Using Machine Learning to Search for Pulsars with LOFAR

Chia Min Tan
University of Manchester

Machine Learning

"A computer is said to learn from experience E
with respect to some class of task
T and performance measure P, if its
performance at tasks in T, as measured by P,
improves with experience E."

–Tom Mitchell

Supervised Machine Learning

- A training data set consisting of examples of pulsar candidate plots is provided to the machine learning algorithm.

- The data set is labelled with candidates that are pulsars and non-pulsars.

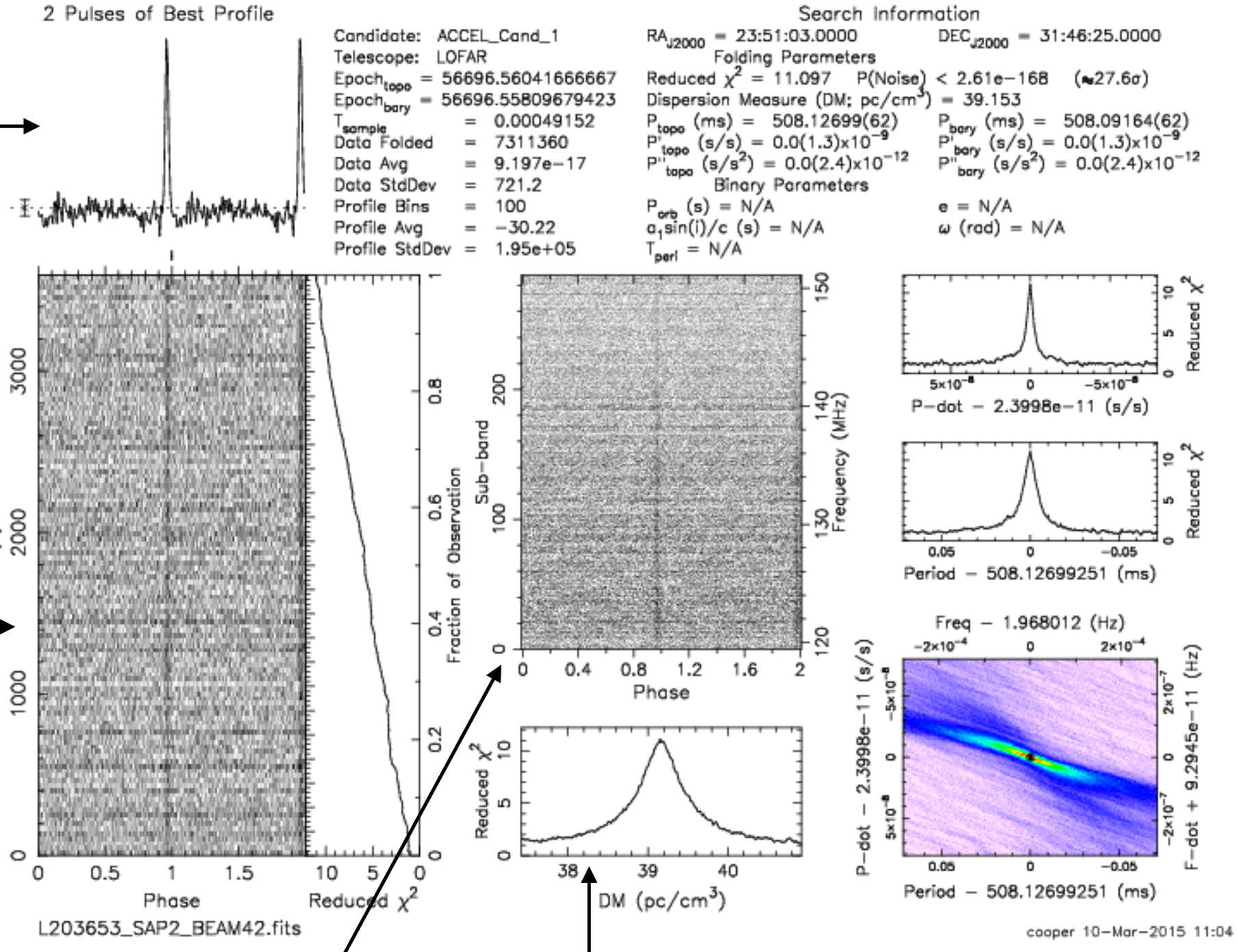
- The candidate plots are represented by a set of variables known as features.

- A classifier is produced by the algorithm to separate the data.

Pulsar Candidate Plot - The Important Bits

Pulse profile shows a peak over the noise.

Signal is consistent over time.



Signal appears across the bandwidth of the observation.

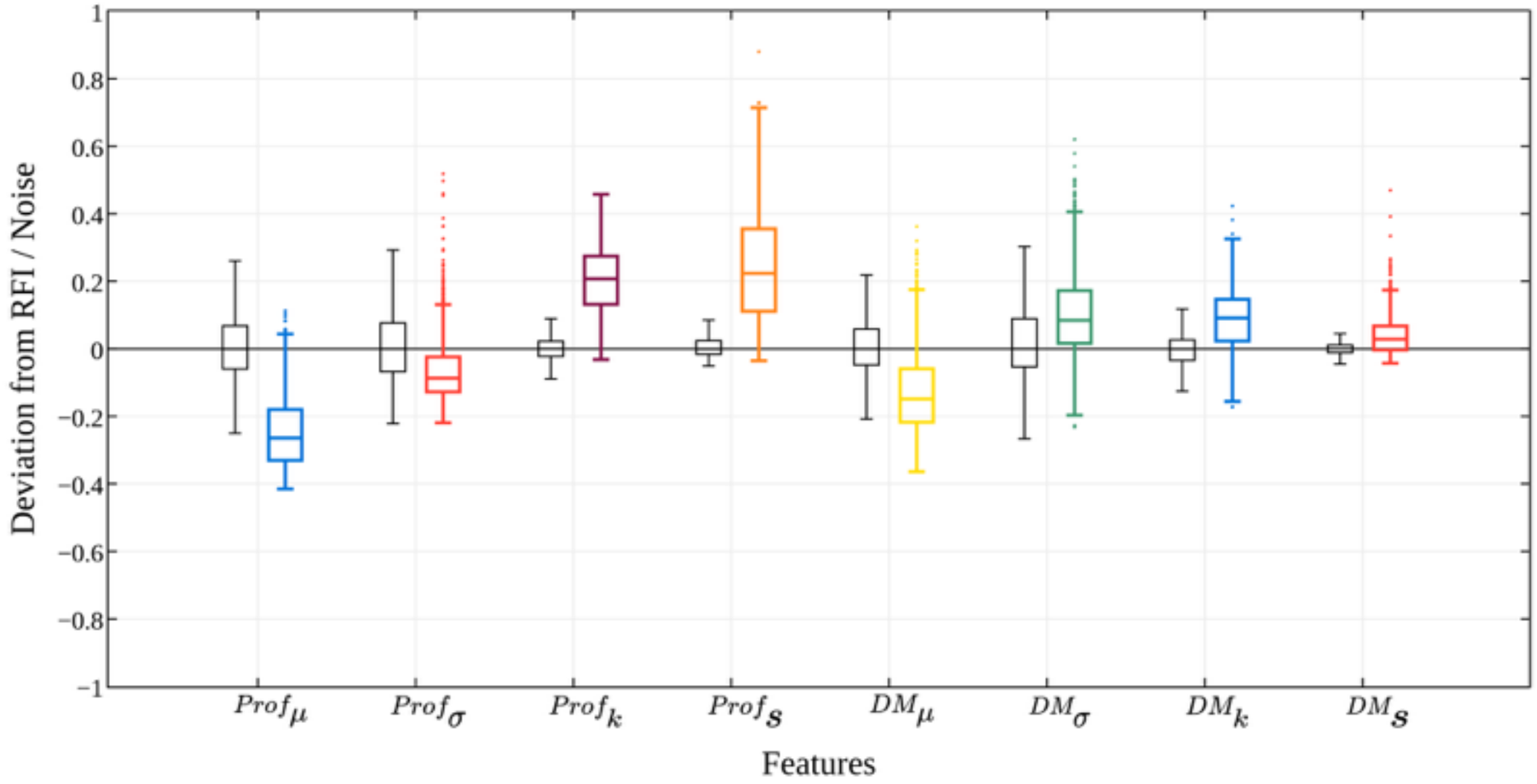
The reduced χ^2 against dispersion measure (DM) curve shows a symmetric peak at a certain DM value

Producing Features from the Candidate Plots

- Lyon et al (2016) used 8 features obtained from just the pulse profile and the reduced χ^2 against DM curve.

- Using the variable on the y-axis of each plot to calculate:
  - sample mean
  - standard deviation
  - skewness
  - excess kurtosis

Boxplots of the Features



Boxplots showing the linear separability of each features used by Lyon et al (2016) on 90000 pulsar and non-pulsar candidates obtained from the HTRU pulsar survey (Morello et al. 2014).

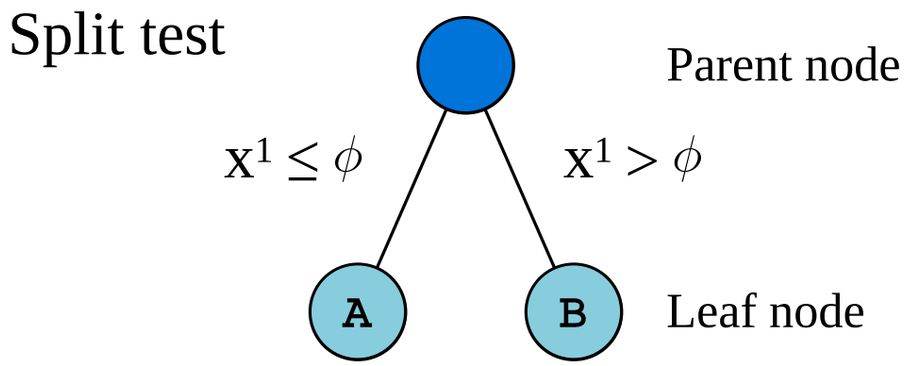
Choosing the data for LOTAAS the training set

- Lyon et al (2016) compiled a training set that consists of 66 pulsars and 4987 non-pulsar candidates.

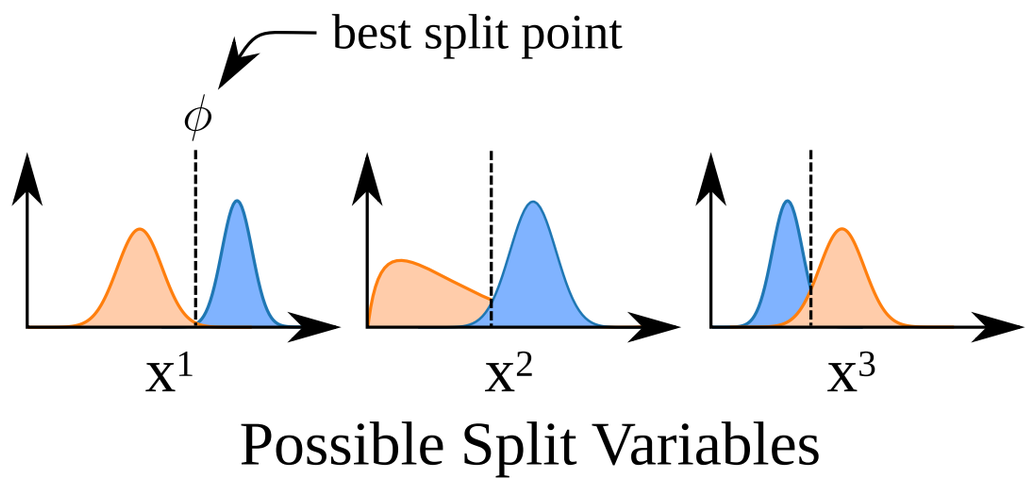
- The candidates were obtained from early survey data, where manual inspection of candidates with high signal-to-noise ratio is done.

Decision Tree Classifier

- The features obtained from the pulsar candidates are then used as input for the decision tree classifier.



- Classifier produced on the LOTAAS data shows a 99% accuracy.

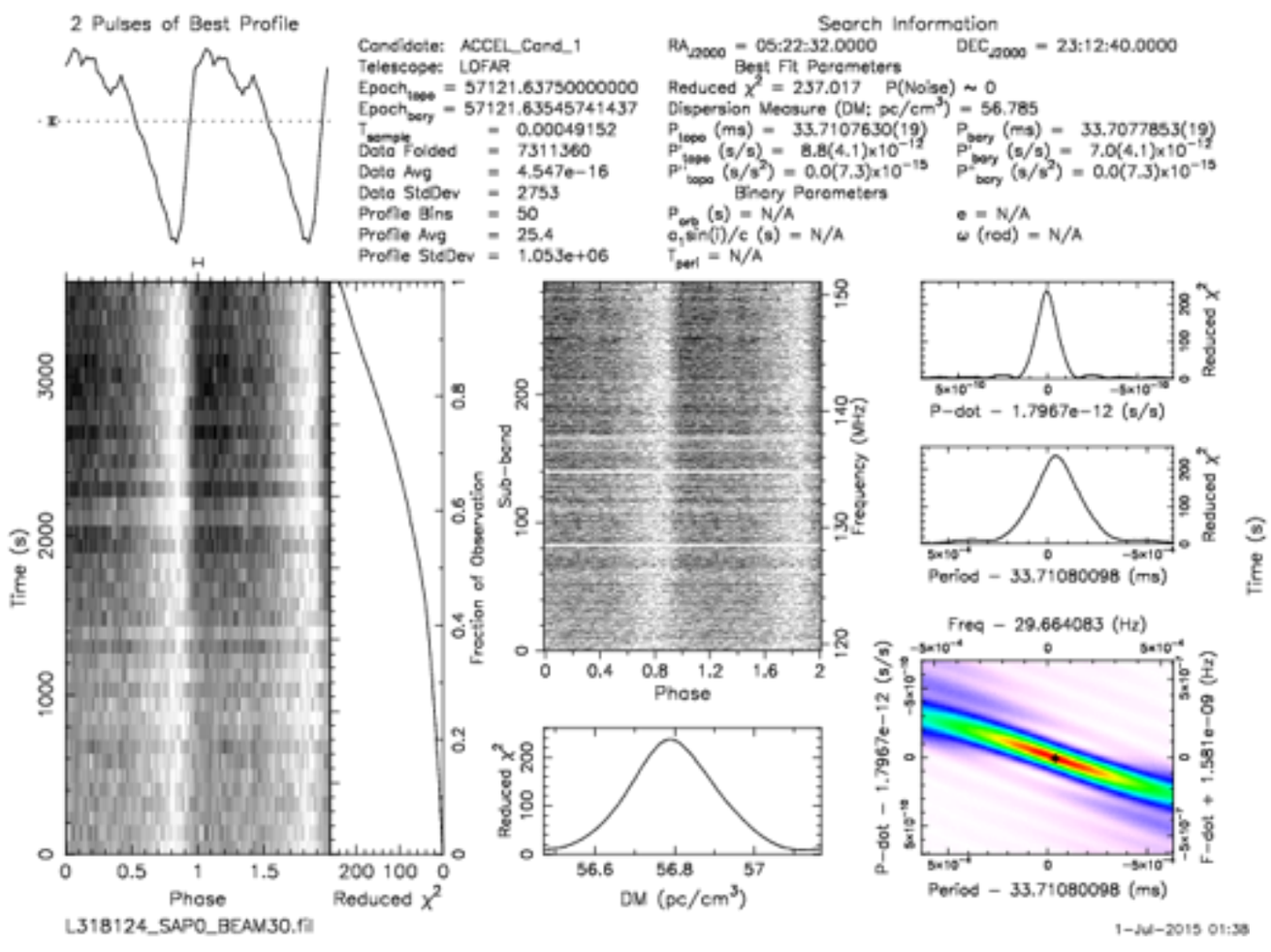


Issues with the Current Classifier

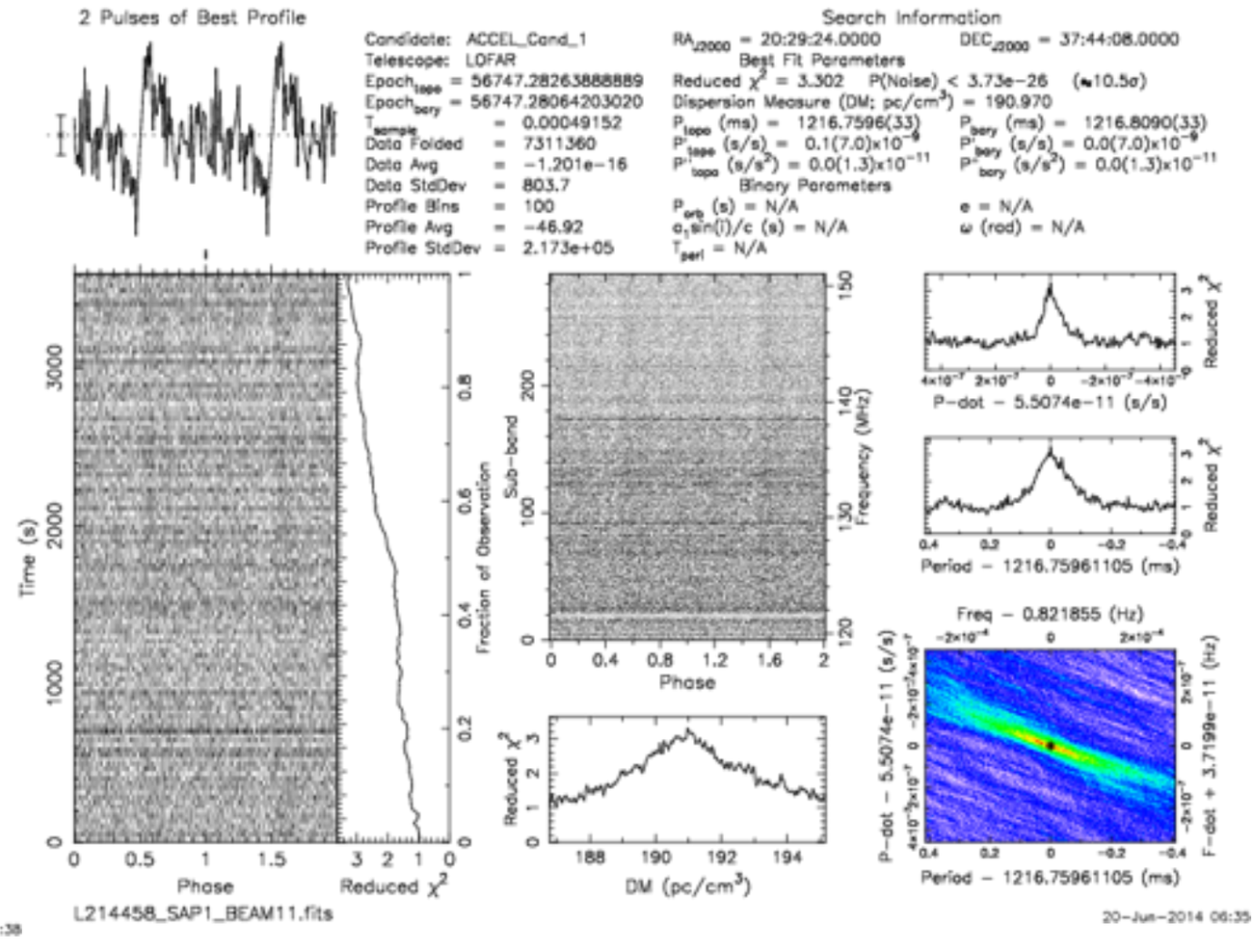
- The current classifier employed appears to be less effective at identifying pulsars with wide pulse profiles and highly scattered pulsars.

- These pulsars are known to be detectable within the field of view of the LOTAAS pointing but are sometimes not classified as pulsars by the classifier.

Examples of Pulsars Not Detected by the Classifier



Crab pulsar - The pulsar has a very wide pulse profile compared to most other pulsars detected.

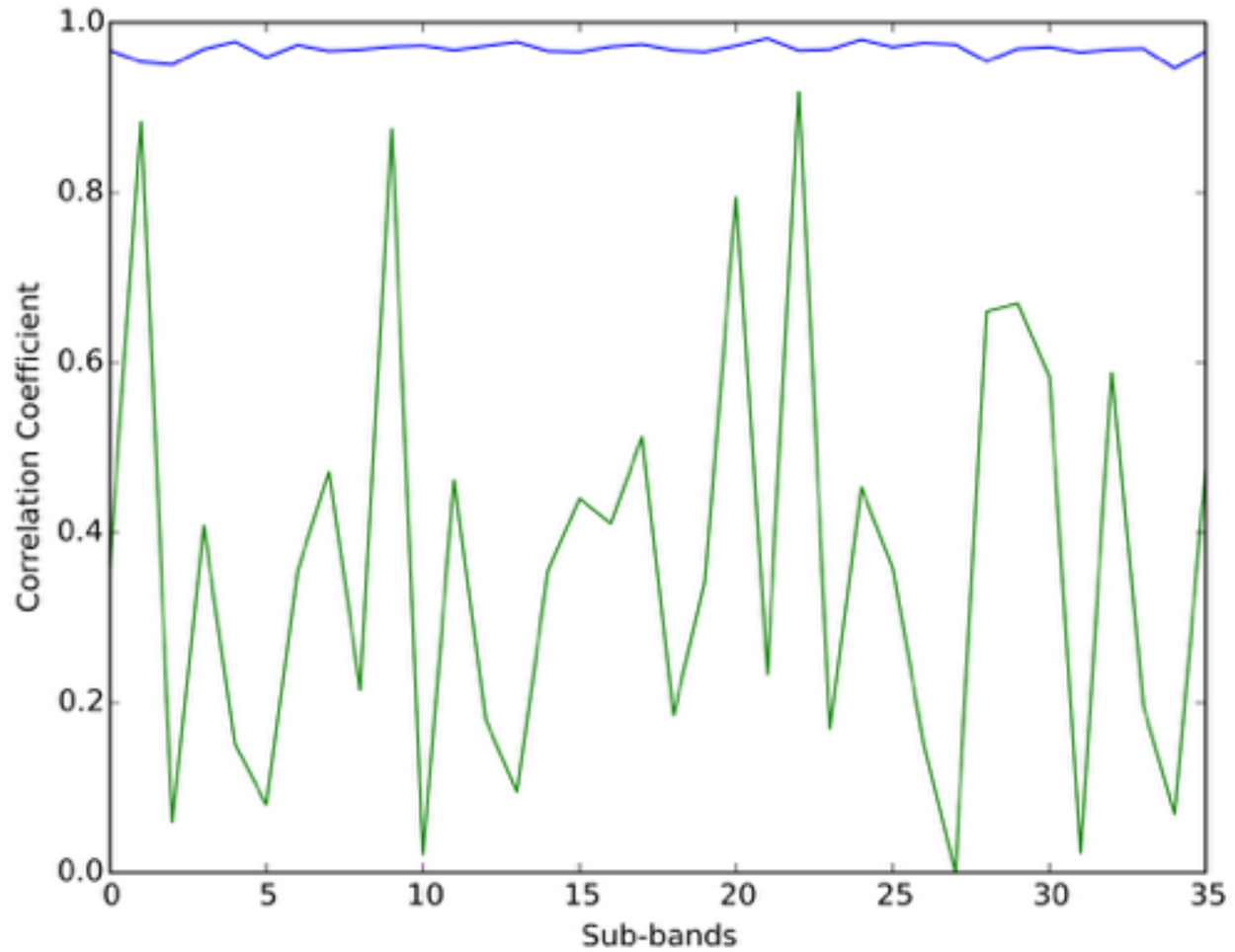


B2027+37 - The pulsar has a long scattering tail, usually seen with pulsars with large DM value.

New Features

- We calculated the correlation coefficients between each sub-band and the pulse profile. The same is done for each sub-integration and the pulse profile.

- The mean, standard deviation, skewness and excess kurtosis of the correlation coefficients are then calculated.

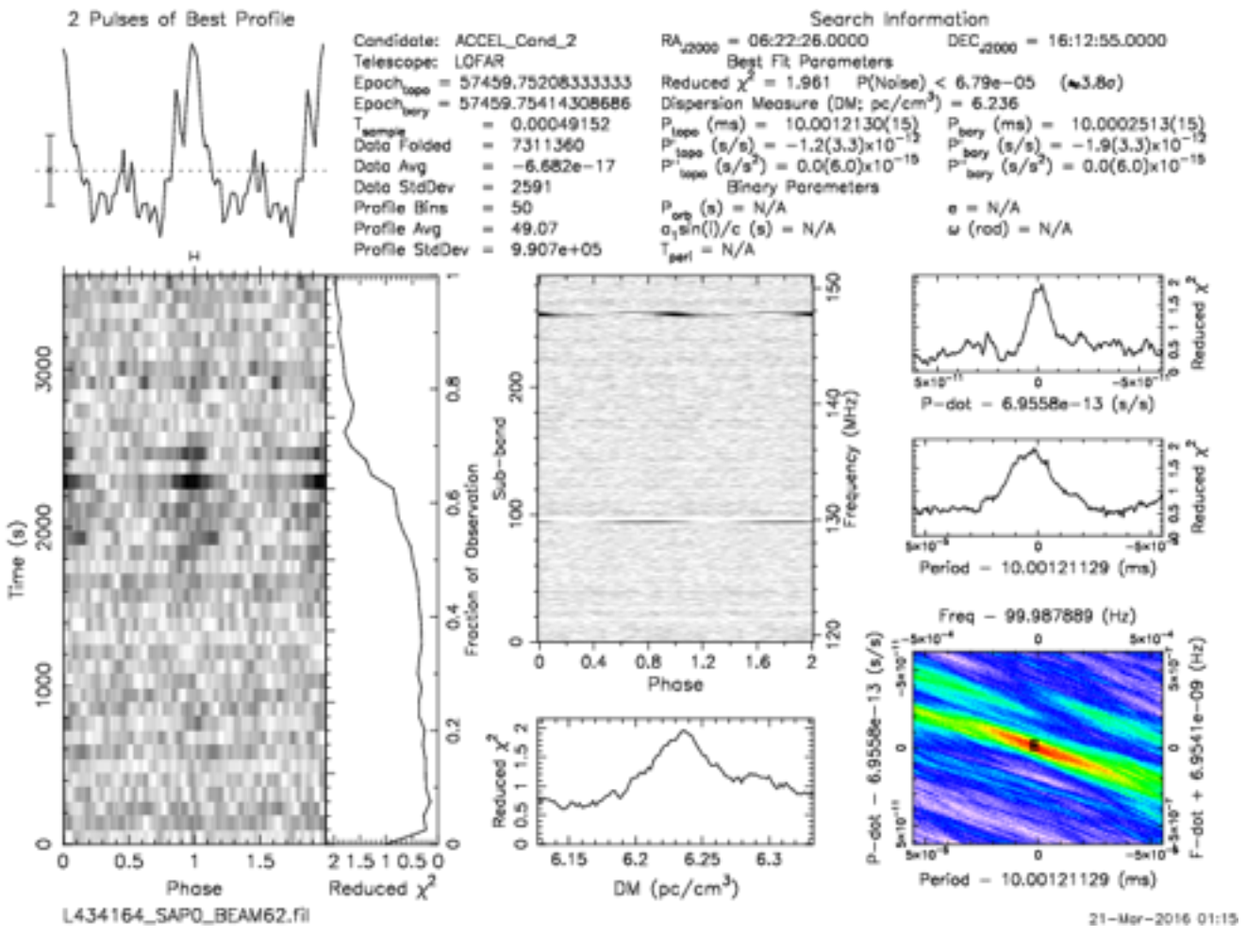


Pulsar

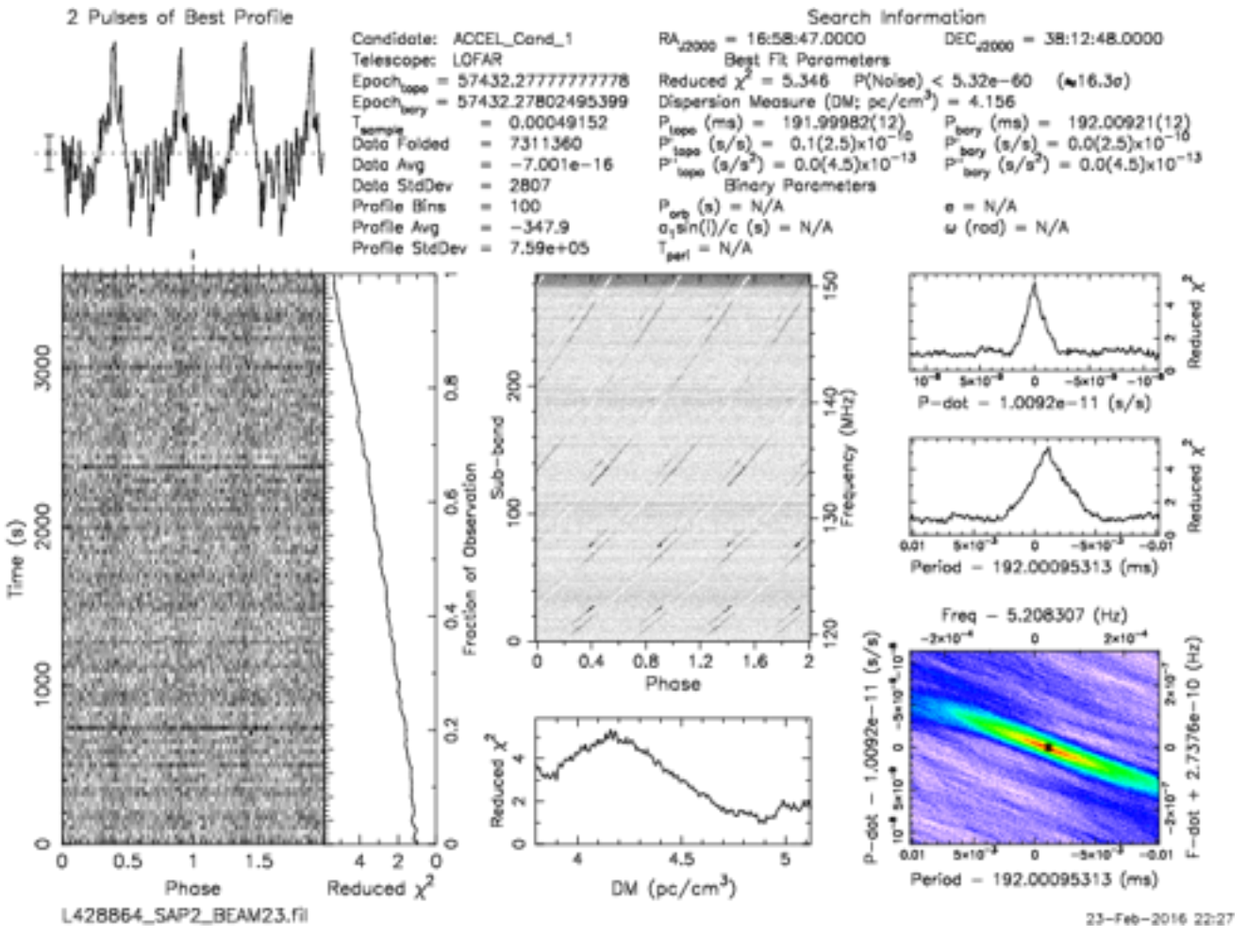
RFI

Classifying RFI candidates

- We assigned a third classification for candidates that are consists of known RFI instances. Plots below show the 2 most common RFI instances.



Narrow band RFI



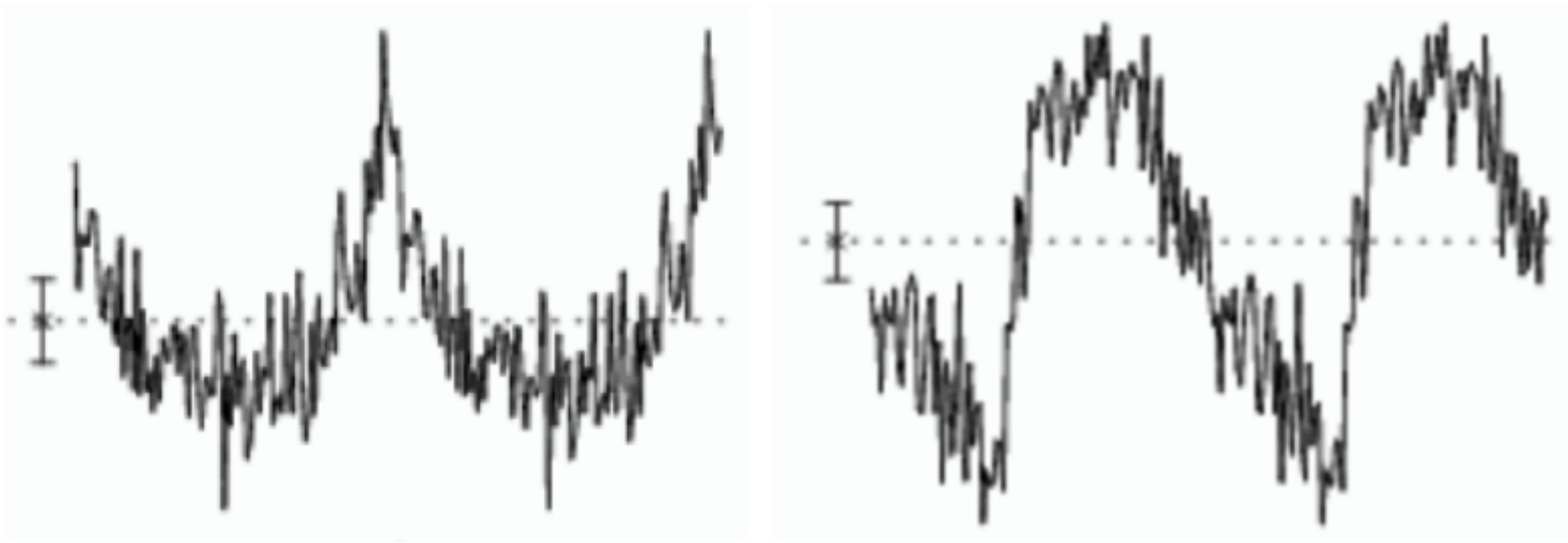
Broad band RFI

New Training Data

- We obtained candidate plots from 236 unique known pulsars detected by LOTAAS, including those that are overlooked by the current classifier.

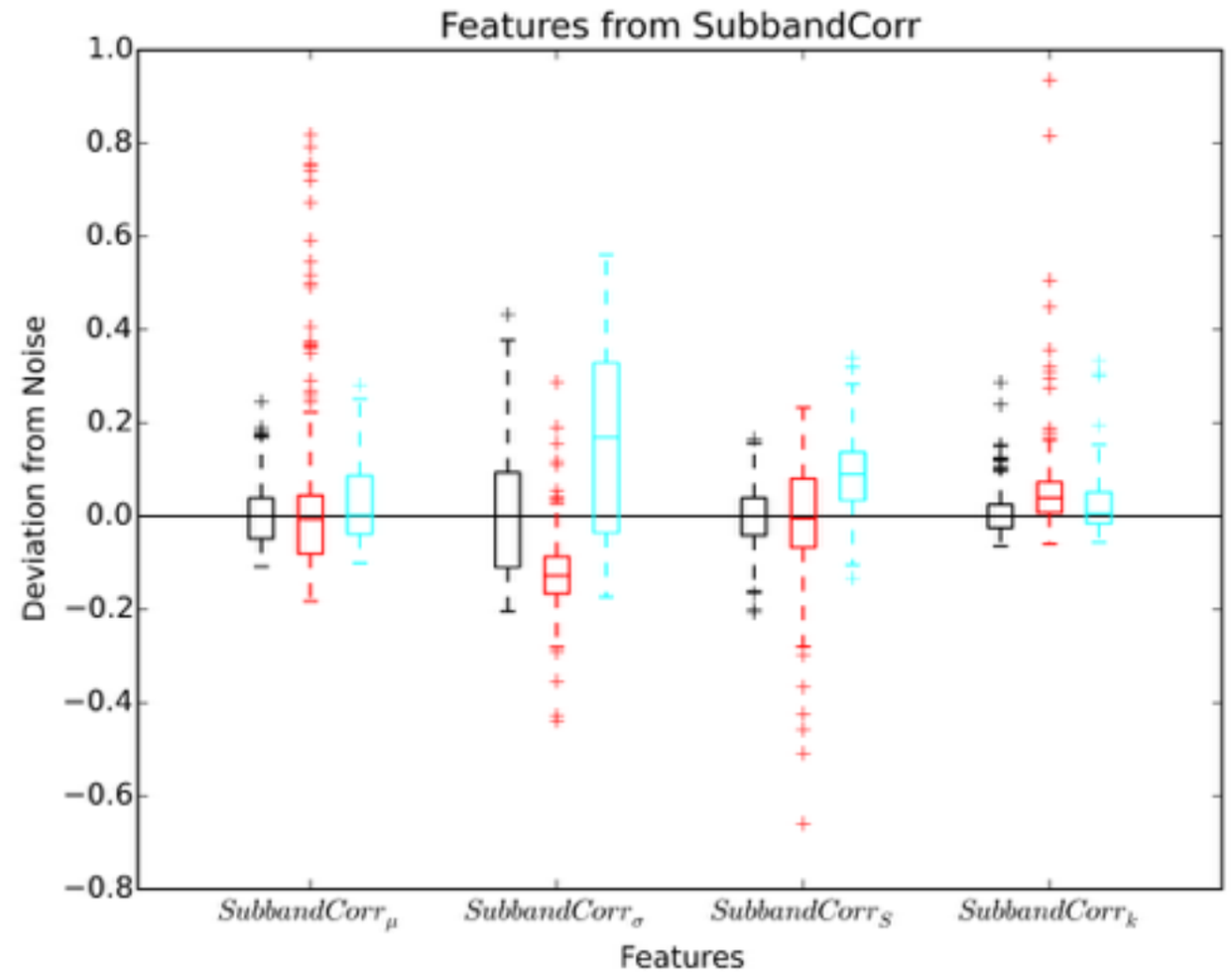
- We also added 26 candidate plots of pulsars observed by LOFAR for the LOFAR HBA census on pulsars (Bilous et al. 2016).

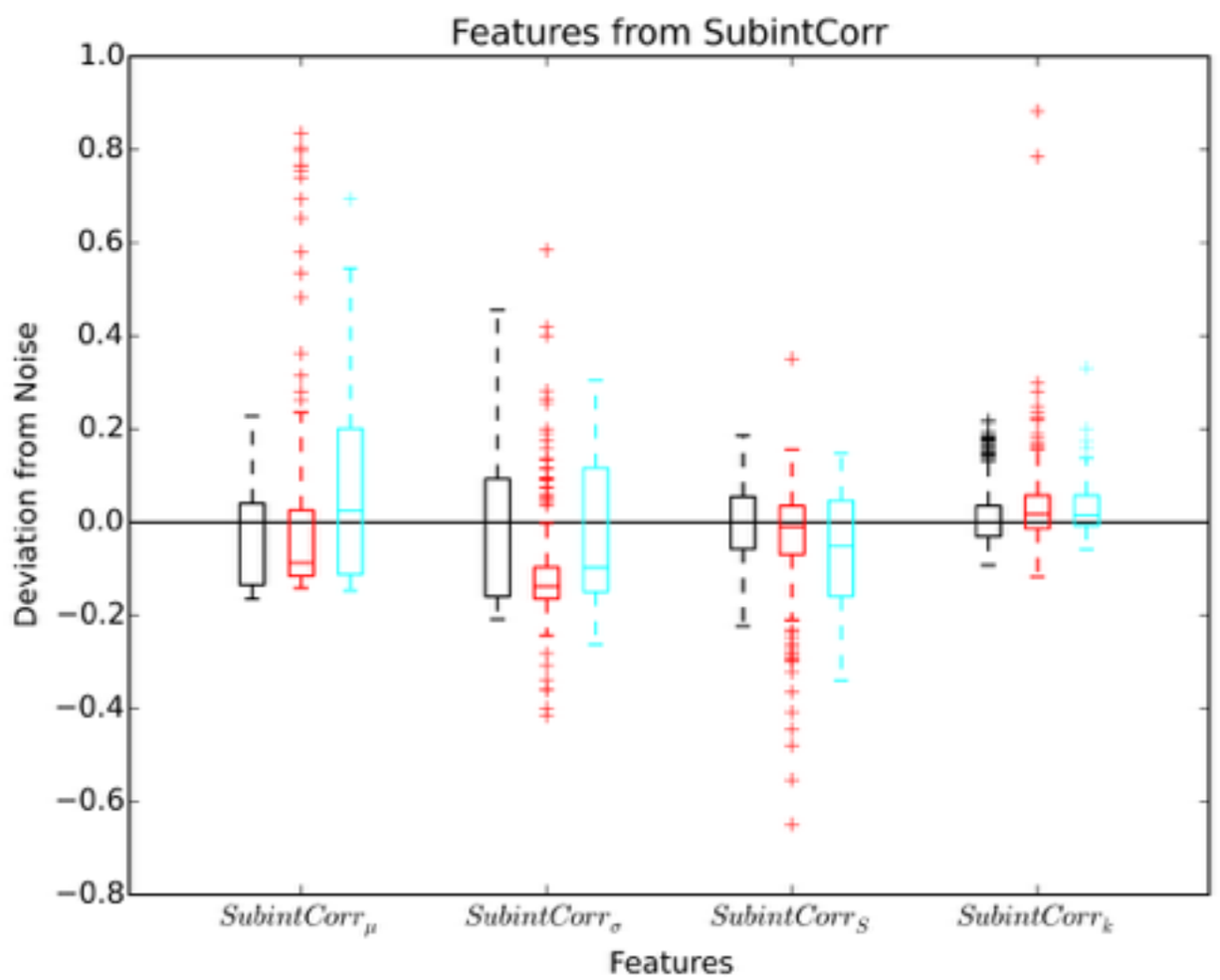
- The non pulsar candidates are consists of 548 noise candidates and 60 RFI instances.



Examples of unique pulse profiles from LOFAR HBA census

Boxplots of the New Features





Black : Noise

Red : Pulsars

Teal : RFI

Current Progress

- A new decision tree classifier is made using the new training data set with the new features. The tree is able to identify pulsars that are misclassified by the previous classifier.

- The validity of the classifier is currently being tested on several independent test data sets.

    1. An second data set using different detections of the pulsars detected by LOFAR and other instances of noise and RFI.

    2. 2 sets of labelled data from the HTRU pulsar survey by Thornton et al. (2014) and Morello et al. (2014)