Feature-based Magnitude Estimates for Small, Nearby Earthquakes in the Yellowstone Volcanic Region

Alysha Armstrong, Ben Baker, Keith Koper

UNIVERSITY OF UTAH SEISMOGRAPH STATIONS

INTRO

• Conventional magnitude methods can fail or be prohibitively time consuming during periods of high seismicity rates, such as the many earthquake swarms in the Yellowstone volcanic region.
• We introduce a machine learning method that uses features derived from short-duration waveform segments of individual phase arrivals and event source parameters to predict local magnitude (M_L).

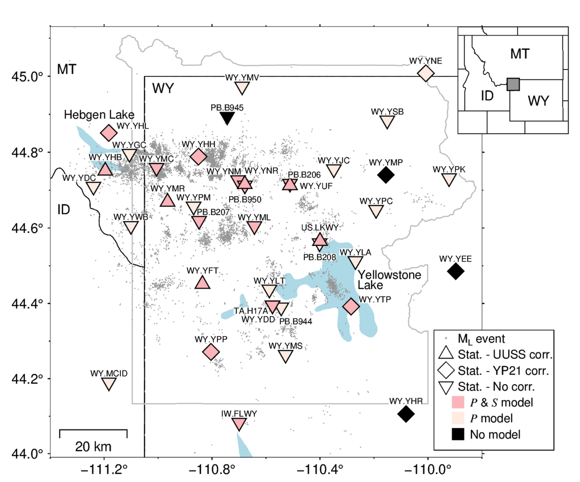
KEY POINTS

Our approach:
• accurately estimates M_L for ~0–3.5
• works for events with small temporal separation
• estimates network M_L using single stations
• does not require 3C broadband stations for M_L
• is easily interrogated, updated, and modified
• maintains consistency with the UUSS catalog
• requires the event location
• can utilize the many phase arrivals, particularly *S*, available in deep-learning enhanced catalogs

DATA

• **Model targets**: High-quality event M_L values from 8,475 earthquakes during 1/10/12–1/1/24
• **Model inputs**: Features derived from 0.95–1.40 s of pre-arrival noise and 2.55–3.60 s of post-arrival signal
  • Separate feature datasets for *P* and *S* arrivals
  • Start with 38 frequency-domain, 4 time-domain, and 3 location-based candidate features:

Name	Abbreviation	Equation or explanation	Transform	Type
Amplitude Ratio	ratio [freq.]	The ratio of the average signal and average noise at the specified corner frequency (freq.) between 1–18 Hz	$\log_{10}$	Time/Freq.
Average Amplitude	amp. [freq.]	The average signal at the specified corner frequency between 1–18 Hz	$\log_{10}$	Time/Freq.
Signal Dominant Frequency	sig. dom. freq.	The dominant frequency in Hz of the phase arrival	$\log_{10}$	Freq.
Signal Dominant Amplitude	sig. dom. amp.	The maximum amplitude of the signal dominant frequency	$\log_{10}$	Freq.
Signal Maximum Amplitude	sig. max. amp.	The difference of the maximum signal amplitude and the minimum signal amplitude	$\log_{10}$	Time
Noise Maximum Amplitude	noise max. amp.	The difference of the maximum amplitude and the minimum amplitude in the noise window	$\log_{10}$	Time
Signal Variance	sig. var.	The variance of the signal time series from zero	$\log_{10}$	Time
Noise Variance	noise var.	The variance of the noise time series from zero	$\log_{10}$	Time
Source-receiver Distance	distance	The distance from the event epicenter to the receiver in km	$\log_{10}$	Event
Source-receiver Back Azimuth	back az.	The distance from the receiver to the event epicenter in degrees. If using a linear model, the sine is used	sine (if linear model)	Event
Source Depth	depth	The depth of the event in km relative to sea-level	-	Event



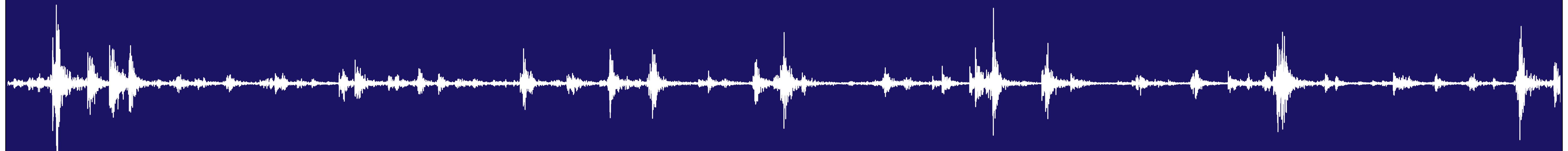
Earthquake and station locations in the Yellowstone region.

Map of earthquakes with an M_L (circles) cataloged by the UUSS near Yellowstone National Park (solid gray line). Triangles, diamonds, and inverted triangles show the locations of stations with a UUSS station correction, with only a Holt *et al.* (2022) station correction (YP21), and with no correction, respectively. The UUSS only uses stations with corrections to compute M_L. The station symbol is filled with a dark shade if it has a *P* and *S* model, filled with a light shade if it only has a *P* model, and filled with black if it has no model.

Contact information:
alysha.armstrong@utah.edu

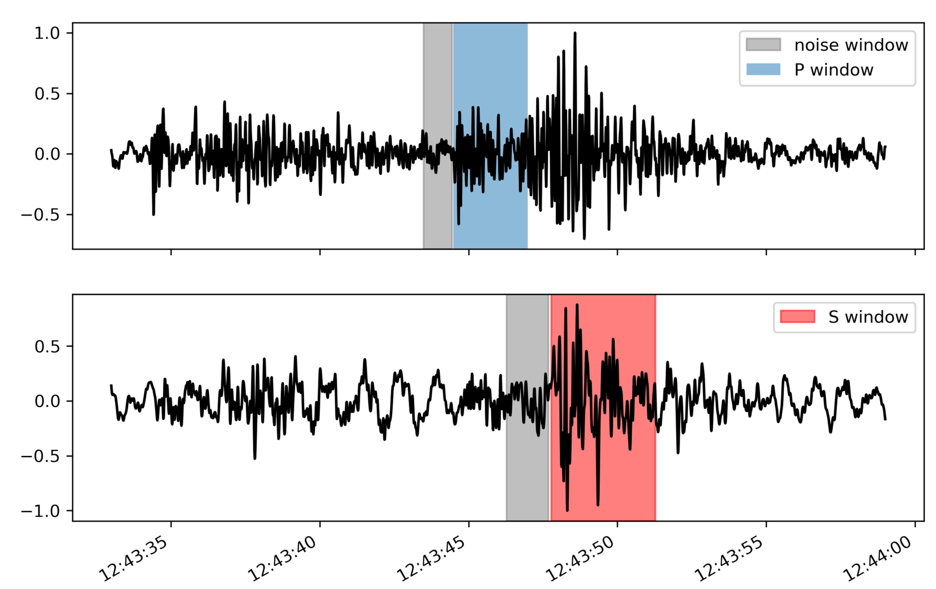
Personal website, optional poster download

Using the earthquake location & 4 waveform features we can accurately compute event magnitude during periods of elevated seismicity.

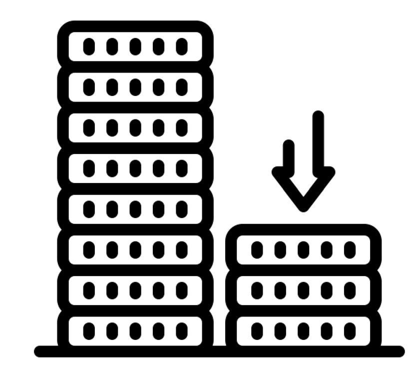


Overview:

1. We **extract time and frequency candidate features** from **short-duration** waveform segments around individual *P* and *S* arrivals.



2.

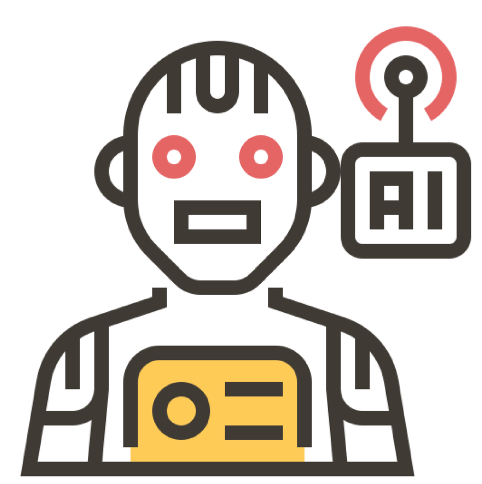


We use our generalizable approach to **reduce the number of features** from 45 to 7:

distance, depth & back-azimuth

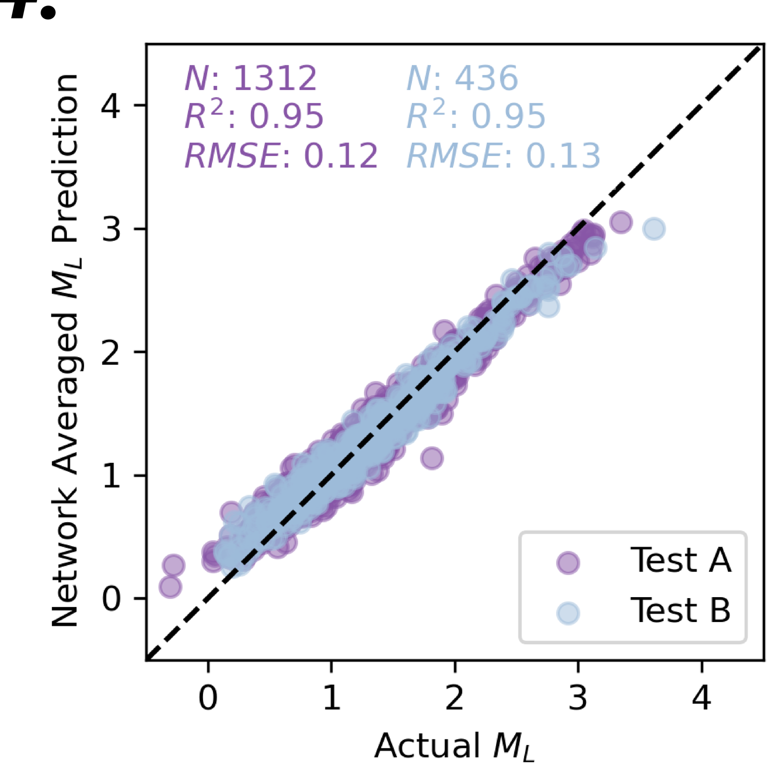
amplitude/SNR proxies: 3 signal & 1 noise

3.



We **train a machine learning model** for each station to **predict local magnitude** using the 7 selected features.

4.



We average all model predictions to create network magnitudes, which **are generally within ~0.13 mu** of the actual magnitudes.

METHODS

Training and Testing Datasets

• Use an 80:20 split of all events in the *P*- and *S*-feature catalogs occurring before 1/1/23 as the training set and **testing set A**.
• Use features computed from events during 1/1/23–1/1/24 as **testing set B**.
• 35 stations in the *P* dataset and 18 in the *S* dataset
  • ≥ 300 *P* station training examples and ≥ 150 *S*

Recursive Feature Elimination Algorithm (RFEA)

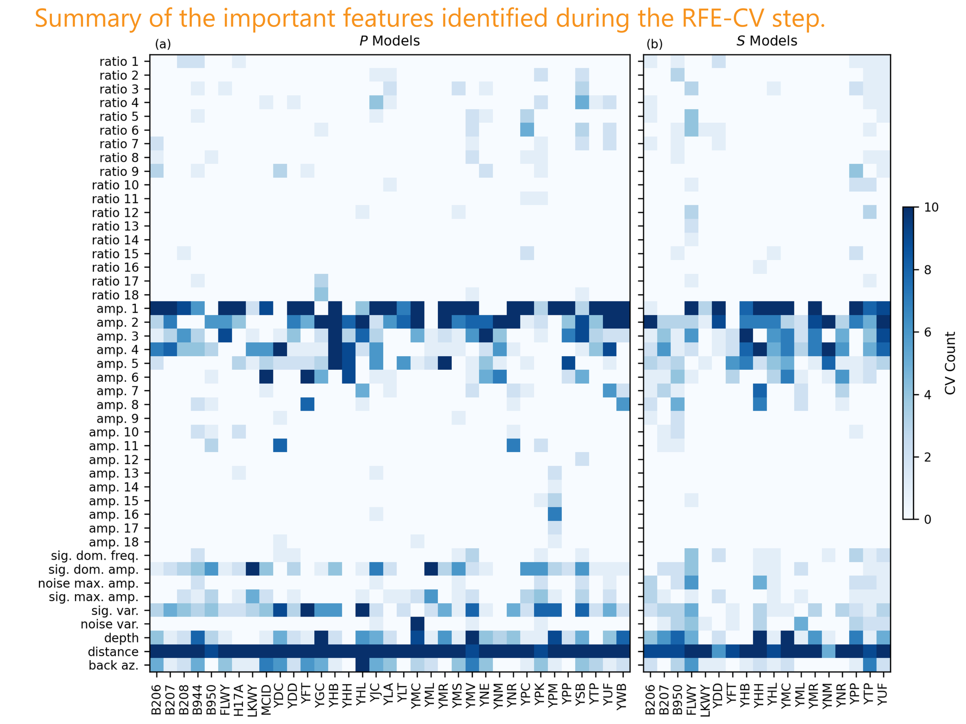
• We use a RFEA that both simplifies and improves the predictive performance of the machine-learning models and limits feature selection bias.

```
for each station
    Split the full training set (X) into K_1 folds
    for each k in K_1 folds
        Create training set k with all examples not in fold k
        Create testing set k with all examples in fold k
        Reduce the number of features from 45 to 19 by selecting the 5 amps. and 5 ratios with
        the largest mutual information in training set k
        Rank the 19 features in training set k using the importance estimator
        for each i in 1,...,19
            Make training and testing set k_i with the i most important features
            Select the predictor model hyperparameters using a K_2-fold cross-validation grid search
            Train the predictor model using training set k_i and the selected hyperparameters
            Evaluate the model on testing set k_i
            Save the testing R^2 value
        end
        Save n_k,max, the number of features corresponding to the max. testing R^2 value
        Save n_k,min, the smallest number of features with a testing R^2 value within one std. err. of
        the max. R^2 value
        Save the feature rankings
    end
    Count the number of times each feature is in n_k,max and n_k,min for all folds
end
Examine the number of times each feature is in n_k,max and n_k,min across all stations
Select features that are generally important for many stations
```
RFE-CV

Pseudocode for our RFEA.

We use a two-step process, in which we first identify the most important features at each station during recursive feature elimination with cross-validation (RFE-CV) and then select a common feature set for all stations. We use a decision tree for the importance estimator and an SVM for the predictor model.

Summary of the important features identified during the RFE-CV step.



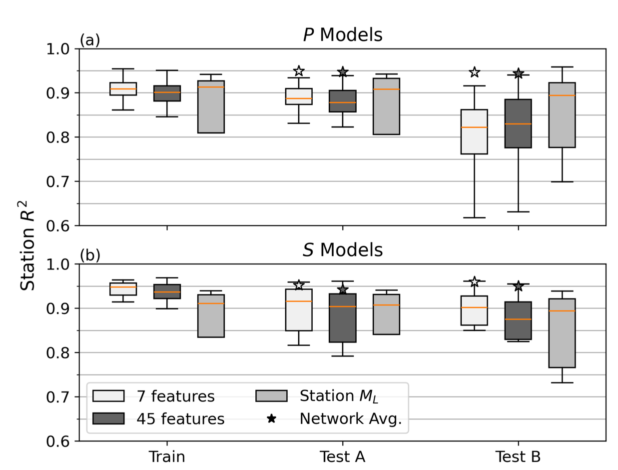
Selected waveform features for all stations:
P: amplitude at 1 & 2 Hz, signal & noise variance
S: amplitude at 1, 2 & 4 Hz, noise variance

Support Vector Machines (SVM)

• We use an SVM with radial basis function kernel to learn a mapping from the features to M_L
• We train one model per station-phase pair
  • *P*: 35 models, *S*: 18 models

RESULTS

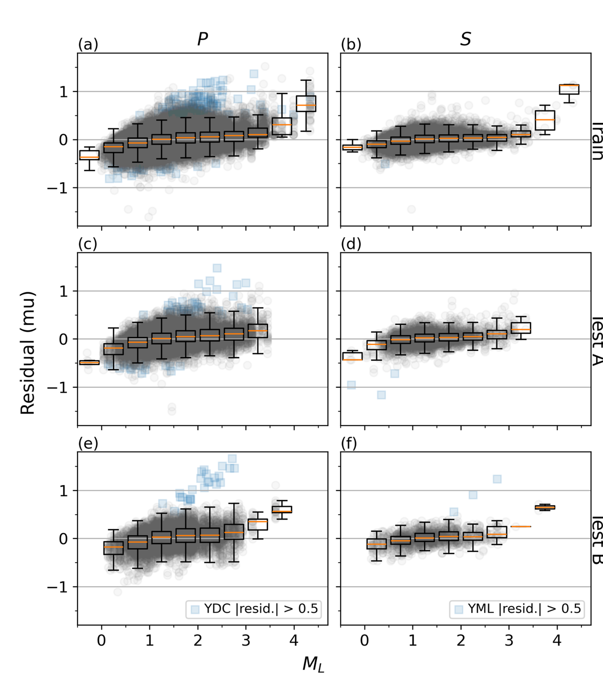
• Generally predict the event M_L within ~0.25 mu for individual stations and ~0.13 mu when averaging



Summary of station R^2 values.

• reducing the features from 45 to 7 leads to a small performance increase
• our *S* and network averaged predictions have better agreement with the event M_L values than the original station M_L values

We show results for the SVM models trained with the selected 7 features (light), the SVM models trained with all 45 features (mid), and the station M_L values (dark). The R^2 values of the network averaged (avg.) test set predictions are marked by stars for the SVM models. All R^2 values are relative to the event M_L.



Residuals (actual – predicted) as a function of M_L.

• generally perform well for events with 0.0 < M_L < 3.5
• slightly overestimate the smaller magnitudes and underestimate the larger magnitudes, likely due to limited training examples.
• we plan to examine probabilistic machine learning models to remove unreliable predictions

Squares show absolute residuals greater than 0.5 mu (|resid.| > 0.5) for *P*-model WY.YDC in a, c, and e and for *S*-model WY.YML in b, d, and f. These two models have anomalously poor performance on testing set B. All other residuals are shown as circles. The boxplots show the distribution of the residuals in 0.5 mu bins starting at -0.5.