

# Automated classification of ground motion record quality using machine learning

Xavier Bellagamba<sup>1</sup>, Robin Lee<sup>1</sup> and Brendon A. Bradley<sup>1</sup>

<sup>1</sup>University of Canterbury

xavier.bellagamba@pg.canterbury.ac.nz



QuakeCoRE  
NZ Centre for Earthquake Resilience  
Te Hirainga Rū

UC  
UNIVERSITY OF  
CANTERBURY  
Te Whare Wānanga o Waitaha  
CHRISTCHURCH NEW ZEALAND

## Motivation

Densification of strong-motion station networks, their increased sensitivity, and the desire to use smaller magnitude data, is leading to exponentially-increasing ground motion datasets. Despite the improving reliability of seismic instrumentation, recorded ground motions are not of uniform quality, and the exponentially-increasing dataset sizes require automated quality assessment in order to be scalable. Here we propose a two-layer neural network that takes key ground motion metrics as inputs to automatically determine the quality of the records.

## Training ground motion dataset from the Canterbury region, New Zealand

As large magnitude earthquakes are rare, their recorded ground motions are of prime importance and should therefore be manually screened. Hence, only ground motions from earthquakes with moment magnitude between 3.5 and 5.0 are considered within this study as shown in Figure 1. The ground motion dataset used to train the neural network is comprised of 3989 records from 168 earthquakes that occurred in Canterbury, New Zealand. These ground motions were recorded between 2003 and 2016 by 55 strong motion stations. Figure 1 shows the map of the strong motion station network, earthquakes sources and schematic ray paths of the observed ground motions.

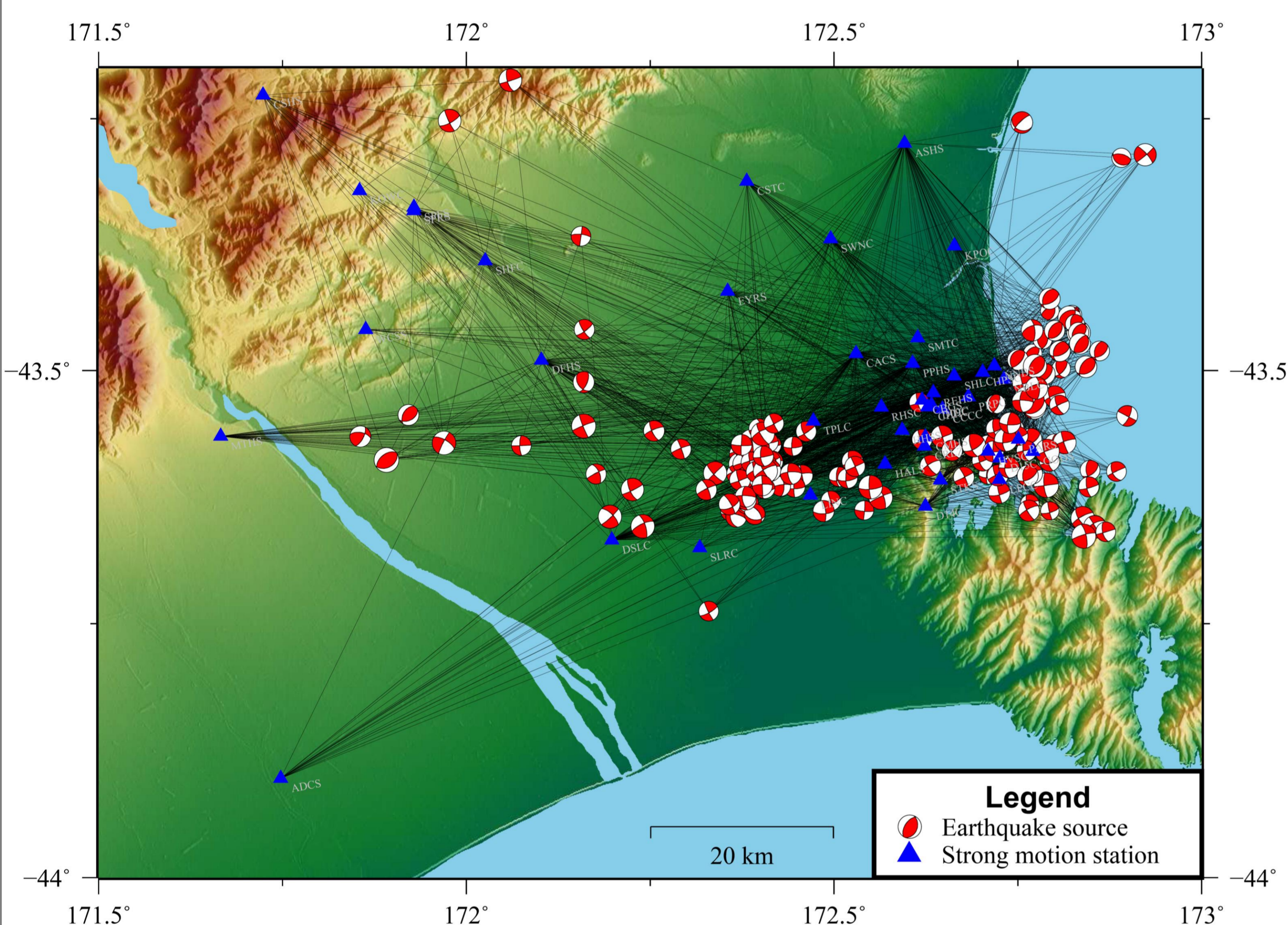


Figure 1: Map plot of 168 earthquake sources which produced observed ground motions at the 55 recording stations considered in this study. Schematic ray paths of observed ground motions are also shown as black lines.

In the dataset, the key ground motion metrics calculated from the records are expressed as the geometric mean of the two horizontal components of the ground motion. The vertical component has been excluded from this analysis. To train the neural network, ground motion record quality has been manually classified as highest quality records; high quality records; average quality records; low quality records; and lowest quality records.

## Trained neural network and future work

The neural network was trained with both the highest and lowest quality records. Its architecture was selected via a grid search combined with a K-Fold cross-validation scheme. Applied to a ground motion, the final neural network assigns two scores: one for resemblance to a highest quality records and one for the resemblance to a lowest quality records. To reduce the number of misclassified ground motions of low quality (i.e. average to lowest quality records must be rejected), multiple acceptance thresholds are tested. It can be observed in Figure 3 that a threshold of 0.9 seems to eliminate most of the average to lowest, while also significantly reducing the number of ground motions available for validation. The developed neural network will further be tested on newly classified datasets (e.g. the Wellington region) to ensure its broader validity. The threshold effect will also be analyzed on the final validation results. Additionally, the introduction of potentially automated-selection-induced biases will also be investigated. Final results will serve to validate results from the New Zealand physics-based ground motion model for small-to-moderate magnitude earthquakes.

Figure 3: Influence of the acceptance threshold on (a) the number of accepted ground motions ; and (b) the proportion of accepted ground motions for each categories. Numbers between brackets indicate the original number of ground motions in each category.

## Differences between high and low quality ground motions and data treatment

To determine the quality of a ground motion record, the primary metric that is checked is the signal-to-noise ratio (SNR). However, there are many other factors which could compromise the quality of a ground motion. Hence to refine the selection process, 20 additional metrics characterizing duration, peak signal and Fourier amplitude are utilized. As neural networks yield better results with uncorrelated and amplitude-like inputs, the variables are standardized and decorrelated using a Mahalanobis transform.

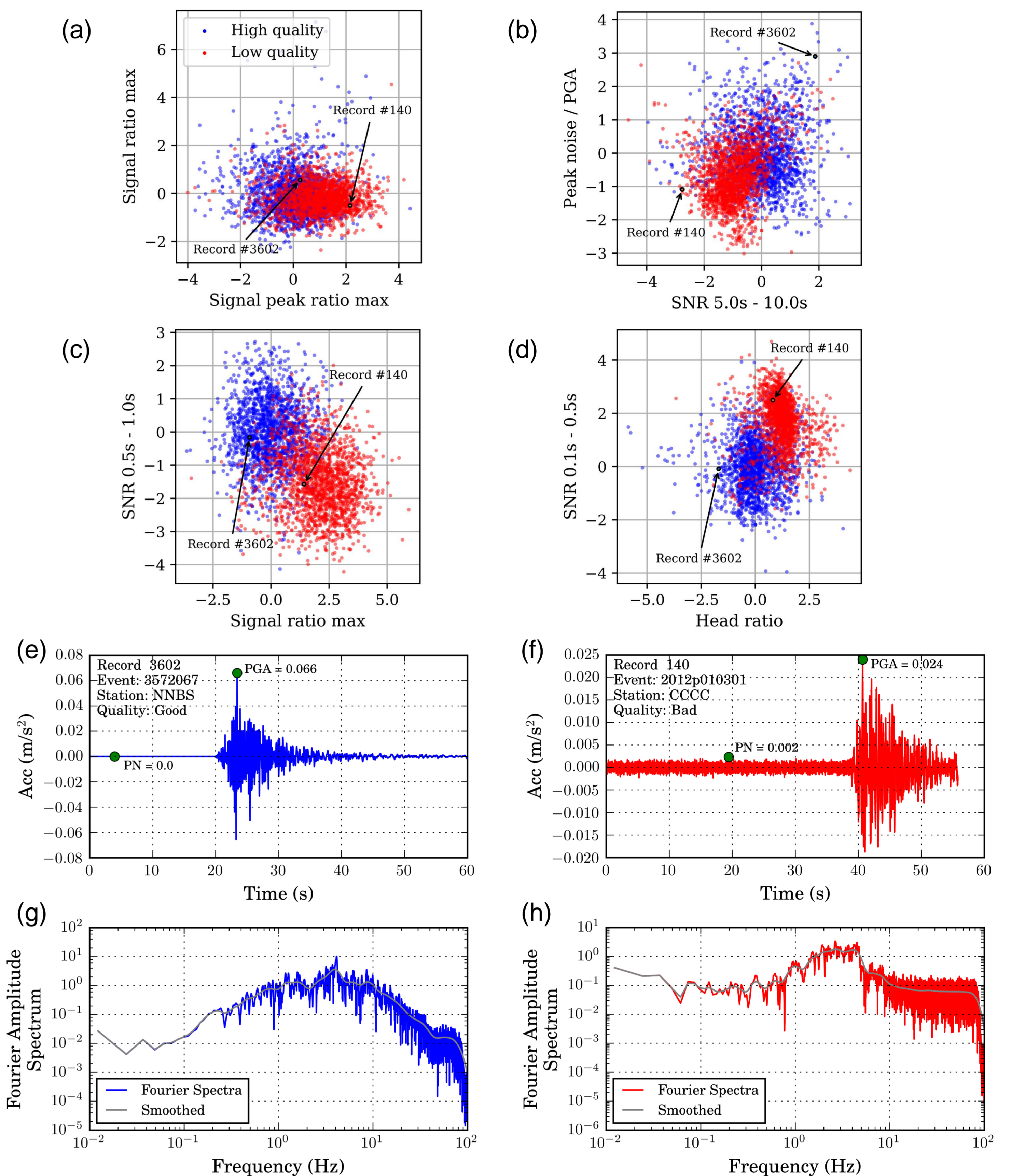


Figure 2 (a)-(d): Scatter plots of four variable couples in their respective transformed spaces showing the differences between the HSQR and LSQR distributions; (e)-(f) Signals of the HSQR and LSQR indicated in Figure 2 (a)-(d); and (g)-(h) Fourier spectra of the same records.

Figures 2 (a)-(d) show four couples of standardized, decorrelated variables, and Figures (e)-(f) the signal and Fourier spectra of a single component of one highest and lowest quality records. It can be observed on Figures 2 (a)-(d) that none of the variable couples could be used to linearly separate high and low quality records. However, neural networks combine classifying planes across multiple dimensions, allowing a more accurate classification. Figures (e)-(h) illustrate some of the potential differences a highest and a lowest quality records can have (e.g. a high peak noise).

