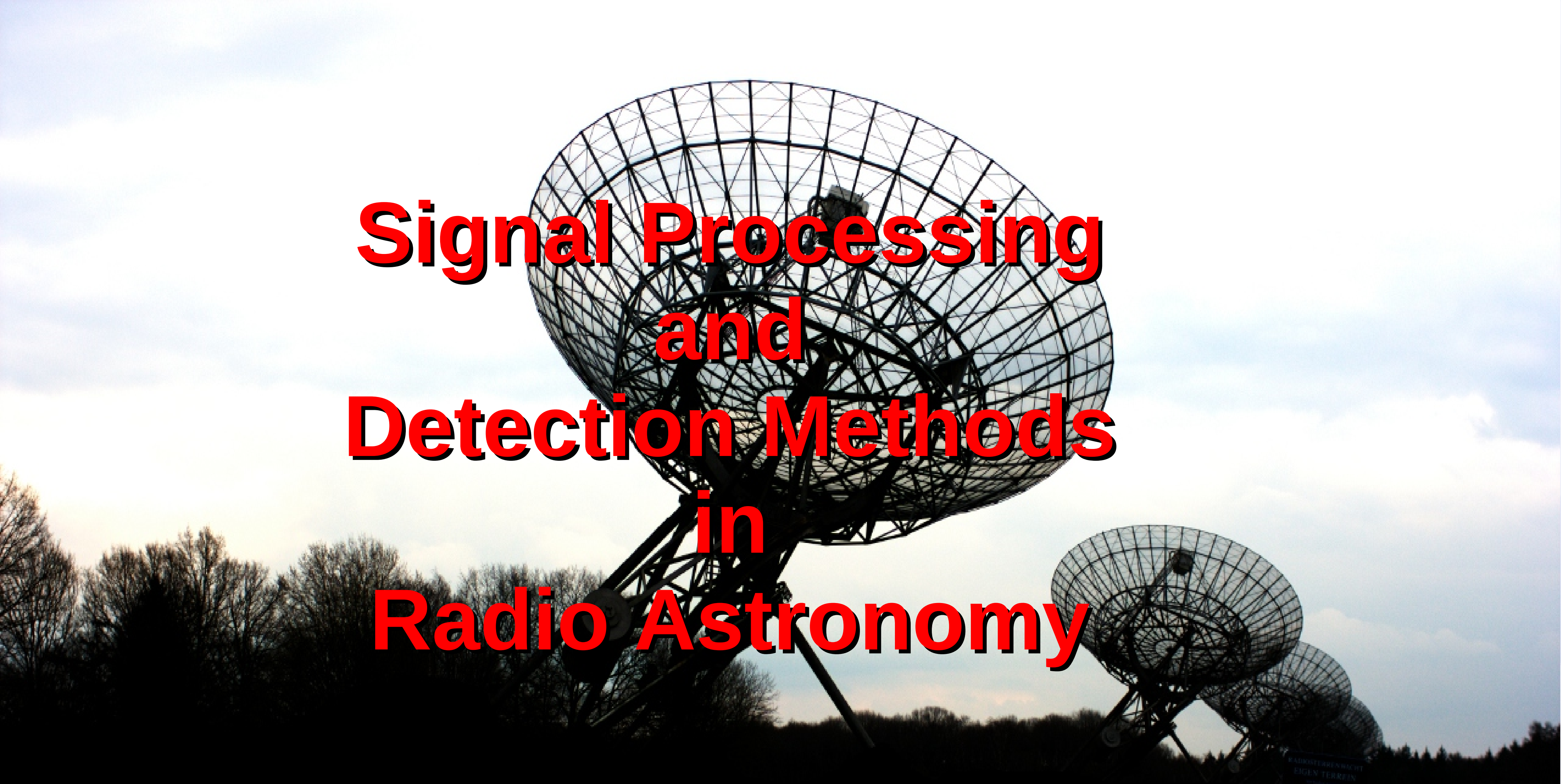


Signal Processing and Detection Methods in Radio Astronomy

André Offringa, Kapteyn Astronomical Institute, RUG

14 December 2011

Course on applied signal processing 2011

Outline

- Introduction on radio telescopes
- Data from telescopes
- Interferometers
- Processing steps of observations
- Interference & detection

- André Gunst wil discuss LOFAR and focus more on the signal chain.

What is a radio telescope?

Arecibo observatory

What is a radio telescope?

(E)VLA observatory

What is a radio telescope?

- A (large) antenna
- Measures EM radiation at radio wavelengths
- Tracks the sky
- Outputs *complex voltages (E)*
- Possibly in multiple polarizations

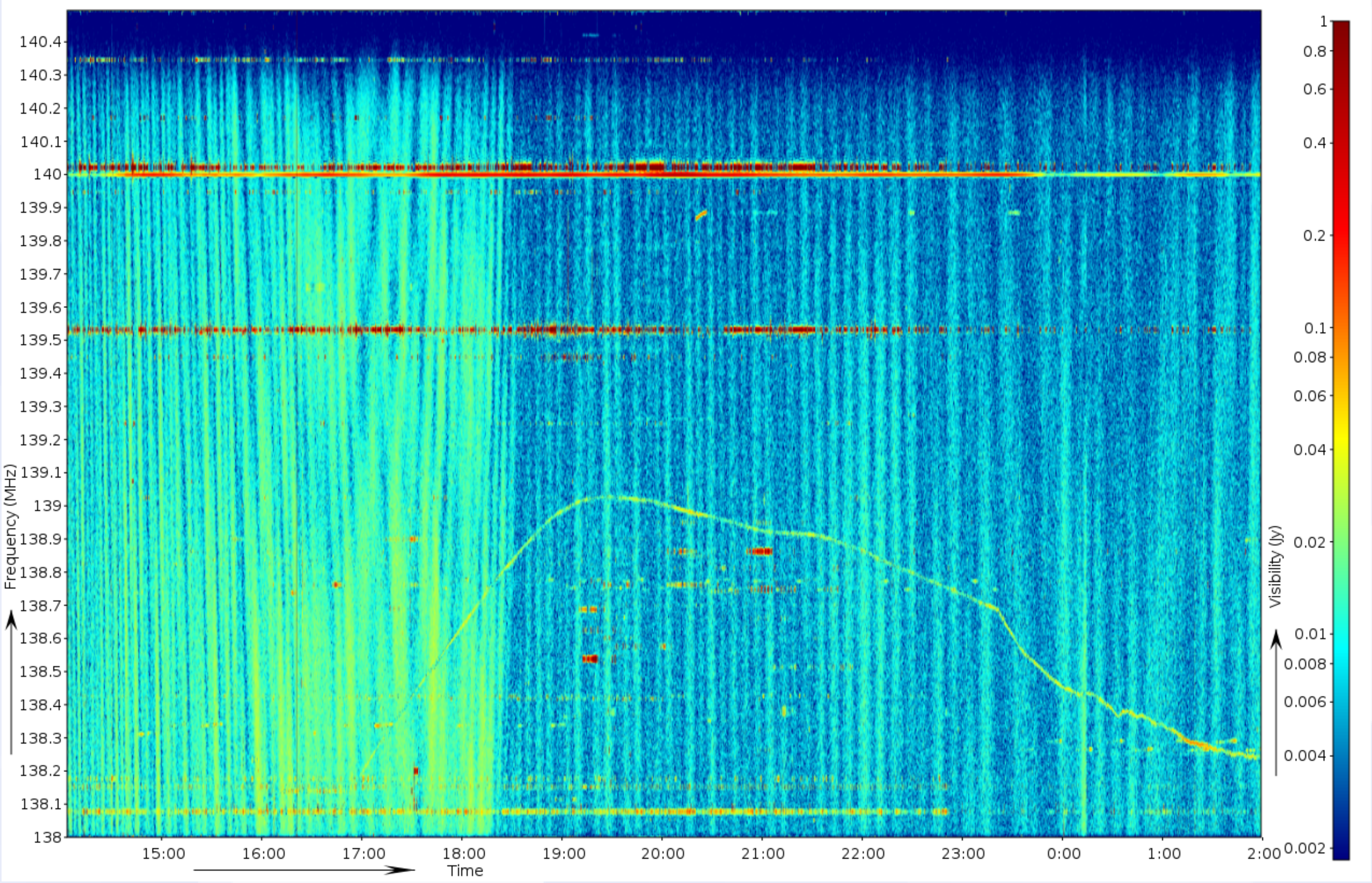
The product of a RT

The data stream from a telescope consists of

- *Complex voltages* as a function of time t and frequency ν: E(t, ν)
- Often with multiple polarizations
- André Gunst will discuss the first part of the data chain

A LOFAR antenna

Data from a RT



Data from a RT

- One quasi monochromatic source:
  $E(t, \nu) = a(l,m) \exp(-i2\pi t \nu)$
- Delayed & correlated...

Data from a RT

- One quasi monochromatic source:
  $E_1(t, \nu) = a(l, m) \exp(-i2\pi\, t\, \nu)$
- Delayed 2nd telescope:
  $E_2(t, \nu) = a(l, m) \exp(-i2\pi\, (t-\Delta t)\, \nu)$
  $E_2(t, \nu) = a(l, m) \exp(-i2\pi\, (t-\tau(u, v))\, \nu)$
- Correlated: $V(u, v) = [\, E(t, \nu)\, E^*(t, \nu)\,]$
  $= I(l, m) \exp(i2\pi\nu\tau(u, v))$
- ...

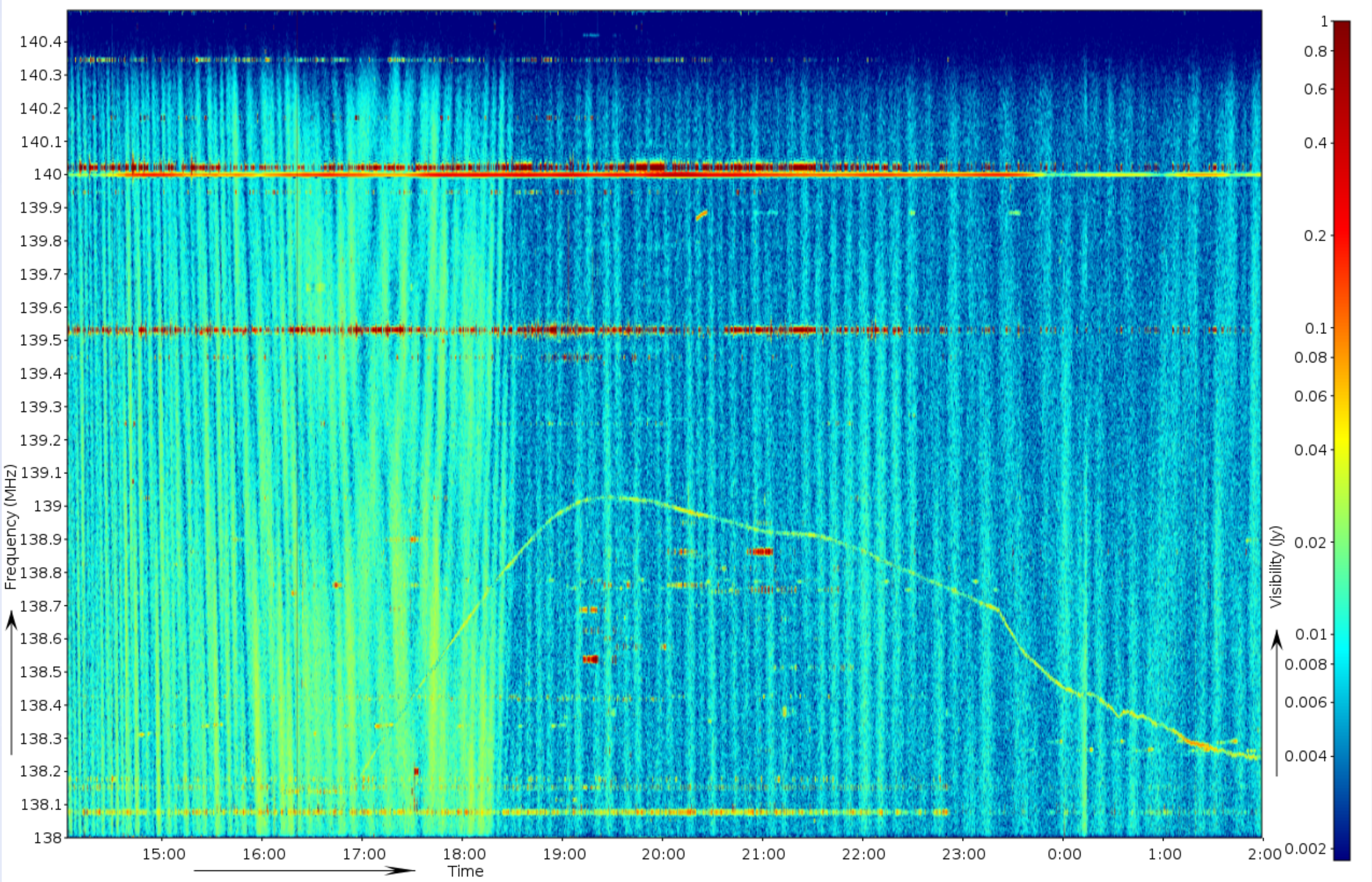
Data from a RT

- $V(u, v) = I(l, m) \exp(i2\pi\nu\tau(u, v))$
- Integrated over sources (all sky):
  $V(u, v) = \iint I(l, m)\, e^{i2\pi\nu\tau(u, v)}\, dl\,dm$
- u, v chosen such that in 2d:
  $V(u, v) = \iint I(l, m)\, e^{i2\pi\nu(lu+vm)}\, dl\,dm$
- Does this look familiar?

Data from a RT

- $V(t, \nu) = E(t, \nu)\, E^*(t, \nu)$
- This does not change rapidly over time
- Gaussian noise in real/imaginary
- Integrate over time:
  $\tilde{V}(t, \nu) = \langle E(t, \nu)\, E^*(t, \nu) \rangle$
- Performed by a “correlator”

Data from a RT



Processing of the correlated data

- Detect interference
- Calibrate
- Image
- Deconvolve
- Extract signal (/ sources)

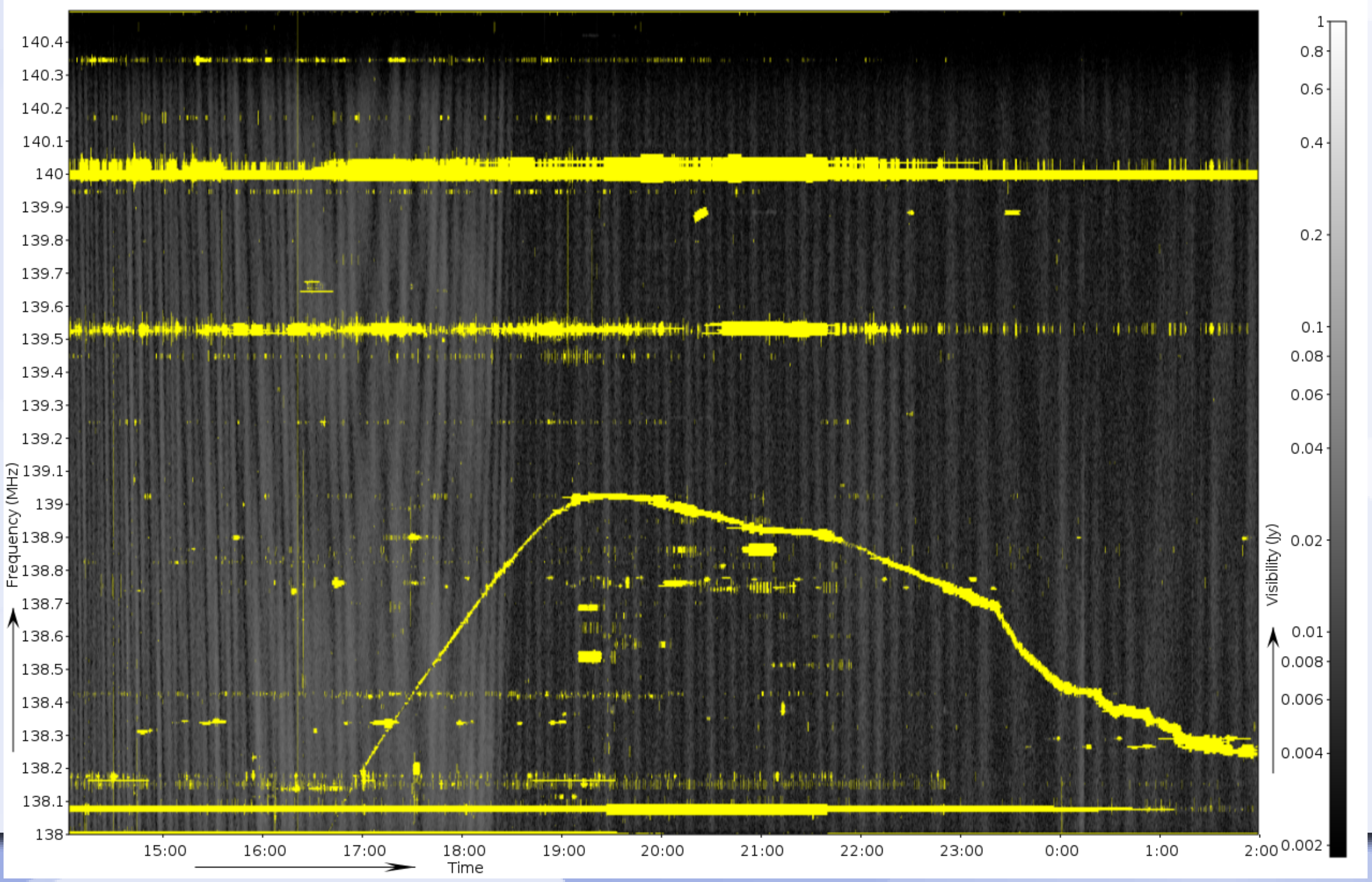
- LARGE(!) data volumes

Step 1: interference detection

- Signal of interest interfered by strong transmitters ('RFI'):
  - Man-made:
    fm radio, (weather) radars, airplanes, satellites, electric fences, HV lines, …
  - But also natural interence:
    Lightning, the sun
- Last resort: detection and ignoring in further steps
- (methods will be explained later)

Step 1: interference detection

- Detect interference:



Step 2: calibration

- We want an accurate brightness measure
- Many instrumental effects:
  - Beam, ionosphere, temperature effects (cable lengths), band filters, …
- Approaches:
  - Use an external calibrator
  - Self-calibration

Step 3: imaging

- Perform the Fourier Transform
- Implemented with FFT

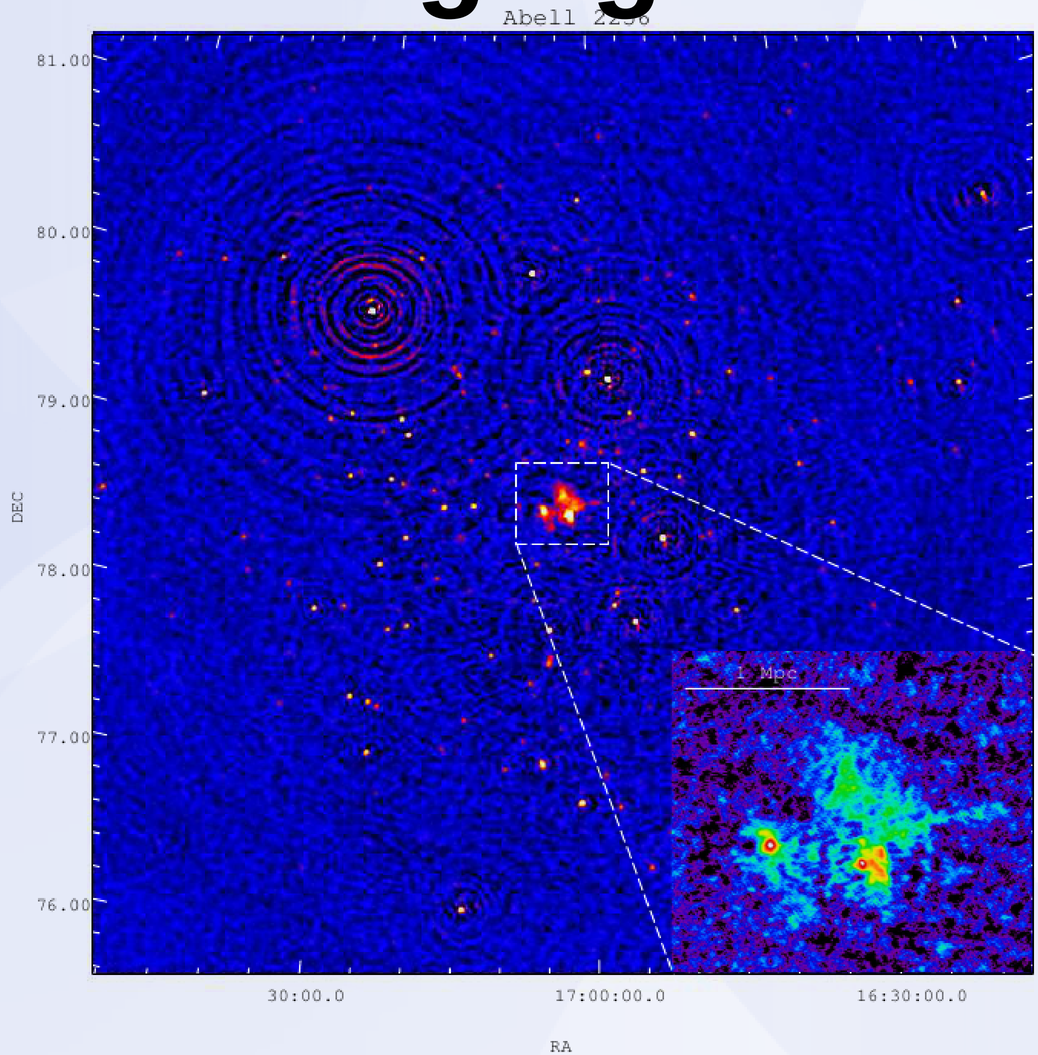
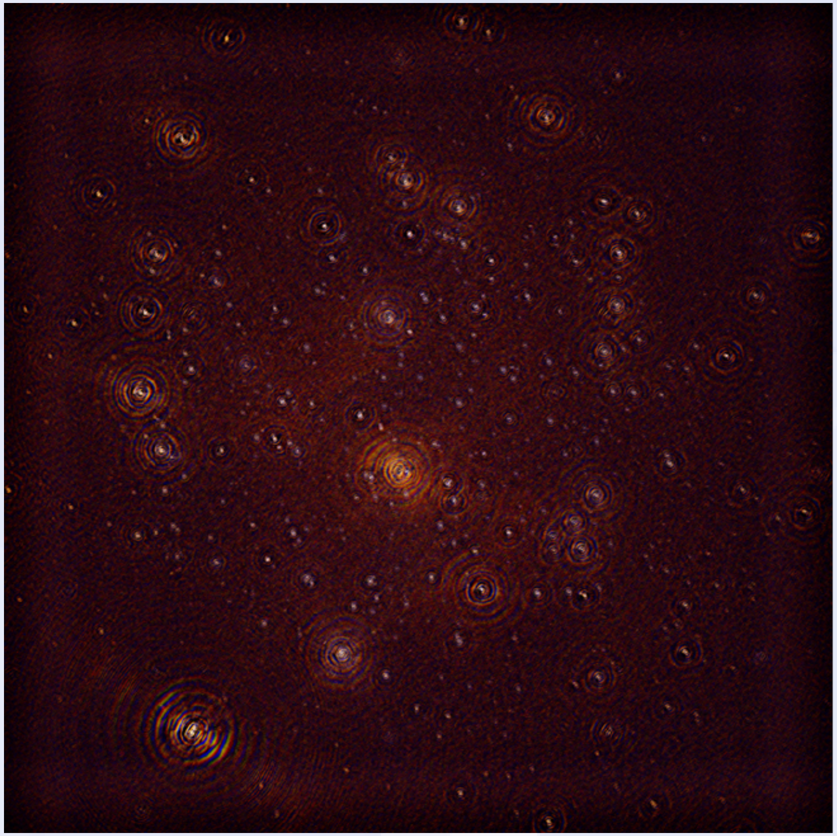


Fig. 4. Overview of the A2256 field at 61–67 MHz compared to the VLSS survey image. Left: FOV centered on A2256 (this image was not corrected for the primary beam attenuation). Top right: the 74 MHz VLSS image with a resolution of 80″. Bottom right: Zoomed version of the 61–67 MHz image with a synthesized beam of 22″ × 26″.

Image: Van Weeren et al., 2011, to be submitted

Another LOFAR example



Sarod Yattawatta, NCPfield, 2011

- LOFAR image
- 6 hr observation
- Not deconvolved

Step 4: deconvolution

- 'Rings' caused by finite FT.
- Deconvolution removes the PSF
- Often combined with imaging & source subtraction

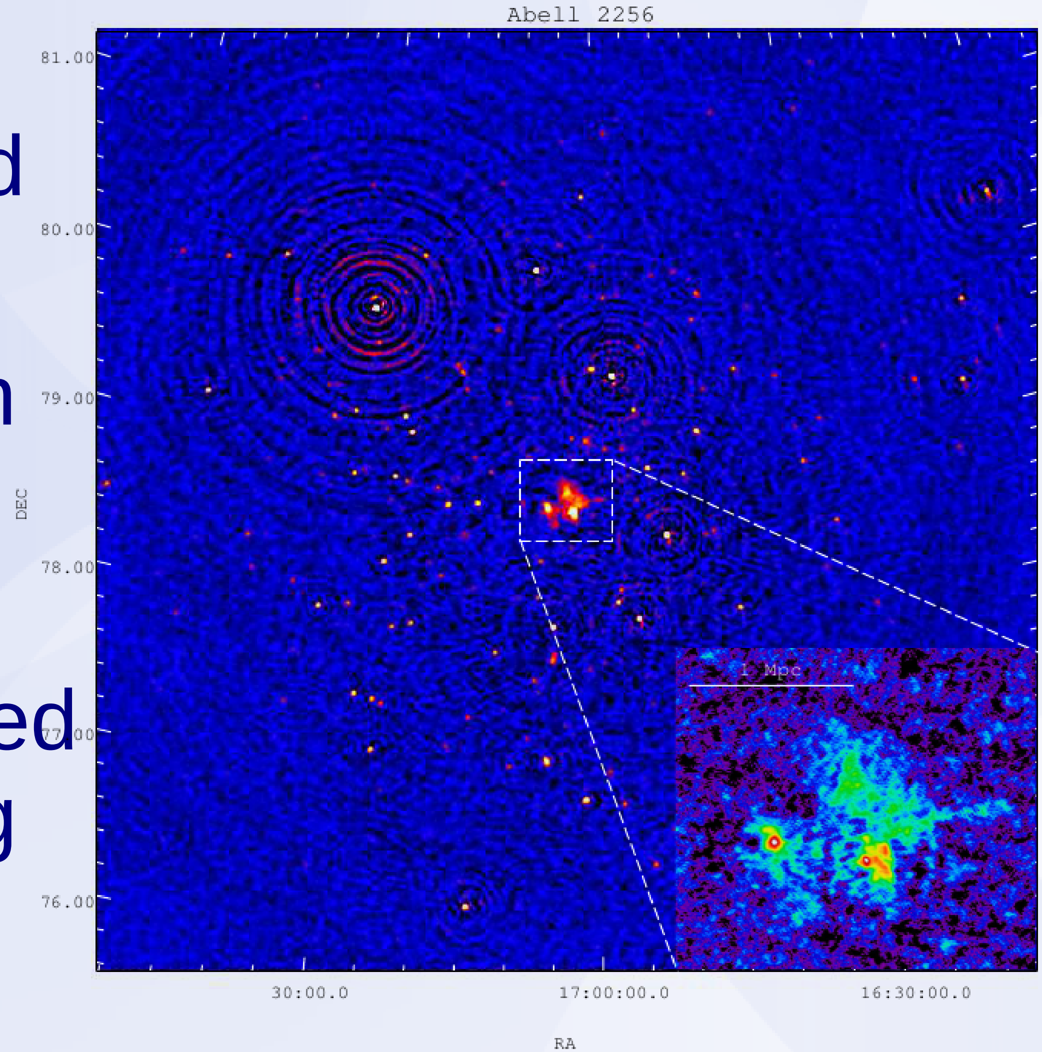
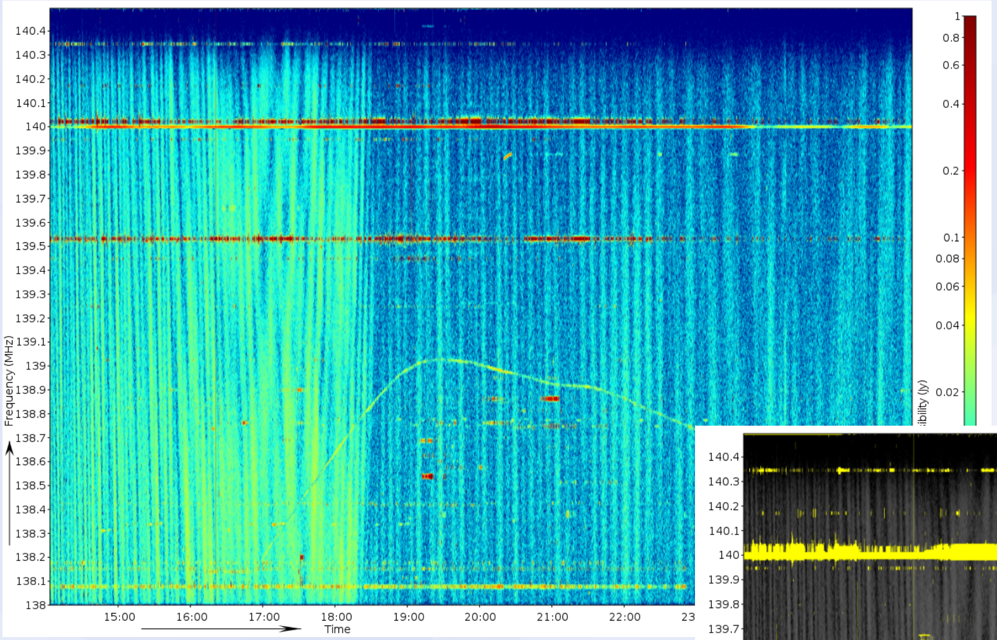
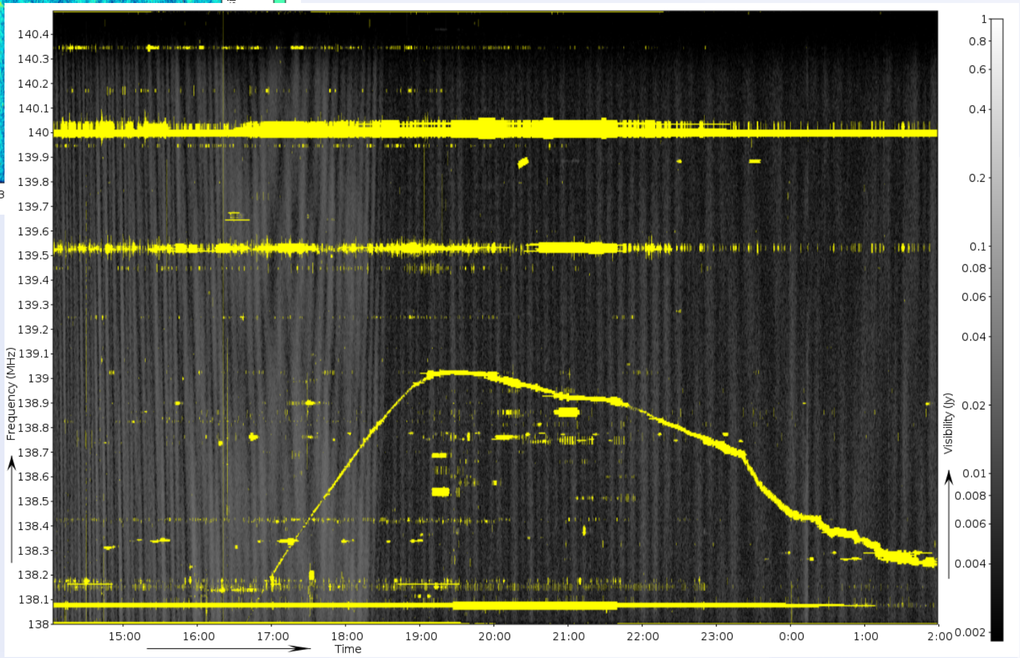


Fig. 4. Overview of the A2256 field at 61–67 MHz compared to the VLSS survey image. Left: FOV centered on A2256 (this image was not corrected for the primary beam attenuation). Top right: the 74 MHz VLSS image with a resolution of 80″. Bottom right: Zoomed version of the 61–67 MHz image with a synthesized beam of 22″ × 26″.

Image: Van Weeren et al., 2011, to be submitted

How to detect RFI?

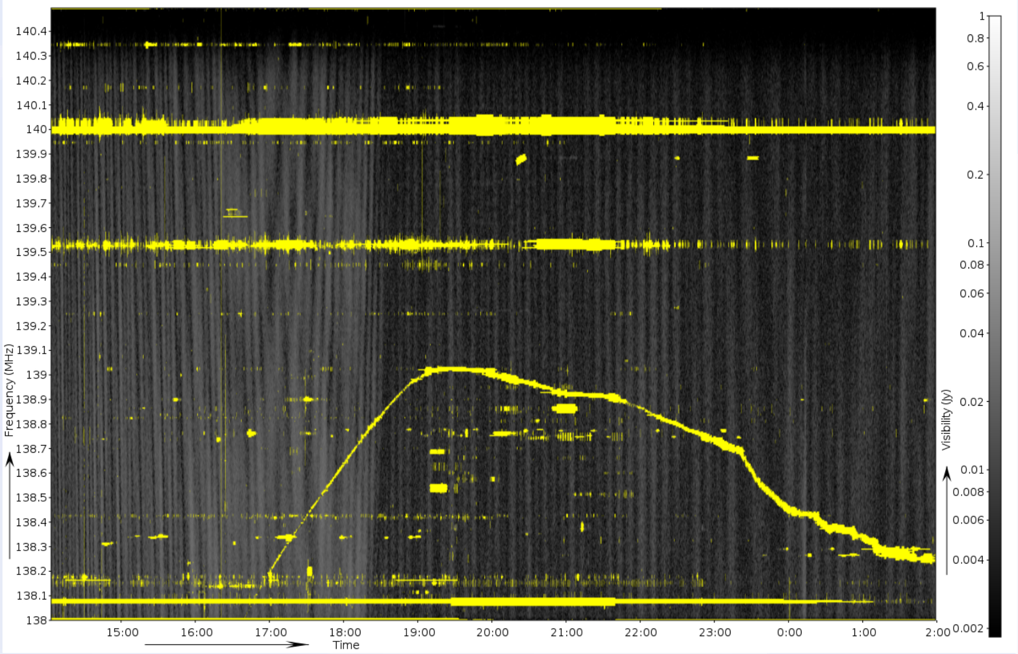




This process is called “data flagging”

How to detect RFI?

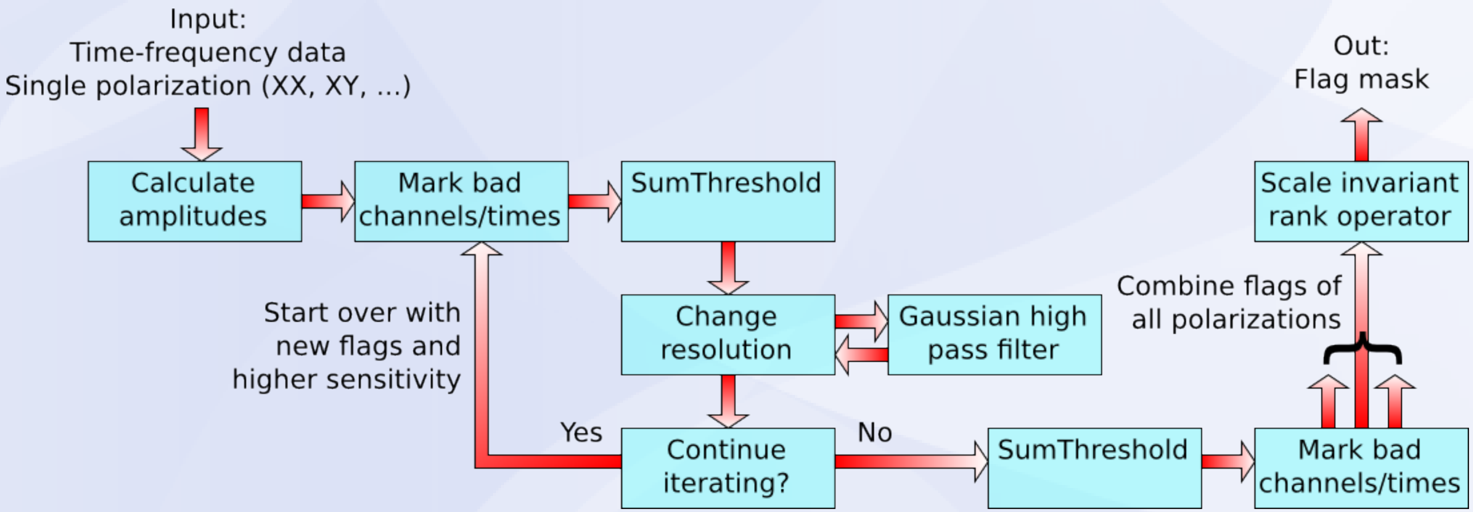
- Used to be done “ad hoc” by astronomer
- (Visually) looking for contaminated baselines, antennas, channels or time ranges.
- Huge observations: No longer feasible!



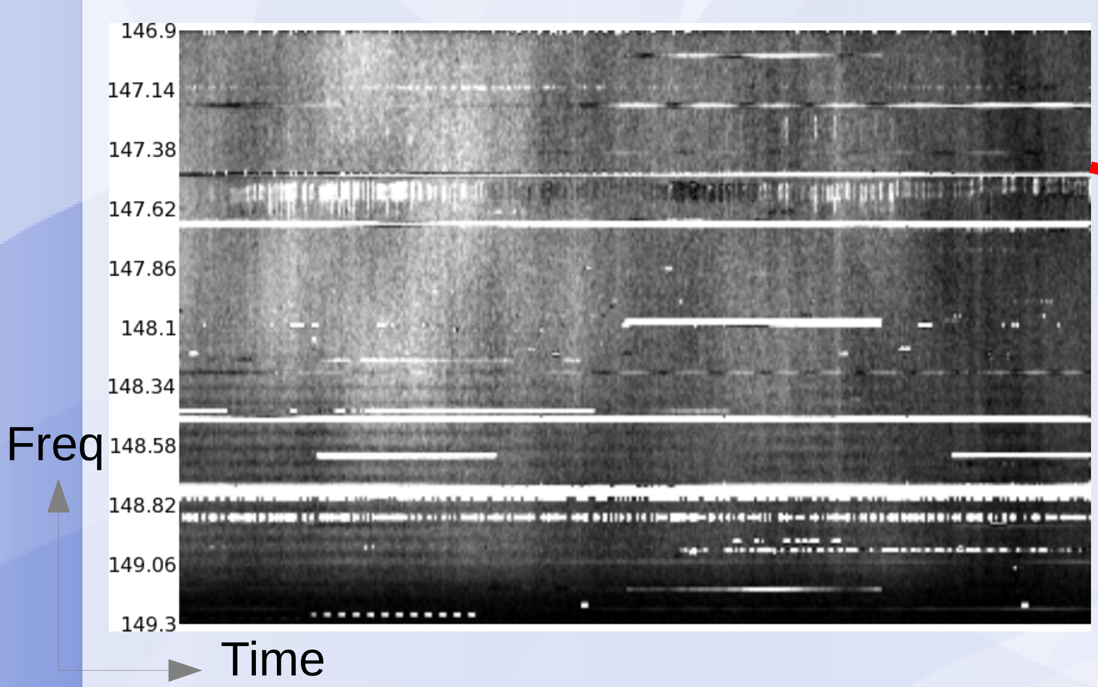
How to detect RFI?

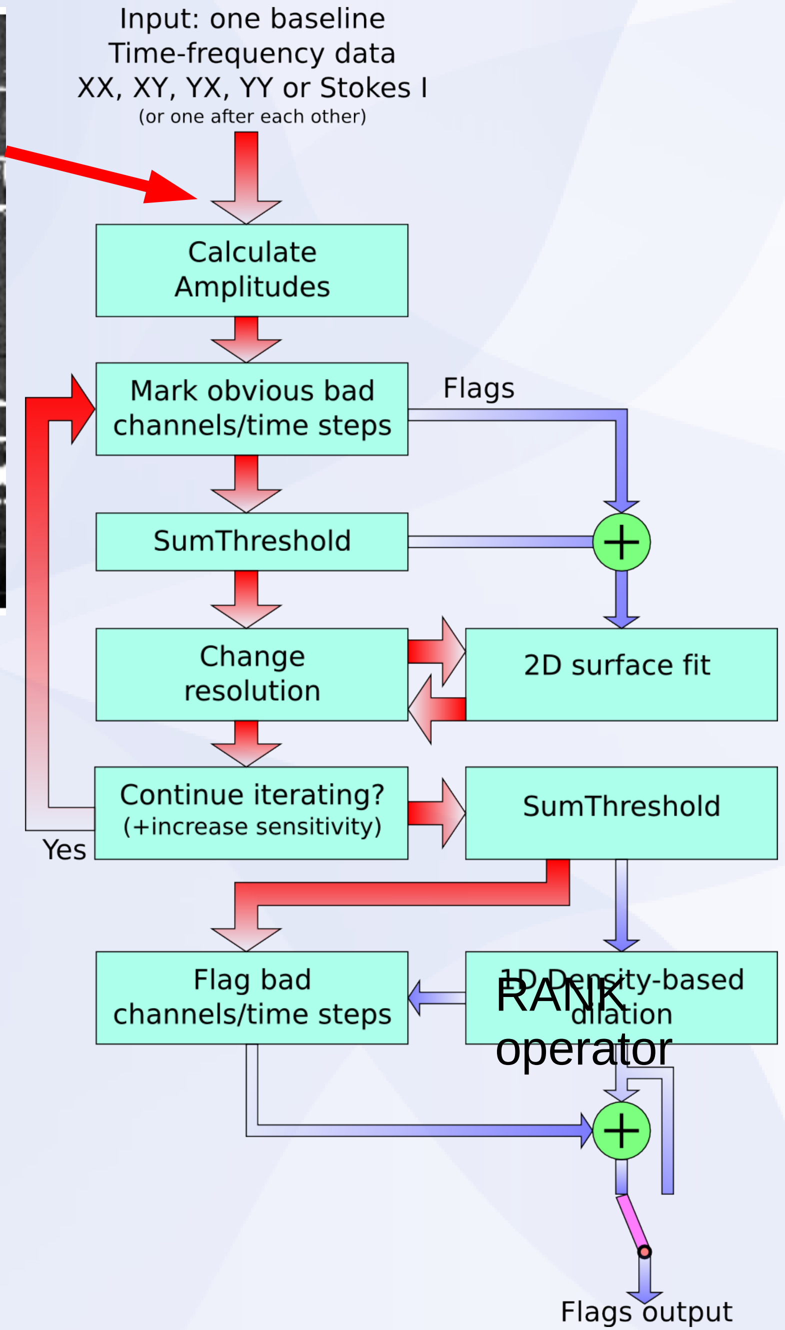
- Detection is a 'last resort'
- Other techniques:
  - Turn radiating devices off (!)
  - Beam shaping
  - (Spatial) Filtering
  - Modeling and subtraction of RFI.
  - Reference antenna

AOFlagger

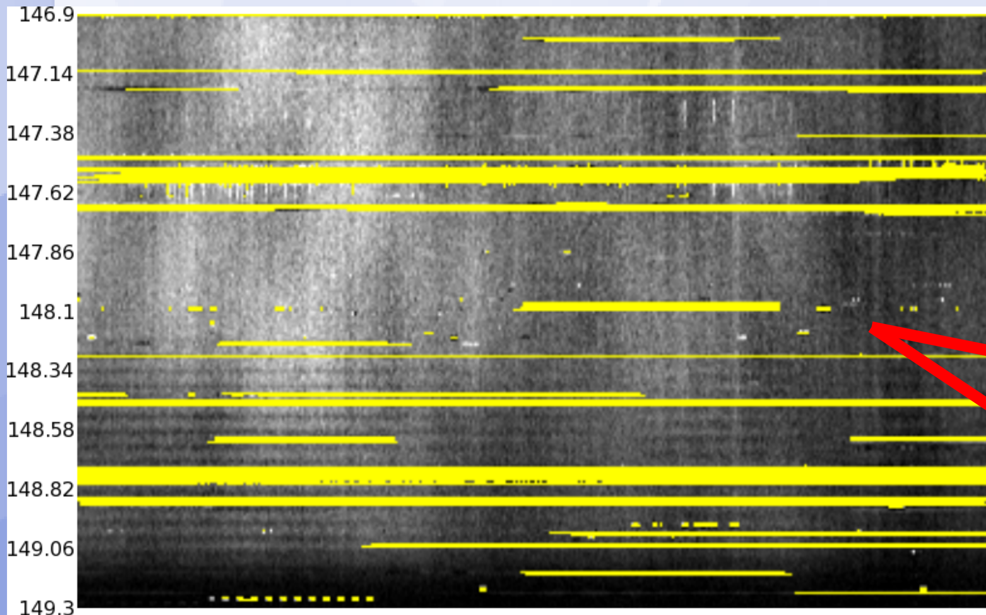


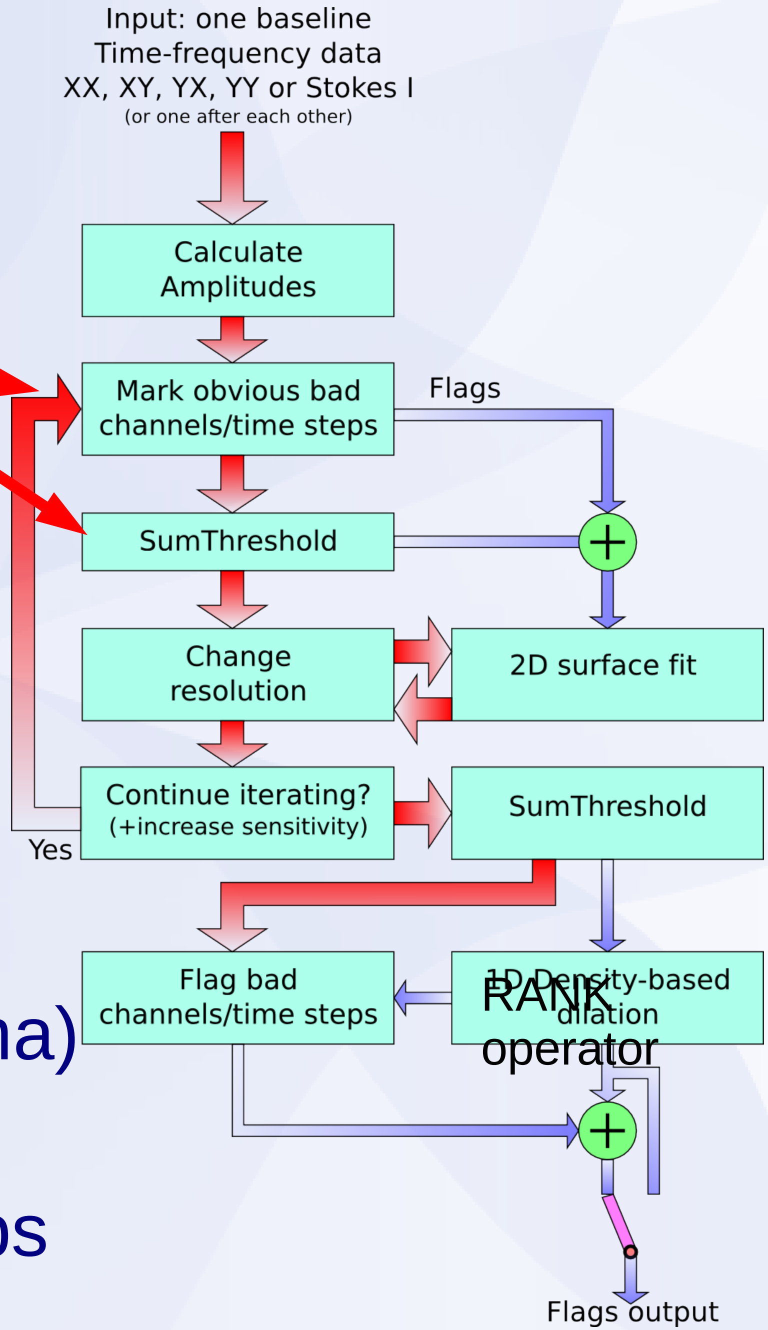
See Offringa et al., MNRAS (2010) & Offringa et al., RFI2010


146.9
147.14
147.38
147.62
147.86
148.1
148.34
148.58
148.82
149.06
149.3
Freq
Time


Input: one baseline
Time-frequency data
XX, XY, YX, YY or Stokes I
(or one after each other)
Calculate Amplitudes
Mark obvious bad channels/time steps
Flags
SumThreshold
Change resolution
2D surface fit
Continue iterating? (+increase sensitivity)
Yes
SumThreshold
Flag bad channels/time steps
1D Density-based dilation
RANK operator
Flags output

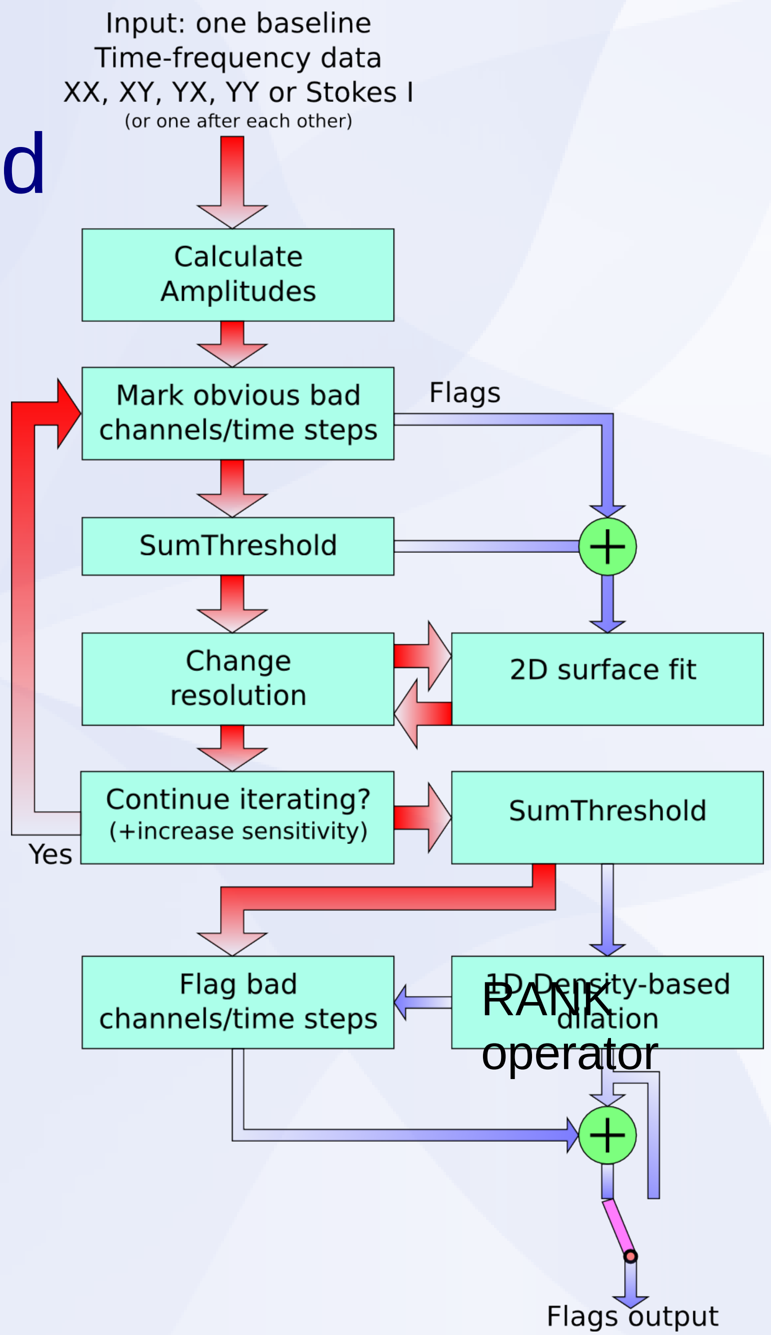
- Input...





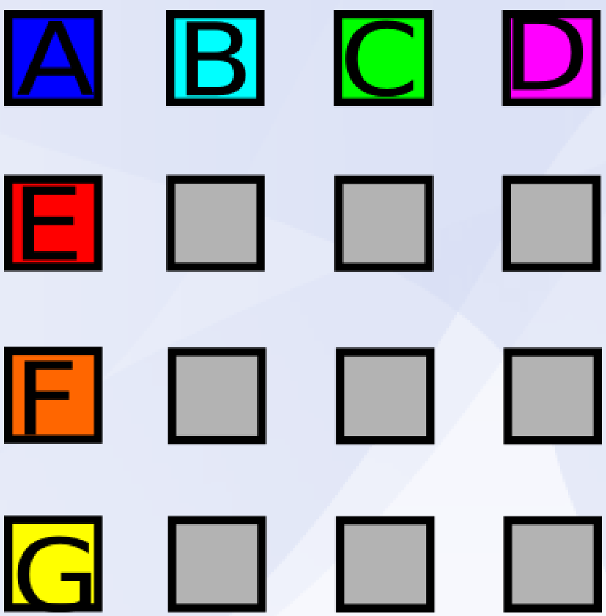
- First quick look at the data:
  - Flag extreme value samples (> 8 x sigma)
  - Flag on power in channels / time steps

- **SumThreshold** is a combinatorial threshold technique...



SumThreshold

- Combinatorial thresholding strategy
- Fast & accurate
- Idea:
  - Sum samples and use different thresholds



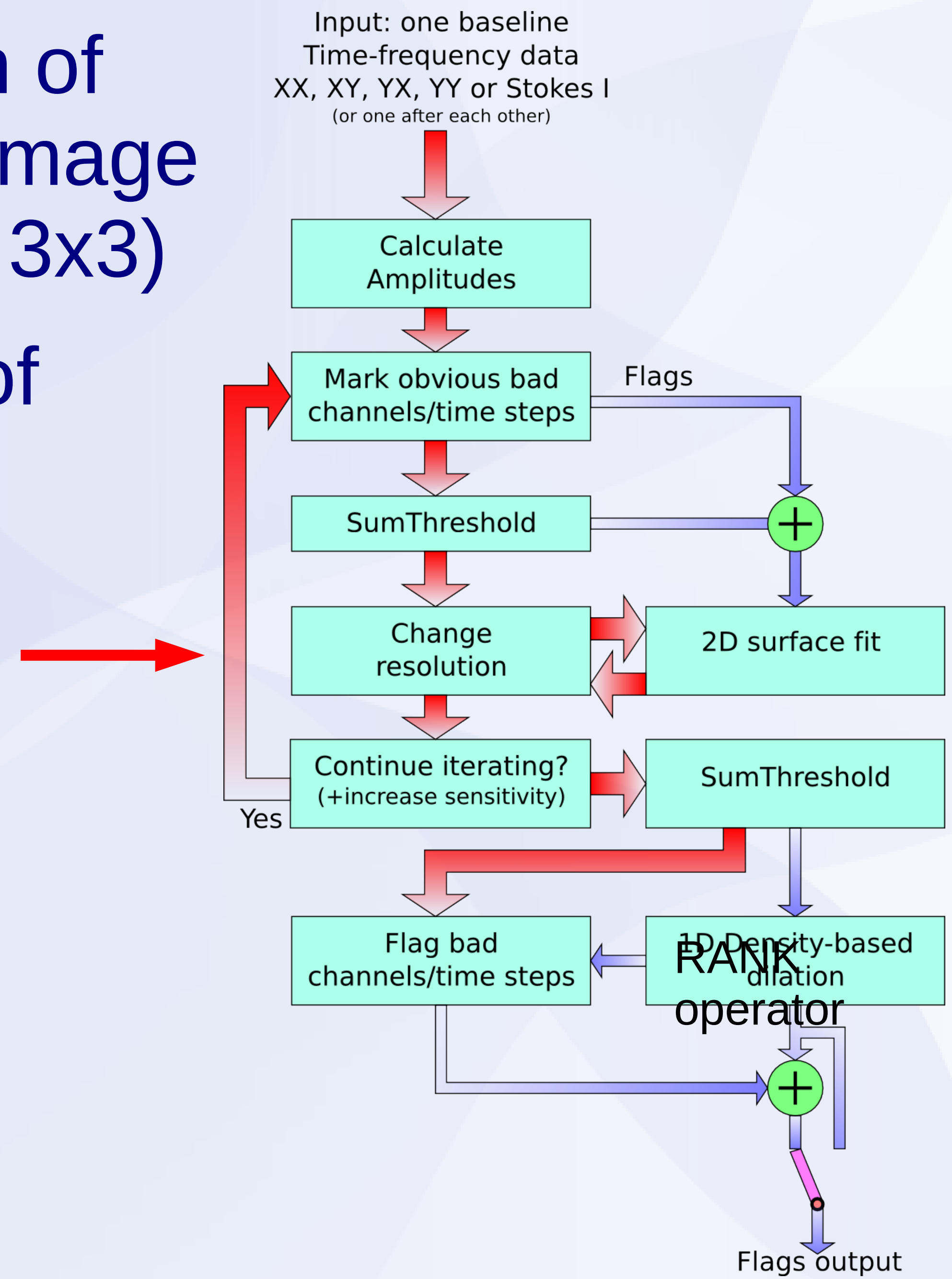
A	> threshold1?	→	FLAG A
A+B	> threshold2?	→	FLAG A, B
A+B+C	> threshold3?	→	FLAG A, B, C
A+B+C+D	> threshold4?	→	FLAG A, B, C, D
A+E	> threshold2?	→	FLAG A, E
A+E+F	> threshold3?	→	FLAG A, E, F
A+E+F+G	> threshold4?	→	FLAG A, E, F, G
B	> threshold1?	→	FLAG B
B+C	> threshold2?	→	FLAG B, C

....

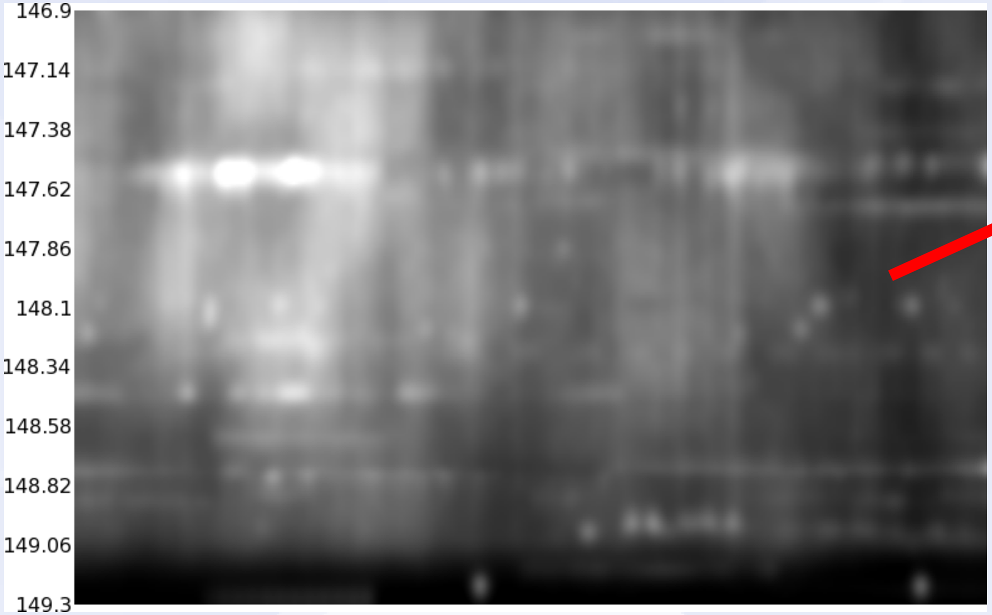
SumThreshold

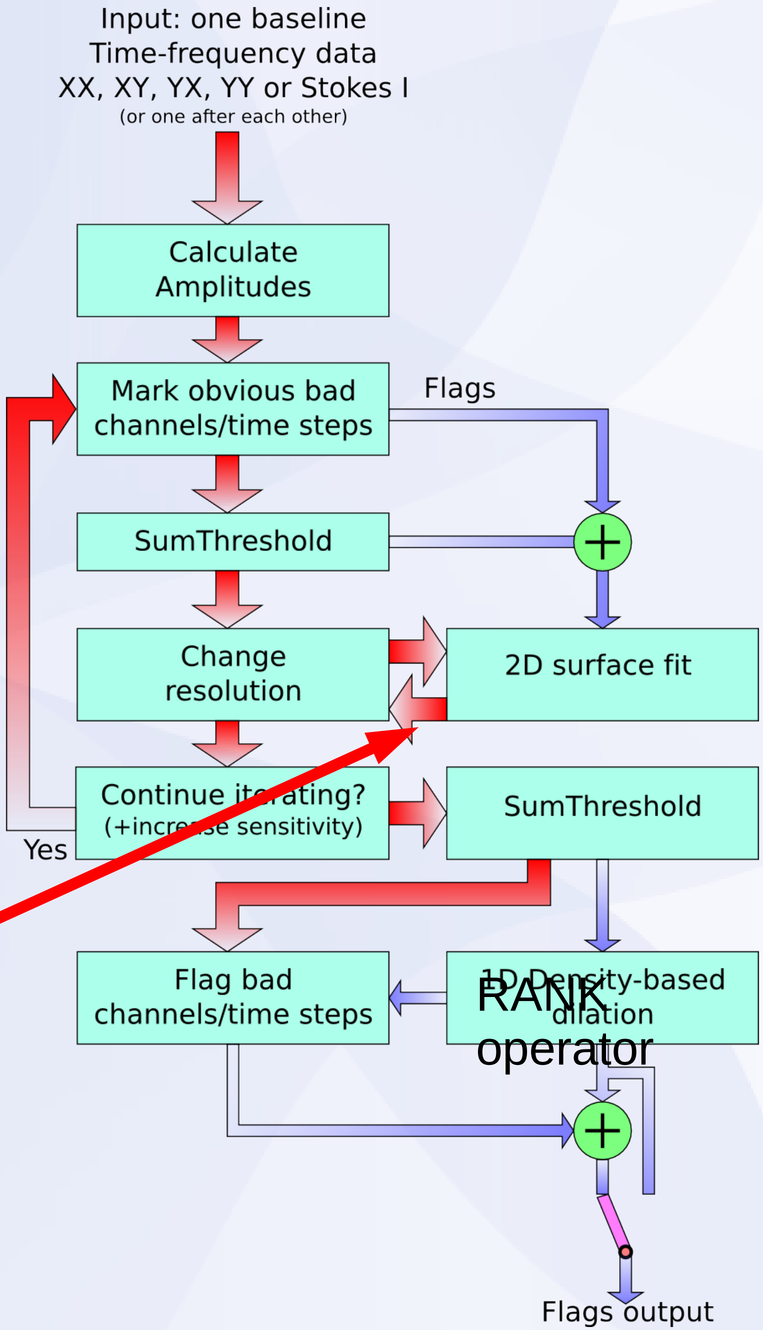
- How to determine 'thresholds'?
- Use the variance of the (residual) data
- Variance strongly biased by RFI...
  - Use “stable” statistics, e.g. trimmed or Winsorized mean&variance.

- Change resolution of time-frequency image (factor of 2x2 or 3x3)
- Only for reasons of speed.

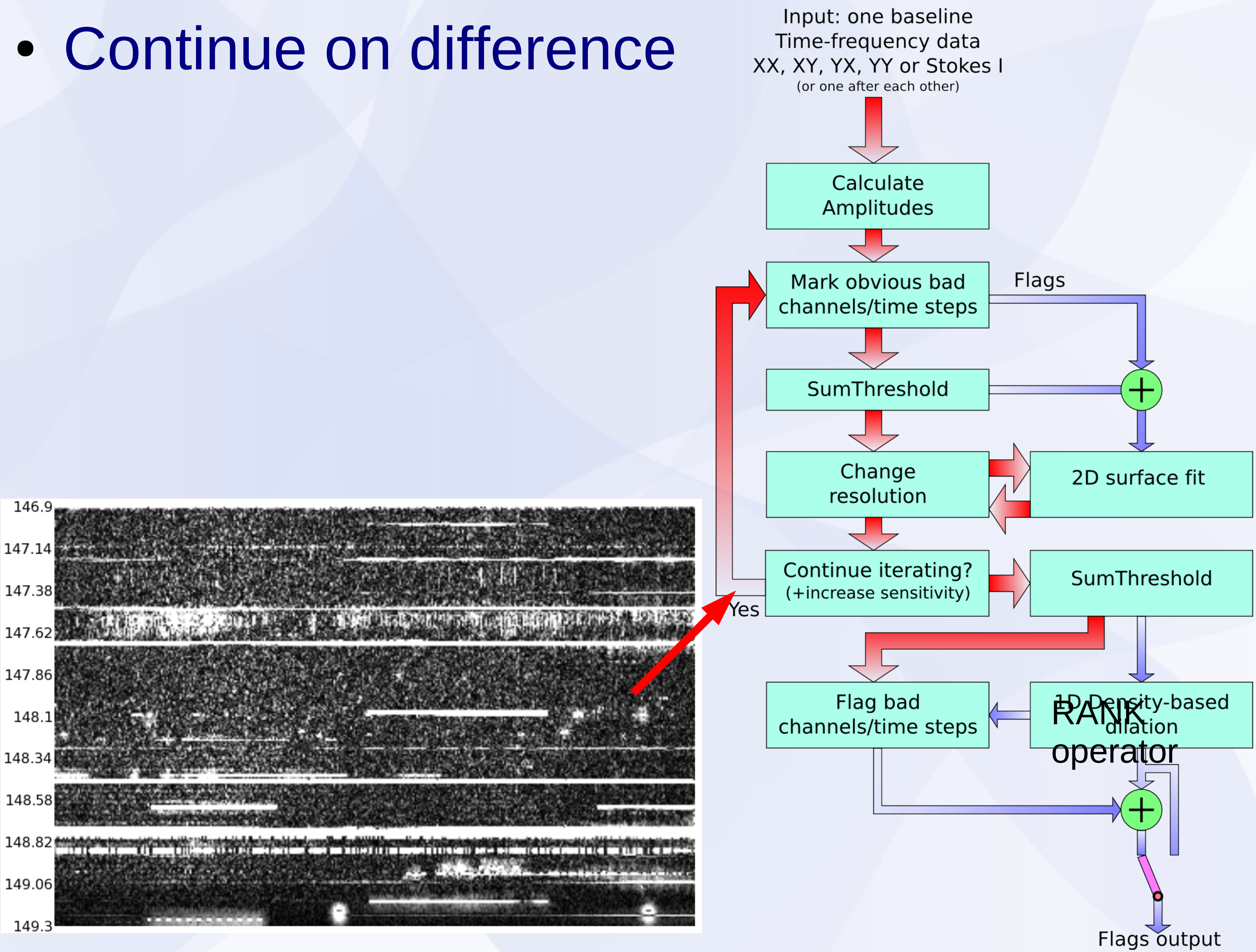


- 2D fit represents signal
- Ignore flagged data
- 2 x 1D Gaussian convolution ("Gaussian weighted local average")
- Fast

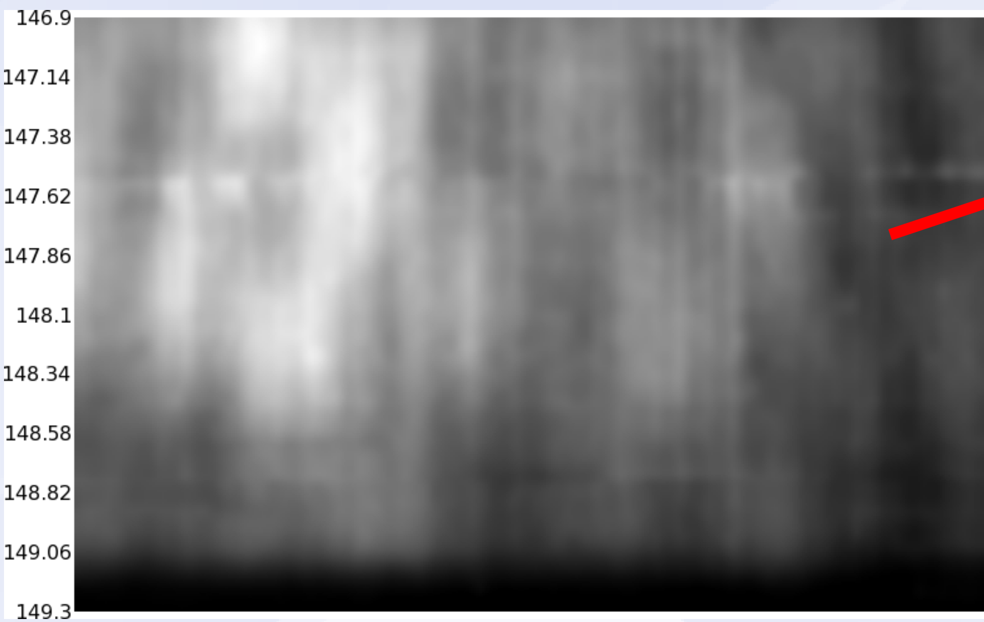


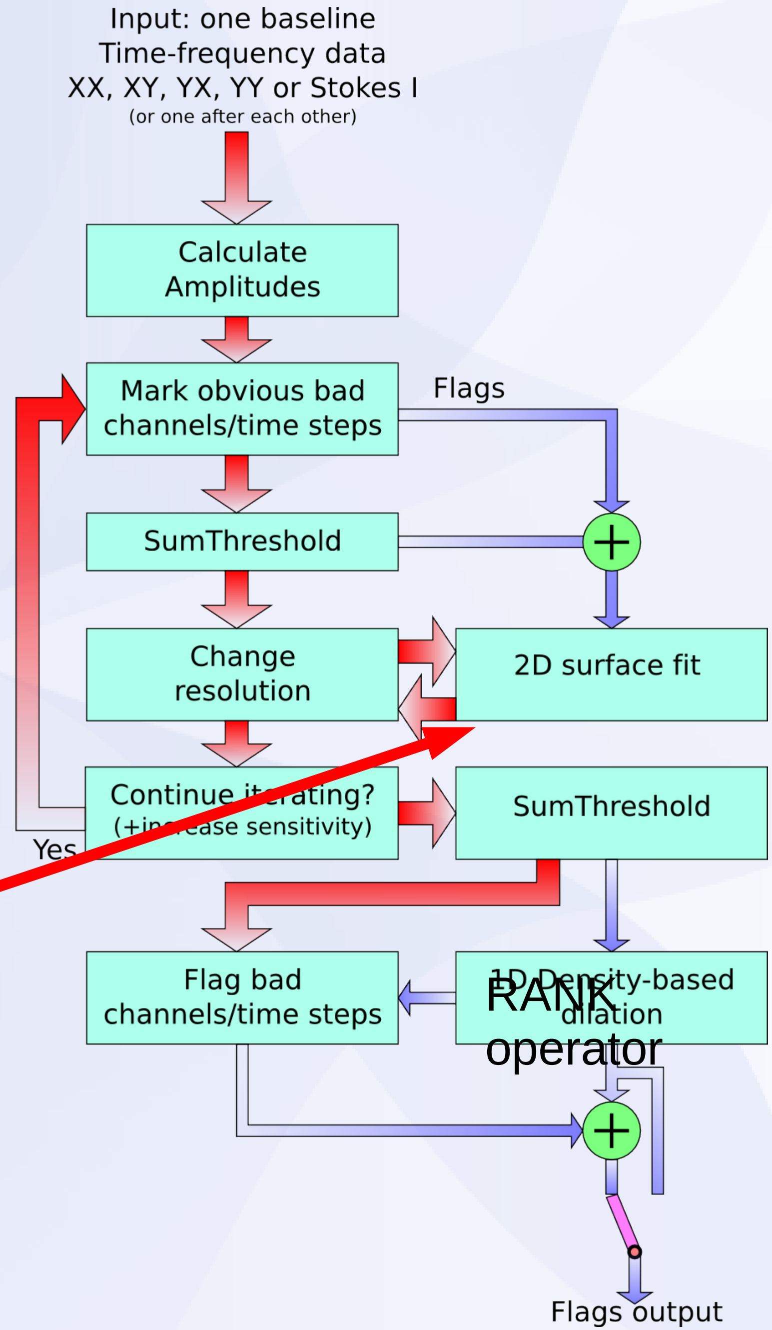


- Continue on difference



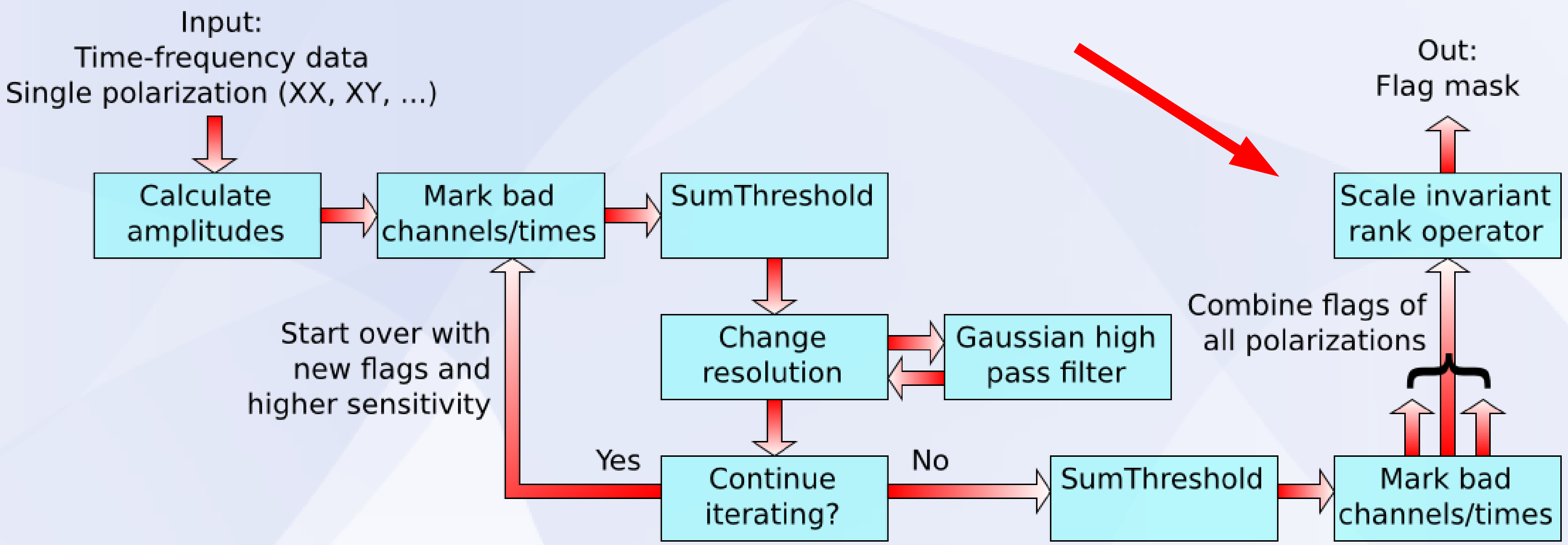
- 2^{nd} fit...


146.9
147.14
147.38
147.62
147.86
148.1
148.34
148.58
148.82
149.06
149.3

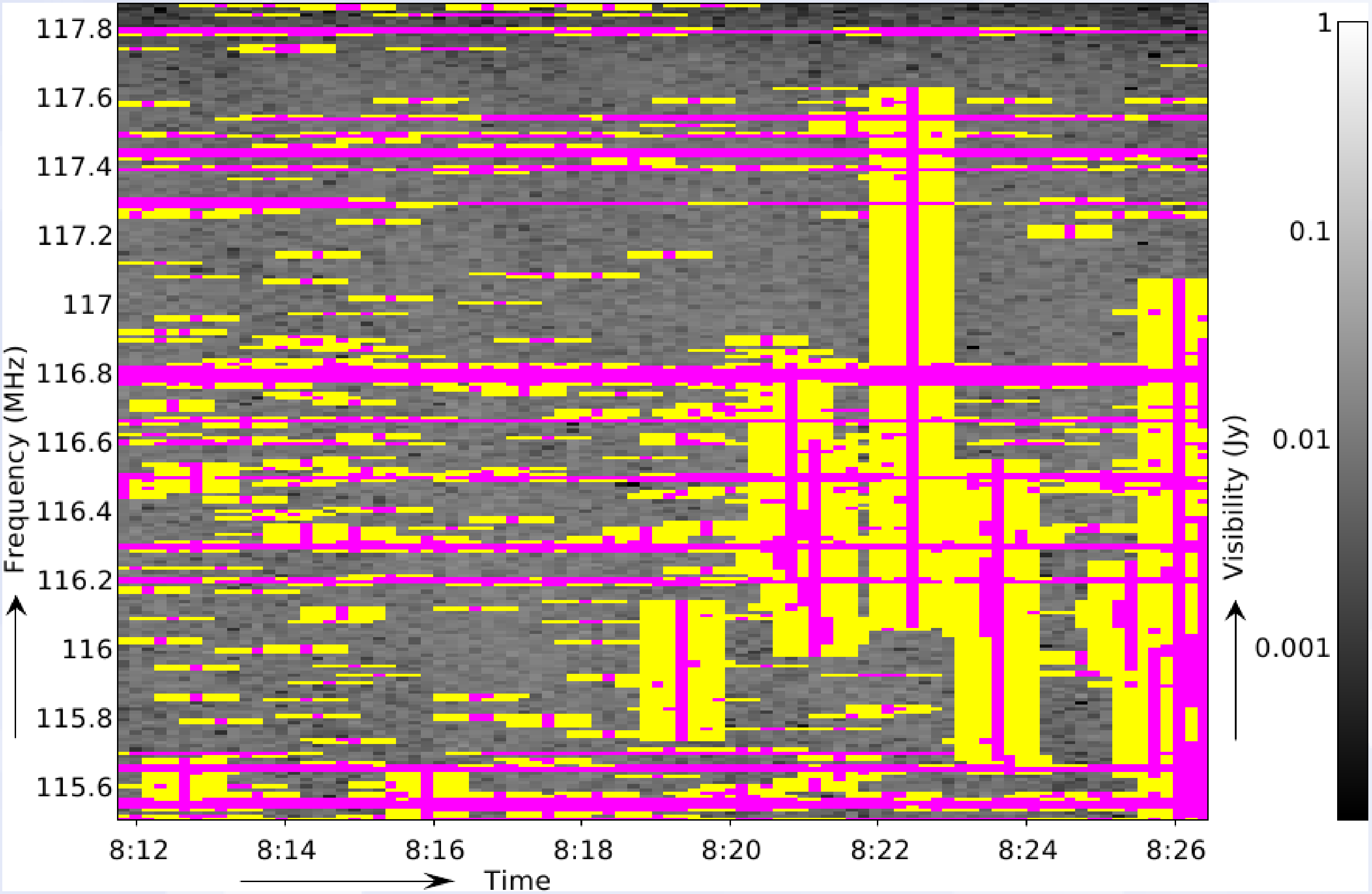

Input: one baseline
Time-frequency data
XX, XY, YX, YY or Stokes I
(or one after each other)
Calculate Amplitudes
Mark obvious bad channels/time steps
Flags
SumThreshold
Change resolution
2D surface fit
Continue iterating? (+increase sensitivity)
Yes
SumThreshold
Flag bad channels/time steps
1D Density-based dilation
RANK operator
Flags output

The scale-invariant rank operator

- No apparent RFI after final SumThreshold
- Might be some unapparent RFI around flagged regions



The SIR-operator: why?



Purple: flags produced by SumThreshold
Yellow: produced by time dilation (i.e., horizontal)

Dilation

- Dilation is “inaccurate”:
  - Flags too much on small RFI scales
  - Flags too little on large-scale RFI
- Dilation efficiency strongly depends on time/frequency resolutions
- Ideally, use a scale-invariant operator...

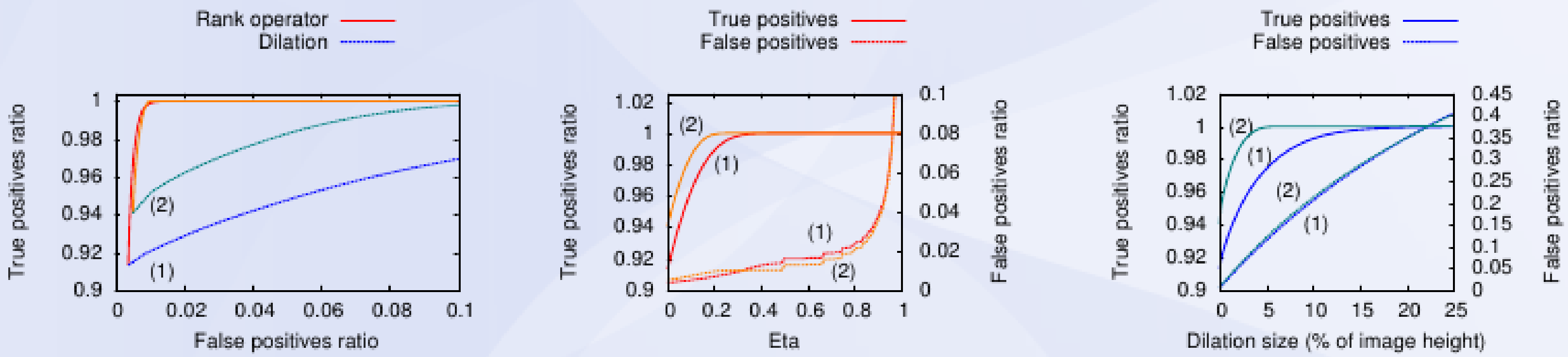
The scale-invariant rank operator

- An “improved” dilation
- Defined on a set of “flags” X:

$$\rho(X) \equiv \bigcup \{[Y1, Y2) \mid \#(X \cap [Y1, Y2)) \geq (1-\eta)(Y2 - Y1)\}$$

- Parameter eta specifies required ratio of good samples in any subsequence

The scale-invariant rank operator



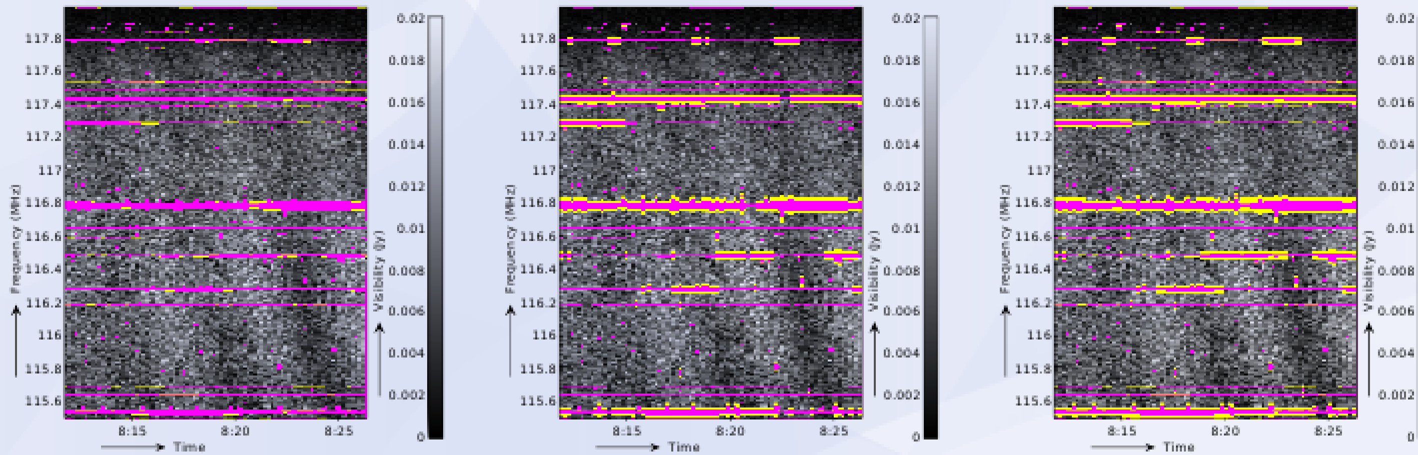
(a) ROC curves for (1) a Gaussian broadband and (2) a sinusoidal RFI feature

(b) Influence of η on the SIR operator

(c) Influence of kernel size on the dilation

Fig. 6: Analysis of the receiver operating characteristics of the SIR operator and a standard dilation on simulated data. Marks (1) and (2) correspond with respectively the Gaussian broadband feature and the sinusoidal feature. Examples of the used features are given in Figs. 2(b) and 2(e).

The scale-invariant rank operator



(a) Horizontal SIR operator (b) Vertical SIR operator (c) SIR operator in both directions

- Scale invariant
- Just submitted faster algorithm for the SIR operator

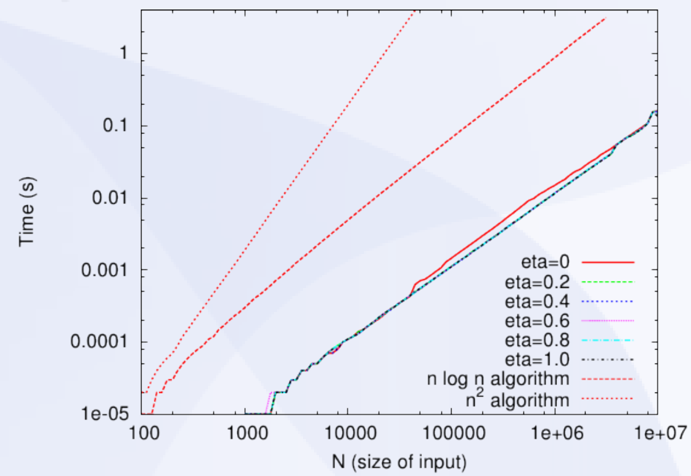
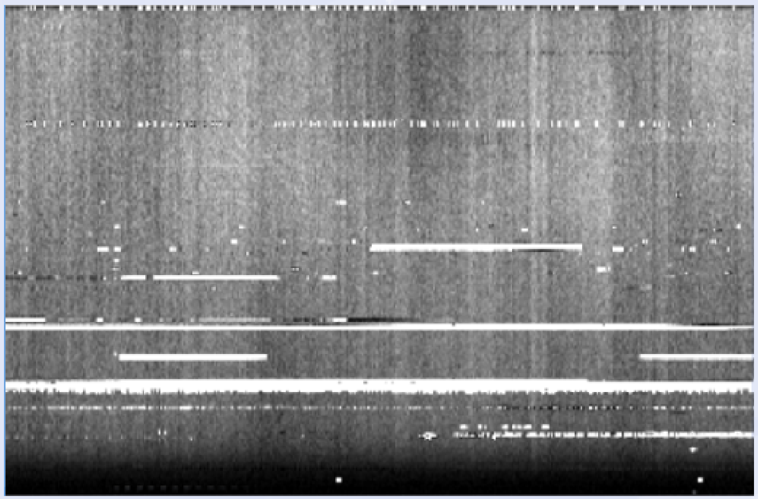
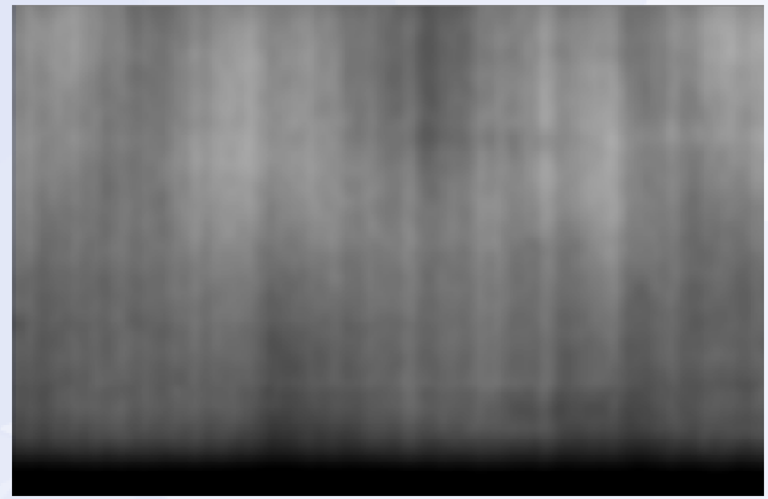


Fig. 5: Computation time versus input size with the different algorithms and fixed $\eta = 0.2$ or, for the linear case, with different η. The average over 1000 runs was taken for each different configuration.

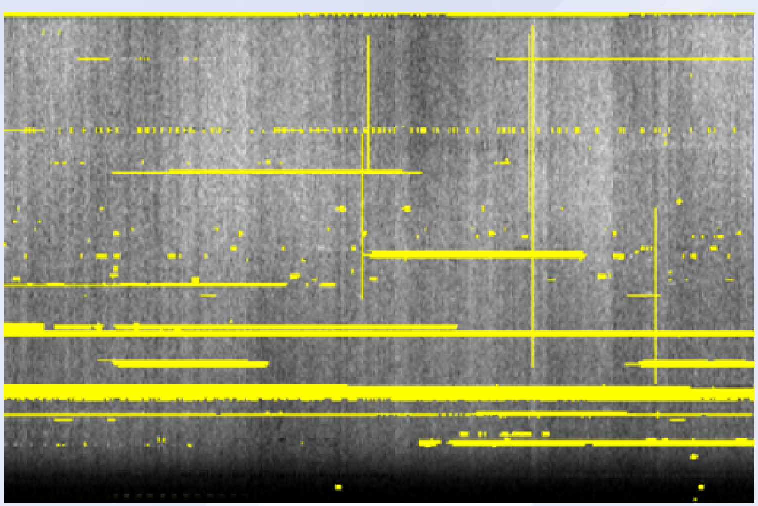
Automatic flagging example



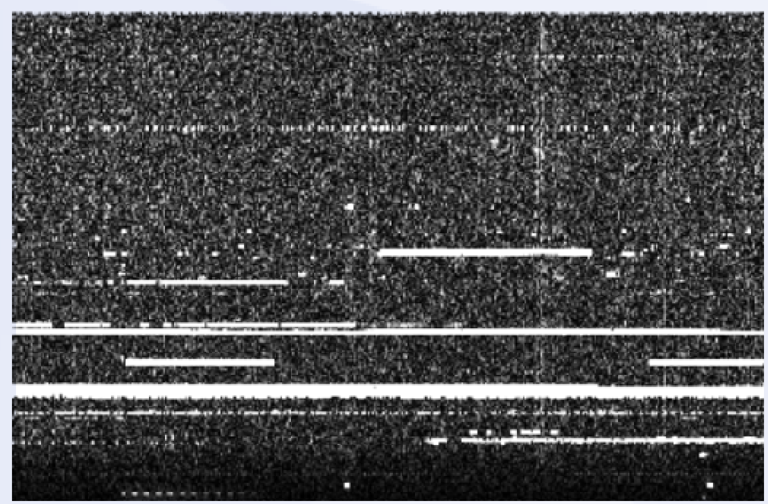
(a) Original



(b) Smoothed



(c) Automated flagging result

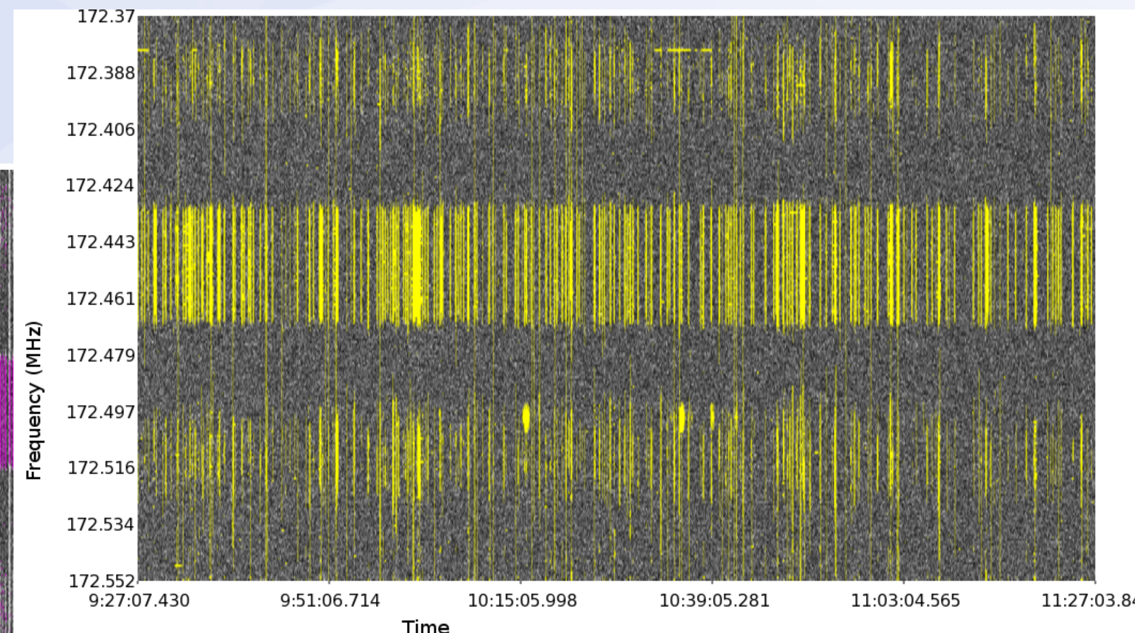


(d) Difference

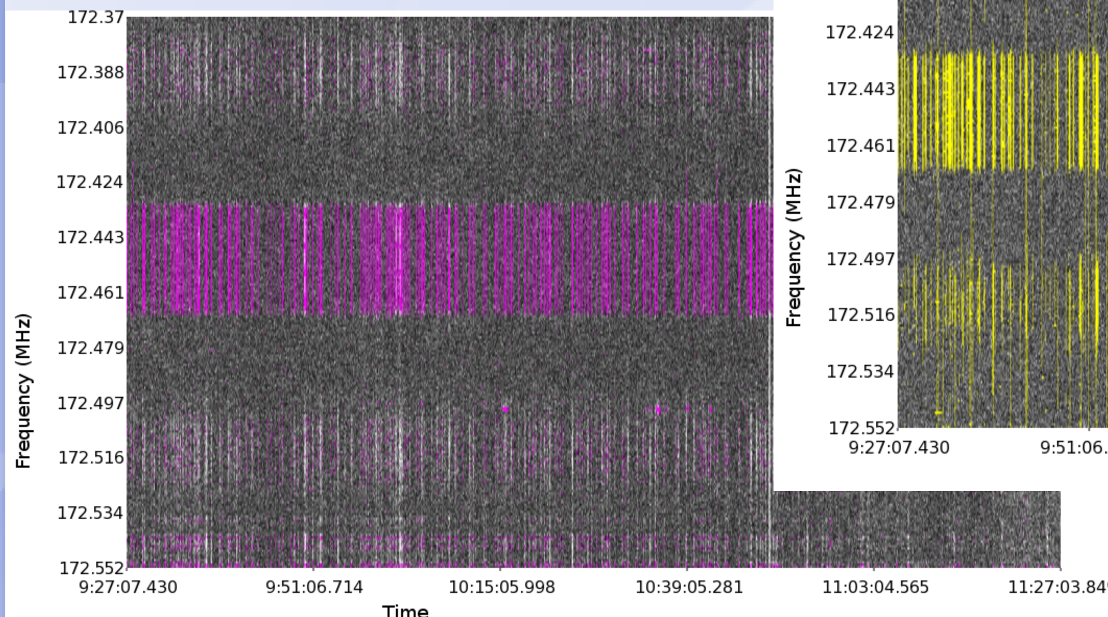
AOFlagger vs other flaggers

- Accuracy higher than other flaggers
- Fast

AOFlagger



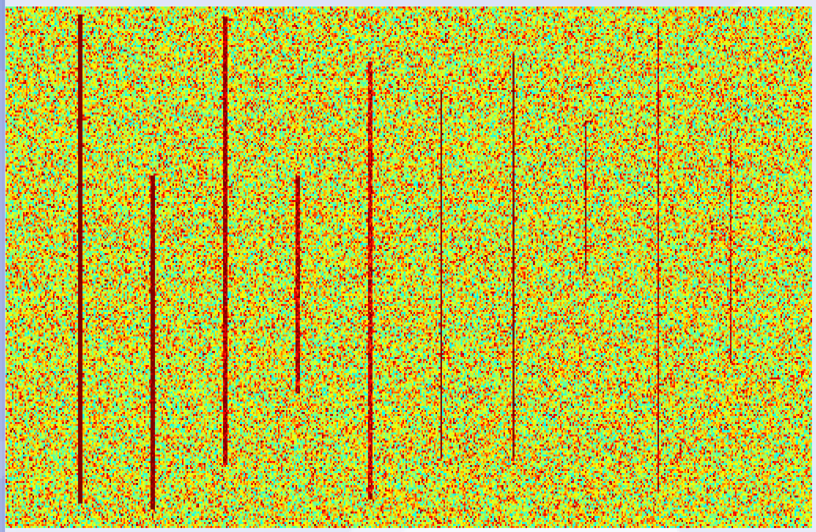
MAD flagger

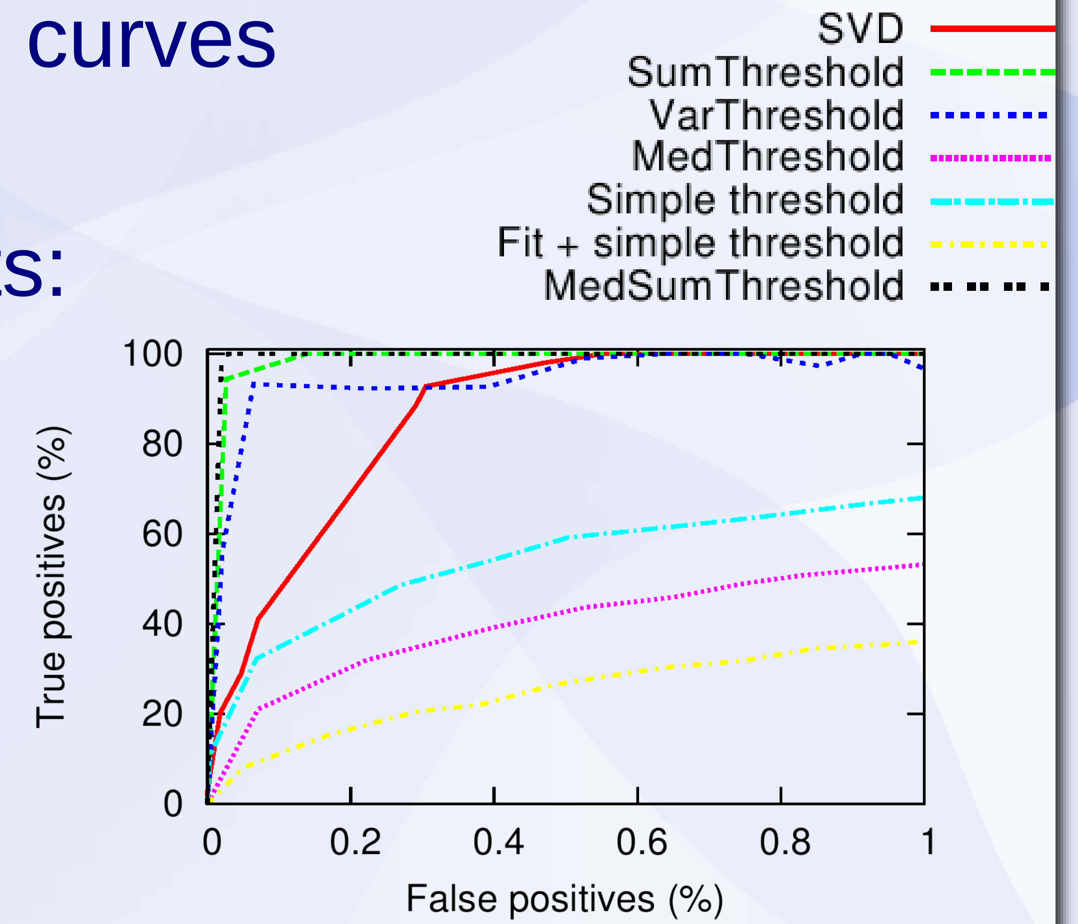


Method comparison

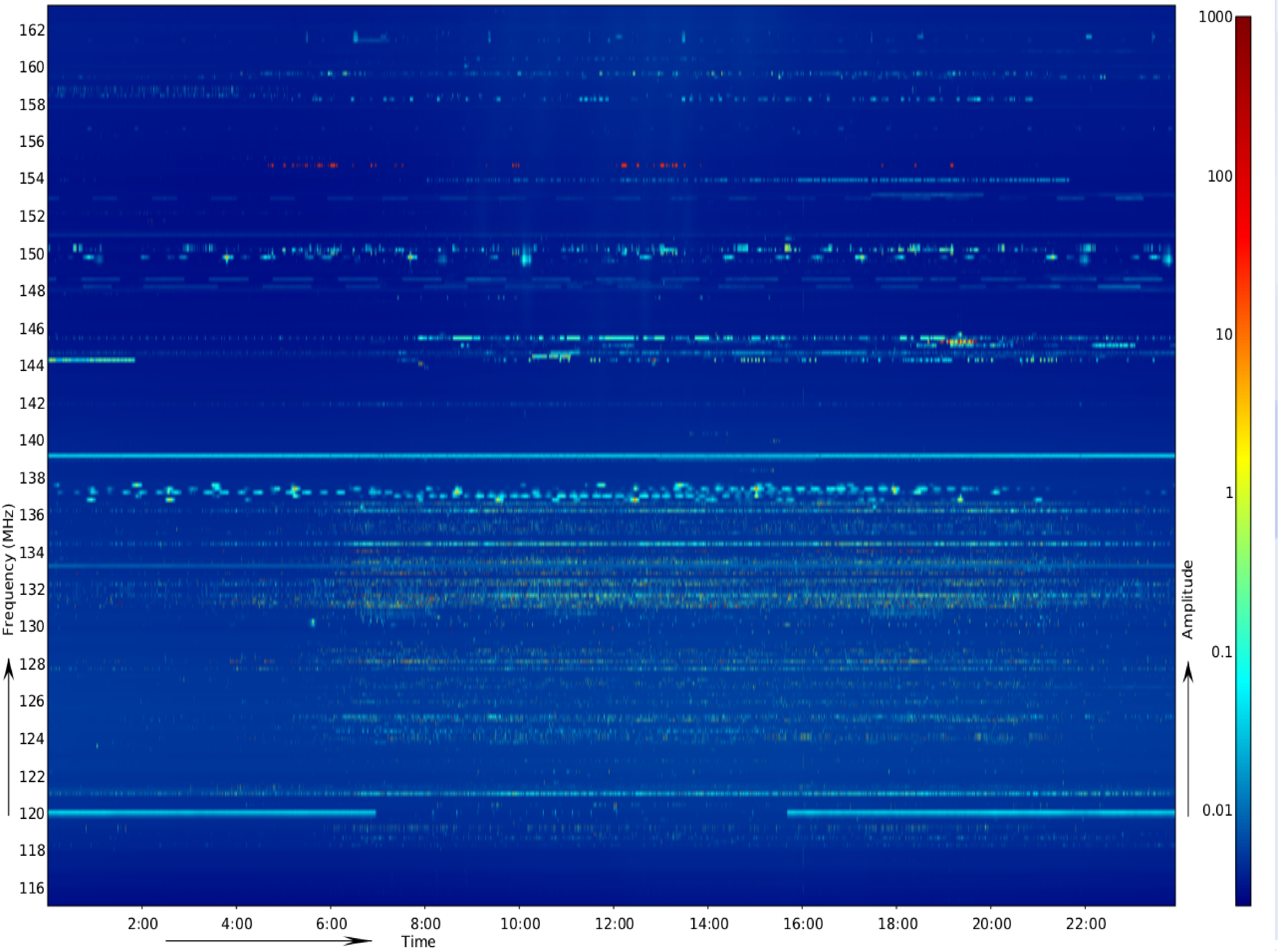
- Compare methods with the help of test sets and ROC curves

- One of the test sets:

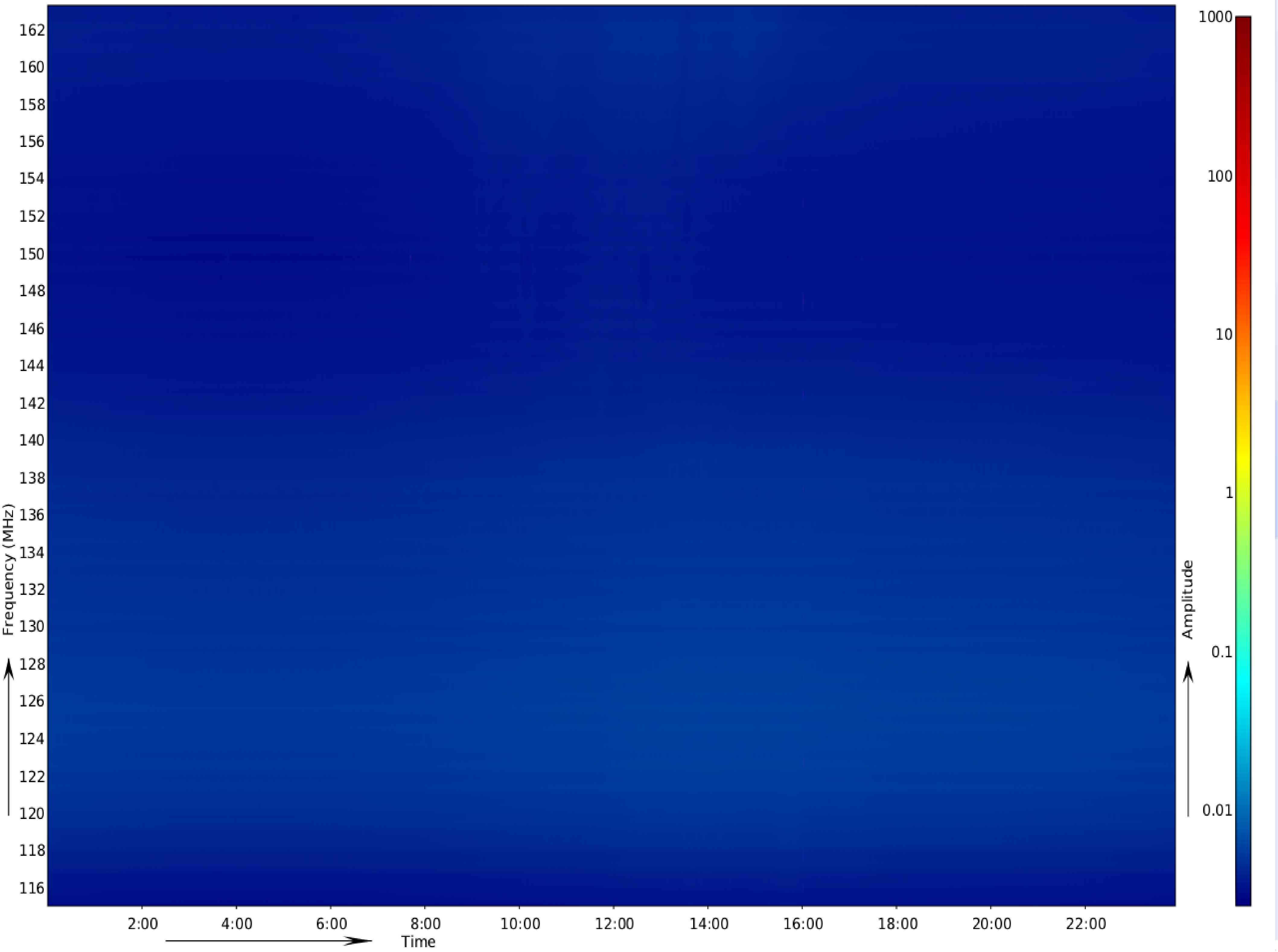


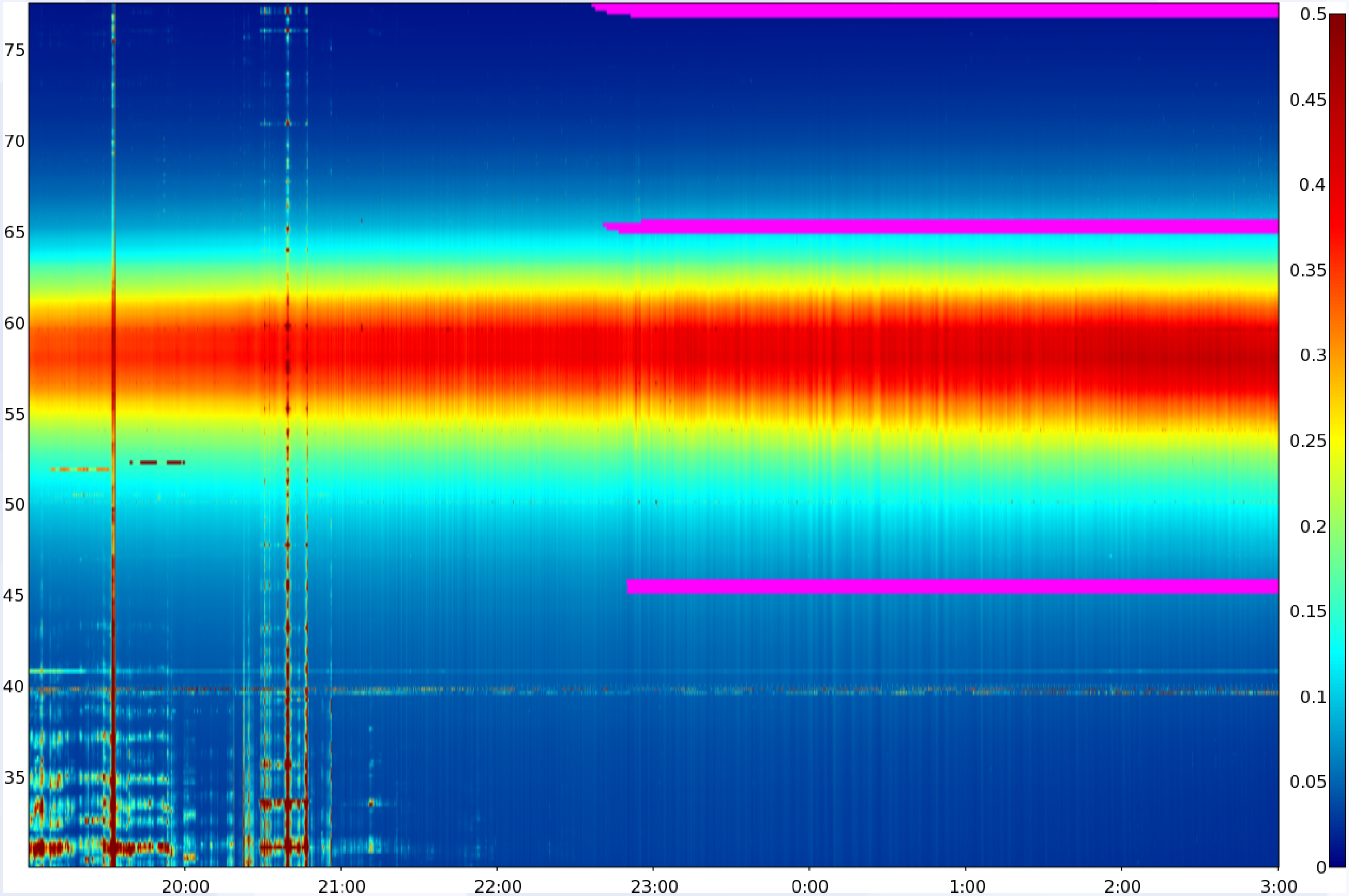


(Offringa et al., MNRAS, 2011)

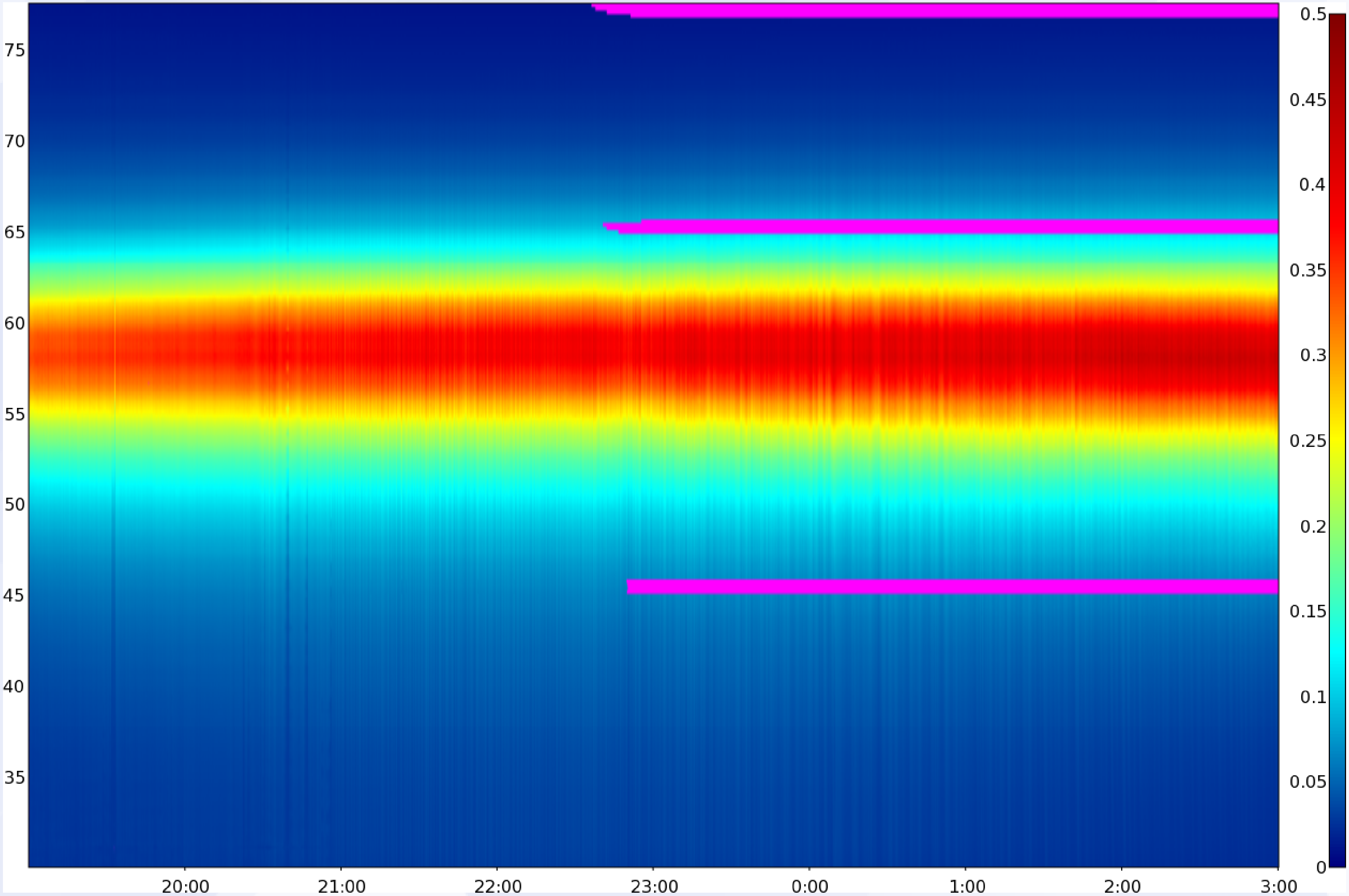


Frequency (MHz)
162
160
158
156
154
152
150
148
146
144
142
140
138
136
134
132
130
128
126
124
122
120
118
116
2:00
4:00
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
Time
1000
100
10
1
0.1
0.01
Amplitude


162
160
158
156
154
152
150
148
146
144
142
140
138
136
134
132
130
128
126
124
122
120
118
116
Frequency (MHz)
2:00
4:00
6:00
8:00
10:00
12:00
14:00
16:00
18:00
20:00
22:00
Time
1000
100
10
1
0.1
0.01
Amplitude

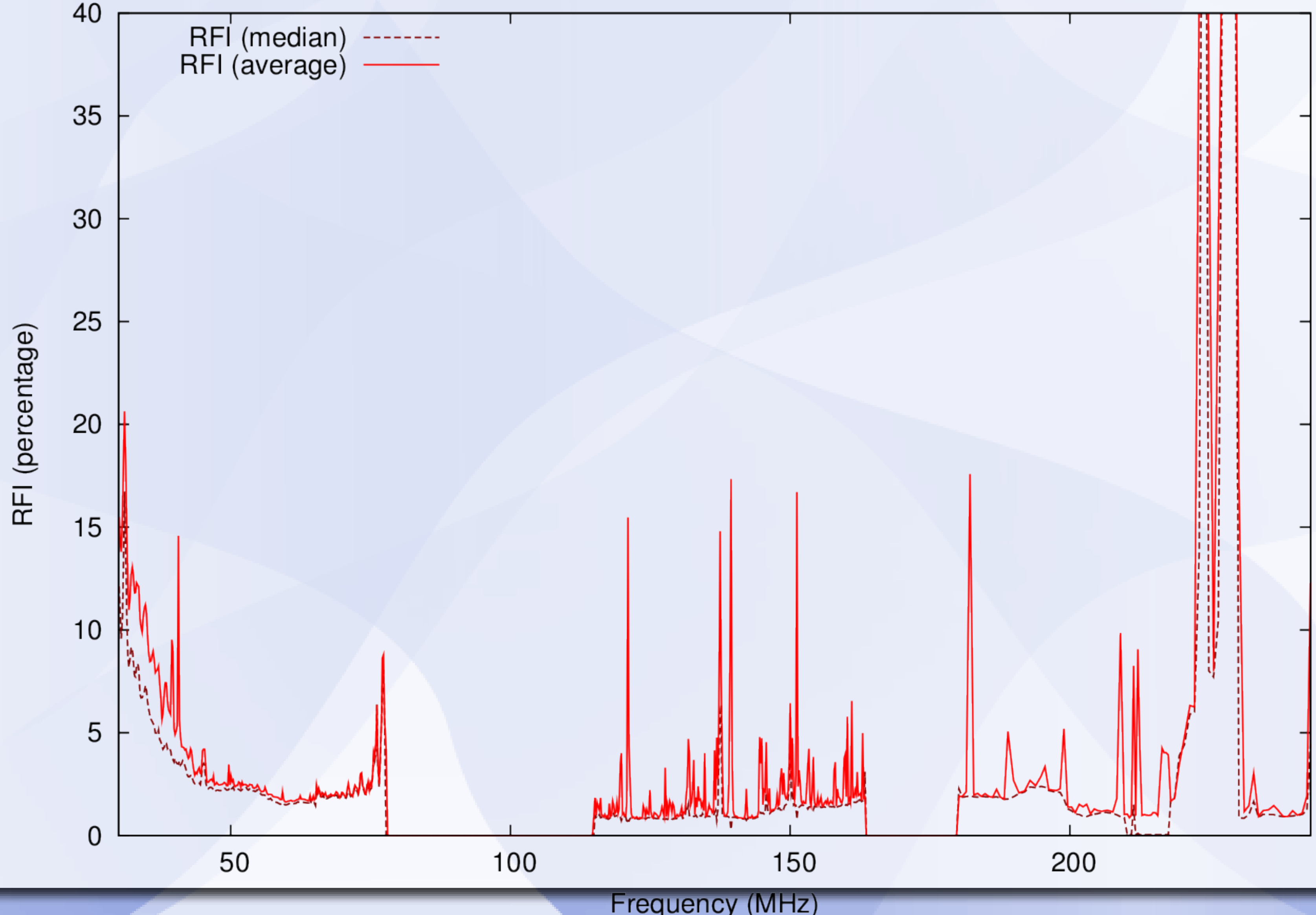


LBA Total power, before flagging



LBA Virgo, total power, after flagging

4 observations combined



LOFAR cookbook chapter 6 (Offringa et al.,)

Flagging performance

- The RFI pipeline needs to be extremely fast:
  - Executed at highest post-correlation resolution
  - On 'all' baselines (~correlations), polarizations, bands
  - LOFAR 24hr observation at high resolution (1s, 1KHz) is ~100 TiB.

Flagging performance

- The RFI pipeline needs to be extremely fast
- Optimized in several ways:
  - Multithreaded
  - Parallelized over ~60 nodes
  - Use of SSE (Streaming SIMD extensions)
  - Flagging is integrated in the next processing step to avoid multiple reads of the data
- Processing time now heavily IO dominated

Flagging performance

- Pipeline is faster than real-time
- With 3 threads (/16 cores), 64 nodes, we flagged a 90 TiB, 24 hr observation in 8 hours...
- But only reading the data already takes 20 hours!

Summary

- A lot of signal processing issues in radio astronomy
- We can automatically detect RFI in an efficient and accurate way
- Moore's law has allowed us to digitize the signal chain and increase time/frequency resolution
- Speed of hard disk did not follow Moore's Law, somewhat of an issue in my field