutive sections and significant changes in dip and rake.

While complex multi-fault ruptures have been observed in New Zealand and worldwide, one of the most notable being the 2016 Kaikoura, New Zealand, earthquake (M_W 7.8), observations of ground motion from such events are still sparse and it cannot necessarily be assumed that existing directivity models can extrapolate to such large magnitudes and/or complex rupture geometries. The presence of multi-fault ruptures requires a different integration process to allow for multiple nucleation points corresponding to each of the different sub-faults each with their own rupture surface, hypocenter distribution, magnitude and rake. This prevents us from

being able to simply transfer to the 2022 NSHM the full randomization of hypocenters approach that we had applied to the 2010 NSHM. As two different calculations processes are needed for the two different fault rupture types inside of the PSHA calculation, there must exist first a multi-fault classifier within the calculation that divides the directivity model evaluation into different workflows depending on whether a rupture is multi-fault (with potentially multiple individual points of nucleation) or simply multi-segment (with a single point of nucleation). The requirement to implement such a drastic change within the PSHA calculation software would inevitably form a barrier to deployment of the models in many cases. Instead, we have adopted a different approach that leverages upon the strategy for directivity modelling on complex ERFs initially proposed by Al Atik et al. (2023) but aims for computational efficiency and scalability to allow for deployment at regional scale.

4.5.2 A machine learning based strategy for application to complex earthquake

4.5.2.1 Ruptures

The challenges of handling directivity models for the complex ERF source models are considerable and initially prohibited application of the randomized hypocenter approach in an efficient manner at a countrywide scale. When considering seismic hazard models with multiple ERFs on different logic tree branches this becomes intractably computationally demanding. Given the modest influence of directivity on seismic hazard for both spectral periods and return periods of engineering interest the cost of explicit implementation may outweigh the benefit. ERFs also have an advantage, however, which is that they describe common and static inventory of ruptures with non-zero occurrence rates. For each given rupture, target site, directivity model and assumed $f_H(h)$ the mean and variance of the amplification remain static for the respective rupture regardless of the probabilities of occurrence assigned to that rupture in each individual logic tree branch.

Recall from section 2 that the *modifier of moments* method requires calibration of the mean amplification factor and its standard deviation owing to $f_H(h)$ and $f_\gamma(\boldsymbol{\theta_{dir}})$, i.e., the *moment modifiers* $\overline{\Delta\mu}_{dir}(m, r, T|f_h, f_\gamma)$ and $\sigma_{f_H f_\gamma}(m, r, T|f_h, f_\gamma)$. Watson-Lamprey (2018) proposed generalized model for the moment modifiers, which is parameterized into a simpler functional form. While this may be appropriate in many contexts, there are potential limitations in the present case. These are, firstly, the ability to reproduce sufficiently the finer resolution spatial features of the moment modifiers using the simpler parametric functional form, and, secondly, the extent to which this can be adapted for extremely geometrically complex and/or multi-fault ruptures. Instead of a parametric modified moments model, we can approach the problem from a different perspective, which is that if we have the methodology to generate the directivity amplification fields in the complex cases we can define the spatial distribution of $\overline{\Delta\mu}_{dir}(m, r, T|f_h, f_\gamma)$ and $\sigma_{f_H f_\gamma}(m, r, T|f_h, f_\gamma)$ across a pre-defined mesh of sites unique to each rupture in the ERF. In a computational sense, this takes the form of a look-up table in which given spatial distributions of the moment modifiers can be calculated and stored in a database prior to the PSHA calculation, which can then be retrieved during the calculation for each respective rupture in the ERF according to the unique rupture identifier. No parametric simplification of the amplification

fields is needed, and the moment modifiers need only be calculated and stored once for a given model configuration and can be re-used in all subsequent calculations *ad infinitum*.

This premise is theoretically straightforward but encounters major computational problems for deployment in practice: data storage and bandwidth. For the ERF described in this section, and assuming a 5 km by 5 km resolution grid of sites covering all onshore New Zealand, the corresponding look-up table occupies approximately 1 Gb on disk. Retrieval of the data from the look-up table requires either data transfer from disk storage for each rupture or storage in shared access memory. Either approach creates calculation bottlenecks when deployed on smaller calculation servers and presents challenges for scalability to larger parallelized computing infrastructures. Instead of adopting look-up tables directly, we try to compact the information required to generate the spatial fields by over-fitting an artificial neural network to the data contained within the look-up tables. This allows us to represent each field in a smaller number of coefficients that can still effectively reproduce even complex fields to a sufficient level of accuracy.

4.5.2.2 A neural network based modified moments implementation

Artificial neural networks (ANNs) are a widely used machine learning tool inspired by the structure of the human brain. Due to the usually exceptionally high number of model-coefficients (neurons), ANNs can be considered generic templates for non-linear models that can be calibrated (trained) to solve a specific task if a large training dataset is available to learn from (Kong et al., 2019). Thereby ANNs have been found to adapt reliably and autonomously to complex, non-linear relations in datasets, with no need to explicitly predefine such relations in the form of a mathematical equation (Lecun et al., 2015). For numerous tasks such as image recognition (Krizhevsky et al., 2017), seismic phase detection in waveforms (Perol et al., 2018) and fault detection from seismic images (Xiong et al., 2018), ANNs have been shown to reliably outperform equation or rule-based models (Mousavi and Beroza, 2022).

4.5.2.3 Overfitting vs. generalization

Due to the large number of model coefficients, ANNs are heavily prone to overfit to data examples in the training dataset. Consequently, emphasis is usually placed on ensuring that the relations inferred during the training also provide valid predictions for examples outside the training dataset, by means of interrupting training as soon as overfitting is detected. The task of modelling moment modifiers for seismic hazard calculations is, however, different to most other use cases, since the ruptures that can be encountered within an ERF-based seismic hazard model is fixed and known *a priori*. Thus, besides a generalizable machine learning based model to generate moment modifiers within arbitrary hazard models, the development of an ERF-specific model can be worthwhile if it is easily and efficiently deployable within the corresponding PSHA. We therefore intentionally overfit an ANN to the amplification fields corresponding to the ruptures in the 2022 NSHM ERFs to obtain highly detailed reproductions of the moment modifiers for each unique rupture, at the cost of sacrificing the ability to generalize to arbitrary earthquake ruptures. Intentionally overfitting an ANN to a given dataset can be considered a form of dimensionality reduction, where the number of coefficients in the ANN architecture determines

the dimensionality of the compressed dataset. Technically speaking, the resulting model is an efficiently compressed lookup table that holds the moment modifiers for all the ruptures in the 2022 NSHM ERF. While the application of ANNs usually takes place in the context of deep learning, we want to emphasize that this is not the case here, since the suggested model does not acquire generalizable knowledge about moment modifiers in the vicinity of earthquake ruptures, but only learns a certain set of examples "by heart". The advantage of such a model compared to an actual lookup table is the compressed size that enables scalability inside seismic hazard calculations. Compared to an empirical model with an equational form, the suggested approach promises more accurate detail in the modelled spatial amplification patterns, especially for the geometrically complex and/or multi-fault ruptures included in the 2022 NSHM.

4.5.2.4 Generating the directivity fields for training the neural network

To develop the modifier of moments model via a neural network we need to first construct a data set of directivity amplification fields for the available ruptures in the ERF, varying the hypocenter position according to $f_H(h)$. We generate for each rupture N_{SAMP} fields of amplification using the Bayless et al. (2020) model, with each field corresponding to a different hypocentre position sampled from $f_H(h)$. As hypocenter position is the only additional aleatory variable in the Bayless et al. (2020) model, only $f_H(h)$ need be considered. A total of $N_{SAMP} = 250$ samples are used in the present case for each rupture in the set of 3,884 unique ruptures found across all 324 logic tree branches of the active shallow crustal ERF. The same distribution for $f_H(h)$ is assumed as that implemented for the 2010 NSHM; however, to allow more variability in the hypocentre position we apply a two-step sampling process where the corresponding along-strike and down-dip bin is selected first (as per the plots in Fig. 4.4), then within the bin X and Z are sampled from a uniform distribution bounded by the bin edges.

ERF-based models allow for the possibility of including complex ruptures with participation from micro-segments on different faults. Given the minimum magnitude of M_W 6.8 assumed in the 2022 NSHM, all ruptures in the ERF contain multiple microsegments. To calculate the directivity field, a different computational process is required depending on whether the rupture is classified as a multi-segment rupture or a multi-fault rupture. For the multi-segment case we assume a single hypocenter distribution and generate the directivity fields directly following an approach similar to that of the PSHA from the fully randomized hypocenter approach (albeit with the specific hypocenter position sampled randomly from $f_H(h)$ rather than enumerating the distinct bins). For the multi-fault case it is assumed that each participatory sub-fault to the rupture, which itself will be multi-segment, will be associated with its own hypocenter. To identify the multi-fault cases and generate the directivity amplification fields we use the following procedure, which is adapted from that proposed by Al Atik et al. (2023) for use with the UCERF3 ruptures in California.

1) A reference 2D grid of sites is first defined spanning all New Zealand and the surrounding region at a resolution of 5 km (161°E to 177°W and -51°N to -30°N). This reference grid has the shape 763 × 362, with each grid cell corresponding to a unique location on Earth.

2) Each rupture from the set of 3,884 unique ruptures found within the full set of ERFs

described by the logic tree is considered in turn. For that rupture we wish to define the directivity across an evenly spaced sub-grid of dimension 256 × 256, which is sliced from the main grid and centered on the cell enclosing the geographic centroid of the rupture.

3) For every unique rupture found within the composite sets of ERFs from the logic tree, each consecutive microsegment is assessed in turn. The rupture will be classified as multi-fault, and a new sub-fault created, if all the following criteria are met: i) the change in dip angle between consecutive fault segments is greater than 50°, ii) the change in first quadrant rake angle between consecutive fault segments is greater than 30°, iii) the gap between consecutive segments exceeds 10 km.

4) If no sub-faults are defined for the rupture, then the rupture is classified as multi-segment and directivity fields will be generated for each sampled hypocenter in the conventional manner.

5) If two or more sub-faults are defined then the rupture is split into N_{SUB} faults, each of which is characterized as an individual earthquake with its own magnitude, rake and fault surface. The magnitude of the rupture for each sub-fault is determined by dividing the total seismic moment of the original rupture proportionally to the rupture area of the sub-fault. Rake for the sub-fault is taken as the mean rake from each of the participating segments weighted by the segment area.

6) Hypocenter position is sampled independently on each sub-fault from $f_H(h)$such that sample i of N_{SAMP} will contain a set of N_{SUB} unique hypocenter positions. For each sub-fault and its respective hypocenter, the directivity amplification parameter is calculated for all target sites and the maximum directivity parameter from each of the sub-faults taken to represent the directivity field at the site for the given rupture.

7) The corresponding directivity fields for each of the N_{SAMP} hypocenter positions are stored for each rupture in a high-density binary file (hdf5).

The outcome of this process is a set of 971,000 directivity fields from 3,884 ruptures and 250 hypocenter positions, with each directivity field represented as a 256 × 256 cell grid of amplification according to the selected directivity model Bayless et al. (2020) and individual spectral period. A total of 11 periods between T = 0.75 s and T = 10 s are considered. To generate the samples for training the neural network, we calculate the mean and variance over the 250 hypocenter realizations for each rupture $\overline{\Delta\mu}_{dir}(m, r, T|f_h, rup_i)$ and $\sigma_{f_h}(m, r, T|f_h, rup_i)$ (here onward $\overline{\Delta\mu}_{dir,i}$ and $\sigma_{f_h,i}$ for short). This finally yields the dataset $Y = \{y_1, y_2, \ldots, y_{3884}\}$ where each y_i is a matrix of shape (256 × 256 × 22). The first and second dimension in y_i describe the 5 km resolution sub-grid for each respective rupture i, and the third dimension holds the 22 data channels (moment modifiers $\overline{\Delta\mu}_{dir,i}$ and $\sigma_{f_h,i}$ at 11 periods respectively). Along with the dataset Y we generate a corresponding vector $X = \{x_1, x_2, \ldots, x_{N_{RUP}}\} = \{1, 2, \ldots, 3884\}$ comprising the corresponding rupture IDs which are utilized as input values for the ANN in deployment.

4.5.2.5 Artificial neural network architecture

We implement a convolutional neural network (CNN), a type of ANN that has been designed specifically for tasks that involve processing of matrix-shaped data, such as multi-channel images (Lecun et al., 2015), and has since been adapted in various studies dealing with spatial data (e.g. Lilienkamp et al., 2022; Waldeland et al., 2018; Xiong et al., 2018). The CNN architecture provides the necessary properties to capture in detail the spatial patterns of the moment modifiers. The first layer of the ANN is designed to receive rupture identifier x_i as input, whereas subsequent layers process the input via embedding, convolution with filter banks, and up-sampling, to eventually generate an output $\hat{y}_i$ of shape $(256 \times 256 \times 22)$. For more details the reader is referred to the appendix (section 4.A, Fig. 4.13 and Table 4.1).

4.5.2.6 Training

After the ANN is initialized, the generated output $\hat{y}_i$ from any input x_i is just a randomly generated matrix. Therefore, the ANN must be trained, i.e., the coefficients in the ANN must be calibrated to generate an output $\hat{y}_i$ that closely resembles y_i by minimizing the mean squared error $MSE_i = \frac{1}{N_{sites}} \sum_n = 1^N_{sites}(y_{i,n} - \hat{y}_{i,n})^2$. Here N_{sites} is the number of sites on the regular sub-grid. Following this procedure repeatedly for all samples in the training dataset eventually leads to the ANN learning the moment modifiers $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{3884}\}$ for all rupture IDs. More detail is provided in the appendix (section 4.A, Table 4.2). The fully trained CNN model has a size of 5.9 MB, which corresponds to a reduction in data by a factor of over 150 compared to the corresponding lookup-tables of $\overline{\Delta\mu}_{dir,i}$ and $\sigma_{f_h,i}$ for all ruptures and spectral periods (∼1Gb). We observe a reproduction of the original amplification fields in detail throughout the dataset as demonstrated in the example shown in Fig. 4.9.

4.5.2.7 Evaluation of model performance

To quantify the similarity of the individual ANN predictions for rupture i, location n, and period T ($\hat{y}_{i,n,T}$) to the precalculated moment modifiers $y_{i,n,T}$, we evaluate the distance between the corresponding Gaussian probability density functions $\phi = \mathcal{N}(\overline{\Delta\mu}_{dir,i,n,T}, \sigma_{f_h,i,n,T})$ and the predicted moment modifiers from the ANN $\hat{\phi} = \mathcal{N}(\hat{\overline{\Delta\mu}}_{dir,i,n,T}, \hat{\sigma}_{f_h,i,n,T})$ via the earth mover's distance (EMD):

$$EMD(\phi, \hat{\phi}) = \int_{-\infty}^{\infty} |\Phi - \hat{\Phi}| \tag{4.8}$$

where Φ and $\hat{\Phi}$ denote the corresponding cumulative density functions. We calculate the EMD individually for each rupture, at each grid point n that is located onshore New Zealand and each period T for which the precalculated $\sigma_{f_h,i,n,T}$ is larger than 10^{-3}. A summary of the resulting EMDs and visualizations of individual comparisons that ease the interpretation of absolute EMD values are given in Fig. 4.10. Over 99 % of EMDs are smaller than 0.03, indicating a high level of agreement between the distributions. Visual comparison of distributions reaching this value indicates acceptable similarity.

In addition to the comparison of moment modifiers, we also evaluated the similarity of

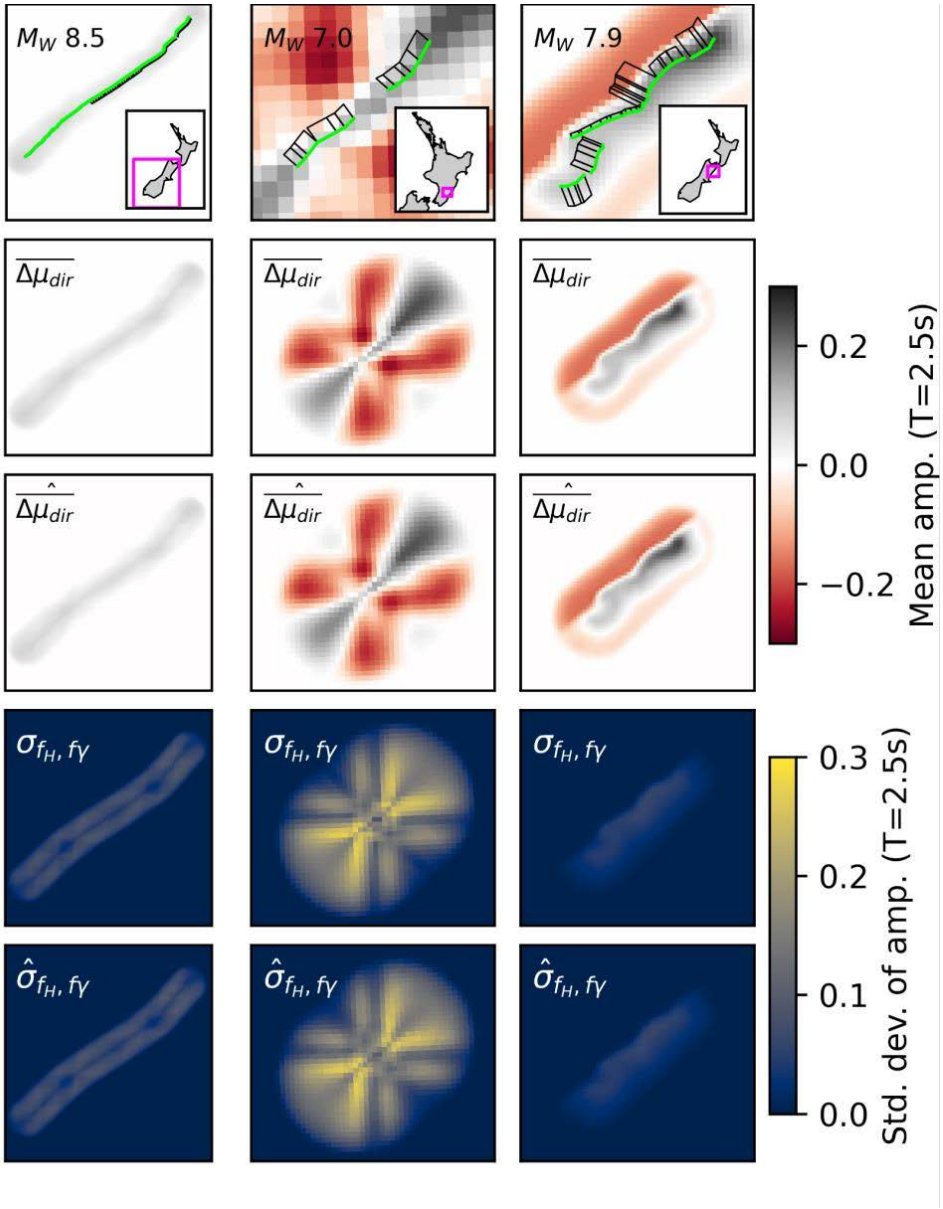

Figure 4.9: Examples of moment modifiers y_i ($Sa(T = 2.5\,\mathrm{s})$) (rows 2 and 4) and the corresponding ANN predictions $\hat{y}_i$ ($Sa(T = 2.5\,\mathrm{s})$) (rows 3 and 5) for the ruptures shown in row 1.

amplification samples obtained via consideration of all relevant ruptures (those for which $\sigma_{f_h,i,n,T}$ is larger than 10^{-3}) at a respective site. To this end, we conduct the following procedure for each grid point onshore New Zealand:

1) Identify all ruptures producing non-zero directivity amplification for the site.
2) Draw sample S and $\hat{S}$ of size 100 from ϕ and $\hat{\phi}$, respectively, for each relevant rupture.
3) Keep drawing until a total sample size of 10,000 is achieved.
4) Draw a final sample of size 10,000 without replacement if the sample size of 10,000 is exceeded.
5) Compare S and $\hat{S}$ via the earth mover's distance.

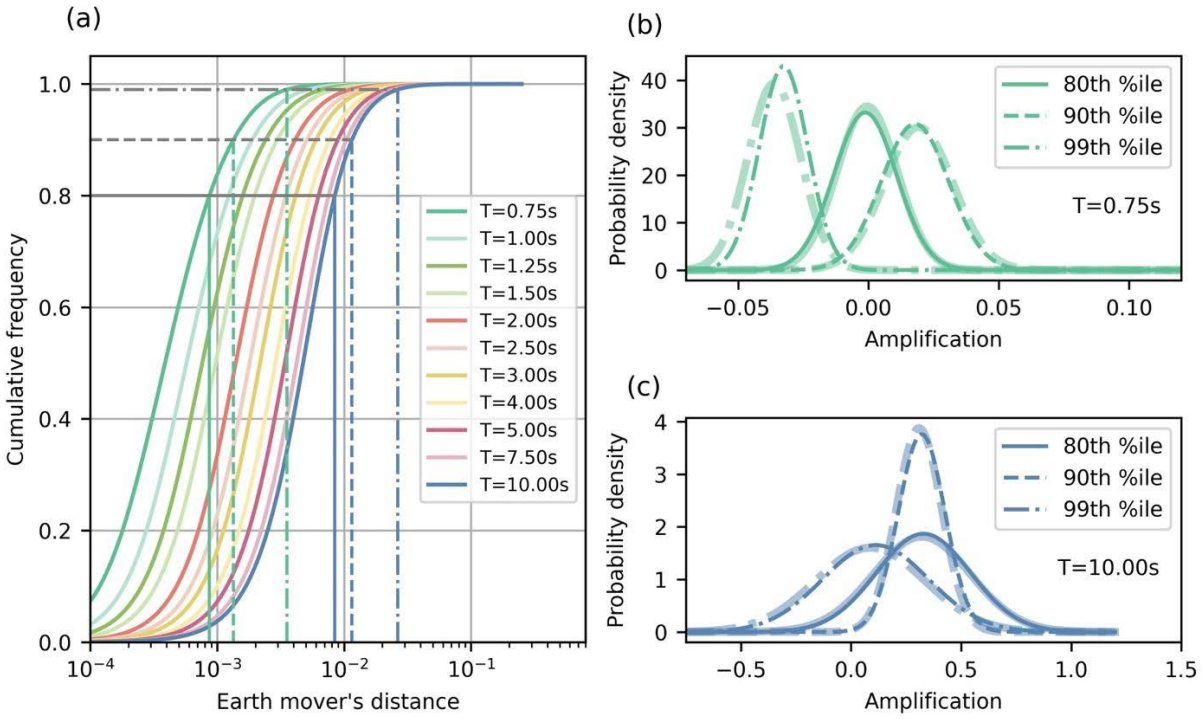

Figure 4.10: (a) Pixelwise comparison of amplification distributions from the precalculated moment modifiers y and the corresponding ANN reproductions $\hat{y}$ via the earth mover's distance (EMD). Example amplification cumulative distributions for EMD values at the 80th, 90th, and 99th percentiles are shown for periods of (b) 0.75 s and (c) 10.0 s. Line styles indicating the percentiles and colors indicating the period are used consistently throughout the sub panels.

A corresponding map of EMD values at a period of T=10 s, as well as additional information, is presented in Fig. 4.11. Over 99 % of EMD values are smaller than 0.01, and the visual comparisons of samples at locations where EMD ≥ 0.01 still indicate acceptable similarity. We therefore conclude that for the purposes of applying the overfit ANN moment modifier in PSHA, its ability to reproduce the spatial distributions of amplification and its variability from the underlying directivity model (Bayless et al., 2020) and $f_H(h)$ is satisfactory.

4.5.3 PSHA implementation for the 2022 NSHM

To demonstrate that the neural network based modifier of moments model can be implemented fully in practice, Fig. 4.12 shows the results of a complete PSHA calculation for New Zealand using the 2022 NSHM source model and the Stafford (2022) GMM, both with and without directivity, for a return period of 2,475 years and intensity measure $Sa(3.0\,\mathrm{s})$. For this calculation we use the *collapsed average* ERF, such that the probability of one or more occurrences of rupture i from each of the 3884 ruptures is defined by:

$$\bar{P}_{rup,i,j}(n \geq 1|t) = \sum_{k=1}^{N_k} \frac{w_k \cdot P_{rup,i,j}(n \geq 1|t,k)}{w_k}, \tag{4.9}$$

where $P_{rup,i,j}(n \geq 1|t,k)$ is the probability of occurence of rup_{ij} in time t according to the kth branch of N_k=324 logic tree branches, each with their respective weight w_k . No background

shallow crustal sources or subduction sources were included here. The implementation of the modified moments model was done using an experimental branch of OpenQuake.

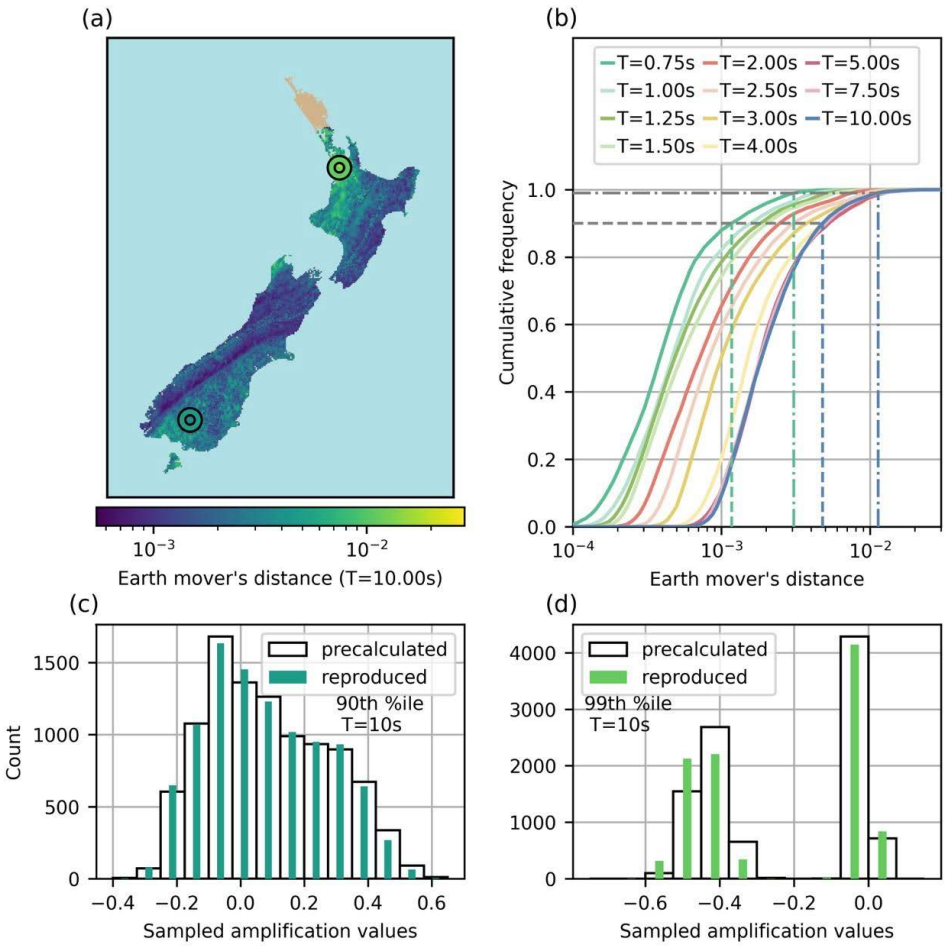

Figure 4.11: Map based comparison of amplification distributions from the precalculated moment modifiers y and the corresponding ANN reproductions $\hat{y}$ via the earth mover's distance (EMD). A comparison of the resulting samples per location via the EMD for the period T=10 s is shown in (a). Cumulative distributions of EMD values per period are shown in (b). Example samples for EMD values at the 90th and 99th percentile for the period T=10 s are shown in (c) and (d), respectively. The locations where these samples are obtained are indicated by the circular markers in (a). The circular markers in (a), as well as the histogram colors in (c) and (d) refer to the colorbar in (a).

In contrast with the spatial patterns of the change in hazard from the 2010 NSHM (compare against Fig. 4.6), the removal of segmentation in the 2022 NSHM results in a smoother, continuous band of amplification along the full extent of the most active strike-slip faults. The highest amplification (on the order of 15 – 20 %) can be seen toward the central and northern end of the Alpine fault, with the branching across the Hope, Clarence and Awatare faults in the northern South Island now very clearly defined. Likewise on the North Island, amplification close to the active strike slip Wellington, Wairarapa and Mohaka faults is clearly visible and now more continuous than before, while potentially significant directivity amplification from the cluster of normal faults close to Mount Egmont (southwest North Island) is now apparent. Elsewhere, the changes in hazard are generally modest, with the normal faults in the Taupo region showing far

less influence for directivity amplification than in the 2010 NSHM in the near source region and contributing to notable reduction in hazard farther away. Similarly modest changes of $< \pm 5\,\%$ can also be seen for other reverse faults in the southern South Island, albeit that the overall contribution of these faults to the hazard at the 2,475-year return period is small.

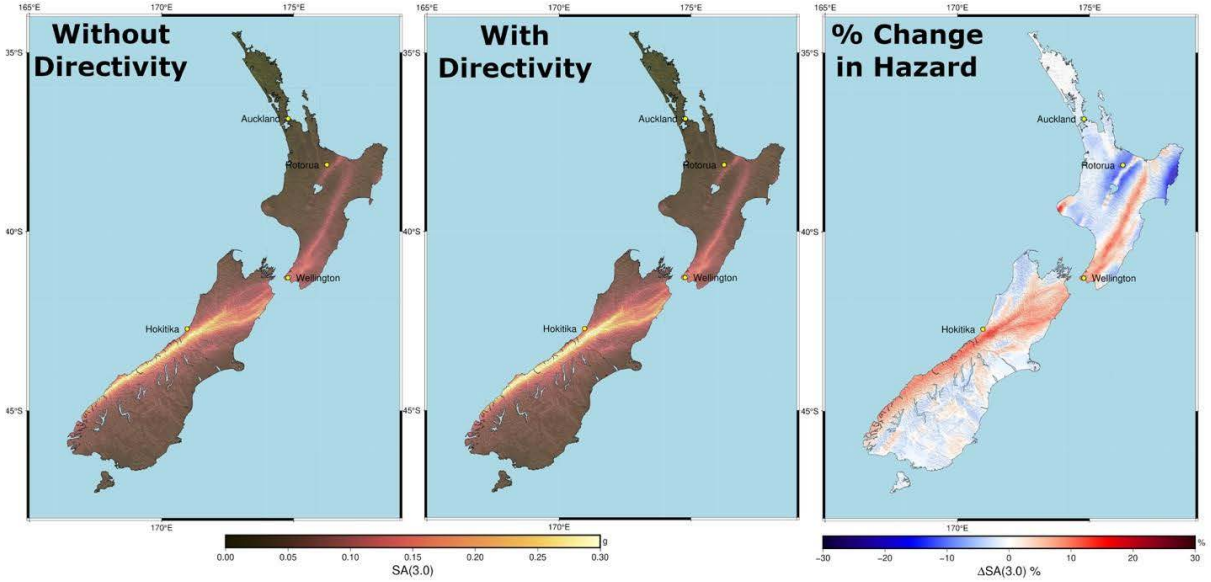

Figure 4.12: $Sa(3.0\,\mathrm{s})$ with a 2,475-year return period modelled using the collapsed average ERF of the 2022 NSHM and the Stafford (2022) GMM: without directivity (left), with directivity using the modified moments model (middle) and change in hazard with explicit directivity modelling (right).

Compared to the spatial pattern of amplification shown when using the 2010 NSHM (Fig. 4.5 and 4.6), the regions of significant increase in hazard for the 2022 NSHM (Fig. 4.12) come into closer agreement with those regions for which $N(T, D) > 1$ in NZS (2004) (shown in Fig. 4.1). While the actual increase in design ground motion when running PSHA with explicit directivity modelling is smaller than that implied by $N(T, D)$, the regions to which it applies are remarkably consistent. For both the 2010 and 2022 NSHM, even in the regions of highest amplification close to the active strike slip fault systems, the increase in hazard of 15 – 20 % is less than $N(T, D)$, which tapers from a maximum increase in hazard 36% at $D \leq 2\,\mathrm{km}$ and $T = 3.0\,\mathrm{s}$ (72 % for $T = 5.0\,\mathrm{s}$) to no increase at $D > 20\,\mathrm{km}$. This may be partly explained by the different directivity models assumed, but it is mostly due to the influence of $f_H(h)$, for which most rupture and hypocentre configurations will produce directivity amplification at a site that is lower than the adverse $M_W 7.5$ scenario on which $N(T, D \leq 2)$ is based (McVerry, 2003). In a probabilistic formulation, less adverse directivity conditions, including $\overline{\Delta\mu} < 0$, are more likely and the net change in hazard is lower. Implications of this analysis for potential modifications $N(T, D)$ in future New Zealand seismic design codes are discussed in more detail in Weatherill (2022).

As Fig. 4.5 and 4.6 showed for the 2010 NSHM, the impact of including explicit directivity modelling for shallow crustal faults diminishes significantly once the background and subduction sources are included, and though not shown here the same holds true for the 2022 NSHM. Disaggregation of seismic hazard according to source type for Wellington using the 2022 NSHM,

for example, showed a major contribution to hazard at long and short periods coming from earthquakes on the Hikurangi subduction zone (Gerstenberger et al., 2022b). This would suggest the changes in seismic hazard owing to directivity from shallow crustal fault ruptures on the Wellington, Wairarapa and Mohaka faults shown in Fig. 4.12 would be tempered by the large contribution of the subduction sources to hazard in this region, for which no explicit directivity modelling is included.

As one of our main obstacles to full implementation of explicit directivity modelling for the 2022 NSHM was computational demand, we can put the scale of the efficiency gain in perspective. Generation of the individual directivity amplification fields for all 3,884 ruptures using 250 hypocenter positions per rupture took almost two weeks of wall clock time (including storing the fields) on a 120 CPU server with 800 Gb RAM. The full size of the 971,000 directivity fields for all 11 periods (and associated metadata) required $\approx$ 3.1 Tb of disk space, with just the moment modifiers $\overline{\Delta\mu}_{dir,i,n,T}$ and $\sigma_{f_H,i,n,T}$ requiring 1 Gb. Training the neural network for the overfitting model required approximately 48 hours on a NVidia A40 GPU (circa 8000 epochs) and results in a model data file of 6 Mb. The PSHA calculation for all New Zealand using the collapsed average ERF for $>$ 17,000 sites took $\approx$ 10 minutes to run the same CPU server without directivity and $\approx$ 12 minutes with directivity applied. While ANN training was conducted on a GPU, regular CPUs were utilized for retrieving moment modifiers in the hazard calculations. The increase in RAM usage during the PSHA calculation was minimal. While the longer calculation time when including directivity using the neural network modifier of moments is inevitable, the use of this approach reduces our additional calculation time from a matter of days/weeks per branch to just a couple of minutes, while also retaining scalability to larger parallel computation infrastructures owing to the smaller memory allocation and data transfer. This gain in computational efficiency may be sufficient to reverse the potentially unfavorable cost-to-benefit ratio of including directivity explicitly in PSHA at regional scale.

4.6 Conclusions

The influence that high-amplitude, pulse-like ground motions in the near-source region of a rupture can have on structures is well observed, but characterization of rupture directivity in PSHA remains challenging, especially at a regional scale. To our knowledge, the implementation of directivity following a fully randomized hypocenter approach using the 2010 NSHM for New Zealand is only the second such regional scale application, with the Al Atik et al. (2023) efforts for California being the first. Regional scale applications such as this not only test the limits of the extent to which explicit directivity modelling in PSHA is feasible, but they also highlight the various factors that can influence the change in seismic hazard, showing that the net effects may not be so straightforward as simply increasing hazard in the near-field region of the fault. The full impact on seismic hazard across a region will depend on both the geometrical configuration of the faults, along with their rates of occurrence, magnitude frequency distributions and the potential interactions of the different fault sources in a region. In complex systems of faults, the expected patterns of amplification and de-amplification of the ground motion with respect to the directivity neutral condition may interfere both constructively and destructively, sometimes

varying considerably over relatively short distances. When combined with the assumption of symmetry in the distributions of hypocenter position with the ruptures, and the magnitude-dependence of t_p that may shift amplification for larger magnitude events to periods longer than those of engineering interest for the structure or application in question, the overall impact of explicit directivity modelling on seismic hazard may be modest. While there still exists considerable epistemic uncertainty in the directivity models, particularly for dip-slip events, and while full hypocenter randomization incurs a significant computational cost, a justifiable argument can be made for retaining still the implicit directivity modelling approach in probabilistic seismic hazard and risk analysis in many cases.

Despite the radical change in the source model, the difference maps for both the 2010 NSHM (Fig. 4.5 and 4.6) and 2022 NSHM (Fig. 4.11) are broadly consistent with one another in terms of both the spatial pattern and degree of change in seismic hazard level. There are notable differences, such as the removal of strict segmentation resulting in a smoother pattern of amplification along the lengths of the large fault systems in the 2022 NSHM rather than the irregular focusing of amplification around the ends of the segments in the 2010 NSHM, but for many applications these may not prove to be especially significant. Similarly, comparisons of the change in probabilistic hazard against the deterministically calibrated NZS 1170.5 near-fault factor $N(T, D)$ (Fig. 4.1) show that although $N(T, D)$ itself may be somewhat conservative in predicting higher amplifications than the probabilistic models, the criteria for application of $N(T, D)$ and the distance-dependent tapering result in similar spatial patterns of amplification. If one accepts that a degree of conservatism in this regard may be warranted for seismic design code applications, the current study based on a full PSHA including rupture directivity does not reveal a significant shortcoming to the near-fault ground motions in current provisions.

Moving to an ERF source model introduces significant challenges for directivity modelling in PSHA, as we have seen here in the 2022 NSHM and observed by (Al Atik et al., 2023) for UCERF3. Ruptures contained within these ERFs can describe source scenarios quite different from those upon which even the most state-of-the-art directivity models are based. These can include complex ruptures with participation from multiple faults with different slip types, along with extremely long ruptures producing earthquakes with magnitudes well above M 8. The work shown here, and in Al Atik et al. (2023), illustrates that the next generation of directivity models will need to consider ruptures of this nature, likely requiring constraint from physics-based simulation given the limited number of near-fault records from observed complex multi-fault ruptures such as 2016 Kaikoura earthquake (M_W 7.8).

An important focus of our efforts in the current analysis has been directed toward feasibility of wider implementation in PSHA, by which we refer to the general ability of hazard modelers to be able to execute part or all of the calculations shown here for their own specific applications potentially with more limited computational resources. Though PSHA using the full randomization of hypocenters may be considered the most desirable approach, in this case it was computationally prohibitive for the 2022 NSHM and would have required extensive modification of the calculation software. The potential advantages of using the modifier of moments approach have been well documented by Watson-Lamprey (2018) and Donahue et al. (2019), but existing approaches have focused on using closed-form parametric models to approximate

$\overline{\Delta\mu}$ and σ_{f_H}. These may be efficient, and applicable in many contexts, but may also struggle to capture the detailed azimuthal and distance dependent characteristics of the underlying models, particularly when the ruptures themselves are complex. The modifier of moments model presented here, which is based on over-fitting a neural network to the mean and variance of the amplification fields from an ERF, capitalizes on the fact that in an ERF source model all the ruptures are known prior to the PSHA calculation. This allows us to define all the amplification fields independently of the PSHA software, which may allow for more flexibility or detail in modelling the directivity (such the multi-fault ruptures considered here), the outputs of which we can then compact into a lower dimensional representation for input into the PSHA calculation. While this compaction will inevitably result in some loss of accuracy, this does not necessarily result in systematic magnitude, distance or azimuthal biases in the mean and standard deviation of the amplification fields and allows for better replication of the amplitude and spatial pattern of amplification in the resulting regional scale seismic hazard analysis.

Despite the developments in implementing explicit directivity modelling in PSHA that we present here, there remain many open questions as to how it should be approached in practice. This study and preceding ones have assumed symmetry in the distribution of hypocenter position within the rupture, thus implying that the direction of rupture is equally probable. This is an ergodic assumption that arises from consideration of global databases of ruptures, but for any given fault there may be a preference for a particular direction of rupture depending on its structure and the properties of the surrounding material (Ben-Zion and Sammis, 2003). A change $f_H(h)$ toward an asymmetric or even exclusively uni-directional distribution would produce a significantly different pattern of amplification around the fault in a PSHA context. Fault-specific hypocenter distributions can be easily accommodated in PSHA using the approaches adopted here, but how to constrain such distributions from available data and how much evidence of preference in rupture direction is needed before a global model$f_H(h)$ can be superseded by a fault-specific model is unclear. Detailed dynamic modelling of rupture evolution on specific faults will be needed to resolve this question, which will take time to develop in practice. Other open questions influencing practical application of PSHA with explicit directivity modelling are whether such models (or newly developed ones) can apply to other types of seismogenic sources, such as subduction interface ruptures or earthquakes from distributed seismicity "background" sources, and how we can constrain epistemic uncertainty in near-fault ground motions without double counting that emerging from the GMMs. For the latter, we note that the modifier of moments framework presented here can be adapted to accommodate scenario-specific (even rupture-specific) epistemic uncertainties, if they can be defined by the modeler.

A final perspective we can take from this work is that we have shown that it is possible to integrate machine learning based strategies to predict spatial characteristics of ground motion into a full PSHA framework, and to do so with enough efficiency to run a regional scale hazard calculations that can adapt to different computational infrastructures. Although we overfit the neural network here to predict amplification for a fixed set of ruptures, efforts are also ongoing to expand this to a generalized model in a "true" deep learning manner; one in which the amplification and its variability can be predicted for any rupture based on a tractable number of predictive parameters normally calculated by PSHA software. For the time being we are using

existing models from the literature to generate the spatial fields of amplification (Bayless et al., 2020; Chiou and Spudich, 2013), but this same framework can also be transferred to consideration of distributions of near-fault ground motions found in large databases of physics-based ground motion simulations. As a data-driven, non-parametric form of directivity amplification model, neural networks like those adopted here have the potential to form a crucial connection between physics-based ground motion simulations and probabilistic seismic hazard modelling. The spatial distributions of expected amplification of ground motion (and its variance) with respect to a parametric, directivity neutral and (potentially) non-ergodic GMM (Abrahamson et al., 2019; Landwehr et al., 2016) can be calibrated by training the neural network from a suite of physics-based simulations for a specific fault or fault system. The neural network itself can then predict the patterns of amplification for each rupture in an ERF, capturing the nuances of the spatial patterns implied by the simulation data. It can also do so at a computational cost that is not significantly greater than using a parametric GMM and is potentially scalable to massively parallel computing infrastructures. A process such as this may prove to be a practical, efficient and computationally scalable pathway for integration of physics-based ground motion simulations into PSHA without the need for radical re-design of existing software.

4.7 Data and Resources

The seismic hazard model files for the 2010 NSHM and 2022 NSHM were provided by GNS Science directly (https://nshm.gns.cri.nz/), while computation of the seismic hazard is undertaken using a customized version of OpenQuake (Pagani et al., 2014). The OpenQuake-compatible implementation of the Bayless et al. (2020) directivity model was verified and tested against code provided by the authors (available for download from https://www.jeff-bayless.com/papers/). Maps shown in this paper were constructed using both QGIS (https://qgis.org) and Generic Mapping Tools (GMT v6) (https://www.generic-mapping-tools.org/). Topography and bathymetry shown in the maps is taken from the 2023 Gridded Bathymetric Dataset produced by the General Bathymetric Chart of the Oceans (GEBCO Bathymetric Compilation Group 2023, 2023). Neural network training, inference, and optimization was implemented with Tensorflow (Abadi et al., 2015), Keras (Chollet et al., 2015), and Intel Neural Compressor (Intel®NeuralCompressor, 2023), while all other data analysis utilized the multiple tools from the Scientific Python ecosystem (Numpy, Scipy, Pandas, GeoPandas, Matplotlib etc.) (https://numfocus.org/)

4.8 Declaration of Competing Interests

The authors declare no competing interests.

4.9 Acknowledgements

The work presented in this manuscript was undertaken as part of the New Zealand National Seismic Hazard Model Revision Project, led by GNS Science, New Zealand. We thank the

project coordinators and ground motion characterization model (GMCM) working group for their support and feedback: Matt Gerstenberger, Brendon Bradley, Gail Atkinson, Peter Stafford, Sanjay Bora, Robin Lee, Elena Manea, Marco Pagani, Jess Hutchinson, Chris de la Torre, Anne Hulsey, Anna Kaiser and Russ van Dissen. Additional guidance, feedback and discussion on directivity in PSHA was provided by Kyle Withers, Jeff Bayless, Linda Al Atik, Nico Luco and Nick Gregor. Data for the 2022 NSHM was provided by Chris Di Caprio. We also thank Fabrice Cotton for feedback and review of the original manuscript and for supervision of the PhD of Henning Lilienkamp. Henning Lilienkamp acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS).

4.A Appendix

4.A.1 ANN architecture

The ANN architecture implemented in this study is presented in Fig. 4.13. The model input is a scalar ID $x_i, i \in [1, 2, \ldots, 3884]$, that uniquely identifies a rupture from the 2022 NSHM ERF model. This ID is transformed into a 128-dimensional embedding vector, a higher dimensional representation of the input, where the transformation procedure itself is subject to optimization during training. In a second step the embedding vector is reshaped to a matrix of size $(4 \times 4 \times 8)$. In the following we will refer to the first two numbers as the spatial dimensions, and to the third number as the number of channels. The following UpConv operations involve upsampling by a factor of 2 in the spatial dimensions, and subsequent convolution with a filter bank of size N_{fil}, that determines the number of channels. After each UpConv operation, the output is processed through a Leaky Rectified Linear Unit (Leaky ReLU) nonlinear activation function. The final step is a 2D convolution followed by a sigmoidal nonlinear activation function that yields the output matrix. More details regarding the configuration of are given in Table 4.1.

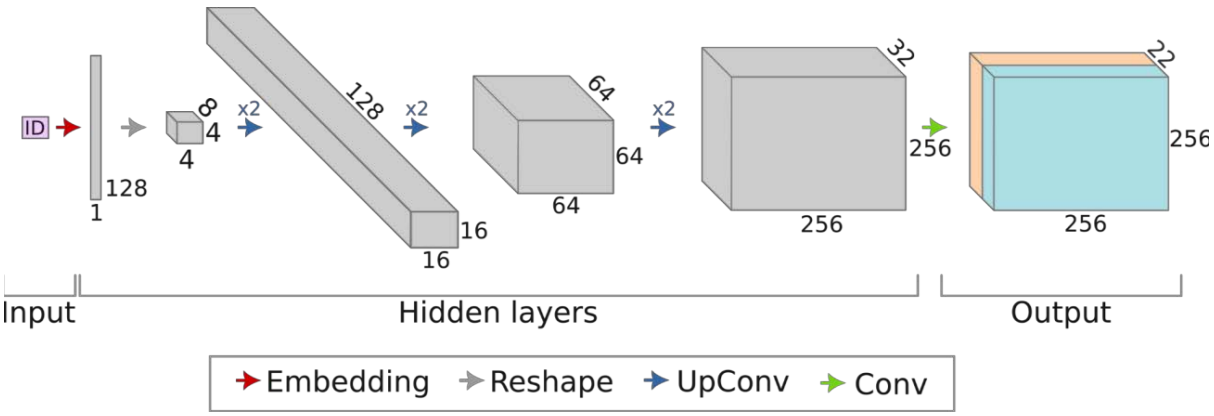

Figure 4.13: Convolutional neural network architecture implemented in this study. The model input is a rupture ID x_i from the New Zealand 2022 NSHM. The model output $\hat{y}_i$ is a matrix of shape $(256 \times 256 \times 22)$, where the maps of shape (256×256) cover a regular grid with a resolution of 5 km, and the 22 channels provide estimates of the moment modifiers at 11 periods, respectively.

Table 4.1: Technical specifications of the artificial neural network utilized in this study.

ID	Layer type	# Filters	Filter size	Input shape	Output shape	Activation	Comment
1	Input			1×1×1			Rupture ID
2	Embedding			1×1×1	1×1×128	Leaky ReLU	Vocabulary size of 3884
3	Reshape			1×1×128	4×4×8	Leaky ReLU	
4	Conv2DTrans	128	5×5	4×4×8	8×8×128	Leaky ReLU	
5	Conv2DTrans	128	5×5	8×8×128	16×16×128	Leaky ReLU	
6	Conv2DTrans	64	5×5	16×16×128	32×32×64	Leaky ReLU	
7	Conv2DTrans	64	5×5	32×32×64	64×64×64	Leaky ReLU	
8	Conv2DTrans	32	5×5	64×64×64	128×128×32	Leaky ReLU	
9	Conv2DTrans	32	5×5	128×128×32	256×256×32	Leaky ReLU	
10	Conv2D	22	9×9	256×256×32	256×256×22	Sigmoid	
11	Output				256×256×22		Moment modifiers at 11 periods

4.A.2 ANN training

In a first step, the untrained ANN is fed with a random rupture ID x_i, such that a random output $\hat{y}_i$ is generated. This output is then evaluated against the respective, precomputed moment modifiers y_i via the mean squared error loss function. Subsequently, backpropagation is conducted in which the Adam optimization algorithm (Kingma and Ba, 2015) is utilized to calculate derivatives with respect to all ANN coefficients, and a subsequent update of coefficients is performed to minimize the value of the loss function. The exact configuration of the Adam optimizer is given in Table 4.2. This procedure is repeated 8,000 times for all rupture IDs in the dataset, such that the ANN iteratively learns to relate the given rupture IDs to the desired moment modifiers.

Table 4.2: Configuration of the Adam optimizer used for training the ANN. One epoch is one training period during which all data examples are utilized once for training. The batch size of 1 indicates that the ANN coefficients are updated for each single sample. The remaining parameters refer to the configuration of the ADAM optimizer and are explained in more detail in Kingma and Ba (2015).

# Epochs	Batch size	Learning rate	Optimizer	β_1	β_2	$\hat{\varepsilon}$
8000	1	1.e-04	ADAM	0.9	0.999	1.e-07

Chapter 5

Towards a Generalizing Deep Learning Based Modifier of Moments Model for Explicit Incorporation of Directivity in Probabilistic Seismic Hazard Analysis

5.1 Introduction

The study presented in the previous chapter is largely concerned with the explicit incorporation of directivity related ground motion amplification in probabilistic seismic hazard analysis (PSHA) calculations, specifically within the 2022 New Zealand National Seismic Hazard Model (Gerstenberger et al., 2022a). My key contribution to this study was the development of an artificial neural network (ANN) based modifier of moments model for which I, contrary to the common practice in ANN studies, decided to intentionally overfit an ANN to learn the moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H} for the 2022 New Zealand earthquake rupture forecast (ERF) (Gerstenberger et al., 2022b; Seebeck et al., 2023) in great detail. As a consequence, the ANN model, hereafter ANN_{ovf}, did not learn meaningful relations between earthquake rupture properties and associated amplification fields, but rather learned a massively compressed representation of a specific set of moment modifiers "by heart". In the case of a seismic hazard model operating on a predefined ERF, this approach poses a practical, efficient, and accurate solution to the problem of explicit incorporation of directivity in PSHA studies at regional scale. However, when facing the situation of simulating directivity for ruptures that are not part of a precalculated inventory, e.g., in ShakeMap calculations or in non-ERF-based PSHA, the necessity of a modifier of moments model generalizing to arbitrary rupture geometries arises. To extend the approach to such cases, I started to explore a deep learning-based approach to derive a model that learns meaningful relations between the geometric and seismic properties of a rupture and the moment modifiers in its vicinity. In this chapter I present this alternative approach that could potentially broaden the applicability of ANN-based modifier of moment models within the engineering seismology community.

5.2 Data

The parameters to be modeled are the moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H} in the vicinity of an earthquake rupture, where the former describes the mean directivity related amplification at a site from averaging over various hypocenter locations h on the rupture plane sampled from the distribution $f_H(h)$, and the latter is the corresponding standard deviation. The moment modifiers utilized for training and validation in model calibration in this study are a subset of the synthetic dataset generated within the scope of section 4.5.2.4 using the Bayless et al. (2020) directivity model. Moment modifiers were calculated on regular grids of shape 256 × 256 with a spacing of 5 km for the $N_{rup} = 3884$ ruptures from the 2022 New Zealand ERF model. Contrary to the study presented in chapter 4, only one period (T=10 s) is considered. Consequently, the dataset $Y = \{y_1, \ldots, y_{3884}\}$ can be described as a matrix of shape (3884 × 256 × 256 × 2), where the 2 channels contain the moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H} at a period of T=10 s, respectively. In modeling contexts, I will refer to these precalculated moment modifiers as "targets".

A total of 9 predictive parameters describing the seismic properties of a rupture and the relative location of a site with respect to it were utilized. The moment magnitude M_W largely determines the overall amplitude of directivity related amplification, as well as the affected frequency band. The style of faulting indicator $\mathbb{I}_{SS}$ distinguishes between primarily strike-slip ruptures ($\mathbb{I}_{SS} = 1$), and others ($\mathbb{I}_{SS} = 0$) which produce fundamentally different spatial amplification patterns. The GC2 coordinates $gc2_u$ and $gc2_t$ (Spudich and Chiou, 2015), the along-strike horizontal distance of a site to the rupture r_y, and the azimuth angle az of a site with respect to the rupture's centroid, represented by $sin(az)$ and $cos(az)$, describe the relative position of a site with respect to the rupture. This information on relative site location is crucial to grasp the direction dependence of rupture forward directivity. The Joyner-Boore distance r_{jb} and the rupture plane distance r_{rup} describe the absolute distance of a site to the rupture, which is a necessary information given that directivity effects are primarily observed in the near field of ruptures. It should be noted that the parameters used are those from the Bayless et al. (2020) directivity model, supplemented by r_{jb} and r_{rup}, and that the calculation of all parameters is obligatory in typical PSHA calculations, such that no additional computational effort is expected in this regard. All predictive parameters are calculated on the same regular grid used for the moment modifiers, resulting in a predictive dataset $X = \{x_1, \ldots, x_{3884}\}$ of shape (3884 × 256 × 256 × 9). An example for an input x_i and a target y_i regarding a single rupture is shown in Fig. 5.1.

In addition to the 2022 New Zealand ERF, also a test set with 550 generic, planar ruptures evenly covering large magnitude, strike, and dip ranges was generated (See Fig. 5.3 e – h). The purpose of this test set is to investigate to which degree the final model is capable of providing reasonable moment modifiers for ruptures that are not included in the New Zealand 2022 ERF in this form. The data generation is the same as for the ERF dataset, therefore the test set can be described by matrices X_{test} and y_{test} of shapes (550 × 256 × 256 × 9) and (550 × 256 × 256 × 2), respectively.

5.3 Methodology

The aim of this study is to train an artificial neural network to provide estimates $\hat{y}_i$ of the moment modifiers y_i for an arbitrary rupture i that is characterized by the predictive parameters x_i. To this end I employ the U-Net neural network architecture (Ronneberger et al., 2015), due to its capability of processing data in the shape of arrays, e.g., maps, that has already been utilized in the ground motion modeling study presented in chapter 3.

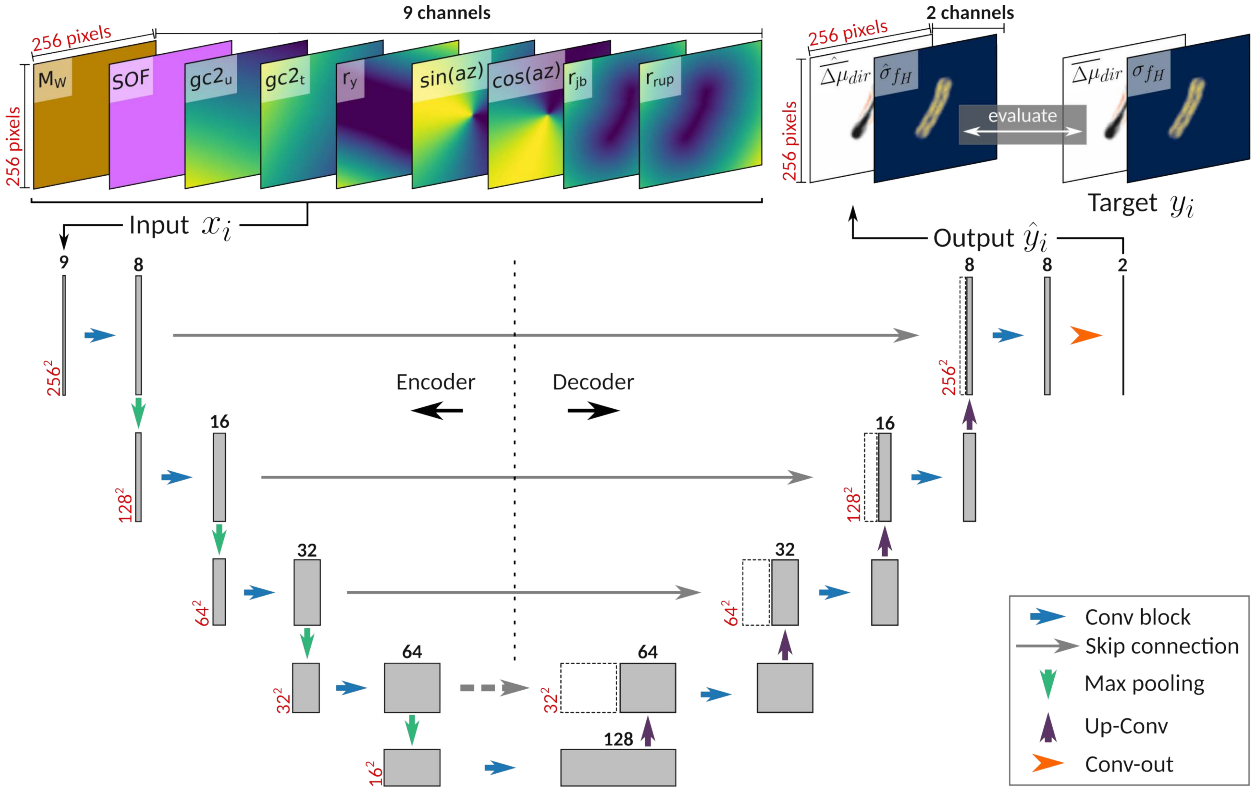

Figure 5.1: The U-Net architecture employed in this study following Ronneberger et al. (2015). The input layer perceives the predictive parameters x_i describing an earthquake rupture in the form of 2D arrays (maps). During training, the input is propagated through the U-Net until an output $\hat{y}_i = (\hat{\overline{\Delta\mu}}_{dir,i}, \hat{\sigma}_{f_H,i})$ is generated. The model coefficients are subsequently updated to minimize the misfit (mean squared error) between the output and the target moment modifiers $y_i = (\overline{\Delta\mu}_{dir,i}, \sigma_{f_H,i})$. Red numbers indicate the map sizes throughout the U-Net. Bold, black numbers indicate the number of data channels. This figure is based on Figure 1 in Ronneberger et al. (2015).

The adaption of the U-Net architecture to the task of estimating directivity related moment modifiers of arbitrary earthquake ruptures is depicted in Fig. 5.1. Maps of the 9 predictive parameters describing a rupture form the input x_i of shape $(256 \times 256 \times 9)$ that is received by the input layer of the U-Net. Subsequently, this input is propagated through the architecture until an output $\hat{y}_i$ of shape $(256 \times 256 \times 2)$ is generated, which is supposed to yield estimates $\hat{\overline{\Delta\mu}}_{dir,i}$ and $\hat{\sigma}_{f_H,i}$ of the moment modifiers y_i once the U-Net has been trained. Additional details regarding the U-Net architecture are given in Table 5.1 and in the appendix (section 5.A).

Before the U-Net can provide meaningful estimates of moment modifiers, training is required in order to teach the U-Net how to extract the relevant information from the predictive param-

eters and how to transform it into the moment modifiers. To this end, 80% of the examples in the available dataset (the training dataset) are used to calibrate the U-Net coefficients, whereas the remaining 20% of the data (the validation set) are used to ensure that the model generalizes to data samples outside the training set, i.e., to detect and avoid overfitting. Each training sample can be used in 4 different orientations, which effectively increases the number of training samples by a factor of 4. The training procedure is implemented as follows:

1 Insert predictive parameters x_i for an arbitrary rupture i to the U-Net

2 Process the input through the U-Net such that a (random) output $\hat{y}_i$ is generated

3 Repeat steps 1-2 for a total of 8 ruptures (a mini data batch of size 8)

4 Evaluate the loss between the outputs and the expected moment modifier maps via the mean squared error (MSE)

5 Employ the Adam optimization algorithm (Kingma and Ba, 2015) together with back-propagation (Rumelhart et al., 1986) to update the model coefficients to minimize the MSE

6 Repeat steps 1-5 for all data samples in the training set

7 Evaluate the MSE loss on the validation dataset (validation loss)

8 Repeat steps 1-7 until the validation loss does not decrease any further.

During this procedure, the U-Net iteratively learns a meaningful relation between the predictive parameters and the moment modifiers. Additional details of the training procedure are given in Table 5.2. Training of a U-Net with the specification given in Table 5.2 takes approximately 8 hours on a machine with an NVidia A40 GPU, and 192GB of RAM. The size of the final model on disk is 2.9 MB (after optimization utilizing IntelNeuralCompressor (Intel®NeuralCompressor, 2023)).

5.4 Results

A summary of the U-Net performance on the training, validation, and test sets is presented in Fig. 5.2 where a normalized mean squared error given by:

$$MSE^* = \frac{1}{2N_s}\sum_{j=1}^{N_s}\left(\frac{\overline{\Delta\mu}_{dir,j} - \hat{\overline{\Delta\mu}}_{dir,j}}{\max(|\overline{\Delta\mu}_{dir,j}|)}\right)^2 + \left(\frac{\sigma_{f_H,j} - \hat{\sigma}_{f_H,j}}{\max(|\sigma_{f_H,j}|)}\right)^2 \tag{5.1}$$

is used as the evaluation metric. Here, max denotes the maximum of a sample, and N_s denotes the number of sites with non-zero amplification. For comparison purposes, the performance of ANN_{ovf} on the validation ruptures from this study is included. Note that the term 'validation ruptures' is somewhat misleading in the context of ANN_{ovf}, because these ruptures were actually used for training of ANN_{ovf}. The smaller MSE^* values on examples from the training set compared to the validation set indicates slight overfitting of the U-Net. ANN_{ovf} yields

considerably smaller MSE^* values on the validation ruptures which is expected, given that it was intentionally overfit massively to these examples. The comparatively large values of MSE^* obtained for the generic test set indicate limitations of the U-Net in terms of generalization to simple, planar ruptures.

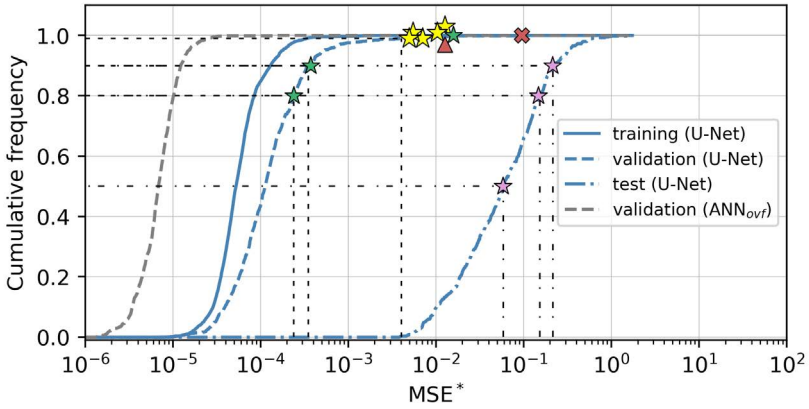

Figure 5.2: Cumulative frequencies of MSE^* scores obtained from the U-Net model (blue) and ANN_{ovf} (grey). Solid/dashed/dash-dotted lines show performance on validation/training/test ruptures, respectively. The green stars indicate example ruptures from the validation set at the 80th, 90th, and 99th %iles presented in Fig. 5.4. The red markers indicate two special cases from the validation set for which the U-Net predictions fundamentally differ from the target moment modifiers (x-shaped marker Fig. 5.6 a–e, triangular marker Fig. 5.6 f–j). The yellow stars indicate the remaining validation samples with MSE^* scores beyond the 99th %ile. The violet stars indicate example ruptures from the test set at the 50th, 80th, and 90th %iles presented in Fig. 5.5.

A more detailed visualization of MSE^* values as a function of moment magnitude M_W and dip angle φ for ruptures from the validation and test sets is presented in Fig. 5.3. Since no systematic variation of MSE^* with strike direction was identified, I excluded this parameter from the analysis. The MSE^* values obtained for ruptures from the validation set, presented in Fig. 5.3 a and b, reveal greater variability of performance for dip slip than for strike slip ruptures, while for both types a slight increase of MSE^* with increasing M_W must be stated. Compared to the validation set, the average misfit on the test set, shown in Fig. 5.3 c and d, is clearly larger. Moreover one can identify that the misfit for strike slip ruptures is in general smaller than for dip slip ruptures. Increasing MSE^*, for $M_W < 6.8$ is clearly visible for both dip slip and strike slip ruptures. This is not surprising, because the minimum magnitude in the New Zealand ERF training set is 6.8 (Fig. 5.3 e, f), and therefore the U-Net is not trained to properly handle smaller magnitude events. Slightly increasing MSE^* for $8.0 < M_W \leq 8.5$ in the test set is also evident, although this magnitude range is well covered in the training set. Increasing misfit for dip slip faulting ruptures in the test set at dip angles $\varphi <\approx 35°$, as seen in Fig. 5.3 c, can be explained by the shortage of such angles in the training set (Fig. 5.3 g. Dip angles $\varphi >\approx 80°$ are completely lacking in the training set for dip slip ruptures, however, performance on such ruptures from the test set is not worse compared to $\approx 35° < \varphi <\approx 80°$.

To provide a more intuitive understanding for the absolute values of MSE^*, I investigate some example predictions at the 80th, 90th, and 99th %iles of MSE^* values for ruptures in

the validation set presented in Fig. 5.4. Subsequently I will focus on a similar set of example ruptures from the test set at the 50th, 80th, and 90th %iles of MSE^* values given in Fig. 5.5.

80% of U-Net predictions for validation ruptures reveal primarily a slightly smoothed out appearance compared to the target moment modifiers (Fig. 5.4 a–e), whereas the actual outlines of the amplification patterns also start to vary at the 90th %ile (Fig. 5.4 f–j). For the few cases with MSE^* beyond the 99th percentile, considerable misfit must be stated (Fig. 5.4 k–o). In particular, we notice two cases for which the U-Net predictions fundamentally differ from the targets (see red markers in Fig. 5.2 and Fig. 5.6), which are discussed in more detail in section 5.5.

The comparably large misfits for ruptures from the test set evident from Fig. 5.2 c and d clearly emerge also in the example predictions shown in Fig 5.5. Even at the 50th %ile, the predicted amplification patterns (Fig. 5.5 c and e are clearly disturbed with respect to the target ones (Fig. 5.5 b and d). This observation is surprising, since the properties of the shown strike slip rupture ($M_W = 8.1$, $\phi = 85°$) are actually well covered in the training dataset (Fig. 5.3 f and h). Systematic deviations between the predicted and target moment modifiers are evident for the example with an MSE^* at the 80th %ile (Fig. 5.5 f–j). The differences for this dip slip rupture can be associated with the rather shallow dip angle of 25°, that is not covered by the training dataset (Fig. 5.3 g). The small magnitude of $M_W = 6.1$ of the example strike slip rupture at the 90th %ile (Fig. 5.5 l–o), which is not covered by the training dataset (Fig. 5.3 f), explains the clearly observable differences between target and predicted moment modifiers for this example. Nonetheless, it should be noted that the comparatively large $MSE*$ values do not indicate complete failure on the test set. The predicted moment modifiers, also at the 90th %ile, still reveal considerable similarity to the targets regarding both the magnitude and the spatial distribution. Large $MSE*$ values therefore indicate lack of detail, rather than lacking adaption to fundamental scaling relations.

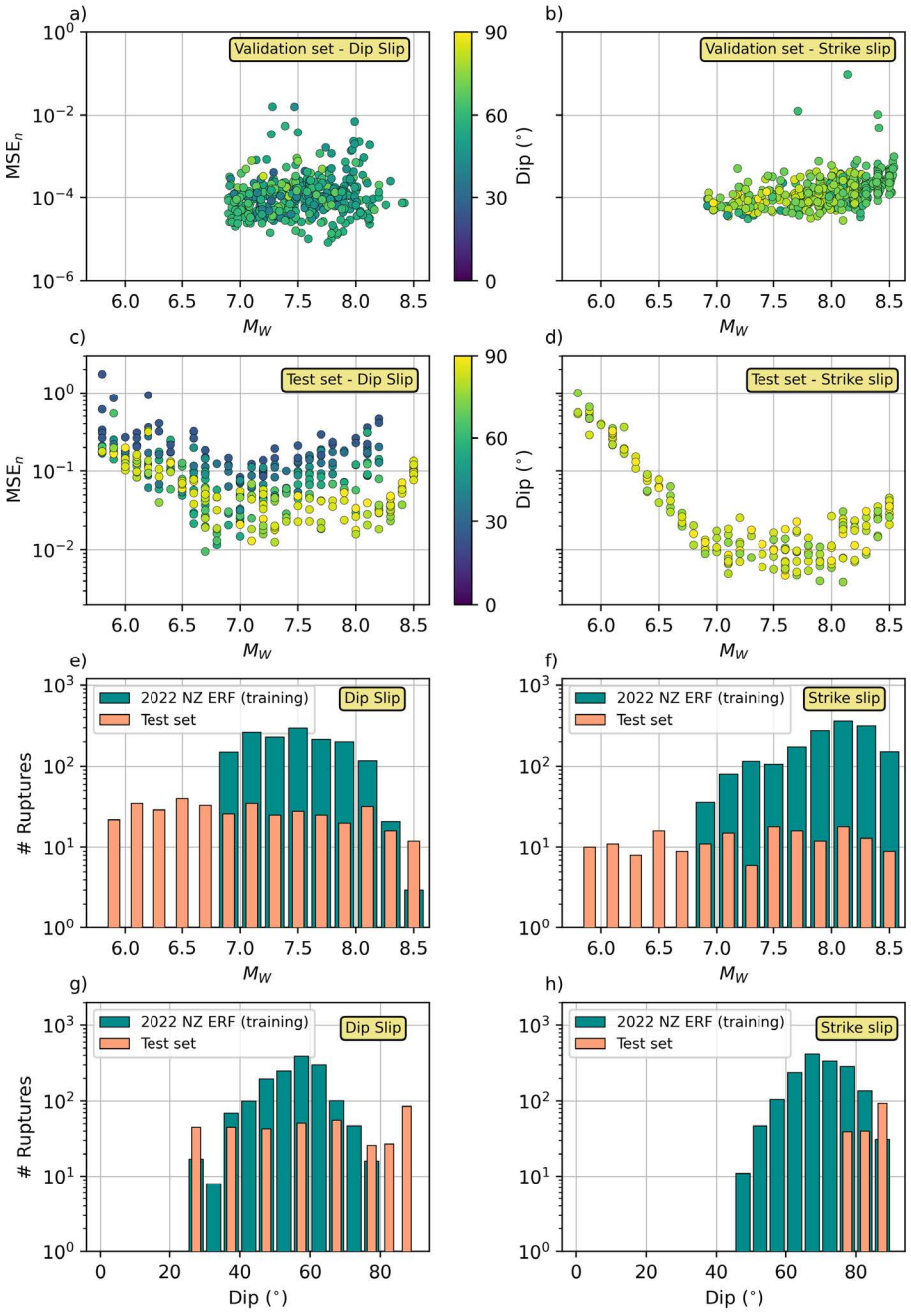

Figure 5.3: Evaluation of the U-Net performance on the ruptures from the validation and test sets. Panels a, b and c, d show the distribution of MSE^* with respect to M_W and dip angle for the validation and test sets, respectively. Panels e–h show the distributions of magnitudes and dip angles in the training and test sets. Strike directions are sampled uniformly in all directions in the test set.

5.5 Discussion

The large MSE^* values obtained for ruptures from the generic test set is a key result of this study. As explained in section 5.4, lacking coverage of predictive parameters describing the rupture properties in the training dataset partially explains this large misfit. However, some robustness against poor coverage in the training dataset can be inferred from the relatively stable performance on steep ($\varphi > 80°$) dip slip ruptures that are completely lacking in the training dataset (compare Fig. 5.3 c and g). One possible explanation for this observation could be that the U-Net inferred the general influence of steep dip angles on the spatial amplification patterns from the available strike-slip ruptures (compare Fig. 5.3 h), and successfully transferred this insight to dip slip ruptures. Nonetheless, it must be stated that also in the well covered ranges $6.8 < M_W < 8.5$ and $50° < \varphi < 90°$, substantially increased MSE^* compared to the validation set is evident. One key difference between the validation and the test set that could explain this difference is the degree of geometric complexity. While almost all ruptures from the 2022 NZ ERF involve at least multiple segments, if not multi faults, even large $M_W > 7$ events are modeled as simple planar ruptures in the generic test set. While one might assume that the U-Net should be capable of generalizing from complex to simple geometries, it seems like a simple geometry is an unknown feature that challenges the U-Net, which learned to predict irregular patterns along large ruptures. An example for this behavior is the example rupture presented in Fig. 5.5 a–e.

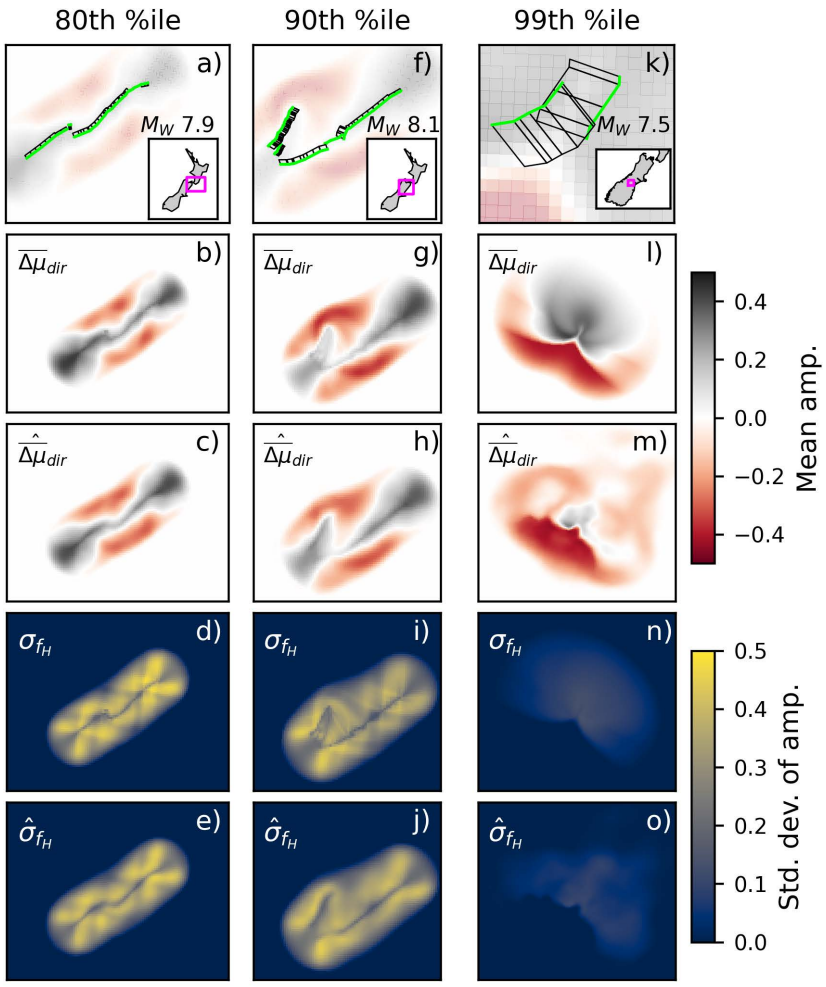

Figure 5.4: Comparison of target moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H} (rows 2, 4) with the U-Net reproductions (rows 3, 5) for example validation ruptures (row 1) with MSE^* scores at the 80th, 90th, and 99th percentile (compare green stars in Fig. 5.2). The magenta squares in the insets indicate the extent of the corresponding main panels. Green lineaments highlight the surface trace of the ruptures.

The misfit on the validation set are found to be considerably smaller than those for the test set. Moreover, the misfit below the 80th %ile of MSE^* values is caused primarily by the slightly washed out appearance of very fine scaled and sharp features such as the criss-cross pattern visible in Fig. 5.4 d. Since the there is no reasonable explanation for such sharp spatial feautures, it might be possible that they are actually artifacts caused by, e.g., the finite sample size of hypocenter locations on the rupture plane. If so, the washed out appearance might actually be a desireable property, even though it causes some misfit in the chosen loss function. Another consideration regarding the comparatively small MSE^* values obtained for validation ruptures is insufficiently strict separation of training and validation ruptures, leading to underestimation of validation misfit. For example, the overlap of fault segments between two ruptures in the 2022 New Zealand ERF can be large, where only few additional segments might distinguish a $M_W = 8.0$ from a $M_W = 8.1$ rupture, that otherwise share the majority of segments. If the random split of data samples assigns one of these ruptures to the training set and one to the validation set, the two sets may end up being too close to each other to accurately quantify a generalization misfit.

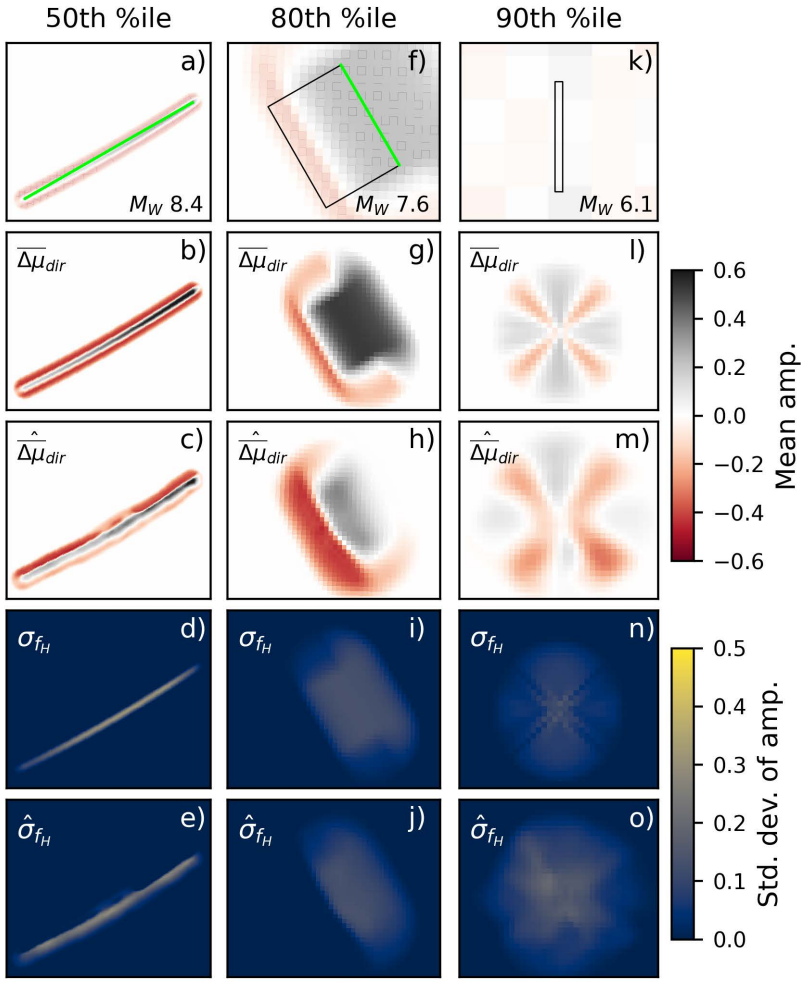

Figure 5.5: Comparison of target moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H} (rows 2, 4) with the U-Net reproductions (rows 3, 5) for example ruptures from the test set (row 1) with MSE^* scores at the 50th, 80th, and 90th %iles (compare violet stars in Fig. 5.2). Green lineaments highlight the surface trace of the ruptures.

Regardless of the reasons responsible for the differing performances on the validation and test sets, it should be noted that the determination of an acceptable misfit level of a modifier of moments model for application in PSHA is yet an unsolved task. Possibly, the rather small MSE^* values obtained for validation ruptures in this study is not necessary for practical implementation in PSHA, and perturbed predictions as those seen for the example in Fig. 5.5 a–e

might actually yield acceptable hazard results.

From the comparison of the overall model performance on the validation and test sets, I come to the conclusion that the usage of a subset of the 2022 New Zealand ERF is insufficient to train a fully generalizing U-Net-based modifier of moments model. Consequently, introduction of both simple, planar ruptures and additional complex ruptures from various ERF models or specific observed or modeled ruptures is a mandatory step prior to future research efforts. Furthermore, determination of the actual requirements of modifier of moment models for practical application in PSHA would be a worthwhile endeavor.

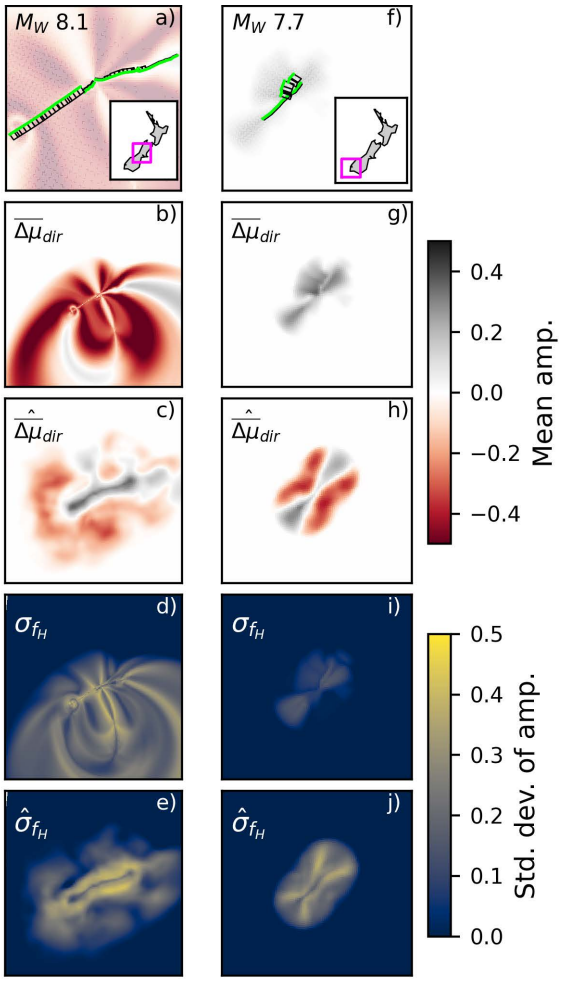

Figure 5.6: Example validation events beyond the 99th percentile of MSE scores for which the predictions deviate fundamentally from the target moment modifiers (compare red markers in Fig. 5.2).

Most of the variability in model performance on validation ruptures can be explained with the varying degree of detail that is captured for the individual ruptures, while the broad scale, overall appearance of predictions usually matches the target moment modifiers well (compare Fig. 5.4). However, there are two specific examples shown in Fig. 5.6 (red markers in Fig. 5.2) for which predictions fundamentally differ from the target moment modifiers, and that, consequently, deserve special attention. The swirly pattern of the target moment modifiers for the $M_W = 8.1$ event shown in Fig. 5.6 a–e (red x-shaped marker in Fig. 5.2) is unique among the ruptures in the validation dataset (although similar examples exist in the training dataset) and cannot be reproduced by the U-Net. Closer inspection of the predictive parameters for this example reveals that this pattern is induced by the GC2 coordinates (Fig. 5.7), where especially the $gc2_t$ coordinate reveals an untypical pattern: The zero isoline is intersecting itself and the fault outline perpendicularly where the two rupture planes are separated, which is not in agreement with the original definition of $gc2_t$ as a generalized strike-normal coordinate.

In fact, as discussed also in Spudich and Chiou (2015), this unreasonable behavior is seen in case of strike discordance among the contributing rupture planes, i.e., due to the change in strike direction by 180 °, caused by individual, almost vertically oriented segments, slightly dipping in opposite directions, which also occurs in this case. While such a change in dip direction has a significant effect on the strike direction, and therefore GC2 coordinates, any practical

impact on rupture propagation across such segments is questionable. Therefore, this property reveals an inherent weakness of GC2 coordinates for the sake of directivity modeling. Despite Spudich and Chiou (2015) warning against reversing the strike direction of individual segments to compensate this issue, due to loss of information regarding dip direction, I consider that such a practice might be an acceptable compromise, as it leads to more reasonable spatial patterns of moment modifiers.

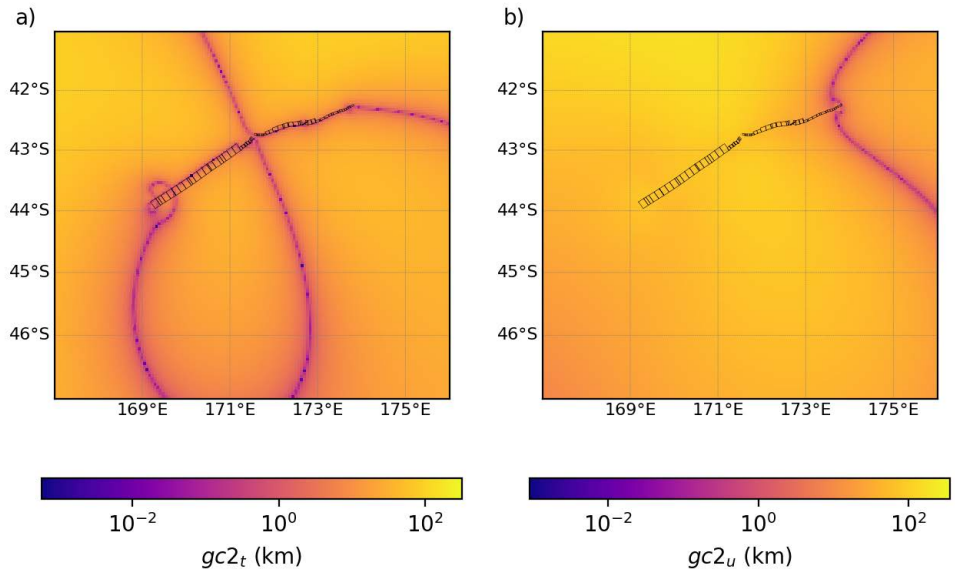

Figure 5.7: GC2 coordinates in the vicinity of the $M_W = 8.1$ rupture presented in Fig. 5.6 a–e. Spudich and Chiou (2015) identify the unreasonable pattern of the $gc2_t$ coordinate as a result of strike discordance among the two rupture planes.

It is worth noticing that, despite the usage of unreasonable GC2 coordinates as predictive parameters in this example, the actual patterns predicted by the U-Net (Fig. 5.6 c and e), although quite different in comparison to the target ones, are still reasonable in predicting general positive amplification in the near field and de-amplification in the far field. At least the near field pattern is thereby similar to past directivity modeling attempts, e.g., in the 2005 New Zealand Standard for seismic loads (NZS, 2004), where a plain near fault factor is employed in order to assume directivity related amplification at short source-to-site distances. Encouragingly, this example suggests that the U-Net model is at least to some degree robust against erroneous input parameters and/or anomolous/esoteric ruptures for which current directivity models are poorly constrained, insofar as it does not provide randomly disturbed moment modifiers in such cases, but maintains at least some degree of reasonability. This robustness can be considered an advantage over equation-based modifier of moments models as suggested by e.g. Watson-Lamprey (2018), as the strict relations defined via an equation would directly transfer erroneous input patterns into the predicted moment modifiers. Nevertheless, this example demonstrates that the degree of geometric complexity that the chosen input parameters can handle has certain limitations, and that care must be taken if exceptionally complex geometries are encountered in application.

The second conspicuous example from the validation dataset is depicted in Fig. 5.6 f–j (compare red triangle in Fig. 5.2). In this case the fundamental differences between the target moment modifiers and the U-Net prediction are caused by the varying styles of faulting across the segments of this multi-fault rupture. In the ERF, 2 of the 4 segments are classified as strike-slip (rake=180°), whereas the other two are considered dip slip (rake=110°). Since the majority of the moment is caused by the strike-slip components, I decided to set the style of faulting indicator $\mathbb{I}_{SS} = 1$, neglecting the dip slip component. Consequently, the U-Net predictions reveal a pure strike-slip pattern that does not actually fit the target. While a redesign of the U-Net input layer to account for varying styles of faulting within one rupture is theoretically possible, the associated effort seems for now disproportionate, considering that multi-fault ruptures in general and multi style-of-faulting ones in particular are rarely encountered.

For the remaining examples from the validation dataset with MSE scores beyond the 99th percentile (yellow stars in Fig. 5.2), the U-Net predictions do not differ fundamentally from the targets, but show considerable perturbations comparable to the example shown in Fig. 5.4 k)-o). Taking a closer look reveals that all affected ruptures share the geometrical property of intersecting fault segments and strike discordance due to change in dip direction from segment to segment by about 180°. Similar to the discussed example shown in Fig. 5.6 a–e, the validity of GC2 coordinates in case of strike discordance is questionable, and strike reversal for individual segments might be a solution. It should be noted that for some examples, e.g. Fig. 5.4 f–j strike discordance does not lead to artificial patterns like those in Fig. 5.6 b and d, thus, further research on the influence of rupture geometry on the moment modifiers is necessary.

In this study I utilized the mean squared error as the loss function during ANN training and to rank the performance on individual examples from the validation dataset. In doing so I generate an non-interpretable misfit mixture over the two moment modifiers $\overline{\Delta\mu}_{dir}$ and σ_{f_H}, neglecting the fact that these two might actually contribute differently due to the different value range that they cover. While the results indicate that predictions for both moment modifiers seem to closely resemble the target ones, it might be worth to adapt a more scale-invariant metric such as the z-score in the future.

5.6 Conclusions

In summary one can state that the U-Net-based modifier of moments model provides satisfying estimates $\hat{y}_i = (\hat{\overline{\Delta\mu}}_{dir,i}, \hat{\sigma}_{f_H,i})$ for about 90% of the ruptures in the validation subset of the 2022 New Zealand ERF. The remaining 8 ruptures show exceptional geometries or source properties that lead to a questionable representation via the chosen set of predictive parameters that causes substantial differences between the U-Net predictions and the (also debatable) target moment modifiers. Particularly conspicuous are the spatial patterns of the GC2 coordinates in case of strike discordance among individual rupture segments, however, further research must be undertaken to uniquely identify these as the source of the deviations and to investigate the impact of strike reversals as suggested carefully also in Spudich and Chiou (2015). Regardless of the cause of the deviations, one can state that the magnitude of amplification and the coarse scaled spatial patterns predicted for the affected ruptures still represent reasonable estimates of

the moment modifiers that are not expected to disturb PSHA calculations considerably. This holds to some degree also true for the generic ruptures from the test set, for which the broad scale spatial amplification patterns are in rough agreement with the target moment modifiers, but for which the reproduction of spatial details seems challenging. These insights suggest that the training dataset should be complemented with both generic, simple ruptures as well as complex observed and/or modeled ruptures before further experiments are conducted. Furthermore, the results from this study suggest that the random split of ruptures from the New Zealand 2022 ERF into a training and a validation set might be insufficient, and that manual revision might be worthwhile. From the preliminary results obtained in this study I finally conclude that with some extensions of the presented methodology and underlying datasets, U-Net-based modifier of moment models are promising candidates to enable the explicit incorporation of directivity effects in PSHA calculations, similar to the approach presented in chapter 4 in the near future.

5.7 Outlook

The results presented in this study are preliminary and many open questions and tasks are yet to be addressed. A first straightforward extension of the methodology is the incorporation of multiple spectral periods, which will slightly increase the size of the resulting U-Net model on disk and the demand in computational resources during training of the ANN, but is not expected to cause any further complications. In addition, the extension of the dataset with additional ruptures seems mandatory to ensure the U-Net's ability to generalize to a wide range of rupture geometries, including also simple planar ruptures that are not included in the 2022 New Zealand ERF. The calculation of GC2 coordinates in case of strike discordance among multiple rupture segments should be revised, and the effect of strike reversals on the resulting target moment modifiers should be investigated in detail. This step may potentially yield more reasonable target moment modifiers, and also fewer outliers regarding the performance of the U-Net on the validation dataset. It should be noted that, from a technical perspective, it would also be within range to utilize a training dataset generated via physics-based waveform simulations instead of empirical directivity models such as the one by Bayless et al. (2020) utilized in this study. Given the fact that actual directivity affected strong motion observations are rare and spatially sparse, a physics-based approach might yield higher resolution and, thus, more realistic representations of directivity related amplification fields. While the performance on the randomly determined validation split of the available dataset gives a good first impression of the model performance, a more sophisticated benchmark framework that systematically challenges the U-Net's ability to generalize to specific types of ruptures might be a desireable step before actual implementation into existing PSHA frameworks. Finally, also an actual exemplary implementation in a PSHA calculation as presented in the previous chapter for the ANN_{ovf} is absolutely crucial to verify that the model's performance is acceptable for application in Probabilistic Seismic Hazard Analysis.

5.A Appendix

5.A.1 U-Net specifications

This sections provides a detailed description of the U-Net neural network (Ronneberger et al., 2015) architecture employed in this study. Additional technical details supporting the following explanations are given in Table 5.1.

The U-Net architecture (Fig. 5.1) consists of three major components: The encoder branch, the decoder branch, and skip connections.

Within the encoder branch, the input is alternately processed through conv blocks and pooling layers. One conv block comprises two 2D convolution layers followed by a non-linear LeakyReLu activation function. The 2D convolutional layers consist of filterbanks that are convolved with the respective input to the layer. The values in the filter masks pose the majority of coefficients in the U-Net and are subject to optimization during the training phase. To avoid boundary value loss from the convolutional operations, symmetric padding of the input maps is conducted before every convolution. Subsequent passing through the non-linear activation function is necessary to enable the U-Net to learn non-linear relations. Pooling layers reduce the size of the data maps, via selecting the largest value of each 2x2 sub-matrix and discarding the remaining values. The interaction between convolution and pooling layers leads to a highly abstract information representation of shape 16x16x128 in the bottleneck of the U-Net. The encoder branch is considered a feature extractor.

The decoder branch consists of alternating up-conv operations and conv blocks. An up-conv operation comprises an up-sampling step, that increases the shape of the data maps by a factor of 2 via bilinear interpolation, and a 2D convolutional layer.

While the input parameters are given as spatial maps of well defined quantities, the information representation in the bottleneck of the U-Net is highly abstract, and a large portion of the spatial context is lost. Therefore, skip connections are employed to concatenate layers in the decoder branch with their pendants from the encoder branch in order to reintegrate this spatial information.

The final conv-out operation is a 2D convolutional layer followed by a linear activation function, generating an output $\hat{y}$ of shape 256x256x2.

Table 5.1: Technical specification of the U-Net architecture utilized in this study and depicted in Fig. 5.1. A conv block as denoted in Fig. 5.1 comprises 2 Conv2D layers (compare the "Comment" column). An up-conv operation as denoted in Fig. 5.1 comprises upsampling and a single Conv2D layer. "cc" stands for concatenation.

ID	Layer type	# Filters	Filter size	Input shape	Output shape	Activation	Comment
L1	Input			256x256x9			9 predictive parameters
L2	Conv2D	8	(3x3)	256x256x9	256x256x8	Leaky ReLU	(2x)
L3	Max Pooling			256x256x8	128x128x8		
L4	Conv2D	16	(3x3)	128x128x8	128x128x16	Leaky ReLU	(2x)
L5	Max Pooling			128x128x16	64x64x16		
L6	Conv2D	32	(3x3)	64x64x16	64x64x32	Leaky ReLU	(2x)
L7	Max Pooling			64x64x32	32x32x32		
L8	Conv2D	64	(3x3)	32x32x32	32x32x64	Leaky ReLU	(2x)
L9	Max Pooling			32x32x64	16x16x64		
L10	Conv2D	128	(3x3)	16x16x64	16x16x128	Leaky ReLU	(2x)
L11	Up-Sampling2D			16x16x128	32x32x128		
L12	Conv2D	64	(3x3)	32x32x128	32x32x64	Leaky ReLU	(1x)
L13	Concatenation			32x32x64	32x32x128		cc with L8
L14	Conv2D	64	(3x3)	32x32x128	32x32x64	Leaky ReLU	(2x)
L15	Up-Sampling2D			32x32x64	64x64x64		
L16	Conv2D	32	(3x3)	64x64x64	64x64x32	Leaky ReLU	(1x)
L17	Concatenation			64x64x32	64x64x64		cc with L6
L18	Conv2D	32	(3x3)	64x64x64	64x64x32	Leaky ReLU	(2x)
L19	Up-Sampling2D			64x64x32	128x128x32		
L20	Conv2D	16	(3x3)	128x128x32	128x128x16	Leaky ReLU	(1x)
L21	Concatenation			128x128x16	128x128x32		cc with L4
L22	Conv2D	16	(3x3)	128x128x32	128x128x16	Leaky ReLU	(2x)
L23	Up-Sampling2D			128x128x16	256x256x16		
L24	Conv2D	8	(3x3)	256x256x8	256x256x8	Leaky ReLU	(1x)
L25	Concatenation			256x256x8	256x256x16		cc with L2
L26	Conv2D	8	(3x3)	256x256x16	256x256x8	Leaky ReLU	(2x)
L27	Conv2D	2	(1x1)	256x256x8	256x256x2	linear	(1x)
L28	Output				256x256x2		Maps of moment modifiers at T=10 s

Table 5.2: Training configuration adopted in this study. Training is conducted for 1000 epochs, i.e., training cycles during which all training samples are used once. Coefficients are updated based on the misfit obtained from 8 training samples (mini batches of size 8). MSE - mean squared error. Remaining parameters refer to the configuration of the Adam optimizer (Kingma and Ba, 2015).

# Epochs	Batch size	Loss function	Learning rate	Optimizer	β_1	β_2	ϵ
1000	8	MSE	1.e-04	Adam	0.9	0.999	1.e-07

Chapter 6

Utilization of Crowdsourced Felt Reports to Distinguish High-Impact from Low-Impact Earthquakes Globally within Minutes of an Event

originally published as:

Lilienkamp, H., R. Bossu, F. Cotton, F. Finazzi, M. Landès, G. Weatherill, and S. von Specht (2023). "Utilization of Crowdsourced Felt Reports to Distinguish High-Impact from Low-Impact Earthquakes Globally within Minutes of an Event". The Seismic Record. 3(1), 29–36, doi: 10.1785/0320220039.

The supplemental material to this article is available online at:

https://gsw.silverchair-cdn.com/gsw/Content_public/Journal/tsr/3/1/10.1785_0320220039/2/tsr-20220039_supplement.pdf?Expires=1695633542&Signature=Rjy8FWsniRq81I2MrWsgPVKApTZMyjyGUdNblJmf~J5w~30MOZ7EazaVTWdYtB4MCuhrHRFl2~bozO2pMUNQ8OAvzTMrdNX9jER0U6AKyG8PZvwYp5o5SJ~AWPS7eOHF3IKzAOFWLwVW4poVWADIsw9hKM6CgEJDpBjPIDW-vEzvS82LwSCKedr0rjRjnCbj-UclE-MetBrFToESub3Xn3ns3E2aYcQo7KxeOPYjReEGuvG7mffG6mWQjbMoJkq99SrAi2rI6GTReCY9cZztLlAv8fqBukyuwQiZc~pb4iJJLo3svQRRsE2Rm0fWhFPEu5DoTbB4P~XmQwbU1P823Q__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA

6.1 Abstract

Rapid assessment of an earthquake's impact on the affected society is a crucial step in the early phase of disaster management, navigating the need for further emergency response measures. We demonstrate that felt reports collected via the LastQuake service of the European Mediterranean Seismological Center can be utilized to rapidly estimate the probability of a felt earthquake being high impact rather than low impact on a global scale. Our data-driven, transparent, and reproducible method utilizing Bayes' theorem and kernel density estimation provides results within 10 min for 393 felt events in 2021. Although a separation of high- and low-impact events remains challenging, the correct and unambiguous assessment of a large portion of low-impact events is a key strength of our approach. We consider our method as an inexpensive addition to the pool of earthquake impact assessment tools, one that is fully independent of seismic data and can be utilized in many populated areas on the planet. Although practical deployment of our method remains an open task, we demonstrate the potential to improve disaster management in regions that currently lack expensive seismic instrumentation.

6.2 Introduction

An urgent question that decision makers and emergency response operatives are facing in the immediate aftermath of felt earthquakes is whether considerable impact on the affected population is to be expected or not. Although a sophisticated answer to this question is crucial to successful long- term disaster management, a preliminary characterization of the situation based on rapidly available, though crude, information in the very first minutes after an earthquake is equally important, because it can determine whether emergency response measures will be initiated or relinquished in the first place.

Rapid impact assessment systems such as PAGER (Jaiswal et al., 2010; Wald et al., 2010) are based on the ShakeMap methodology (Wald et al., 2005) and typically provide the first quantitatively reliable estimate of expected impact such as financial losses, destroyed and damaged buildings, and number of casualties. The ShakeMap methodology requires a description of the earthquake source, ground acceleration data from a dense strong-motion network, and/or macroseismic intensity observations collected, for example, via the "Did You Feel It?" service of the U.S. Geological Survey (Wald et al., 1999). Consequently, it takes on average 30 min until a first impact assessment is available after an earthquake (Wald et al., 2010), although in rare cases this number can be as low as 5 min if a dense, regional real-time strong-motion network is operated (e.g. Poggi et al., 2021).

Complementary to this indispensable framework, we suggest an approach in which we circumvent the intermediate step of a ShakeMap and classify an earthquake as either high impact or low impact solely based on the reported level of shaking inferred from felt reports collected globally by the European Mediterranean Seismological Centre's service LastQuake (Bossu et al., 2017, 2018b). Because LastQuake felt reports are collected numerously and fast via websites, and a smartphone application (50% of felt reports are collected within 10 min after an event), our method is independent of seismic instrumentation and can provide a rapid, preliminary characterization of the situation before more sophisticated and quantitatively reliable estimates

from ShakeMap-based approaches are available. Furthermore, results of our model may still serve as an independent control mechanism, in which large deviations between our results and later approaches could indicate under- or overestimation by either of the methods. Although independence of seismic data makes our method applicable in poorly instrumented regions, a technology-affine and participating population is a limiting necessity.

We use a dataset comprising over 1.5 million globally collected felt reports from over 10,000 earthquakes of any magnitudes between 2014 and 2021. Our easily comprehensible method utilizes Bayes' theorem and kernel density estimators (KDEs), and is therefore fully data driven, transparent, and reproducible. We discuss our model using well-reported validation earthquakes from 2022, that is, earthquakes that occured in 2022 and have not been used to calibrate the model and therefore allow an estimation of the model performance on future earthquakes.

6.3 Data and processing

Over 1.5 million felt reports collected globally from over 10,000 felt earthquakes (i.e., earthquakes with at least 10 felt reports) between 2014 and 2021 form our data foundation (Bossu et al., 2023). One felt report comprises a pseudointensity value that quantifies the level of shaking, the timing, and the report location. We use the term pseudointensity, because the reported value is inferred from mapping a single macroseismic observation to the EMS-98 macroseismic scale (Grünthal, 1998), whereas a true, quantitatively reliable macroseismic intensity would be obtained from averaging multiple observations across a region. Reliable information can only be obtained from felt reports when collected in large numbers and if pseudointensity values are averaged spatially. Because the magnitude of an event might be unknown by the time when first felt reports are available, and because we focus on the impact rather than the physics of earthquakes, we include earthquakes of any magnitude. An example collection of felt reports for the M_W 5.7 event in Bosnia and Herzegovina on 22 April 2022, comprising 14,000 felt reports, the first 50 of which were collected within 95 s, demonstrates the capability of the LastQuake collection procedure to meet the preceding requirements (Fig. 6.1).

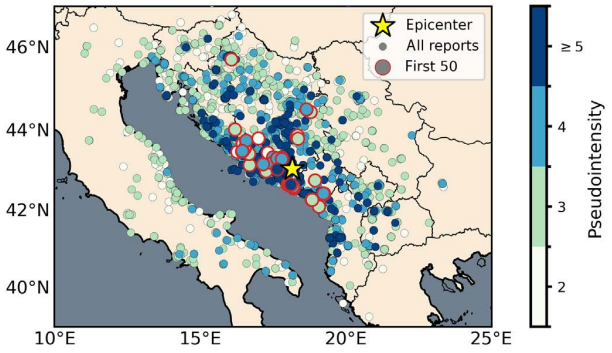

Figure 6.1: Example collection of felt reports from the M_W 5.7 event on 22 April 2022 in Bosnia and Herzegovina. Large pseudointensities overlay small ones. The first 50 reports overlay later ones.

The key goal of this study is to develop a probabilistic model that classifies an earthquake as either high impact or low impact based on felt reports rapidly after an event. We briefly outline here the dataset preparation, and additional details are found in the supplemental material S1. In a first step, we transform felt reports into representative features such as the average pseudointensity $\bar{I}$ or the average distance $\bar{R}$ of reporting locations to the barycenter (their geometric centroid). A list of all considered features is provided in Table S1, available in the

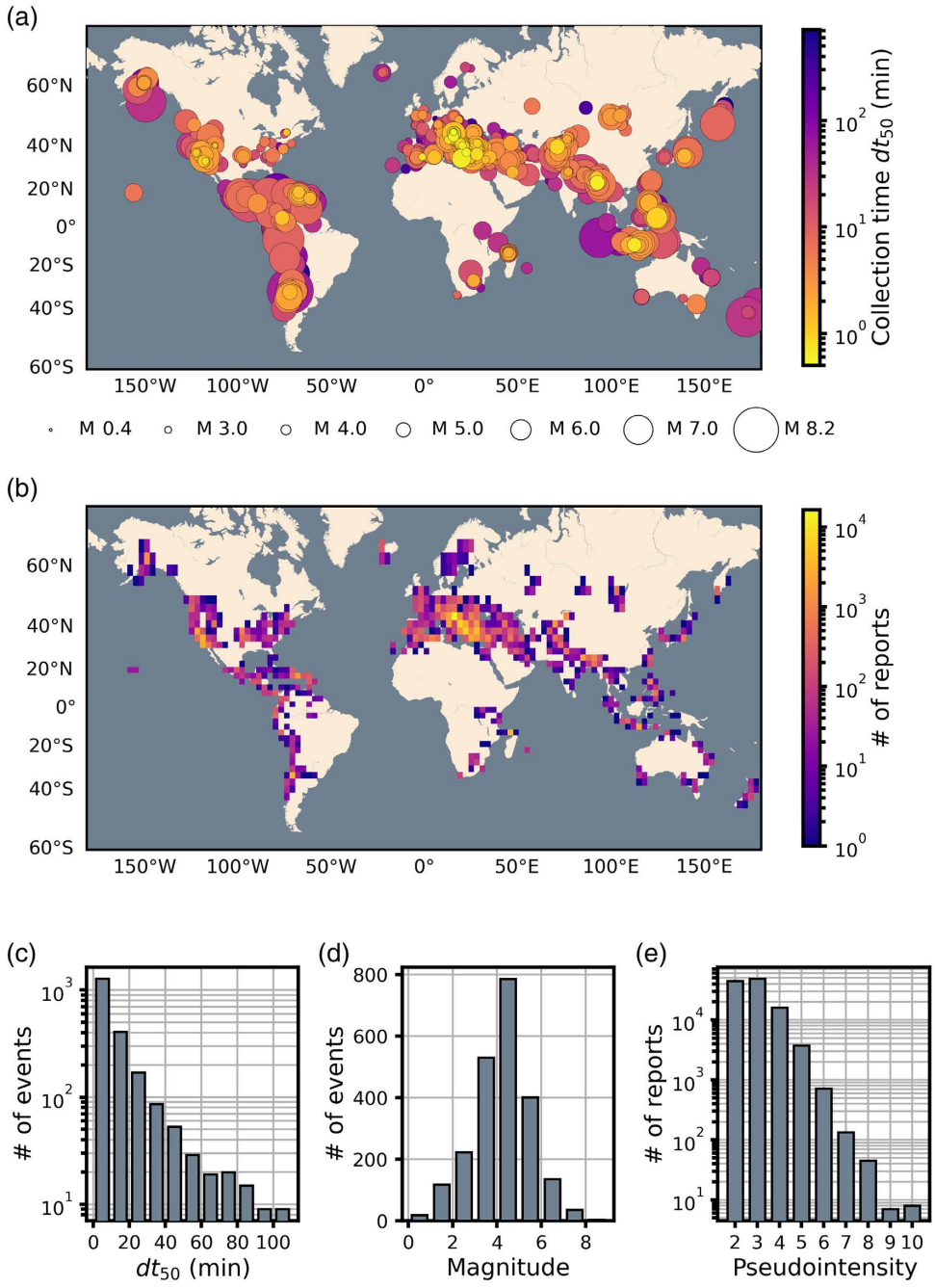

Figure 6.2: Geographical distribution of (a) events and (b) felt reports in the processed LastQuake database. Small dt_{50} overlay large dt_{50} in panel (a). The distributions of dt_{50}, magnitude (mixed types), and reported pseudointensities are given in panels (c), (d), and (e), respectively. The histogram in panel (c) comprises 94 % of events; for the sake of clarity, we refrained from visualizing the remaining 6 % of events with $120 \leq dt_{50} \leq 720$ min.

supplemental material to this article. To assure rapid availability and accuracy of the predictive features, we derive them from the first 50 felt reports that are available for an earthquake and remove the event from the database if it has fewer reports. In a second step, we label earthquakes as high impact or low impact based on impact measures that are documented in the NCEI/WDS Global Significant Earthquake Database (GSED), the Emergency Events Database (EM-DAT), and the Earthquake Impact Database (EID, see Data and resources). We define an

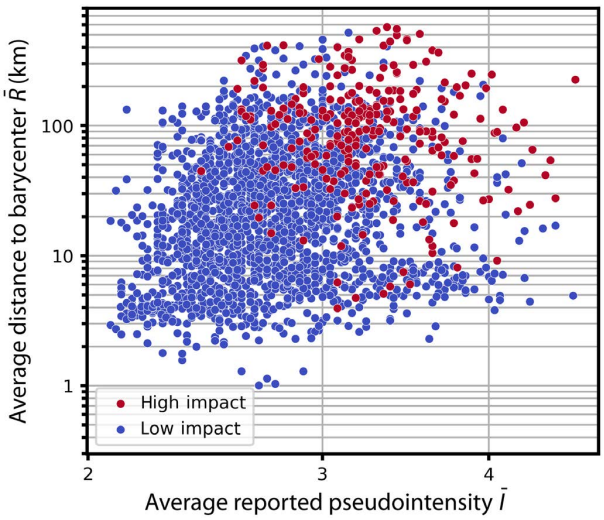

Figure 6.3: Distribution of high-impact (red) and low-impact (blue) events with respect to $\bar{I}$ and $\bar{R}$.

event as high impact if it caused at least one of the following impacts:

- at least 1 destroyed building;
- at least 50 damaged buildings;
- at least 2 fatalities;
- any documented financial losses.

These rather small threshold values reflect our intention to distinguish the majority of felt earthquakes that have no impact on the society from the minority of events that do. Although focusing on the most devastating events might be an equally interesting strategy, the necessity for a considerable amount of data examples in both the classes to obtain stable modeling results led us to choose the adopted scheme. By choosing a threshold value of two fatalities instead of one, we avoid classifying a considerable amount of earthquakes as high impact, in which single people died due to incidents that happened during, but cannot directly be related to an earthquake (e.g. Nievas et al., 2020; Shoaf et al., 1998).

The final database comprising 254 high-impact and 1994 low-impact events is summarized in Fig. 6.2. The geographic distribution of events and felt reports in Fig. 6.2 a, b reveals the global utilization of the LastQuake service, albeit with a substantial bias toward Europe, in which ∼75 % of felt reports are collected. The distribution of durations to collect 50 felt reports dt_{50} (Fig. 6.2 c) shows that for over 1000 events the required data are collected within 10 min, emphasizing the efficiency of felt report collection via LastQuake. The distribution of events according to the features $\bar{I}$ and $\bar{R}$ is depicted in Fig. 6.3. Despite the overlap of the two types of events, visual inspection indicates differing underlying distributions for high- and low- impact events. We identify the trend that larger impact is expected for strong shaking felt over large areas. This reason- able finding is in agreement with the conclusions of, for example, Atkinson and Wald (2007).

6.4 Probabilistic classification of earthquakes

We derive a probabilistic model providing the probability p of an earthquake being high impact (H), rather than low impact (L), given its features X derived from felt reports. In the following, we consider the case in which X comprises two predictive parameters $X_{2D} = (\ln \bar{I}, \ln \bar{R})$, in which ln denotes the natural logarithm.

The desired posterior probability $p(H|X_{2D})$ is calculated via Bayes' theorem:

$$p(H|X_{2D}) = \frac{f(X_{2D}|H)p(H)}{f(X_{2D})}, \tag{6.1}$$

in which $f(X_{2D}|H)$ and $f(X_{2D})$ denote the densities of the likelihood and the marginal, respectively. We estimate the prior $p(H)$ probability of occurrence of a high-impact event from the numbers N_H and N_L of high-impact and low-impact events in our database, respectively:

$$p(H) = \frac{N_H}{N_H + N_L}. \tag{6.2}$$

We infer the likelihood $f(X_{2D}|H)$ (and also$f(X_{2D}|L)$) from data. Visual inspection of Fig. 6.3 suggests to model $f(X_{2D}|H)$ and $f(X_{2D}|L)$ as bivariate Gaussians; however, Kolmogorov-Smirnoff tests indicate that normality cannot be assumed for most predictive features (see Fig. S4 in the supplemental material). We therefore choose kernel density estimation (KDE) with Gaussian kernels to estimate $f(X_{2D}|H)$ and $f(X_{2D}|L)$. When fitting a KDE to data, optimal choice of the kernel bandwidth h is crucial. To ensure smoothness of the resulting density functions, we first generate the density function of a bivariate Gaussian $\hat{m}_{gauss}$ from the mean and covariance of the data. We then select the bandwidth h_g that leads to a density function generated from the KDE $\hat{m}_{h_g}$ that is most similar to that of the Gaussian in terms of mean-squred error:

$$h_g = \min_h \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} [\hat{m}_h(x_{ij}) - \hat{m}_{gauss}(x_{ij})]^2. \tag{6.3}$$

$N_x = N_y = 50$ denotes the horizontal and vertical dimensions of the regular grid x_{ij} on which the densities are compared (the grid spans exactly the value range of the predictive parameters). This way we ensure a smooth shape of the estimated density and simultaneously account for the non-Gaussian properties of the data.

The marginal $f(X_{2D})$ is modeled as a mixture of $f(X_{2D}|H)$ and $f(X_{2D}|L)$:

$$f(X_{2D}) = (f(X_{2D}|H) \cdot N_H + f(X_{2D}|L) \cdot N_L) \cdot C, \tag{6.4}$$

in which C is a normalization constant:

$$C = \frac{1}{N_H + N_L}. \tag{6.5}$$

The posterior $p(H|X_{2D})$ is presented in Fig. 6.3. We additionally calculated alternative solutions for $p(H|X_{2D})$, in which we modeled the likelihoods in equation (6.1) as (1) bivariate gaussians and (2) with an alternative KDE approach in which we optimize the kernel bandwidth in a leave-one-out cross-validation procedure to maximize the likelihood of the data. Visualizations of resulting likelihoods, posteriors, and uncertainty estimates derived from bootstrapping with 5,000 draws are given in Fig. S5 in the supplemental material.

6.5 Results

Confirming the previous interpretations, the posterior presented in Fig. 6.4 suggests that the stronger the shaking and the larger the area over which the shaking is felt, the more likely it is for an earthquake to be of high impact. The posterior $p(H|X_{2D})$ provides fairly low values

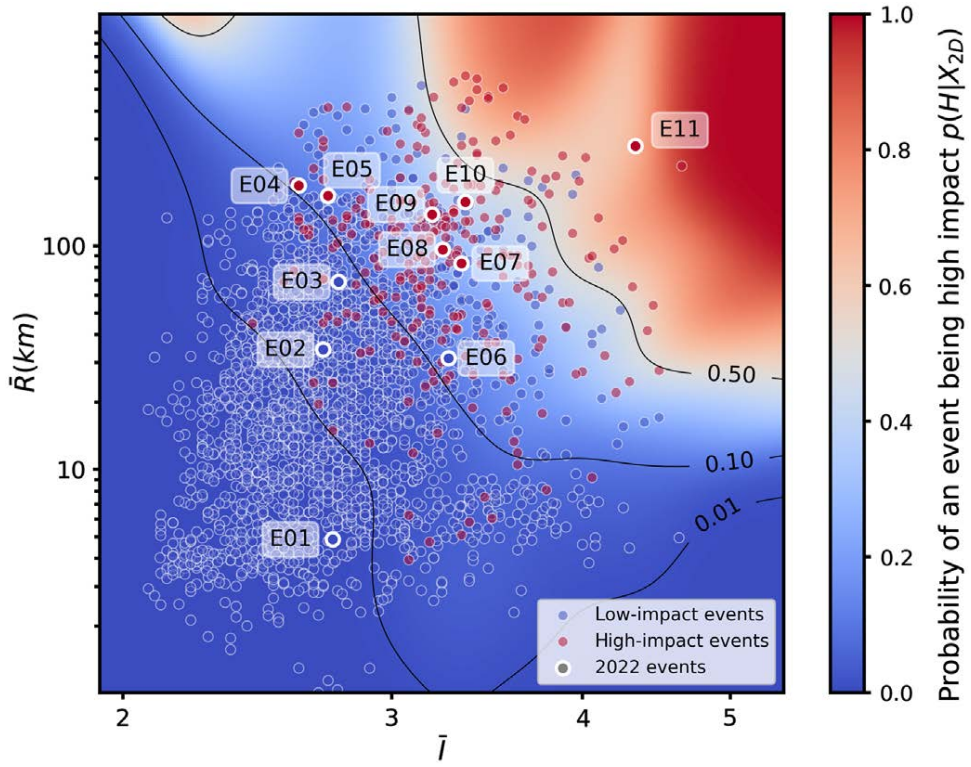

Figure 6.4: Visualization of the posterior probability $p(H|X_{2D})$ of an earthquake being high impact, given the average reported pseudointensity $\bar{I}$ and the average barycentral distance $\bar{R}$ obtained from its collection of felt reports. Transparent markers represent events between 2014 and 2021 that were used to calibrate the model, in which the classification into high- and low-impact events is based on the impact databases from GSED, EM-DAT, and EID (see section 6.8). Opaque markers represent validation events from 2022 in which the IDs correspond to those in Table 6.1.

where the density of high-impact events is high and only exceeds the value of 0.5 occasionally. This is caused by the small value of the prior $p(H)$ (equation 6.2), that is, as the vast majority of ∼92 % of felt earthquakes are actually low impact (see the supplemental material S1). Although the orientations of isolines of $p(H|X_{2D})$ in Fig. 6.4 seem reasonable where data density is large, considerable influence of individual data samples is obvious in regions where data density is low, indicating slight overfitting of the KDEs to the data.

For validation purposes, we applied our model to a selection of well-recorded validation earthquakes from 2022 (see Fig. 6.4; Table 6.1). We notice that high-impact events are generally assigned higher values of $p(H|X_{2D})$ compared with low-impact events. Distinct and accurate classification is seen for the low-impact events from Croatia (E01) and California (E02). The intermediate values of $0.07 \leq p(H|X_{2D}) \leq 0.50$ assigned to events E03–E10 indicate a nonnegligible chance of impact that would be difficult to interpret in a real-time operation of the system. Also the comparatively large value of $p(H|X_{2D}) = 0.68$ assigned to the event from Japan (E11) is ambiguously interpretable. We notice that the high-impact events from Sumatra (E04) and Nepal (E05) are assigned relatively small values compared to their impact, whereas the event from Chile (E06) with a slightly larger value is actually of low impact.

6.6 Discussion

6.6.1 Modeling decisions

We conducted a fully data-driven modeling approach via the use of kernel density estimators. As expected, the resulting posterior probability $p(H|X_{2D})$ is poorly constrained where data density is low (e.g., wiggles in isolines in Fig. 6.4), which is reflected by increased uncertainty estimates in these regions (compare Fig. S5 in the supplemental material). Fig. S5 shows that even though normality of predictive features is formally not given, assuming bivariate gaussian distributions still leads to a useful and even more reasonable (straight isolines) model that might be more applicable in a practical implementation of our model.

Table 6.1: Validation Events from 2022, Impact Data from GSED and Earthquake Impact Database (EID). Abbreviations L and H in the column "True Class" refer to low impact and high impact, respectively.
* – EID, Earthquake Impact Database. † – GSED, Global Significant Earthquake Database.

ID	Date (yyyy/mm/dd)	Region	Magnitude	dt_{50} (s)	Number of Reports	Impact	True Class	$P(H\|X_{2D})$
E01	2022/06/18	Croatia	M_L 2.7	37	5,358	–	L	< 0.01
E02	2022/01/30	California	M_W 4.1	96	570	1 damaged*	L	0.02 ± 0.01
E03	2022/09/10	France/ Germany	M_L 4.7	109	1,845	22 damaged*	L	0.07 ± 0.03
E04	2022/02/25	Sumatra	M_W 6.2	424	1,267	11 dead, 103 destroyed†	H	0.11 ± 0.03
E05	2022/07/31	Nepal	m_b 5.1	595	123	3 destroyed, 475 damaged*	H	0.12 ± 0.04
E06	2022/08/18	Chile	m_b 5.5	154	334	–	L	0.13 ± 0.01
E07	2022/02/13	Armenia	M_L 5.4	183	1,333	80 damaged†	H	0.28 ± 0.04
E08	2022/04/22	Bosnia and Herzegovina	M_W 5.7	96	14,087	1 dead, 350 damaged†	H	0.28 ± 0.05
E09	2022/10/05	Iran	M_W 5.6	463	292	∼75 destroyed†	H	0.33 ± 0.07
E10	2022/06/21	Afghanistan	M_W 5.9	492	154	1,150 dead, 3,000 destroyed†	H	0.41 ± 0.08
E11	2022/03/16	Japan	M_W 7.3	741	133	4 dead†	H	0.68 ± 0.14

6.6.2 Interpretation of the posterior

The most interesting property of our model is that 39 % of low-impact events in the calibration dataset and not a single high-impact event fall in the region in which $p(H|X_{2D}) < 0.01$ (Fig. 6.4). Consequently, considerable impact can almost certainly be ruled out for future earthquakes with similar appearance. Bearing in mind that these events are still largely felt and may cause considerable public anxiety (e.g. Becker et al., 2019; Casey et al., 2018), the ability to comfort

the affected population in these cases is a key strength of our methodology.

For few earthquakes with $0.5 < p(H|X_{2D}) \leq 0.93$, the analysis suggests that the occurrence of impact is more likely than its absence, although uncertainties are still large in most cases.

The overlap of high- and low-impact events, and the subsequent small-to-medium values of the posterior $p(H|X_{2D})$ (Fig. 6.4) raise the question of how to utilize modeling results for $0.01 < p(H|X_{2D}) < 0.5$. One possible solution would be to introduce a traffic light system that suggests a decision maker to not take any further action at low $p(H|X_{2D})$ (green), to suggest further investigations at intermediate levels of $p(H|X_{2D})$ (yellow), or to raise an alert at large $p(H|X_{2D})$ (red). Because the exact thresholds that define the boundaries between "green", "yellow", and "red" events would largely depend on the intended use case, we will not suggest any particular values.

6.6.3 Performance and applicability

Because of the exclusive selection of events with at least 50 reports, our suggested methodology is applicable to ∼22 % of 2746 felt earthquakes in 2021, in which for ∼14 % (393 events) a result can be obtained within 1 min. We expect these numbers to increase over time according to the increasing usage of the LastQuake service (the number of reported earthquakes with $4 \leq M \leq 5$ increased on average by 23 % per year from 155 in 2014 to 646 in 2021). Because every smartphone user is a potential contributor of felt reports, we are still far from what could possibly be achieved once the value of dense and inexpensive felt report collection is properly acknowledged and encouraged by governments and emergency response operatives. Admittedly, our model will be of minor impact in regions where dense real-time strong-motion networks and automatized impact assessment are already in place, as is the case in the Friuli Venezia Giulia region in Italy, for example (Poggi et al., 2021). However, the M_W 5.9 event in Afghanistan on 22 June 2022 (E10) in Table 6.1 is a striking example of the potential impact that our model could have in remote regions that lack seismic instrumentation. The required amount of 50 reports was in this case collected within about 8 min, and even though the corresponding $p(H|X_{2D}) = 0.41$ does not unambiguously hint at the extreme impact of this event, at least the considerable probability of impact could have been noticed rapidly after the earthquake. In such regions, promoting the low-cost usage of LastQuake might be a worthwile option as long as the installation of strong-motion instruments is infeasible.

6.6.4 Geographic and operational prerequisites

The validation events from Sumatra and Nepal (E04 and E05 in Table 6.1) are assigned relatively small values of $p(H|X_{2D})$,despite their considerable impact. In the first case, only few felt reports were issued from Sumatra, whereas the majority were submitted from Kuala Lumpur, Malaysia, some 400 km away, causing distorted distributions of pseudointensities and barycentral distances. In Nepal, our approach suffers from the inaccessibility of LastQuake in China (Bossu et al., 2018b), causing a lack of reports beyond the Chinese border. These two cases emphasize the necessity of active participation of LastQuake users, on the one hand, and the transnational collection of felt reports, on the other hand. Furthermore, the Nepal case indicates

limited applicability to coastal regions, where the azimuthal distribution of felt reports is likewise highly nonuniform. The validation events in Afghanistan (E10) and at the French–German border (E03) are two counterexamples where reports were successfully derived across borders between Afghanistan and Iran, and Germany and France, respectively. Subsequently, the derived $p(H|X_{2D})$ is equally valid for all affected countries.

6.6.5 Outlook

For the sake of interpretability, we utilize only two predictive parameters to describe an earthquake. However, the simple formulation of the modeling task in equation (6.1) allows for a straightforward extension to additional parameters obtained from felt reports or other sources, such as population density data products. Additional crowdsourced datasets, such as the one collected by the earthquake network initiative (EQN, Bossu et al., 2022), could contribute more information and improve modeling results. With the database of felt reports increasingly growing in coming years, also calibration of our model to specific continents or regions will be within reach in the near future.

6.7 Conclusions

In this study, we have presented the development of a data-driven, probabilistic model to rapidly distinguish high-impact from low-impact earthquakes based on LastQuake felt reports. For 14 % of 2,740 felt earthquakes in 2021, our model could have provided a classification estimate within 10 min of the event. The key strength of our model is the ability to correctly classify a large portion of 39 % of low-impact events with high confidence, such that urgent necessity for comprehensive emergency measures can be ruled out reliably and rapidly after such an event. Active participation of LastQuake users is a key prerequisite for the proper functionality of our model. If reports are collected numerously and fast, our model might be among the first available information sources to independently characterize the situation after a felt earthquake. Our inexpensive and easily implementable approach could be an effective option to potentially improve rapid response in regions where the installation of strong-motion networks in the near future is unlikely or unaffordable.

6.8 Data and resources

Earthquake impact data used in this study was derived from the Global Significant Earthquake Database (GSED) of the National Geophysical Data Centre and the World Data Service (NGDC/WDS) provided by the National Centers for Environmental Information (NCEI) available at https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.hazards:G012153; the International Events Database (EM-DAT) of the Université Catholique de Louvain, Belgium, available at https://public.emdat.be/; the Earthquake Impact Database (EID) available at https://erdbebennews.de/earthquake-impact-database-2021/. All impact sources were visited last on 26 October 2022. The U.S. Geological Survey's earthquake catalog was available at https://earthquake.usgs.gov/earthquakes/search/ to

fill gaps in the EM-DAT database. The python code developed within the scope of this study is available at https://git.gfz-potsdam.de/lilienka/lq_impact. All websites were last accessed in January 2023. The supplemental material to this article contains two additional texts, five figures, and two tables providing details concerning the data processing and modeling decisions.

6.9 Declaration of competing interests

The authors declare no competing interests.

6.10 Acknowledgements

The authors warmly thank David Wald and Danielle Sumy for their careful reviews and well-targeted suggestions that considerably improved the quality of our study. The authors also express their thanks to Cecilia Nievas for fruitful discussions regarding the availability of earthquake impact data. The first author acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS). This article was partially funded by the European Union (EU)'s Horizon 2020 Research and Innovation Program under Grant Agreement RISE Number 821115 and Grant GEO-INQUIRE Number 101058518. Opinions expressed in this article solely reflect the authors' views; the EU is not responsible for any use that may be made of information it contains. European-Mediterranean Seismological Centre (EMSC) thanks the SCOR Foundation for Science for its support. Part of this research is funded by the European Commission, ITNMarie Sklodowska-Curie New Challenges for Urban Engineering Seismology URBASIS-EU project, under Grant Agreement 813137.

Chapter 7

Discussion

The purpose of this chapter is to embed the studies and research questions (compare section 1.5.1) presented in this thesis in a broader scientific context. The research questions 1–3 concerning the suitability of artificial neural networks (ANNs) for ground motion modeling, the ANN assisted explicit incorporation of directivity in PSHA, and the optimal usage of pseudo-intensity reports in RRE are primarily addressed in the core studies of my thesis presented in chapters 3 – 6. However, for the sake of clarity, I have refrained from an exhaustive elaboration of some noteworthy, though uncritical aspects, which I would like to pick up in the following sections 7.1 – 7.3. Special emphasis is put on the subsequent discussion of research question 4, regarding the association of my studies to the field of Big Data Seismology, presented in section 7.4, as it has not been addressed explicitly in the core chapters.

7.1 Opportunities and limitations of the U-Net-based ground motion model

7.1.1 Groundbreaking advancements in ground motion modeling through ANNs?

No matter how cautiously and critically one approaches the topic of deep learning, bearing in mind breakthroughs like AlphaFold (section 2.2.2), there is always a glimmer of hope that the application of an ANN to a known problem could lead to breakthrough advances over earlier, more conventional modeling approaches. For the case of ground motion modeling, this would manifest in the form of dramatically reduced misfit between ground motion observations and predictions, i.e., considerably reduced aleatory uncertainty of ground motion models (GMMs) that has been found nearly unchanged for decades (e.g. Douglas and Edwards, 2016; Strasser et al., 2009). Indeed there is some evidence, such as the observation of spatial correlation of ground motion residuals (e.g. Loth and Baker, 2013), that a certain portion of what is currently considered aleatory uncertainty might actually be attributed to mechanisms that are not yet implemented in GMMs. However, one should bear in mind that the approach of empirical ground motion modeling itself is actually a massive simplification of the underlying physical processes, because the complexities of the source, path, and shallow geological properties of the site are not captured appropriately in the formulation of empirical ground motion models.

Therefore, it is to be expected that the aleatory uncertainty of GMMs will remain substantial, even if additional, previously unconsidered mechanisms and parameters are included. This deliberation has been confirmed by several studies in the literature (e.g. Derras et al., 2014; Dhanya and Raghukanth, 2018; Pozos-Estrada et al., 2014), where the authors could achieve GMM performance comparable to previous studies with relatively little effort and simple ANN architectures, but no advancement per se could be stated in terms of misfit. The results of my study regarding the development of an ANN-based, nonergodic GMM for the Kanto basin in Japan (chapter 3), are no exception in this regard, as the comparison to the GMMs by J. X. Zhao et al. (2016a,b,c) revealed comparable performance, insofar as one can compare these models at all due to their fundamentally different modeling approaches (nonergodic vs. ergodic). In summary I come to the conclusion that deep learning-based ground motion models will most likely not lead to a dramatic reduction in aleatory uncertainty in the future, and that coming studies in this direction should rather focus on the promising explorative properties that are discussed in the following paragraphs.

7.1.2 U-Net compared to other nonergodic modeling frameworks

The opportunity of (partially) waiving the ergodic assumption in ground motion modeling due to the abundance of strong motion data collected in the last years has put forth a number of modeling frameworks, the most prominent of which are linear (Abrahamson and Youngs, 1992) and non-linear (e.g. Kotha et al., 2016, 2020) mixed effects regression (MER), and the varying-coefficient model (VCM) based on gaussian process regression (Bussas et al., 2015; Landwehr et al., 2016). As an additional contribution to the pool of modeling frameworks, it is reasonable to compare some properties of the U-Net-based approach presented in chapter 3 to such previous concepts. Thereby I want to focus on the properties of the frameworks rather than specific model implementations.

As a consequence of treating the problem of ground motion modeling as an image processing task via the usage of a U-Net convolutional neural network, the information used to predict an IM is not limited to the set of predictive parameters at the target site, but also includes, for example, geological characterization in its vicinity. Methodologically, this is achieved via the use of finite sized filter masks, and the max-pooling operations in the U-Net (compare Fig. 3.1), that increase the size of the geographic area covered by the filter masks in the deeper layers of the U-Net. Considering the usage of site-parameters such as the depth to seismic bedrock, V_{S30}, or the thickness of a sedimentary layer, this means that, the 3D geological structure in the vicinity of a site is taken into account when predicting ground shaking. While the practical usefulness of this property is to be investigated in more detail in the future, I want to emphasize that, to my knowledge, the U-Net approach is the only GMM framework so far that natively provides such functionality without additional effort.

One property that distinguishes the VCM and the U-Net on the one hand, and the MER framework on the other hand is the way that the moderation of the ergodic assumption is achieved. While both VCM and U-Net introduce the absolute, continuous coordinates of individual sources and sites as predictive parameters, the MER framework allows the introduction of random effects, i.e. adjustments of model coefficients, at different scales, ranging from re-

gional adjustments according to tectonic setting or country (e.g. Kotha et al., 2016), to local calibration for individual station and earthquake locations (e.g. Kotha et al., 2020). Contrary to the VCM and the U-Net, if random effects are inferred in MER at a local level, e.g., station wise adjustments for site amplification, they cannot be simply transferred to locations other than those present in the training dataset. The development of an external, spatial correlation model, a native component of the VCM and the U-Net frameworks, would be necessary in this case.

A key feature that sets the U-Net approach aside from previous frameworks is the independence of any a priori defined functional form. While this fully data-driven method prohibits the introduction of theoretical knowledge to a ground motion model, it offers a fully objective perspective on the problem, which might lead to the identification of potentially existing, previously unconsidered systematic relations in strong motion datasets. Furthermore, being both independent of an equation and nonergodic, the U-Net framework is the first to offer the opportunity of investigating the predictive power of untypical predictive parameters, for which no closed form, physics-based description is available, in a nonergodic framework. Such novel relations or powerful predictive parameters could be identified via a considerable reduction in aleatory variability when compared to a different model with a comparable degree of ergodicity. However, the actual physical reasoning from such a relation would not be obvious to the modeler, because the coefficients of an ANN lack physical interpretability such that more detailed research would be required.

Another consequence related to the two previously discussed aspects is the fact that the U-Net framework assumes all relations in a strong motion dataset as dependent on both source and site coordinates, i.e., the U-Net-based GMM is fully nonergodic. In comparison, both the VCM and MER frameworks allow for the selection of individual terms in the GMM that shall be considered nonergodic (e.g. anelastic attenuation in Kotha et al. (2016)), whereas other terms can be chosen to remain ergodic (e.g. magnitude scaling in Landwehr et al. (2016)) if severe underdetermination of the resulting model is expected otherwise. Subsequently, it is reasonable to assume that a U-Net-based GMM will be insufficiently constrained outside the core range of the training dataset. While the precise impact on the predictive performance can be approximated provisionally from the application to a validation dataset, the possibility of obtaining completely unreasonable predictions for certain earthquake scenarios can never be ruled out. Consequently, when compared to the VCM and MER approaches, which fall back to an ergodic backbone model if evidence from data is lacking, great care must be taken when considering the application of a U-Net-based GMM in PSHA.

Finally, the practicality of the frameworks in terms of computational efficiency varies considerably. The necessity to invert a covariance matrix of shape $(N_D \times N_D)$, where N_D denotes the number of observations in a dataset, leads to demand in computational resources scaling with ${N_D}^3$, posing a severe computational challenge to the calibration of a VCM given a large dataset (Seeger, 2004). While various approximating methods have been suggested to reduce the computational load, utilization of high-performance computing equipment is indispensable when developing a VCM at regional scale. In contrast, models based on the MER and the U-Net frameworks can be calibrated on a decent desktop computer with a consumer-end GPU (for the

U-Net approach) within a reasonable amount of time.

In summary, I do consider the MER and the VCM frameworks to be the best options for application in PSHA due to the superior control that they grant to the modeler in terms of choosing the degree of nonergodicity, and the physical interpretability of the model coefficients. However, the U-Net-based approach offers some interesting explorative features regarding the assessment of novel parameters and the exploration of large datasets, as well as the incorporation of predictive information from the spatial vicinity of the target site.

7.1.3 Future perspectives

The key prerequisite for the successful training of a U-Net-based GMM is a large ground motion dataset recorded by a dense strong motion network operated over a long period of time in a seismically active region. While KiK-net is one of the most dense strong motion networks on the planet, I found that the misfit of the U-Net GMM in locations between training stations is still considerably increased compared to the training sites, and that the associated uncertainty is underestimated. Therefore, even higher station density would be a desireable feature. In the Kanto basin, and also in the rest of Japan, one possible option to achieve this would be to incorporate the equally dense K-net network. While a comprehensive flat file containing observations from both KiK-net and K-net over the entire operation time is not yet available, work is currently in progress to provide such a database (Sebastian von Specht, Karina Loviknes, personal communication), such that a substantial update of the suggested U-Net model is within reach in the near future.

Another possible future use case for U-Net-based GMMs is currently arising from the development of low-cost instruments based on MEMS (Micro-Electro-Mechanical-Systems, e.g., Holland, 2003) and the targeted deployment in huge numbers at city scale. For example, future plans for the instrumentation of the city of Tokyo encompass the deployment of 2,000 MEMS sensors in 1,000 different locations (Danijel Schorlemmer, personal communication). With such a high station density, interpolation between observations sites might not be a severe problem anymore.

In addition, the dense instrumentation of urban areas opens up new possibilities with regard to the investigation of the influence of urbanization on seismic wavefields. Utilizing data products such as land-use maps or maps of buildings and infrastructure as predictive parameters for a GMM is an easy to implement option in the U-Net framework as I demonstrated with the usage of the rather untypical site parameter "depth to bedrock" in my study. Since a theoretical model for such interaction is not in sight as of today, the U-Net offers a unique future opportunity in this regard.

While the necessity of spatial density of strong motion sensors can be addressed with the installation of more instruments, the lack of observations from large magnitude earthquakes due to limited observation times is a more persistent problem, that cannot be solved within the coming decades. However, the acquisition of a substantial dataset of low- to medium sized earthquakes from extremely dense low-cost networks that are deployed today might be achievable within few years, and therefore allow first experiments regarding the development of GMMs for urban areas in the near future. Although such a model would not be suitable for application

in PSHA, for the numerous reasons outlined above, I see the potential that such experiments will lead to insights that might subsequently influence the development of more conventional, equation-based GMMs. In summary I come to the conclusion that while U-Net-based GMMs will most likely not find entry into the realm of PSHA in the near future, this explorative way ground motion modeling is an appealing option to support the development of robust ground motion models in the long term.

7.2 ANN assisted incorporation of directivity effects in PSHA

7.2.1 Universality vs. reliability

In chapters 4 and 5 I have explored two different strategies for the explicit incorporation of rupture forward directivity in PSHA calculations via the utilization of ANNs. The first one is based on an overfitting approach that aims at a very accurate representation of a fix set of moment modifiers for ruptures from the 2022 New Zealand ERF model, and the second one targets a more universal model that is capable of predicting moment modifiers from representations of arbitrary earthquake ruptures. Although the latter approach seems to be the superior one, due to its universal applicability to arbitrary ruptures, I want to emphasize here that I consider the two approaches actually equivalent solutions for two different types of problems. Admittedly, a model that is capable of generalizing without restrictions to arbitrary rupture geometries could be considered the ultimate goal of this research efforts. However, as outlined, e.g., in section 5.5, unforeseen challenges such as the limitation of a model to generalize to ruptures, the predictive properties of which are principally well represented in the training set, but have a simpler parameterization, can never be ruled out. Therefore, the guaranteed functionality and full control over moment modifiers granted by the overfitting approach might be considered worth the considerable, though manageable, effort to generate ERF specific training datasets for PSHA studies at regional scale. On the other hand, there are use cases such as Shakemap calculations or seismic hazard models without a fix inventory of earthquake ruptures, for which the less accurate modeling via the generalizing approach and the limitations regarding generalization might be acceptable in order to achieve explicit incorporation of directivity effects at reasonable efforts. I therefore conclude that both approaches are worth pursuing in future research.

7.2.2 Relation between ANN_{ovf} and U-Net

While the ANN architecture used for the overfitting approach (ANN_{ovf}) and the U-Net architecture used for the generalizing approach seem to be unrelated at first glance, ANN_{ovf} can actually be interpreted as the decoder branch of a U-Net. The convolutional layers in the bottleneck of a U-Net contain a highly abstract, and compressed representation of the information provided in the input layer, which is subsequently decoded in the decoder branch to derive maps of moment modifiers. In the same way, one may consider the rupture ID given as the input to (ANN_{ovf}) to be an extremely compressed representation of an earthquake rupture, that is then decoded by (ANN_{ovf}) to derive the corresponding moment modifiers. Of course this comparison is slightly misleading, because a rupture ID does not provide any actual rupture characteristics,

and the overfitting strategy leads to made up pseudo-relations, but from a technical point of view this is how the design of ANN_{ovf} was inspired.

7.2.3 Integration of modifier of moments model and a GMM

The U-Net architectures utilized in chapter 3 for the sake of ground motion modeling and in chapter 4 for the sake of directivity modeling are conceptually identical, which opens up an opportunity to actually integrate ground motion modeling and explicit incorporation of directivity into a single computing step within a PSHA calculation. Integration of, e.g., the Stafford (2022) GMM utilized for PSHA calculations in chapter 4 into the modifier of moments model would basically require two modifications: The first one concerns the generation of the training dataset described in section 4.5.2.4, where instead of the moment modifiers, the actual modified moments, i.e., $(\mu + \Delta\mu_{dir}, \sigma_{dir})$ (compare section 4.3.2 and equations 4.3, 4.4) would have to be generated. The second modification concerns the input layer of the U-Net, where three additional channels would have to be included (the depth to the top of rupture z_{TOR}, the depth at which the shear wave velocity adapts a value of 1 km/s $Z1.0$, and V_{S30}) to complement the set of input parameters to the Stafford (2022) GMM. Subsequent training of a U-Net with this synthetic dataset would then lead to the simultaneous adaption of both the Stafford (2022) GMM and the moment modifiers, which could then efficiently be generated in a single step in a PSHA calculation.

7.3 Utilization of pseudo-intensity reports for rapid impact assessment

7.3.1 Performance on earthquakes from the Kahramanmaras sequence in February 2023

The M_W 7.8 and M_W 7.5 earthquakes from the Kahramanmaras sequence occuring on February 6th 2023 close to Pazarcik and Elbistan (Turkey) were the by far most incisive events over the course of my doctoral studies. The resulting 56,000 casualties and over 87 billion dollars of economic loss in Turkey and Syria places this sequence among the 5 most destructive earthquakes of the 21st century, the others being Sumatra 2004 (M_W 9.1, over 225.000 dead (Tsunami)), Kashmir 2005 (M_W7.6, over 75.000 dead), Sichuan 2008 (M_W 7.9, over 85.000 dead), and Haiti 2010 (M_W 7.0, over 310.000 dead)(data from National Geophysical Data Center / World Data Service, 2023).

Occuring close in time with the publication of our study in chapter 6, the Kahramanmaras earthquakes represented, from a scientific perspective, a rare opportunity to test the performance of our model on an extremely 'high-impact' earthquake. The collections of pseudo-intensities for the two events are presented in Fig. 7.1, whereas the characterization in terms of average pseudo-intensity $\bar{I}$ and average distance to the barycenter $\bar{R}$ together with the classification results according to our model are presented in Fig. 7.2. The assigned probabilities of being high impact ($P(H|X_{2D}) = 0.68$ and $P(H|X_{2D}) = 0.51$ for the Pazarcik and Elbistan events, respectively) are among the largest compared with the earthquakes from the training dataset,

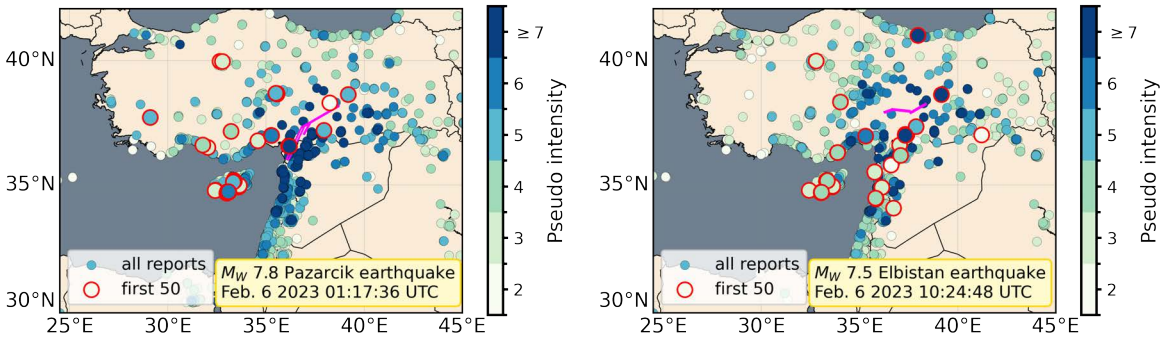

Figure 7.1: LastQuake pseudo-intensities collected during the Kahramanmaras earthquake sequence on Feb. 6th 2023. Large pseudo-intensities overlay small ones. The first 50 reports overlay later ones. Magenta lines represent finite fault solutions from U.S. Geological Survey (2023a,b).

and reasonably suggest that the events are more likely 'high-impact' than 'low-impact'. However, in the light of the extraordinarily disastrous consequences that these earthquakes finally had and the low impact thresholds that we defined to consider an earthquake as 'high-impact', these results are not fully satisfactory. As pointed out in section 6.5,

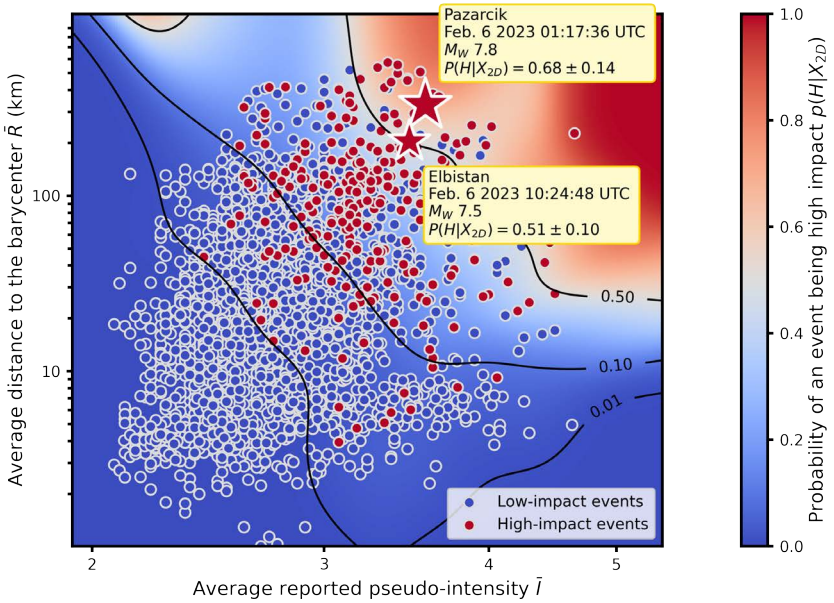

Figure 7.2: Probabilistic classification of the Kahmaranmaras earthquakes (compare Fig. 7.1) of being 'high-impact' according to our model presented in chapter 6.

the small absolute values of $P(H|X_{2D}) = 0.68$ and $P(H|X_{2D}) = 0.51$ can be explained with the severe imbalance of 'high impact' and 'low impact' events in the training database, that basically leads to typical 'high-impact' events being more likely 'low-impact events with untypically large $\bar{I}$ and $\bar{R}$. However, the values for $\bar{I}$ and $\bar{R}$ are also not outstanding compared to the training events with significantly smaller impact. I identified two possible reasons why the values $\bar{I}$ and $\bar{R}$ are actually smaller than one would expect. First of all, from Fig. 7.1, one can see that the first

50 reports utilized to characterize the events barely sample the strongest affected regions in the close vicinity of the earthquake rupture, thus, leading to the average reported pseudo-intensity $\bar{I}$ biasing low. This so called doughnut effect (Bossu et al., 2018a), is also visible if all available reports are considered, especially for the Elbistan event. The second reason could be the sparse availability of pseudo-intensity reports from Syria, Iraq, and Jordan, potentially leading to the average distance to the barycenter $\bar{R}$ being biased low. As discussed already in section 6.6.4, both issues are already well known and motivate ongoing research to identify predictive features that are less affected by such geographic distortions.

Finally it should be noted that the Kahramanmaras earthquakes are actually not prime examples for the profitable usage of our model. The disastrous impact in a densely populated and urbanized region was obvious from the beginning, thus, the probabilistic evaluation whether these events have been 'high-impact' would have been superfluous in practice. As a counterexample, I would like to refer again to the M_W 5.9 event that occured on June 22nd in Afghanistan discussed in section 6.6.3, where the event hit a sparsely populated rural area where the occurence of severe impact was unclear for several hours after the event.

7.4 Connection of this thesis with Big Data Seismology

7.4.1 U-Net-based ground motion model

The development of a deep learning-based GMM for the Kanto region in Japan (chapter 3) reveals two key features that let this study appear as a contribution to Big Data Seismology. For one, there is the utilization of an advanced ANN architecture (U-Net), that is trained on consumer-end, though deep learning optimized, computational resources. The second aspect is the fully data-driven approach in which the *Value* is extracted from data, one in which we pretend to not have any theoretical knowledge on ground motion modeling and thus consider the *Value* to be "hidden". In discarding all the theoretical knowledge that has been acquired on the topic of ground motion modeling in the last decades, we adopt a Big Data perspective that considerably differs from that of studies that choose well established equation-based modelling approaches, such as those by J. X. Zhao et al. (2016a,b,c) that were used for comparative purposes. Starting from this perspective, the Bahrampouri et al. (2021) dataset was chosen due to the density and long-term operation of the underlying KiK-net strong motion network that is absolutely crucial to at least partially constrain the comparatively large amount of coefficients in the U-Net. Therefore, one can also state that the selection of the dataset was Big Data inspired.

However, when evaluating the Bahrampouri et al. (2021) database objectively, it exhibits very little Big Data characteristics. The flat file including all metadata has in fact a size of about 15 gigabytes, and is therefore easily manageable on a decent laptop or desktop computer. This rather small *Volume* does not prevent the application of traditional modeling frameworks such as mixed effects regression, and the application of advanced algorithms such as deep learning is not necessary from a technical perspective. Rapid processing of records from ongoing earthquakes is also not critical, since there is not much value in updating modern GMMs with few records from a specific earthquake, therefore *Velocity* is also not a challenging aspect when processing utilizing this type of data. Furthermore, all issues regarding *Veracity* have been addressed in

the previous strong motion processing (Bahrampouri et al., 2021), such that the data quality or missing values do not play a role at this stage. On top of that, the flat file is highly structured, very easily accessible, and contains a manageable amount of well defined parameters (strong motion features), thus, one has to state low *Variety*. In summary, one can state that the strong motion flat file is actually a prime example for a small dataset.

One should mention that a different picture emerges for the underlying waveform database the flat file is obtained from. With a size of ∼1 terabyte, manual identification and inspection of earthquake signals becomes infeasible (*Volume* issue) and a small scale server is a worthwhile computational option to conveniently manage the required automated processing steps. Strong motion processing involves a number of typical *Veracity* issues that need to be solved in an automatic way. For example, strong motion instruments in KiK-net do not acquire continuous waveforms, but are triggered by exceedance of a certain ground acceleration threshold, which sometimes leads to missing bodywaves in the seismograms. Automatic evaluation of the signal to noise ratio and the detection of phase arrivals from multiple earthquakes are additional typical *Veracity* issues in strong motion processing.

In summary, I follow that the presented study is an example for the application of an advanced deep learning method on a small dataset. The choice to utilize an ANN was motivated by curiosity how the results would compare to those of a traditional modeling approach, and not because the dataset suggested this to be the most appropriate modeling choice to derive a ready-to-use, robust GMM, at the present data situation. In my opinion, the severe limitations regarding the applicability of our model in PSHA is a direct consequence of this circumstance, and highlights the fact that this is not a Big Data Seismology study in the strictest sense. However, this might change for similar studies in the future: The installation of extremely dense networks of low-cost sensors in urban areas as planned currently, e.g., with 2,000 sensors for the city of Tokyo (Danijel Schorlemmer, Personal Communication), might lead to rapidly growing strong motion flat files in seismically active areas and therefore shift the available data *Volumes* towards Big Data dimensions. Furthermore, in areas such as Tokyo, also the urbanization itself interacting with seismic wavefields should be considered in an eventual GMM derived from such datasets. Since theoretical modeling of such interaction is not in sight, the fictional Big Data perspective on the *Value* of ground motion data adopted in this study might in fact be a realistic one for incorporation of, e.g., land-use data in GMMs for urban areas in the future.

7.4.2 Explicit incorporation of directivity in PSHA

Both studies regarding the incorporation of explicit modeling of rupture forward directivity in Probabilistic Seismic Hazard Assessment (PSHA) presented in chapters 4 and 5, make extensive use of artificial neural networks. While in chapter 4 an ANN is adopted for the uncommon task of compressing and storing a database for efficient retrieval in its hidden layers, in chapter 5 I present a more traditional usage in which a U-Net ANN is trained to learn how to predict average site amplification and its variability in the vicinity of an earthquake rupture, given its geometric and seismic properties.

The database underlying both studies is synthetically generated, comprises only two target parameters, and takes up a size of approximately 40 gigabytes on hard disk, therefore, no

challenges regarding *Volume* or *Variety* can be stated. Synthetic generation of a dataset grants optimal control and knowledge of the final outcome, thus, challenges in terms of *Veracity* and *Velocity* do not exist.

In chapter 4, we utilize the ANN to learn the samples from the dataset "by heart", i.e., we suggest a purely technically motivated procedure to efficiently compress a dataset on hard disk and make it easily accessible during PSHA calculations. Thus, one can state that no *Value* is extracted from the data. The study presented in chapter 5 is, however, a different case, one in which the U-Net ANN is trained to actually learn a meaningful relation between the earthquake rupture properties and the directivity related, spatial amplification patterns in their vicinities. This endeavor may be considered an aggregation of *Value* in the sense of Big Data, since no physical theory is available to model the relations in question, yet, and therefore the *Value* of the dataset is 'hidden'.

Considering the above thoughts, I come to the conclusion that, despite the Big Data style aggregation of *Value* in chapter 5, the operation on a small, synthetically generated dataset, and the unconventional usage of an ANN in chapter 4 prohibits the consideration of the directivity related studies in this thesis as contributions to Big Data Seismology.

7.4.3 Utilization of pseudo-intensity reports for rapid impact assessment

Among the various databases utilized within the scope of this thesis, the database acquired via the LastQuake service is in my opinion the most interesting one with regard to Big Data. While only the numerical data in the form of reported pseudo-intensity values are utilized in this thesis, the service offers also the opportunity to submit short text messages and photos characterizing the post-earthquake situation. Joint analysis of these three datatypes may be considered a challenge in terms of *Variety*. Non-expert citizens should not be considered calibrated instruments, and the perception of ground shaking varies significantly from person to person. Especially at high levels of shaking, personal experiences are potentially emotionally biased, leading to increased uncertainties at the upper end of the macroseismic scale. Subsequently, pseudo-intensities do not provide a high degree of *Veracity*. Also *Velocity* plays an important role in the light of real-time usage of LastQuake data for Earthquake Early Warning or Rapid Response to Earthquakes: Processing the abrupt onset of incoming reports after an earthquake, on the fly identification of outliers, assigning reports to the correct earthquake, and yielding the product to subsequent steps, e.g., Shakemap, is a pipeline of significant complexity, the efficient implementation of which is crucial.

While the confrontation with Big Data challenges is consequentially inevitable for potential follow-ups of my study presented in chapter 6, I want to clarify that only the pseudo-intensity values, and a fixed version of the database (2014–2021) were considered so far, ignoring the *Velocity* and *Variety* challenges as far as my study is concerned. The limited *Veracity*, however, was considered both in the data selection (removing intensities greater than 10), and, perhaps more importantly, in the selection of the modeling target: While high quality seismic observations and macroseismic intensities offer a data quality that enables actual quantification of expected economic losses and number of casualties via loss assessment frameworks such as PAGER, with the classification into 'high-impact' and 'low-impact' events, we chose a rougher

modeling target according to the limited *Veracity* of pseudo-intensities. Relating collections of LastQuake pseudo-intensities to the impacts of earthquakes is a previously non-attempted task for which no theoretical model is available. Subsequently, one may consider that the *Value* that we gather with our fully data-driven methodology is 'hidden' in the LastQuake database, adding a glimpse of Big Data to this study. On the other hand, I must state that the *Volume* of the utilized database is actually very small ($\sim$ 150 megabytes), a fact that may, however, change when incorporating also the text messages and photos.

Again, while revealing some Big Data Seismology typical aspects, the presented study shows, in my opinion, primarily characteristics of a classical, statistical work on a small dataset, yielding a proof of concept that may enable similar, but more sophisticated follow-up studies in the future. Incorporating photos and text messages, and potentially additional data sources as suggested in section 6.6.5, may then truly move this kind of study in the realm of Big Data in the future.

7.4.4 Summary

Seismology is evidently entering the era of Big Data, increasingly facing datasets revealing Big Data characteristics and relying on advanced algorithms for processing and gathering new insights. Especially on the methodological side, I have often come into contact with this development during my studies, in particular regarding the utilization of ANNs and the data-centric perspectives that I took on the addressed research questions. However, I would characterize the analyzed datasets primarily as Small Data according to the limited overlap with the properties known as "The 5 Vs of Big Data". As a consequence of applying advanced algorithms to Small Data, the results presented in this study are explorative, and represent foremost proofs of concepts the value of which will become clear in the course of time.

However, it is a common property of all databases considered in this thesis, that they will most likely develop into Big Data in the foreseeable future. The deployment of large and dense seismic networks of classical high-quality seismometers but also emerging new instruments such as low-cost micro electro mechanical systems (MEMs) (e.g. Holland, 2003) or fibre-optic cables for usage in distributed acoustic sensing (DAS) (e.g. Lindsey and Martin, 2021), is in large swing and will lead to a rapid increase in the *Volume* and *Variability* of future seismic databases. Moreover, incorporation of various types of crowdsourced data will also lead to more *variable*, but also potentially less reliable databases in terms of *Veracity*. The insights gained from this thesis may therefore represent the fundament to develop stable and practical solutions upon in the years to come.

Chapter 8

Conclusions and outlook

The key concept of my thesis was to take data centric prespectives on various challenges regarding seismological contributions to mitigation of earthquake consequences in the fields of probabilistic seismic hazard assessment and rapid response to earthquakes. The exploration of the potential benefits from novel algorithms, primarily deep learning, was thereby a key focus. My doctoral studies have lead me to the following key conclusions:

The U-Net neural network architecture allows the computationally efficient implementation of fully data-driven, nonergodic ground motion models. The key property that sets this novel approach aside from previous attempts is the fact that no a priori assumptions of any form are introduced during the model calibration and, thus, the resulting model can be considered a fully objective interpretation of strong-motion data. From the practical implementation of a prototype model for the Kanto region, Japan, I found that the overall scalings of ground motion intensity with magnitude and source-to-site distance were adopted successfully, and that predictions are comparable to those from well established modeling frameworks. However, the applicability of the model to large magnitude earthquakes and short source-to-site distances is limited due to the highly nonuniform distribution of strong motion datasets. While previous methods like the varying-coefficient model and non-linear mixed-effects regression have been identified as the more suitable options for standard PSHA, I see the future potential of the U-Net method primarily in supporting development of equation-based ground motion models, given the ability to include also atypical characteristics of ground motion for which closed-form physical equations cannot be constructed.

I have shown that artificial neural networks can be utilized to incorporate explicit modeling of directivity into PSHA calculations. In a first step I have demonstrated, both conceptually and in a practical PSHA application in a New Zealand case study, that a database of moment modifiers, i.e., ground motion adjustment terms accounting for the average influence of rupture forward directivity, can be stored inside an artificial neural network, such that it is both easily distributable and efficiently retrievable for thousands of ruptures in a PSHA calculation. In an alternative approach, I started to explore the possibility to train an ANN such that it learns meaningful relations between geometric and seismic properties of arbitrary ruptures on the one hand, and the corresponding moment modifiers on the other hand. The preliminary results indicate that such a model is clearly within reach, but that considerable extension of the training dataset is a necessary step before application in PSHA can be suggested.

Our study of LastQuake pseudo-intensity report collections revealed considerable correlation between the statistical properties of such and the severity of societal impact of the corresponding earthquake. I demonstrated that this correlation can be used to formulate a probabilistic model that is capable of assigning a probability of an earthquake to have considerable societal impact within minutes of the origin time. While the unambiguous classification of high-impact events turned out to be a challenging task, I identified the ability to reliably assign small probabilities of impact to low-impact earthquakes as a key strength of the model. Furthermore I come to the conclusion, that the task of moving the suggested model into practical application cannot be solved solely from the seismological perspective, but that social scientists, authorities, and disaster managers must be incorporated in the next steps.

Despite the extensive use of deep learning, and the adaption of data-centric perspectives on modeling problems, the datasets utilized in this thesis are in many respects representatives of Small Data, leading to the categorization of the studies presented in this thesis more towards classical rather than Big Data Seismology. As a consequence, the majority of the results presented are to be understood as proofs of concepts whose significance for the mitigation of earthquake disasters are expected to unfold in the foreseeable future as the size and complexity of the datasets concerned increase.

Among the various open questions that arose over the cause of my doctoral studies, I identified the following as the most promising ones to invest further research into:

The U-Net ground motion modeling framework poses a unique opportunity to investigate the influence of incorporating not only the geophysical site properties at a specific site of interest, but also in its vicinity, into the prediction of ground motion intensity measures. This unique property potentially allows for the incorporation of the 3D subsurface structure in empirical ground motion modeling and therefore poses, in my opinion, a promising opportunity to improve ground motion models in the future.

While the suitability of the deep learning approach for linking the average effect of rupture forward directivity in the vicinity of an earthquake rupture to its geometrical and seismic properties has been demonstrated, a fully operational model for usage in arbitrary seismic hazard models and rupture geometries is yet to be established. The generation of the required additional synthetic datasets of moment modifiers is in my opinion the critical next step that seems achievable with reasonable effort in the near future. An ANN trained on such an extended dataset would enable the applicability of the resulting model to use cases, where no fix inventory of earthquake ruptures is available, e.g., ShakeMap calculations or non ERF-based PSHA. Such a model would, therefore, considerably ease the cost of explicit incorporation of directivity for the entire engineering seismology community.

Following the demonstration of how the correlation between statistical properties of pseudo-intensity collections and earthquake impact severity can be utilized in a statistical model to predict the impact of an earthquake, incorporation of additional datasets such as macroseismic intensities from "Did you feel it?" (Wald et al., 1999), or smartphone-based ground motion records from the EQN service (Bossu et al., 2022), as well as refined interpretation of modeling results in the form of a traffic light system are among the next steps. The practical implementation of such an improved framework might potentially be a valuable extension to previously

established rapid impact assessment methodologies, especially in remote regions of the planet that currently lack expensive seismic instrumentation.

With regard to the overarching question of the potential of data-centric perspectives and the application of machine learning methods in ground motion modelling and RIA, I come to the conclusion that the insights gained today from the investigation of yet small datasets provide the foundation for profitable, appropriate usage in the future, when large, diverse datasets will actually pose challenges to the application of conventional approaches.

Bibliography

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. URL: https://zenodo.org/record/6574269.

Abbott, E., N. Horspool, M. Gerstenberger, R. Huso, C. Van Houtte, G. McVerry, and S. Canessa (2020). "Challenges and opportunities in New Zealand seismic hazard and risk modeling using OpenQuake". In: *Earthquake Spectra* 36.S1, pp. 210–225. DOI: 10.1177/8755293020966338.

Abercrombie, R. E., P. Poli, and S. Bannister (2017). "Earthquake Directivity, Orientation, and Stress Drop Within the Subducting Plate at the Hikurangi Margin, New Zealand". In: *Journal of Geophysical Research: Solid Earth* 122.12, pp. 10, 176–10, 188. DOI: 10.1002/2017JB014935.

Abrahamson, N. A. and R. R. Youngs (1992). "A stable algorithm for regression analyses using the random effects model". In: *Bulletin of the Seismological Society of America* 82.1, pp. 505–510. DOI: 10.1785/bssa0820010505.

Abrahamson, N. A. (2000). "Effects of rupture directivity on probabilistic seismic hazard analysis". In: *Proceedings of the Sixth International Conference on Seismic Zonation: Managing Earthquake Risk in the 21st Century, Palm Springs, CA, 12-15 November 2000*.

Abrahamson, N. A., N. M. Kuehn, M. Walling, and N. Landwehr (2019). "Probabilistic seismic hazard analysis in california using nonergodic ground-motion models". In: *Bulletin of the Seismological Society of America* 109.4, pp. 1235–1249. DOI: 10.1785/0120190030.

Abrahamson, N. A., W. J. Silva, and R. Kamai (2014). "Summary of the ASK14 ground motion relation for active crustal regions". In: *Earthquake Spectra* 30.3, pp. 1025–1055. DOI: 10.1193/070913EQS198M.

Al Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). "The variability of ground-motion prediction models and its components". In: *Seismological Research Letters* 81.5, pp. 794–801. DOI: 10.1785/gssrl.81.5.794.

Al Atik, L., N. Gregor, S. Mazzoni, and Y. Bozorgnia (2023). *Directivity-Based Probabilistic Seismic Hazard Analysis for the State of California : Report 2, Directivity Implementation*. Tech. rep. Los Angeles: University of California, Los Angeles, p. 211. DOI: 10.34948/N3KS3B.

Allen, R. V. (1978). "Automatic earthquake recognition and timing from single traces". In: *Bulletin of the Seismological Society of America* 68.5, pp. 1521–1532. DOI: 10.1785/bssa0680051521.

Ameri, G., S. Drouet, P. Traversa, D. Bindi, and F. Cotton (2017). "Toward an empirical ground motion prediction equation for France: accounting for regional differences in the source stress parameter". In: *Bulletin of Earthquake Engineering* 15.11, pp. 4681–4717. DOI: 10.1007/s10518-017-0171-1.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue (2013). *Peer NGA-West2 database*. Tech. rep. May 2013. Berkeley, CA: Pac. Earthq. Eng. Res. Ctr, pp. 1–136.

Ancheta, T. D., R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue (2014). "NGA-West2 database". In: *Earthquake Spectra* 30.3, pp. 989–1005. DOI: 10.1193/070913EQS197M.

Anderson, J. G. and J. N. Brune (1999). "Probabilistic seismic hazard analysis without the ergodic assumption". In: *Seismological Research Letters* 70.1, pp. 19–28. DOI: 10.1785/gssrl.70.1.19.

Arrowsmith, S. J., D. T. Trugman, J. MacCarthy, K. J. Bergen, D. Lumley, and M. B. Magnani (2022). "Big Data Seismology". In: *Reviews of Geophysics* 60.2, pp. 1–55. DOI: 10.1029/2021RG000769.

Atkinson, G. (2022). *Backbone Ground-Motion Models for Crustal, Interface and Slab Earthquakes in New Zealand*. Tech. rep. GNS Science Report 2022/11. Lower Hutt: GNS Science, pp. 1–68.

Atkinson, G. and D. J. Wald (2007). ""Did you feel it?" Intensity data: A surgprisingly good measure of earthquake ground motion". In: *Seismological Research Letters* 78.3, pp. 362–368. DOI: 10.1785/gssrl.78.3.362.

Bahrampouri, M., A. Rodriguez-Marek, S. Shahi, and H. Dawood (2021). "An updated database for ground motion parameters for KiK-net records". In: *Earthquake Spectra* 37.1, pp. 505–522. DOI: 10.1177/8755293020952447.

Baker, J. W. (2007). "Quantitative classification of near-fault ground motions using wavelet analysis". In: *Bulletin of the Seismological Society of America* 97.5, pp. 1486–1501. DOI: 10.1785/0120060255.

Baker, J. W., B. A. Bradley, and P. J. Stafford (2021a). "Magnitude and Distance Distribution". In: *Seismic Hazard and Risk Analysis*. 1st ed. Cambridge: Cambridge University Press, pp. 163–165.

Baker, J. W., B. A. Bradley, and P. J. Stafford (2021b). "Probabilistic Seismic Hazard Analysis". In: *Seismic Hazard and Risk Analysis*. 1st ed. Cambridge: Cambridge University Press, pp. 5–10. DOI: 10.1017/9781108425056.

Bayless, J. and P. G. Somerville (2013). *Bayless and Somerville Model*. Tech. rep. PEER Report No. 2013/09. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 29–46.

Bayless, J., P. G. Somerville, and A. Skarlatoudis (2020). *A Rupture Directivity Adjustment Model Applicable to the NGA-West2 Ground Motion Models and Complex Fault Geometries*. Tech. rep. U. S. Geological Survey Technical Report, pp. 1–72.

Becker, J. S., S. H. Potter, S. K. McBride, A. Wein, E. E. Doyle, and D. Paton (2019). "When the earth doesn't stop shaking: How experiences over time influenced information needs, communication, and interpretation of aftershock information during the Canterbury Earthquake Sequence, New Zealand". In: *International Journal of Disaster Risk Reduction* 34.December 2018, pp. 397–411. DOI: 10.1016/j.ijdrr.2018.12.009.

Ben-Zion, Y. and C. G. Sammis (2003). "Characterization of fault zones". In: *Pure and Applied Geophysics* 160.3-4, pp. 677–715. DOI: 10.1007/PL00012554.

Bertero, V. V., S. A. Mahin, and R. A. Herrera (1978). "Aseismic design implications of near-fault san fernando earthquake records". In: *Earthquake Engineering & Structural Dynamics* 6.1, pp. 31–42. DOI: 10.1002/eqe.4290060105.

Bilham, R. (2009). "The seismic future of cities". In: *Bulletin of Earthquake Engineering* 7.4, pp. 839–887. DOI: 10.1007/s10518-009-9147-0.

Bishop, C. M. (2006). "Pattern Recognition and Machine Learning". In: *Pattern Recognition and Machine Learning*. Springer New York. DOI: 10.1007/978-0-387-45528-0.

Blitzstein, J. K. and J. Hwang (2014). *Introduction to probability*. 1st. New York: Chapman and Hall/CRC, pp. 385–401. DOI: 10.1201/b17221.

Bommer, J. J., J. Douglas, F. Scherbaum, F. Cotton, H. Bungum, and D. Fäh (2010). "On the selection of ground-motion prediction equations for seismic hazard analysis". In: *Seismological Research Letters* 81.5, pp. 783–793. DOI: 10.1785/gssrl.81.5.783.

Boore, D. M. (2010). "Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion". In: *Bulletin of the Seismological Society of America* 100.4, pp. 1830–1835. DOI: 10.1785/0120090400.

Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014). "NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes". In: *Earthquake Spectra* 30.3, pp. 1057–1085. DOI: 10.1193/070113EQS184M.

Bossu, R., F. Finazzi, R. Steed, L. Fallou, and I. Bondár (2022). ""Shaking in 5 Seconds!"—Performance and User Appreciation Assessment of the Earthquake Network Smartphone-Based Public Earthquake Early Warning System". In: *Seismological Research Letters* 93.1, pp. 137–148. DOI: 10.1785/0220210180.

Bossu, R., M. Landès, F. Roussel, R. Steed, and G. Mazet-Roux (2023). *Crowdsourced earthquake felt reports collected via the LastQuake service (2014 - 2021 + examples from 2022)*. DOI: https://doi.org/10.5880/GFZ.2.6.2023.001. URL: https://doi.org/10.5880/GFZ.2.6.2023.001.

Bossu, R., M. Landès, F. Roussel, R. Steed, G. Mazet-Roux, S. S. Martin, and S. Hough (2017). "Thumbnail-based questionnaires for the rapid and efficient collection of macroseismic data from global earthquakes". In: *Seismological Research Letters* 88.1, pp. 72–81. DOI: 10.1785/0220160120.

Bossu, R., M. Landès, F. Roussel, and R. Steed (2018a). "Felt reports for rapid mapping of global earthquake damage: The doughnut effect?" In: *Seismological Research Letters* 89.1, pp. 138–144. DOI: 10.1785/0220170129.

Bossu, R., F. Roussel, L. Fallou, M. Landès, R. Steed, G. Mazet-Roux, A. Dupont, L. Frobert, and L. Petersen (2018b). "LastQuake: From rapid information to global seismic risk reduc-

tion". In: *International Journal of Disaster Risk Reduction* 28.February, pp. 32–42. DOI: 10.1016/j.ijdrr.2018.02.024.

Bradley, B. A. (2012). "Strong ground motion characteristics observed in the 4 September 2010 Darfield, New Zealand earthquake". In: *Soil Dynamics and Earthquake Engineering* 42.September 2010, pp. 32–46. DOI: 10.1016/j.soildyn.2012.06.004.

Bradley, B. A. and J. W. Baker (2015). "Ground motion directionality in the 2010-2011 Canterbury earthquakes". In: *Earthquake Engineering and Structural Dynamics* 44.3, pp. 371–384. DOI: 10.1002/eqe.2474.

Bradley, B. A., D. Pettinga, J. W. Baker, and J. Fraser (2017). "Guidance on the utilization of earthquake- induced ground motion simulations in engineering practice". In: *Earthquake Spectra* 33.3, pp. 809–835. DOI: 10.1193/120216EQS219EP.

Bradley, B. A., M. C. Quigley, R. Van Dissen, and N. J. Litchfield (2014). "Ground motion and seismic source aspects of the canterbury earthquake sequence". In: *Earthquake Spectra* 30.1, pp. 1–15. DOI: 10.1193/030113EQS060M.

Bussas, M., C. Sawade, T. Scheffer, and N. Landwehr (2015). "Varying-coefficient models with isotropic Gaussian process priors". In: pp. 1–17. arXiv: 1508.07192.

Callaway, E. (2020). "'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures". In: *Nature* 588.7837, pp. 203–204. DOI: 10.1038/d41586-020-03348-4.

Campbell, K. W. and Y. Bozorgnia (2014). "NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra". In: *Earthquake Spectra* 30.3, pp. 1087–1114. DOI: 10.1193/062913EQS175M.

Casey, J. A., S. Goldman-Mellor, and R. Catalano (2018). "Association between Oklahoma earthquakes and anxiety-related Google search episodes". In: *Environmental Epidemiology* 2.2, e016. DOI: 10.1097/EE9.0000000000000016.

Chiou, B. S. and P. Spudich (2013). *The Chiou and Spudich NGA-West 2 Directivity Predictor DPP*. Tech. rep. PEER Report No. 2013/09. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 85–94.

Chiou, B. S. and R. R. Youngs (2014). "Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra". In: *Earthquake Spectra* 30.3, pp. 1117–1153. DOI: 10.1193/072813EQS219M.

Chollet, F. et al. (2015). *Keras*. URL: https://keras.io.

Cornell, C. (1968). "Engineering Seismic Risk Analysis." In: *Bulletin of the Seismological Society of America* 58.5, pp. 1583–1606.

Cotton, F., F. Scherbaum, J. Bommer, and H. Bungum (2006). "Criteria for selecting and adjusting ground-motion models for specific target regions: Application to central Europe and rock sites". In: *Journal of Seismology* 10.2, pp. 137–156. DOI: 10.1007/s10950-005-9006-7.

Cremen, G. and C. Galasso (2020). "Earthquake early warning: Recent advances and perspectives". In: *Earth-Science Reviews* 205.March. DOI: 10.1016/j.earscirev.2020.103184.

Derras, B., P. Y. Bard, and F. Cotton (2014). "Towards fully data driven ground-motion prediction models for Europe". In: *Bulletin of Earthquake Engineering* 12.1, pp. 495–516. DOI: 10.1007/s10518-013-9481-0.

Derras, B., P. Y. Bard, F. Cotton, and A. Bekkouche (2012). "Adapting the neural network approach to PGA prediction: An example based on the KiK-net data". In: *Bulletin of the Seismological Society of America* 102.4, pp. 1446–1461. DOI: 10.1785/0120110088.

Dewey, J. and P. Byerly (1969). "The early history of Seismometry (to 1900)". In: *Bulletin of the Seismological Society of America* 59.1, pp. 183–227. DOI: 10.1785/BSSA0590010183.

Dhanya, J. and S. T. Raghukanth (2018). "Ground Motion Prediction Model Using Artificial Neural Network". In: *Pure and Applied Geophysics* 175.3, pp. 1035–1064. DOI: 10.1007/s00024-017-1751-3.

Donahue, J. L., J. P. Stewart, N. Gregor, and Y. Bozorgnia (2019). *Ground-Motion Directivity Modeling for Seismic Hazard Applications*. Tech. rep. PEER Report No. 2019/03. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 1–62.

Douglas, J. (2003). "Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates". In: *Earth-Science Reviews* 61.1-2, pp. 43–104. DOI: 10.1016/S0012-8252(02)00112-5.

Douglas, J. (2022). *Ground Motion Prediction Equations 1964-2022*. URL: http://www.gmpe.org.uk/ (visited on Sept. 11, 2023).

Douglas, J. and B. Edwards (2016). "Recent and future developments in earthquake ground motion estimation". In: *Earth-Science Reviews* 160, pp. 203–219. DOI: 10.1016/j.earscirev.2016.07.005.

Esteva, L. and E. Rosenblueth (1964). "Espectros de temblores a distancias moderadas y grandes". In: *Boletin Sociedad Mexicana de Ingenieria Sesmica* 2.1, pp. 1–18.

Field, E., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, K. R. Milner, M. T. Page, T. Parsons, P. M. Powers, B. E. Shaw, W. R. Thatcher, R. J. Weldon, and Y. Zeng (2014). "Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model". In: *Bulletin of the Seismological Society of America* 104.3, pp. 1122–1180. DOI: 10.1785/0120130164.

Field, E., T. Jordan, and C. Cornell (2003). "OpenSHA: A developing community-modeling environment for seismic hazard analysis". In: *Seismological Research Letters* 74.4, pp. 406–419. DOI: 10.1785/gssrl.74.4.406.

Folesky, J., J. Kummerow, and S. A. Shapiro (2018). "Patterns of Rupture Directivity of Subduction Zone Earthquakes in Northern Chile". In: *Journal of Geophysical Research: Solid Earth* 123.12, pp. 10, 785–10, 796. DOI: 10.1029/2018JB016331.

Fujimoto, K. and S. Midorikawa (2006). "Relationship between Average Shear-Wave Velocity and Site Amplification Inferred from Strong Motion Records at Nearby Station Pairs". In: *Journal of JAEE* 6.1, pp. 11–22. DOI: 10.5610/jaee.6.11.

Gabriel, A.-A., T. Ulrich, M. Marchandon, J. Biemiller, and J. Rekoske (2023). "Rapid 3D dynamic rupture modeling of the February 6, 2023, Kahramanmaras, Turkey, Mw7.8 and Mw7.7 earthquake doublet". In: pp. 1–29. arXiv: 2308.02144. URL: http://arxiv.org/abs/2308.02144.

GEBCO Bathymetric Compilation Group 2023 (2023). *The GEBCO_2023 Grid - a continuous terrain model of the global oceans and land*. URL: https://doi.org/10.5285/F98B053B-0CBC-6C23-E053-6C86ABC0AF7B.

GeoScienceWorld (2023). *GeoScienceWorld*. URL: https://pubs.geoscienceworld.org/ (visited on Aug. 23, 2023).

Gerstenberger, M., S. Bora, B. A. Bradley, C. DiCaprio, R. Van Dissen, G. M. Atkinson, C. Chamberlain, A. Christophersen, K. Clark, and G. Coffey (2022a). *New Zealand National Seismic Hazard Model 2022 Revision: model, hazard and process overview*. Tech. rep. GNS Science Report 2022/57. Lower Hutt: GNS Science, pp. 1–106.

Gerstenberger, M., R. Van Dissen, C. Rollins, C. DiCaprio, C. Chamberlain, A. Christophersen, G. Coffey, S. Ellis, P. Iturrieta, K. Johnson, N. Litchfield, S. J. Rastin, H. Seebeck, M. Stirling, and K. K. Thingbaijam (2022b). *The Seismicity Rate Model for the 2022 New Zealand National Seismic Hazard Model*. Tech. rep. GNS Science Report 2022/47. Lower Hutt: GNS Science, pp. 1–156.

Gerstenberger, M., G. McVerry, D. Rhoades, and M. Stirling (2014). "Seismic hazard modeling for the recovery of christchurch". In: *Earthquake Spectra* 30.1, pp. 17–29. DOI: 10.1193/021913EQS037M.

Graves, R. W. and D. J. Wald (2001). "Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions 1. Strong motions". In: *Journal of Geophysical Research: Solid Earth* 106.B5, pp. 8745–8766. DOI: 10.1029/2000jb900436.

Grünthal, G. (1998). "European Macroseismic Scale 1998". In: *European Center of Geodynamics and Seismology* 15, pp. 1–100. DOI: 10.2312/EMS-98.full.en.

Guérin-Marthe, S., P. Gehl, C. Negulescu, S. Auclair, and R. Fayjaloun (2021). "Rapid earthquake response: The state-of-the art and recommendations with a focus on European systems". In: *International Journal of Disaster Risk Reduction* 52.July 2021. DOI: 10.1016/j.ijdrr.2020.101958.

Gutenberg, B. and C. F. Richter (1944). "Frequency of earthquakes in California*". In: *Bulletin of the Seismological Society of America* 34.4, pp. 185–188. DOI: 10.1785/bssa0340040185.

Hall, J. F., T. H. Heaton, M. W. Halling, and D. J. Wald (1995). "Near-Source Ground Motion and its Effects on Flexible Buildings". In: *Earthquake Spectra* 11.4, pp. 569–605. DOI: 10.1193/1.1585828.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant (2020). "Array programming with NumPy". In: *Nature* 585.7825, pp. 357–362. DOI: 10.1038/s41586-020-2649-2. arXiv: 2006.10256.

Holland, A. (2003). "Earthquake data recorded by the MEMS accelerometer: Field testing in Idaho". In: *Seismological Research Letters* 74.1, pp. 20–26. DOI: 10.1785/gssrl.74.1.20.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science and Engineering* 9.3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

Hutchinson, J., B. A. Bradley, R. Lee, L. Wotherspoon, M. Dupuis, C. Schill, J. Motha, A. Kaiser, and E. Manea (2022). *2021 New Zealand Ground-Motion Database*. Tech. rep. GNS Science Report 2021/56. Lower Hutt: GNS Science, pp. 1–52.

Idrissa, I. M. (2014). "An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes". In: *Earthquake Spectra* 30.3, pp. 1155–1177. DOI: 10.1193/070613EQS195M.

Iervolino, I. and C. Cornell (2008). "Probability of occurrence of velocity pulses in near-source ground motions". In: *Bulletin of the Seismological Society of America* 98.5, pp. 2262–2277. DOI: 10.1785/0120080033.

Inkscape Project (2020). *Inkscape*. URL: https://inkscape.org.

Intel®NeuralCompressor (2023). *Intel® Neural Compressor*. URL: https://github.com/intel/neural-compressor.

IRIS (2023). *IRIS DMC Data Statistics*. URL: https://ds.iris.edu/data/distribution/.

Jaiswal, K., D. J. Wald, and K. Porter (2010). "A global building inventory for earthquake loss estimation and risk management". In: *Earthquake Spectra* 26.3, pp. 731–748. DOI: 10.1193/1.3450316.

Jayaram, N. and J. W. Baker (Dec. 2009). "Correlation model for spatially distributed ground-motion intensities". In: *Earthquake Engineering and Structural Dynamics* 38.15, pp. 1687–1708. DOI: 10.1002/eqe.922. arXiv: 1403.5481.

Jordan, T. H., Y. T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka, and J. Zschau (2011). "Operational earthquake forecasting: State of knowledge and guidelines for utilization". In: *Annals of Geophysics* 54.4, pp. 319–391. DOI: 10.4401/ag-5350.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis (2021). "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596.7873, pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

Kingma, D. P. and J. L. Ba (2015). "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15. arXiv: 1412.6980.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2019). "Machine learning in seismology: Turning data into insights". In: *Seismological Research Letters* 90.1, pp. 3–14. DOI: 10.1785/0220180259.

Kotha, S. R., D. Bindi, and F. Cotton (2016). "Partially non-ergodic region specific GMPE for Europe and Middle-East". In: *Bulletin of Earthquake Engineering* 14.4, pp. 1245–1263. DOI: 10.1007/s10518-016-9875-x.

Kotha, S. R., G. Weatherill, D. Bindi, and F. Cotton (2020). "A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe". In: *Bulletin of Earthquake Engineering* 18.9, pp. 4091–4125. DOI: 10.1007/s10518-020-00869-1.

Kramer, S. L. (1996a). "Code-Based Development". In: *Geotechnical earthquake engineering*. New Jersey: Prentice Hall, pp. 328–329.

Kramer, S. L. (1996b). "Development of Predictive Relationships". In: *Geotechnical earthquake engineering*. New Jersey: Prentice Hall, pp. 86–88.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2017). "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6, pp. 84–90. DOI: 10.1145/3065386.

Kuehn, N. M. and N. A. Abrahamson (2020). "Spatial correlations of ground motion for nonergodic seismic hazard analysis". In: *Earthquake Engineering and Structural Dynamics* 49.1, pp. 4–23. DOI: 10.1002/eqe.3221.

Kuo, T.-C. (1957). "On the Shensi earthquake of january 23, 1556". In: *Chinese Journal of Geophysics (in Chinese)* 6.01, pp. 59–68.

L'Heureux, A., K. Grolinger, H. F. Elyamany, and M. A. Capretz (2017). "Machine Learning with Big Data: Challenges and Approaches". In: *IEEE Access* 5, pp. 7776–7797. DOI: 10.1109/ACCESS.2017.2696365.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). "A nonergodic ground-motion model for California with spatially varying coefficients". In: *Bulletin of the Seismological Society of America* 106.6, pp. 2574–2583. DOI: 10.1785/0120160118.

Lay, T. and T. C. Wallace (1995). "Historical Development of Global Seismology". In: *Modern Global Seismology*. Vol. 58, p. 7.

LeCun, Y., B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. D. Jackel (1989). "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation*. Vol. 1. 4, pp. 541–551. DOI: 10.1162/neco.1989.1.4.541.

Lecun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444. DOI: 10.1038/nature14539.

Legge, J. (1865). *The Chinese Classics: Vol. 3.1: The first parts of the Shoo-king, or the books of T'ang; the books of Yu; the books of Hea; the books of Shang; and the Prolegomena*. Bd. 3. at the author's, p. 125. URL: https://books.google.de/books?id=jtKzhaCspc8C.

Levie, R., C. Yapar, G. Kutyniok, and G. Caire (2020). "Pathloss Prediction using Deep Learning with Applications to Cellular Optimization and Efficient D2D Link Scheduling". In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* 2020-May, pp. 8678–8682. DOI: 10.1109/ICASSP40776.2020.9053347.

Lilienkamp, H., S. von Specht, G. Weatherill, G. Caire, and F. Cotton (2022). "Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach". In: *Bulletin of the Seismological Society of America* 112.3, pp. 1565–1582. DOI: 10.1785/0120220008.

Lindsey, N. J. and E. R. Martin (2021). "Fiber-optic seismology". In: *Annual Review of Earth and Planetary Sciences* 49, pp. 309–336. DOI: 10.1146/annurev-earth-072420-065213.

Loth, C. and J. W. Baker (2013). "A spatial cross-correlation model of spectral accelerations at multiple periods". In: *Earthquake Engineering and Structural Dynamics* 42.3, pp. 397–417. DOI: 10.1002/eqe.2212. arXiv: 1403.5481.

Mai, P. M., T. Aspiotis, T. A. Aquib, E. V. Cano, D. Castro-Cruz, A. Espindola-Carmona, B. Li, X. Li, J. Liu, R. Matrau, A. Nobile, K. H. Palgunadi, M. Ribot, L. Parisi, C. Suhendi, Y. Tang, B. Yalcin, U. Avşar, Y. Klinger, and S. Jónsson (2023). "The Destructive Earthquake Doublet of 6 February 2023 in South-Central Türkiye and Northwestern Syria: Initial Observations and Analyses". In: *The Seismic Record* 3.2, pp. 105–115. DOI: 10.1785/0320230007.

Mai, P. M., P. Spudich, and J. Boatwright (2005). "Hypocenter locations in finite-source rupture models". In: *Bulletin of the Seismological Society of America* 95.3, pp. 965–980. DOI: 10.1785/0120040111.

Markhvida, M., L. Ceferino, and J. W. Baker (2018). "Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics". In: *Earthquake Engineering and Structural Dynamics* 47.5, pp. 1107–1123. DOI: 10.1002/eqe.3007.

Marzocchi, W. and A. M. Lombardi (2009). "Real-time forecasting following a damaging earthquake". In: *Geophysical Research Letters* 36.21, pp. 1–5. DOI: 10.1029/2009GL040233.

Mavroeidis, G. P., G. Dong, and A. S. Papageorgiou (2004). "Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems". In: *Earthquake Engineering and Structural Dynamics* 33.9, pp. 1023–1049. DOI: 10.1002/eqe.391.

Mazzoni, S., L. Al Atik, N. Gregor, and Y. Bozorgnia (2023). *Directivity-Based Intensity-Measure Interactive Maps*. DOI: 10.34948/N3101P. URL: https://doi.org/10.34948/N3101P.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The Bulletin of Mathematical Biophysics* 5.4, pp. 115–133. DOI: 10.1007/BF02478259.

McGuire, R. K. (1976). *FORTRAN computer program for seismic risk analysis*. Tech. rep. 67. United States Geological Survey, p. 90. URL: http://pubs.er.usgs.gov/publication/ofr7667.

McVerry, G. (2003). "From hazard maps to code spectra for New Zealand". In: *Proceedings of the Pacific Conference on Earthquake Engineering* 1992.034, pp. 1–9.

McVerry, G., J. X. Zhao, N. A. Abrahamson, and P. G. Somerville (2006). "New Zealand acceleration response spectrum attenuation relations for crustal and subduction zone earthquakes". In: *Bulletin of the New Zealand Society for Earthquake Engineering* 39.1, pp. 1–58. DOI: 10.5459/bnzsee.39.1.1-58.

Melgar, D. and G. P. Hayes (2019). "The correlation lengths and hypocentral positions of great earthquakes". In: *Bulletin of the Seismological Society of America* 109.6, pp. 2582–2593. DOI: 10.1785/0120190164.

Miller, A. C. and T. R. Rice (1983). "Discrete Approximations of Probability Distributions." In: *Management Science* 29.3, pp. 352–362. DOI: 10.1287/mnsc.29.3.352.

Milner, K. R., B. E. Shaw, and E. Field (2022). "Enumerating Plausible Multifault Ruptures in Complex Fault Systems with Physical Constraints". In: *Bulletin of the Seismological Society of America* 112.4, pp. 1806–1824. DOI: 10.1785/0120210322.

Mousavi, S. M. and G. C. Beroza (2022). "Deep-learning seismology". In: *Science* 377.6607. DOI: 10.1126/science.abm4470.

Nakano, K. and H. Kawase (2021). "Spatial properties of the site amplifications evaluated by generalized inversion technique with Fourier spectra and response spectra". In: *The 6th IASPEI / IAEE International Symposium: Effects of Surface Geology on Seismic Motion, GS5-P03*.

National Geophysical Data Center / World Data Service (2023). *NCEI/WDS Global Significant Earthquake Database*. DOI: 10.7289/V5TD9V7K. URL: https://www.ngdc.noaa.gov/hazel/view/hazards/earthquake/search.

National Research Institute for Earth Science and Disaster Resilience (2019). *NIED K-NET, KiK-net*. DOI: 10.17598/NIED.0004. URL: https://www.doi.org/10.17598/NIED.0004.

Nievas, C. I., J. J. Bommer, H. Crowley, J. van Elk, M. Ntinalexis, and M. Sangirardi (2020). "A database of damaging small-to-medium magnitude earthquakes". In: *Journal of Seismology* 24.2, pp. 263–292. DOI: 10.1007/s10950-019-09897-0.

NZS (2004). *Structural Design Actions Part 5: Earthquake Actions - Commentary*. Technical Report 1170.5 Supp 1. Wellington: Standards New Zealand. URL: papers3://publication/uuid/45FCE1E1-E438-405C-AA05-51DAF4A7C05A.

Ogata, Y. (1998). "Space-Time Point-Pocess Models For Earthquake Occurrences". In: *Ann. Inst. Statist. Math.* 50.2, pp. 379–402.

Okabe, S. (1956). "Time variation of the atmospheric radon content near the ground surface with relation to some geophysical phenomena." In: *Memoirs of the College of Science* 28.2, pp. 99–115.

Pagani, M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, M. Simionato, and D. Vigano (2014). "Openquake engine: An open hazard (and risk) software for the global earthquake model". In: *Seismological Research Letters* 85.3, pp. 692–702. DOI: 10.1785/0220130087.

Page, M. T., E. Field, K. R. Milner, and P. M. Powers (2014). "The UCERF3 grand inversion: Solving for the long-term rate of ruptures in a fault system". In: *Bulletin of the Seismological Society of America* 104.3, pp. 1181–1204. DOI: 10.1785/0120130180.

Paolucci, R., C. Smerzini, and M. Vanini (2021). "Bb-speedset: A validated dataset of broadband near-source earthquake ground motions from 3d physics-based numerical simulations". In: *Bulletin of the Seismological Society of America* 111.5, pp. 2527–2545. DOI: 10.1785/0120210089.

Paté-Cornell, M. E. (1994). "Quantitative safety goals for risk management of industrial facilities". In: *Structural Safety* 13.3, pp. 145–157. DOI: 10.1016/0167-4730(94)90023-X.

Perol, T., M. Gharbi, and M. Denolle (2018). "Convolutional neural network for earthquake detection and location". In: *Science Advances* 4.2, pp. 2–10. DOI: 10.1126/sciadv.1700578. arXiv: 1702.02073.

Poggi, V., C. Scaini, L. Moratto, G. Peressi, P. Comelli, P. L. Bragato, and S. Parolai (2021). "Rapid damage scenario assessment for earthquake emergency management". In: *Seismological Research Letters* 92.4, pp. 2513–2530. DOI: 10.1785/0220200245.

Pozos-Estrada, A., R. Gómez, and H. P. Hong (2014). "Use of neural network to predict the peak ground accelerations and pseudo spectral accelerations for Mexican Inslab and Interplate Earthquakes". In: *Geofisica Internacional* 53.1, pp. 39–57. DOI: 10.1016/S0016-7169(14)71489-8.

QGIS Development Team (2021). *QGIS Geographical Information System*. URL: https://www.qgis.org.

Quitoriano, V. and D. J. Wald (2020). "USGS "Did You Feel It?"—Science and Lessons From 20 Years of Citizen Science-Based Macroseismology". In: *Frontiers in Earth Science* 8.May, pp. 1–13. DOI: 10.3389/feart.2020.00120.

Reid, H. (1910). *The California Earthquake of April 18, 1906 . State Earthquake Investigation Committee A. C. Lawson*. Tech. rep. 3. Washington, D.C.: Carnegie Institution of Washington, pp. 288–290. DOI: 10.1086/621732.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28. arXiv: 1505.04597.

Ross, Z. E., M. A. Meier, E. Hauksson, and T. H. Heaton (2018). "Generalized seismic phase detection with deep learning". In: *Bulletin of the Seismological Society of America* 108.5, pp. 2894–2901. DOI: 10.1785/0120180080. arXiv: 1805.01075.

Rowshandel, B. (2013). *Rowshandel Model*. Tech. rep. PEER Report No. 2013/09. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 47–67.

Rowshandel, B. (2018). "Capturing and PSHA Implementation of Spatial Variability of Near-Source Ground Motion Hazard". In: *Geotechnical Earthquake Engineering and Soil Dynamics V*, pp. 56–63. DOI: 10.1061/9780784481462.006.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). "Learning Representations by Back-Propagating Errors". In: *Nature* 323.9, pp. 533–536. DOI: 10.7551/mitpress/1888.003.0013.

Sagiroglu, S. and D. Sinanc (2013). "Big data: A review". In: *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, pp. 42–47. DOI: 10.1109/CTS.2013.6567202.

Savage, J. (1982). "Current Research in Earthquake Prediction. Vol. 1". In: *Tectonophysics* 88.1-2, pp. 196–197. DOI: 10.1016/0040-1951(82)90214-1.

Seebeck, H., R. Van Dissen, N. Litchfield, P. M. Barnes, A. Nicol, R. Langridge, D. J. Barrell, P. Villamor, S. Ellis, M. Rattenbury, S. Bannister, M. Gerstenberger, F. Ghisetti, R. Sutherland, H. Hirschberg, J. Fraser, S. D. Nodder, M. Stirling, J. Humphrey, K. J. Bland, A. Howell, J. Mountjoy, V. Moon, T. Stahl, F. Spinardi, D. Townsend, K. Clark, I. Hamling, S. Cox, W. de Lange, P. Wopereis, M. Johnston, R. Morgenstern, G. Coffey, J. D. Eccles, T. Little, B. Fry, J. Griffin, J. Townend, N. Mortimer, S. Alcaraz, C. Massiot, J. V. Rowland, J. Muirhead, P. Upton, and J. Lee (2023). "The New Zealand Community Fault Model–version 1.0: an improved geological foundation for seismic hazard modelling". In: *New Zealand Journal of Geology and Geophysics*, pp. 1–21. DOI: 10.1080/00288306.2023.2181362.

Seeger, M. (2004). *Gaussian processes for machine learning*. Vol. 14. Adaptive computation and machine learning 2. MIT Press, pp. 69–106. DOI: 10.1142/S0129065704001899.

Senna, S., T. Maeda, Y. Inagaki, H. Suzuki, H. Matsuyama, and H. Fujiwara (2013). "Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband

strong motion evaluation". In: *J. Disast. Res.* 8.5, pp. 889–903. DOI: 10.20965/jdr.2013.p0889.

Senna, S., A. Wakai, H. Fujiwara, K. Jin, A. Yatagai, H. Suzuki, Y. Inagaki, and H. Matsuyama (2019). "Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband strong motion evaluation in japan". In: *Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions*, pp. 4931–4938. DOI: 10.20965/jdr.2013.p0889.

Sgobba, S., G. Lanzano, and F. Pacor (2021). "Empirical nonergodic shaking scenarios based on spatial correlation models: An application to central Italy". In: *Earthquake Engineering and Structural Dynamics* 50.1, pp. 60–80. DOI: 10.1002/eqe.3362.

Shahi, S. and J. W. Baker (2013). *A Probabalistic Framework to Include the Effects of Near-fault Directivity in Seismic Hazard Assessment*. Tech. rep. PEER Report No. 2013/15. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 1–217.

Shahi, S. and J. W. Baker (2014). "An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions". In: *Bulletin of the Seismological Society of America* 104.5, pp. 2456–2466. DOI: 10.1785/0120130191.

Shoaf, K. I., H. R. Sareen, L. H. Nguyen, and L. B. Bourque (1998). "Injuries as a result of California earthquakes in the past decade". In: *Disasters* 22.3, pp. 218–235. DOI: 10.1111/1467-7717.00088.

Somerville, P. G., N. F. Smith, R. W. Graves, and N. A. Abrahamson (1997). "Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity". In: *Seismological Research Letters* 68.1, pp. 199–222. DOI: 10.1785/gssrl.68.1.199.

Spudich, P., J. Bayless, J. W. Baker, B. S. Chiou, B. Rowshandel, S. Shahi, and P. G. Somerville (2013). *Final Report of the NGA-West2 Directivity Working Group*. Tech. rep. PEER Report No. 2013/09. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 1–162.

Spudich, P. and B. S. Chiou (2008). "Directivity in NGA earthquake ground motions: Analysis using isochrone theory". In: *Earthquake Spectra* 24.1, pp. 279–298. DOI: 10.1193/1.2928225.

Spudich, P. and B. S. Chiou (2013). *Spudich and Chiou Model*. Tech. rep. PEER Report No. 2013/09. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), pp. 77–82.

Spudich, P. and B. S. Chiou (2015). *Strike-Parallel and Strike-Normal Coordinate System Around Geometrically Complicated Rupture Traces— Use by NGA-West2 and Further Improvements*. Tech. rep. Reston, Virginia: United States Geological Survey, pp. 1–28.

Spudich, P., B. Rowshandel, S. Shahi, J. W. Baker, and B. S. Chiou (2014). "Comparison of NGA-West2 directivity models". In: *Earthquake Spectra* 30.3, pp. 1199–1221. DOI: 10.1193/080313EQS222M.

Stafford, P. J. (2014). "Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models". In: *Bulletin of the Seismological Society of America* 104.2, pp. 702–719. DOI: 10.1785/0120130145.

Stafford, P. J. (2022). *A model for the distribution of response spectral ordinates from New Zealand crustal earthquakes based upon adjustments to the Chiou and Youngs (2014) response spectral model*. Tech. rep. GNS Science Report 2022/15. Lower Hutt: GNS Science, pp. 1–106.

Stafford, P. J., D. M. Boore, R. R. Youngs, and J. J. Bommer (2022). "Host-region parameters for an adjustable model for crustal earthquakes to facilitate the implementation of the backbone approach to building ground-motion logic trees in probabilistic seismic hazard analysis". In: *Earthquake Spectra* 38.2, pp. 917–949. DOI: 10.1177/87552930211063221.

Stirling, M., G. McVerry, M. Gerstenberger, N. Litchfield, R. Van Dissen, K. Berryman, P. Barnes, L. Wallace, P. Villamor, R. Langridge, G. Lamarche, S. Nodder, M. Reyners, B. A. Bradley, D. Rhoades, W. Smith, A. Nicol, J. Pettinga, K. Clark, and K. Jacobs (2012). "National seismic hazard model for New Zealand: 2010 update". In: *Bulletin of the Seismological Society of America* 102.4, pp. 1514–1542. DOI: 10.1785/0120110170.

Strasser, F. O., N. A. Abrahamson, and J. J. Bommer (2009). "Sigma: Issues, insights, and challenges". In: *Seismological Research Letters* 80.1, pp. 40–56. DOI: 10.1785/gssrl.80.1.40.

The Headquarters for Earthquake Research promotion (2021). "Modeling concept of subsurface structures from seismic bedrock to ground surface (in Japanese)". In: URL: https://www.jishin.go.jp/evaluation/strong%7B%5C_%7Dmotion/underground%7B%5C_%7Dmodel/integration%7B%5C_%7Dmodel%7B%5C_%7Dkanto%7B%5C_%7D2021/.

Tothong, P., C. Cornell, and J. W. Baker (2007). "Explicit directivity-pulse inclusion in probabilistic seismic hazard analysis". In: *Earthquake Spectra* 23.4, pp. 867–891. DOI: 10.1193/1.2790487.

Trifunac, M. D. (1979). "Preliminary empirical model for scaling fourier amplitude spectra of strong ground acceleration in terms of modified mercalli intensity and recording site conditions". In: *Earthquake Engineering & Structural Dynamics* 7.1, pp. 63–74. DOI: 10.1002/eqe.4290070106.

Trugman, D. T. and P. M. Shearer (2018). "Strong correlation between stress drop and peak ground acceleration for recent M 1–4 earthquakes in the San Francisco bay area". In: *Bulletin of the Seismological Society of America* 108.2, pp. 929–945. DOI: 10.1785/0120170245.

U.S. Geological Survey (2023a). *M 7.5 - Elbistan earthquake, Kahramanmaras earthquake sequence, USGS Earthquake Hazards Program*. URL: https://earthquake.usgs.gov/earthquakes/eventpage/us6000jlqa (visited on Aug. 22, 2023).

U.S. Geological Survey (2023b). *M 7.5 - Elbistan earthquake, Kahramanmaras earthquake sequence, USGS Earthquake Hazards Program*. URL: https://earthquake.usgs.gov/earthquakes/eventpage/us6000jlqa (visited on Aug. 22, 2023).

Van Dissen, R., H. Seebeck, L. Wallace, C. Rollins, M. Gerstenberger, A. Howell, C. DiCaprio, and C. Williams (2022). *New Zealand National Seismic Hazard Model 2022: Geologic and Subduction Interface Deformation Models*. Tech. rep. GNS Science Report 2022/31. Lower Hutt: GNS Science, pp. 1–30.

Virtanen, P. et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17.3, pp. 261–272. DOI: 10.1038/s41592-019-0686-2. arXiv: 1907.10121.

Wakamatsu, K. and M. Matsuoka (2013). "Nationwide 7.5-arc-second Japan engineering geomorphologic classification map and Vs30 zoning". In: *Journal of Disaster Research* 8.5, pp. 904–911. DOI: 10.20965/jdr.2013.p0904.

Wakamatsu, K. and M. Matsuoka (2020). "Update of the Nationwide 7.5-Arc-Second Japan Engineering Geomorphologic Classification Map and Vs30 Zoning". In: *Bulletin of JAEE* 40, pp. 24–27.

Wald, D. J., K. Jaiswal, K. D. Marano, D. B. Bausch, and M. G. Hearne (2010). *PAGER–Rapid Assessment of an Earthquake's Impact: U.S. Geological Survey Fact Sheet*. 2010-3036 2010-3036. U.S. Geological Survey, pp. 1–4. DOI: 10.3133/fs20103036. URL: https://pubs.usgs.gov/fs/2010/3036/pdf/FS10-3036.pdf.

Wald, D. J., V. Quitoriano, L. A. Dengler, and J. W. Dewey (1999). "Utilization of the Internet for Rapid Community Intensity Maps". In: *Seismological Research Letters* 70.6, pp. 680–697. DOI: 10.1785/gssrl.70.6.680.

Wald, D. J. and T. H. Heaton (1994). "Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake". In: *Bulletin - Seismological Society of America* 84.3, pp. 668–691. DOI: 10.1785/bssa0840030668.

Wald, D. J., B. Worden, V. Quitoriano, and K. Pankow (2005). *ShakeMap ® Manual*. Tech. rep. Version 1.0, pp. 1–156.

Waldeland, A. U., A. C. Jensen, L.-J. Gelius, and A. H. S. Solberg (2018). "Convolutional neural networks for automated seismic interpretation". In: *The Leading Edge* 37.7, pp. 529–537. DOI: 10.1190/tle37070529.1. URL: https://doi.org/10.1190/tle37070529.1.

Watson-Lamprey, J. A. (2018). *Capturing directivity effects in the mean and aleatory variability of the NGA-West2 ground-motion prediction equations*. Tech. rep. 2018-04. Berkeley, CA: Pacific Earthquake Engineering Research Center (PEER), p. 112.

Weatherill, G. (2022). *Impact of Directivity on Probabilistic Seismic Hazard Calculations in New Zealand*. Tech. rep. GNS Science Report 2022/01. Lower Hutt: GNS Science, pp. 1–85.

Woith, H., G. M. Petersen, S. Hainzl, and T. Dahm (2018). "Review: Can animals predict earthquakes?" In: *Bulletin of the Seismological Society of America* 108.3, pp. 1031–1045. DOI: 10.1785/0120170313.

Wood, H. O. and F. Neumann (1931). "Modified Mercalli intensity scale of 1931". In: *Bulletin of the Seismological Society of America* 21.4, pp. 277–283. DOI: 10.1785/bssa0210040277.

Xie, J. (2019). "Strong-motion directionality and evidence of rupture directivity effects during the chi-chi Mw 7.6 earthquake". In: *Bulletin of the Seismological Society of America* 109.6, pp. 2367–2383. DOI: 10.1785/0120190087.

Xiong, W., X. Ji, Y. Ma, Y. Wang, N. M. Benhassan, M. N. Ali, and Y. Luo (2018). "Seismic fault detection with convolutional neural network". In: *Geophysics* 83.5, O97–O103. DOI: 10.1190/geo2017-0666.1.

Yao, W., Z. Zeng, C. Lian, and H. Tang (2018). "Pixel-wise regression using U-Net and its application on pansharpening". In: *Neurocomputing* 312, pp. 364–371. DOI: 10.1016/j.neucom.2018.05.103.

Yen, M. H., S. von Specht, Y. Y. Lin, F. Cotton, and K. F. Ma (2022). "Within-and Between-Event Variabilities of Strong-Velocity Pulses of Moderate Earthquakes within Dense Seismic

Arrays". In: *Bulletin of the Seismological Society of America* 112.1, pp. 361–380. DOI: 10.1785/0120200376.

Yenier, E. and G. Atkinson (2015a). "An equivalent point-source model for stochastic simulation of earthquake ground motions in California". In: *Bulletin of the Seismological Society of America* 105.3, pp. 1435–1455. DOI: 10.1785/0120140254.

Yenier, E. and G. Atkinson (2015b). "Regionally adjustable generic ground-motion prediction equation based on equivalent point-source simulations: Application to central and eastern North America". In: *Bulletin of the Seismological Society of America* 105.4, pp. 1989–2009. DOI: 10.1785/0120140332.

Zhao, J. X., J. Hu, F. Jiang, J. Zhou, Y. Zhang, X. An, M. Lu, and D. A. Rhoades (2015). "Nonlinear site models derived from 1D analyses for ground-motion prediction equations using site class as the site parameter". In: *Bulletin of the Seismological Society of America* 105.4, pp. 2010–2022. DOI: 10.1785/0120150019.

Zhao, J. X., F. Jiang, P. Shi, H. Xing, H. Huang, R. Hou, Y. Zhang, P. Yu, X. Lan, D. A. Rhoades, P. G. Somerville, K. Irikura, and Y. Fukushima (2016a). "Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions". In: *Bulletin of the Seismological Society of America* 106.4, pp. 1535–1551. DOI: 10.1785/0120150056.

Zhao, J. X., X. Liang, F. Jiang, H. Xing, M. Zhu, R. Hou, Y. Zhang, X. Lan, D. A. Rhoades, K. Irikura, Y. Fukushima, and P. G. Somerville (2016b). "Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions". In: *Bulletin of the Seismological Society of America* 106.4, pp. 1518–1534. DOI: 10.1785/0120150034.

Zhao, J. X., S. Zhou, J. Zhou, C. Zhao, H. Zhang, Y. Zhang, P. Gao, X. Lan, D. Rhoades, Y. Fukushima, P. G. Somerville, and K. Irikura (2016c). "Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions". In: *Bulletin of the Seismological Society of America* 106.4, pp. 1552–1569. DOI: 10.1785/0120150063.

Zhao, L., P. Chen, and T. H. Jordan (2006). "Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies". In: *Bulletin of the Seismological Society of America* 96.5, pp. 1753–1763. DOI: 10.1785/0120050253.

Zhu, W. and G. C. Beroza (2019). "PhaseNet: A deep-neural-network-based seismic arrival-time picking method". In: *Geophysical Journal International* 216.1, pp. 261–273. DOI: 10.1093/gji/ggy423. arXiv: 1803.03211.