Can machine learning determine physical source properties of earthquakes from a single station?

Shanna Chu and Jeremy Maurer
schu3@stanford.edu jlmaurer@stanford.edu

MOTIVATION

In seismology it is important to determine the certain information about the source of an earthquake such as its origin depth and moment magnitude (a measure of the earthquake's size.) This usually requires performing calculations on data collected from multiple stations. We are interested in determining if, using features extracted from the time-series of ground velocity collected at a single seismometer, we are able to classify earthquake events according to depth or magnitude. These include features such as the ratios of different windows of the frequency spectrum and time difference between maximums in the timeseries.

DATASET

We use data from Japan's Hi-Net seismometer array, specifically all catalogued earthquake events from a year of data recorded at Station UCNH on Kagoshima island. The dataset is highly inhomogenous in that smaller magnitude earthquakes occur much more frequently than larger ones and shallower earthquakes occur more frequently than deeper ones.

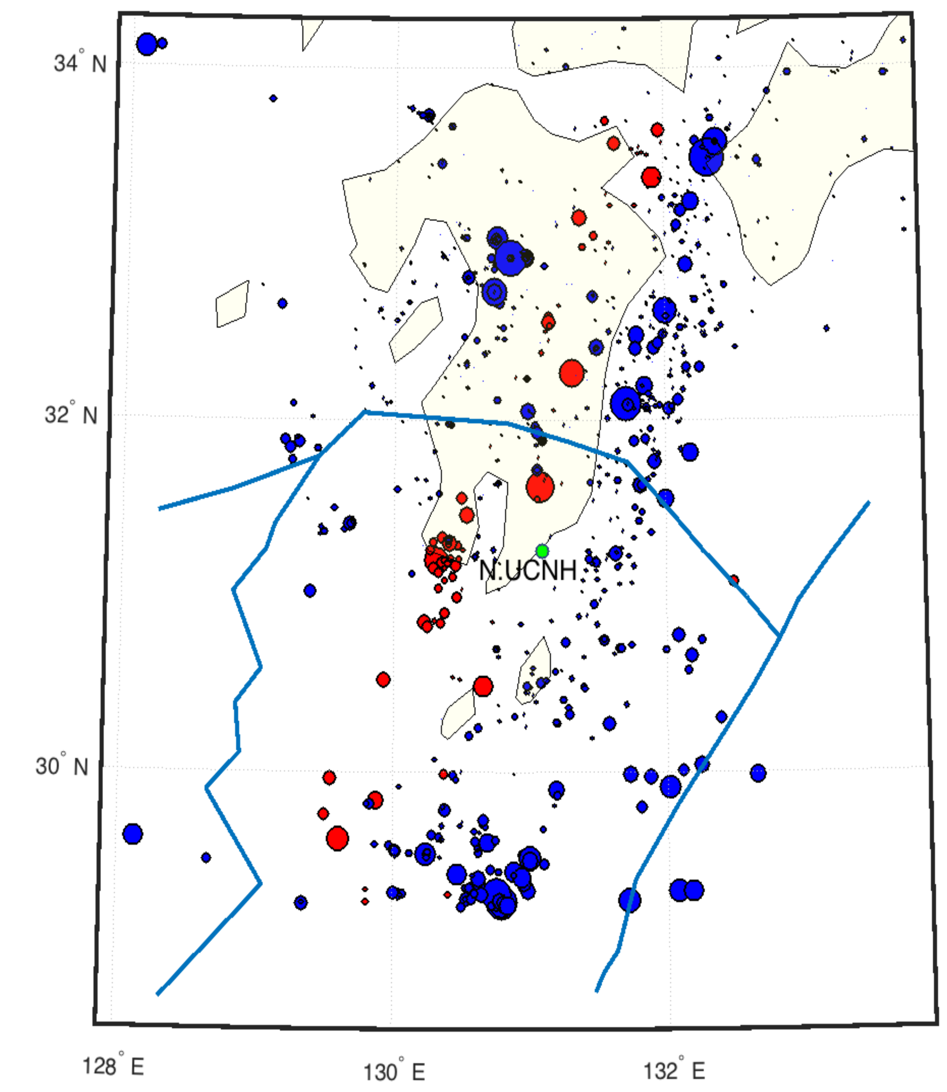


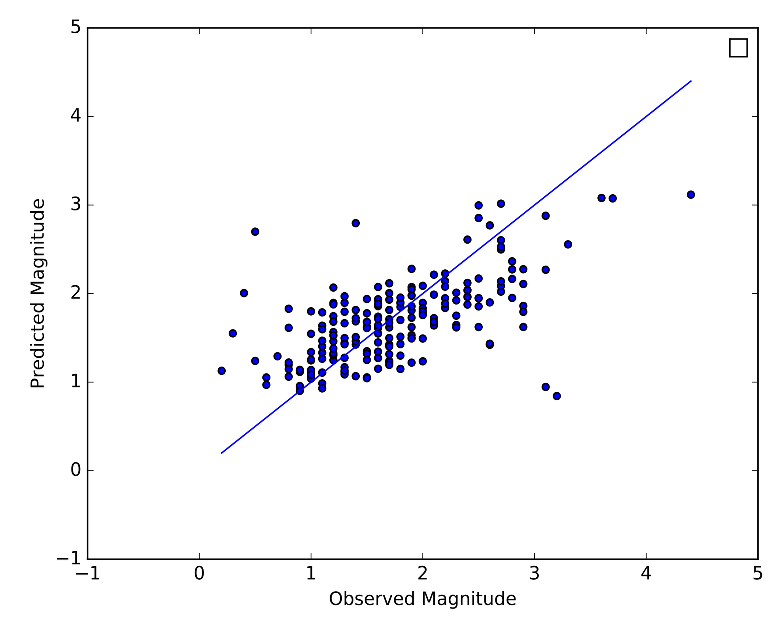
Fig.1. Map of earthquake data used in this study. Red indicates events deeper than 60 km and blue are shallower. Size of dots are proportional to earthquake magnitude. The green dot is the location of station UCNH.

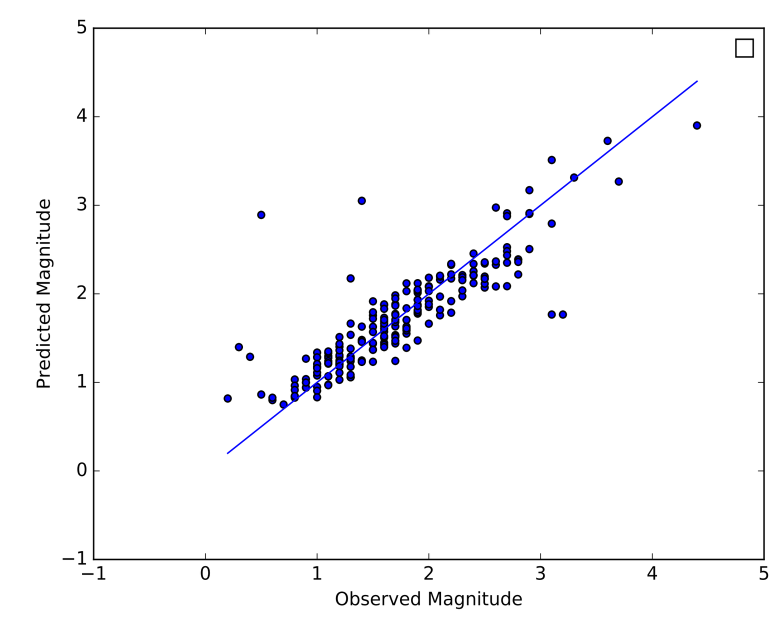
MODEL AND RESULTS

Magnitude Regression

Models and Performance

Model	Equations	# Features Used	Training R^2	Test R^2
Bayesian Ridge Regression	$p(\theta\|y^{(i)}) = \frac{p(y^{(i)}\|\theta)p(\theta)}{p(\theta)}$ where $p(\theta) \sim \mathcal{N}\left(0, \lambda^{-1}\mathbf{I}\right)$ and $p(y^{(i)}\|\theta) \sim \mathcal{N}\left(\theta^T x^{(i)}, \alpha\mathbf{I}\right)$	16	0.49	0.39
Random Forest Regressor	Finds the feature/ threshold that minimizes the total mean squared error for all the data	14	0.94	0.75





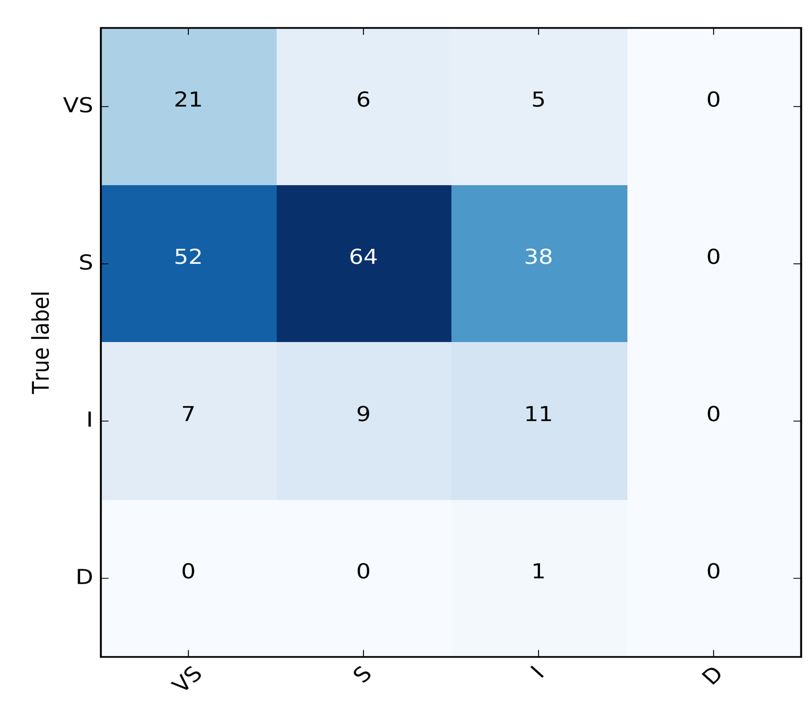
Features for Magnitude Regression

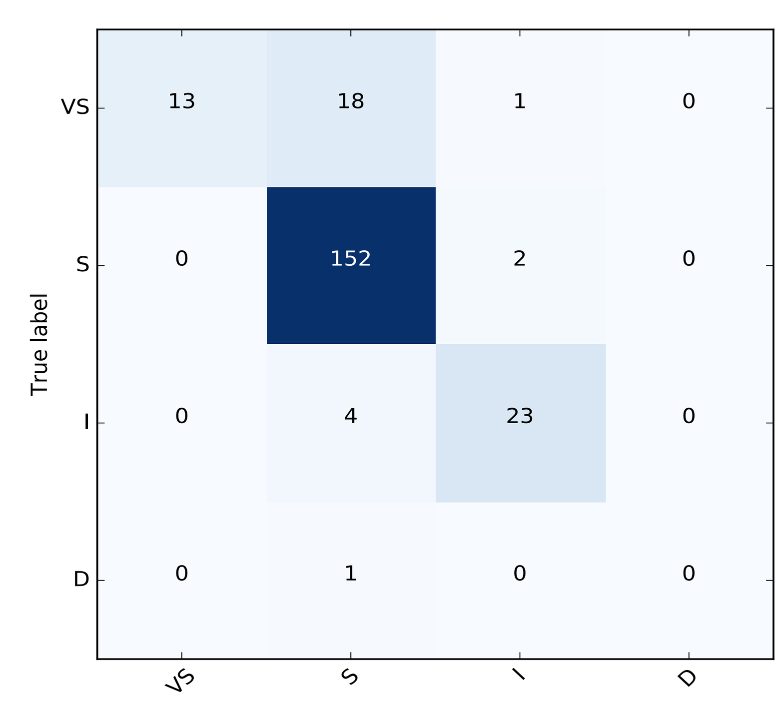
The features that we used to predict earthquake magnitude were based on the features extracted using tsfresh. We used Cross-Validation to determine the features that were informative for predicting magnitude. Magnitude is related to the amplitude (height) of the wavefor, so features that turned out to be important were a number that quantify the amplitude of the waveform: maximum, minimum, quantile ratios, The two regression models we used did not use the same set of features, and the features used by the RFR were more related to amplitude. The fact that the RFR did much better than the linear BRR could indicate non-linear relation between predictors and magnitude.

Depth Classification

Models and Performance

Model	Equations	# Features Used	Training accuracy	Test accuracy
SVM	$\text{minimize} \|\|w\|\|$ $\text{s.t.} y^{(i)}(w^T x^{(i)} - b) \geq 1$	200	0.48	0.45
Random Forest Classifier	Finds the feature/ threshold that minimizes the mis-classification rate at each branch	49	0.97	0.87





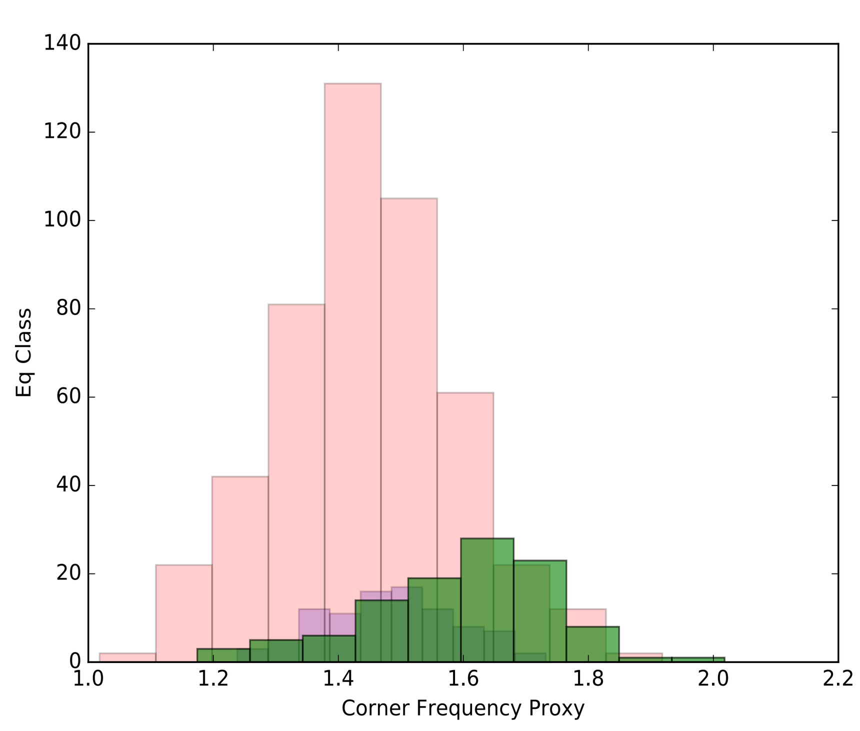
Features for Depth Classification

The earthquakes in our study area are clustered at different depths, so instead of regressing depth we divide them into a set of depth bins (Very Shallow [VS: 0-20 km], Shallow [S: 20-70 km], Intermediate [I: 70-300 km], and Deep [D: >300 km]). We use the features generated by tsfresh, as well as creating our own features relating to peaks in the frequency spectrum, an estimate of the number of 'corner frequencies' in the log-amplitude vs. log-angular frequency plot, and the first three principal components. We use 5-fold cross validation to choose the most predictive features for the Random Forest Classifier (RFC), and an F-test scoring function to select the top 200 features directly for use with the SVM. We use balanced class weights with the SVM because there are many fewer VS, I, and D events than S.

FEATURES

We extract features from the raw timeseries data using tsfresh (Christ et al., 2016), as well as compute several additional features ourselves from the data such as the ratio of subsequent windows in the power spectra, which represents a measure of the shape of the spectrum. Tsfresh extracts various types of statistics related to the raw data and various transforms of the data (e.g. Discrete Fourier Transform). Since there are many features (~700) and not all are relevant, we use various kinds of feature selection techniques (F-test score, cross-validation, and random forest classifiers) to extract only the relevant features and use those for our problem. Examples of relevant features include basic statistics like kurtosis, number of peaks, minimum and maximum, the quantiles of the time derivative; autoregressive, autocorrelation, and different frequency-domain coefficients also were important.

Fig.2. Windowed ratio of power spectrum which serves as a proxy for corner frequency, shown here in depth clusters. Dark pink, light pink and green represent very shallow, shallow, and intermediate-depth earthquakes repsectively.



DISCUSSION

We obtained better results for both magnitude and depth prediction using random forest models. This makes sense because individual features themselves are not extremely indicative of either magnitude or depth. But by combining the features, we get a better predictor. Physically, we understand why some features predict certain labels - earthquake magnitude, for example, is partially related to timeseries amplitude, especially for earthquakes at a common depth. Depth is related to various measures of the power spectrum because high frequency data is attenuated at depth. However, some of the features that are either related to amplitude or frequency may actually be correlated with each other, and we might have gotten a better model by removing some features that correlate with each other.

FUTURE DIRECTIONS

As stated in the Discussion, we would like to analyze our features more to see which features are independent of each other. Secondary features derived from physics, such as the windowed ratio of the power spectra proved to be an interesting , but we did not have enough time to completely separate the values for the labels (see Figure 2). We could try a projection with SVM to try to separate out the values for each label or a deep neural network to learn the important features directly.

REFERENCES

Christ, M., Kempa-Liehr, A.W. and Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. ArXiv e-print 1610.07717, https://arxiv.org/abs/1610.07717

Package we used: scikit learn