Predicting Eruption Events at Volcanoes from Earthquake Data

Ben Mullet (bmullet), Ankush Singh (ankushs), Ying-Qi Wong (yqwong)

Introduction

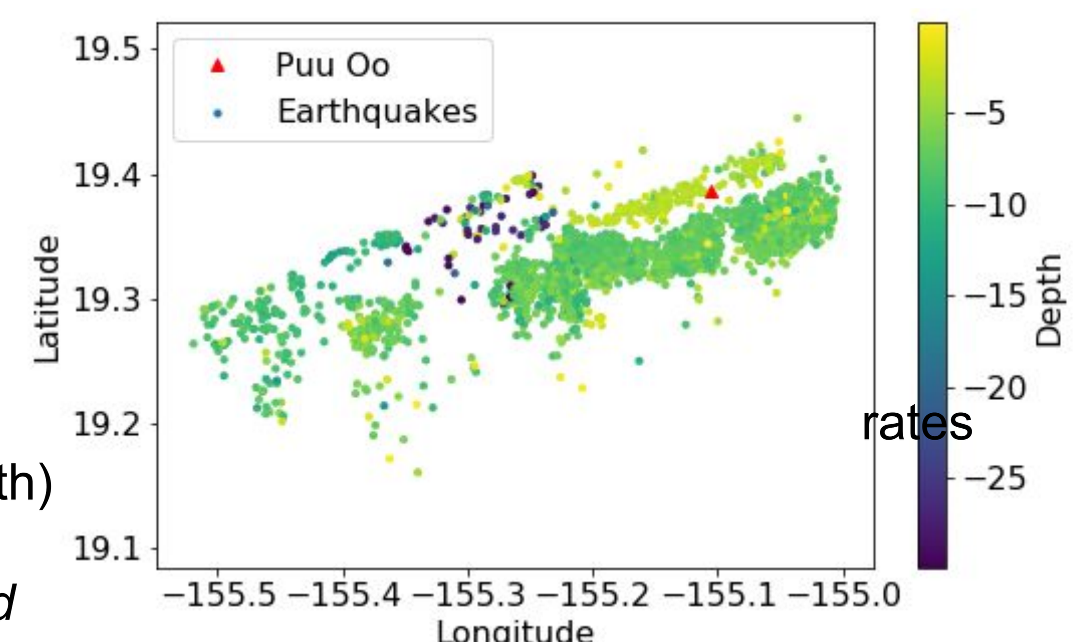
Anticipating the timing and location of volcanic eruptions can reduce the hazard to nearby communities. Many volcanic eruptions are preceded by unrest that includes increased earthquake rates, pronounced ground inflation and elevated gas emissions. However, the implications of these signals on eruption timing are still unclear. In this project, we applied machine learning algorithms to elucidate the **relationship between earthquakes and volcanic eruptions** for Kilauea, a highly active volcano on the Big Island of Hawaii. We first use the earthquakes to **classify if a volcano is erupting**, then explore methods to see if we can **predict the time to eruption.**

Dataset

We employ an earthquake catalog of Kilauea from 1983 – 1986 from the World Volcano Database. This period has good records of eruption events (timing, location, duration) for labels of our dataset. We randomly split the dataset 70%-20%-10% for the training, development and test sets respectively while ensuring a consistent proportion of eruptions in each set.

Features of each earthquake are:

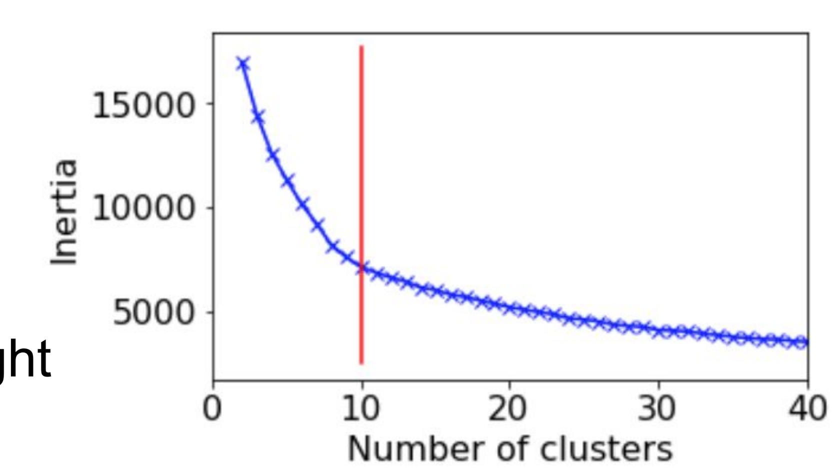
- Longitude
- Latitude
- Depth
- Magnitude
- Earthquake rates (per day, week, month)



Each feature was *scaled* before applying the algorithms.

Algorithms deployed

- **Logistic regression**
- **Random Forest**
- **Neural network**
  - Built using pytorch
  - 4 layer network
  - SGD optimizer with weight decay regularization
- **K-means clustering**
  - Ideal number of clusters chosen by elbow
  - Classification: clustering during eruptions and repose done separately, then closest centroids to test set were calculated to classify.
  - Eruption timing: Time to eruption included as a clustering variable.

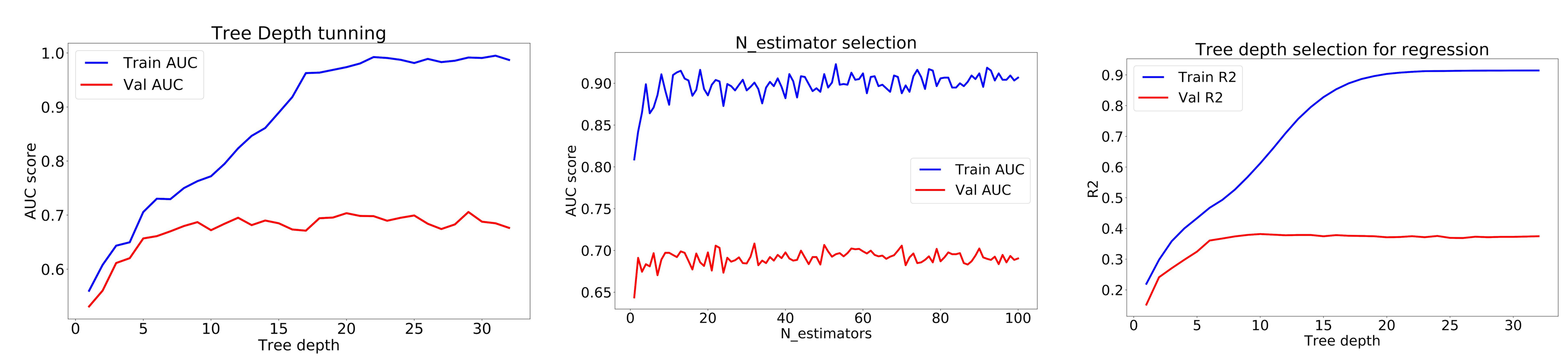


Binary Classification: Is it erupting?

Algorithm	Dataset	Kappa	AUROC	Confusion matrix
Logistic Regression	Train	0.37	0.64	[[2165 45] [295 128]]
	Dev	0.25	0.59	[[630 17] [83 22]]
	Test	0.28	0.60	[[315 9] [42 13]]
Kmeans	Train	0.31	0.67	[[1895 315] [219 204]]
	Dev	0.31	0.67	[[559 88] [54 51]]
	Test	0.24	0.63	[[275 49] [32 23]]
Random Forest	Train	0.84	0.89	[[2043 0] [0 590]]
	Dev	0.46	0.66	[[575 9] [102 66]]
	Test	**0.52**	**0.71**	[[288 5] [50 36]]
Neural Network	Train	0.35	0.63	[[2025 18] [429 161]]
	Dev	0.27	0.60	[[572 12] [131 37]]
	Test	0.29	0.61	[[290 3] [67 19]]

Parameter tuning example

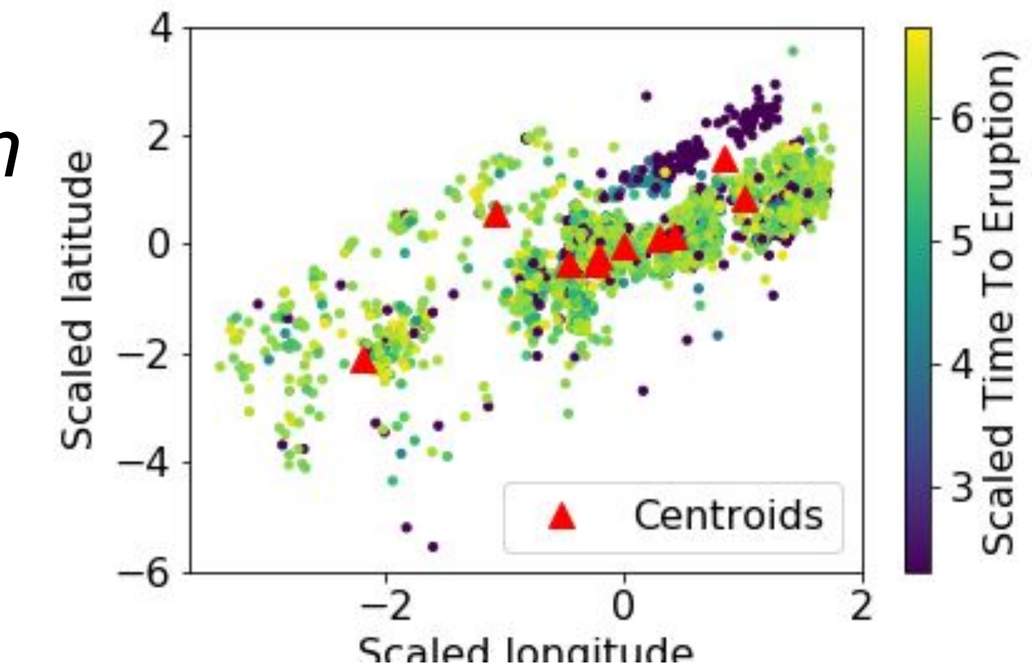
Random Forest classifier and regressor



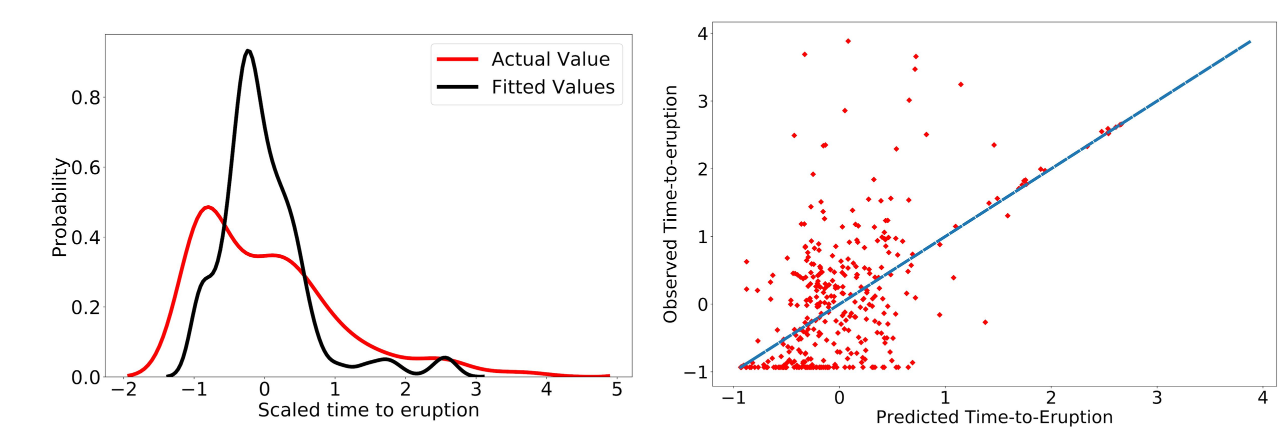
Predicting the time to eruption

Algorithm	RMSE	R^2
Kmeans (train/dev/ test)	0.95/0.95/1.02	0.097/0.059/0.037
Randm Forest (train/dev/ test)	**0.16/0.68/0.64**	**0.82/0.37/0.38**
Neural Net (train, dev, test)	0.86/0.92/0.90	0.26/0.15/0.19

Time to eruption from K-means clustering



Time to eruption from Random Forest



Challenges and Future Work

Challenges

→ **Imbalances:** Curse of the 0 label prevalence

→ **Earthquake triggering**: Larger earthquakes trigger and are triggered by smaller earthquakes

→ **Limited Features**: Only lat-lon, depth, magnitude, rate: Challenging to get additional relevant features

→ **Variable time scales of relevance**: Not clear which time scale is more relevant for the predictive parameters

Future work

- Expanded the earthquake and eruption catalog beyond the time frame considered here.
- Gather data for additional features: continuous seismic recordings, GPS measurements, gas emission rates, etc.
- Include volcanoes from other part of the world with variable eruption frequencies