Prediction of Ground-Motion Time-Series at an arbitrary location using Gaussian Process Interpolation: Application to the Ridgecrest Earthquake

Aidin Tamhidi[1]; Nicolas Kuehn[1]; Yousef Bozorgnia[1]; Ertugrul Taciroglu[1]; Tadahiro Kishida[2]

[1] University of California, Los Angeles, USA; [2] Khalifa University, UAE

Contact: Aidin Tamhidi (aidintamhidi@ucla.edu)

Abstract

- In many engineering applications a variety of seismic intensity measures (IMs) are required as input, that are estimated either through ground motion prediction equations (GMPEs) or spatial interpolations of their recorded values.
- To understand dynamic responses of the structures, estimation of the entire ground motion (GM) time-series at specific geographic locations is required as well.
- In the present study, we examine the accuracy of the Gaussian Process (GP) interpolation method, by predicting GM time series at an arbitrary location using available spatially sparse records.
- To achieve a quantitative assessment, we used simulated GM time series of the 1906 San Francisco earthquake. Our current work is aimed at applying the developed and verified procedure on GM time series recorded during the Ridgecrest 2019 earthquake.

Methodology

Each GM time series is transformed into the frequency domain to be used in GP interpolation.

↓

We place 6 equally spaced boundaries for frequencies in logarithmic scale to construct 5 frequency windows

↓

We would pass the values of real and imaginary parts of each frequency window at different geographic locations into the GP interpolation model

↓

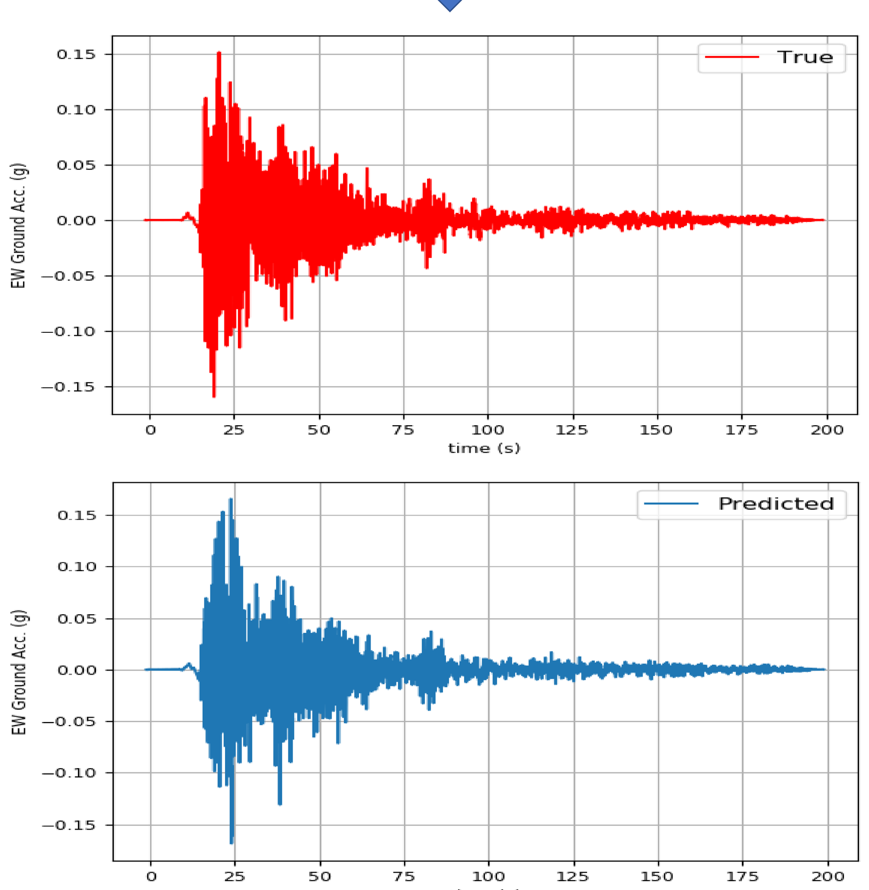
GP regression method would optimize the hyperparameters of the model through maximizing the marginal log-likelihood of the observed data points

↓

Optimized GP interpolation model would predict the real and imaginary parts of the corresponding frequencies at the desired location

↓

The predicted frequency domain content of the target location will be inverted into the time domain to construct the GM time series.



I. Gaussian Process

- A Gaussian Process defines a prior over functions. After having observed some function values it can be converted into a posterior over functions.
- In a GP, any point $x \in R^d$ is assigned a random variable $f(x)$ where the joint distribution of a finite number of these variables $p(f(X_1), \dots, f(X_N))$ is itself Gaussian.

$$p(f|X) = N(f|\mu, K) \quad (1)$$

Where $\mu = (m(X_1), \dots, m(X_N))$ and $K_{ij} = k(x_i, x_j)$.

- After having observed some data y, the posterior can then be used to make predictions f_* given new Input X_*:

$$p(f_*|X_*, X, y) = N(f_*|\mu_*, \Sigma_*) \quad (2)$$

$$\begin{pmatrix} y \\ f_* \end{pmatrix} \sim N(\mu, \begin{pmatrix} K_y & K_* \\ K_*^T & K_{**} \end{pmatrix}) \quad (3)$$

By definition of the GP, the mean vector and covariance matrix at a set of unobserved locations can be computed with [1][2]:

$$\mu_* = K_*^T K_y^{-1} y \quad (4)$$

$$\Sigma_* = K_{**} - K_*^T K_y^{-1} K_* \quad (5)$$

II. Prediction Model

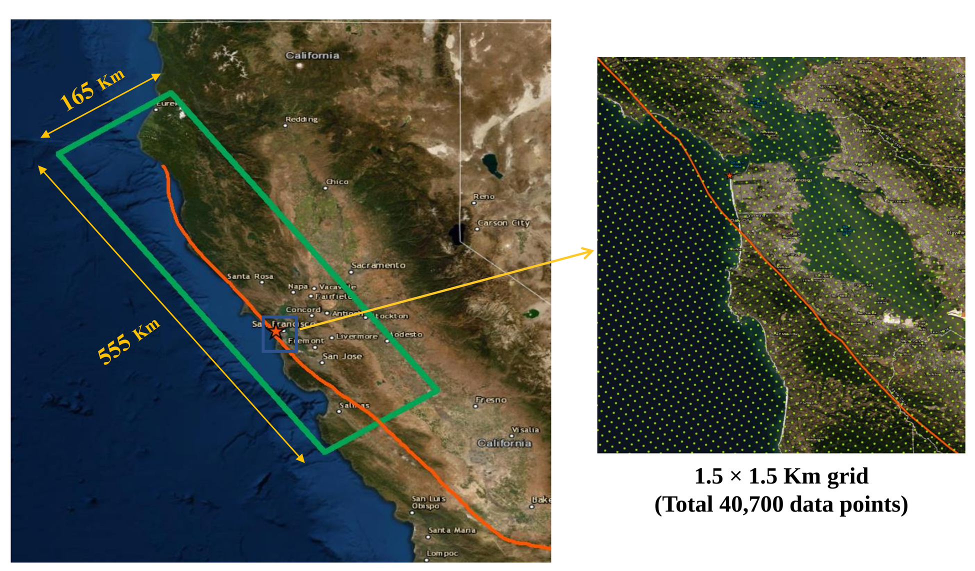
In this study, Matern kernel is used as kernel function in GP interpolation method. Equation 6 shows Matern kernel function.

$$k(x_i, x_j) = \sigma^2 \left(1 + \gamma\sqrt{3}d\left(\frac{x_i}{l}, \frac{x_j}{l}\right)\right) \exp(-\gamma\sqrt{3}d\left(\frac{x_i}{l}, \frac{x_j}{l}\right)) \quad (6)$$

σ and l are hyperparameters which need to be optimized based on the dataset.

Data

- **San Francisco 1906 Simulated Ground motions** [3].



1.5 × 1.5 Km grid (Total 40,700 data points)

Fig. 1, Bounding Box (green rectangular box) of the domain used by Aagard ground-motion modeling group

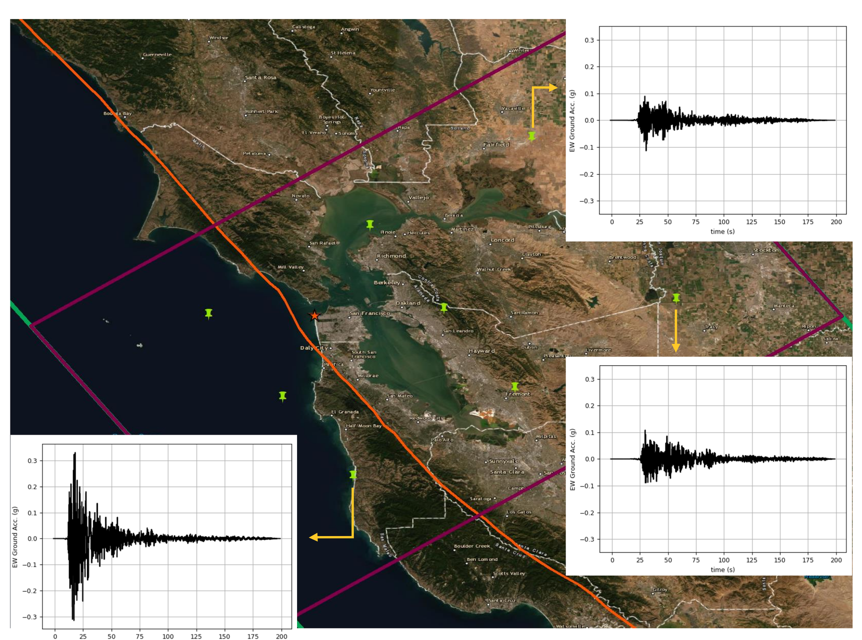


Fig. 2, Simulated ground-motion time-series in 1906 San Francisco event by Aagard GM modeling group.

- In order to be able to use GP regression method, prior must follow Gaussian distribution. In our study, real and imaginary parts of frequency content over different geographic locations must follow Gaussian distribution.

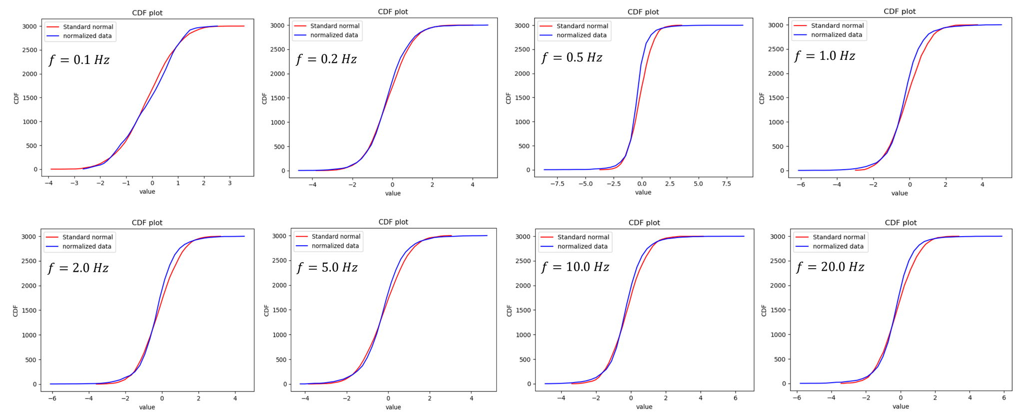
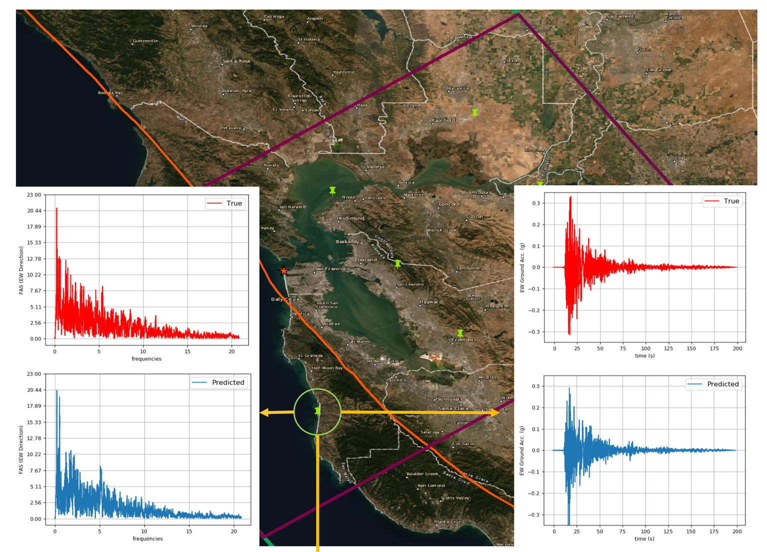


Fig. 3, CDF plots of the real parts of the frequency content at $f = 0.1, 0.2, 0.5, 1, 2, 5, 10, 20\ Hz$ for EW GM time series considering 3000 data points in radius of neighborhood R ≈ 45 Km.

Results

I. San Francisco 1906

- We have used aforementioned model to be trained on the simulated San Francisco 1906 earthquake GM time series.
- Boundaries of the length scale hyperparameter are determined through minimizing RotD50 RMSE between Predicted and True GM time series at different geographic locations.
- Radius of neighborhood around each target point, R = 8.5 Km, (Number of neighbors = 100.)



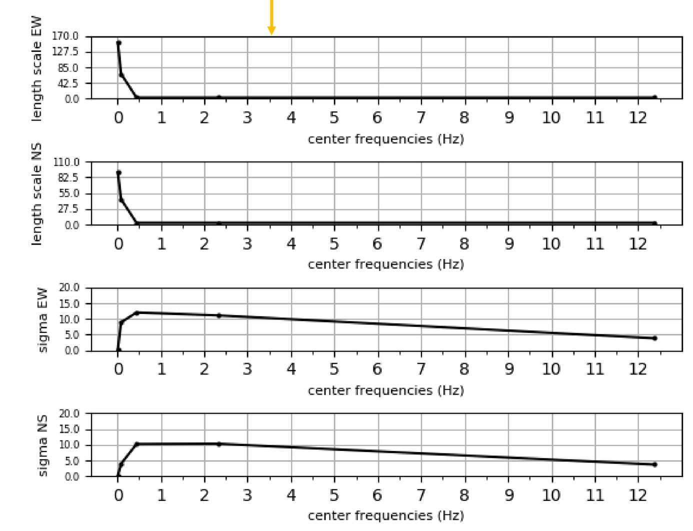


Fig. 4, Trained model hyperparameters (lower bound for length scales = 3.0)

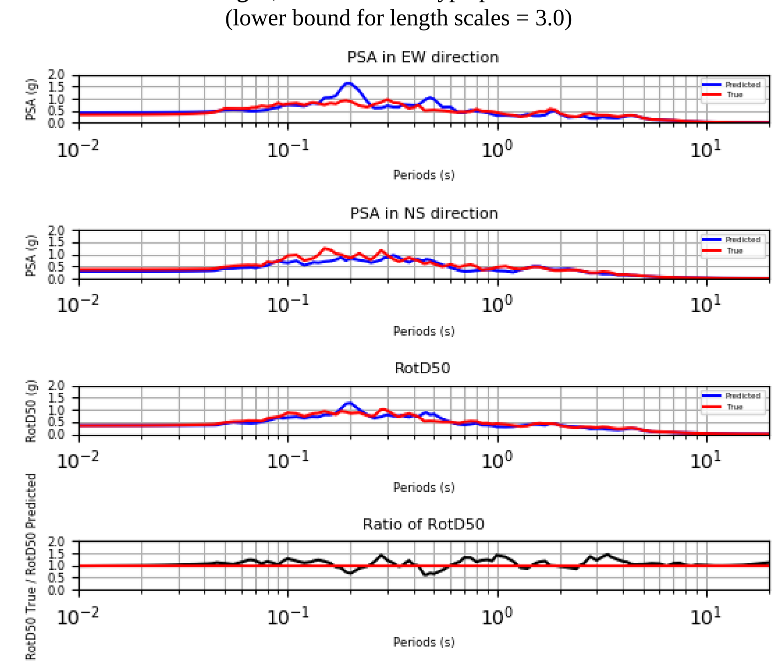
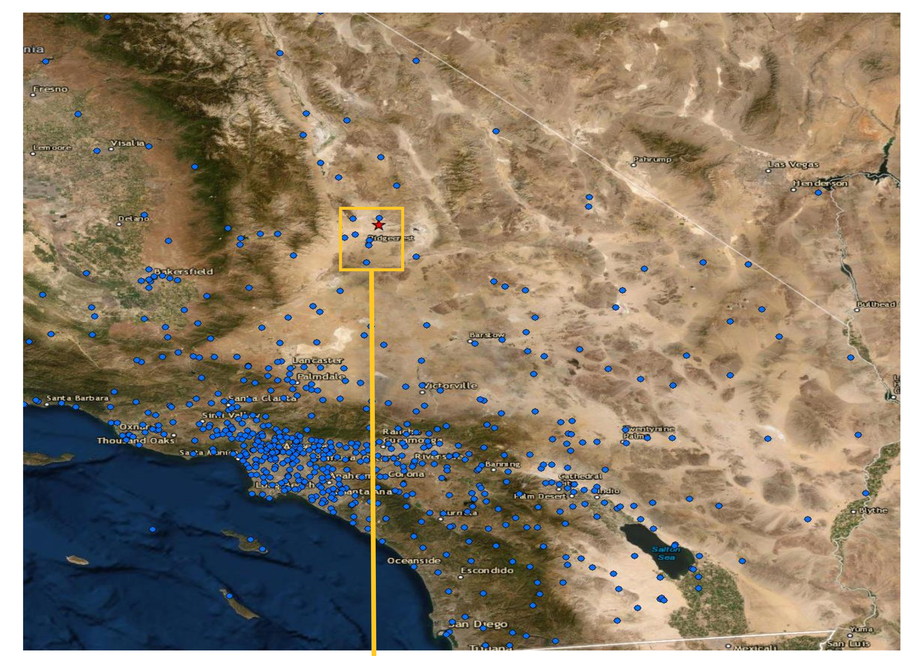
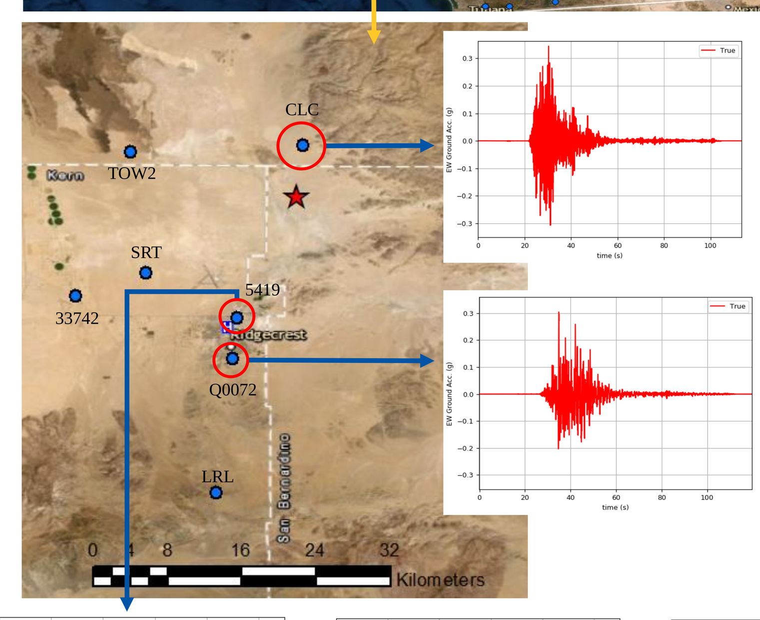


Fig. 5, PSA and RotD50 spectrums of Predicted and True GM time series

II. Ridgecrest 2019

- Ridgecrest M7.1
- Using trained model on San Francisco 1906 simulated data, we would predict the GM time-series at the Ridgecrest Hospital location (Radius of Neighborhood = 20 Km).





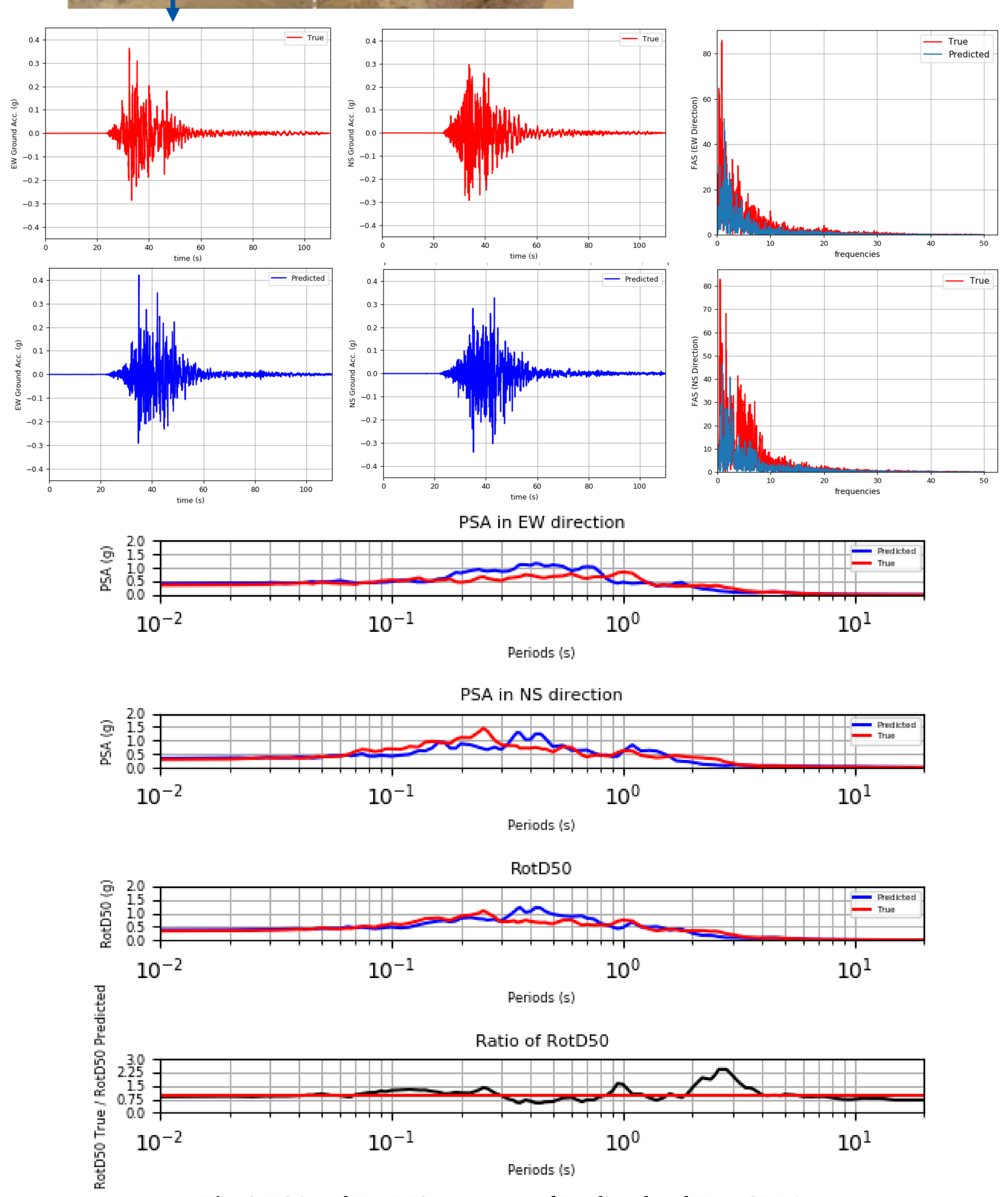
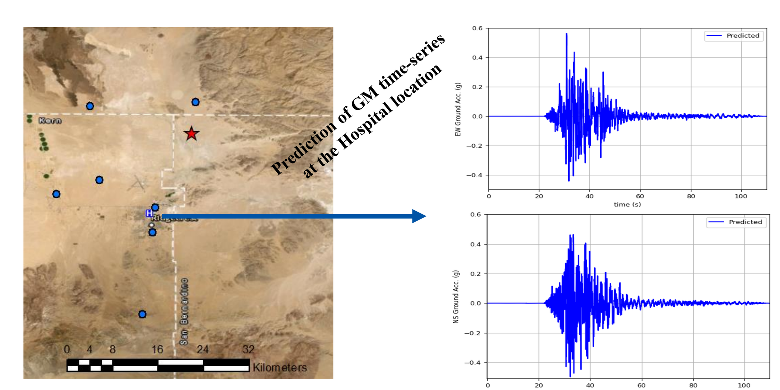


Fig. 6, PSA and RotD50 spectrums of Predicted and True GM time series for station "5419"



Summary

- Trained model shows higher l values for lower frequencies, which leads to smoother prediction functions
- After a specific frequencies, optimization leads to lower l values for higher frequencies, which make prediction functions more wiggly with wide confidence intervals between training data points
- Trained model could predict reasonable results for periods up to 0.2 s

References

1- Williams, Christopher KI, and Carl Edward Rasmussen. Gaussian processes for machine learning. Vol. 2. No. 3. Cambridge, MA: MIT press, 2006.

2- Krasser, Martin, "Gaussian Processes", GitHub, 19 March 2018, krasserm.github.io/2018/03/19/gaussian-processes/

3- Aagaard, Brad T., et al. "Ground-motion modeling of the 1906 San Francisco earthquake, part II: Ground-motion estimates for the 1906 earthquake and scenario events." Bulletin of the Seismological Society of America 98.2 (2008): 1012-1046

Acknowledgements

We would like to thank Dr. Robert W. Graves for providing us with 40,700 broadband simulated ground-motion time-series for San Francisco 1906 earthquake.