

Time Series: Earthquake Forecasting

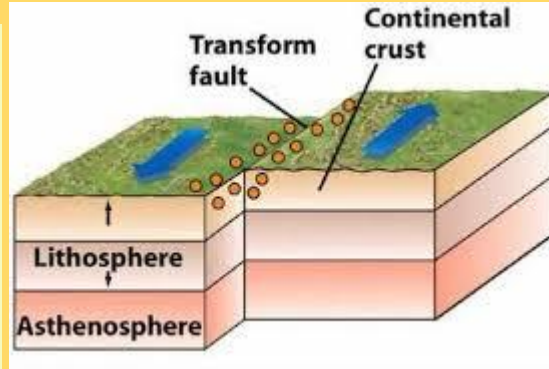


By: Jeffrey Ng

Business Case

The SF/Bay Area is home to the San Andreas fault and 6 others making it an active earthquake belt. Forecasting earthquakes so the area can prepare, will minimize injuries, and save economic/financial losses due to property damage.

Earthquake Genealogy

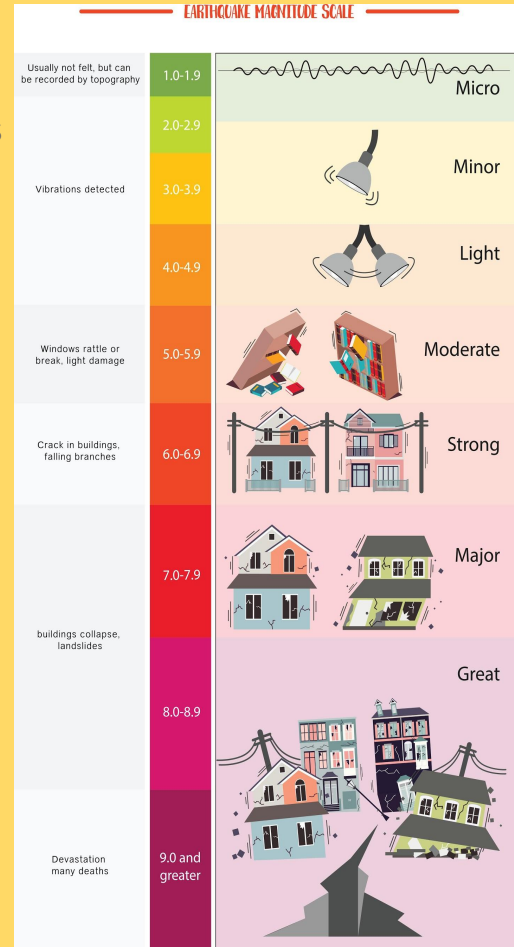


An earthquake is caused by the friction of the plates moving against each other

Earthquakes in developed countries	Average damage (\$mil, 2015 dollars)	Median damage (\$mil, 2015 dollars)
All 5.5+ quakes	\$12,146	\$484
Magnitude > 6.5	\$23,966	\$628
Magnitude 5.5 to 6.5	\$2,145	\$178
Population > 250,000	\$20,705	\$1,980
Population < 250,000	\$172	\$28

A foreshock is a series of smaller seismic events before a main larger earthquake. Foreshocks and aftershocks are relative terms that can't be named until the main earthquake occurs.

Richter Scale



Data Gathering & Cleaning

- www.usgs.gov
- Time series of seismic data from quakes > 0.1 for the San Francisco/Bay Area from Jan. 1, 2010 to Aug 2, 2020 (totaling 14,116 observations, 22 columns)
- Packages Used

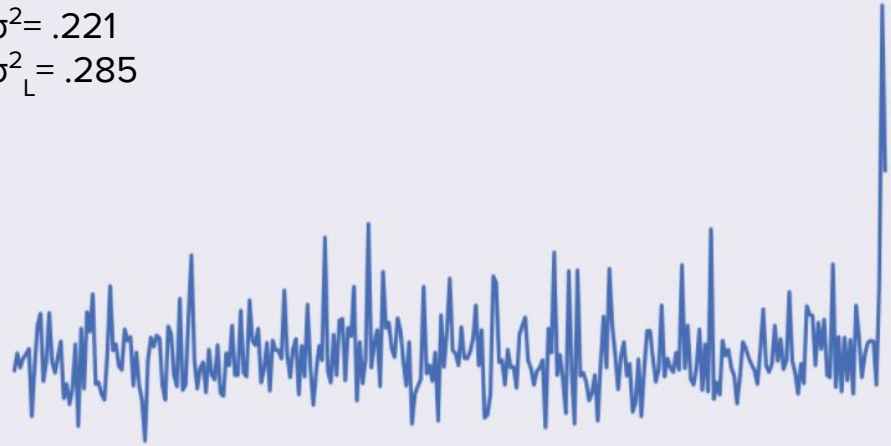


Napa, CA Earthquake 2014

- Only MAJOR earthquake in area for past 10 years
- 6.0 magnitude
- Aug 24, 2014 (3:20pm local time)
- 1 hr. NE of SF



$$\sigma^2 = .221$$
$$\sigma_L^2 = .285$$



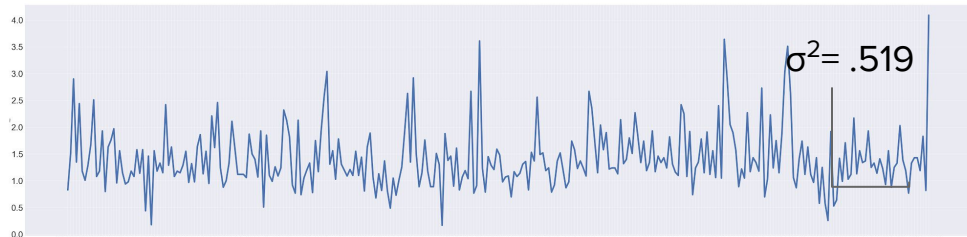
- H_0 - There is no significant difference in the variance of measured magnitude 2 months before quake, and 2 weeks before quake.

EDA of 4.0 quakes

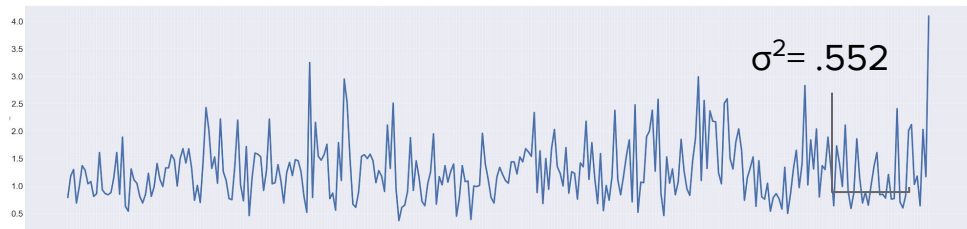
Quake
A



Quake
B



Quake
C



$$\sigma^2_L = .287$$



$$\sigma^2_L = .325$$

$$\sigma^2_L = .301$$

Reject Null
Hypothesis

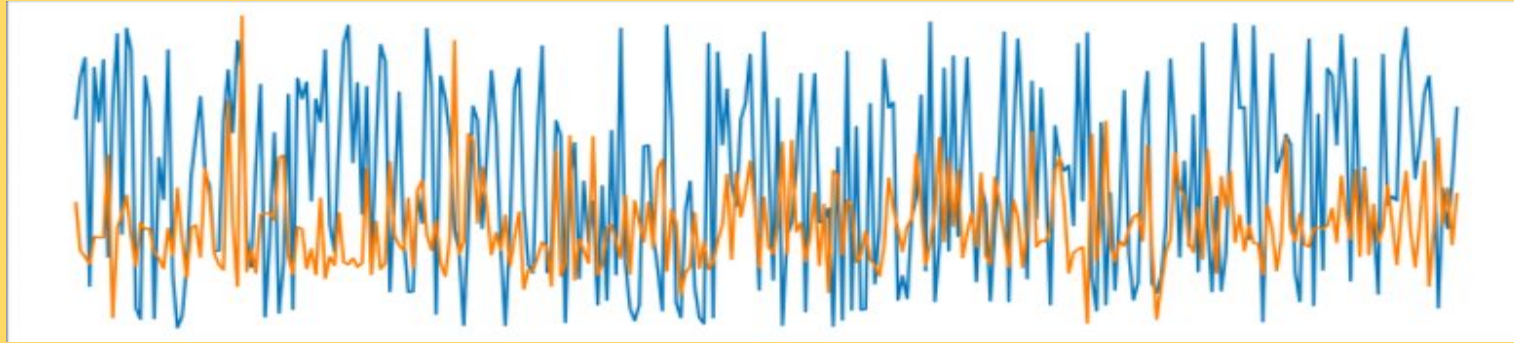
Further EDA



- Higher seismic activity 2 weeks before a feelable quake.
- Other notable features
4.0 happen 8 in 10 years
3.0 happen 10/year
2.0 happen 1 / 2.5 days

With so much seismic activity, is it even possible to forecast an earthquake?

Modeling...

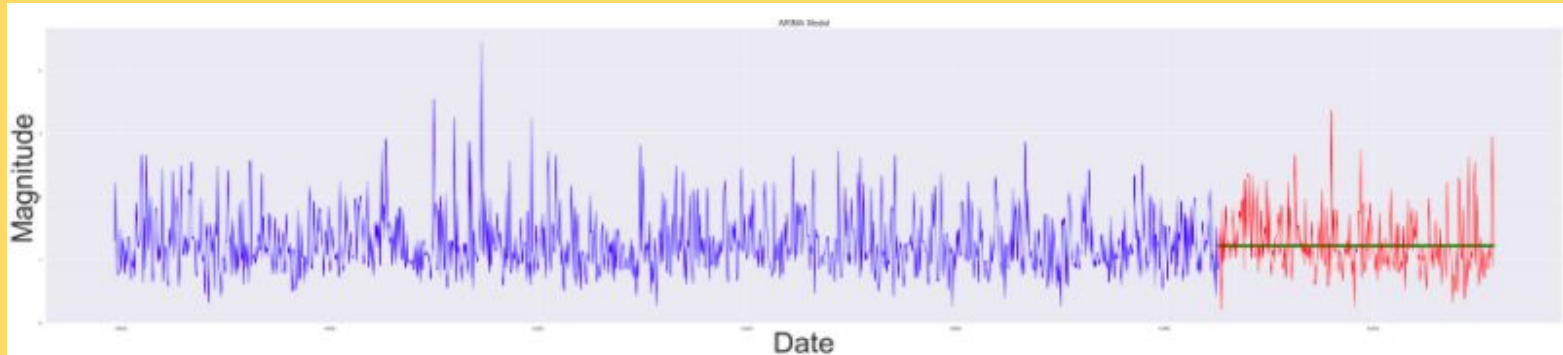


Dummy
Model

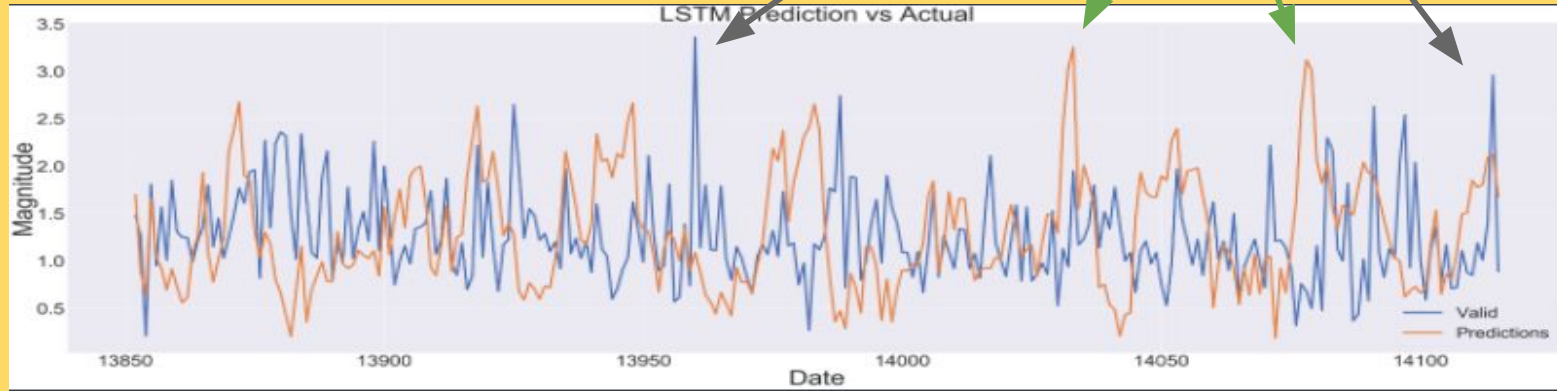
RMSE: .673
SMAPE: 79.29

ARIMA
(1,0,0) Order

RMSE : 0.0363
SMAPE: 28.9



LSTM model



LSTM

RMSE = 0.47
SMAPE = 13.8

Forecasts (Testing Data, 2 month span)

- **Despite their error in time, green arrows are similar to grey arrows**
 - **Important quakes are forecasted**
- The peaks are represented in different periods of time
- 10 months of training data were used

Evaluation

	Dummy (Baseline)	ARIMA	LSTM*
RMSE	0.227	0.0363	0.47
SMAPE	84.24	28.9	13.8

- SMAPE (Symmetric Mean Absolute Percentage Error) measures mean %age error in forecast, better measure
- RMSE = measures average predicted magnitude \pm RMSE
- *LSTM : neural network of 2 layer, 100 nodes/layer + dense layer of 35 nodes ran for 4000 epochs

Key Takeaways

- We can only forecast, not predict
- One major earthquake in 10 years
 - 4.0 quakes happened 8 times in 10 years
 - 3.0 quakes happen 10x / year
- Hypothesis supports notion of foreshocks
- LSTM can forecast within a given timeframe
 - Can forecast approximate magnitude of quake within 2 months

Future Steps

- Fine tune LSTM
- Using raw seismic data
- Investigation into foreshocks
- Further testing on hypothesis

Thank You!
Questions?

Github: <https://github.com/halujeff5/EarthquakeProj>

LinkedIn: <https://www.linkedin.com/in/jeffrey-ng-606406197/>