Multidimensional classification of burst triggers from LIGO S5 run

Soma Mukherjee for the LSC
Center for Gravitational Wave Astronomy
University of Texas at Brownsville, USA.
Amaldi Conference, Sydney, July 8-14, 2007

LIGO-G070434-00-0

A quick walk through …

•This study involves implementation of classification methods (non-parametric/hierarchical and parametric, k-means) to see presence of structures in higher dimensional parameter space. Often features embedded in higher dimensions are not elucidated in simple 1 or 2 dimensional study.

•References for LIGO classification analysis: S. Mukherjee : Past LSC, F2F and GWDAW talks, Burst and Glitch group telecons, published paper in CQG).

•kleine Welle *(Blackburn, L. et. al. 2005 LIGO-050158-00-Z)* 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.

Pipeline : What is new since March 2007 LSC meeting

a. We start with the kleineWelle trigger database from all channels.

b. Data conditioning is applied to clean the narrowband features and filter in the appropriate frequency bands.

c. Analysis database is constructed with **logarithmically transformed** Δt, frequency and snr and shape data. **Shape information of the triggers taken into account from the processed time series**.

d. A hierarchical classification algorithm and **a time series similarity-based classification** is applied to the reconstructed database.

e. At the end of the pipeline, we have information on existing classes (statistics, members,properties) from all channels.

- **We directly access raw data uninterruptedly**. This involves :

a. Fast connection for rapid data transfer.
b. Data storage.
c. Storage for pipeline output.

a. Identification of the class properties.

b. Correlation between channels.

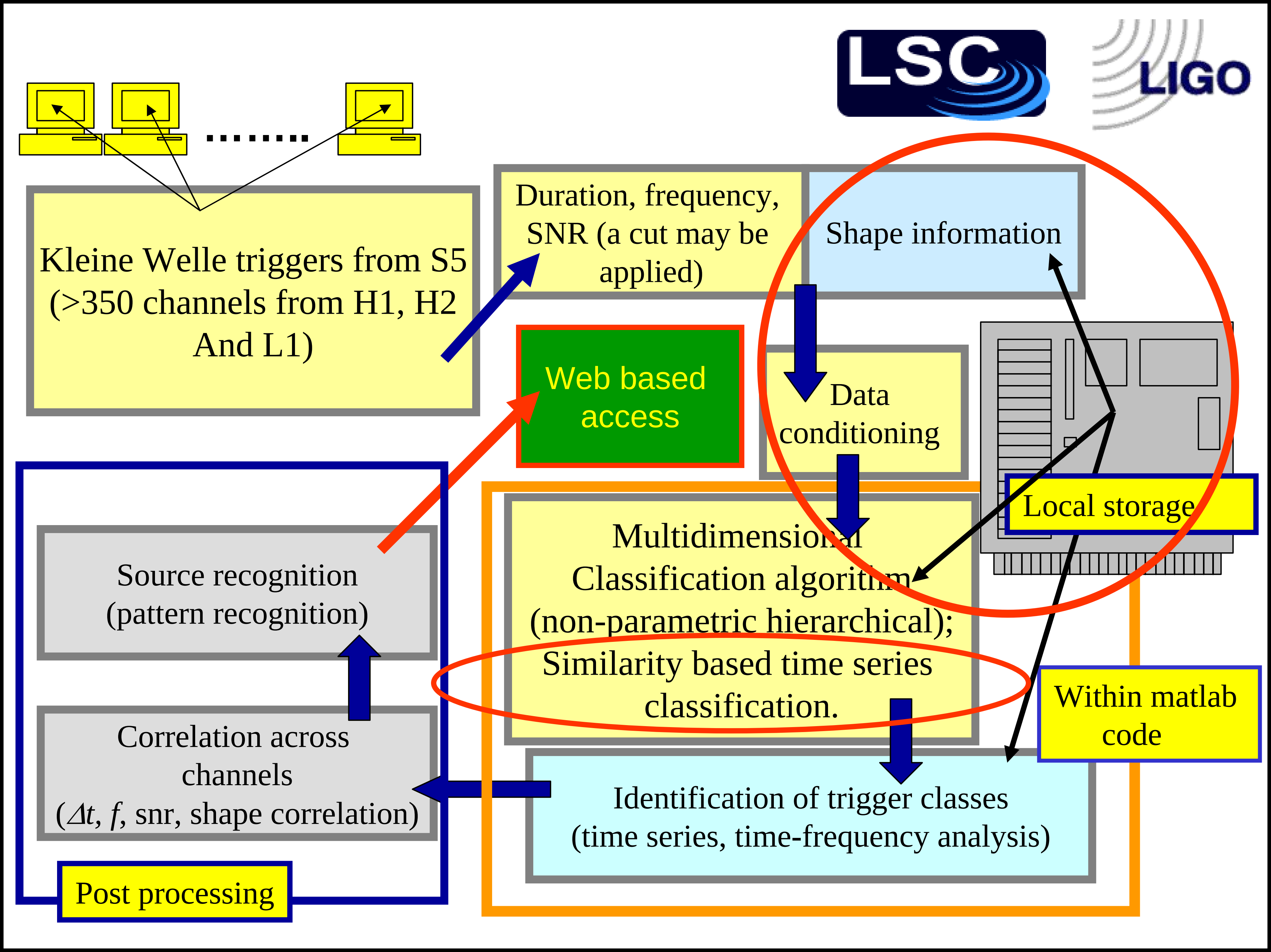
c. Formation of class basis.

d. Final step towards direct classification based on pattern recognition.

Veto applications.

LSC

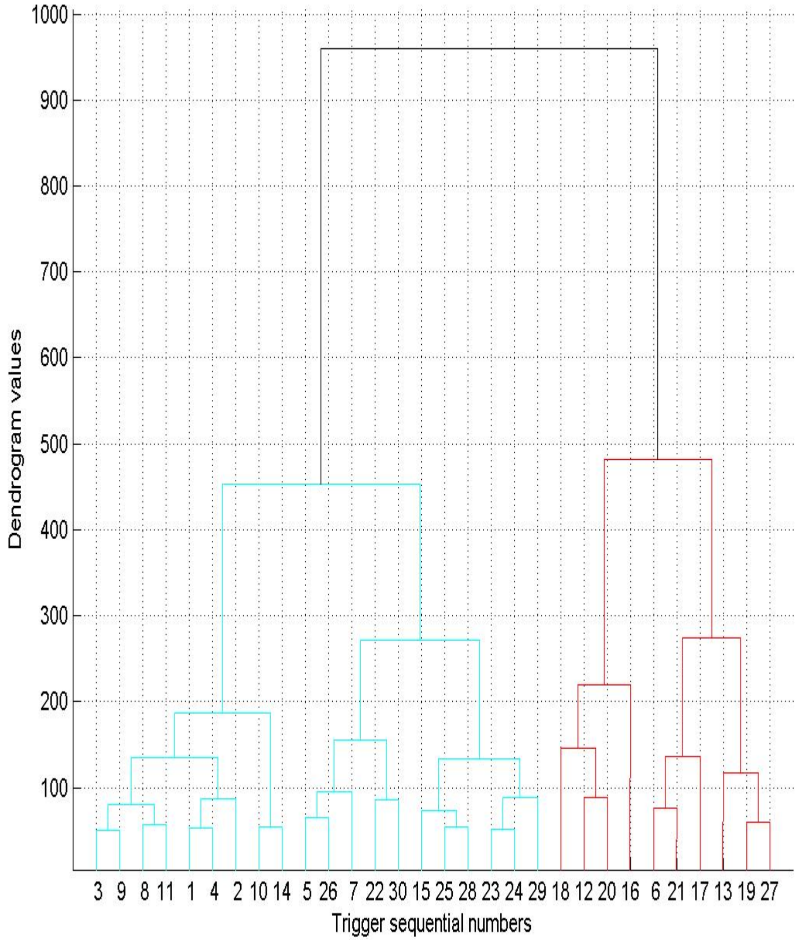
LIGO


LSC
LIGO
Kleine Welle triggers from S5 (>350 channels from H1, H2 And L1)
Duration, frequency, SNR (a cut may be applied)
Shape information
Web based access
Data conditioning
Local storage
Multidimensional Classification algorithm (non-parametric hierarchical); Similarity based time series classification.
Within matlab code
Source recognition (pattern recognition)
Correlation across channels (Δt, f, snr, shape correlation)
Identification of trigger classes (time series, time-frequency analysis)
Post processing

Algorithm : Hierarchical classification

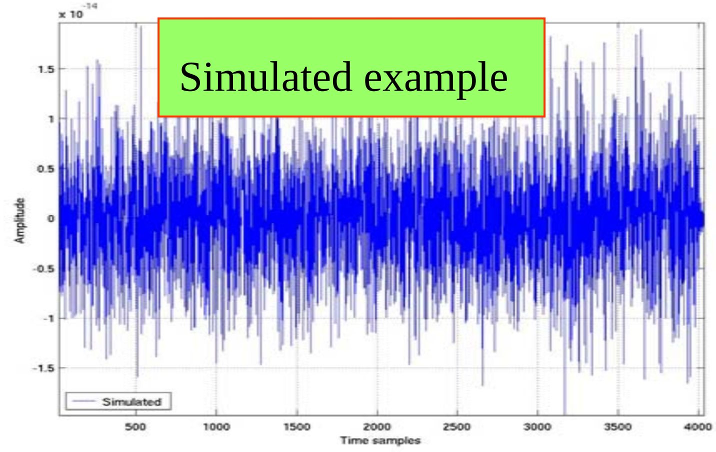
The algorithm is based on computation of distances between data points in the multi-dimensional space. A variance minimization criterion is used to group the objects into statistically distinct classes. The metric may be chosen in several different ways. The group formation stage is guided by *complete-linkage* criterion, i.e. largest distance between objects in two groups.

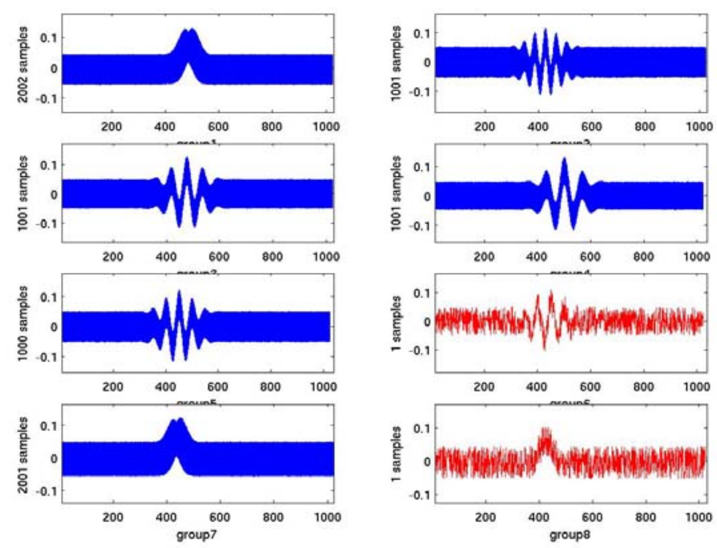
Example

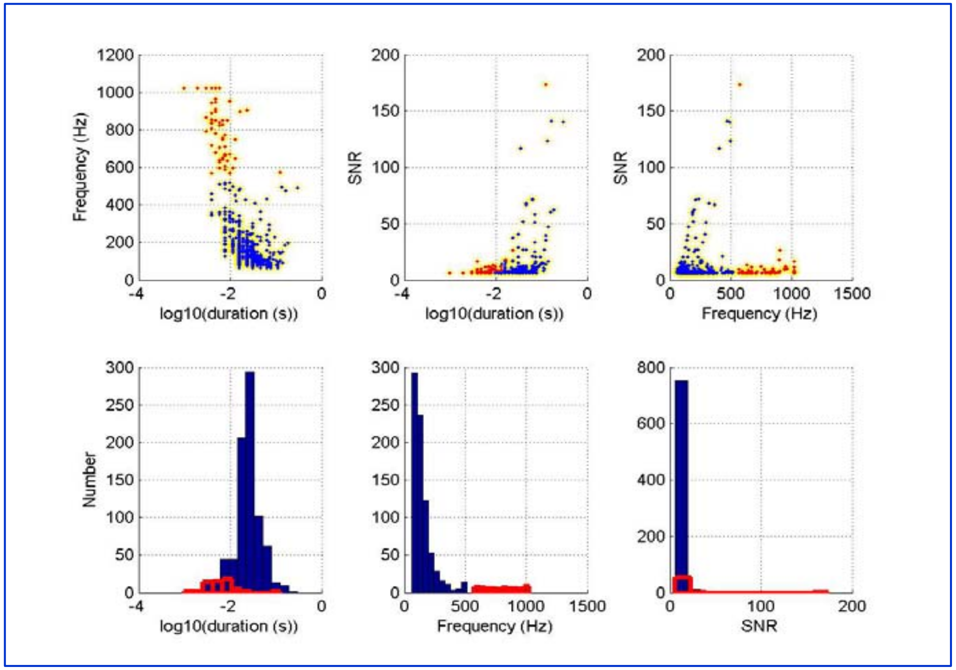


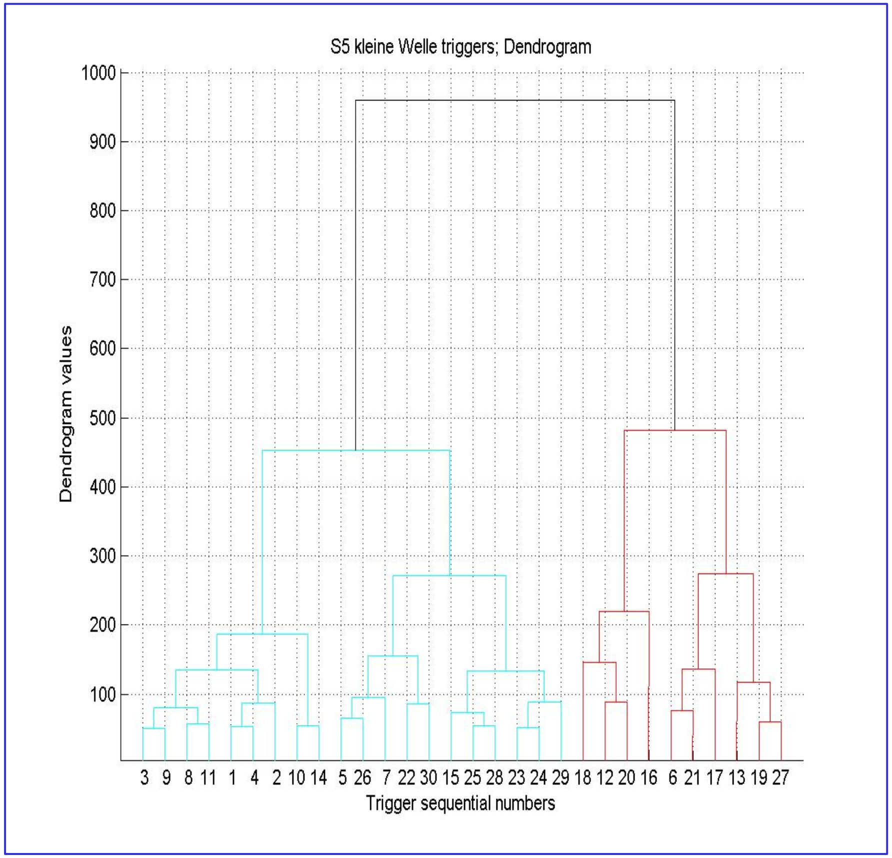
Algorithm : Similarity driven time series classification

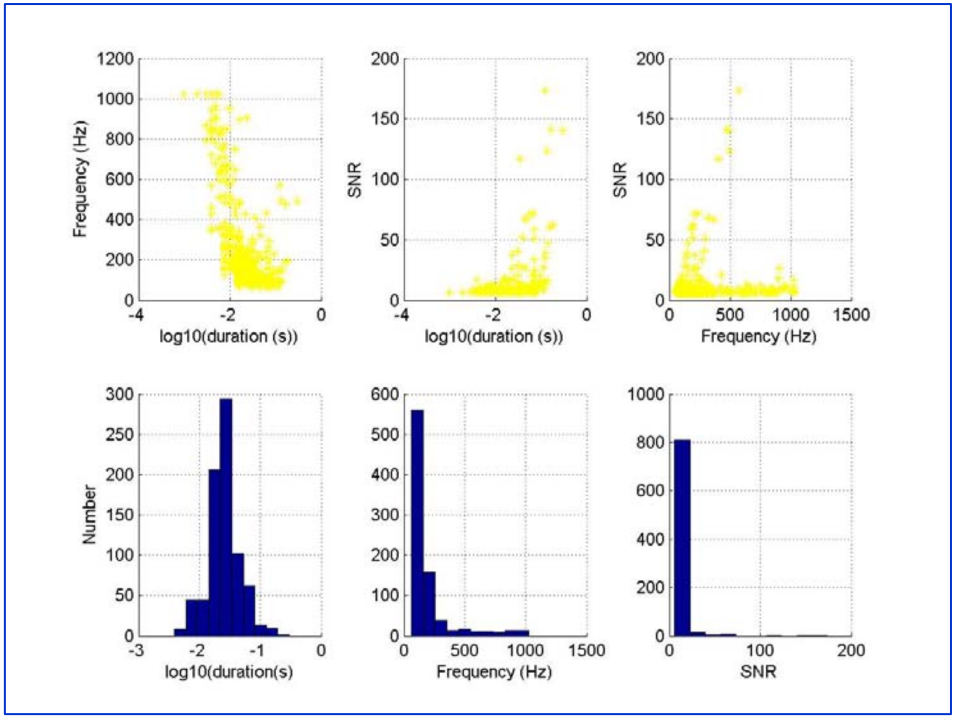
This algorithm is based on *k-means* which classifies data by assignment of *k* centroids chosen *a priori* and then partitioning data based on association of data points to the nearest centroid. In the final step, the algorithm minimizes an objective function, which in this case is a squared error function. The method is an iterative one where the centroids are re-calculated until stability is reached.





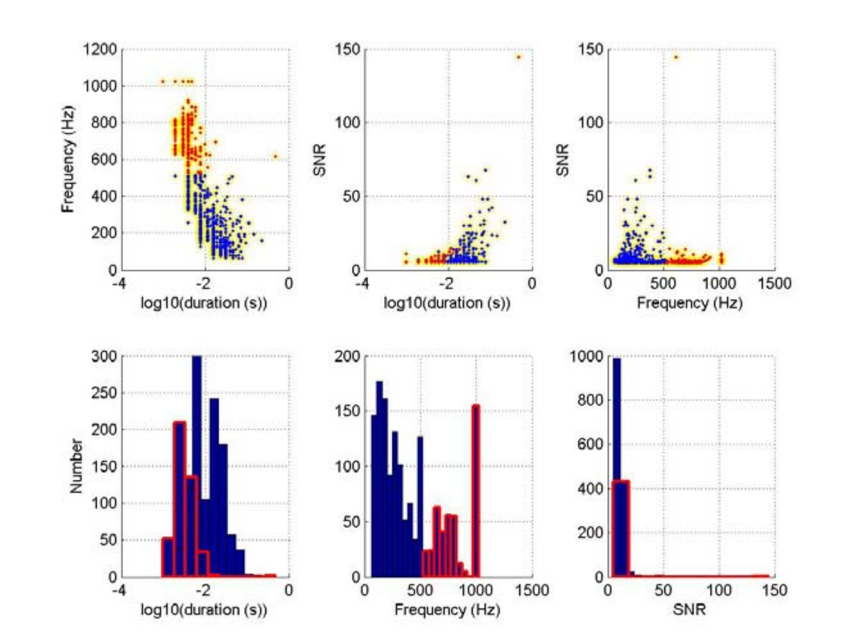


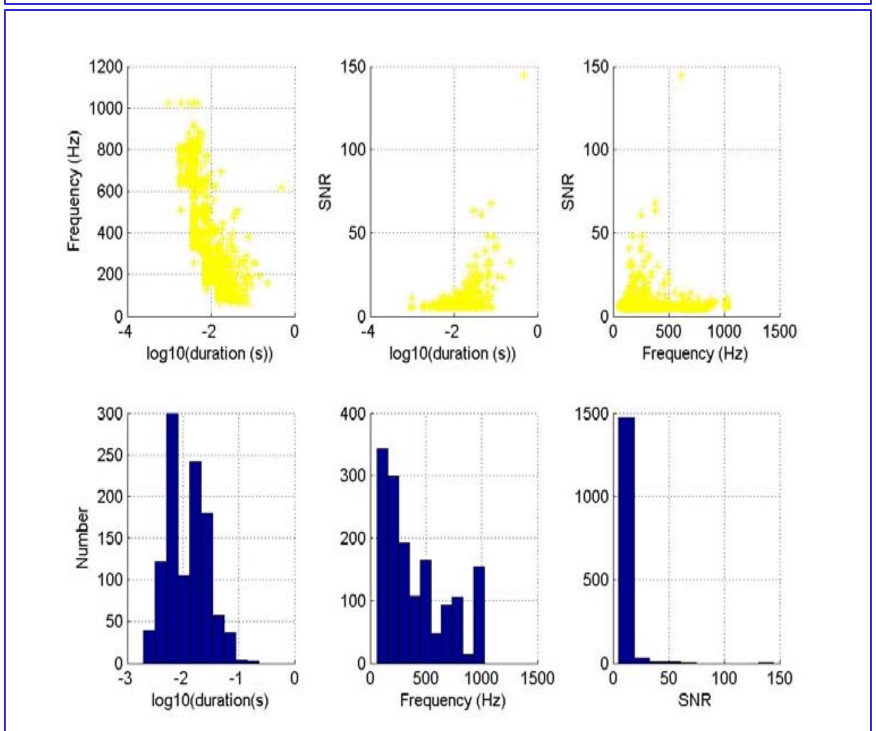


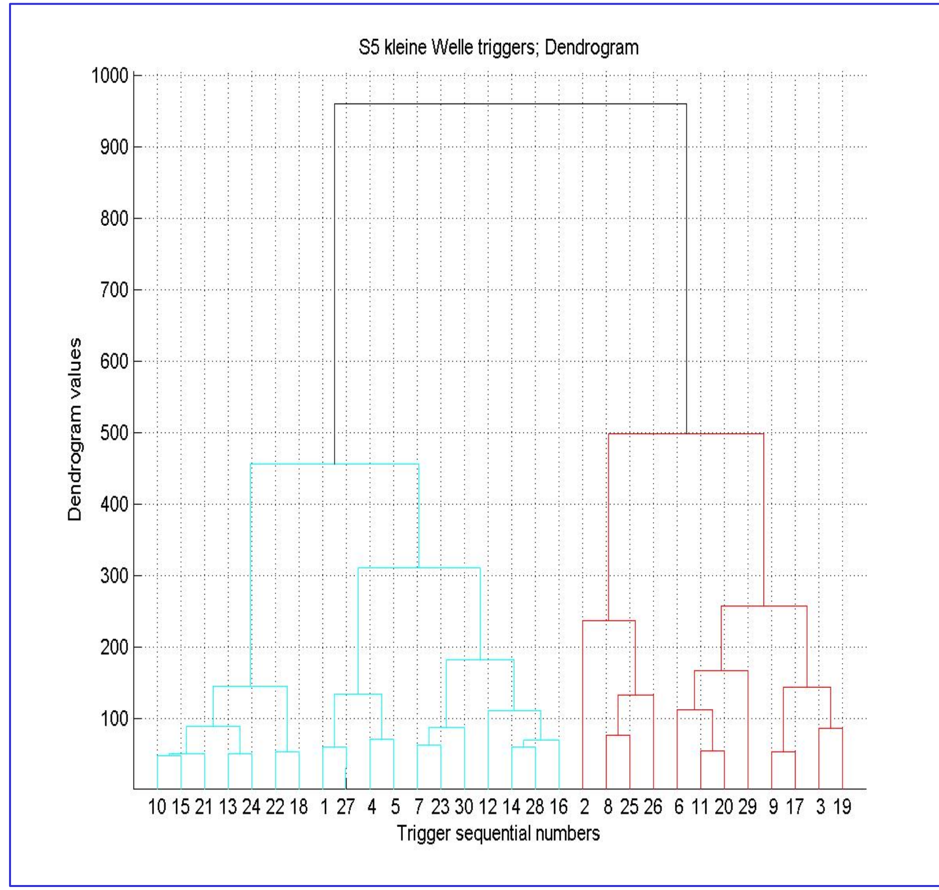


H1:LSC-DARM_ERR

Groups=2, $p<0.003$, $r=0.93$

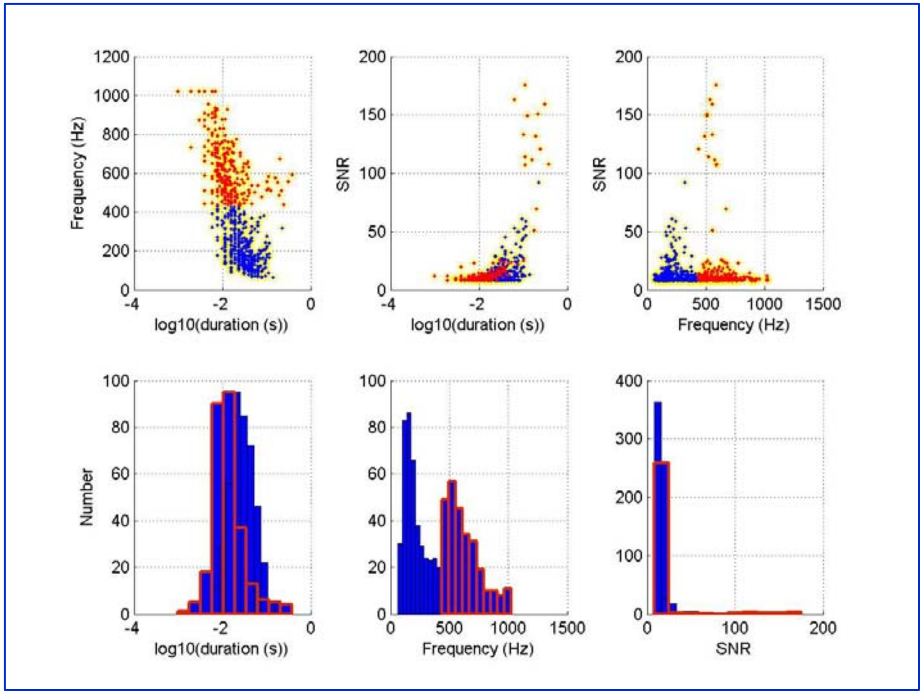


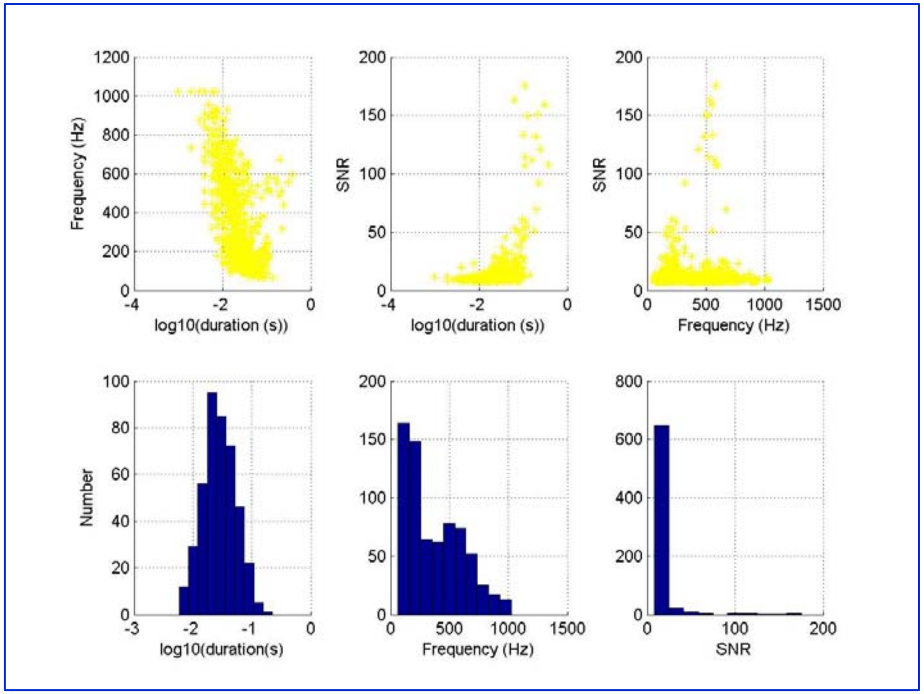


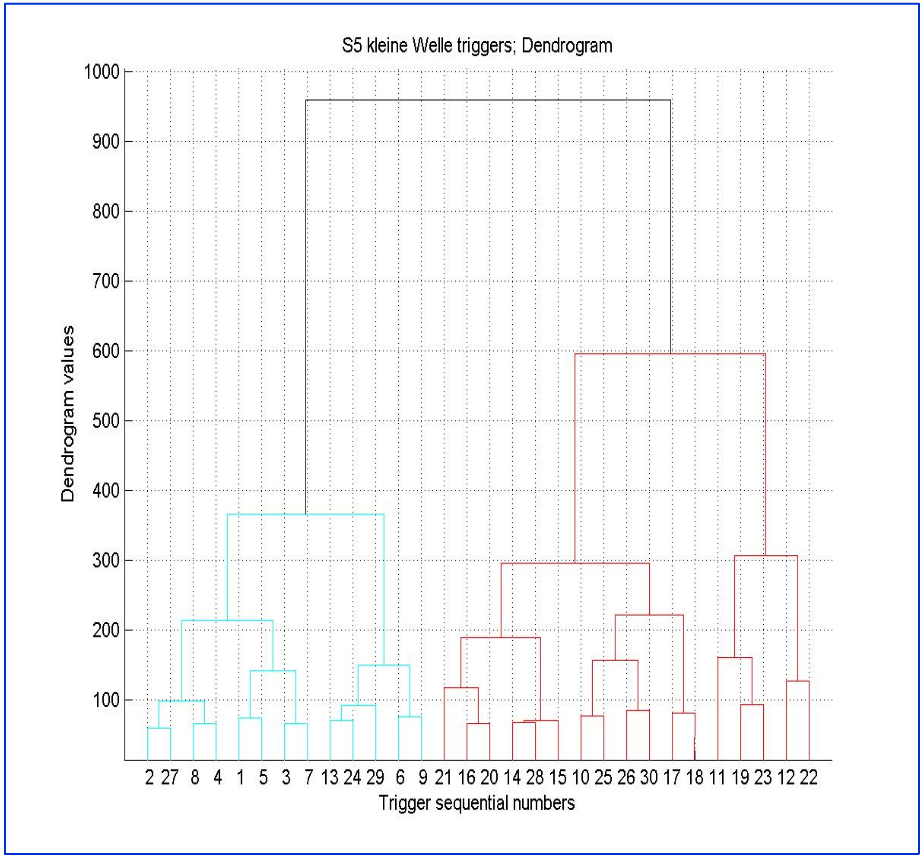


H2:LSC-DARM_ERR

Group=2, $p<0.004$, $r=0.81$

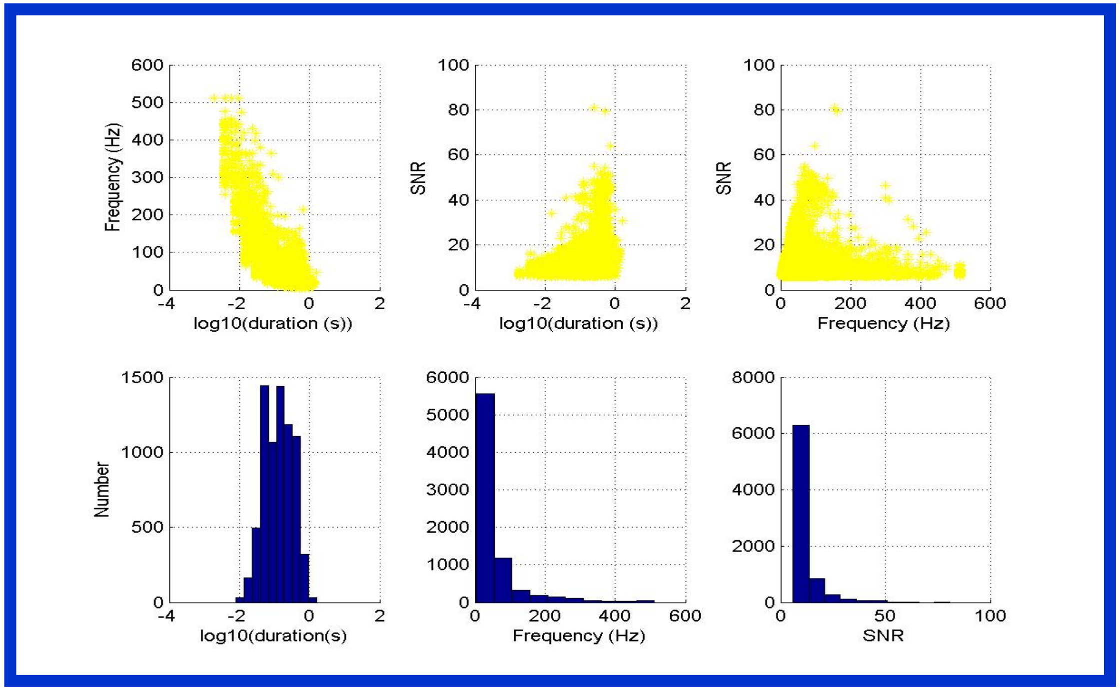


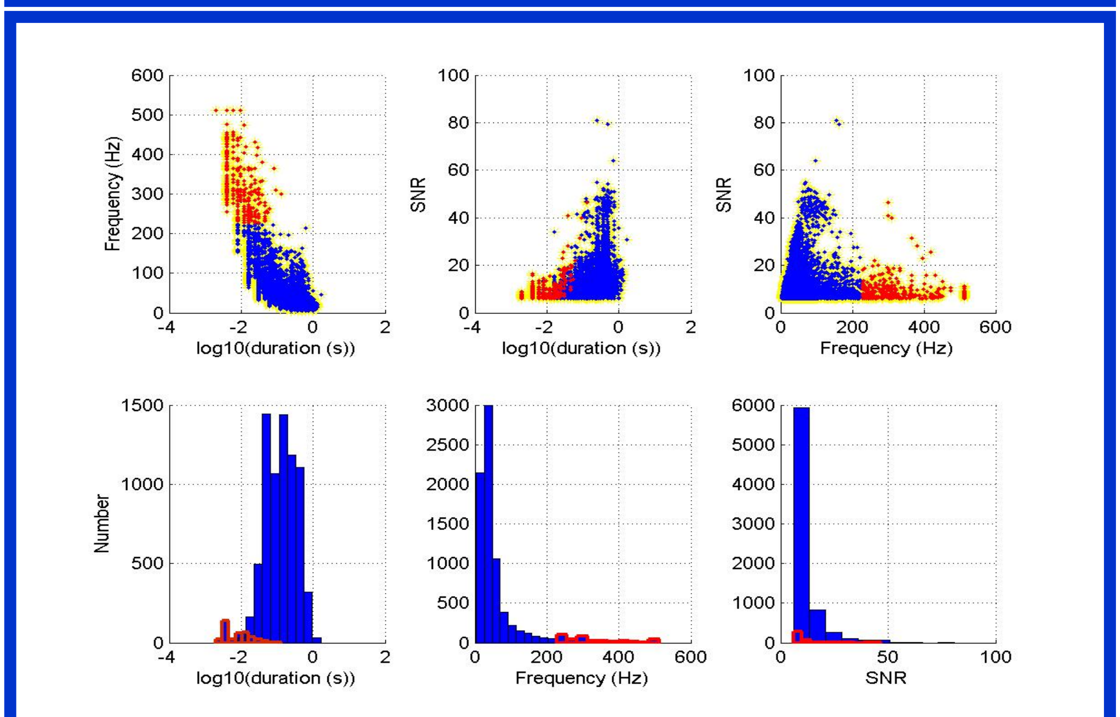




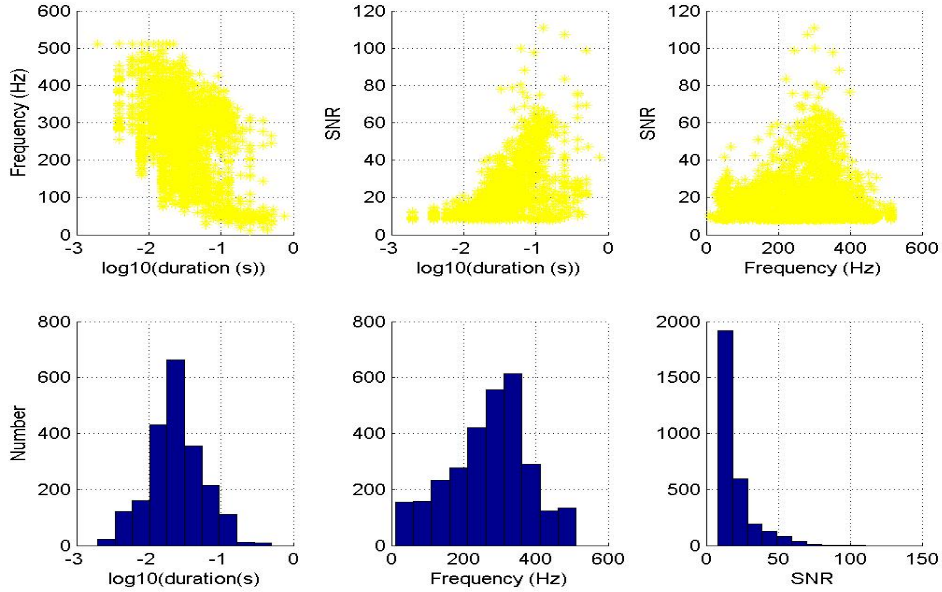
Groups=2, p<0.04, r=0.80

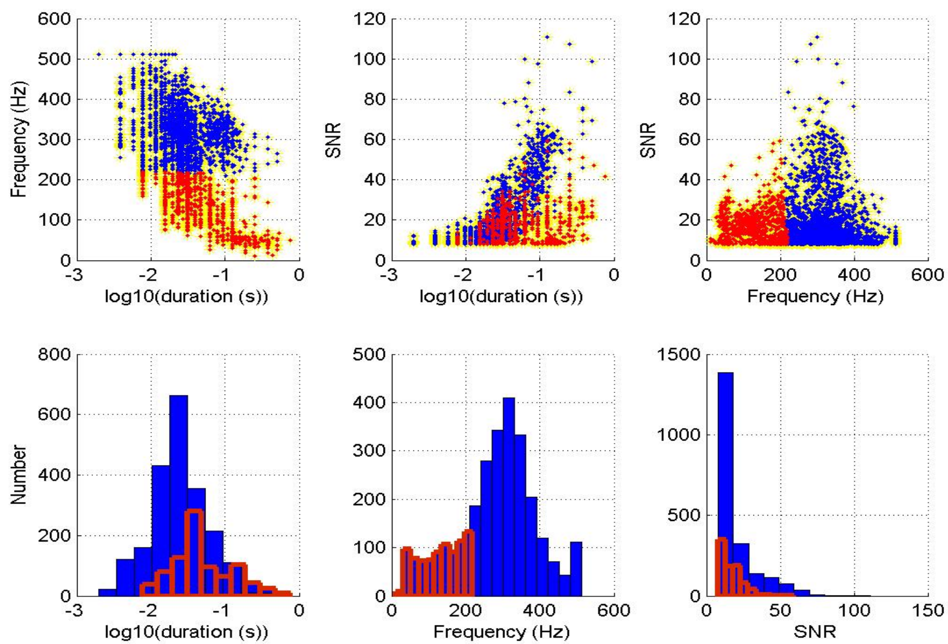
L1:LSC-DARM_ERR



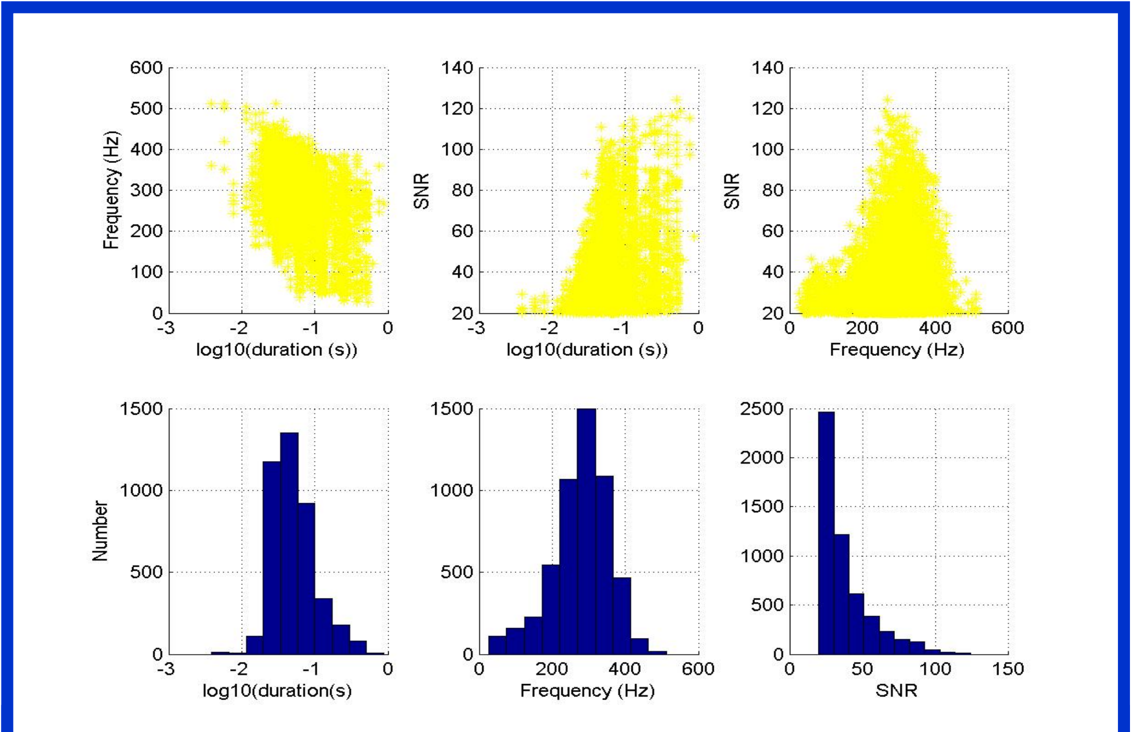


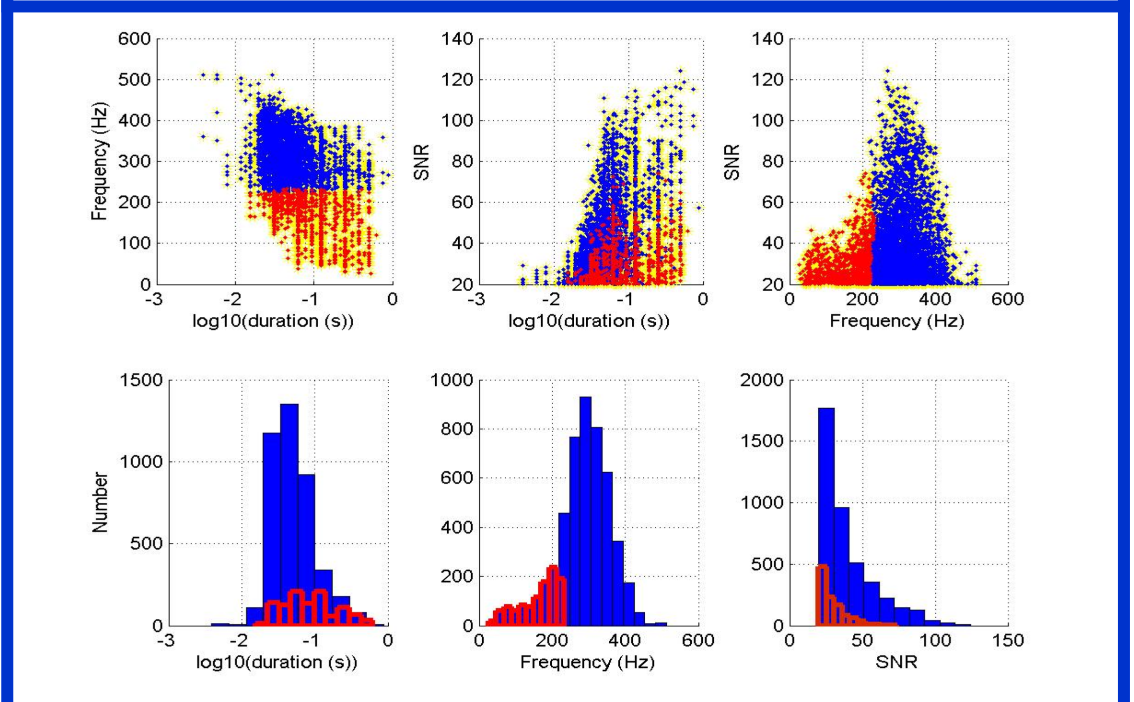
Detector : H0
Channel : BSC6_MAGZ
Number of groups =2
$p < 10e-6$
r=0.92
n=26304
snr cut = 6
n-surviving = 7669





Detector : H1
Channel : COIL_MAGZ
Number of groups =2
p < 10e-8
r=0.70
n=31086
snr cut = 8
n-surviving = 2963





Detector : H1

Channel : COIL_MAGY

Number of groups =2

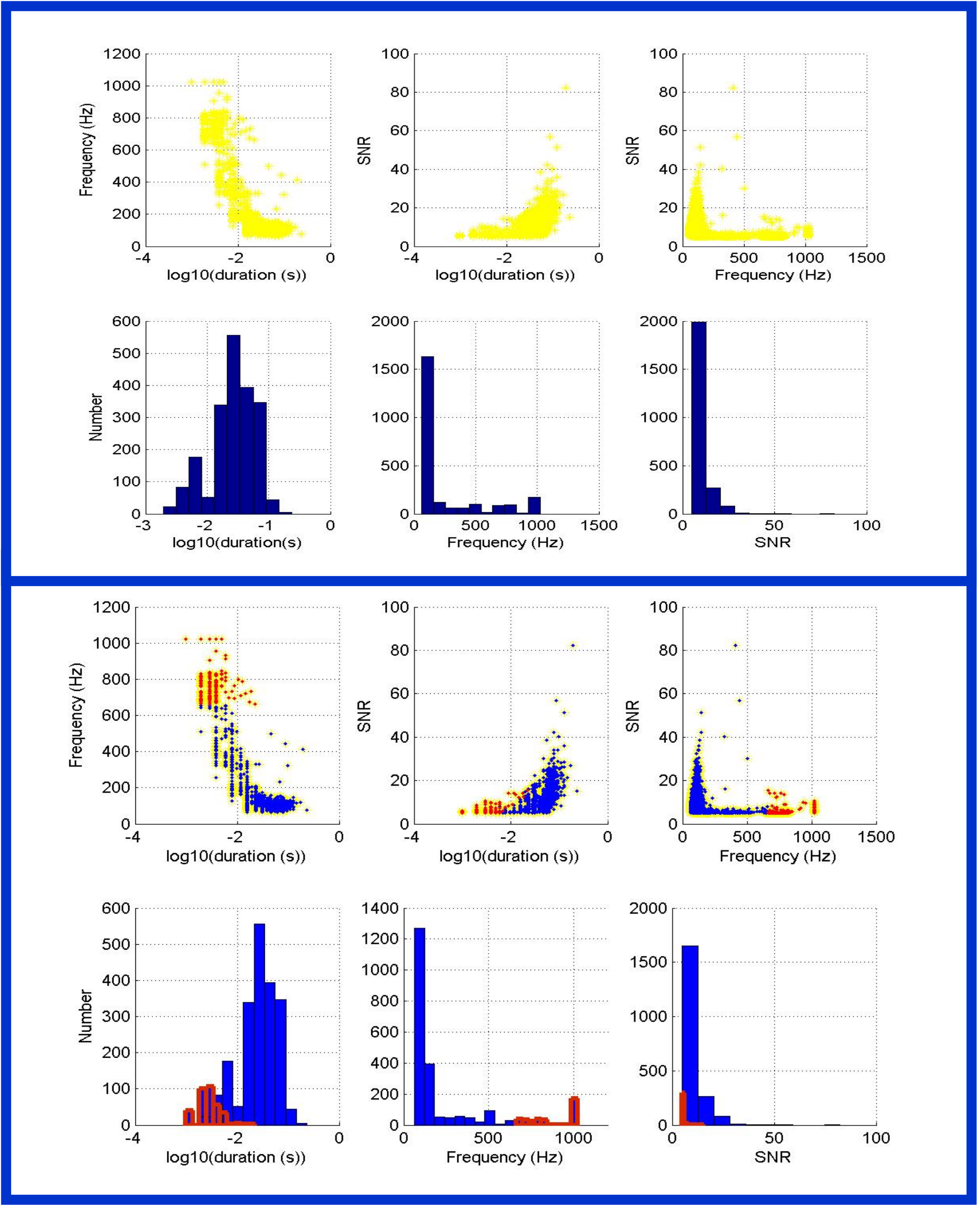
$p < 10e-19$

r=0.70

n=64220

snr cut = 20

n-surviving = 5270



Detector : H1

Channel : MICH_CTRL

Number of groups =2

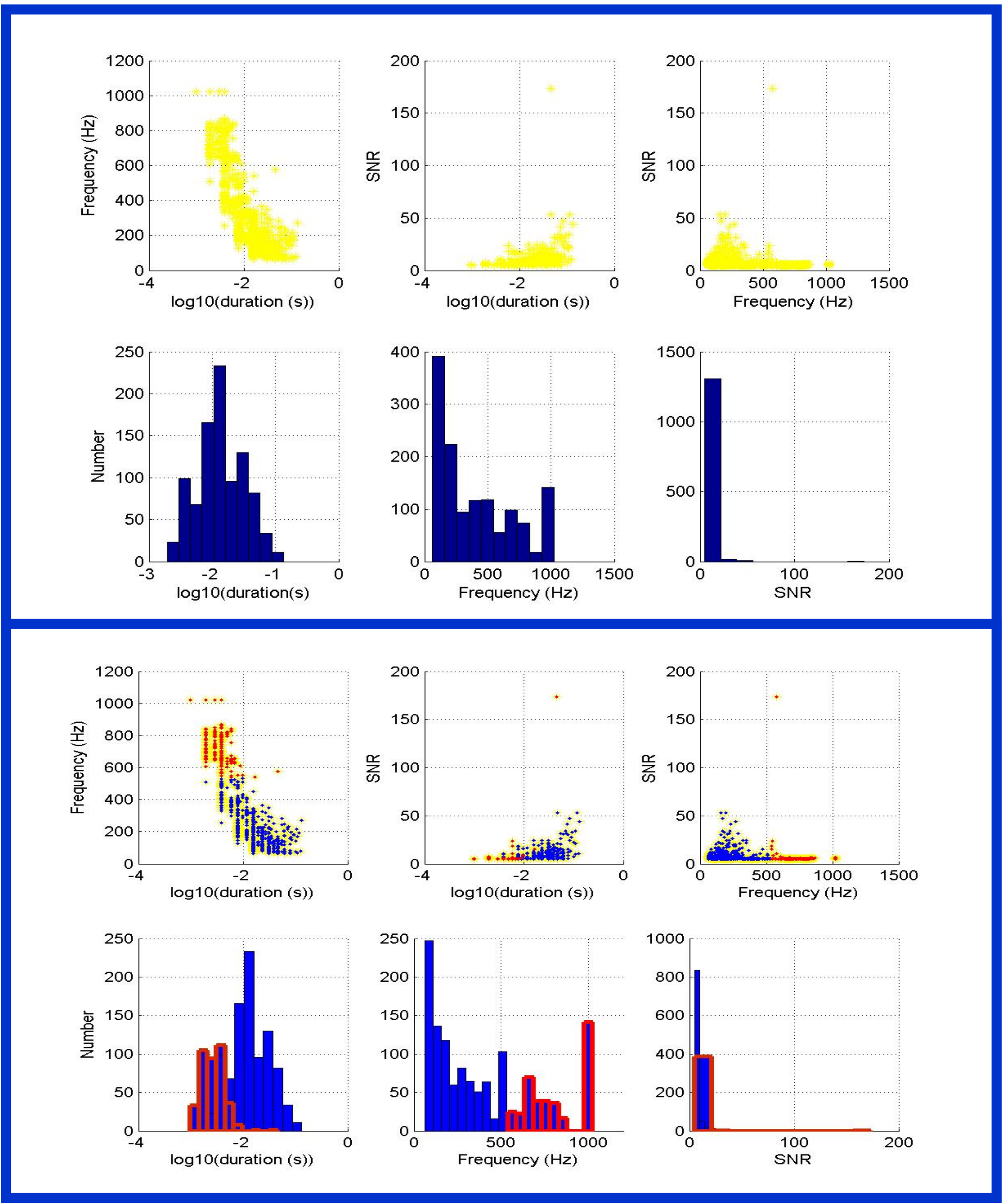
p < 10e-19

r=0.94

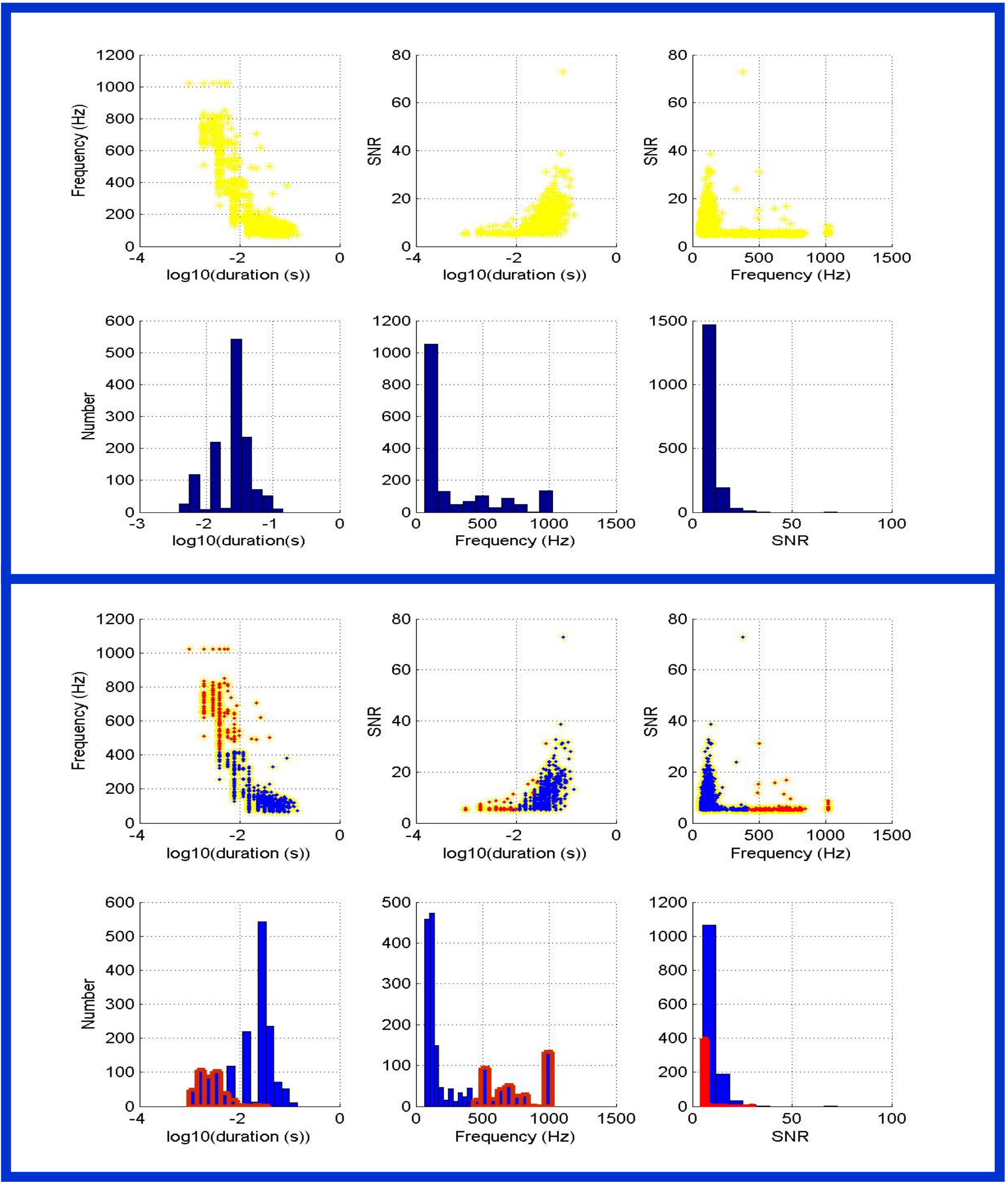
n=17068

snr cut = 5

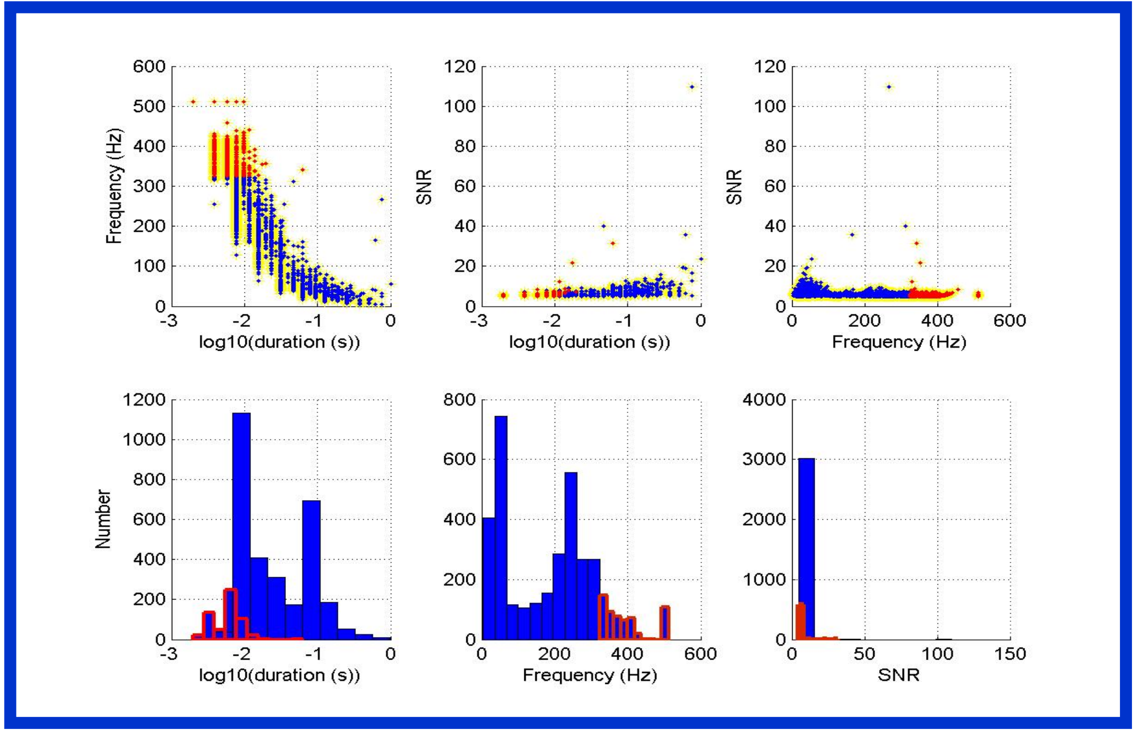
n-surviving = 2364

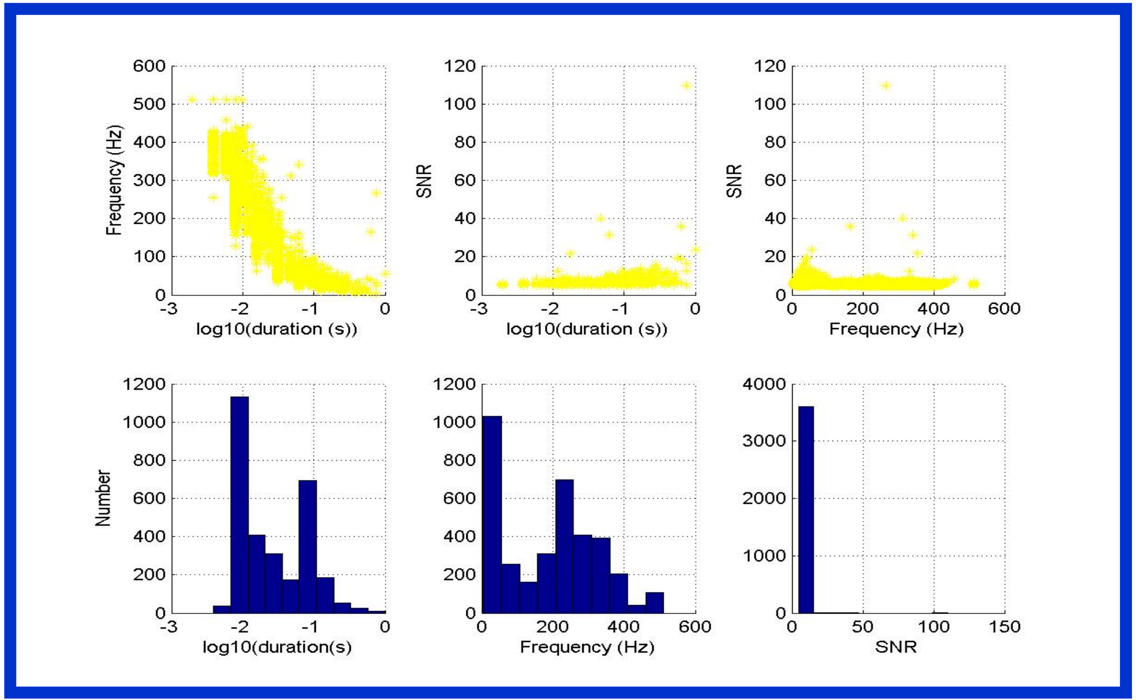


Detector : H2
Channel : POB_I
Number of groups =2
$p < 10e\text{-}15$
$r=0.82$
$n=17096$
snr cut = 5
n-surviving = 1331

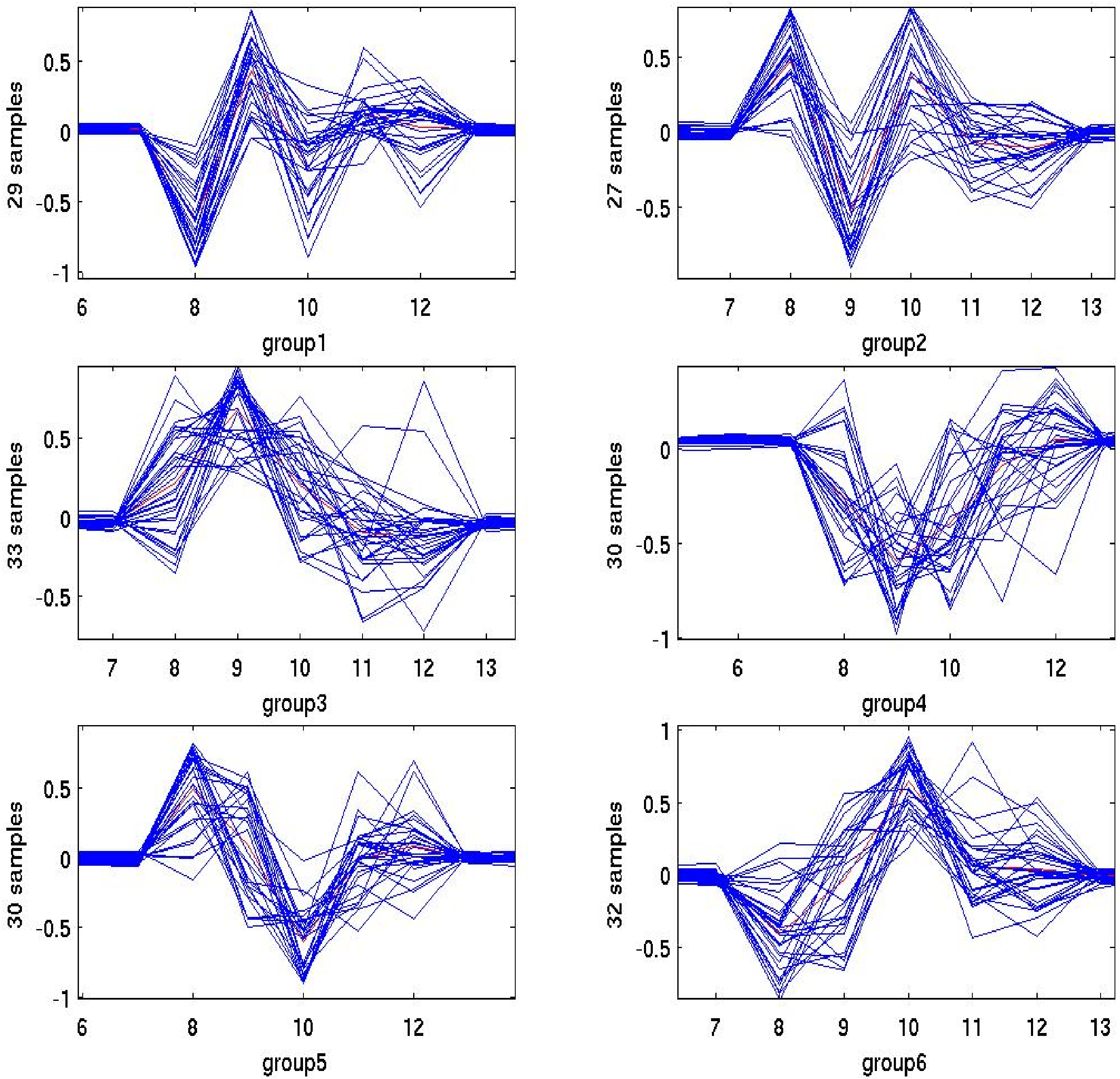


Detector : H1
Channel : POB_I
Number of groups =2
p < 10e-8
r=0.88
n=17779
snr cut = 5
n-surviving = 1706





Detector : L1
Channel : ITMX_Y
Number of groups =2
$p < 10e\text{-}9$
r=0.74
n=16508
snr cut = 5
n-surviving = 3609

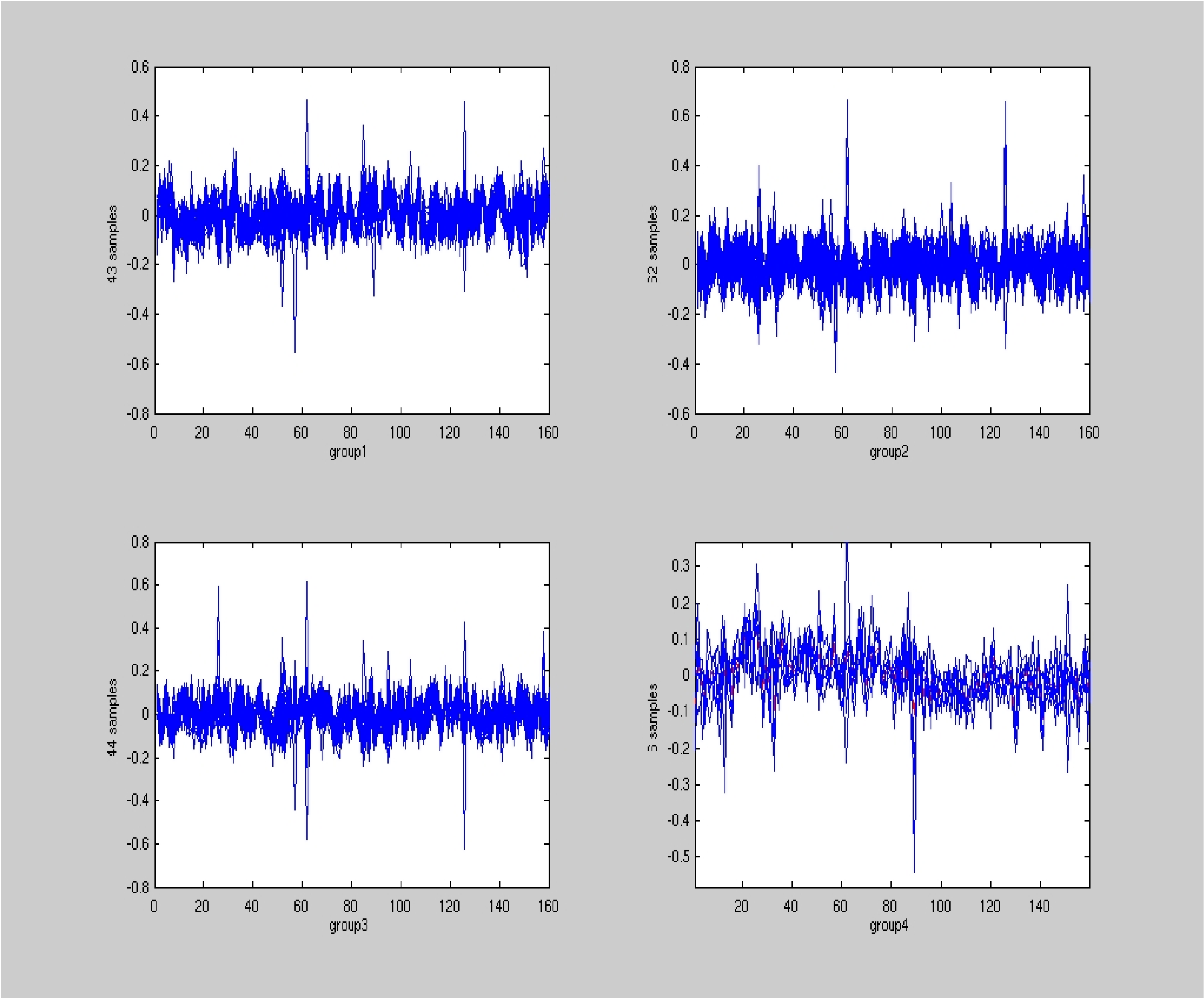


L1:LSC-DARM_ERR

**GPS : 862588800-
862675200 s.**

6 distinct groups found by Similarity driven classification algorithm. These classes are based on the shape of the triggers obtained by retaining 128 Hz around the central frequency and band passing.
SNR cut= 30

L1:PEM-LVEA_MAGZ



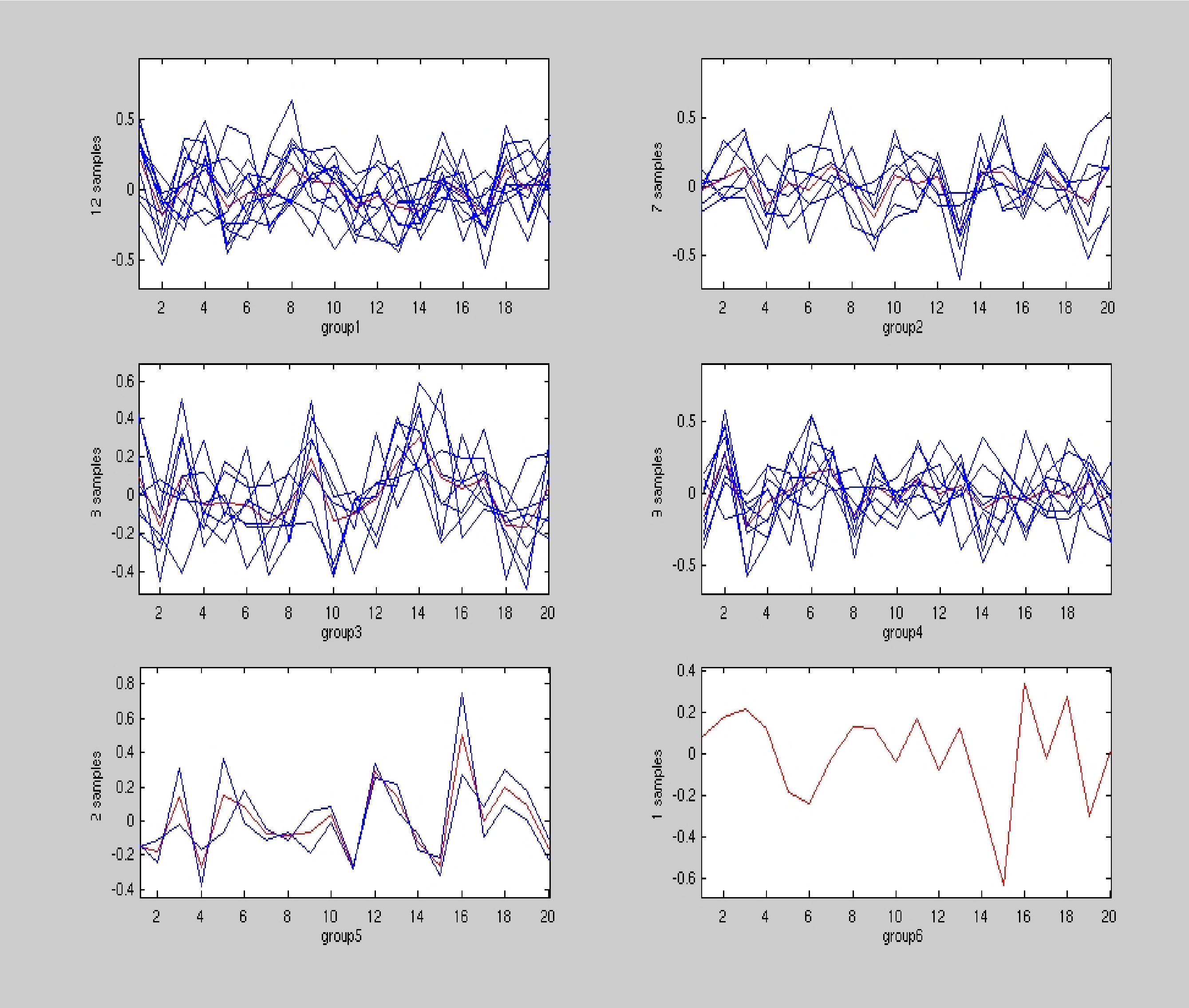
GPS : 862588800-862675200 s.

4 distinct groups found by Similarity driven classification algorithm.

These classes are based on the shape of the triggers obtained by retaining 128 Hz around the central frequency and band passing.

SNR cut= 15

L0:PEM-EY_MAGX



GPS : 862588800-862675200 s.

6 distinct groups found by Similarity driven classification algorithm.

These classes are based on the shape of the triggers obtained by retaining 32 Hz around the central frequency and band passing.

SNR cut= 5

Conclusions

Hierarchical classification algorithm shows existence of more than one statistically significant classes in the kleine-Welle trigger database.

Development of time series classification algorithms and incorporation of trigger shape into the analysis shows more structures present in the multi-dimensional space.

Knowledge from both the analyses being used for pattern recognition across all channels. *(In progress)*

Target & timeline : what's been met and what next

- •Main hierarchical classification code in Matlab.
- •Hardware/software/bandwidth for uninterrupted data transfer.
- •Data storage.
- •Code modification to access data for shape information.
- •Development of similarity based time series shape classification.
- •End-to-end test.

- •Extend study to coincident triggers.
- •Archiving of results in a password protected web page accessible to the collaboration. [October 2007 LSC].
- •BNS trigger classification.

- •Trigger identification.
- •Automation.
- •Veto. (November/December 2007)

Acknowledgment

Similarity based clustering algorithm is developed by H.Lei, L. R. Tang, S. Mukherjee, S. D. Mohanty, 2007. Use of LIGO S5 data is gratefully acknowledged.