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), Marius Kriegerowski (2)

(1) University of Potsdam, Potsdam, Germany,
(2) GFZ, Potsdam, Germany

EGU General Assembly 2020

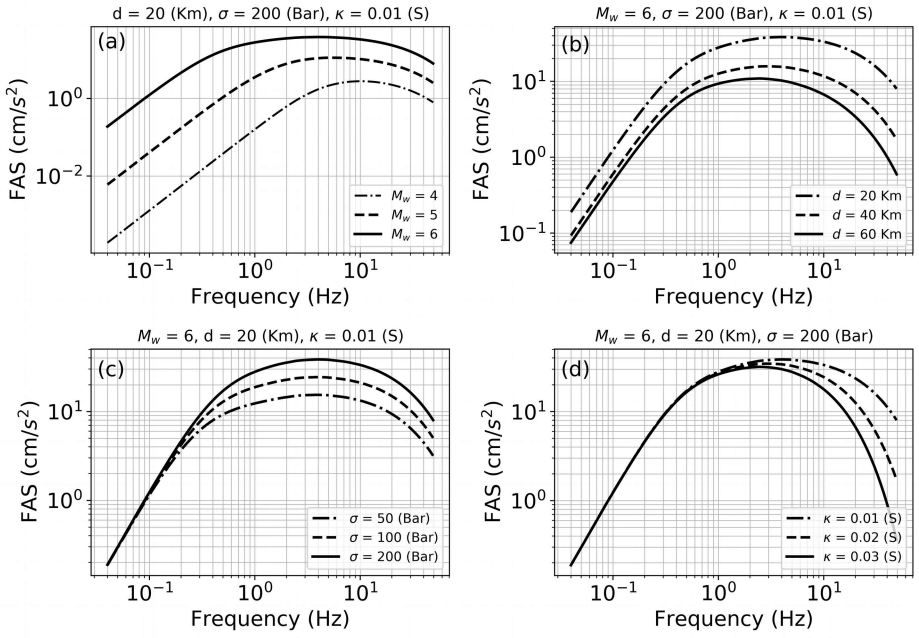
Universität Potsdam

Natural Hazards and Risks in a Changing World

GFZ Helmholtz-Zentrum Potsdam

Background: Beyond GMPE

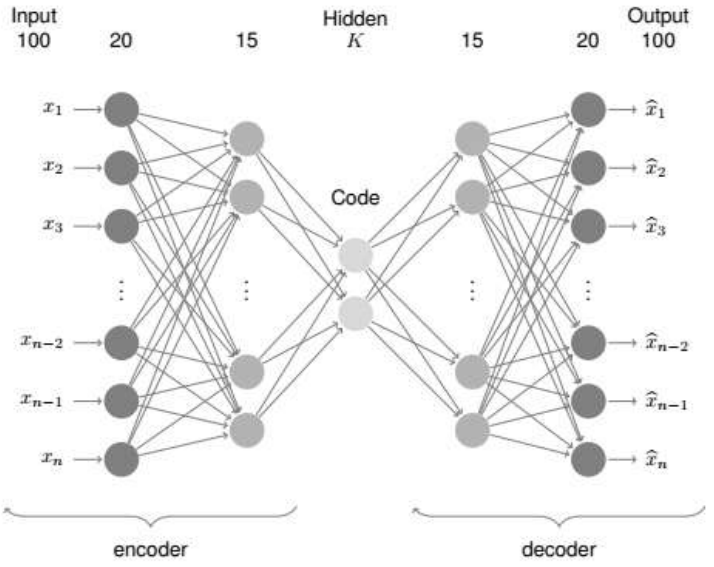
- The ground motion are commonly predicted by using regression analysis as a function of set of predictive variables.
- Considering the complexity of the involved physical processes, the selection of the set of predictive variables is a major challenge.
- In Fourier amplitude spectrum (FAS), different parameters (e.g., Distance, and Magnitude) effect different frequency range of FAS (Fig a-d).



Question: What is the **smallest number** of predictive variables needed to reconstruct the distribution of ground motion data?

Method: Using autoencoder to reduce the dimensionality of FAS

- Autoencoder are a class of artificial neural networks contains encoder and decoder.
- Encoder maps the data to a low dimensional manifold- the code and decoder reconstructs the orginal data from the manifold.
- The code can be understood as a bottleneck, that allows only a limited amount of information to pass through.
- The size of the code, k, which is the number of nodes in the code layer, corresponds the dimensionality that can be reconstructed in the output data.
- To determine the dimensionality of the input data, we increase k successively from one to nine and plot the total residual between input and output data at the y-axis against the used code size at the x-axis



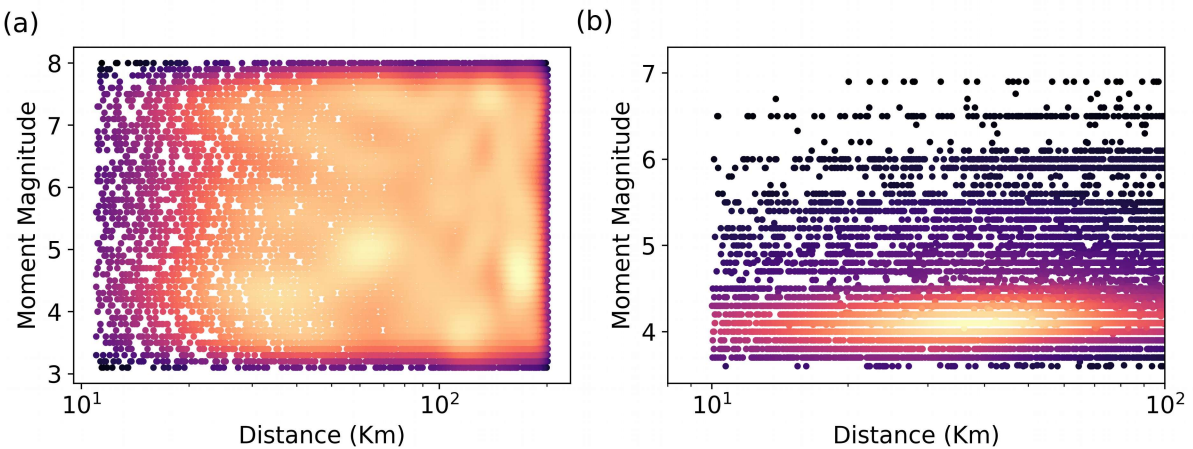
Dataset: Synthetic and real data

Using two synthetic data simulated by Boore method (2003):

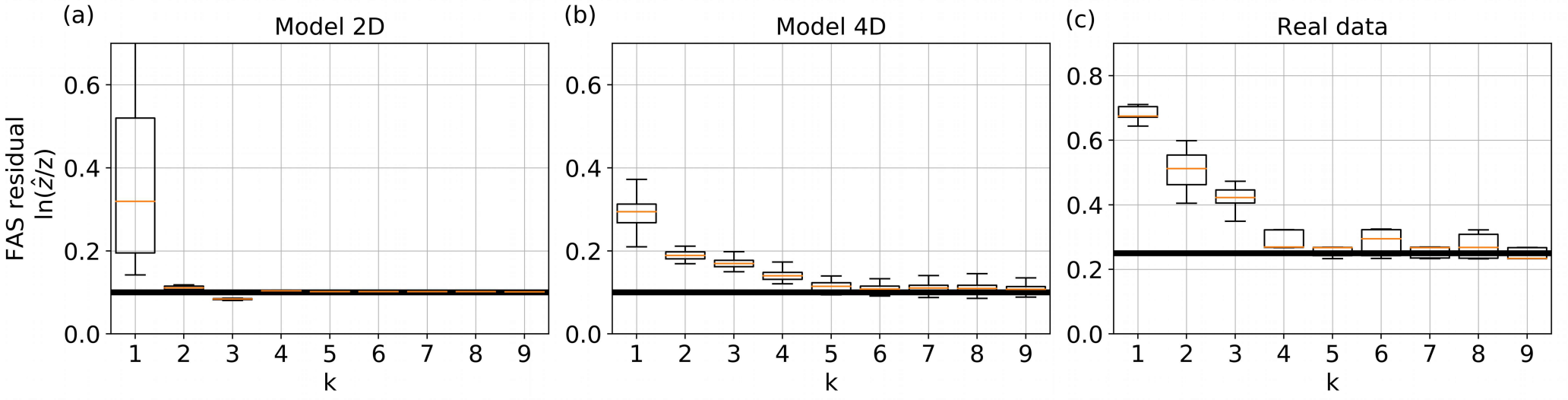
- The Brune model is used as a source spectrum depends on two parameters: the moment magnitude and the stress drop.
- The spectral attenuation depends on the distance between source and recorded site.
- The site amplification depends on high frequency attenuation.

Dataset	Dimension	Magnitude	Distance (km)	Stress drop (bar)	κ_0 (s)	size
2D	2	3 - 8	10 - 200	150	0.02	10 000
4D	4	3 - 8	10 - 200	50 - 150	0.01-0.03	10 000
ESM	unknown	3.5 - 7	10 - 100	usually unknown		9 637

Magnitude-distance density of (a) synthetic, and (b) real data.



Results: Dimensionality of FAS

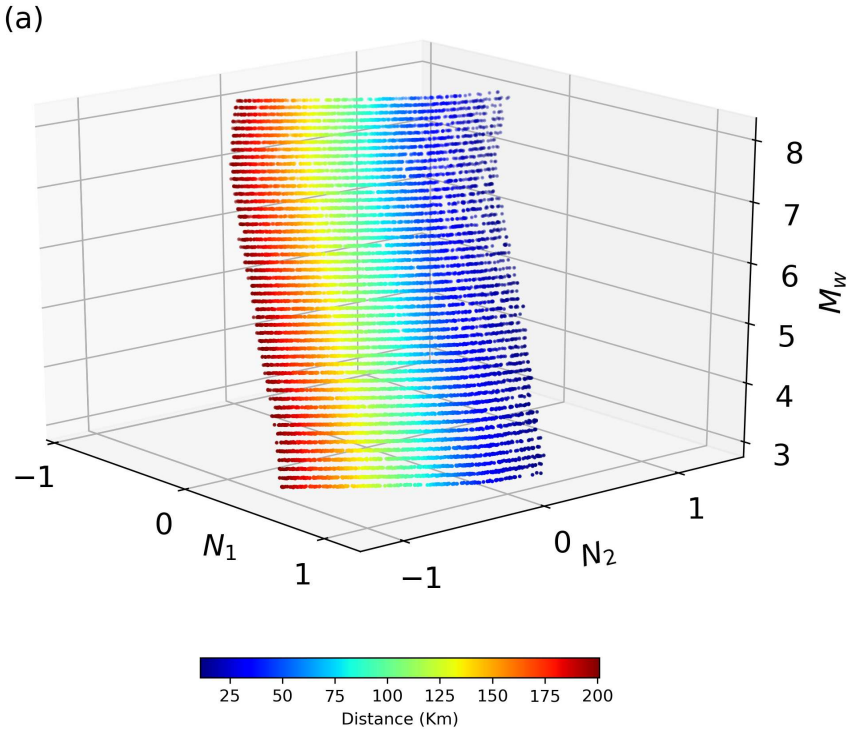


- **The results of residual analysis based on the number of bottleneck codes show two codes in the bottleneck can capture the dominant features of FAS.**

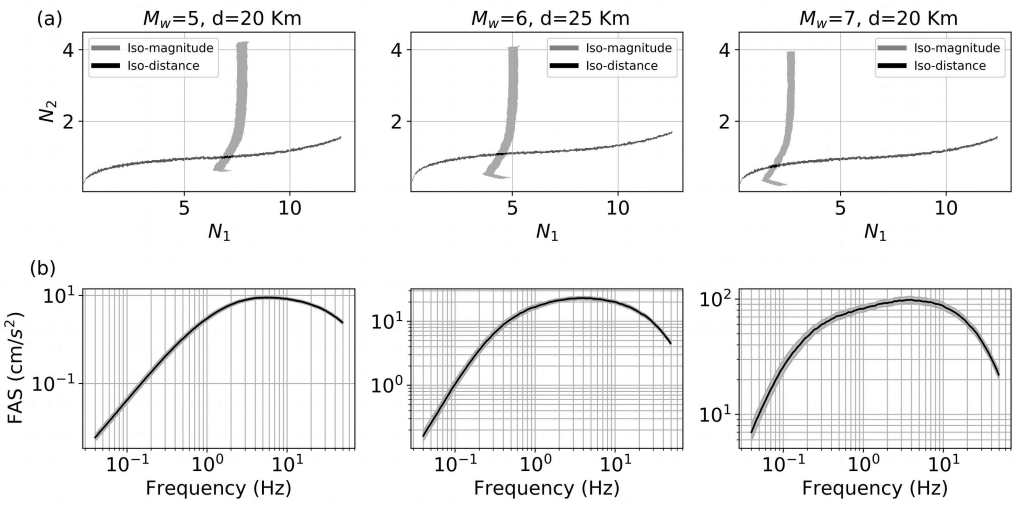
- **The real data is much more complex to capture with two codes.**

Results: Synthetic data (2D)

- The bottleneck visualization for 2D data. N1 and N2 is the code values. Mw is moment magnitude and color shows distance.

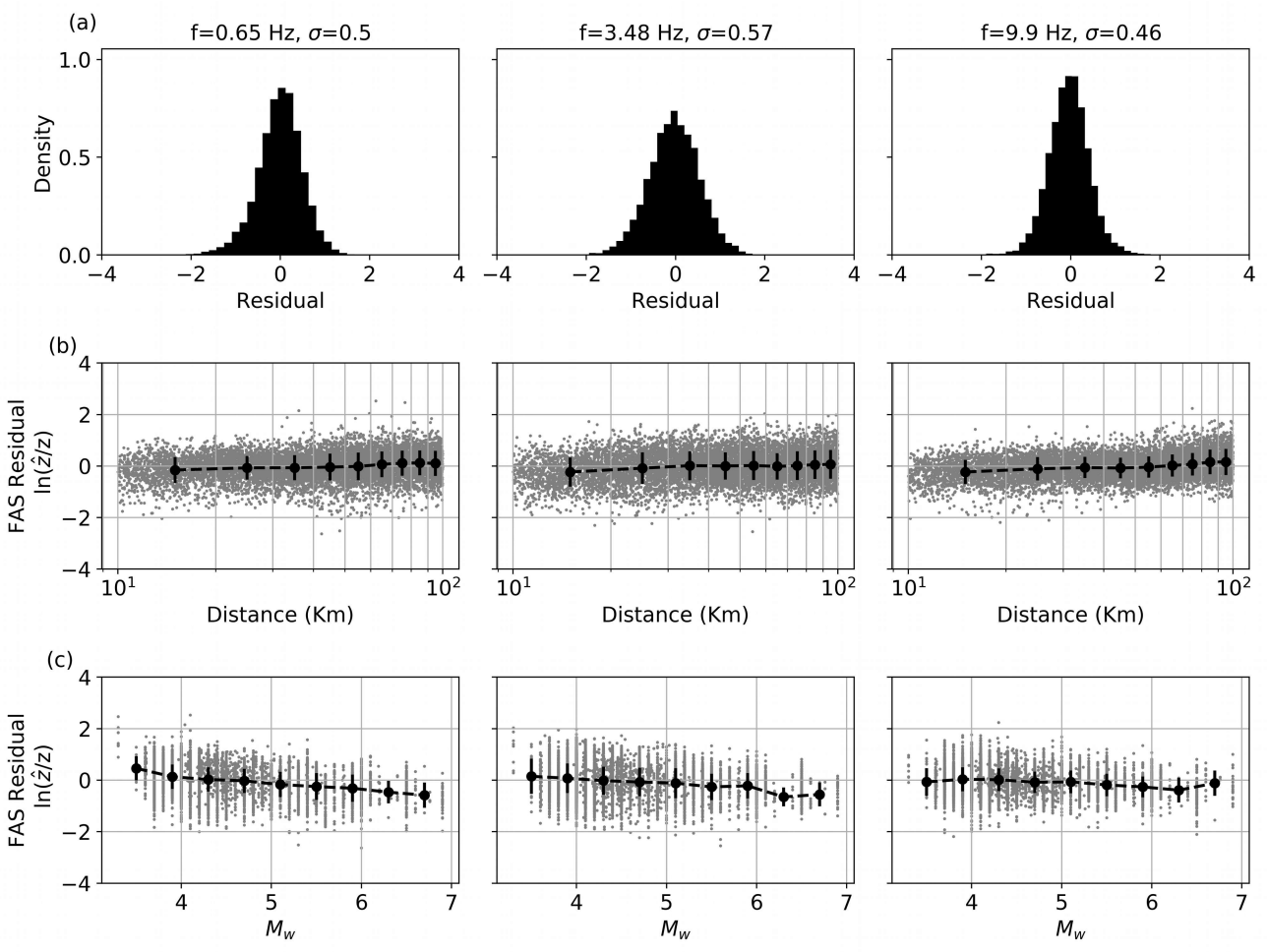


- The ground motion prediction based on the bottleneck values (Mw, Distance) and decoder.



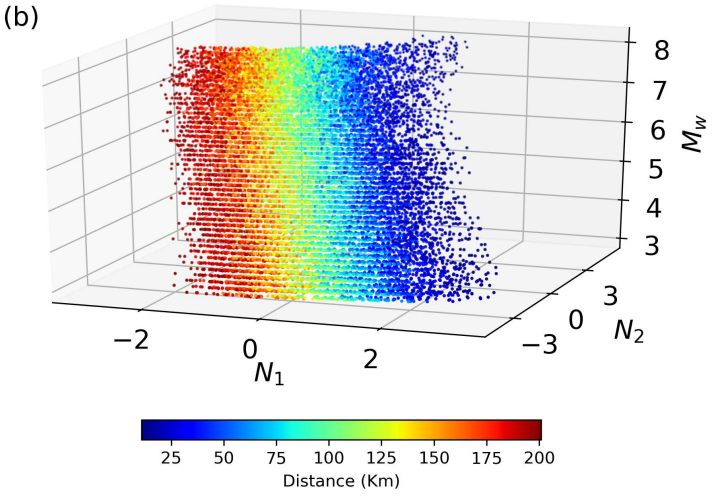
Results: Residual analysis of 2D synthetic data

- The residual analysis between observed and predicted FAS. (a) The histogram of residual. The variation of residual with respect to the model parameters (b) distance, and (c) moment magnitude.

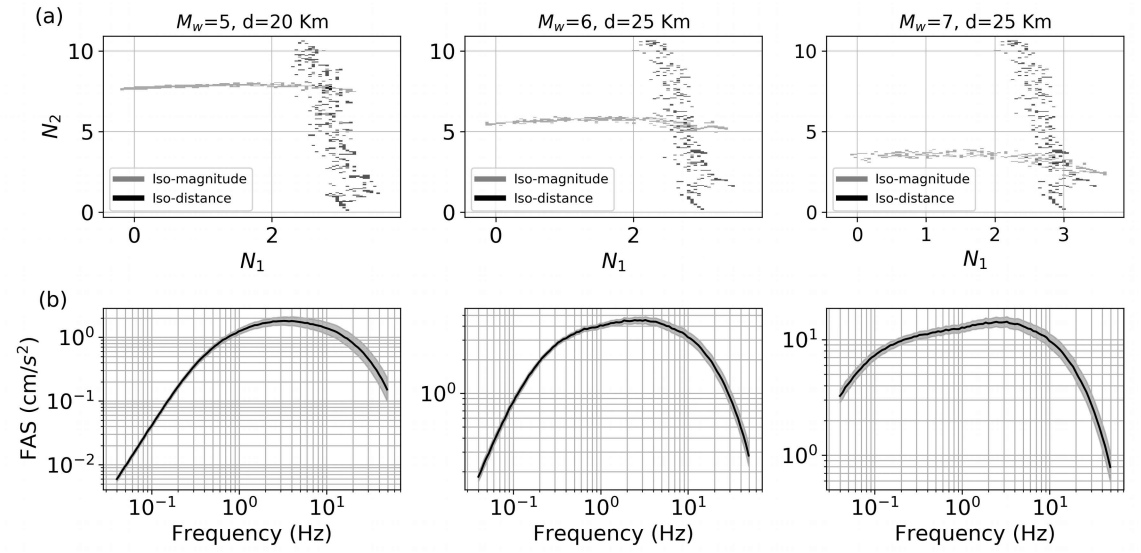


Results: synthetic data (4D)

- The bottleneck visualization for 4D data. N1 and N2 is the code values. Mw is moment magnitude and color shows distance.

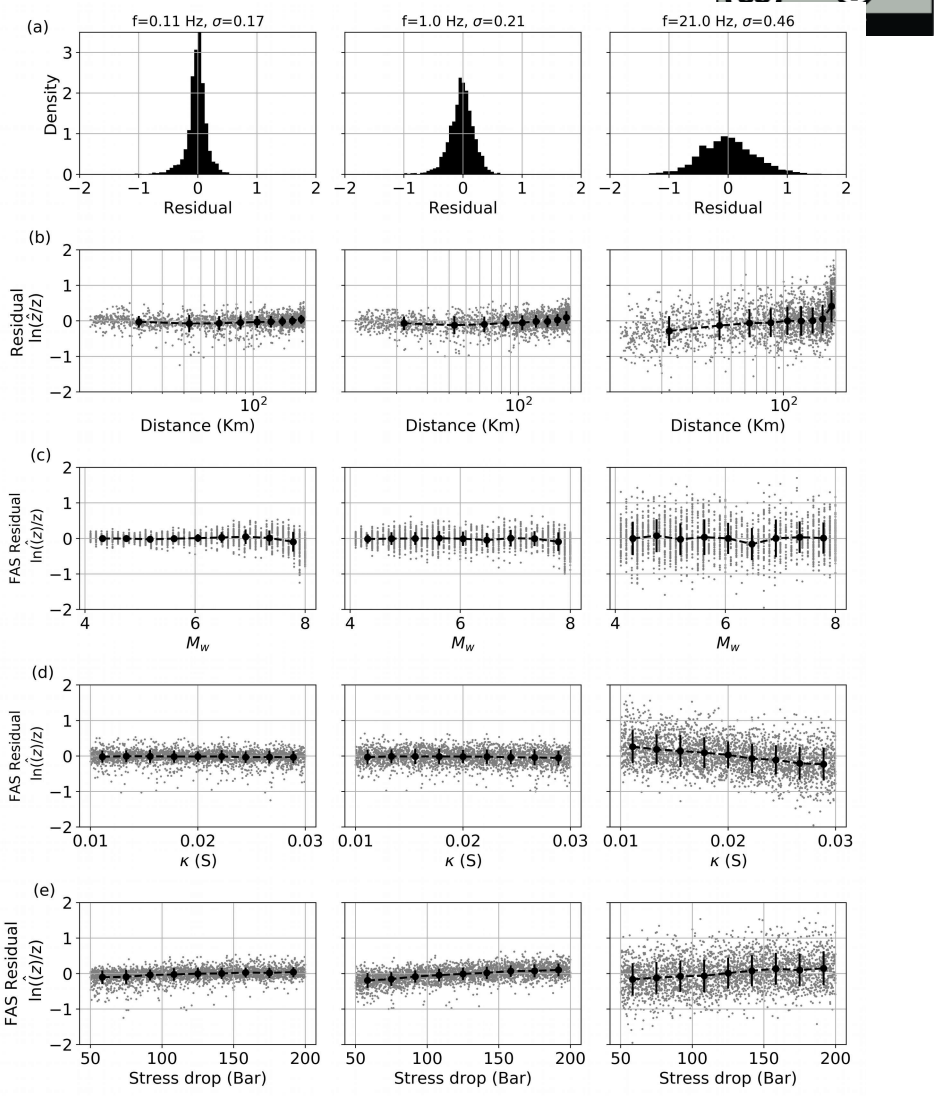


- The ground motion prediction based on the bottleneck values (Mw, Distance) and decoder.



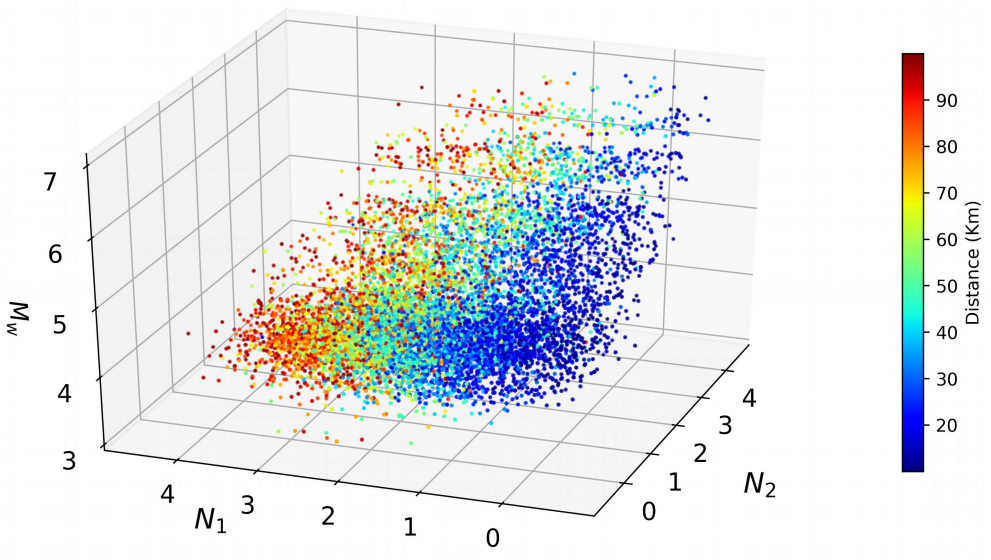
Results: Residual analysis of 4D synthetic data

- The residual analysis between observed and predicted FAS. (a) The histogram of residual. The variation of residual with respect to the model parameters (b) distance, (c) moment magnitude, (d) kappa, and (e) stress drop.

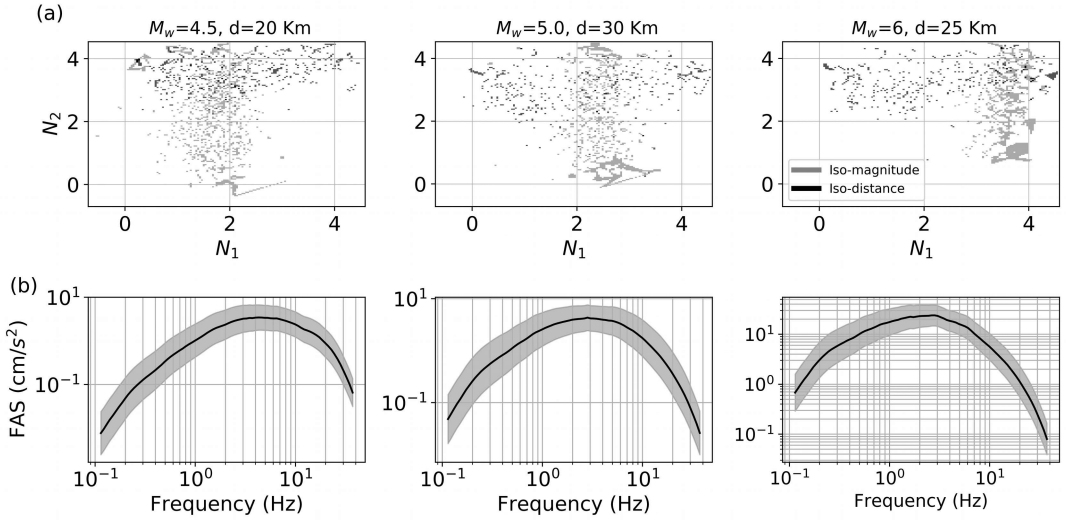


Results: Real data

- The bottleneck visualization for real data. N1 and N2 is the code values. Mw is moment magnitude and color shows distance.



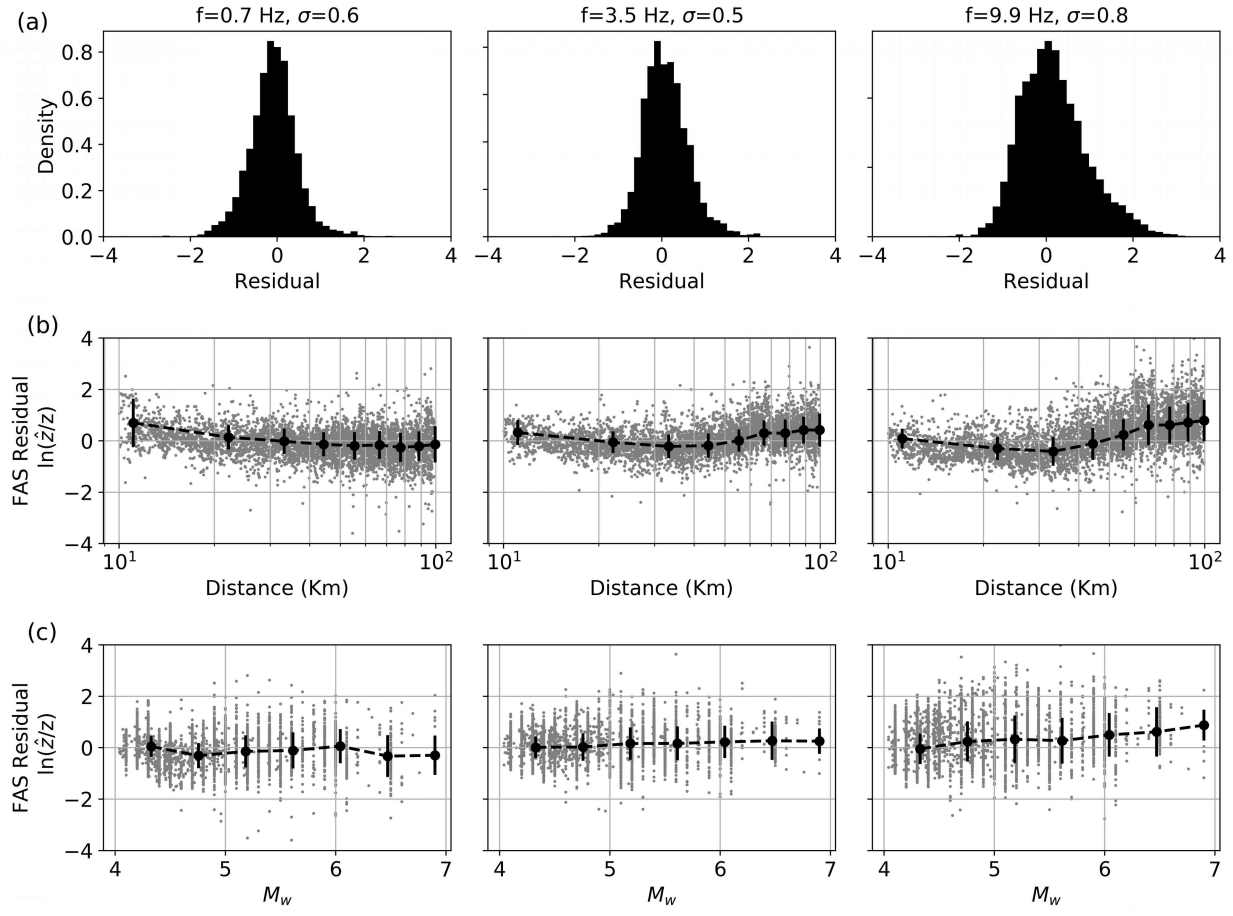
- The ground motion prediction based on the bottleneck values (Mw, Distance) and decoder.



Results: Residual analysis of real data

- The residual analysis between observed and predicted FAS. (a) The histogram of residual. The variation of residual with respect to the model parameters (b) distance, and (c) moment magnitude.

- Two codes can not capture information in the high frequency range of FAS.



Conclusions

- A data-driven method is proposed for analyzing the dimensionality of ground motion data

- The results show that the ground motions are controlled mainly by **two parameters**. However two **additional parameters** are needed to fully capture ground shaking.

- The comparison between observed and predicted data shows that the **aleatory variable** of proposed method is closed to classical analysis.