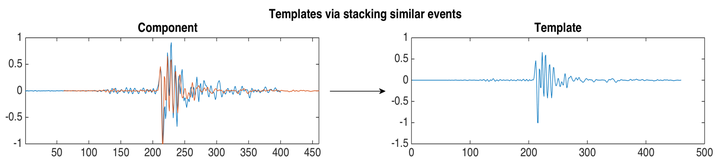
Event Identification in Continuous Seismic Data

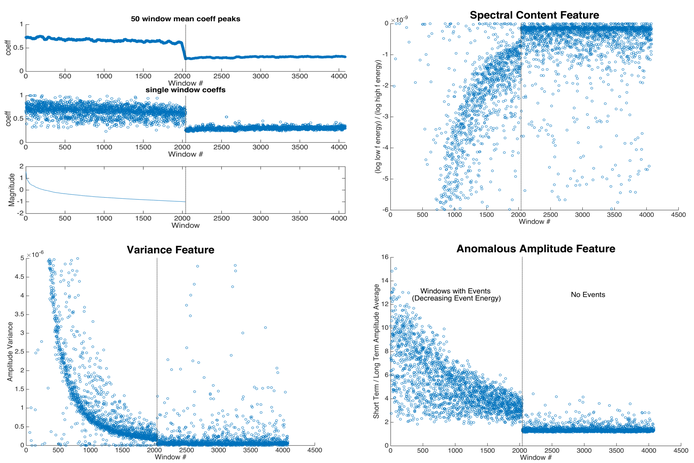
Alex Hakso (ahakso@stanford.edu) & Fatemeh Rassouli (frasouli@stanford.edu) *Department of Geophysics, Stanford University*

Robust Single-Receiver Seismic Event Detection

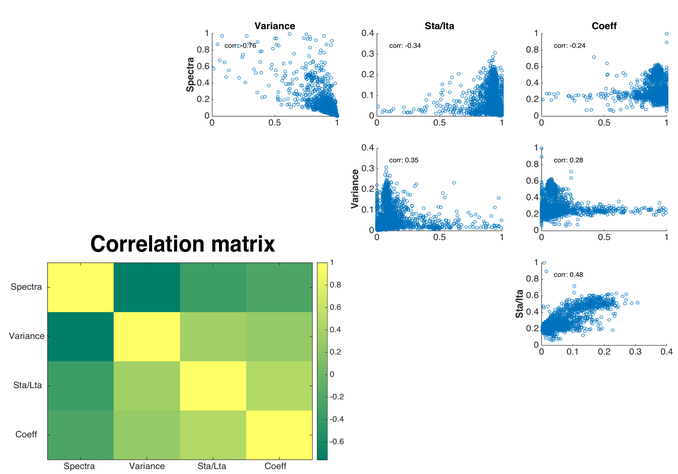
A single 3-component seismometer sampled continuously at 100 Hz for 1 month during a high-density earthquake sequence near Guy, Arkansas. This data comprises nearly 3 million candidate time windows, of which only ~8000 contain events.



Feature Detection

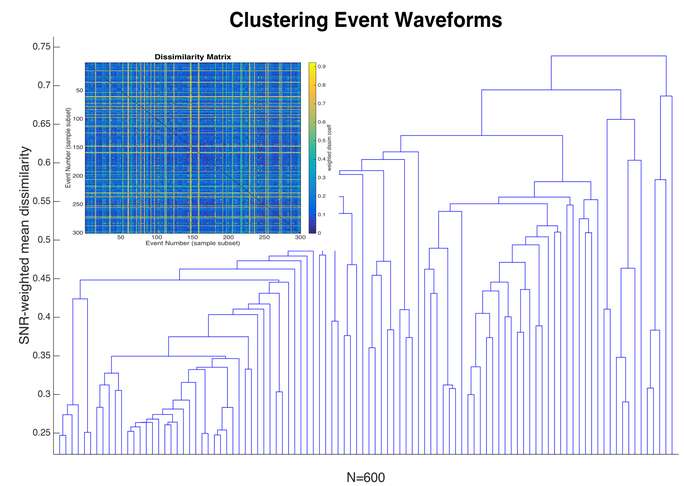


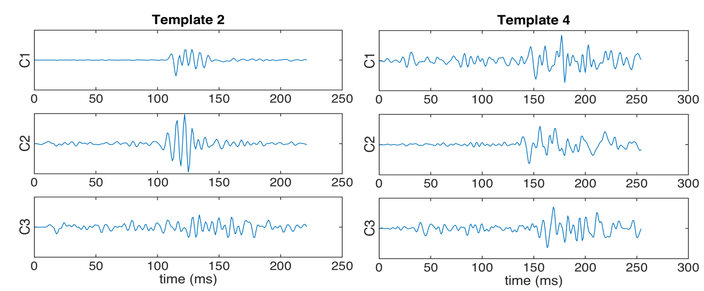
Correlation matrix: sample normalized feature values



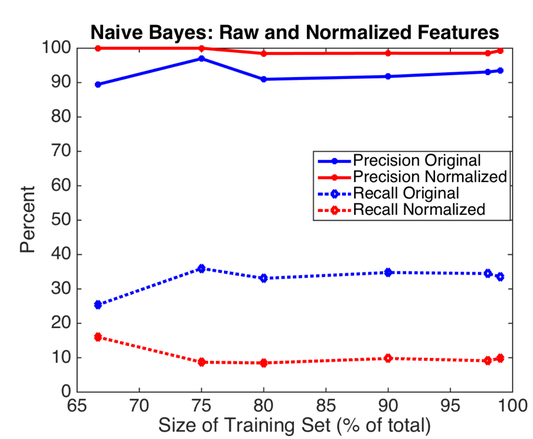
Unsupervised learning to extract cross-correlation features

$$t(x,y) \star f(x,y) = \frac{1}{n}\sum_{x,y}\frac{(f(x,y)-\bar{f})(t(x,y)-\bar{t})}{\sigma_f \sigma_t}$$



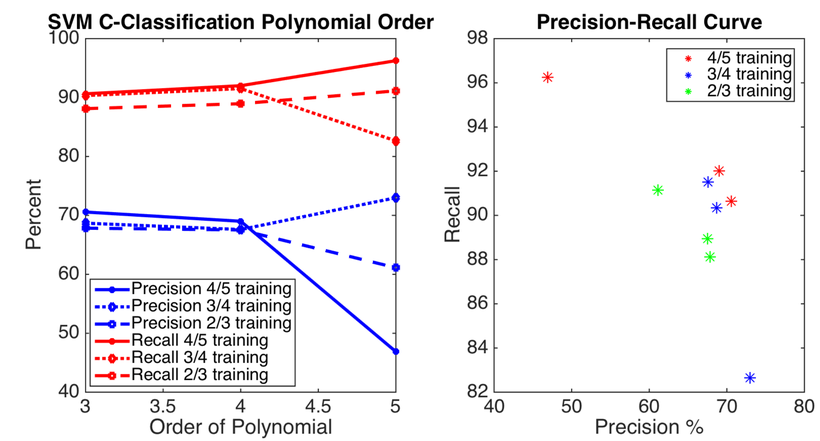


Feature Normalization



Supervised Learning Classification

Logistic Regression	$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}.$
Naïve Baye's	$\Pr(Y=k\|X=x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K}\pi_l f_l(x)}$
KNN	$\Pr(Y=j\|X=x_0) = \frac{1}{K}\sum_{i\in\mathcal{N}_0} I(y_i=j)$
SVM	C-Classification Nu-classification



It was determined that this highly skewed data set was most effectively characterized using the non-parametric KNN approach. Despite the relatively high computational cost, the unparalleled precision and recall achieved by KNN makes it the practical choice for most applications. In particular, having developed the model, real time monitoring is achievable at low cost using KNN.