

Multidimensional classification of burst triggers from the fifth science run of LIGO

Soma Mukherjee
CGWA, UTB
GWDAW11, Potsdam, 12/18/06
LIGO Tech doc # LIGO-G060648-00-0

Why and What

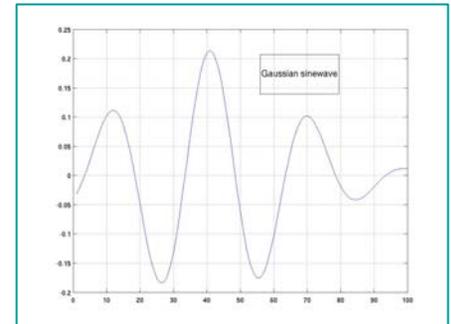
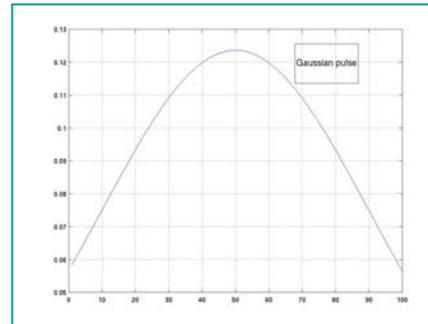
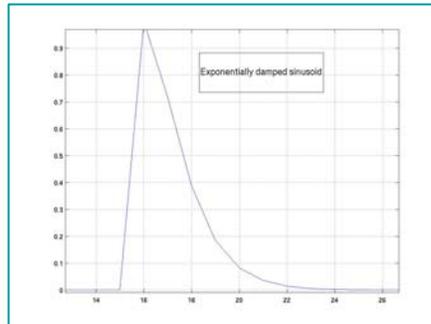
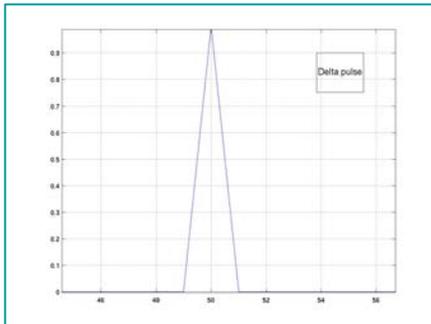
- *kleine Welle* (*Blackburn, Katsavounidis et. al.*) is an algorithm that picks up burst triggers from the gravitational wave, auxiliary and environmental channels in LIGO. It generates several gigabytes of trigger database containing information about the physical properties of the burst triggers. The purpose of this analysis is to mine the trigger database to see if the triggers can be categorized in different groups that share common properties. This will lead to effective dimensionality reduction of the problem since the number expected groups will be a countable small number and each group, to some extent, uniform in character. The physical motivation here is that this could become a powerful veto mechanism.
- Out of many existing classification approaches, we adopt the hierarchical classification algorithm. This scheme is based on computation of *distances* among the objects in n -dimensions and then grouping them (with a variance accountability criterion) in successive stages. The algorithm thus starts with N objects of one cluster and ends with N clusters of one object each. Statistical tests determine how many clusters are actually significant.
- The study is carried out in a multiple parameter space to include physical properties and shape information of the triggers.

What information to use

A trigger in the kleineWelle database is described by several time, frequency and energy values. Most important properties are *duration* of the triggers, *central frequency* and *signal-to-noise ratio* (snr, calculated from the energy values and pixel numbers).

Additional information can be provided by the shape of the triggers in the time domain i.e. the waveform. A wavelet transform can be used as a reasonable shape representation. However, the best representation of the waveform is the *data itself*.

Since the multidimensional hierarchical classification does not in principle pose any restriction on the number of parameters, the waveform can be folded in into the parameter space.



What we have so far and what comes next

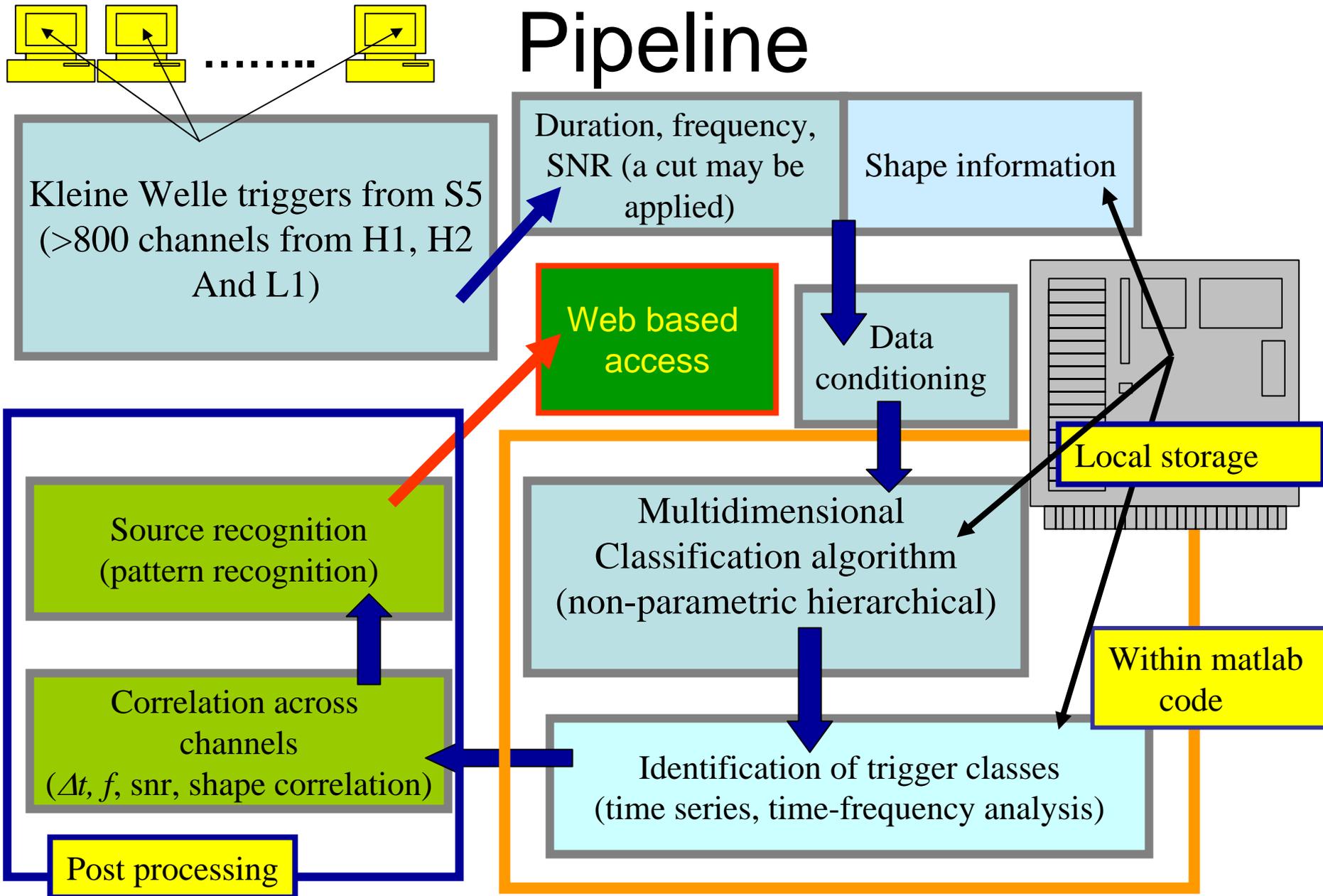
- We start with the kleineWelle trigger database from all channels.
- Data conditioning is applied to whiten the data and clean the narrowband features.
- Analysis database is constructed with Dt, frequency and snr and shape data. Shape information is retained in the form of 32 data points around the peak time.
- A hierarchical classification algorithm is applied.
- At the end of the pipeline, we have information on existing classes (statistics, members, properties) from all channels.

- We need to directly access raw data uninterruptedly. This involves :
 - a. Fast connection for rapid data transfer.
 - b. Data storage.
 - c. Storage for pipeline output.

- Identification of the class properties.
- Correlation between channels.
- Formation of class basis.
- Final step towards direct classification based on pattern recognition.

Veto applications.

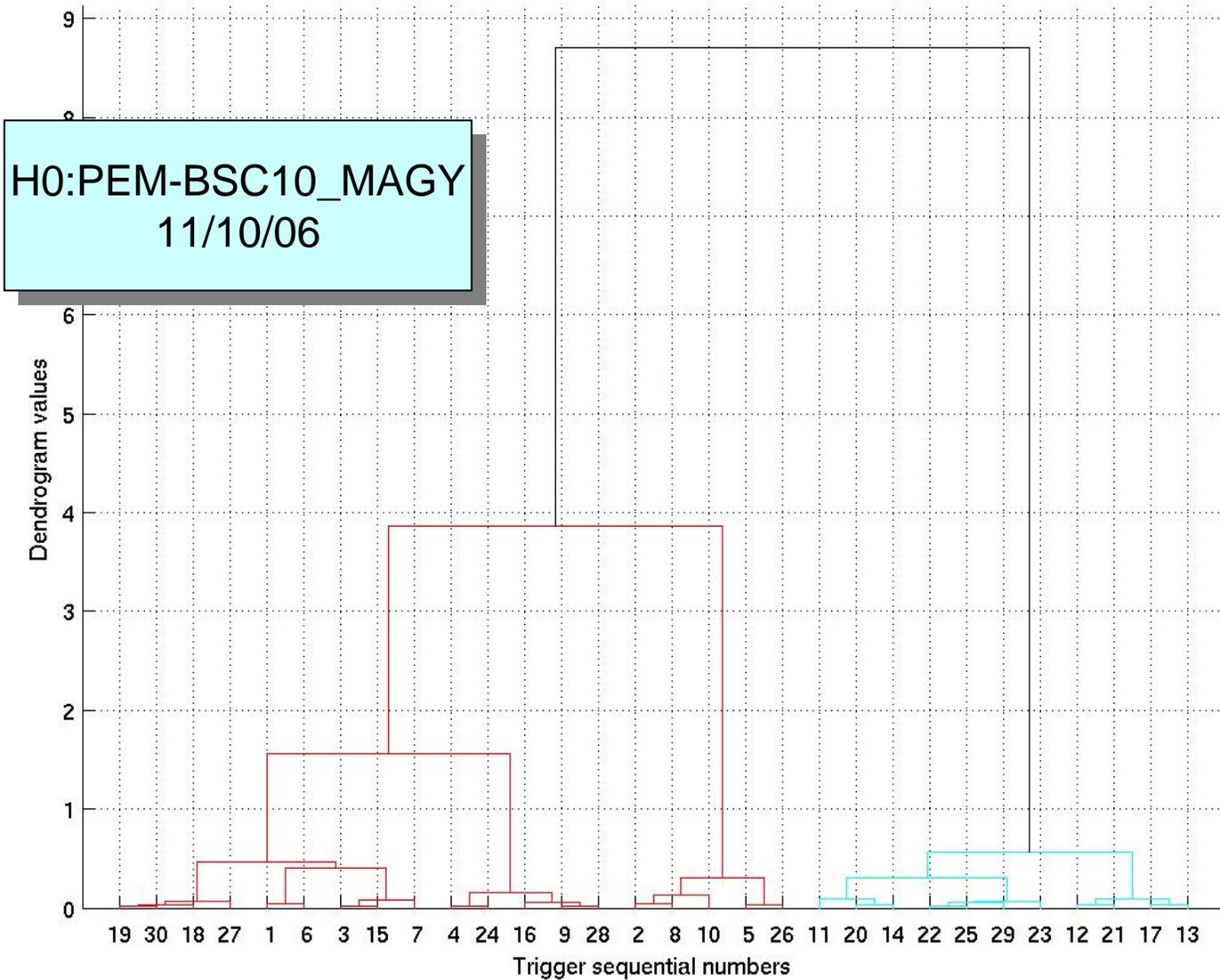
Pipeline



Code status and infrastructure development

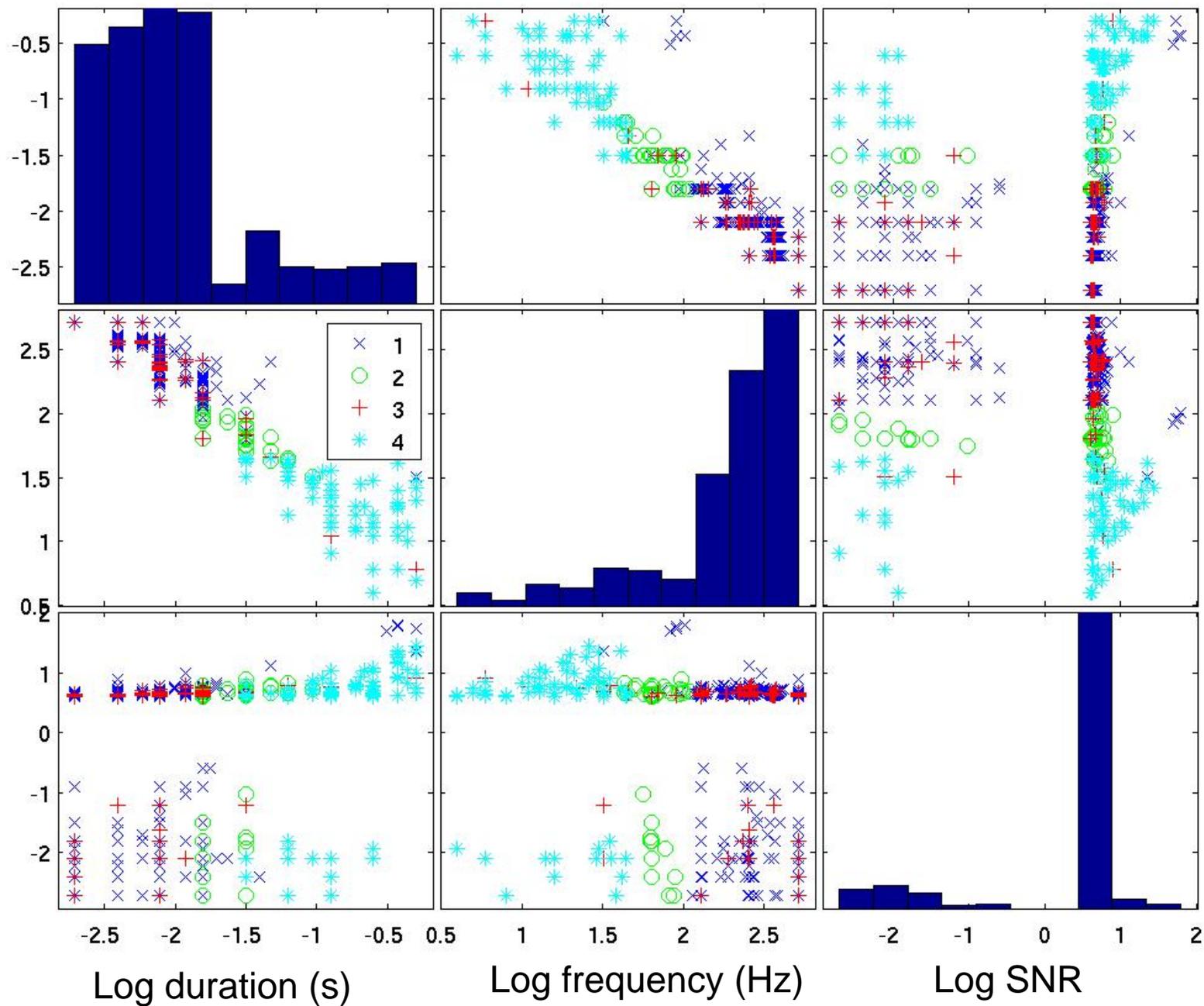
- Scripts in place for distributed run on 16 node lab cluster at UTB. Successful e2e test run performed.
- Hardware acquisition complete for LIGO S5 data storage on a weekly basis. Efforts towards campus bandwidth increase for efficient data transfer.
- The Matlab code upgraded to include the shape information generation, trigger group property recognition.
- Automated archiving of the results in a central repository (triggers in each class, cluster statistics, parameter correlations, inter-channel shape correlations ...)
- Results available over secured web pages.
- Timeline for full pipeline in operation → *very soon* .

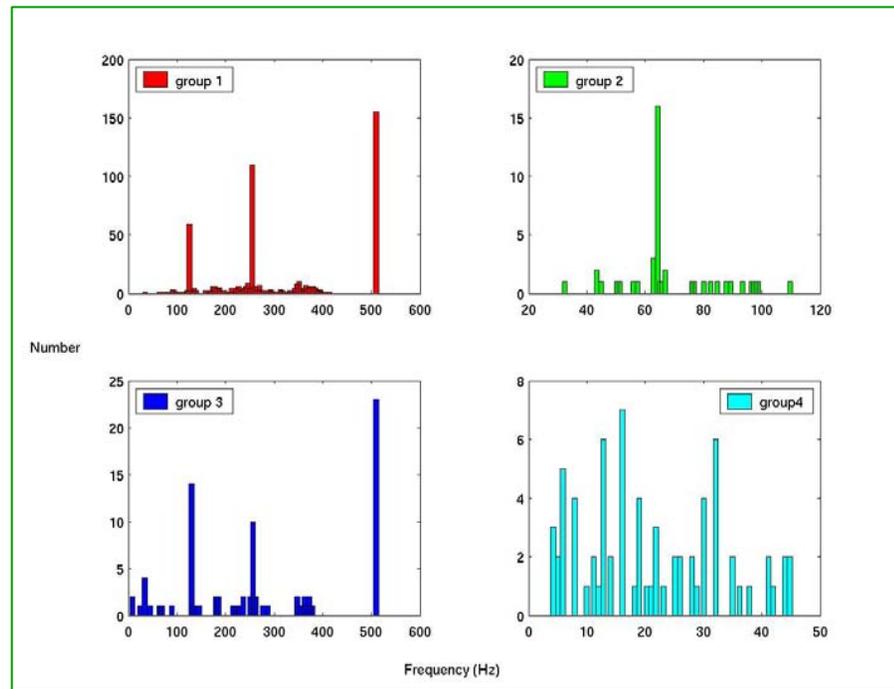
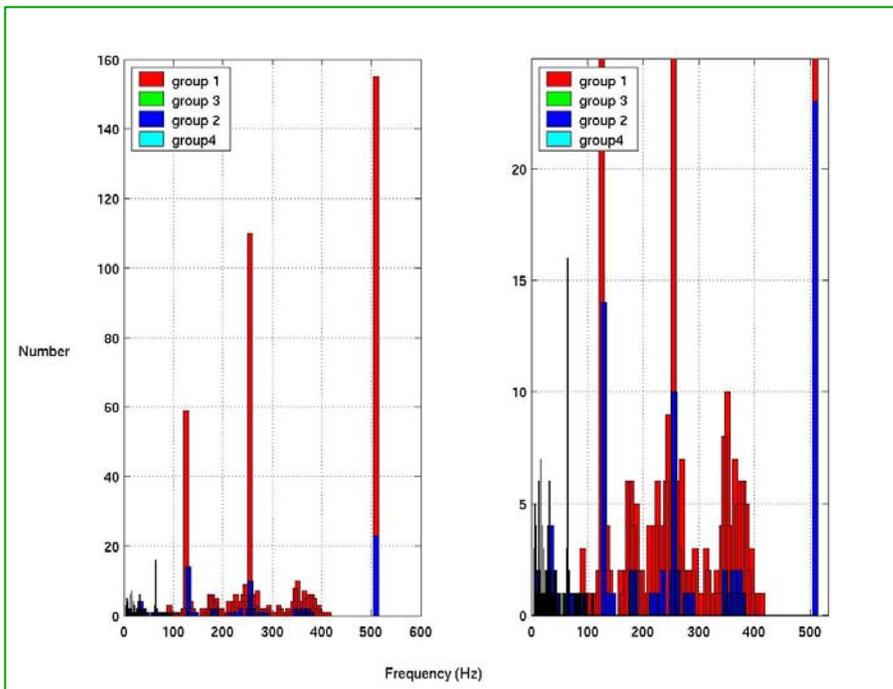
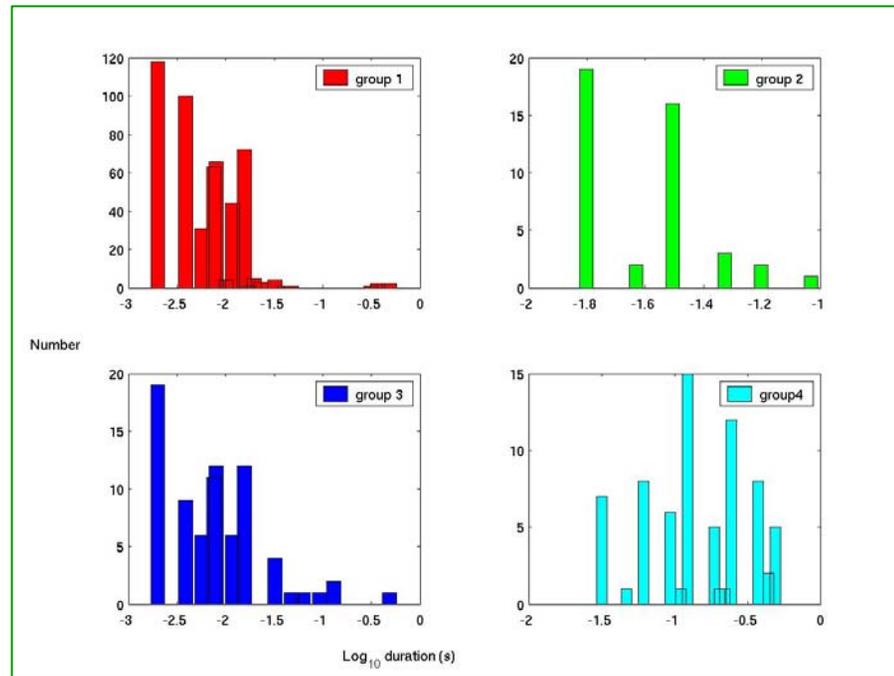
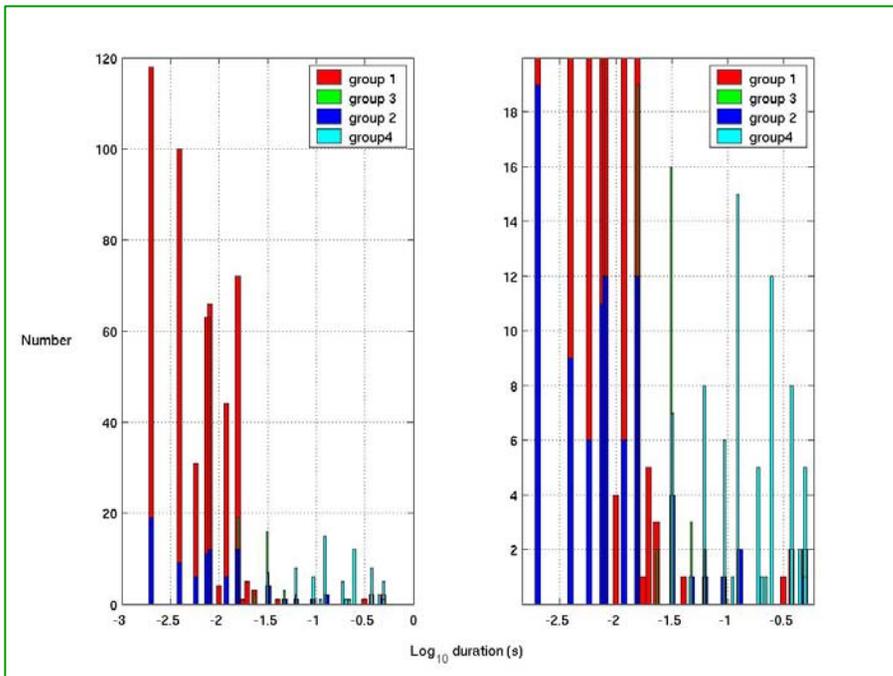
S5 kleine Welle trigger; Dendrogram

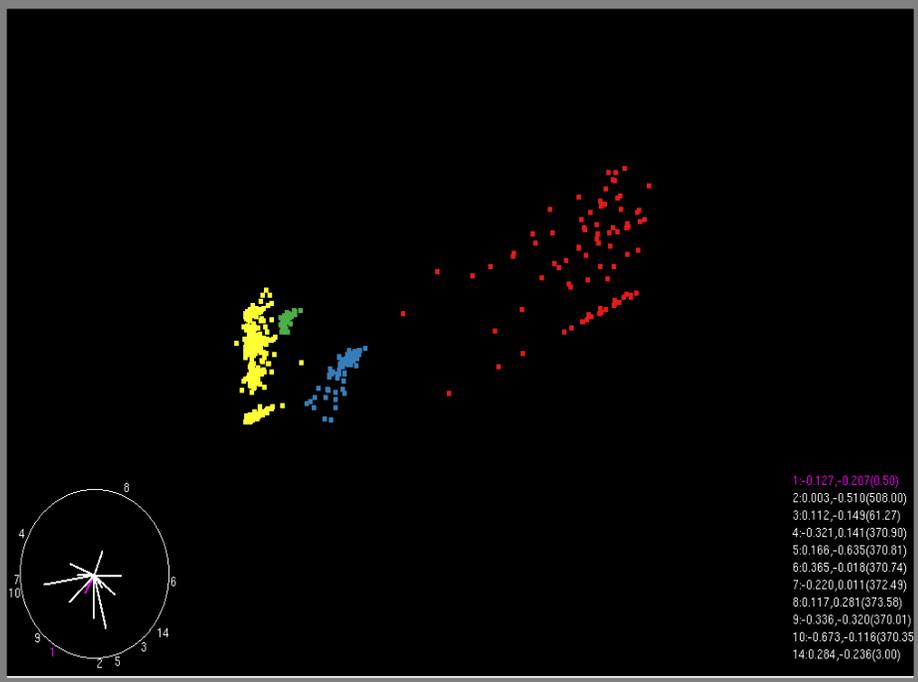
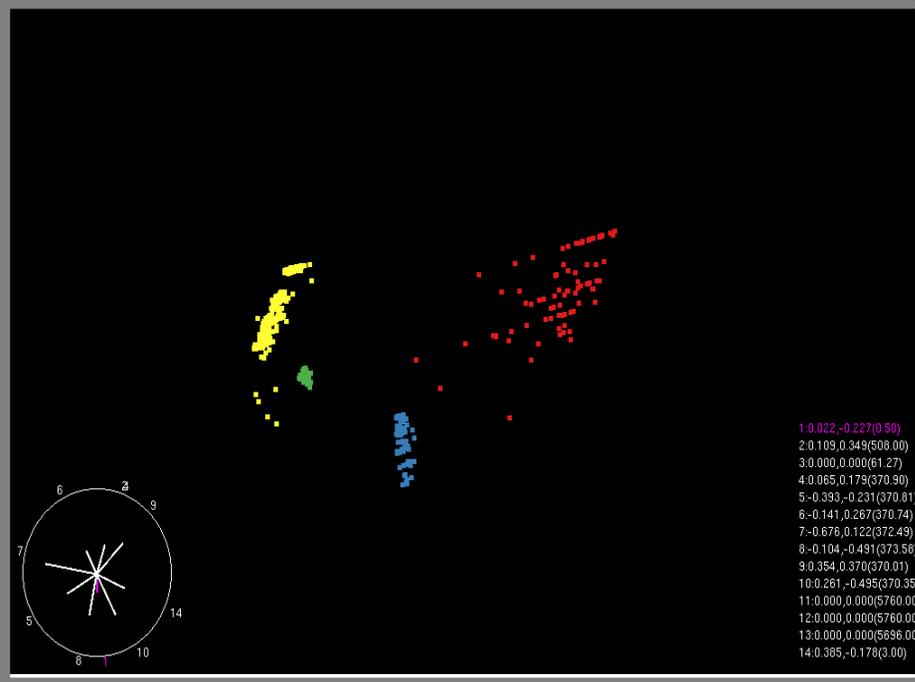
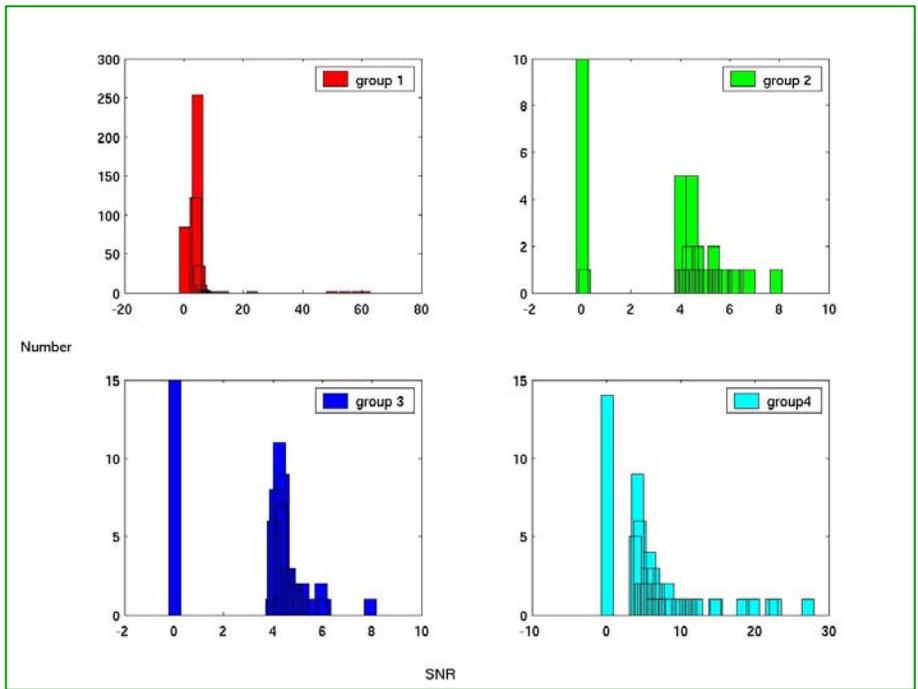
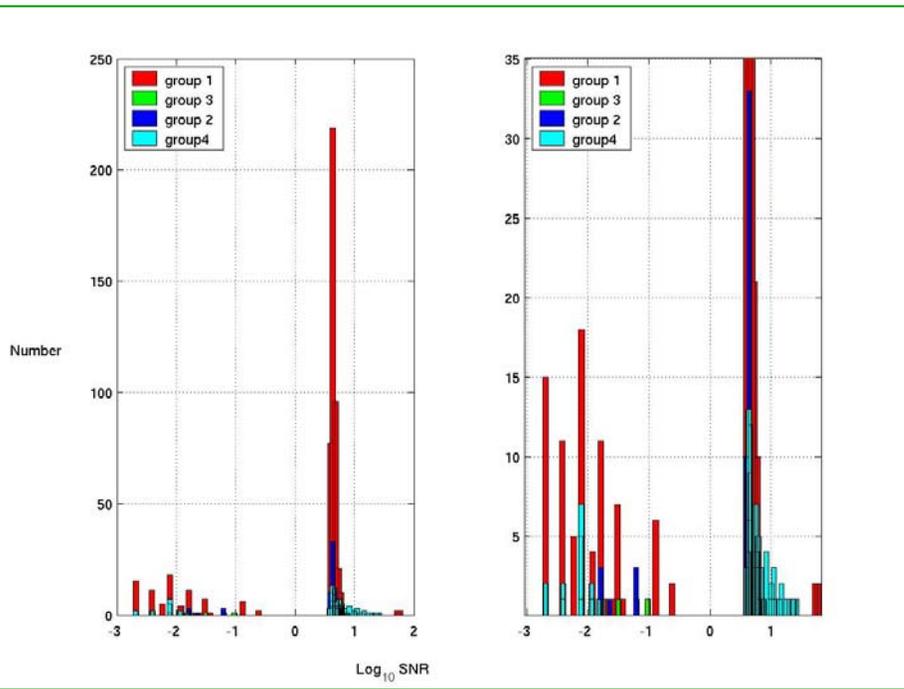


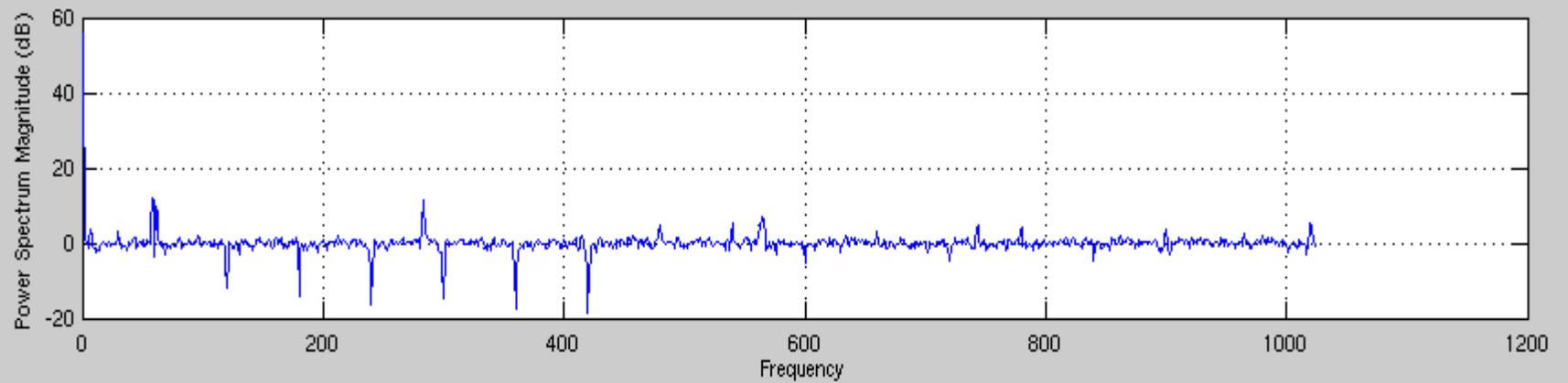
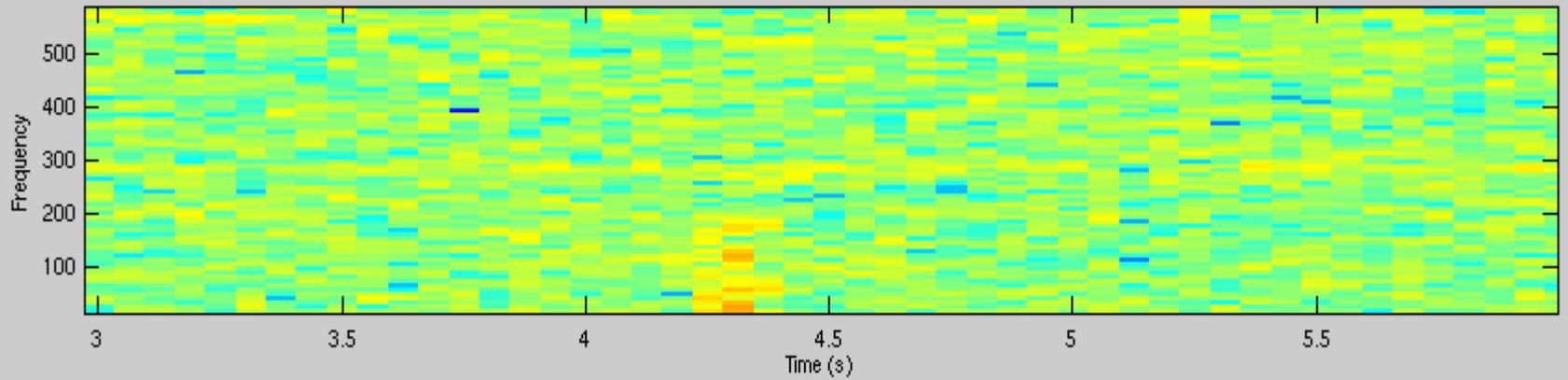
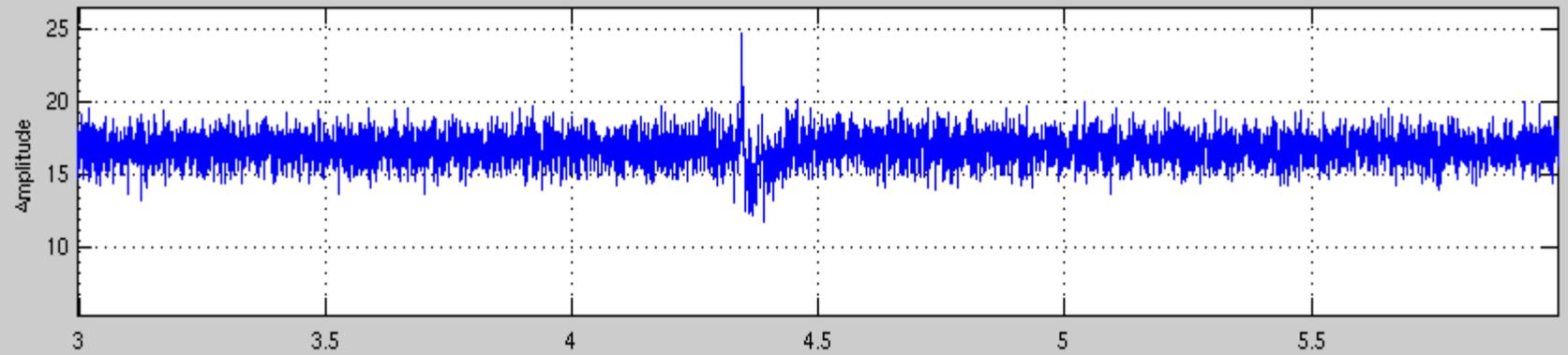
$r = 0.89$ p - values = 0, 0 , 0.05 Number of significant clusters = 4

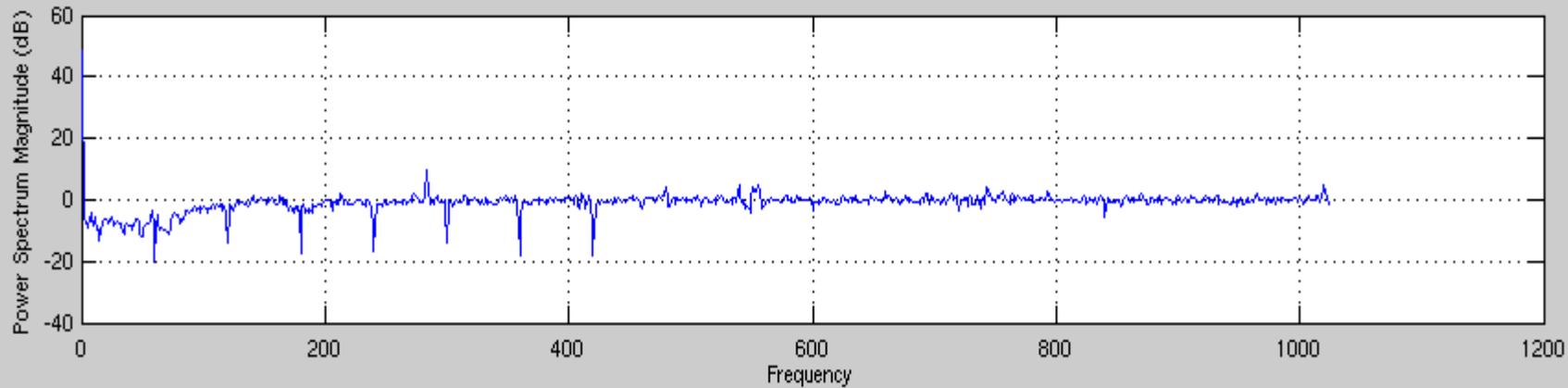
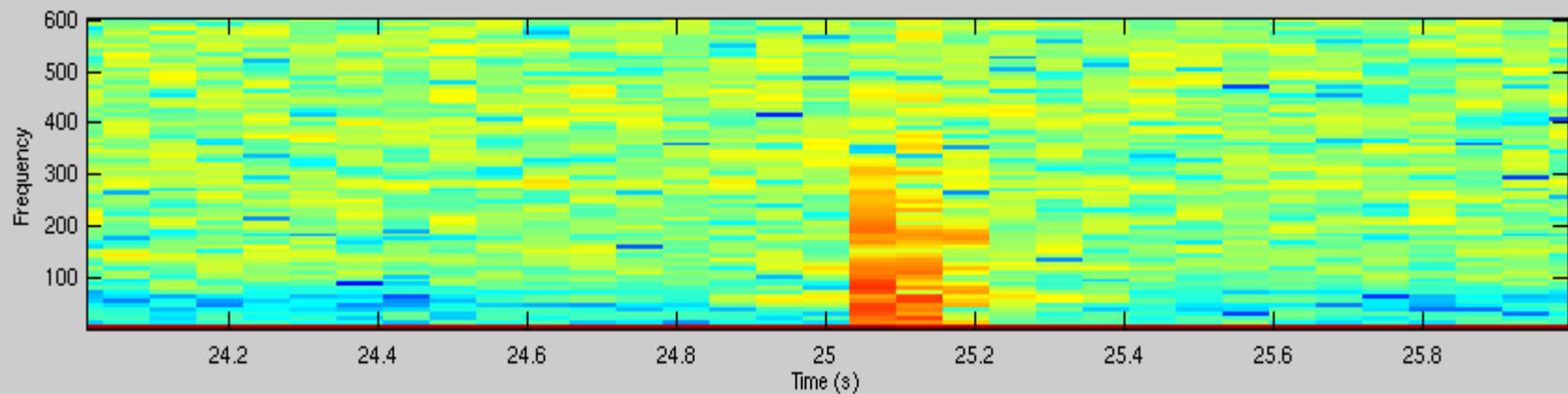
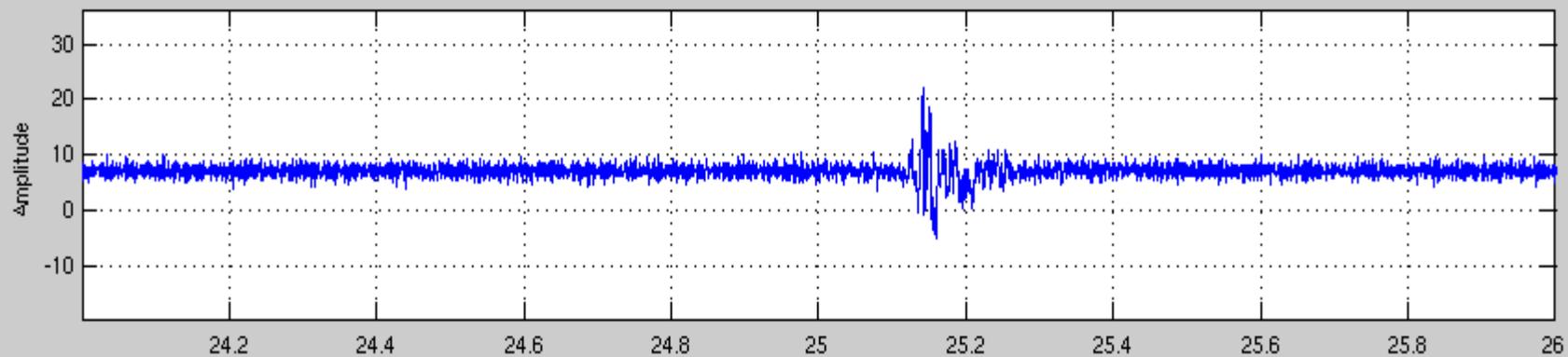
S5 kleineWelle triggers; Group plot and correlations between duration, central frequency & SNR

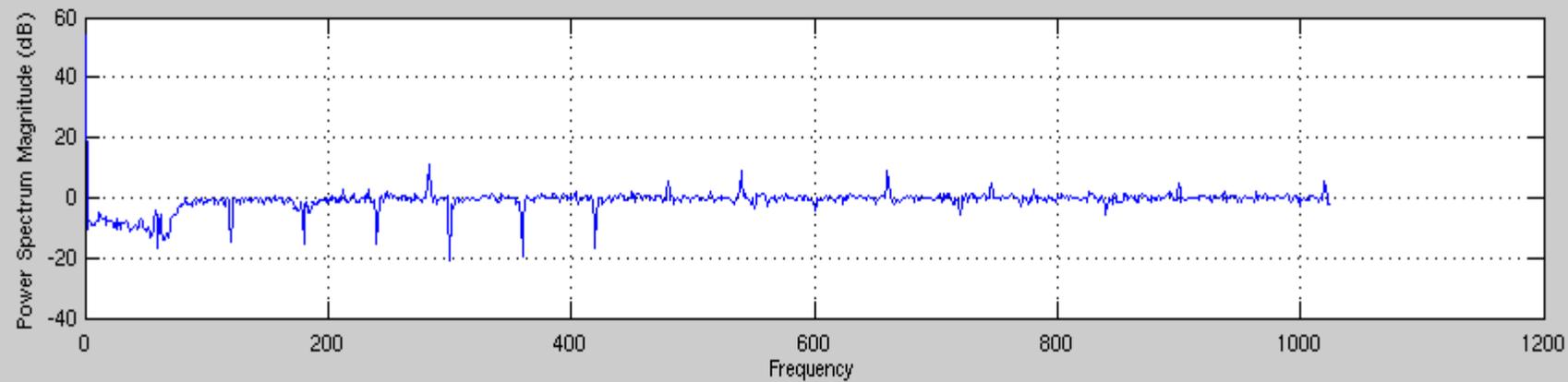
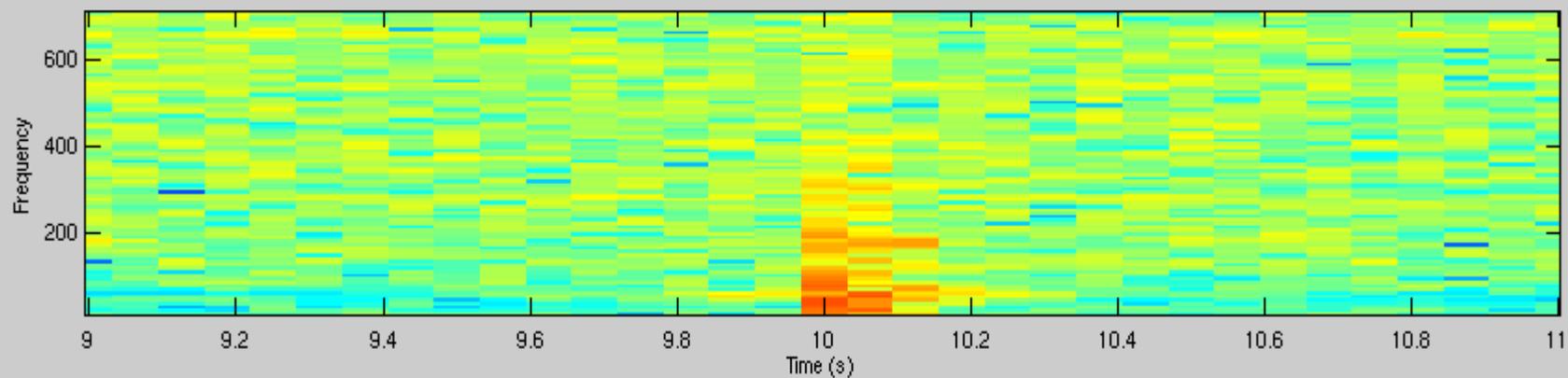
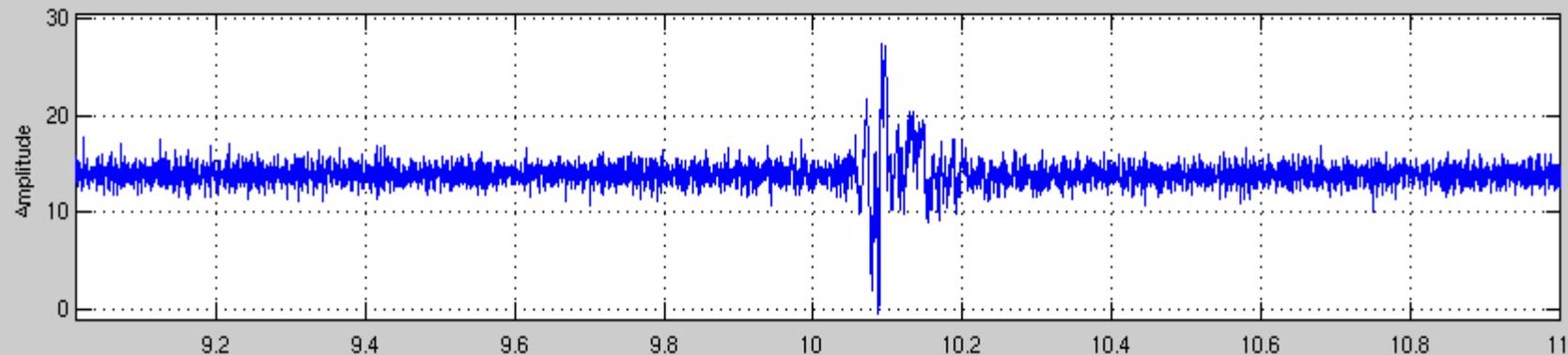






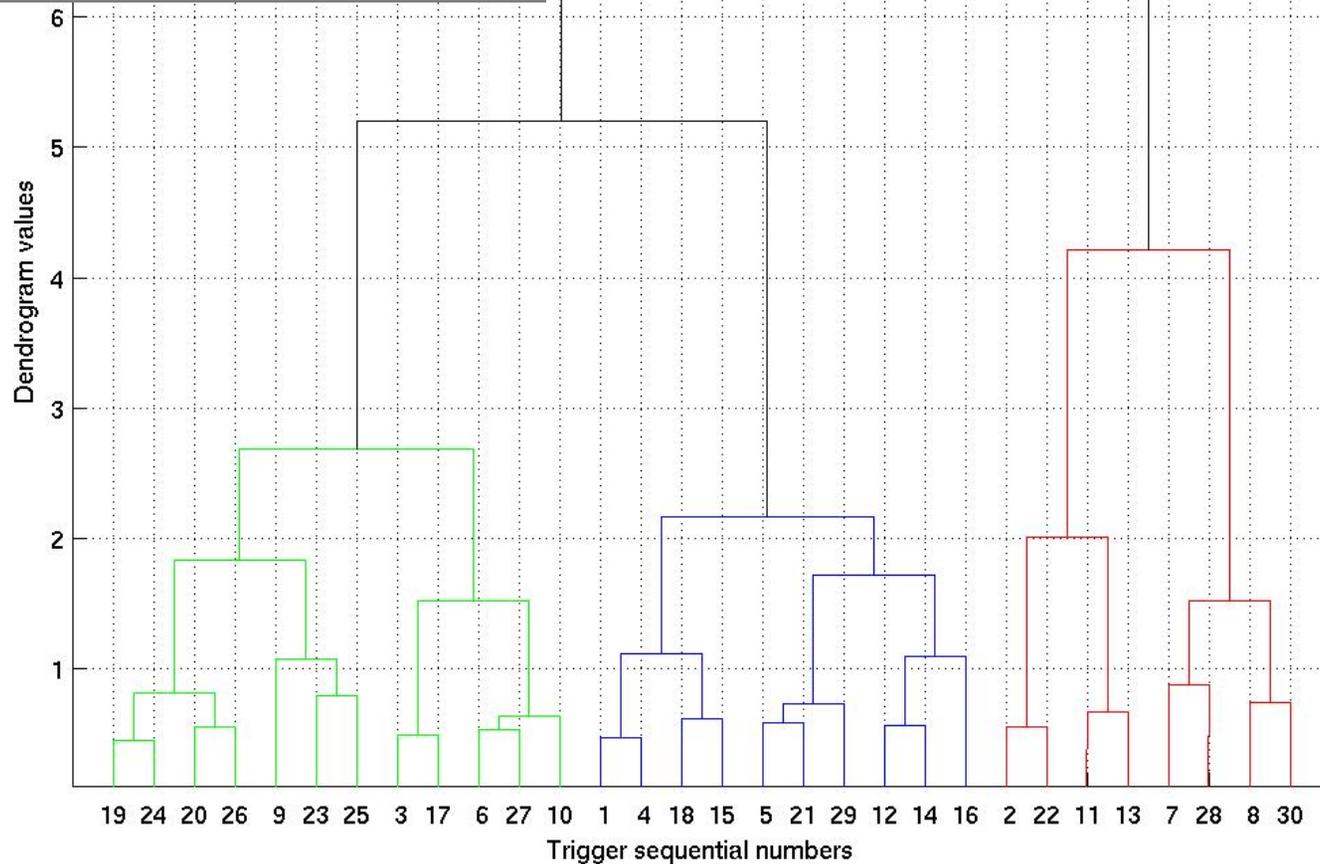






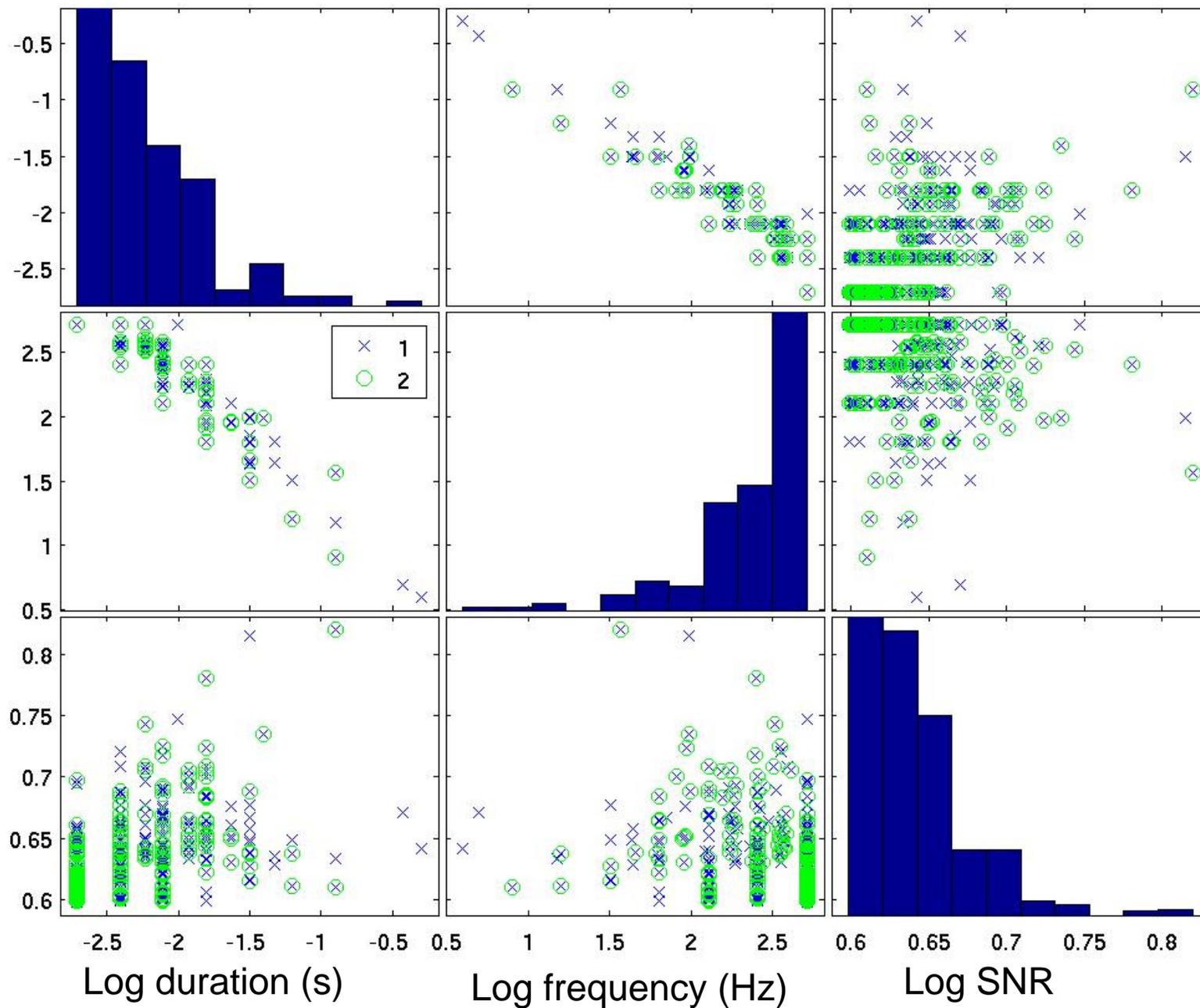
S5 kleine Welle triggers; Dendrogram

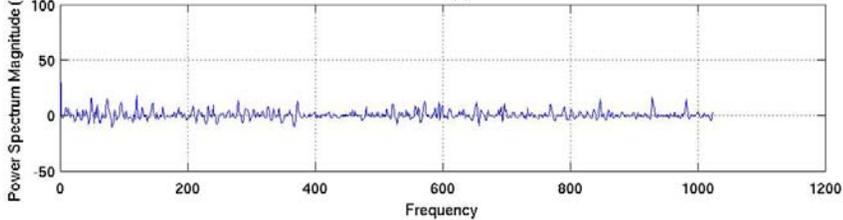
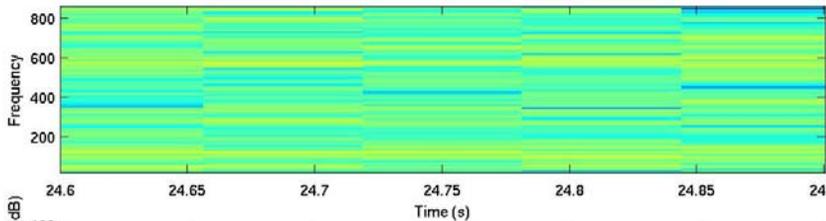
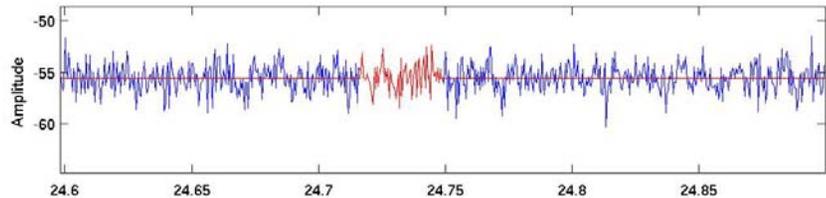
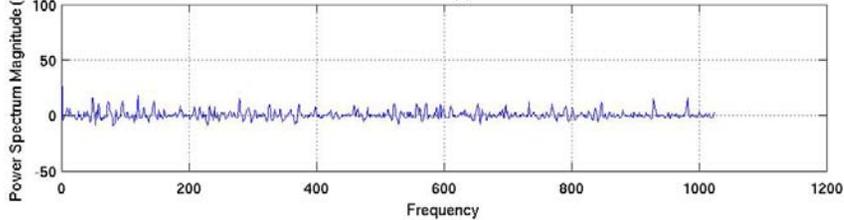
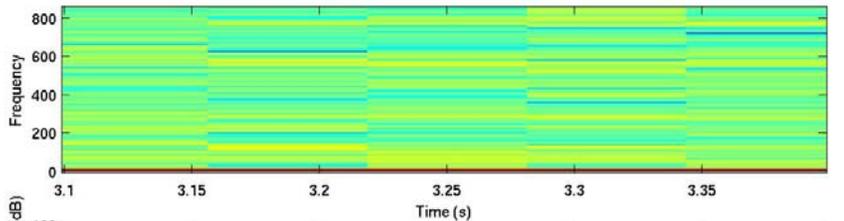
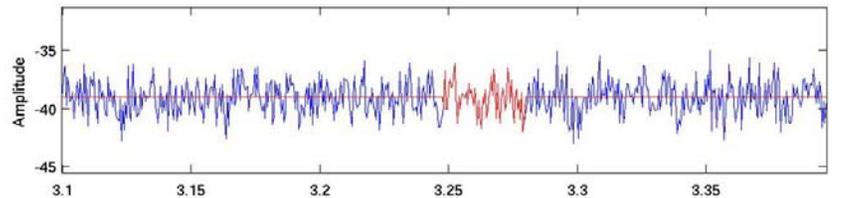
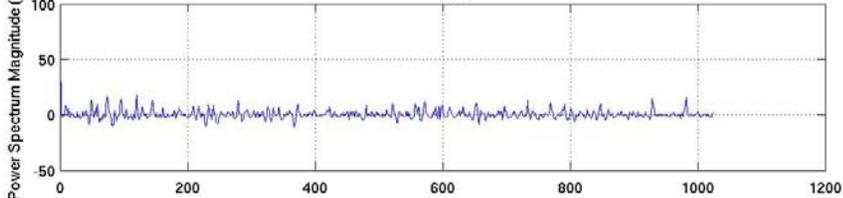
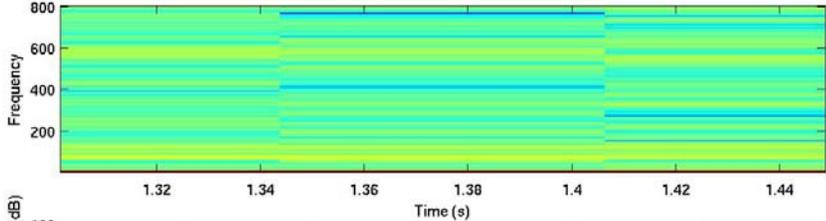
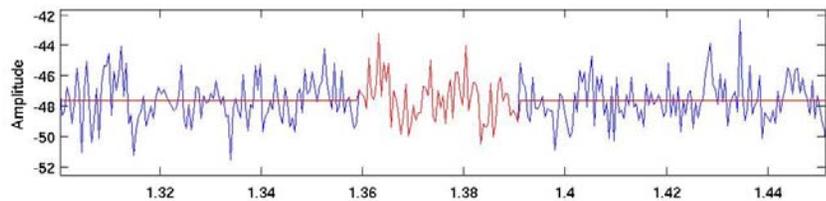
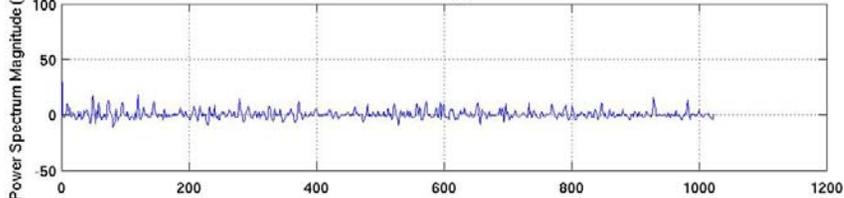
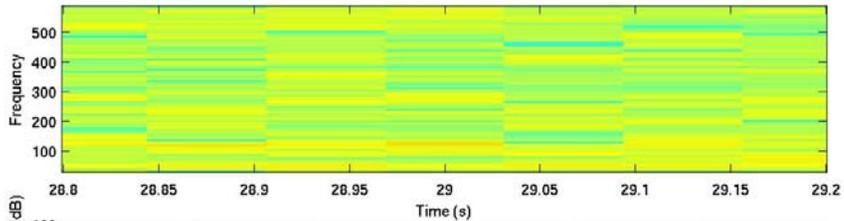
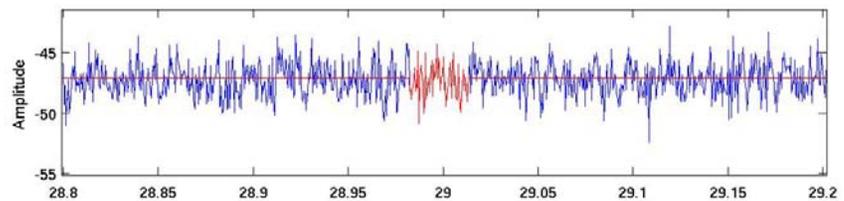
H0:PEM-BSC10_ACC1Y
11/10/06



$r = 0.67$ p - values = 0.58 Number of significant clusters = 1

S5 kleineWelle triggers; Group plot and correlations between duration, central frequency & SNR





Post classification

- Construct class *basis* from initial knowledge of the class properties (especially from studying the auxiliary channel classes).
- Construct training algorithm that classifies each trigger into a given class.
- Validate by comparison of physical properties within bounds of our instrumental knowledge.
- To look for more efficient ways of time series classification, once the pipeline starts full-mode operation.