12th National Conference on Earthquake Engineering
Salt Lake City, Utah
27 June - 1 July 2022
Hosted by the Earthquake Engineering Research Institute

A Deep Learning based Generalized Ground Motion Model for the Chilean Subduction Seismic Environment

J. Fayaz[1], M. Medalla[2], P. Torres-Rodas[3], and C. Galasso[4]

ABSTRACT

This paper proposes a deep learning-based generalized ground motion model (GGMM) for interface and inslab subduction earthquakes recorded in Chile. A total of ~7000 ground-motion records from ~1700 events are used to train the GGMM. Unlike common ground-motion models (GMM), which generally consider individual ground-motion intensity measures such as spectral acceleration at a given period, the proposed GGMM is a data-driven framework that coherently uses recurrent neural networks (RNN) and hierarchical mixed-effects regression to output a cross-dependent vector of 35 ground-motion intensity measures (**IM**). The **IM** vector includes geomean of Arias intensity, peak ground velocity, peak ground acceleration, and significant duration, and RotD50 spectral accelerations at 32 periods between 0.05 to 5 seconds (denoted as $S_a(T)$). The inputs to the GMM include six causal seismic source and site parameters. The statistical evaluation of the proposed GGMM shows that the proposed framework results in high prediction power with coefficient of determination R^2 > 0.7 for most IMs while maintaining the cross-IM dependencies. Furthermore, it is observed that the proposed GGMM leads to better goodness of fit for all periods of S_a(T) compared to two state-of-the-art Chilean GMMs (on average 0.2 higher R^2).

Introduction

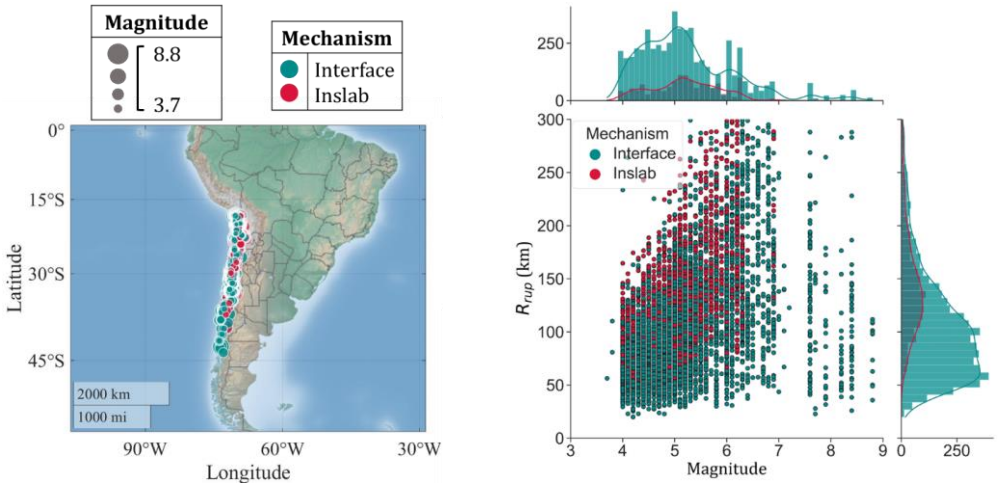
Motivated by Fayaz et al. [1], this study proposes an recurrent-neural-network- (RNN) and mixed-effects-regression- based GGMM specifically for the earthquakes recorded in Chile (subduction environment). A total of ~7000 ground motion records from ~1750 events are used to train a data-driven RNN model that uses six source and site parameters (including fault slab mechanism (F), magnitude (M_w), closest rupture distance (R_{rup}), Joyne-Boore distance (R_{JB}), soil shear-wave velocity (V_{s30}), and hypocentral depth (Z_{hyp}) to predict a vector of 35 IMs (denoted as **IM** hereafter), including geomean of Arias intensity (I_a), peak ground velocity (PGV), peak ground acceleration (PGA), significant duration (D_{5-95}) (denoted as $I_{a_{geom}}$, PGV_{geom}, PGA_{geom}, and $D_{5-95_{geom}}$, respectively), and RotD50 spectral acceleration at 32 periods between 0.05 to 5 seconds (for 5% damped oscillator; denoted as $S_a(T)$).The proposed GGMM is a general tool that can be helpful for various applications such as: 1) multi-IM-based record selection for time-history structural analysis; 2) the assessment of geotechnical structures, where the parameters of interest are commonly not the spectral coordinates $S_a(T)$, but other IMs, such as PGV [2], I_a [3]; 3) analysis and validation of simulated ground motions [4–7]; 4) vector-valued PSHA [8], etc.

[1]Research Fellow, Department of Civil, Environmental, and Geomatic Engineering, University College London, UK WC1E 6BT (email: j.fayaz@ucl.ac.uk)
[2]Lecturer, Faculty of Engineering and Applied Sciences, Universidad de los Andes, Chile
[3]Assistant Professor, Universidad San Francisco de Quito USFQ, Colegio de Ciencias e Ingenierías, Campus Cumbayá, Quito 170901, Ecuador.
[4]Professor, Department of Civil, Environmental, and Geomatic Engineering, University College London, UK WC1E 6BT

Fayaz J., Medalla M., Torres-Rodas P., Galasso C. A Deep Learning based Generalized Ground Motion Model for the Chilean Subduction Seismic Environment. *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.

Ground-Motion Database

As indicated previously, this study uses a set of ~7000 two-component ground-motion acceleration time series from ~1700 seismic events available in the SIBER-RISK strong motion database [9]. The locations of these events are shown in Figure 1a, and the corresponding $M_w - R_{rup}$ metadata is presented in Figure 1b. In terms of the two mechanisms, *i.e.*, interface and inslab, ~5000 ground motion records are from interface events, and ~2000 ground motion records are from intraslab events. As observed from Figure 1(b), the records include events with M_w ranging from 3.7 to 8.8, while R_{rup} ranges from 15 to 300 km. The statistical details of the other event parameters obtained from the metadata include i) $0.9 \leq R_{JB} \leq 298.2$ km; ii) $2 \leq Z_{hyp} \leq 213$ km; and iii) $108 \leq V_{s30} \leq 2127$ m/s.



(a) (b)

Figure 1. Earthquake events: (a) epicenters; and (b) $M_w - R_{rup}$ metadata.

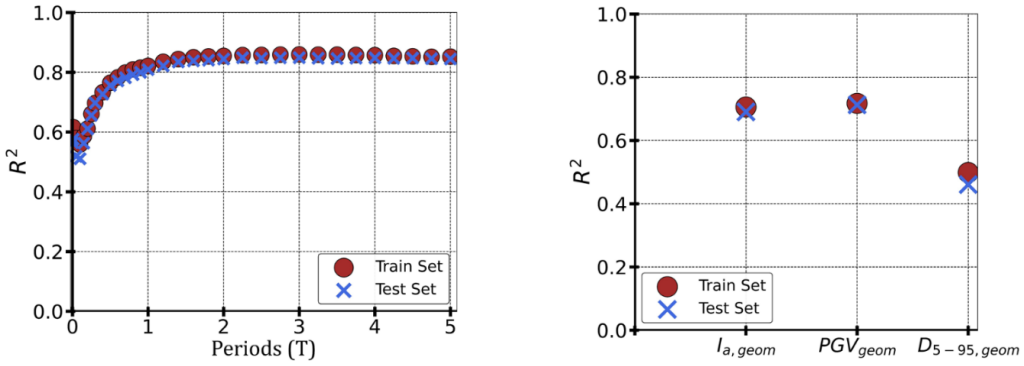
Generalized Ground Motion Model

Mean model using Recurrent Neural Networks

Similar to Fayaz *et al.* [1], this study employs long-short-term-memory (LSTM)-based RNNs to predict the considered **IM** vector and model their sequence-dependent behavior which means that the values are related to each other [10]. The neural network structure is trained with cross-validation using randomly selected 80% of the total dataset. The remaining 20% of the dataset is used as the test set, which is not used in the training of the RNN. The RNN structure is trained using six seismic source and site parameters, including F, M_w, R_{rup}, R_{JB}, V_{s30},and Z_{hyp} (collectively denoted as $\boldsymbol{\theta}$). The parameter F is inputted as a one-hot vector with [1,0] for the interface mechanism and [0,1] for the inslab mechanism. These features basically describe the physical mechanism of the earthquake source (*i.e.*, F, M_w, Z_{hyp}) and the site characteristics (*i.e.*, R_{rup}, R_{JB}, V_{s30}). Also, as it is well known fact that unlike traditional machine learning and statistical models, neural networks don't need explicit feature engineering as the weights of the neurons are automatically optimized through gradient descent depending on the sensitivity of the target variables with respect to the input features.

The prediction power of the trained RNN is analysed by comparing the measured and predicted values of the **IM** vector through the coefficient of determination R^2. The R^2 for the 32 periods of $S_a(T)$ and the other three IMs are shown in Figures 2a and 2b, respectively. It can be observed from the figures that in most cases the R^2 is above 0.7 except for $D_{5-95,geom}$ and $S_a(T)$ for very short periods of less than 0.5 seconds for both train and test sets. This is because subduction ground motions tend to have very high variabilities for $S_a(T)$ for shorter periods and significant duration, which makes any pattern recognition very difficult in such cases. As an example, Montalva *et al.* [11] indicates a total standard deviation (in log-space) close to 0.86 for the short period domain (natural periods < 0.5s) and 0.79 for natural periods between 0.5-2.5s. The R^2 values are also observed to be very close between the test and train sets, showing that the trained RNN is not overfitting the data and is capable of modelling IMs for the ground motions that are not used to train the model (*i.e.*, the test set). It should be noted that the RNN framework is not only trained to possess good prediction power for each IM but also

maintains the internal cross-dependencies within **IM**.



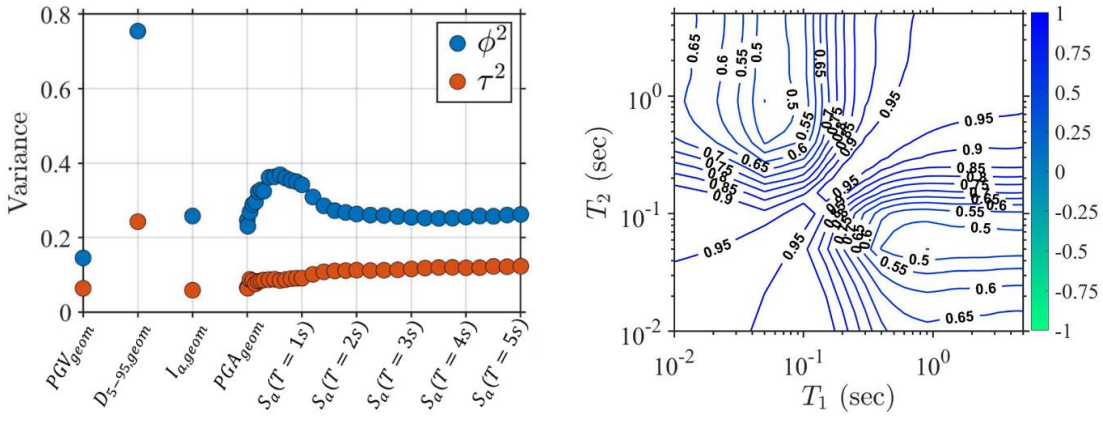
(a) (b)

Figure 2. R^2 for: a) $S_a(T)$; and b) $I_{a,geom}$, PGV_{geom}, and $D_{5-95,geom}$

Covariance Matrices for Residuals

Due to the hierarchical structure of the ground motions arising from multiple recordings of the same event at different stations and recordings from different events at the same stations, the residuals between the true log-scaled $\mathbf{IM}_{ij}$ vector and RNN-predicted log-scaled $\widehat{\mathbf{IM}}_{ij}$ vector are used to compute 35 values of between-event and within-event variabilities for i^{th} event and j^{th} recording. This is done by fitting a mixed-effects regression [12] model to the residuals as given in Equation 1 where η_i represents between-event variability with $\mathbf{T^2}$ variance matrix for the 35 IMs (with ${\tau^2}_k$ for k^{th} IM in the **IM** vector), ε_{ij} represents within-event variability with $\mathbf{\Phi}^2$ variance matrix for the 35 IMs (with ${\phi^2}_k$ for k^{th} IM in the **IM** vector) and $\boldsymbol{c_0}$ represents any pending bias in the residuals for the 35 IMs. $\boldsymbol{c_0}$ was observed to be very close to zero (failing the regression hypothesis *t*-test at 5% significance level [13]) and hence dropped in the overall analysis. Also, empirical Pearson correlations are computed for the residuals of **IM** vector, which are then used to convert the between-event and within-event variance matrices into their respective covariance matrices. The independent variances and the correlation structure (only for 32-period $S_a(T)$ spectrum) of the residuals are presented in Figures 3a and 3b, respectively. It is assumed that the correlation structure for ε_{ij} and η_i is the same. In summary, the overall RNN framework is developed for log-scaled mean predictions, and the residuals are used to construct between-event and within-event covariance matrices. The overall model (*i.e.*, mean and covariances) is called the generalized ground motion model (GGMM).

$$\mathbf{IM}_{ij} - \widehat{\mathbf{IM}}_{ij} = \boldsymbol{c_0} + \eta_i(\mathbf{0}, \mathbf{T^2}) + \varepsilon_{ij}(\mathbf{0}, \mathbf{\Phi}^2) \quad (1)$$



(a) (b)

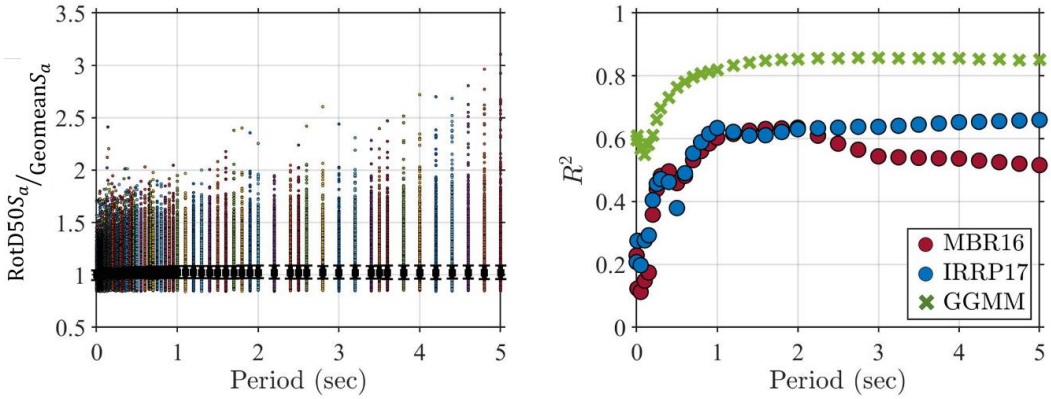
Figure 3. a) Between- and Within- Event Variances; and b) Residual Correlation Structure

Comparison against other Chilean GMMs

In this section, the spectral prediction of the proposed mean GGMM (*i.e.*, RNN predictions) is compared against two popular state of the art GMMs utilized in Chile, *i.e.*, (Montalva et al. [10]; Idini et al. [11]), denoted as MBR16 and IRRP17, respectively). However, both the GMMs were originally developed for the estimation of the geometric mean spectrum

($GeomeanS_a$). As mentioned earlier, in this study $RotD50S_a$ has been considered due to its popularity in the earthquake engineering community because of its unbiased nature and non-dependence on the orientation of sensors. Hence, to provide a fair comparison between the proposed GGMM and existing GMMs, it is essential to convert the $GeomeanS_a$ predictions into $RotD50S_a$ without addition of any additional variability. Figure 4a presents the ratios of the $RotD50S_a$ and $GeomeanS_a$ for all the ~7000 Chilean subduction ground motions for the 32 periods used in this study. It can be observed that while the mean ratios for all periods lie very close to 1, the ratios can go as high as ~3 and as low as ~0.75. Hence, $GeomeanS_a$-based GMM cannot be directly used for $RotD50S_a$ estimates. Therefore, the mean predictions of the used GMMs (i.e., MBR16 and IRRP17) (denoted as $\mu_{GeomeanS_a}$) are converted to $RotD50S_a$ with no additional variability. This is done by using Equation 2. By computing the product in Equation 2, no variability is added to the mean estimates of the GMM and leads to the mean GMM estimates of $RotD50S_a$ (denoted as $\mu_{RotD50S_a}$). This is equivalent to the GMMs developed for $RotD50S_a$ having the same variance and errors that are associated with the GMM $GeomeanS_a$.

$$\mu_{RotD50S_a} = \mu_{GeomeanS_a} \times {}^{\text{True } RotD50S_a}/_{\text{True } GeomeanS_a} \quad (2)$$



(a) (b)

Figure 4. a) Ratio of $RotD50S_a$ and $GeomeanS_a$; and b) R^2 GGMM *vs.*GMMs

Figure 4b shows the comparison of the two GMMs and the GGMM estimates using the coefficient of determination R^2. It is observed that for all periods, the proposed GGMM leads to better goodness of fit and performs better than the other two GMMs. The two GMMs are very close in their prediction power, with IRRP17 having a slight edge for longer periods. It can be further observed that the proposed GGMM, on average, leads to 0.2 higher R^2 than the other two GMMs. Hence, it can be quickly concluded that the proposed GGMM performs statistically better than both GMMs (i.e., MBR16 and IRRP17) while maintaining the internal cross-dependencies between the spectral accelerations at various periods, which is not present in the GMM estimates.

Conclusions

The RNN- and mixed-effects-regression- based GGMM proposed herein offers a robust seismic hazard-consistent tool for the Chilean subduction environment that can be used for several purposes, including structural and geotechnical design and analysis (e.g., ground motion selection using multi-objective IM criteria), risk- and reliability-based decision-making (where hazard consistent IMs are required), validation of artificial/simulated ground motion records (by checking if the simulated ground motions comply with a set of possible **IM** vector corresponding to causal parameters), etc. Tools, such as the GGMM model, will provide the users (both academic and professional engineers) a practical tool capable of estimating different IMs without the necessity to generate complicated routines to link the different tools that currently exist for the hazard consistent assessment. The GGMM will allow the users to use simple deterministic estimates of a seismic source and site parameters to generate a probabilistic distribution of the corresponding **IM** vector. The proposed GGMM can also be easily re-trained with new ground motion records or extended to a larger **IM** vector and can be appended to a seismic hazard analysis toolbox. Finally, GGMM proposed here has shown better goodness of fit regarding common GMMs, while at the same time it can predict another hazard consistent **IMs** (e.g., as $I_{a_{geom}}$, PGV_{geom}, PGA_{geom}, and $D_{5-95_{geom}}$)

References

1. Fayaz J, Xiang Y, Zareian F. Generalized ground motion prediction model using hybrid recurrent neural network. Earthq Eng Struct Dyn 2021;50:1539–61.
2. Bommer JJ, Alarcon JE. The prediction and use of peak ground velocity. J Earthq Eng 2006;10:1–31.
3. Macedo J, Abrahamson N, Bray JD. Arias intensity conditional scaling ground-motion models for subduction zones. Bull Seismol Soc Am 2019;109:1343–57.
4. Petrone F, Abrahamson N, McCallen D, Miah M. Validation of (not-historical) large-event near-fault ground-motion simulations for use in civil engineering applications. Earthq Eng Struct Dyn 2021;50:116–34.
5. Ming D, Huang C, Peters GW, Galasso C. An advanced estimation algorithm for ground-motion models with spatial correlation. Bull Seismol Soc Am 2019;109:541–66.
6. Tsioulou A, Taflanidis AA, Galasso C. Validation of stochastic ground motion model modification by comparison to seismic demand of recorded ground motions. Bull Earthq Eng 2019;17:2871–98.
7. Fayaz J, Azar S, Dabaghi M, Zareian F. Methodology for validation of simulated ground motions for seismic response assessment: Application to CyberShake source-based ground motions. Bull Seismol Soc Am 2021;111:226–41.
8. Baker, J. W., and C. A. Cornell. "Vector-Valued Ground Motion Intensity Measures for Probabilistic Seismic Demand Analysis, PEER Report 2006/08, Pacific Earthquake Engineering Research Center." *University of California, Berkeley* (2006).
9. Castro, Sebastián, Roberto Benavente, Jorge Gustavo Federico Crempien De La Carrera, Gabriel Candia, and Juan Carlos de la Llera Martin. A Consistently Processed Strong Motion Database for Chilean Earthquakes. 2021..
10. Schmidhuber J, Hochreiter S. Guessing can outperform many long time lag algorithms 1996.
11. Montalva GA, Bastías N, Rodriguez-Marek A. Ground-motion prediction equation for the Chilean subduction zone. Bull Seismol Soc Am 2017;107:901–11.
12. Demidenko E, Stukel TA. Influence analysis for linear mixed-effects models. Stat Med 2005;24:893–909.
13. Sheynin O. Helmert's work in the theory of errors. Arch Hist Exact Sci 1995;49:73–104. https://doi.org/10.1007/BF00374700.
14. Idini B, Rojas F, Ruiz S, Pastén C. Ground motion prediction equations for the Chilean subduction zone. Bull Earthq Eng 2017;15:1853–80.