EGU2020-6911, updated on 23 Oct 2020
https://doi.org/10.5194/egusphere-egu2020-6911
EGU General Assembly 2020

Exploring the dimensionality of ground motion data

Reza Dokht Dolatabadi Esfahani[1,2], Kristin Vogel[1], Fabrice Cotton[1,2], Matthias Ohrnberger[1], Frank Scherbaum[1], and Marius Kriegerowski[2]
[1]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany (dokhtdolatabadie@uni-potsdam.de)
[2]GFZ German Research Center for Geociences, Section 2.6, Potsdam, Germany

For years, engineering seismologists aim to reduce the epistemic uncertainty related to ground motion prediction. Assuming that simple models with few variables are not sufficient to describe the complex phenomena, there is a trend in present-day science to increase complexity of ground motion models. Therefore, some of the most recent ground motion prediction equations use more than 20 variables to improve the predictive power of the model. However, the legitimate question to ask is whether the inclusion of additional variables leads to an improved predictive power of the model. In other words, what is the smallest number of predictive variables needed to reconstruct the distribution of ground motion induced shaking observed in data? In this study, by taking advantage of the exponential growth of ground motion data and new machine learning methods, we present a data-driven approach to derive the dimensionality of ground motion data in the Fourier amplitude spectrum (FAS) metric. We apply an autoencoder architecture, which is commonly used for mapping high dimensional data to a lower dimensional space (bottleneck) and search for the lowest dimensionality (minimum number of nodes in the bottleneck) required to reconstruct the FAS input data. The approach is tested on synthetic ground motion data with known dimensionality (2D and 4D) and finally applied to the FAS of recorded ground motion data. A simple autoencoder with variable nodes in the bottleneck is used to explore the dimensionality of the ground motion data. We use the relation between the total residual of the network with the number of codes in the bottleneck as an indicator of dimensionality. Its numerical value is estimated based on the reduction of residuals by increasing the number of codes in the bottleneck layer. In addition, we use the low dimensional manifold of the ground motion data to predict the ground motion shaking for a given scenario. The residual analyses between observed and reconstructed data and observed and predicted data are used to validate the training and prediction steps. We applied the method on different scenarios in two synthetic data sets which are simulated by a stochastic simulation method and secondly the Pan-European engineering strong motion data (EMS) to show the performance of the proposed method. The results show that the statistical properties of ground motion data can be captured by using a limited number of three to five parameters. Especially for low frequency data the most dominant features are already captured by two parameters (codes), which roughly correspond to magnitude and distance. For higher frequencies additional parameters, e.g. corresponding to stress drop and kappa, become more relevant. The standard deviation of the residuals can be reduced to its lower

bound in comparison with the standard deviations of conventional methods. Finally, we use a two-dimensional manifold to predict the FAS for given magnitude and distance values.