REWFERS: A Regional Early Warning Framework for Estimating Response Spectra

Jawad Fayaz  
Assistant Professor (Lecturer), School of Computing, Engineering, and Digital Technologies, Teesside University, Teesside, UK

Rodrigo Astroza  
Associate Professor, Faculty of Engineering and Applied Sciences, Universidad de los Andes, Santiago, Chile

Miguel Medalla  
Lecturer, Faculty of Engineering and Applied Sciences, Universidad de los Andes, Santiago, Chile

ABSTRACT: Earthquake early warning (EEW) systems aim to provide critical information (i.e., magnitude, ground-shaking intensity, and/or potential consequences) about the characteristics of incoming earthquake waves at a target site to mitigate the hazard in real-time/near-real-time rapidly. In general, EEW systems are classified into two groups: 1) on-site systems, which use early features of the incoming waves at the target site itself; 2) regional systems, which use the seismic waves recorded at sensors located closer to the epicenter to estimate ground shaking intensity at the target site and implement appropriate mitigation measures or actions. This study introduces a hybrid deep learning- and Bayesian statistics-based regional EEW framework called REWFERS: Regional Early Warning Framework for Estimating Response Spectra. The proposed framework is based on variational autoencoders (VAE), deep neural networks (DNN), and Gaussian process regressions (GPR) trained to utilize the early non-damage-causing p-waves recorded at the station closest to the epicenter to estimate the acceleration response spectrum \( S_a(T) \) at a farther-located target site in real-time. The framework is trained and assessed using a carefully selected extensive database of ~14800 Japanese strong ground motions with peak ground accelerations > 0.01g. The proposed framework works in two phases. The first phase uses GPR and VAE to obtain prior \( S_a(T) \) estimate at the target site using the early waves picked at the station closest to the rupture. The second phase takes over once the waves arrive at the target site, which are then used to obtain an on-site \( S_a(T) \) estimate using the DNN and VAE. The on-site estimate is then combined with the prior estimate through Bayesian updating to provide a more informative posterior estimate. The framework is statistically tested, and it is observed that the framework offers a highly accurate prediction of expected \( S_a(T) \) in real-time and can help improve EEW.

1. INTRODUCTION
The objective of earthquake early warning (EEW) systems is to provide accurate and timely warnings to support risk management and decision support systems during an earthquake event. In a real-time setting, EEW systems generally utilize the time difference between the arrival of faster non-damaging longitudinal p-waves and slower destructive transverse s-waves (Kramer, 1996). Since the information waves travel faster than the seismic waves, EEW systems exploit this phenomenon to trigger alarms. However, due to the slight delay between the arrival of the two waves and the short duration of such events, the efficacy of EEW systems play a crucial role in ensuring that the warnings are rapid, accurate, and structurally useful. This warning time, although short, can mitigate the impacts of an earthquake on many sectors of society ranging from individuals following the “drop, cover, and hold on” to shutting down potential threats like gas in pipelines, high-speed trains, etc. (Velazquez et al., 2020).

EEW systems can be broadly divided into two sub-classes: (1) regional and (2) on-site. Regional EEW systems consist of a network of seismograms located within the region of high seismicity, which
use the recorded waves to predict ground shaking at the target site before the arrival of damaging waves. On the other hand, the on-site EEW system is a standalone system that uses the early recorded waves to estimate ground shaking on or in the vicinity of the recording station (which typically is the target site). In both cases of EEW, the recorded waves are generally used to estimate the source parameters, which are then fed to pre-calibrated ground motion models (GMMs, e.g., Campbell & Bozorgnia, 2013; Fayaz, Xiang, et al., 2021) to estimate intensity measures (IMs) like peak ground acceleration (PGA) (Cremen & Galasso, 2020).

Among the various ground-shaking parameters of interest (Fayaz, Azar, et al., 2021), the acceleration response spectrum ($S_a(T)$) has been one of the most widely used IMs that can effectively integrate the features of the ground motion waveform (such as amplitude, frequency content, etc.) with the dynamic behavior of a structural system idealized as a single-degree-of-freedom (SDOF) system (Bazzurro et al., 1998). Over the years, $S_a(T)$ has grown in its utility for structural and earthquake engineering purposes, including ground motion simulation validation (Fayaz, Dabaghi, et al., 2020; Fayaz, Rezaeian, et al., 2021), ground motion selection (Fayaz et al., 2019; Vamvatsikos & Allin Cornell, 2002), etc.

Aiming to improve the EEW systems, Fayaz & Galasso, (2022) proposed a data-driven deep learning-based framework called Real-time On-Site Estimation of Response Spectra (ROERS) to estimate the expected on-site ground motion $S_a(T)$ using IMs of early p-waves and site characteristics. Based on a robust feed-forward deep neural network (DNN) (Chollet, 2017), ROERS uses the initial 3 seconds of the arriving ground motion (after p-waves detection) to estimate two sufficient and efficient latent variables (LVs) that can sufficiently and efficiently construct $S_a(T)$ spectrum, arising from crustal type earthquakes, using the decoder of a pre-trained variational autoencoder (VAE) (Kingma & Welling, 2019).

This paper extends the on-site ROERS framework to a regional EEW system called REWFERS (Regional Early Warning Framework for Estimating Response Spectra). The framework is trained and tested on a comprehensive Japanese ground motion database composed of subduction and crustal ground motions. REWFERS comprises training of: i) VAE to obtain two-dimensional LV space ($LV$) of the Japanese ground motions, ii) DNN to estimate the on-site LVs in real-time using a vector of IMs of early 10-seconds of ground motion waves ($IM_{10s}$) and site characteristics (SC), and iii) Gaussian process regression (GPR) (Williams & Rasmussen, 1995) based spatial regression model which uses $LV$, $IM_{10s}$, and SC of the first station to undergo ground motion combined with the SC of the target site and the distance between the first and target site to estimate the $LV$ at the target site.

![Figure 1: Phase 1 of REWFERS framework; at the onset of the earthquake event, the early 10 seconds of ground motion (GM) received at the first station are used to obtain: i) the on-site $S_a(T)$ estimate for station 1 using ROERS framework, and ii) the prior $S_a(T)$ estimate for the target station 'n' using GPR-based spatial correlation model. This provides initial estimate for EEW at the n<sup>th</sup> station.](image-url)
2. REWFERS FRAMEWORK
The REWFERS framework is based on the surrogate LVs proposed by Fayaz & Galasso (2022). The framework works in two phases:

Phase 1: After the onset of the earthquake rupture, as the early 10 seconds of seismic waves arrive at the first station (after \( p \)-wave detection), as done in ROSERS, the pre-trained on-site DNN is used to estimate the LV using the SC and \( IM_{10s} \) vectors at the first station (denoted as \( LV^1, SC^1 \) and \( IM_{10s} \)). Following the process conducted by the ROSERS framework, The estimated \( LV^1 \) is then used as input to the pre-trained VAE decoder to estimate the on-site \( S_a(T) \) spectrum at the first station. The \( LV^1 \) is combined with \( SC^1, SC^n \) (SC at the \( n \)th station), \( IM^n_{10s} \), and the distance between the first and \( n \)th station \( (d_{1n}) \). The combination is then used as inputs to the GPR-based spatial regression model to estimate \( LV \) at the \( n \)th station (denoted as \( LV^n \)). The \( LV^n \) is then used in the pre-trained VAE decoder to obtain a prior estimate of \( S_a(T) \) spectrum at the \( n \)th station. Hence the \( S_a(T) \) spectrum is estimated at the \( n \)th station before the seismic waves arrive, thereby offering a regional early warning. This is illustrated in Figure 1.

Phase 2: As 10 seconds of waves arrive at the \( n \)th station (after \( p \)-wave detection), the pre-trained DNN and VAE decoder are used to compute the on-site estimate of \( LV^n \) and \( S_a(T) \) spectrum at the \( n \)th station using \( SC^n \) and \( IM^n_{10s} \). The prior estimate and on-site estimates of \( LV^n \) are combined through Bayesian updating to obtain the final posterior \( S_a(T) \) spectrum. The proposed REWFERS framework provides on-site and regional EEW capabilities to alert the community in real time. The framework utilizes less than 3 secs on average, providing end-users with ample time for decision-making through risk informed EEW decision support systems and shake maps. This is illustrated in Figure 2.

3. GROUND MOTION DATABASE
A comprehensive database of unprocessed bi-directional 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 REWFERS 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 \( PGA > 0.01 \text{g} \) are finally selected for the analyses. This results in 14,865 ground motion components obtained from 1860 earthquake events between 1996 and 2022. Figure 3 shows a description of the ground motion database in terms of moment magnitude \((M)\) and epicentral distance \((R_{epi})\). While the large section of data belongs to \( 4 < M < 8 \), many ground motions...
belong to $M > 8$ from the well-known Tokachi and Tohoku earthquakes, thereby making the dataset exhaustive for deep learning.

LVs for the 88-period $S_a(T)$ spectra (including PGA). 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 possess continuous and smooth representations. Consequently, nearby LVs correspond to similar reconstructions using the decoder.

The $S_a(T)$ spectra of the 14865 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 $\mu_{z_1}$ and $\mu_{z_2}$) in the sampling layer. The vector of $\mu_{z_1}$ and $\mu_{z_2}$ is denoted as LV. The trends of LVs with $M$ and $R_{epi}$ are presented in Figures 4a and 4b. The goal here is to encode the LVs so that are sufficient and efficient to reconstruct the $S_a(T)$ spectra using the decoder. The VAE is trained through cross-validation using a randomly selected 80% of the events while the remaining 20% is used as the test set. The VAE is trained using a log transformation of the $S_a(T)$ spectra. The reconstruction power of the final VAE can be observed in Figure 5a, 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

![Figure 3: Metadata of the ground motion database](image)

### 4. DEVELOPMENT OF REWFERS

The proposed REWFERS framework relies on four major machine-learning-based components explained in the following sections, and the results are discussed.

#### 4.1 Spectral Surrogacy through VAE

The underlying reason for using VAE (Kingma & Welling, 2019) is to develop statistical surrogate...
indicating the sufficiency and efficiency of the LV to reconstruct $S_a(T)$ spectra.

4.2 DNN for on-site prediction of LVs

As mentioned by Fayaz & Galasso (2022), an accurate and rapid estimation of LV is required for construction of accurate $S_a(T)$ spectrum for the target site. To allow ample warning time for both on-site and regional settings during the occurring ground motion, which can last up to 2 to 4 minutes in most subduction sources (Fayaz, Medalla, et al., 2020), only a few early seconds of the arriving waveform can be practically used. Various initial time windows (ranging from 1 sec to 30 seconds) after detection of $p$-wave arrival were considered for computing the amplitude-, duration-, energy-, and frequency-based ground-motion IMs (as used by Fayaz & Galasso, 2022) and then using them to estimate $\mu_z_1$ and $\mu_z_2$. During this exercise, the time window of 10 seconds was observed to be a good trade-off between the prediction power to estimate LV and the requirement of a short time window (Fayaz & Galasso, 2022). The correlation matrix between IM$_{10s}$ and LV is shown in Figure 5b. It can be observed that most IMs are well correlated with the LVs and hence can play a vital role in the prediction process. It should be noted that the DNNs

**Figure 5:** (a) $R^2$ of reconstructed $S_a(T)$; (b) correlation matrix between IM$_{10s}$ and LV

**Figure 6:** True vs. Predicted (a) $\mu_z_1$ and (b) $\mu_z_2$, from the trained DNN
can capture highly nonlinear relations that may not be observed in the correlation matrix.

Finally, a DNN was trained using the SC (includes only $V_{30}$) and $\text{IM}_{10s}$ vectors as inputs to predict the LV vector in real-time. The DNN is trained with cross-validation using the training dataset (randomly selected 80% of the events), while evaluations are conducted on the remaining test dataset. After thorough hyperparameter tuning, 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 Figures 6a and 6b. Furthermore, both mean LVs are estimated simultaneously, thereby ensuring the cross-correlations. The DNN provides the on-site LV estimate (i.e. $P(LV^n | IM^n_{10s}, SC^n)$).

4.3 GPR-based spatial regression model for LVs
In order to provide regional estimates of LVs, in this study GPR-based model is used to develop the spatial relationship between the LV at various sites. This is done by using the events from the training set which have been recorded at more than 10 stations to fit the GPR model (Williams & Rasmussen, 1995). GPR serves as an efficient tool for performing inference and kriging both passively (i.e., interpolation or extrapolation) as well as actively (i.e., filtering and smoothing) (Rasmussen

---

**Figure 7:** Normalized mean values of (a) $\mu_{z_1}$ and (b) $\mu_{z_2}$, and standard deviations of (a) $\sigma_{\mu_{z_1}}$ and (b) $\sigma_{\mu_{z_2}}$, for prior and posterior distributions.
& Williams, 2006). Hence it can serve as a good resource to serve as spatial regression model.

The $\text{IM}_{10s}$ and $\text{LV}$ vectors computed at the first station (i.e., station closest to the rupture), $\text{SC}$ of the $n^{th}$ and first station, and the distance between the stations is used to develop the covariance structure by using them as the inputs to the GPR. While various kernels are available to develop GPR’s covariance structure (Fayaz et al., 2023), this study uses the summation of matérn and white kernels (Rasmussen & Williams, 2006). The train and test datasets of the GPR lead to an average $R^2$ of $0.7$ for both LVs. The trained GPR is used in phase 1 of the REWFERS framework to obtain the prior estimate of $\text{LV}$ at the target site based on the spatial correlation (i.e., $P(\text{LV}^n|\text{LV}^1, \text{SC}^n, \text{SC}^1, d_{n,1})$).

4.4 Development of LVs’ posterior distributions through Bayesian updating
As mentioned in Section 2, phase 2 of the REWFERS framework involves Bayesian updating of the prior $\text{LV}$ distribution ($P(\text{LV}^n|\text{LV}^1, \text{SC}^n, \text{SC}^1, d_{n,1})$) using the on-site $\text{LV}$ distribution($P(\text{LV}^n|\text{IM}_{10s}^n, \text{SC}^n)$), thereby obtaining the posterior $\text{LV}$ and then the $\text{S}_a(T)$ spectrum at the target station. The Bayesian updating is conducted via Markov-Chain Monte Carlos (MCMC) simulation (Vats et al., 2015).

A likelihood function is evaluated through simulation, and conditional updating of the quantity of interest is conducted. In this case, the problem may be stated as follows:

- Prior beliefs: The prior $\text{LV}$ distribution for the target site is obtained using the GPR-based prediction, defined by $P(\text{LV}^n|\text{LV}^1, \text{SC}^n, \text{SC}^1, d_{n,1}) \sim N(\mu(\mu_{\text{LV}}^{GPR}), \sigma(\mu_{\text{LV}}^{GPR}))$, where ($\alpha, \beta$) denotes a normal distribution with mean $\alpha$ and standard deviation $\beta$ and $\mu(\mu_{\text{LV}}^{GPR})$ and $\sigma(\mu_{\text{LV}}^{GPR})$ are the mean and standard deviation of the mean values of the $\text{LV}$ estimated by the GPR model, respectively.

- New information: once the 10 seconds of waves reach the target site, the DNN is used to establish $P(\text{LV}^n|\text{IM}_{10s}^n, \text{SC}^n) \sim N(\mu(\mu_{\text{LV}}^{DNN}), \sigma(\mu_{\text{LV}}^{DNN}))$, where $\mu(\mu_{\text{LV}}^{DNN})$ and $\sigma(\mu_{\text{LV}}^{DNN})$ denotes the mean and standard deviation of the mean values of the $\text{LV}$ estimated by the DNN, respectively.

- The likelihood of observing $\mu_{\text{LV}}^{\text{True}}$ (a deterministic quantity) in the prediction is modeled by an error function as $P(\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{GPR}))|\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{DNN}))) \sim N(\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{GPR})), N(\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{GPR})), 0, \sigma(\mu_{\text{LV}}^{DNN}))$, where the error function is defined as: $\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{GPR})) = \sqrt{(\mu_{\text{LV}}^{\text{True}} - \mu(\mu_{\text{LV}}^{GPR}))^2 + (\mu_{\text{LV}}^{\text{True}} - \mu(\mu_{\text{LV}}^{DNN}))^2}$, and $\sigma(\mu_{\text{LV}}^{DNN})$ serves as the Gaussian discrepancy model. In this problem, the evidence or observed data corresponds to the error function $\text{error}(\mu_{\text{LV}}^{\text{True}}, \mu(\mu_{\text{LV}}^{GPR}))$.

Figure 7 shows the histogram of the statistics where the corresponding true LVs normalize the mean LVs of the predicted distributions. It can be observed that mean estimates of both the LVs are significantly improved after the updating process, and standard deviations are substantially reduced. Therefore, merging information from the spatial regression GPR and on-site DNN provides more accurate LV estimates for all stations.

5. CONCLUSIONS
This study introduced a deep learning-based on-site EEW framework named REWFERS (Regional Early Warning Framework for Estimating Response Spectra). REWFERS extends the concept of the on-site ROSERS framework to regional level. The proposed framework is based on VAE (trained to spatially estimate the $\text{S}_a(T)$ spectrum), DNN (trained to estimate $\text{LV}$ using $\text{SC}$ and $\text{IM}_{10s}$ vectors accurately), and GPR (trained to estimate $\text{LV}$ using $\text{SC}$ at the target $n^{th}$ site using $\text{LV}^1, \text{SC}^n, \text{SC}^1$, and $d_{n,1}$).

REWFERS works in two phases. Phase 1 involves prior estimation of $\text{S}_a(T)$ spectrum at the target site using information from the site closest to rupture and the distance between the target and closest sites. The prior $\text{S}_a(T)$ estimate can be used to conduct some EEW drills until the seismic waves reach the target site. Once the seismic waves reach the target site from rupture source, phase 2 of the framework is implemented. This involves obtaining the on-site estimate of the $\text{S}_a(T)$ using the ROSERS framework which is then combined with the prior estimate through Bayesian updating
thereby sampling the more informed posterior $S_a(T)$ spectrum.

The trained framework is trained and tested using an extensive Japanese ground motion dataset and leads to high prediction power. The prediction of the complete $S_a(T)$ spectrum (while inherently maintaining the cross-correlations) in a regional setting in a highly accurate and rapid manner makes the proposed framework valuable for real-time EEW decision-making and rapid response systems.

6. REFERENCES
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, May.