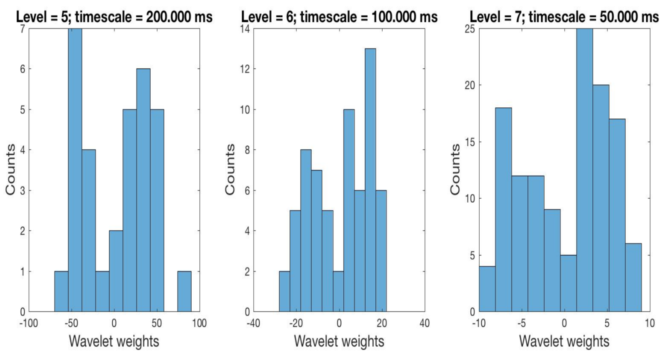
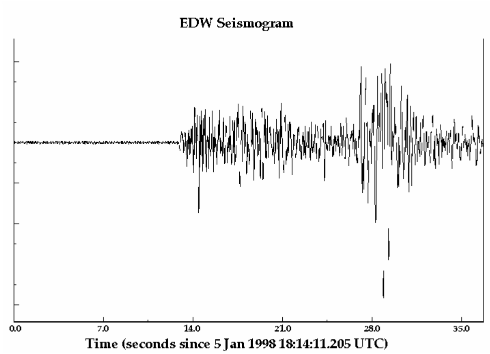
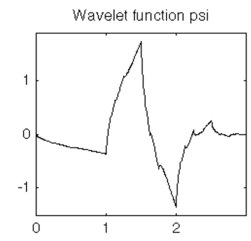
Exceptional service in the national interest

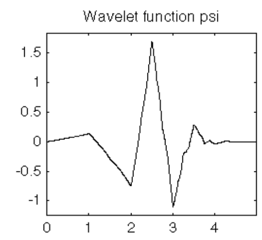
Sandia National Laboratories







db2



db3

Using discrete wavelet transforms to discriminate between noise and phases in seismic waveforms

J. Ray, C. Hansen, R. Forrest & C. J. Young

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

SAND2017-3960C

Introduction

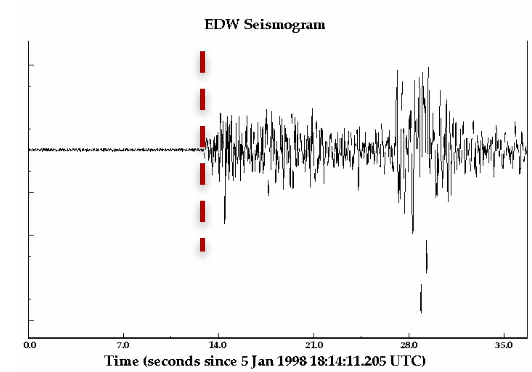
- Aim: Define features of a seismic waveform that can be used to discriminate between noise and P/S phases
  - Particularly, use the multi-resolution information in the waveform to define these features
- Focus on:
  - Regional scales (300-3000 km), wave travel through the upper mantle
  - 3-component data only; no use of data from arrays
- Hypothesis
  - Regional waves have structures that are localized in the coda
    - Global methods (like Fourier decompositions) are not very good for localized phenomena
    - Using energies measured over a given time-scale may be too coarse a measure to detect these structures
- These multi-resolution (MR) features would augment current features to discriminate noise/signal i.e., picking

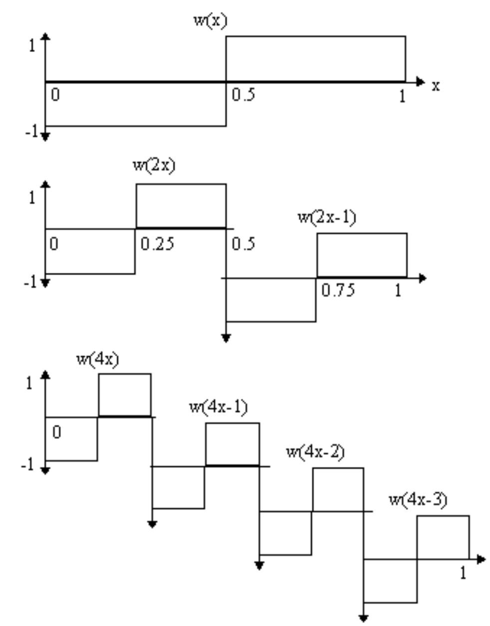
Current method

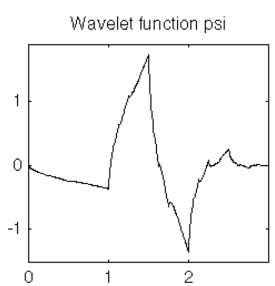
- Noise-signal discrimination done using STA/LTA algorithm or fitting 2 AR models via AIC criteria
  - STA/LTA requires one to specify windows of length *S* and *L* and abutting at *t*, compute the energy ratio, and search for an t_{max} that maximizes the ratio
    - t_{max} is the arrival time
    - *S, L* vary between station
  - AIC method requires one to make auto-regressive models in a noise & signal window, abutting at a pick time *x*
    - *t* is varied along the time axis till we minimize AIC
- P/S discrimination done using polarization metrics
  - *S* waves also have a bigger amplitude
  - Signal duration, horizontal-vertical amplitude ratio, rectilinearity, frequency, time difference between arrivals etc. are the classifier's features

What are multi-resolution features?

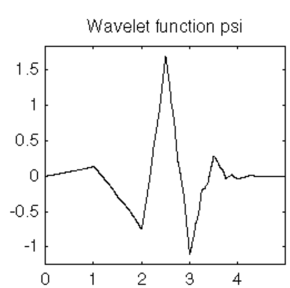
- MR features are derived from a wavelet decomposition of $f(t_i)$, i = 1…N, the waveform
  - Wavelets are orthogonal functions Φ(t) with compact support
  - Discrete wavelets are defined at resolution levels l, which change by powers of 2
  - $f(t) = \sum_{l=0}^{L} \sum_{j=0}^{2^l} w_{lj} \Phi_{lj}(t)$, L = $\log_2$(N)
- Many types of wavelets, Haars, Daubechies 4, ….
  - Ideally, we'd like to choose a wavelet which maximizes the number of w_{lj} = 0



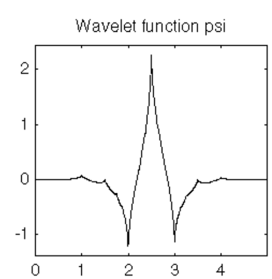




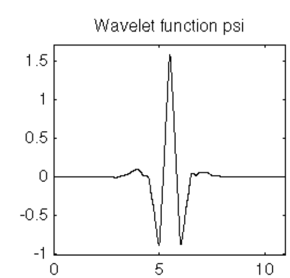
db2



db3



coif1

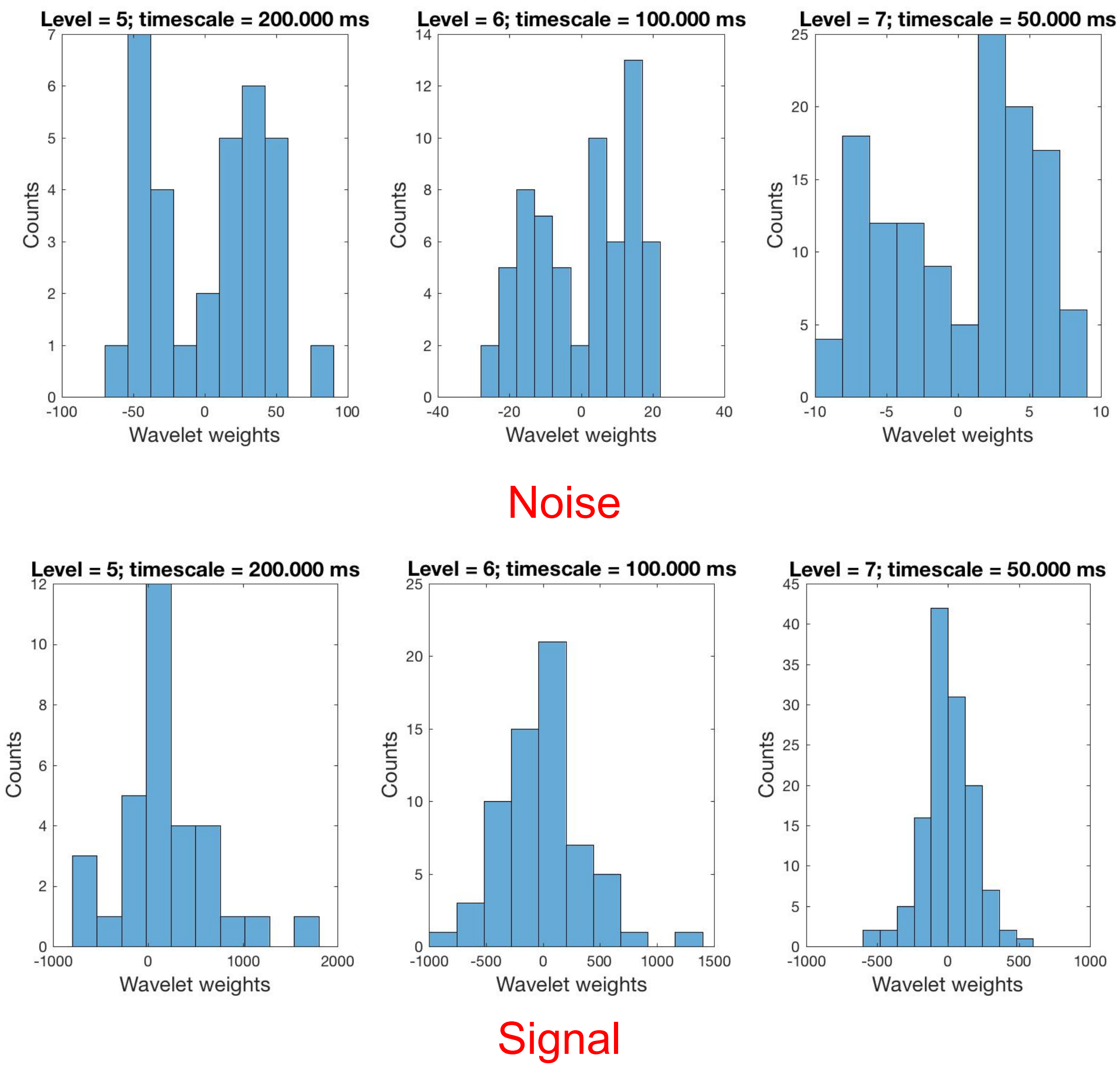


coif2

Thanks, www.wavelet.org/

Illustration – MR decomposition

- We perform a wavelet decomposition of a waveform, 6.4 seconds (256 samples, 40 Hz sampling rate) window, pre- and post-pick time
- Histograms of wavelet weights w_{lj} are different, pre- and post-pick
  - How to turn this into a feature?



Designing discriminating features

- Feature # 1: A boolean summary of wavelet distribution $g(\mathfrak{W})$
  - At each level *l*, do a two-sided Kolmogorov-Smirnov test to check, at α significance level, if the 2 histograms are different
  - This will provide you with 3 true/false results (from the 3 components)
  - Assumption: It's unlikely that an arriving wave is registered only on 1 axis/component
  - So, at least 2 'trues' (score > 2) are required to indicate the arrival of a wave. *But does this feature work (discriminate) for all levels l?*
- Feature #2: A ratio of pre- and post-pick wavelet energies
  - For each level compute $E_l^{(k)} = \sum_j w_{lj}^2, k \in \{\text{noise}, \text{signal}\}$
  - Compute energy ratio $\kappa_l = E^{(signal)}_l / E^{(noise)}_l$
  - *A ratio > 1 indicates arrival of a signal, but what is that threshold κ^*?*
  - *Also, which levels l does this work for?*
- Basically, we are interrogating the waveforms at a range of timescales ("multiresolution") for discriminating features

Tests using g($\mathfrak{W}$)

- Data obtained from Coronel Fontana, Argentina (CFAA); 74 x 3 waveforms
  - Manually labelled data, with 74 arrivals and picks
- Check how well g($\mathfrak{W}$) discriminates noise & signal. Errors: false negatives
- Use g($\mathfrak{W}$) on waveforms *without* an arrival. Errors: false positives
- Opt. parameters: α (KS2 significance); for now , α = 5%

	Level 5 (τ = 200 ms)	Level 6 (τ = 100 ms)	Level 7 (τ = 50 ms)	Summary (from 50 & 100 ms levels)	STA/LTA algorithm
Data with a signal	66.2%	83%	92%	**False negative rate: 12.5%**	**77% FN: 23%**
Data without a signal	100%	100%	84%	**False positive rate: 8%**	**51% FP: 49%**

Percentage of correct discrimination & false negative/positive rates

Tests using wavelet energy ratio κ

- Use the same CFAA data.
- Opt. parameters: $\kappa*$ (threshold ratio); for now , $\kappa* = 2$

	Level 5 (τ = 200 ms)	Level 6 (τ = 100 ms)	Level 7 (τ = 50 ms)	Summary (from 50 & 100 ms levels)	STA/LTA algorithm
Data with a signal	92%	94%	94%	**False negative rate: 6%**	**77% FN: 23%**
Data without a signal	72.5%	75.5%	76.5%	**False positive rate: 24%**	**51% FP: 49%**

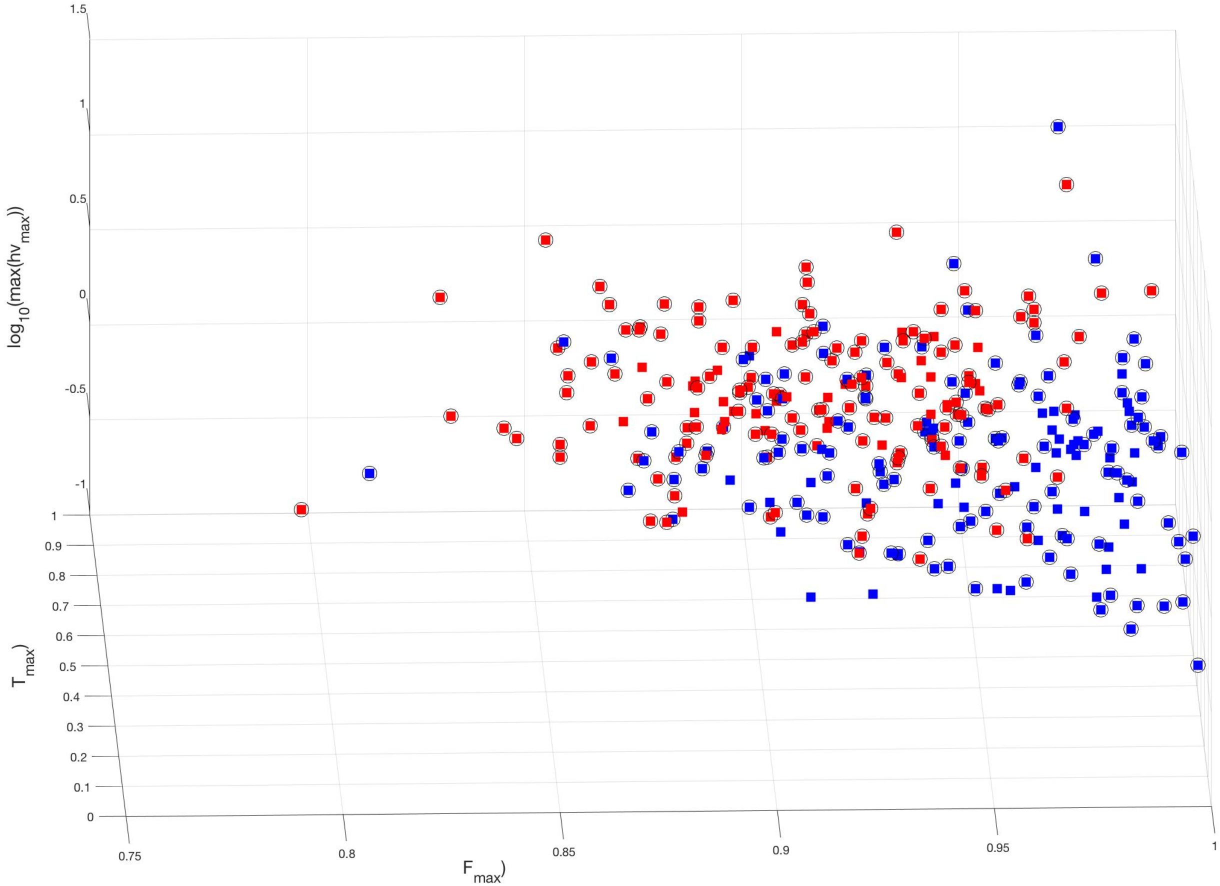
Percentage of correct discrimination & false negative/positive rates

- The 50 ms & 100 ms timescales contain the best discriminating features
- Combining $g(\mathfrak{W})$ and κ could give a better classification accuracy

Discriminating P/S waves

- P and S waves have different polarization
  - However, this information could be hiding at certain timescales; don’t know which
- So, borrow an old idea from Anant & Dowla, BSSA 1997
  - Create multi-resolution versions of the (Z, N E) waveforms by
    - doing a wavelet transform (Daubechies-20 worked best)
    - zeroing out all wavelets except level l
    - Inverse wavelet transform, to get signal at level *l*
- Compute rectilinearity (F), transverse-to-radial ratios (T) and horizontal-to-vertical ratio (hv) in a moving window (1 sec) at each multi-resolution level *l (these are time-series $g_l(t)$)*
  - Composite across scales (25 ms – 400 ms) as $G(t) = \Pi_l \, g_l(t)$
  - For each P / S arrival compute G_{max} = max(G(t)) in a 1 second window after pick time
    - The idea is that G_{max} are predictors of P versus S waves, G = {rectilinearity, transverse-to-radial ratio, horizontal-to-vertical ratio}

Tests with CFAA data



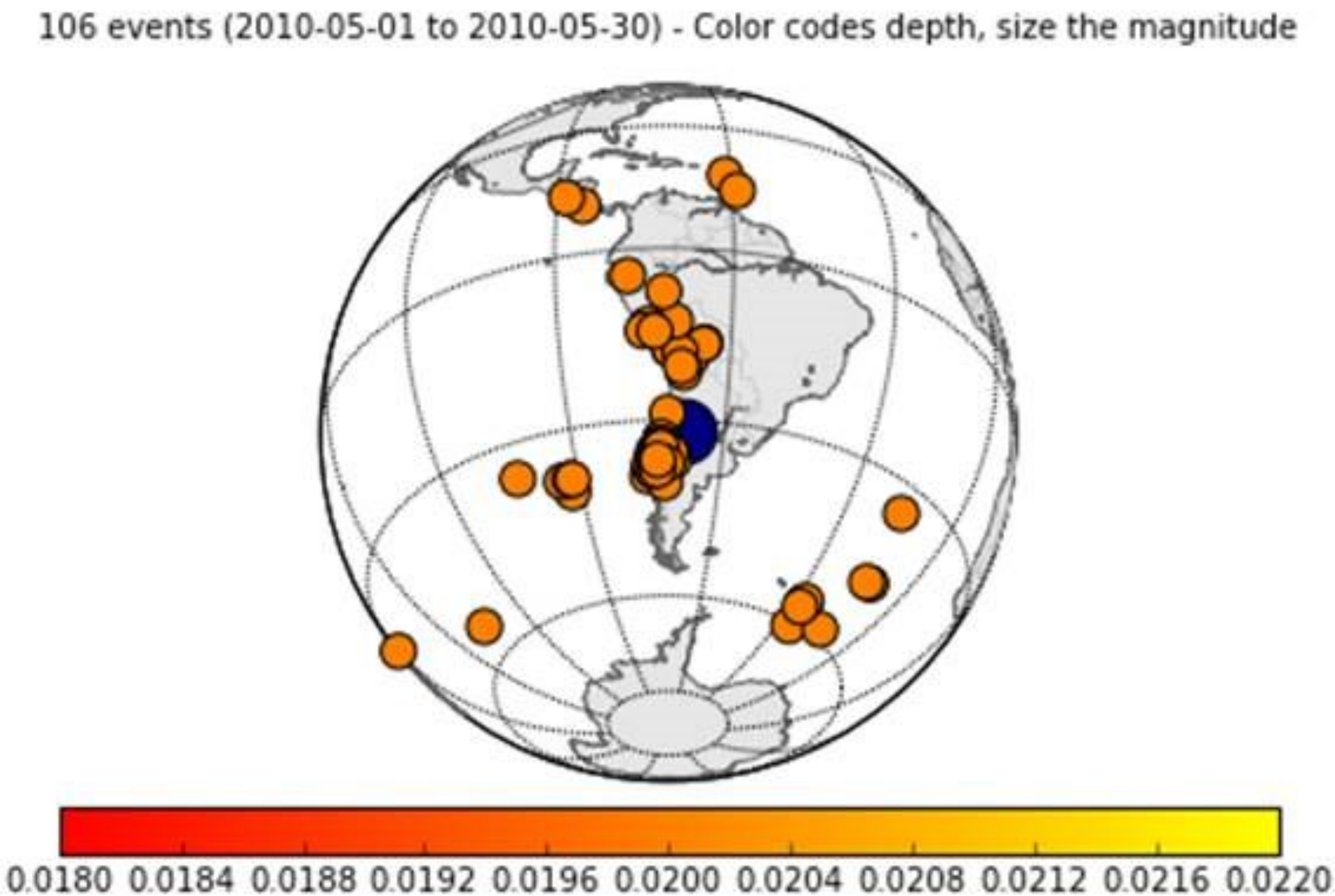
- Classifier trained with 259 samples each of P & S phases
  - SVM with Gaussian kernels
  - Misclassification rate: 12% (7-fold CV)

Conclusions

- We have been investigating MR features to include in a classifier to discriminate noise/P/S waves in 3-component seismograms
- We find that
  - Distribution of wavelet coefficients and the ratio of pre- and post-pick wavelet energies are good discriminators for noise/P
  - The relevant features lie at the 50 ms and 100 ms timescales
- For P/S classification we need a wavelet-enhanced version of 3 polarization metrics
  - Wavelets are used to zero-in on information in the (25, 400) ms timescales
    - Daubechies-20 worked across all 2 x 259 signals from CFAA
  - Composited rectilinearity, transverse-to-radial and horizontal-to-vertical ratios are the discriminators
  - SVM classifier gives 12% misclassification (7-fold CV)

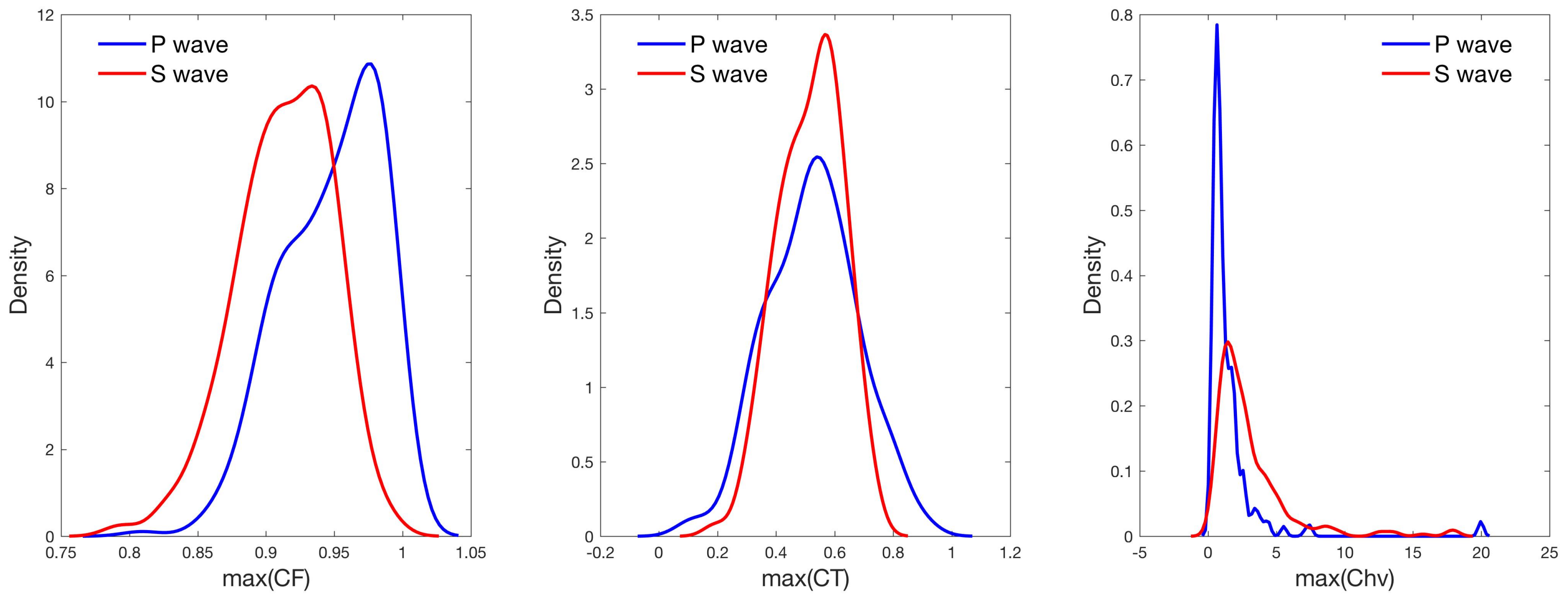
BACKGROUND

Station & events



- Regional events mostly

Tests with composite's max values



- If F_{max}, T_{max}, and hv_{max} are predictors, their distributions for P and S waves must be very different
  - And unlike Anant & Dowla, did not have to select a different wavelet for each event
  - So, if $\{F_{max}, T_{max}, hv_{max}\}$ are predictors, can we make a P/S classifier?