Using Machine Learning Techniques to Forecast Major Earthquake Probabilities Along the San Andreas Fault System After a 100-Year Drought

Yousaf, Fatimah (School: Little Rock Central High School)

California experiences reoccurring earthquakes along its fault lines; however, since 1906, the southern region of the San Andreas Fault System has not produced a 6.5 or higher magnitude (M) earthquake in over 100 years despite statistical predictions that a major earthquake would occur (Wald et al., 2019). Computational algorithms remain largely a novel field in statistical predictions. To improve earthquake predictions and investigate the likelihood of the current seismic drought ending in the next 30 years, machine learning techniques–neural networking, logistic regression, and random forest classifier–were compared using about 200,000 earthquake catalogs from the Rate State earthQuake Simulator, a physics-based earthquake simulator. As a continuation project, Lasso regularization was included for each code to analyze clustered earthquakes' stress tension occurring in one region. Time-independent Poisson statistics were used to find the overall probability of an M6.5+ earthquake occurring over the next 30 years. UCERF3's time-dependent model of past earthquakes was used as a control. Each algorithm measured magnitude over a specific fault region to predict earthquake tension based on rupture percentages. After measuring each algorithms' accuracy, logistic regression was found to be most efficient in future earthquake predictions; the model discovered that the probability of a 6.5 magnitude drought-ending event in the next 30 years is 94%. Each table has a p-value less than 0.05, making the data statistically significant. Therefore, California is very likely to see the drought end soon. Extended research analyzed frequency likelihoods of earthquakes in Japan, Mexico, and the United States.

Awards Won: