DeepQuake

Artificial Intelligence for Earthquake Forecasting Using Fine-Grained Climate Data

Yash Narayan, Jeonghyeop Kim, William Holt

THE NUEVA SCHOOL

Stony Brook University

Background

- Every year, ~20,000 earthquakes occur worldwide
- In the last 20 years, earthquakes have caused 750,000 deaths
- More than 125 million people have been displaced or injured
- Earthquakes usually occur without much warning
- Trigger other natural disasters, e.g., tsunamis and landslides

Traditional physics-based approaches

- Seismic researchers have tried to identify predictors for earthquakes with little success; current methods inaccurate and unreliable

- Relationships between the occurrence of an earthquake and seismic features of earthquake are **complex** and **highly non-linear**

- In recent years, AI approaches have started gaining interest

Connection: Climate and Earthquakes

- Recent studies have established climate variables can cause deformations within the Earth's crust, leading to earthquakes

- For example, California's wet winters creates extra hydrologic load leading to the lowest vertical heights of the elastic crust

- While dryer summer months lead to the loss of surface loads, causing the crust to typically reaches the highest vertical point.

DeepQuake: hybrid physics and AI-based approach to forecast earthquakes
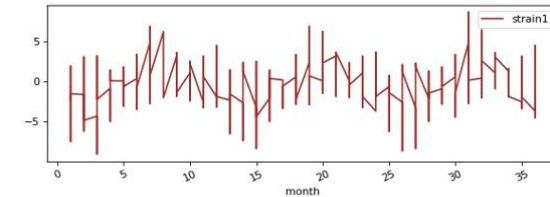
- Physical model: inverts GPS measurements into a time series of month-over-month horizontal strain values

- Deep learning model: feed this data into Recurrent Neural Network (RNN) to provide earthquake forecasts
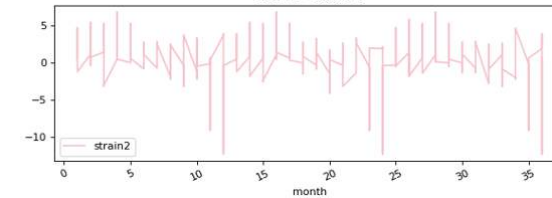
Input Dataset

Inverted cGPS into time-series data of month-over-month horizontal displacement strain at each cell

- Strain x direction
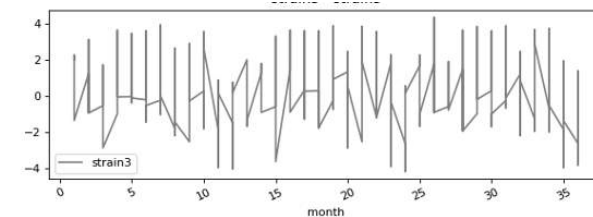- Strain y direction
- Strain z direction

Figure 1: Input dataset (strain in x, y, z direction) versus time

Ground Truth Dataset

Original Earthquake Catalog: 2085 earthquakes near Napa Valley & Long Valley Caldera from 2007-2019 (magnitude, depth, and location)
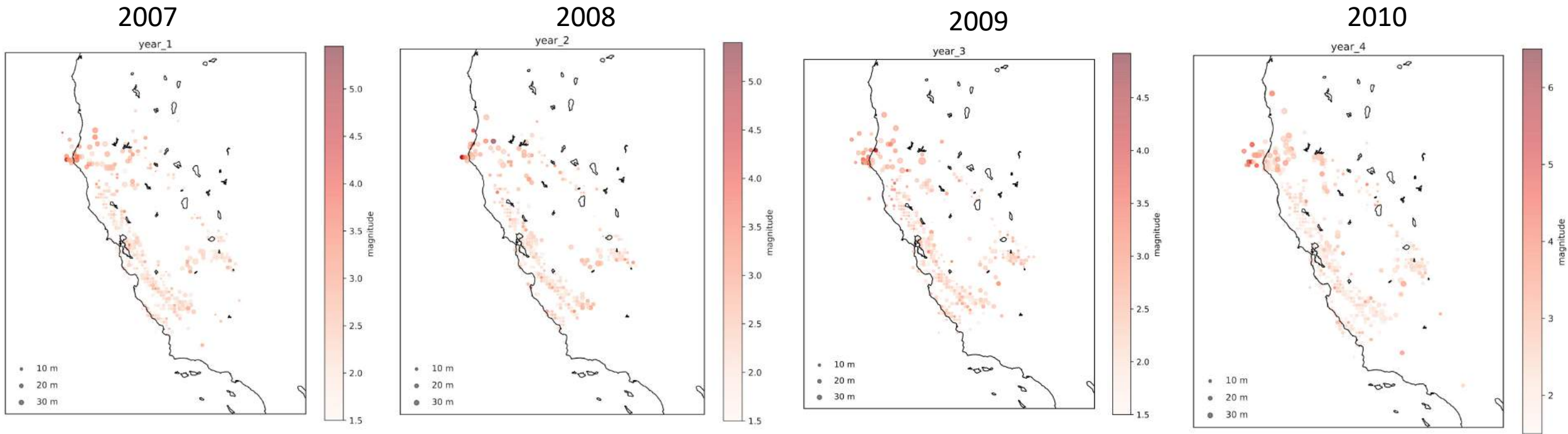

Figure 2: Year-over-year earthquake catalog (magnitude, depth, and location) near Long Valley Caldera

Preprocessing algorithm

Filter this catalog to evaluate localized regions near faults that have a high density of earthquake
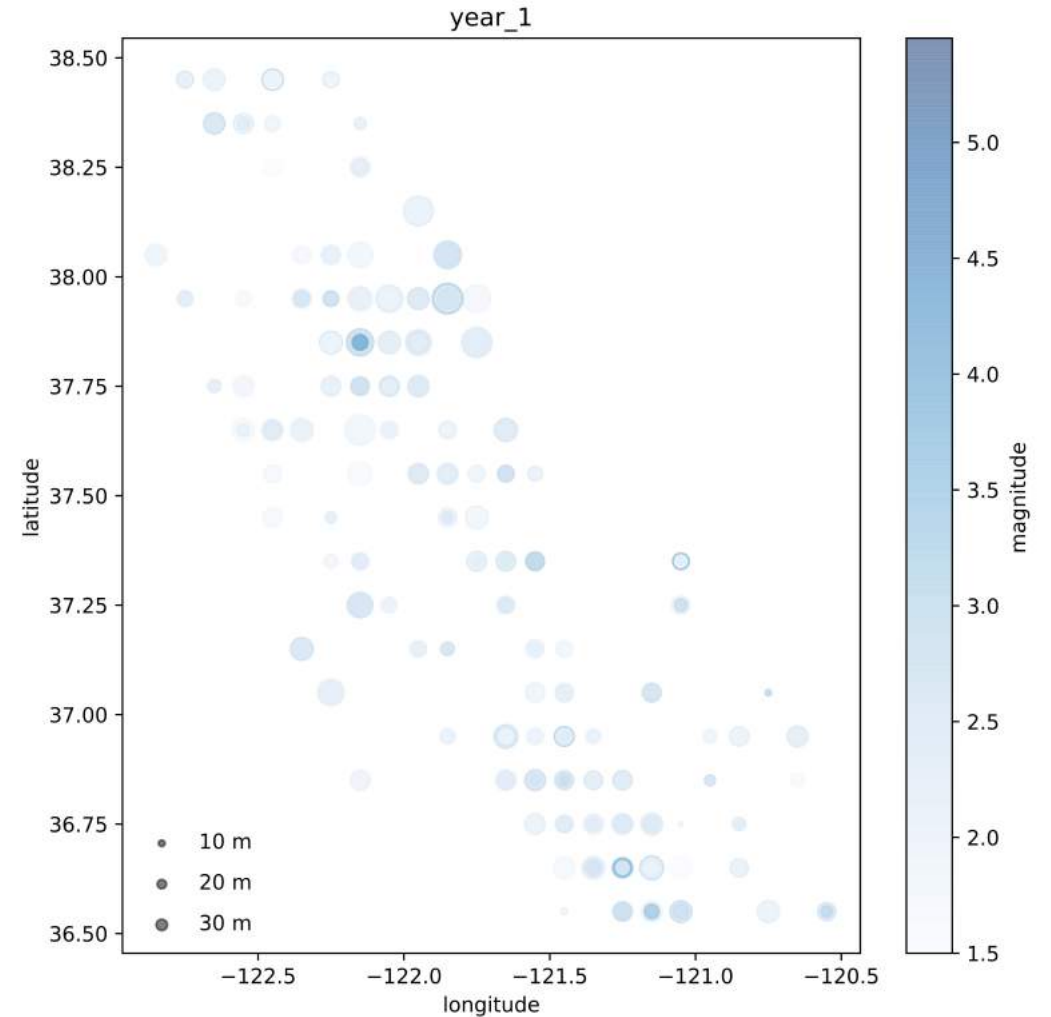

Figure 3: Example of a filtered down localized region of interest called a "grid"

Visualizing physical relationships

Correlation strength between the inputs (strain values in the x, y, and z directions, location, & cell) and outputs (magnitude, depth, & location)
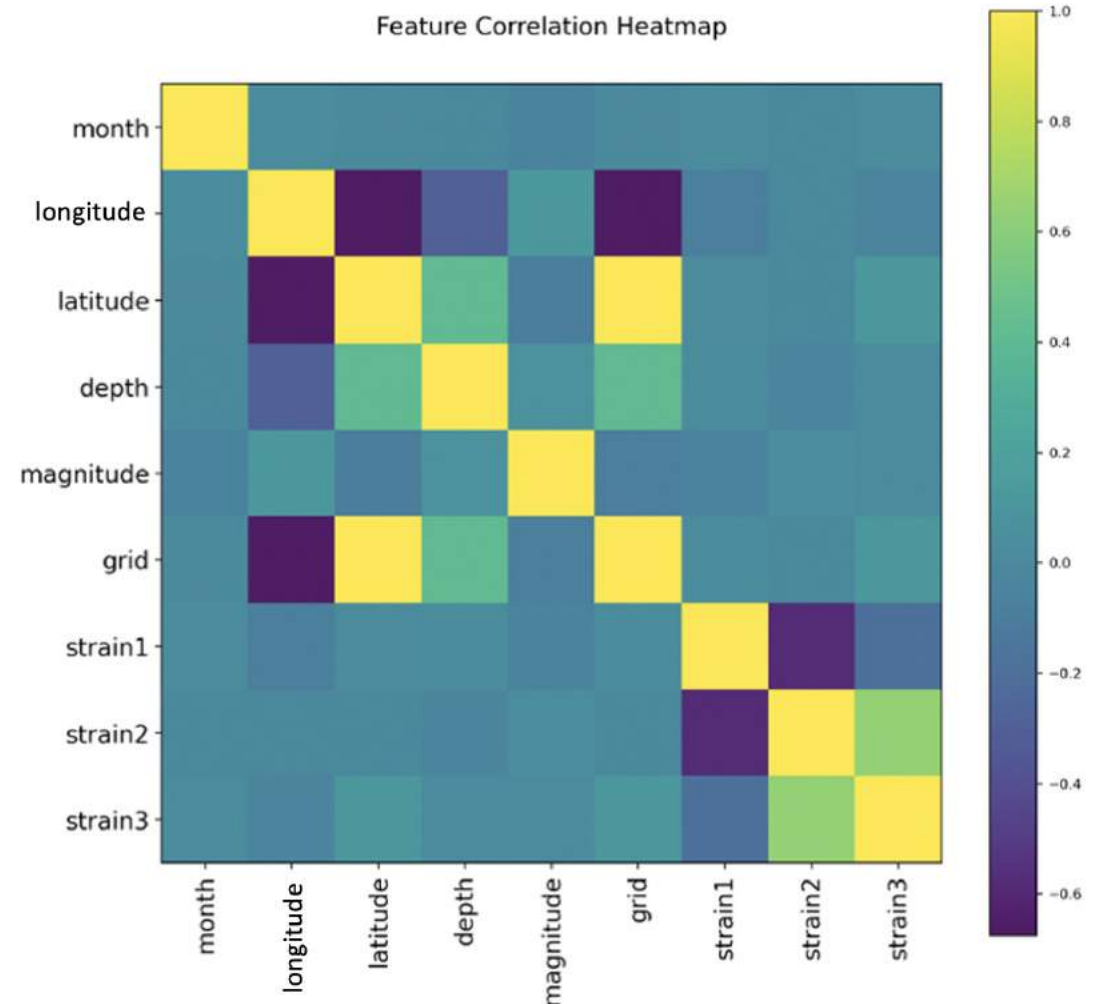

Figure 4: Linear heat map of correlation strength between the inputs and outputs

Model

- **Problem:** Sequence-to-sequence classification problem

- **Model:** Recurrent Neural Networks (**RNNs**) have shown considerable success in such sequence-to-sequence learning tasks

- **LSTM Network:** type of RNN able to capture long-term dependencies

DeepQuake architecture

Two LSTM networks:

- First LSTM processes input sequence of strain values at each cell generating a prediction of future strain values at each grid.

- Second LSTM uses these future strain values and the historical earthquake catalog to produce the output sequence consisting of the predicting earthquake forecasts.

Results

Comparing DeepQuake's predictions of key earthquake variables against the ground truth earthquake catalog (Figure 5)
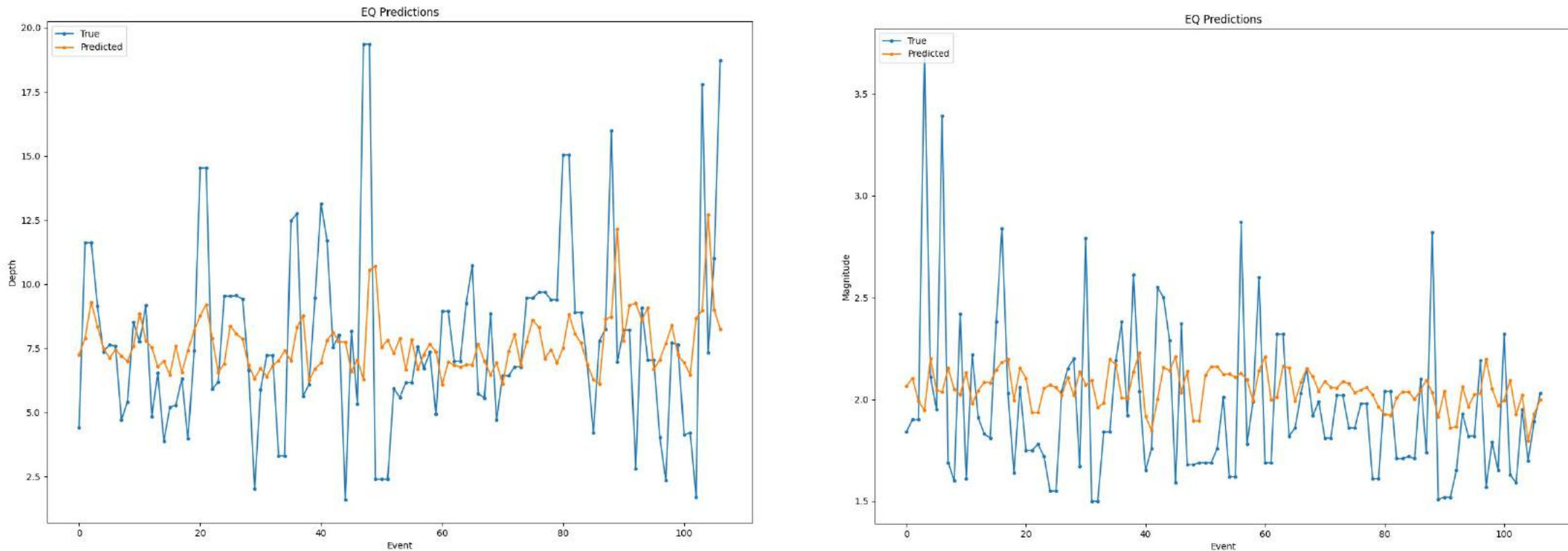

Figure 5: Ground truth vs DeepQuake predictions for depth and magnitude in Long Valley Caldera

Future work

Results

Hybrid model to predict key seismic variables of a future earthquake

Limitations & Future Work

- Include other climate variables such as historical precipitation to improve the accuracy of the model further

- DeepQuake is trained in a localized region of California

- Black-box nature of AI algorithms makes it difficult to understand underlying physical phenomena of earthquakes