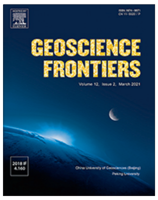
Geoscience Frontiers 15 (2024) 101839

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf



Research Paper

Interpretability and spatial efficacy of a deep-learning-based on-site early warning framework using explainable artificial intelligence and geographically weighted random forests

Jawad Fayaz [a,b,*], Carmine Galasso [b]

[a] Department of Computer Science, University of Exeter, UK
[b] Department of Civil, Environmental, and Geomatic Engineering, University College London (UCL), London, UK

ARTICLE INFO

Article history:
Received 5 July 2023
Revised 23 February 2024
Accepted 9 April 2024
Available online 12 April 2024
Handling Editor: Masaki Yoshida

Keywords:
Earthquake early warning systems
Spatial regression
Neural networks
Japanese subduction
Explainable artificial intelligence

ABSTRACT

Earthquakes pose significant risks globally, necessitating effective seismic risk mitigation strategies like earthquake early warning (EEW) systems. However, developing and optimizing such systems requires thoroughly understanding their internal procedures and coverage limitations. This study examines a deep-learning-based on-site EEW framework known as ROSERS (Real-time On-Site Estimation of Response Spectra) proposed by the authors, which constructs response spectra from early recorded ground motion waveforms at a target site. This study has three primary goals: (1) evaluating the effectiveness and applicability of ROSERS to subduction seismic sources; (2) providing a detailed interpretation of the trained deep neural network (DNN) and surrogate latent variables (LVs) implemented in ROSERS; and (3) analyzing the spatial efficacy of the framework to assess the coverage area of on-site EEW stations. ROSERS is retrained and tested on a dataset of around 11,000 unprocessed Japanese subduction ground motions. Goodness-of-fit testing shows that the ROSERS framework achieves good performance on this database, especially given the peculiarities of the subduction seismic environment. The trained DNN and LVs are then interpreted using game theory-based Shapley additive explanations to establish cause-effect relationships. Finally, the study explores the coverage area of ROSERS by training a novel spatial regression model that estimates the LVs using geographically weighted random forest and determining the radius of similarity. The results indicate that on-site predictions can be considered reliable within a 2–9 km radius, varying based on the magnitude and distance from the earthquake source. This information can assist end-users in strategically placing sensors, minimizing blind spots, and reducing errors from regional extrapolation.

1. Introduction

Earthquakes are one of the most devastating natural hazards that can hit without long warning times, causing loss of life, significant damage, and disruptions to physical infrastructure and social networks. Given the current knowledge, it is impossible to deterministically estimate when and where future earthquakes will occur and their resulting severity. Nevertheless, recent technological advancements have enabled the development and implementation of earthquake early warning (EEW) systems (e.g., Cremen and Galasso, 2020) across the world. These systems can provide crucial seconds or even minutes of warning time (depending on the source-to-site distance) to take protective actions (e.g., Hsu et al., 2013; Bhardwaj et al., 2016; Caruso et al., 2017; Iaccarino et al., 2020; Münchmeyer et al., 2021; Fayaz and Galasso, 2022; Galasso et al., 2023) before strong ground shaking reaches target individuals and critical infrastructure at risk.

EEW relies on a network of seismic sensors that detect the initial seismic waves produced by an earthquake rupture and, in some configurations (i.e., regional EEW; see below), transmit data in real time to a centralized processing center. The processing center then analyzes the data to estimate the earthquake's location, magnitude, and/or expected shaking intensity (and related uncertainties). Using this information, the system can quickly send out alerts via various communication channels such as mobile apps, text messages, or loudspeakers to warn people in affected areas, allowing them to take protective actions (such as "drop, cover, and hold on" – DCHO – or move to safer locations within a

* Corresponding author at: Department of Computer Science, University of Exeter, UK.
E-mail address: j.fayaz@exeter.ac.uk (J. Fayaz).

https://doi.org/10.1016/j.gsf.2024.101839
1674-9871/© 2024 China University of Geosciences (Beijing) and Peking University. Published by Elsevier B.V. on behalf of China University of Geosciences (Beijing).

building, stopping machinery, or shutting down critical systems) before the strong ground shaking reaches them (e.g., McBride et al., 2020).

The concept of EEW has been around for several decades. Yet, recent technological advancements have significantly improved earthquake detection and real-time characterization as well as warning speed and accuracy. EEW can be broadly allocated to three sub-classes: (1) regional; (2) on-site; and (3) hybrid. Regional EEW systems consist of a network of seismic stations located within a region of high seismicity and use the early-recorded seismic waves at some stations to estimate ground shaking (directly or using predicted source magnitude/location) at the target sites, within the region, before the arrival of damaging waves. Largely, a regional EEW system uses the early-recorded waves at a number of seismic sensors to estimate the source parameters, which are then fed into pre-calibrated ground-motion models (GMMs; e.g., Campbell and Bozorgnia, 2013, 2019; Fayaz et al., 2021, 2023) to calculate intensity measures (IMs) like peak ground acceleration (PGA) (e.g., Lin et al., 2012; Wu et al., 2013; Bhardwaj et al., 2016; Caruso et al., 2017; Cremen and Galasso, 2020; Münchmeyer et al., 2021) and several others. On the other hand, an on-site EEW system is a standalone system that uses the early-recorded waves to directly estimate ground shaking at or near the recording station (typically coincident with the target site). Finally, hybrid EEW systems combine the advantages of both regional and on-site sensors to provide a more comprehensive approach to earthquake detection, characterization, and warning. These systems use regional data of recorded ground motions to quickly estimate the location and magnitude of an earthquake and on-site data to provide more detailed information on the expected ground shaking at specific locations.

One of the critical challenges in developing effective EEW systems is the need for swift data processing, communication, and reliable algorithms.Due to the fast traveling speed of seismic waves (6–10 km/s) (e.g., Kramer, 1996), there is only a short time window between the initial detection of an earthquake rupture and the arrival of damaging ground shaking at nearby locations. As a result, EEW systems need to process and disseminate warning information accurately in seconds, often in the face of incomplete or highly uncertain data. Against this backdrop, data-driven and machine-learning (ML)-based EEW approaches have emerged as promising alternatives to enhance the accuracy and effectiveness of the process. An early study by Hsu et al. (2013) used seismic *p*-waves and the vertical ground motion components to predict PGA through support vector machines (SVM) regression. Similarly, Caruso et al. (2017) used the peak displacement (P_d) and dominant period (τ_c) of the vertical ground-motion component to obtain the event's magnitude (M) and rupture distance through multiple regression models. Furthermore, Münchmeyer et al. (2021) used initial ground-motion time series from multiple stations to estimate PGA value at target locations through Transformer neural networks. Jozinović et al. (2020) introduced a method using convolutional neural networks (CNNs) with raw waveform data for rapid earthquake ground shaking intensity prediction, emphasizing the model's prediction capability without prior earthquake location and magnitude knowledge. Jozinović et al. (2022) then explored transfer learning to enhance neural network predictions of earthquake ground shaking in regions with insufficient data, demonstrating improved prediction accuracy by leveraging pre-trained models on larger or different datasets. Furthermore, Bloemheuvel et al. (2023) proposed a novel graph neural network architecture designed for multivariate time series regression, specifically tested on seismic datasets to predict maximum intensity measurements of ground shaking at seismic stations. Collectively, these studies represent significant strides in utilizing advanced ML models to enhance earthquake prediction capabilities.

Recently, Fayaz and Galasso (2022) proposed a highly accurate deep learning (DL) on-site EEW framework based on a variational autoencoder (VAE) and deep neural networks (DNNs) to estimate the spectral acceleration response spectrum ($S_a(T)$) at a target location. All these ML/DL-based methods, while providing efficient tools for EEW, come with certain limitations. One of the key challenges is the lack of interpretability and explainability of ML models. Due to their complex algorithms and the large number of parameters involved, it can be challenging to understand why a particular prediction or alert is made. Additionally, the ML models' "black box" nature can make incorporating domain expertise and expert knowledge into the system challenging. This lack of transparency hinders the ability to validate and verify the reasoning behind the model's decisions, making gaining trust and acceptance from users and stakeholders difficult.

Furthermore, both on-site and regional EEW systems/approaches present certain limitations. For example, on-site EEW has limited coverage since it relies on the sensors installed at specific locations. This means that the coverage area is limited to the vicinity of these sensors, leaving areas outside such range without EEW capabilities. Furthermore, on-site systems are more susceptible to false alarms triggered by non-earthquake events, such as construction activities or vehicle-induced/traffic vibrations, that can produce ground motion. These false alarms can lead to reduced complacency and trust in the system. On the other hand, there are some challenges related to regional EEW. Specifically, regional EEW systems rely on seismic sensors spread across a wide area. However, for earthquakes occurring close to the monitoring network, the detection time may not be significantly reduced, limiting the lead time (i.e., warning time) for issuing alerts to nearby areas (e.g., Tajima and Hayashida, 2018). In addition, the transmission and processing of data from multiple sensors to a central processing facility can introduce communication delays. This can impact the speed at which alerts are issued, reducing the potential lead time for warning recipients. Therefore, it is paramount to understand the extent of coverage required for the effective operation of EEW (mainly on-site systems) and ensure their real-time operational efficiency.

This study presents a detailed investigation of the real-time on-site EEW framework known as ROSERS (Real-time On-Site Estimation of Response Spectra) proposed by Fayaz and Galasso (2022). ROSERS utilized DL techniques and was initially trained and evaluated on ~7000 ground motion records obtained from the crustal Next-Generation Attenuation West 2 (NGA-West2) project (Ancheta et al., 2014). In particular, the premise of the ROSERS framework is based on two statistically derived surrogate latent variables (LVs) that can sufficiently and efficiently construct the $S_a(T)$ spectrum of single degree of freedom (SDoF) for the earthquake-induced ground motions. In this study, ROSERS is further advanced by retraining and testing it on a dataset of ~11,000 subduction ground motions from the Kik-net and K-net Japanese databases (National Research Institute for Earth Science and Disaster Resilience, 2019). Goodness-of-fit testing is conducted, and the trained VAE and DNN of ROSERS are analyzed for interpretability using explainable artificial intelligence (XAI) and game theory-based Shapley additive explanations (SHAP) (Roth, 1988). SHAP is employed to gain insights into the two LVs by establishing a cause-effect relationship with the $S_a(T)$ spectrum and IMs of the early recorded seismic waves, thereby providing interpretability to end-users.

Furthermore, this study explores the coverage area of an on-site EEW system by developing a novel spatial regression model that estimates the LVs using geographically weighted random forest (GWRF), utilizing M and epicentral distance (R_{epi}) as input features. The estimated LVs are then geographically clustered through a density-based spatial clustering of applications with noise (DBSCAN) algorithm to determine the radius of similarity, which defines the coverage area of an on-site EEW system based on the M of the earthquake event and R_{epi} of the recording station. Hence, this study aims at conducting a more comprehensive analysis of on-site EEW systems (specifically ROSERS) to interpret the underlying DL models and assess their efficacy in terms of coverage area.

In summary, this study provides three major contributions: (1) extension and testing of ROSERS framework on a different seismic environment i.e., Japanese subduction; (2) XAI based interpretation of the trained neural networks for user explainability and more transparent decision-making; and (3) EEW coverage area assessment by training novel spatial regression model i.e., GWRF and clustering the predictions through DBSCAN for similarity analysis.

Specifically, this study contributes to understanding the suitable application range for on-site EEW and identifying the transition point where regional EEW can be more beneficial/reliable. The findings of this research can ultimately contribute to mitigating the societal impact of earthquakes, supporting real-time risk-reduction-oriented decision-making processes, and empowering individuals to take necessary protective actions. By delineating the strengths and limitations of on-site EEW, this study provides valuable insights to optimize the allocation of resources (for instance, in terms of sensor location) and improve the overall effectiveness of earthquake preparedness and response strategies.

2. Ground-motion database

A comprehensive database of unprocessed bi-directional subduction ground motions from the strong motion seismograph networks K-Net and Kik-Net (National Research Institute for Earth Science and Disaster Resilience, 2019) is employed to train and test the ROSERS framework. The ground-motion component time histories are minimally processed with baseline correction and linear trend removal (as a similar process is followed in real-time). The ground motion components with geomean PGA > 0.01 g and R_{epi} < 300 km are finally selected for the analyses to preserve most strong motion dataset (conventionally used distance metric closest rupture distance R_{rup} is not readily available in the databases). This results in ∼11,000 ground motion components (including north–south, NS, and east–west, EW, components) obtained from ∼1500 earthquake events between 1996 and 2022. Fig. 1a shows a synthetic description of the ground motion database in terms of M and R_{epi}. Furthermore, Fig. 1b shows the epicentral locations with scatter sizes and colors representing M and hypocentral depth (Z_{hyp}). While a large section of data belongs to $4 < M < 8$, many ground motions belong to $M > 8$ from the well-known Tokachi and Tohoku earthquakes, making the dataset exhaustive for DL. It is recognized that generally the closest rupture distance (R_{rup}) is used as a potentially more accurate metric. However, due to the availability constraints of the R_{rup} in K-net and Kik-net databases, the study uses R_{epi} as the best available parameter for distance (it is noteworthy that R_{rup} and R_{epi} are inherently correlated). The earthquake events and corresponding source and site information is divided into training and testing sets. The split is done by hand-picking the events with $M > 7.6$ to be in the testing set, and the remaining events are randomly split into training (80%) and testing (20%) sets. The high-magnitude events are purposely kept in the testing set to assess the framework's extrapolation capability.

3. Background

Fig. 2 illustrates the on-site EEW ROSERS framework. After detecting the p-wave at a given recording site, ROSERS captures the early seismic waveforms at the same sensor location. In the original ROSERS work by Fayaz and Galasso (2022), the framework was trained on groundmotions originating from crustal sources (which, on average, last 1–2 min (Fayaz et al., 2020)), requiring three seconds of on-site waveform to estimate the $S_a(T)$ spectrum. However, the subduction ground motions are typically longer, ranging from 2–4 min of strong shaking (Fayaz et al., 2020, 2023). Thus, in this study focusing on subduction ground motions, internal trials were conducted (described in Section 4), and 10 s of ground-motion waveform following p-wave detection was selected for the retraining of the ROSERS framework.

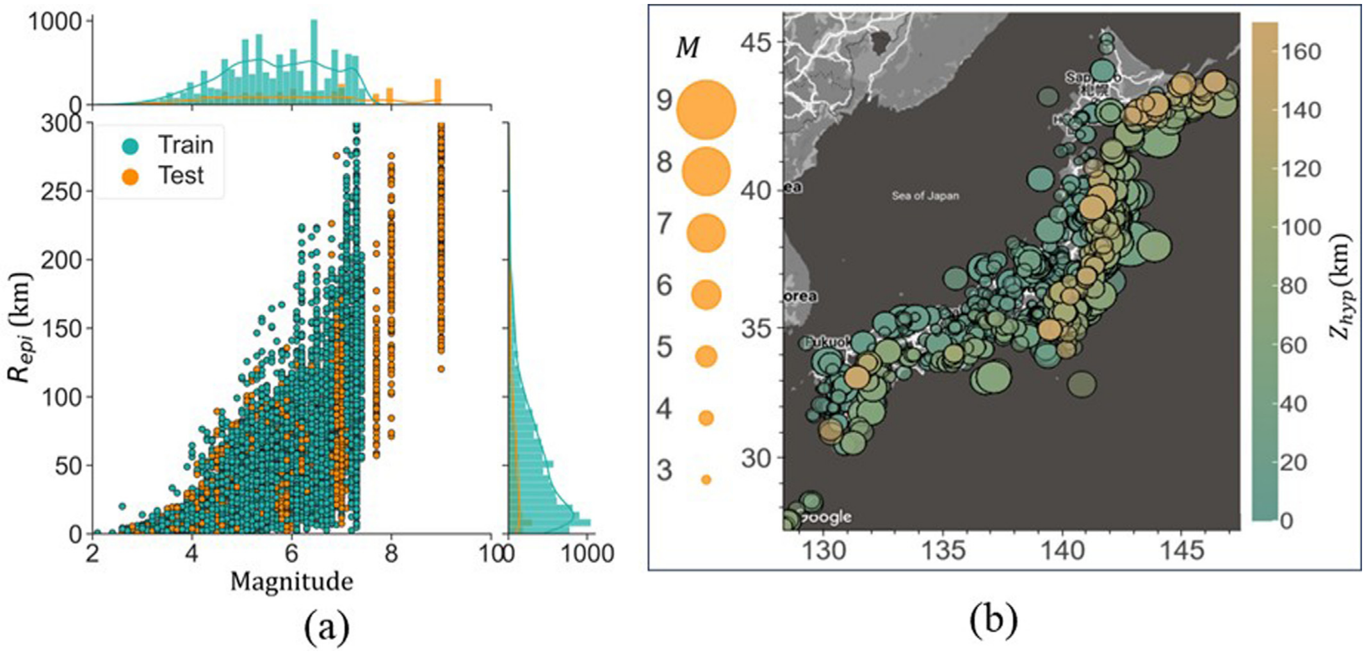


Fig. 1. Ground motion database: (a) M vs. R_{epi}; and (b) epicentral locations with scatter sizes and colors representing M and Z_{hyp}.

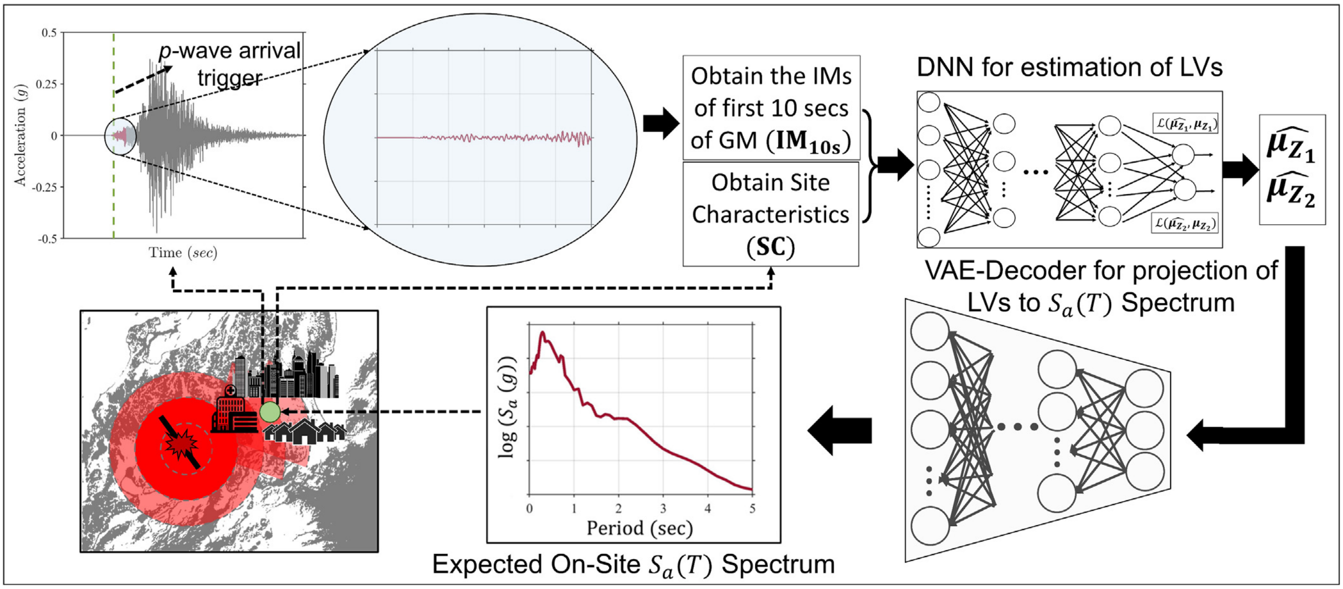

Acceleration (g)
0.5
0.25
0
-0.25
-0.5
Time (sec)
p-wave arrival trigger
Obtain the IMs of first 10 secs of GM ($\mathbf{IM_{10s}}$)
Obtain Site Characteristics ($\mathbf{SC}$)
DNN for estimation of LVs
$\mathcal{L}(\widehat{\mu_{Z_1}}, \mu_{Z_1})$
$\mathcal{L}(\widehat{\mu_{Z_2}}, \mu_{Z_2})$
$\widehat{\mu_{Z_1}}$
$\widehat{\mu_{Z_2}}$
VAE-Decoder for projection of LVs to $S_a(T)$ Spectrum
$\log(S_a(g))$
0
1
2
3
4
5
Period (sec)
Expected On-Site $S_a(T)$ Spectrum

Fig. 2. Illustration of ROSERS framework (Fayaz and Galasso, 2022).

Once the p-waves are detected by the sensor (Akazawa, 2004; Kalkan, 2016), the ROSERS framework waits to receive 10 s of the waveform data. It then computes a vector of seven ground-motion IMs (i.e., $\mathbf{IM_{10s}}$) representing the amplitude, energy, frequency, and significant duration of the initial 10 s of the ground-motion waveform. Specifically, the $\mathbf{IM_{10s}}$ includes seven IMs: Arias intensity (I_a), significant duration (D_{5-95}), mean period (T_m), PGA, peak ground velocity (PGV), peak ground displacement (PGD), and cumulative absolute velocity (CAV). The framework combines $\mathbf{IM_{10s}}$ with the apriori-known site characteristics (**SC**); however, in this case, the database lacked the **SC** information for most database stations; hence only $\mathbf{IM_{10s}}$ is used. This is not expected to severely impact the framework's performance as the **SC** information is also inherently linked in the IMs of the ground motions; in addition, Fayaz and Galasso (2022) showed that **SC** has a low predictive power on the response-spectrum estimation. Hence, only $\mathbf{IM_{10s}}$ is used as the input to the train and utilize the feed-forward DNN. The DNN is trained to estimate two statistically derived surrogate LVs (i.e., μ_{z_1} and μ_{z_2}) that are projected into the $S_a(T)$ spectrum of the expected complete waveform using a pre-trained VAE decoder.

The basis of the framework used in this study mainly relies on two DL-based models, including (1) the VAE (Kingma and Welling, 2019) (which provides regularized surrogate parameters, whose encoder projects $S_a(T)$ into two surrogate LV spaces, and the decoder reconstructs $S_a(T)$ using the LVs); and (2) the DNN (which utilizes $\mathbf{IM_{10s}}$ to compute the two mean LVs μ_{z_1} and μ_{z_2}). Further details of the underlying ROSERS framework can be obtained from Fayaz and Galasso (2022).

4. Training ROSERS on Japanese database

The underlying reason for using a VAE is to develop statistical surrogate LVs for 88-period $S_a(T)$ spectra (including PGA). A VAE provides a probabilistic approach to describe vectorial observation in their LV space. Using a neural network-based encoder and decoder framework, the latent space is compelled to have continuous and smooth representations. Consequently, nearby LVs correspond to similar reconstructions using the decoder.

The $S_a(T)$ spectra of the ~11,000 ground motion components are used as the inputs and outputs in the VAE, and the VAE is bottlenecked to have two independent normally distributed LVs (denoted as z_1 and z_2 with means μ_{z_1} and μ_{z_2}) in the sampling layer. The trends of μ_{z_1} and μ_{z_2} (collectively denoted as **LV**) with M and R_{epi} are presented in Fig. 3a and b. It should be noted that the trends in **LV** in this study are not the same as in Fayaz and Galasso (2022). The variability in the observed patterns is attributed to the inherent randomness (due to the Bayesian sampling layer) in training VAE, impacting the reproducibility of precise trends across studies and also to the differences in the characteristics of ground motions originating from the different seismic environments, i.e., crustal US West coast (Fayaz and Galasso, 2022) and the Japanese subduction (this study).

The goal of training VAE is to encode the LVs so that they are sufficient and efficient (Bosq, 2007) to reconstruct the $S_a(T)$ spectra using the decoder. The selected configuration of the VAE is trained through hyperparameter optimization (Bergstra and Bengio, 2012) with cross-validation. The training is conducted using a log transformation of the $S_a(T)$. The reconstruction power of the final VAE is presented in Fig. 3c, where the coefficients of determination (R^2) for different periods are presented for both train and test sets. On average, $R^2 > 0.9$ is observed across all periods, thereby indicating the sufficiency and efficiency of the **LV** to reconstruct $S_a(T)$ spectra of the Japanese subduction ground motions. A drop in R^2 particularly observed in the very-short-period domain ($S_a(T < 0.2\mathrm{s})$) can be attributed to the minimal ground motion processing utilized in this study (to better replicate the conditions of real-time records) and the highly stochastic nature of the ground motions originating from subduction sources. This drop in R^2 is deemed insignificant for the practical EEW applications since these periods are generally outside the range of typical short-to-long-period structures/critical infrastructure (i.e., including school buildings, high-rise buildings, factories, nuclear power plants, etc.) (e.g., Whittake et al., 2014; Aydınoğlu and Vuran, 2015; Xiang et al., 2020).

As discussed in Fayaz and Galasso (2022), an accurate and rapid estimation of **LV** is required for the construction of the accurate $S_a(T)$ spectrum at a target site. Only a few early seconds of the arriving waveform can be practically used to allow ample warning time for both on-site and regional settings during an occurring ground motion, which can last up to 2–4 min in most subduction sources (Fayaz et al., 2020). Various initial time windows (ranging

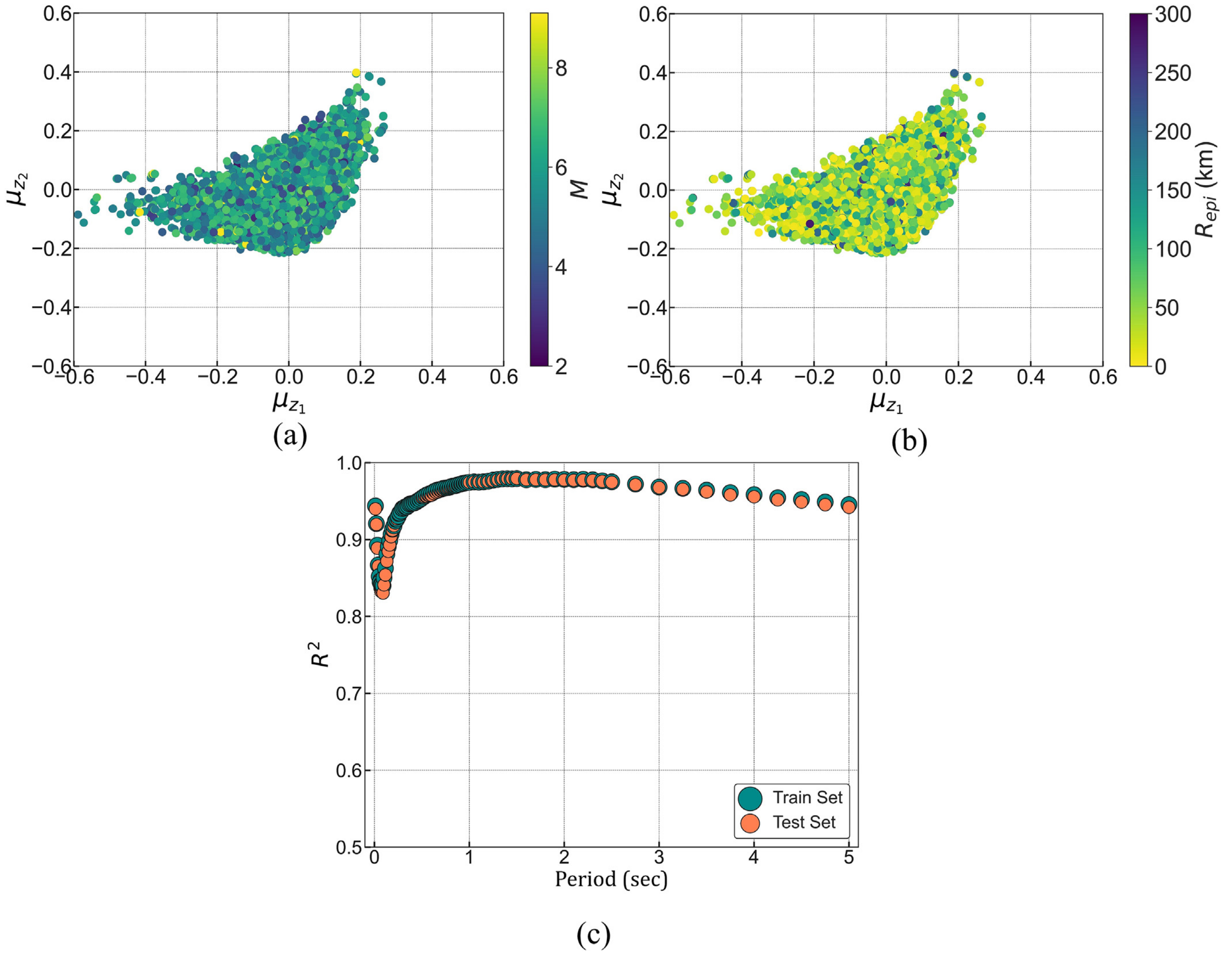


Fig. 3. LVs *vs.* (a) M; and (b) R_{epi}; and (c) R^2 of VAE predictions.

from 1 to 30 s) after detection of the *p*-wave arrival were considered for computing the amplitude-, duration-, energy-, and frequency-based ground-motion IMs and then using them to estimate **LV**. During this exercise, the time window of 10 s was observed to be a good trade-off between the prediction power and the requirement of a short time window (Fayaz and Galasso, 2022). The correlation matrix between $\mathbf{IM_{10s}}$ and **LV** is shown in Fig. 4a. It can be observed that most IMs are well correlated with the **LV** and hence can play a vital role in the prediction process. It should be noted that the DNNs can capture highly nonlinear relations that may not be observed in the correlation matrix.

Finally, a DNN is trained using the $\mathbf{IM_{10s}}$ vectors as inputs to predict the **LV** vector in real time. The DNN is trained through hyperparameter optimization with cross-validation using the training dataset (randomly selected 80% of the events). The final DNN led to an average $R^2 \sim 0.9$ for both LVs. The goodness of fit is shown through predicted *vs.* true values in Fig. 4b and c. Furthermore, mean LVs are estimated simultaneously, thereby ensuring cross-correlations.

5. Interpretation of the neural networks of ROSERS

Due to the versatility of DL models, they have been widely used in engineering applications. However, due to the "black box" nature of these models, there is a general reluctance in the research community to recommend and employ such models. Hence it is critical to provide sufficient analytics for model interpretability and its response in terms of predictions based on variability in the input features. With the onset of XAI, various algorithms have been developed that provide different methods that allow interpretability of these "black-box" models and step towards a "grey-box" and even "white-box" nature (such as linear regression, decision trees).

This study uses SHAP to analyze and interpret the nature of the developed framework. SHAP is a post-hoc model-agnostic procedure that provides insights to explain individual predictions of the model based on the game's theoretically optimal Shapley values. Shapley values are a widely used approach from cooperative game theory with desirable characteristics (Roth, 1988). The Shapley value is the average marginal contribution of a feature value across all possible coalitions where the feature values of a data instance act as coalition players. Shapley values are the only solution that satisfies the properties of efficiency, symmetry, dummy, and additivity, which form the basis of the most reliable explanation methods (Lundberg and Lee, 2017). SHAP belongs to the class of models called "additive feature attribution methods", where the SHAP values explain the contribution of the features to the respective outputs in a quantitative manner, thereby allowing interpreta-

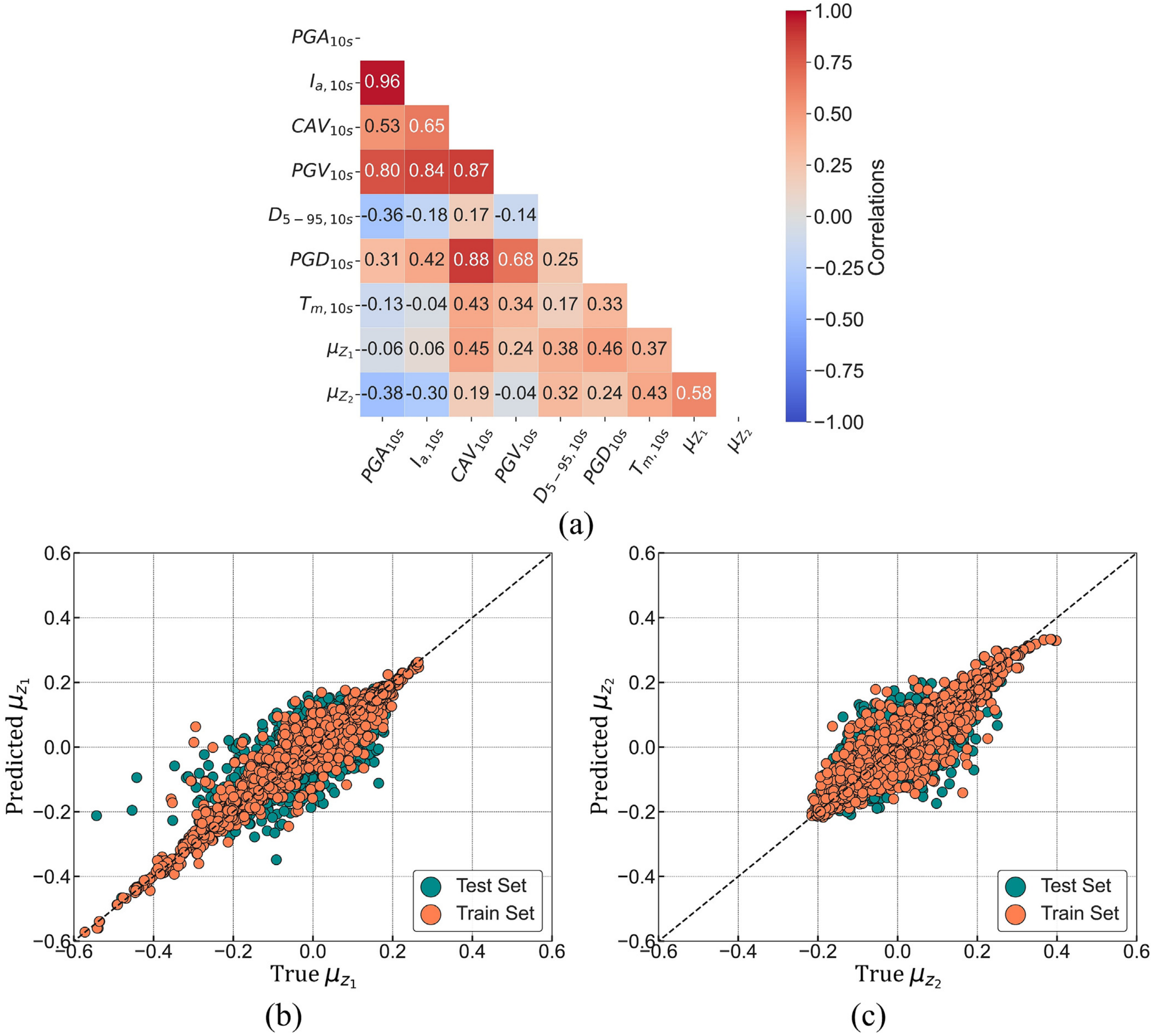


Fig. 4. (a) Correlations between $\mathbf{IM_{10s}}$ and **LV**; true *vs.* DNN-predicted (b) μ_{z_1}; and (c) μ_{z_2}.

tion. These are analogically similar to the coefficients of a regression model, which provide the impact of the corresponding feature on the target variable. Due to the computational complexity, SHAP values are approximated using various types of explainers such as kernel-explainer, tree-explainer, deep-explainer, etc. (Molnar, 2020). In this study, kernel-explainer uses a special weighted linear regression to compute the importance of each feature. The computed importance values are Shapley values from game theory and coefficients from a local linear regression (Molnar, 2020). The following sections apply SHAP analysis to interpret the cause-effect relationship between the predictors and targets of the trained VAE decoder and DNN.

5.1. Interpretation of the VAE decoder

The decoder of the trained VAE is analyzed using SHAP analysis with μ_{z_1} and μ_{z_2} of the ~11,000 ground motions as the inputs (features) and the corresponding predictions of $S_a(T)$ spectra as the outputs (targets). For each input–output combination, the SHAP value is computed. Hence, a total of ~11,000 SHAP values are calculated for each combination of the 88 outputs ($S_a(T)$ spectra) and two inputs (μ_{z_1} and μ_{z_2}). The values are presented for *PGA*, $S_a(T = 0.5\text{s})$, $S_a(T = 1\text{s})$, and $S_a(T = 2.5\text{s})$ in Fig. 5, where the color of the data points represents the magnitude of the feature values. In this case, 'low' represents values close to −0.6, and 'high' refers to values close to +0.3 for μ_{z_1} and similarly, 'low' refers to values close to −0.25, and 'high' refers to values close to +0.4 for μ_{z_2} (based on Fig. 3a and b). It can be observed that, in general for $S_a(T > 0.5\text{s})$, with an increase in the value of both features, their corresponding SHAP values tend to move from negative values to positive values. Similar behavior is observed for the $S_a(T \le 0.5\text{s})$ for μ_{z_1}; however, the SHAP behavior is observed to be more convoluted for $S_a(T \le 0.5\text{s})$ for μ_{z_2} (as will be discussed in Section 5.2, μ_{z_2} seems to contain minimal information related to stiff and short-period SDoFs). This means that as the values of the LVs increase, they tend to increase the predicted $S_a(T)$ values (positive correla-

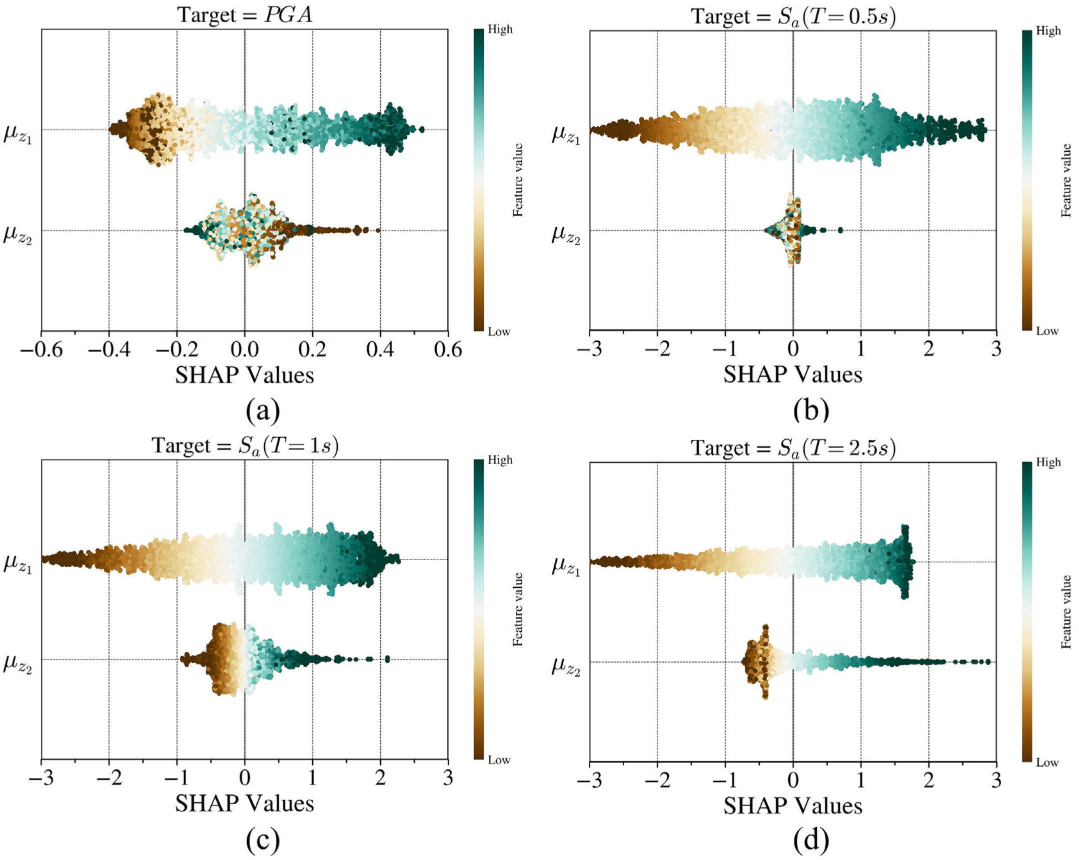


Fig. 5. VAE-decoder SHAP values of μ_{z_1} and μ_{z_2} for prediction of (a) *PGA*; (b) $S_a(T = 0.5\text{s})$; (c) $S_a(T = 1\text{s})$; and (d) $S_a(T = 2.5\text{s})$.

tion) obtained from the decoder; in contrast, lower values of LVs lead to lower predicted $S_a(T)$ values (negative correlation). Also, it is observed that the SHAP values tend to be symmetric on both sides of zero for μ_{z_1} for all four target variables. This means that for both extreme values of the LVs (i.e., −0.6 and +0.3), their absolute impact on the four target variables is similar (since the absolute SHAP value is similar).

However, for μ_{z_2}, it is observed that the SHAP values are not highly distinct for $S_a(T)$ with shorter periods (*PGA* and $S_a(T = 0.5\text{s})$). This phenomenon is observed to be lower for $S_a(T)$ with longer periods ($S_a(T = 1\text{s})$, and $S_a(T = 2.5\text{s})$), thereby indicating the higher importance of μ_{z_2} for prediction of $S_a(T)$ at longer periods. Furthermore, it is observed that the ranges of SHAP values for the two LVs for the four periods are generally different, indicating different impacts of μ_{z_1} and μ_{z_2} on the different periods of $S_a(T)$. Comparing the SHAP values across the four target variables, it is noticed that the contribution range of the **LV,** particularly μ_{z_2}, increases (due to an increase in the spread) with an increase in the period of the SDoF. In a nutshell, μ_{z_1} is observed to have a higher impact on shorter periods while μ_{z_2} has a higher impact on longer periods. A large variation in μ_{z_1} computed at different sites during an earthquake event thereby signifies high amplitude and frequency content of ground motions while large variation in μ_{z_2} indicates low frequency and high energy content of ground motions.

5.2. Interpretation of the DNN

Similar to the interpretation process of the VAE decoder, the trained DNN is analyzed using SHAP analysis by using the $\mathbf{IM_{10s}}$ features of the ~11,000 ground motions as the inputs and the corresponding mean latent variables μ_{z_1} and μ_{z_2} as the outputs. The respective SHAP values are computed for each case for the trained DNN. Hence, a total of ~11,000 SHAP values are calculated for each combination of the seven inputs ($\mathbf{IM_{10s}}$) and two outputs (μ_{z_1} and μ_{z_2}). Fig. 6 presents the SHAP values for the $\mathbf{IM_{10s}}$ corresponding to the two target mean latent variables μ_{z_1} and μ_{z_2} in descending order of mean contribution. The color of the data points represents the magnitude of the corresponding feature values. It can be observed from the sub-figures that *CAV*, *PGA*, and D_{5-95} lead to the highest SHAP values for μ_{z_1} while for μ_{z_2} the SHAP values of D_{5-95} get reduced significantly and I_a is observed to be the highest contributor, followed by *CAV* and *PGA*. This indicates that μ_{z_1} is affected by the amplitude (*PGA*), frequency (*CAV*), and duration (D_{5-95}) of the ground motion and μ_{z_2} is impacted by the energy (I_a), frequency (*CAV*), and amplitude (*PGA*) of the ground motion. This bolsters the observations made in the previous section about μ_{z_1} containing higher information related to *PGA* and stiffer period $S_a(T)$ and signifies the importance of μ_{z_2} in capturing the energy and long period $S_a(T)$ of the ground motion waveform. In both cases of μ_{z_1} and μ_{z_2}, T_m is observed to be the least contributor. Also, the DNN is observed to be more impacted by the *PGV* of the initial ground motion for the prediction of μ_{z_2} as compared to μ_{z_1}.

Unlike the previous section, the SHAP values of the features are not observed to be symmetric around zero (especially for μ_{z_1}), thereby indicating that the value of different features leads to different contributions to the DNN predictions. In general, it is observed that for features *AV*, *PGV*, D_{5-95}, and *PGD* lower values of the features lead to a negative contribution, and higher values of the features lead to a positive contribution (positive correlation).

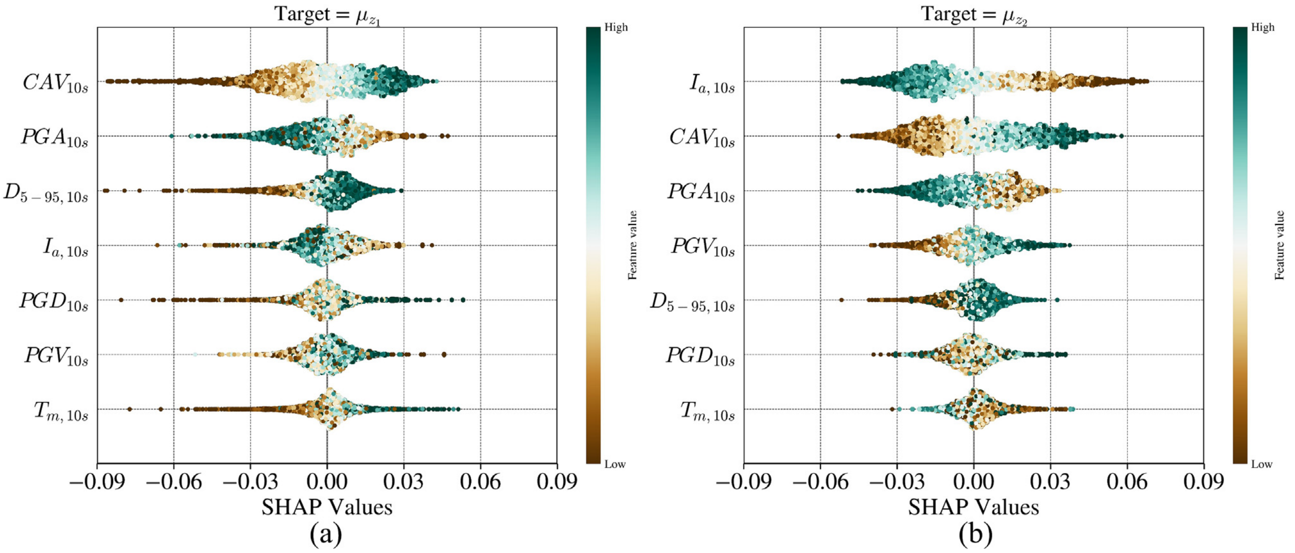


Fig. 6. DNN SHAP values of $\mathbf{IM_{10s}}$ for prediction of (a) μ_{z_1}; and (b) μ_{z_2}.

While for I_a and *PGA* opposite behavior is observed (negative correlations). This is in line with the observations made in Fig. 4a. It should be noted here negative sense does not mean lower contribution but rather specifies that the corresponding feature value lowers the prediction from the average prediction value of the DNN.

5.3. Summary of interpretations

To present the details more concisely, Fig. 7 illustrates the relative feature importance of μ_{z_1} and μ_{z_2} on the VAE decoder predictions and the relative feature importance of $\mathbf{IM_{10s}}$ on the DNN predictions of μ_{z_1} and μ_{z_2}, in terms of their mean absolute SHAP values (|SHAP|). This is done by computing the mean |SHAP| values for both VAE and DNN for the ~11,000 samples for each target and then dividing them by the sum of the mean |SHAP| for each target. Since the SHAP values represent the contribution of the features (i.e., μ_{z_1} and μ_{z_2} for VAE and $\mathbf{IM_{10s}}$ for DNN) to the model output, computing their relative sum for the two features in an absolute sense signifies the importance of the respective feature in predicting the target (i.e., $S_a(T)$ for VAE and μ_{z_1} and μ_{z_2} for DNN).

It is observed from Fig. 7a that μ_{z_1} has a dominant influence on $S_a(T)$ for shorter to mid-range spectral periods (*PGA*, 0.2 s, 0.5 s, and 1 s). This dominance gradually decreases for longer periods (2.5–5 s), suggesting that μ_{z_1} is more indicative of the ground motion behavior where higher frequencies are more prominent. Typically, it is known that the amplitude of the ground motions controls the short-period behavior (e.g., Bozorgnia et al., 2004), thereby reaffirming the observations made in Section 5.1 for μ_{z_1}. Conversely, μ_{z_2} shows a different trend, with its relative importance increasing for $S_a(T)$ with longer spectral periods (2.5 s and 5 s). This trend aligns with the understanding that the energy content of the ground motions, which is more closely associated with μ_{z_2} (as discussed in Section 5.2), significantly influences flexible and long-period structures (e.g., Bozorgnia et al., 2004). Exceptions to general trends are noted in the mid-range periods of 0.5 and 1 s, where μ_{z_1}'s importance peaks predominately compared to μ_{z_2}, suggesting complex interactions between the amplitude and energy content of the ground motions at these intermediate periods. These will require further analysis, which is beyond the scope of this study.

In general, it is concluded that the observations made in Fig. 7a are consistent with the discussions of the previous section and the summary plot shown in Fig. 7b. It is observed from Fig. 7b that *CAV* tends to have similar importance for both μ_{z_1} and μ_{z_2}, and μ_{z_1} and μ_{z_2} are dominated by amplitude-based *PGA* and energy-based I_a, respectively. Thus, in general, it is observed that μ_{z_1} inherits the capabilities to capture the acceleration and amplitude effects on stiffer SDoFs, and with an increase in flexibility of SDoF, μ_{z_2} increases in its importance, thereby inheriting energy-based characteristics. Hence this process of interpreting DL-based models

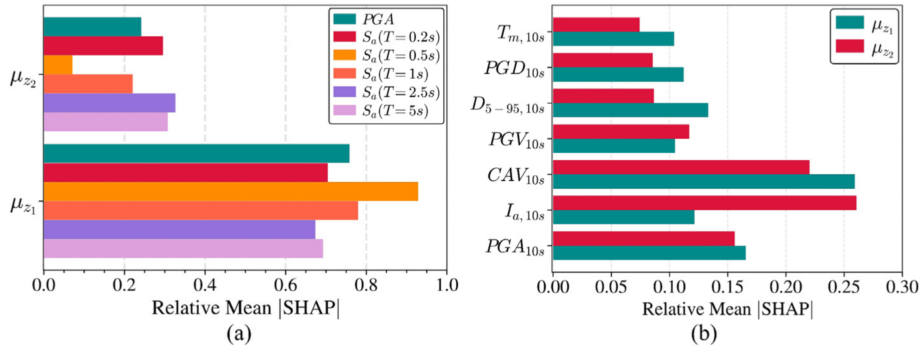


Fig. 7. Relative mean absolute SHAP values of (a) **LV** for $S_a(T)$ prediction (VAE); (b) $\mathbf{IM_{10s}}$ for **LV** prediction (DNN).

provides insights into the cause-effect relationship between the features and targets, thereby promoting transparency and providing details about the LVs in capturing the behavior of SDoF systems with different periods. Furthermore, this encourages understanding the "black box" nature of the predictions made by the neural networks and how the target variables are affected by the changes in the features (similar to classical regression analysis).

6. Spatial efficacy analysis of the ROSERS latent variables

This section analyzes the spatial efficacy of the ROSERS EEW framework in terms of coverage area. The coverage area of an on-site EEW is an important consideration for its effective operation and implementation. Herein, the coverage area refers to the geographical extent within which the warning based on a single EEW sensor is reliable. Since the sensors are geographically distributed over a region, it is important to know how far the warnings from a station can be used for EEW at nearby locations. Hence, understanding the coverage area is crucial for determining the reach and applicability of the system, as well as for informing decision-making processes and response strategies.

The coverage area of an on-site EEW can be influenced by various factors, including M, distance from the source, soil conditions, earthquake parameters of interest, and the system's capabilities. Ideally, a higher density of seismic stations can resolve this issue by providing more comprehensive monitoring of ground shaking throughout the region. However, such provisions are impractical due to cost and social (such as privacy) considerations. Hence, the strategic placement of sensors in seismically active zones based on the coverage area estimates can ensure that the system effectively serves at-risk communities.

Various methodologies and techniques are employed to estimate the coverage area, including spatial regression models, geostatistics, and statistical analysis of historical earthquake data (e.g., Páez and Wheeler, 2009; Bi and Hao, 2012; Caramenti et al., 2022; Meng and Goulet, 2022). These approaches help determine the range of applicability and the spatial distribution of warning capabilities. In this study, a spatial regression model (i.e., GWRF; Georganos et al., 2021) is trained and validated to represent spatial variability of the LVs (which are sufficient and efficient surrogate variables representing $S_a(T)$, as discussed above). The seismic event parameters (i.e, M and R_{epi}) serve as the inputs and μ_{z_1} and μ_{z_2} of NS and EW components are used as the target variables (hence two inputs and four outputs). It should be noted here that although the training and interpretation of the ROSERS framework hinge on the LVs and IMs of both ground motion components (NS and EW), they are merged in the dataset and not explicitly mentioned in the paper uptill now. This omission is due to the developed VAE and DNN models relying on component-specific inputs and outputs which inherently do not possess common parameters (e.g. $S_a(T)$ or $\mathbf{IM}_{10s}$ are distinct for the two ground motion components of the same station). However, since the two ground motion components correspond to same M and R_{epi}, for clarity it is explicitly mentioned in this section.

To accurately incorporate the spatial correlations, only the events from the dataset recorded at multiple stations must be considered. Hence a preliminary check is made in Fig. 8a to check the number of events in the database that are recorded at various minimum numbers of stations. Based on the results, the spatial regression analysis is conducted using the ~65 events ($5 \leq M \leq 9$), which are recorded at a minimum of 20 stations, thereby leading to > 5000 ground motion components.

The trained spatial regression model is utilized to estimate the LVs for both components for a grid-based inputs of M and R_{epi}. The predictions of each component LV are utilized for clustering similar LVs together while accounting for the station's distance from the source across different magnitudes. This allows stepped discretization of the predicted LVs which is then used to compute the half-width of each discretized section in terms of R_{epi} and obtain a function form that relates coverage area in terms of coverage radius (r_c) with the M, R_{epi} and LVs. Fig. 8b illustrates an earthquake event recorded at stations 1 to 5. The coverage area of a station represents the area around any station up till which the predictions can be statistically considered to be the same as that of the station point (in this case, it is considered to be a circle shown in red color around the station 2) and r_c represents the radius of the circle. Hence, the coverage area (in terms of r_c) of an on-site EEW sensor is based on the similarity of the LVs as predicted by the spatial regression model. A schematic flowchart of spatial efficacy analysis of ROSERS on-site sensors is provided in Fig. 9. The details of the models and procedure are provided in the following two sections.

6.1. Geographically weighted random forests (GWRF)

GWRF is an innovative approach that combines the strengths of two powerful techniques: random forests (RF) and geographically weighted regression (GWR) (Georganos et al., 2021). It aims to improve the modeling of spatially correlated data by incorporating both spatial autocorrelation and the nonlinear relationships between the features and target variables.

Traditional RF and other regression models treat the data as independent and identically distributed, ignoring the spatial structure and potential spatial heterogeneity. On the other hand, GWR (Páez and Wheeler, 2009), such as the spatially varying coefficients model (SVCM) (Georganos et al., 2021), considers spatial nonstationarity by estimating local models for each observation, capturing the spatially varying relationships between the features and targets. However, GWR has limitations due to the underlying relationship model form requirement, computational complexity, the lack of stability, and high bias and variance (Georganos et al., 2021). A GWRF addresses these limitations by integrating RF and GWR. It applies RF locally, considering a subset of the data within a specific geographic neighborhood defined by a spatial weighting kernel. This allows the model to capture the spatial variations of the relationships between predictors and the response variable while benefiting from RF's computational efficiency and stability.

One of the main differences between traditional GWR and GWRF models is the non-stationarity coupled with a flexible nonlinear model, which is difficult to overfit due to its bootstrapping nature, thus relaxing the assumptions of traditional Gaussian statistics. By combining multiple decision trees trained on different subsets of data, a GWRF captures the complex interactions and nonlinear relationships between predictors and the response variable. Furthermore, including spatial weighting ensures that the model gives more weight to nearby observations, considering their proximity in the analysis and accounting for spatial autocorrelation. Essentially, GWRF was designed to bridge machine learning and geographical models, combining inferential and explanatory power.

As a first study in seismology, the proposed GWRF model is kept simple, and ground motion waveforms are considered ergodic and isotropic to provide better interpretability and usability. The models are trained using the two most essential earthquake parameters M and R_{epi} of the training set as the input features and μ_{z_1} and μ_{z_2} of the NS and EW ground motion components as the target variables. For a good bias-variance trade-off, after hyperparameter tuning, the GWRF is trained with random bootstrapping using a maximum depth of 10 trees, a bandwidth of five stations, and a weighing kernel based on the haversine distance between the stations. The weighing kernel ($w(d)$) is expressed in Eqs. (1)–(3)

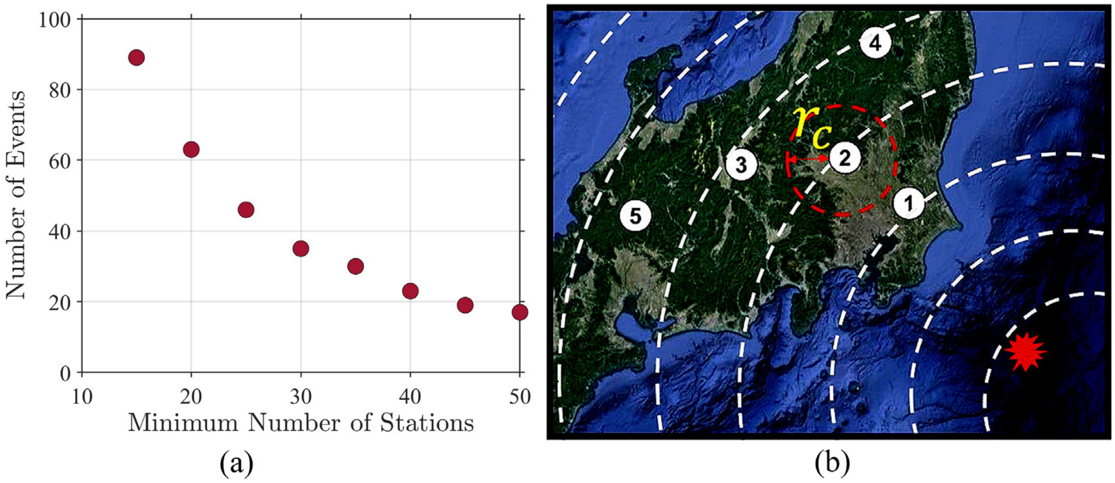


Fig. 8. (a) Number of events *vs.* minimum number of recording stations in the dataset; (b) illustration of coverage area and coverage radius (r_c) of an on-site EEW sensor.

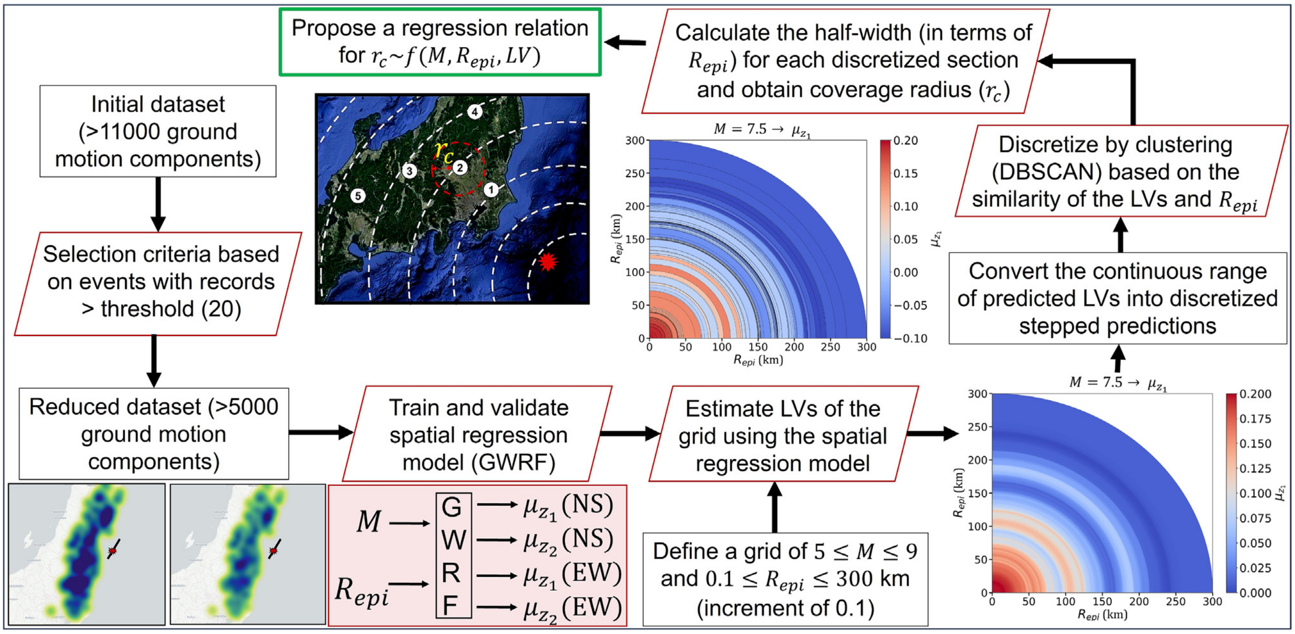

Propose a regression relation for $r_c \sim f(M, R_{epi}, LV)$
Calculate the half-width (in terms of R_{epi}) for each discretized section and obtain coverage radius (r_c)
Initial dataset (>11000 ground motion components)
Selection criteria based on events with records > threshold (20)
Reduced dataset (>5000 ground motion components)
Train and validate spatial regression model (GWRF)
Estimate LVs of the grid using the spatial regression model
Define a grid of $5 \le M \le 9$ and $0.1 \le R_{epi} \le 300$ km (increment of 0.1)
Discretize by clustering (DBSCAN) based on the similarity of the LVs and R_{epi}
Convert the continuous range of predicted LVs into discretized stepped predictions
$M = 7.5 \rightarrow \mu_{z_1}$
R_{epi} (km)
μ_{z_1}
M
R_{epi}
G
W
R
F
μ_{z_1} (NS)
μ_{z_2} (NS)
μ_{z_1} (EW)
μ_{z_2} (EW)
r_c
1
2
3
4
5
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
50
100
150
200
250
300

Fig. 9. Flowchart for spatial efficacy analysis of ROSERS LVs.

where $\Delta\phi$ and $\Delta\lambda$ are the differences in latitude and longitude between the two points (in radians), respectively, ϕ_1 and ϕ_2 are latitudes of the two points (in radians), R represents the radius of Earth (=6371 km), d is the haversine distance, and σ is the bandwidth parameter. The hyperparameter tuning included grid-based analysis for the maximum depth of trees ranging from 5 to 20 trees, bandwidth of stations ranging from 3 to 10 stations, different weighing kernels including Gaussian, bisquare, and exponential kernels with Euclidean and haversine distances. The residuals of the two LVs from the model predictions of NS component are presented against M and R_{epi} of all ~65 events in Fig. 10, where the red blocks show the 5 to 95 percentile grouped box plots (similar results are obtained for the EW component). It is observed that the residuals of both LVs for both components have a mean close to 0, indicating the white noise characteristics and demonstrate no evident bias against any values of M and R_{epi}.

$$a = \sin^2\left(\frac{\Delta\phi}{2}\right) + \cos(\phi_1)\cdot\cos(\phi_2)\cdot\sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (1)$$

$$d = R\cdot 2\cdot \operatorname{atan2}(\sqrt{a}, \sqrt{1-a}) \quad (2)$$

$$w(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right) \quad (3)$$

Fig. 11 further presents the true values and GWRF predictions of μ_{z_1} and μ_{z_2} of NS component for an event of $M = 7.3$ (origin time = 13:30:18 JST on October 6th, 2000; latitude = 35.270°N; longitude = 133.352°E; focal depth = 12.2 km) from the test split of the dataset. In general, it can be observed from the figure that the true and predicted values of the two LVs are very close to each other across the region of recordings. Although the predicted values are lower than the true values for a particular minority of the regions, this is not expected to affect the results of this study significantly. It should be noted that even though more complex and intriguing predictive models can be proposed for the application, the goal of this study is not to offer a highly accurate model. Instead, this study focuses on using a model that appropriately

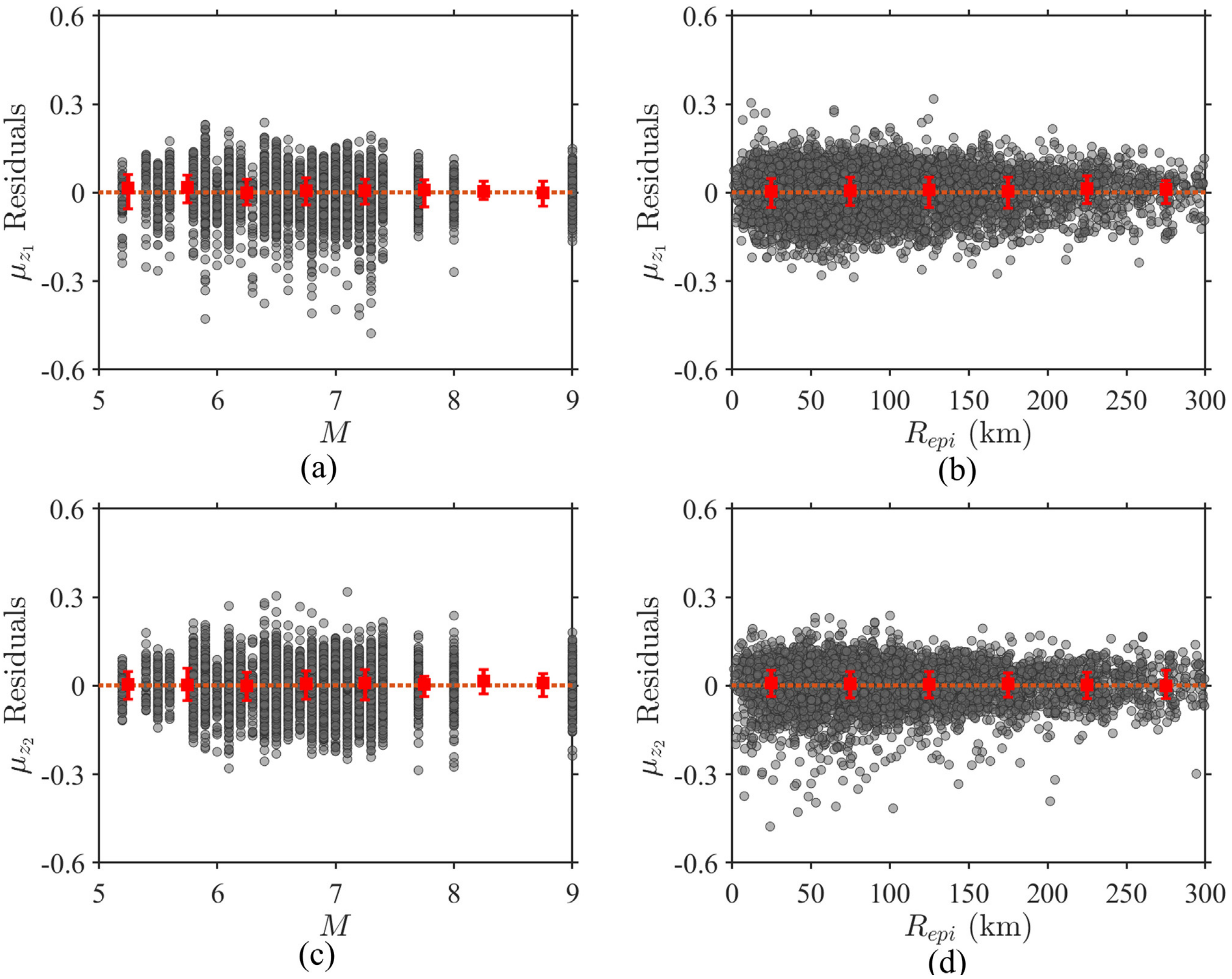


Fig. 10. GWRF residuals for μ_{z_1} *vs.* (a) M; and (b) R_{epi}; and μ_{z_2} *vs.* (c) M; and (d) R_{epi} of the NS component.

captures the spatial trends of LVs with the least number of input features to assess the coverage area of each sensor under the ROSERS EEW framework.

6.2. Estimation of coverage radius ($\boldsymbol{r_c}$)

The trained GWRF is used to estimate the LVs for a grid of $5 \leq M \leq 9$ and $0.1 \leq R_{\text{epi}} \leq 300$ km at an increment of 0.1 and ~0.1 km, respectively. This is done by defining hypothetical earthquake events of varying magnitudes and assuming stations throughout the geographical region of Japan at increments of 0.01 latitude and longitude. The input grid is used to make predictions of LVs using the trained GWRF. Fig. 12a and b showcases the predictions of μ_{z_1} and μ_{z_2} of NS component for an event with $M = 7.5$. It should be noted here the values of the LVs do not represent the intensity of shaking, so larger values do not need to be estimated at stations with small R_{epi}.

To define the coverage area and obtain r_c, the continuous change in the LVs, as observed in Fig. 12a and b, must be discretized based on the similarity with the LVs of the neighboring station (closer values of R_{epi} based on the assumption of ergodicity and isotropy). Hence in other terms, the predicted LVs are required to be clustered together based on the similarity of the LVs and R_{epi}. In this case, DBSCAN (Ester et al., 1996) is used for the clustering. DBSCAN is a density-based clustering algorithm that groups data points based on their density and proximity. It can help identify clusters of similar values within a certain radius. In this case, the maximum distance parameter of 10 km and 30 km is used to define the maximum threshold radius of similarity for R_{epi} < 150 km and R_{epi} > 150 km, respectively. Also, the minimum number of points is kept as 5 to form a cluster. This is done by hyperparameter tuning to optimize the algorithm.

The DBSCAN algorithm is utilized to discretize the continuous LV predictions of the GWRF into stepped predictions, thereby allowing for a more interpretable representation. For the LV predictions of $M = 7.5$ event in Fig. 12a and b, the discretized results are presented in Fig. 12c and d, respectively. The boundaries marked in black indicate the points where the LV values change, and the half-width of each discretized section represents the r_c associated with the central point. From Fig. 12, it is observed that DBSCAN successfully discretizes the continuous predictions into cluster-based stepped predictions. This behavior is consistent across various grid events analyzed in this study.

In general, for $R_{\text{epi}} < 100$ km, $100 < R_{\text{epi}} < 200$ km, and $R_{\text{epi}} > 200$ km, mean r_c of ~2 km, ~4 km, and ~9 km, respectively, are observed to be optimal. Hence the findings suggest that for EEW sensors lying in close proximity to the epicenter ($R_{\text{epi}} < 100$ km), an effective warning can be issued for a relatively small coverage radius ($r_c \sim 2$ km), providing a high level of similar-

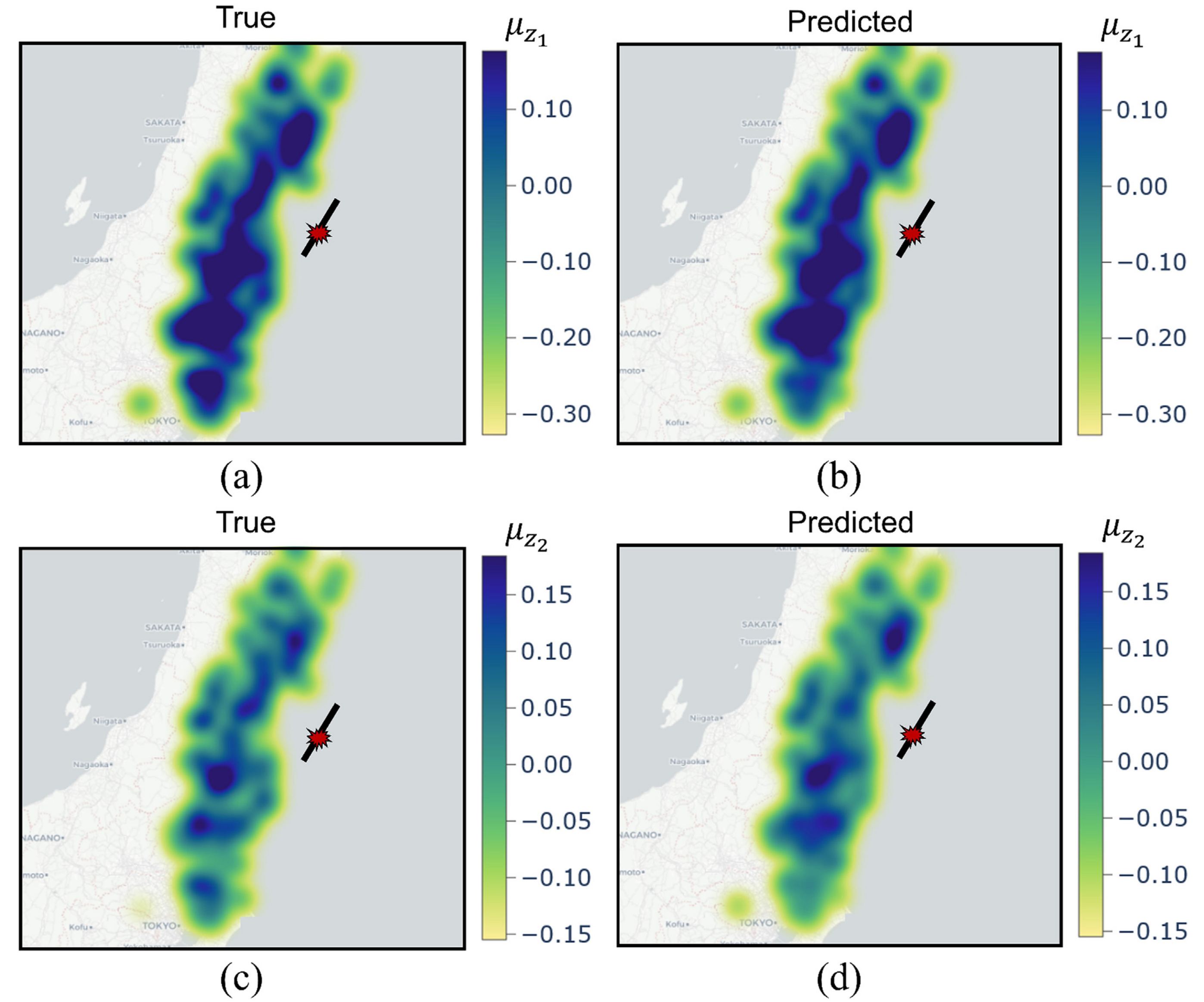


Fig. 11. (a) True μ_{z_1}; and (b) GWRF-predicted μ_{z_1}; and (c) true μ_{z_2}; and (d) GWRF-predicted μ_{z_2} for NS component for an event with $M = 7.3$.

ity accuracy. For sensors with R_{epi} between 100 km and 200 km, the optimal coverage radius of similarity increases to approximately 4 km ($r_c \sim 4$ km). This suggests that as the distance from the earthquake source increases, the on-site warnings are reliable for a larger coverage area. Finally, for the sensors that are at a considerable distance from the epicenter, a much wider coverage area ($r_c \sim 9$ km) can rely on the on-site warnings. It should be noted that the values mentioned here are averaged for the example event $M = 7.5$ and the range of r_c for different R_{epi} can be seen from the Fig. 12. Furthermore r_c values are observed to fluctuate with respect to the M of the earthquake event.

Subsequently, the obtained r_c is then regressed against the corresponding M, R_{epi}, and μ_z as per Eq. (4). The objective is to provide users with a simple yet predetermined representation to estimate r_c based on the expected M, R_{epi}, and μ_z. This information can assist end users in strategically placing sensors in seismically active zones, thereby minimizing blind spots and mitigating errors caused by regional extrapolation.

In a pre-event scenario, the M and R_{epi} can be easily obtained from earthquake rupture forecasting tools (Field et al., 2009, 2017). The value of μ_z can be derived by inverting the predictions of $S_a(T)$ from probabilistic seismic hazard analysis (PSHA) using the pre-trained encoder of the VAE. The results of fitting the regression model are presented in Table 1, demonstrating the regression coefficients. Although more complex prediction models could be employed for regression purposes, this study proposes a simple model for improved interpretability and usability. Overall, the combination of DBSCAN-based discretization and the regression model provides a practical framework to estimate r_c based on readily available inputs, enabling effective sensor placement and enhancing EEW in seismically active regions. It should be noted that the results presented herein are based on Japanese geography, site conditions, and source characteristics. The results can deviate for other regions, and similar thorough analysis is needed.

$$r_c = \beta_0 + \beta_1 \cdot M + \beta_2 \cdot R_{\mathrm{epi}} + \beta_3 \cdot M \cdot R_{\mathrm{epi}} + \beta_4 \cdot \ln(\mu_z + 1) \qquad (4)$$

7. Conclusions

The potential risks posed by earthquakes to communities worldwide underscore the critical need for developing dependable EEW systems as part of a holistic earthquake mitigation strategy,

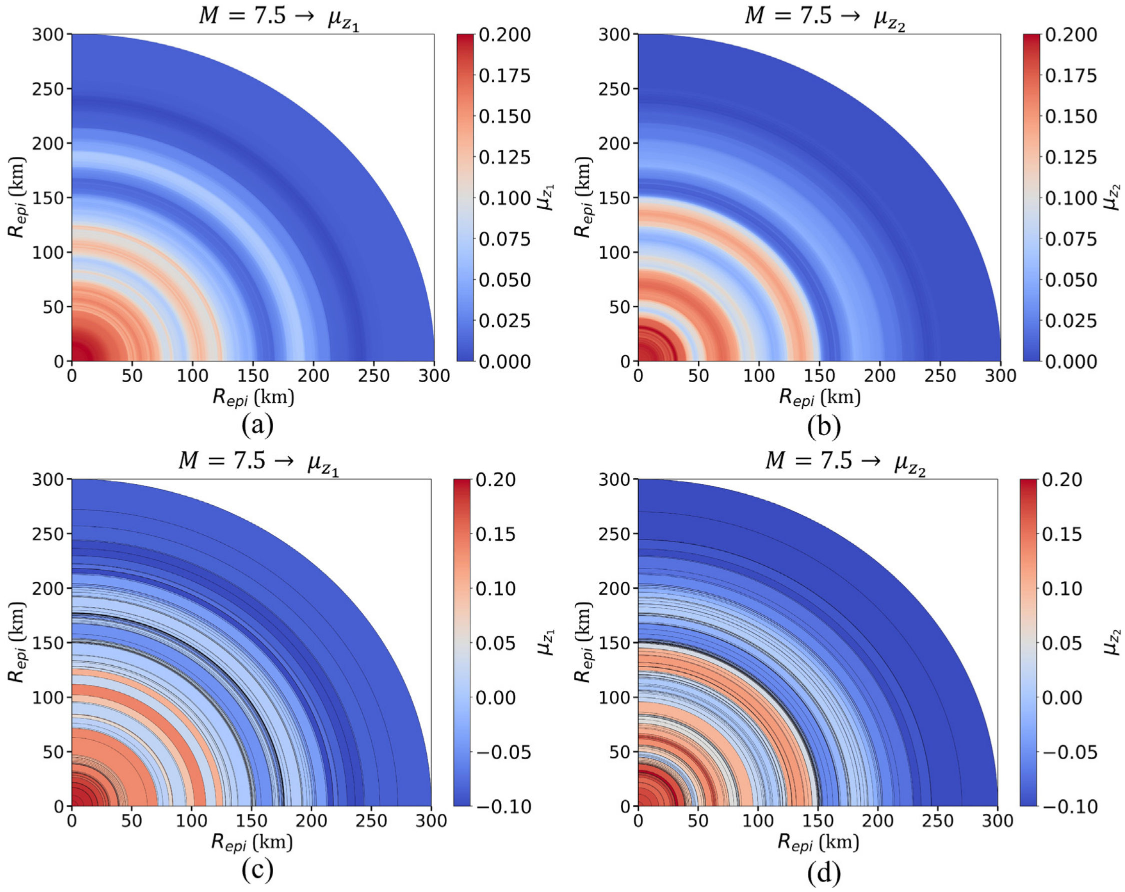


Fig. 12. GWRF LV predictions *vs* R_{epi} for (a) μ_{z_1}; and (b) μ_{z_2}; and DBSCAN smoothened LV predictions *vs* R_{epi} for (c) μ_{z_1}; and (d) μ_{z_2} for NS component.

Table 1
Coefficients for Eq. (4) for r_c estimation.

r_c	β_0	β_1	β_2	β_3	β_4
μ_{z_1} (NS)	−86.92	0.93	15.32	−0.13	−79.52
μ_{z_1} (EW)	−128.16	0.95	21.85	−0.13	−86.55
μ_{z_2} (NS)	−52.53	0.70	9.83	−0.10	−67.62
μ_{z_2} (EW)	−101.01	0.79	17.36	−0.11	−74.60

where EEW's role, among other earthquake risk-reduction tactics, is critical. Recent technological advancements have enabled the training and implementation of data-driven and ML/DL based EEW systems. However, due to the black-box nature of such models, the internal peculiarities of the models remain unknown, leading to reluctance in their use by the community. Furthermore, due to the lack of resources (and other potential reasons) to lay a fine grid of EEW sensors in a region, it is vital to understand the extent of coverage each EEW sensor can provide for effective real-time operational efficiency.

This study addressed the urgent need for effective EEW systems by investigating the real-time on-site EEW framework known as ROSERS. Firstly, the ROSERS framework was evaluated for effectiveness and applicability for a different seismic region than previously used Western U.S crustal sources by Fayaz and Galasso (2022). The framework was extended to the Japanese subduction zone by retraining it on a dataset of ∼11,000 unprocessed subduction ground motion components. The goodness-of-fit testing revealed that ROSERS framework accurately estimates response spectra of the on-site incoming ground motion waveform from early recorded 10 s of the ground motion with $R^2 > 0.9$. This finding futher established ROSERS as a reliable and robust framework for subduction earthquake scenarios.

Secondly, the interpretability analysis using XAI techniques, particularly the application of game theory-based SHAP, provided valuable insights into the two LVs of the VAE and their supplementary correlations with short- and long-period $S_a(T)$. The SHAP analysis of the trained DNN further established a cause-effect relationship between the LVs and the IMs of the early recorded seismic waves, thereby enhancing the interpretability of the framework. In general, it was observed that μ_{z_1} inherits the capabilities to capture the acceleration and amplitude effects on stiffer SDoFs, and with an increase in flexibility of SDoF, μ_{z_2} increases in its importance, thereby inheriting energy-based characteristics This interpretability aspect is crucial for end-users, as it allows them to understand the underlying mechanisms and make informed decisions based on the predictions provided by ROSERS.

Finally, this study explored the spatial efficacy of ROSERS and assessed the coverage area of the on-site EEW stations. By training a novel spatial regression model using GWRF, the study estimated the LVs using the M and R_{epi} as the input features. The LVs were estimated for a grid of $5 \leq M \leq 9$ and $0.1 \leq R_{\text{epi}} \leq 300$ km at an increment of 0.1 and ~0.1 km, respectively. The estimated LVs were then clustered and averaged using DBSCAN to determine the radius of similarity (i.e., r_c) that defines the coverage area of the ROSERS system. The results indicated that in general, for $R_{\text{epi}} < 100$ km, $100 < R_{\text{epi}} < 200$ km, and $R_{\text{epi}} > 200$ km, mean r_c of ~2 km, ~4 km, and ~9 km, respectively, are observed to be optimal. Furthermore, linear regression equations were provided to estimate r_c based on the M, R_{epi} and LV value. This information is valuable for optimizing the placement of sensors in seismically active zones, minimizing blind spots, and improving the overall effectiveness of on-site EEW systems.

The findings of this research have practical implications for mitigating the societal impact of earthquakes, aiding real-time risk reduction-oriented decision-making processes, and empowering individuals to take necessary protective actions. By delineating the strengths and limitations of ROSERS, this study paves the way for allocating resources and enhancing earthquake preparedness and response strategies. Future research can build upon these findings to further optimize on-site EEW systems and improve their effectiveness in providing timely warnings and protecting vulnerable communities.

CRediT authorship contribution statement

Jawad Fayaz: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. **Carmine Galasso:** Investigation, Validation, Writing – review & editing.

Data availability

The ground motion records utilized in this study are openly available through the Kyoshin Network operated by the National Research Institute for Earth Science and Disaster Resilience (NIED) in Japan. The data can be accessed and downloaded from the Kyoshin Network's website at https://www.kyoshin.bosai.go.jp/kyoshin/docs/overview_kyoshin_en.shtml. Researchers and interested parties are encouraged to explore and utilize these openly accessible ground motion records for further analysis and scientific investigations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors extend their appreciation to Mr. Owais Javed, graduate student at Teesside University, for his contributions to the organization of data and his diligent efforts in the preliminary trials of spatial regression modeling. The authors also acknowledge the assistance of Dr. Rodrigo Astroza and his group in compiling the Japanese subduction ground-motion database.

References

Akazawa, T., 2004. A technique for automatic detection of onset time of P- and S-phases in strong motion records, in: Proceed of the 13th World Conference on Earthquake Engineering, Vancouver, B.C., Canada, August 1-6, Paper No. 786.

Ancheta, T.D., Darragh, R.B., Stewart, J.P., Seyhan, E., Silva, W.J., Chiou, B.-S.-J., Wooddell, K.E., Graves, R.W., Kottke, A.R., Boore, D.M., Kishida, T., Donahue, J.L., 2014. NGA-West2 database. Earthq. Spectra 30 (3), 989–1005. https://doi.org/10.1193/070913EQS197M.

Aydınoğlu, M.N., Vuran, E., 2015. Developments in Seismic Design of Tall Buildings: Preliminary Design of Coupled Core Wall Systems. In: Ansal, A. (Ed.), Perspectives on European Earthquake Engineering and Seismology. Geotechnical, Geological and Earthquake Engineering, vol 39. Springer, Cham, 227–243. https://doi.org/10.1007/978-3-319-16964-4_9.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305. http://scikit-learn.sourceforge.net.

Bhardwaj, R., Sharma, M.L., Kumar, A., 2016. Multi-parameter algorithm for earthquake early warning. Geomat. Nat. Haz. Risk 7 (4), 1242–1264. https://doi.org/10.1080/19475705.2015.1069409.

Bi, K., Hao, H., 2012. Modelling and simulation of spatially varying earthquake ground motions at sites with varying conditions. Probab. Eng. Mech. 29, 92–104. https://doi.org/10.1016/J.PROBENGMECH.2011.09.002.

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., Atzmueller, M., 2023. Graph neural networks for multivariate time series regression with application to seismic data. Int. J. Data Sci. Anal. 16 (3), 317–332. https://doi.org/10.1007/s41060-022-00349-6.

Bosq, D., 2007. Sufficiency and efficiency in statistical prediction. Statist. Probab. Lett. 77 (3), 280–287. https://doi.org/10.1016/J.SPL.2006.07.021.

Bozorgnia, Y., Vitelmo V. Bertero, V.V., 2004. Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering. CRC Press, Boca Raton. https://doi.org/10.1201/9780203486245.

Campbell, K.W., Bozorgnia, Y., 2013. NGA-West2 Campbell-Bozorgnia ground motion model for the horizontal components of PGA, PGV, response spectra for periods ranging from 0.01 to 10 Sec. PEER Report 2013/06. Peer Report, no. May.

Campbell, K.W., Bozorgnia, Y., 2019. Ground motion models for the horizontal components of Arias Intensity (AI) and cumulative absolute velocity (CAV) using the NGA-West2 database. Earthq. Spectra 35 (3), 1289–1310. https://doi.org/10.1193/090818EQS212M.

Caramenti, L., Menafoglio, A., Sgobba, S., Lanzano, G., 2022. Multi-source geographically weighted regression for regionalized ground-motion models. Spatial Statistics 47,. https://doi.org/10.1016/J.SPASTA.2022.100610 100610.

Caruso, A., Colombelli, S., Elia, L., Picozzi, M., Zollo, A., 2017. An on-site alert level early warning system for Italy. J. Geophys. Res. Solid Earth 122 (3), 2106–2118. https://doi.org/10.1002/2016JB013403.

Cremen, G., Galasso, C., 2020. Earthquake early warning: recent advances and perspectives. Earth Sci. Rev. 205,. https://doi.org/10.1016/j.earscirev.2020.103184 103184.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Knowl. Discov. Data Min. 96 (34), 226–231.

Fayaz, J., Galasso, C., 2022. A deep neural network framework for real-time on-site estimation of acceleration response spectra of seismic ground motions. Comput. Aided Civ. Inf. Eng. 38 (1), 87–103. https://doi.org/10.1111/mice.12830.

Fayaz, J., Medalla, M., Zareian, F., 2020. Sensitivity of the response of Box-Girder Seat-type bridges to the duration of ground motions arising from crustal and subduction earthquakes. Eng. Struct. 219,. https://doi.org/10.1016/j.engstruct.2020.110845 110845.

Fayaz, J., Xiang, Y., Zareian, F., 2021. Generalized ground motion prediction model using hybrid recurrent neural network. Earthq. Eng. Struct. Dyn. 50 (6), 1539–1561. https://doi.org/10.1002/eqe.3410. T.H.

Fayaz, J., Medalla, M., Torres-Rodas, P., Galasso, C., 2023. A recurrent-neural-network-based generalized ground-motion model for the Chilean subduction seismic environment. Struct. Saf. 100,. https://doi.org/10.1016/j.strusafe.2022.102282 102282.

Field, E.H., Dawson, T.E., Felzer, K.R., Frankel, A.D., Gupta, V., Jordan, T.H., Parsons, T., Petersen, M.D., Stein, R.S., Weldon, R.J., Wills, C.J., 2009. Uniform California earthquake rupture forecast, version 2 (UCERF 2). Bull. Seismol. Soc. Am. 99 (4), 2053–2107. https://doi.org/10.1785/0120080049.

Field, E.H., Jordan, T.H., Page, M.T., Milner, K.R., Shaw, B.E., Dawson, T.E., Biasi, G.P., Parsons, T., et al., 2017. A synoptic view of the third uniform California earthquake rupture forecast (UCERF3). Seismol. Res. Lett. 88 (5), 1259–1267. https://doi.org/10.1785/0220170045.

Galasso, C., Zuccolo, E., Aljawhari, K., Cremen, G., Melis, N.S., 2023. Assessing the potential implementation of earthquake early warning for schools in the Patras region, Greece. Int. J. Disaster Risk Reduct. 90,. https://doi.org/10.1016/j.ijdrr.2023.103610 103610.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 36 (2), 121–136. https://doi.org/10.1080/10106049.2019.1595177.

Hsu, T.Y., Huang, S.K., Chang, Y.W., Kuo, C.H., Lin, C.M., Chang, T.M., Wen, K.L., Loh, C. H., 2013. Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan. Soil Dyn. Earthq. Eng. 49, 210–217. https://doi.org/10.1016/j.soildyn.2013.03.001.

Iaccarino, A.G., Picozzi, M., Bindi, D., Spallarossa, D., 2020. Onsite earthquake early warning: predictive models for acceleration response spectra considering site effects. Bull. Seismol. Soc. Am. 110 (3), 1289–1304. https://doi.org/10.1785/0120190272.

Jozinović, D., Lomax, A., Štajduhar, I., Michelini, A., 2020. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. Geophys. J. Int. 222 (2), 1379–1389. https://doi.org/10.1093/gji/ggaa233.

Jozinović, D., Lomax, A., Štajduhar, I., Michelini, A., 2022. Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. Geophys. J. Int. 229 (1), 704–718. https://doi.org/10.1093/gji/ggab488.

Kalkan, E., 2016. An automatic P-phase arrival-time picker. Bull. Seismol. Soc. Am. 106 (3), 971–986. https://doi.org/10.1785/0120150111.

Kingma, D.P., Welling, M., 2019. An introduction to variational autoencoders. Found. Trends Mach. Learn. 12 (4), 307–392. https://doi.org/10.1561/2200000056.

Kramer, S.L., 1996. Geotechnical Earthquake Engineering. Prentice Hall.

Lin, J.C.-C., Lin, P.-Y., Chang, T.-M., Lin, T.-K., Weng, Y.-T., Chang, K.-C., Tsai, K.-C., 2012. Development of on-site earthquake early warning system for Taiwan. In: D'Amico S. (Ed.), Earthquake Research and Analysis - New Frontiers in Seismology. InTech. https://doi.org/10.5772/28056.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30.

McBride, S.K., Bostrom, A., Sutton, J., de Groot, R.M., Baltay, A.S., Terbush, B., Bodin, P., et al., 2020. Developing post-alert messaging for Shakealert, the earthquake early warning system for the West Coast of the United States of America. Int. J. Disaster Risk Reduct. 50,. https://doi.org/10.1016/j.ijdrr.2020.101713 101713.

Meng, X., Goulet, C.A., 2022. Lessons learned from applying varying coefficient model to controlled simulation datasets. Bull. Earthq. Eng. 21, 5151–5174. https://doi.org/10.1007/s10518-022-01512-x.

Molnar, C., 2020. Interpretable Machine Learning. Lulu.com.

Münchmeyer, J., Bindi, D., Leser, U., Tilmann, F., 2021. The transformer earthquake alerting model: a new versatile approach to earthquake early warning. Geophys. J. Int. 225 (1), 646–656. https://doi.org/10.1093/gji/ggaa609.

National Research Institute for Earth Science and Disaster Resilience, 2019. NIED K-NET, KiK-Net. National Research Institute for Earth Science and Disaster Resilience.

Páez, A., Wheeler, D.C., 2009. Geographically weighted regression. Int. Encyclopedia Human Geogr. 47 (3), 407–414. https://doi.org/10.1016/B978-008044910-4.00447-8.

Roth, A.E., 1988. The Shapley Value: Essays in Honor of Lloyd S. Shapley. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511528446.

Tajima, F., Hayashida, T., 2018. Earthquake early warning: What does "seconds before a strong hit" mean? Prog. Earth Planet Sci. 5 (1), 63. https://doi.org/10.1186/s40645-018-0221-6.

Whittake, A.S., Kumar, M., Kumar, M., 2014. Seismic isolation of nuclear power plants. Nucl. Eng. Technol. 46 (5), 569–580. https://doi.org/10.5516/NET.09.2014.715.

Wu, S., Beck, J.L., Heaton, T.H., 2013. ePAD: Earthquake probability-based automated decision-making framework for earthquake early warning. Comput. Aided Civ. Inf. Eng. 28 (10), 737–752. https://doi.org/10.1111/mice.12048.

Xiang, Y., Farzad, N., Farzin, Z., 2020. Evaluation of natural periods and modal damping ratios for seismic design of building structures. Earthq. Spectra 36 (2), 629–646. https://doi.org/10.1177/8755293019900776.