ASTRON

Netherlands Institute for Radio Astronomy

# Flagging, demixing & averaging

Tammo Jan Dijkema

LOFAR Data Processing School, 5 September 2016

# Outline

Steps performed on raw correlated data:

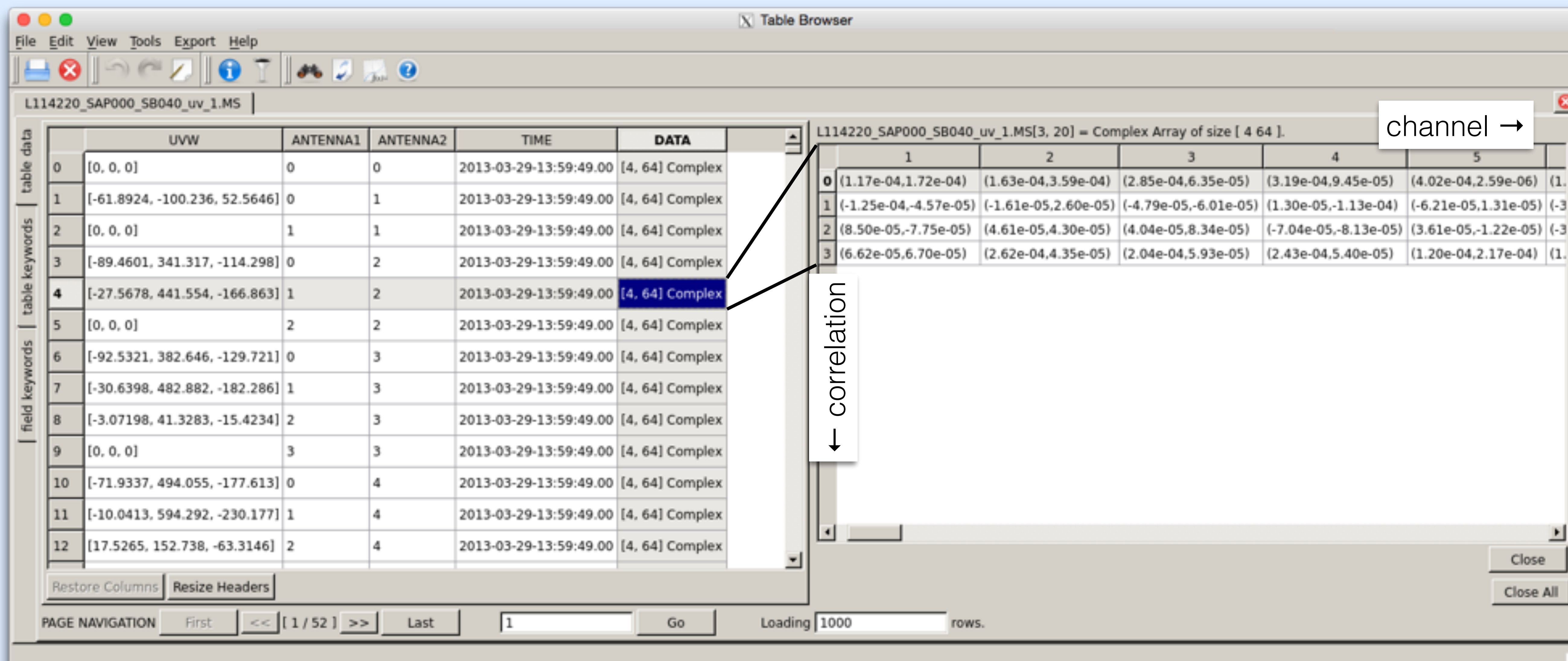
1. Add weights
2. Flag and remove RFI (radio frequency interference)
3. Remove ("demix") contribution of off-axis bright sources
4. Average / compress the data to lower time and freq. resolution

Compression is critical to decreasing processing time, which is necessary given the data volumes.

# Visibilities, output of correlator

Correlator outputs a measurement set (MS) per subband, containing visibilities: a complex number for each

- Timeslot
- Baseline (combination of two antennas)
- Channel
- Correlation (XX, XY, YX, YY)



# Overview of DPPP (or DP$^3$)

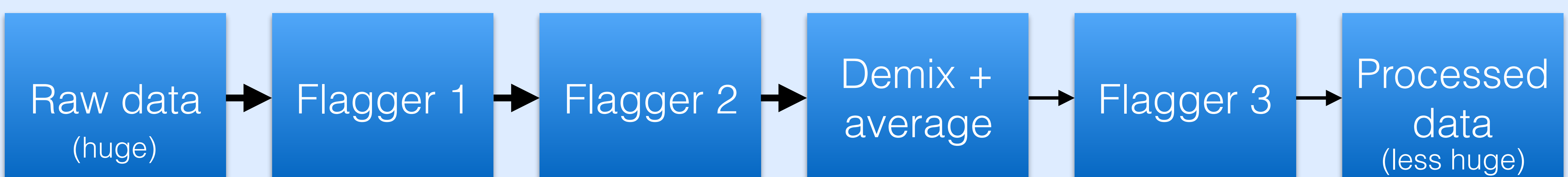
Default Preprocessing Pipeline: perform some operations on data.

It's a **pipeline**, so only read and written once. Data is piped through all steps as soon as possible: first data can be written to disk when last data has not been read yet.

DPPP is the only program that can read raw correlated LOFAR data and writes it out as the standard (CASA) MS-format.

A **step** operates on the output of the previous step.

Typical pipeline:



# DPPP: user interface

Command-line tool, input as a parset, output as feedback on screen.

```
> DPPP myreduction.parset
```

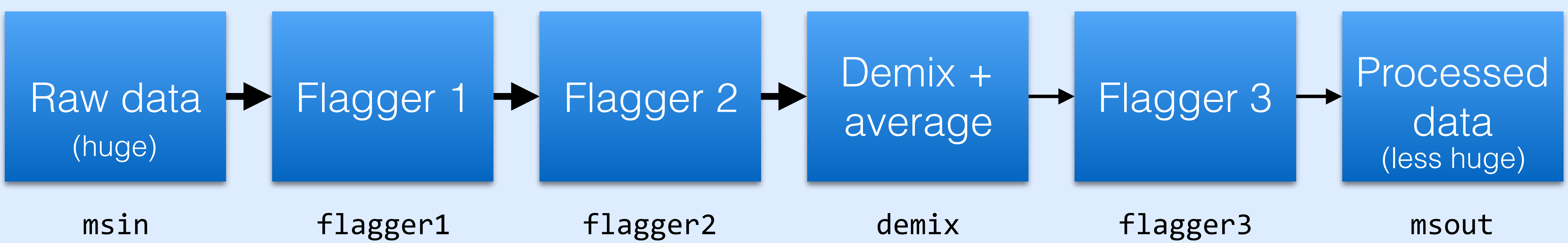
Overriding some parameters from the command line:

```
> DPPP myreduction.parset msin=L91.MS
```

Documentation:
http://www.lofar.org/wiki/doku.php?id=public:user_software:ndppp

```
msin=L123.MS
msout=L123_dppp.MS
steps=[flagger1, flagger2, demix, flagger3]

flagger1.type=aoflagger

flagger2.type=preflagger
flagger2.baseline=*&&& # autocorrelations

demix.type=demixer
demix.subractsources=[CygA,CasA]
demix.skymodel=Ateam.sourcedb

flagger3.type=aoflagger
```

Typical pipeline:



# Weights, autoweight

Associated to each visibility $v_{i,j} = \hat{v}_{i,j} + n_{i,j}$ (between station $i$ and $j$) is a weight $w_{i,j}$.

To exploit the data as much as possible, weights should be set such that noisy visibilities get down-weighted.

$$w_{i,j} = \frac{N_{\text{samples}}}{\sigma_i^2 \sigma_j^2}$$

Variance of noise of one station $\sigma_i$ estimated from autocorrelation $v_{i,i}$.

Weights are computed when DPPP reads raw data and `msin.autoweight=true`.
Autoweight should be performed only once on the data.

Weights are stored in the column `WEIGHT_SPECTRUM`.

# Good time and frequency resolution

Statement for frequency resolution (channel width):

$$\Delta\theta\Delta\nu \ll \theta_s\nu$$

For a single subband of LOFAR data (195 kHz), bandwidth smearing generally does not produce a large effect when imaging sources in primary beam.
However: bright sources outside the primary beam are affected!

Statement about time resolution ($P \approx$ 24 hours):

$$\Delta\theta\Delta t \ll \frac{\theta_s P}{2\pi} \approx \theta_s \times 1.37 \times 10^4 \text{ s}$$

Actual limit on time scale is the time behavior of ionosphere: 5–10 sec.

For efficient processing, time and frequency resolution should be **low**!
Raw data is taken at **high resolution**, for flagging and demixing.

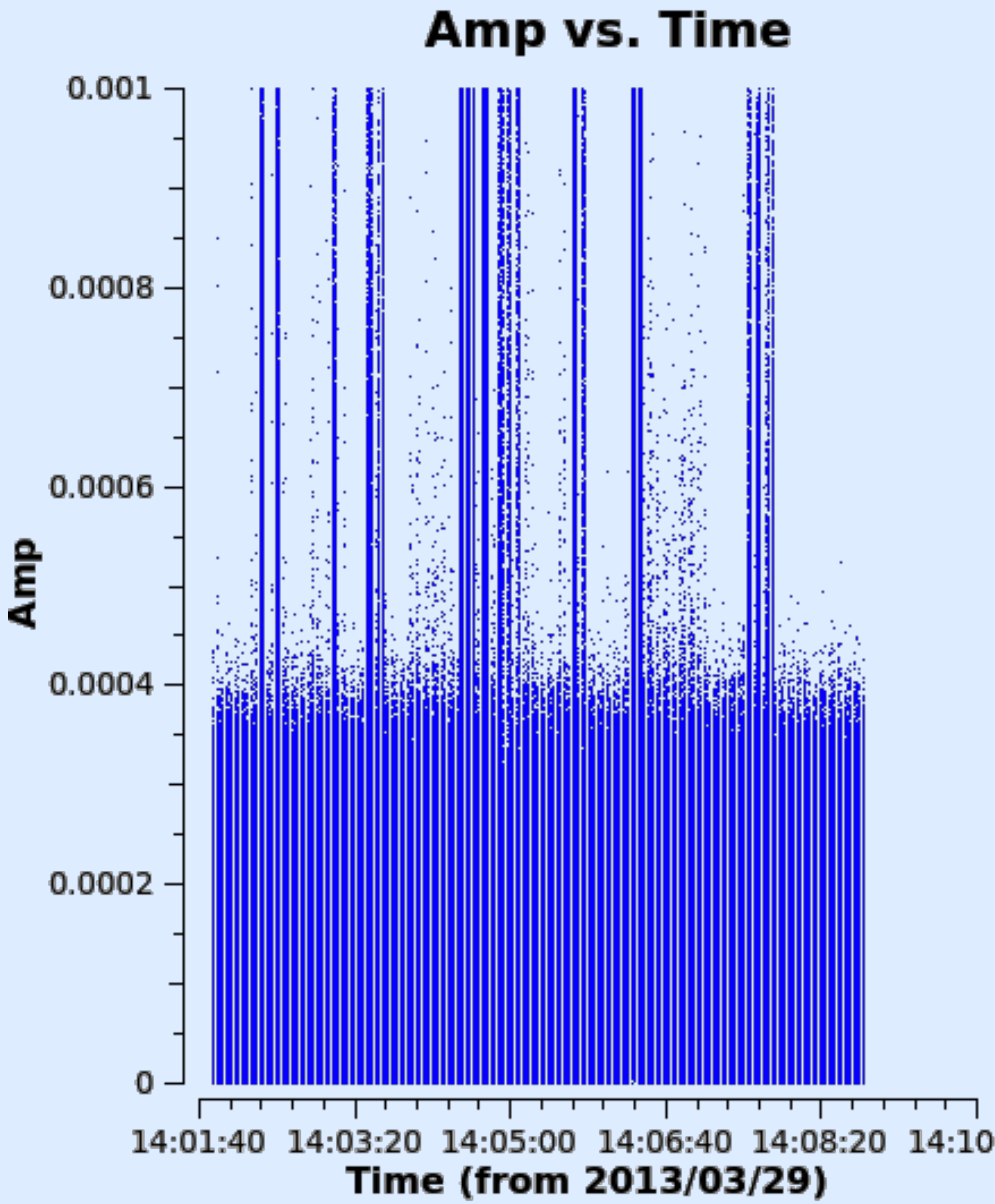
# Flagging

Some samples affected by radio frequency interference (RFI).

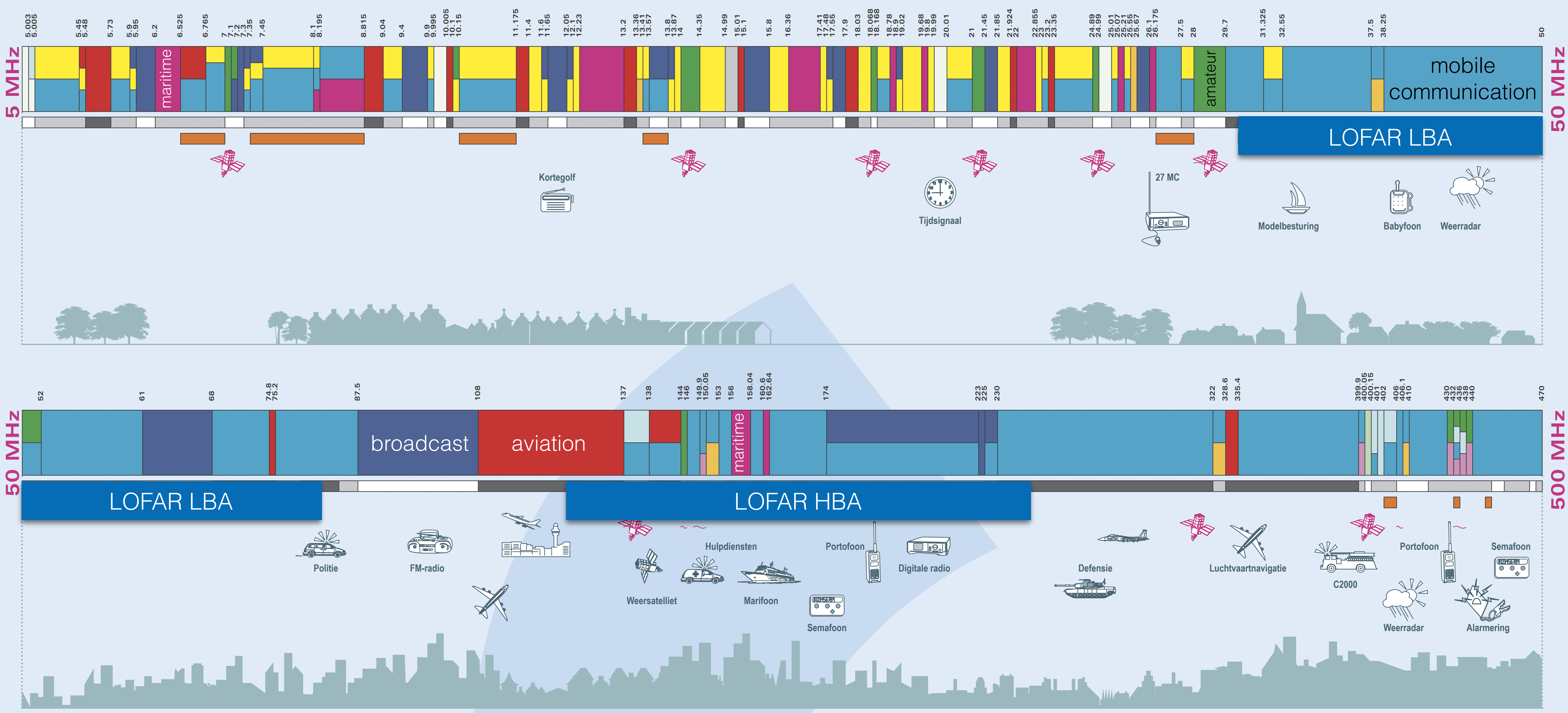
This makes these samples unsuitable for further processing.

We will flag them, and pretend they were never there.

Data is not deleted, just a check is put in `FLAG` column in MS.



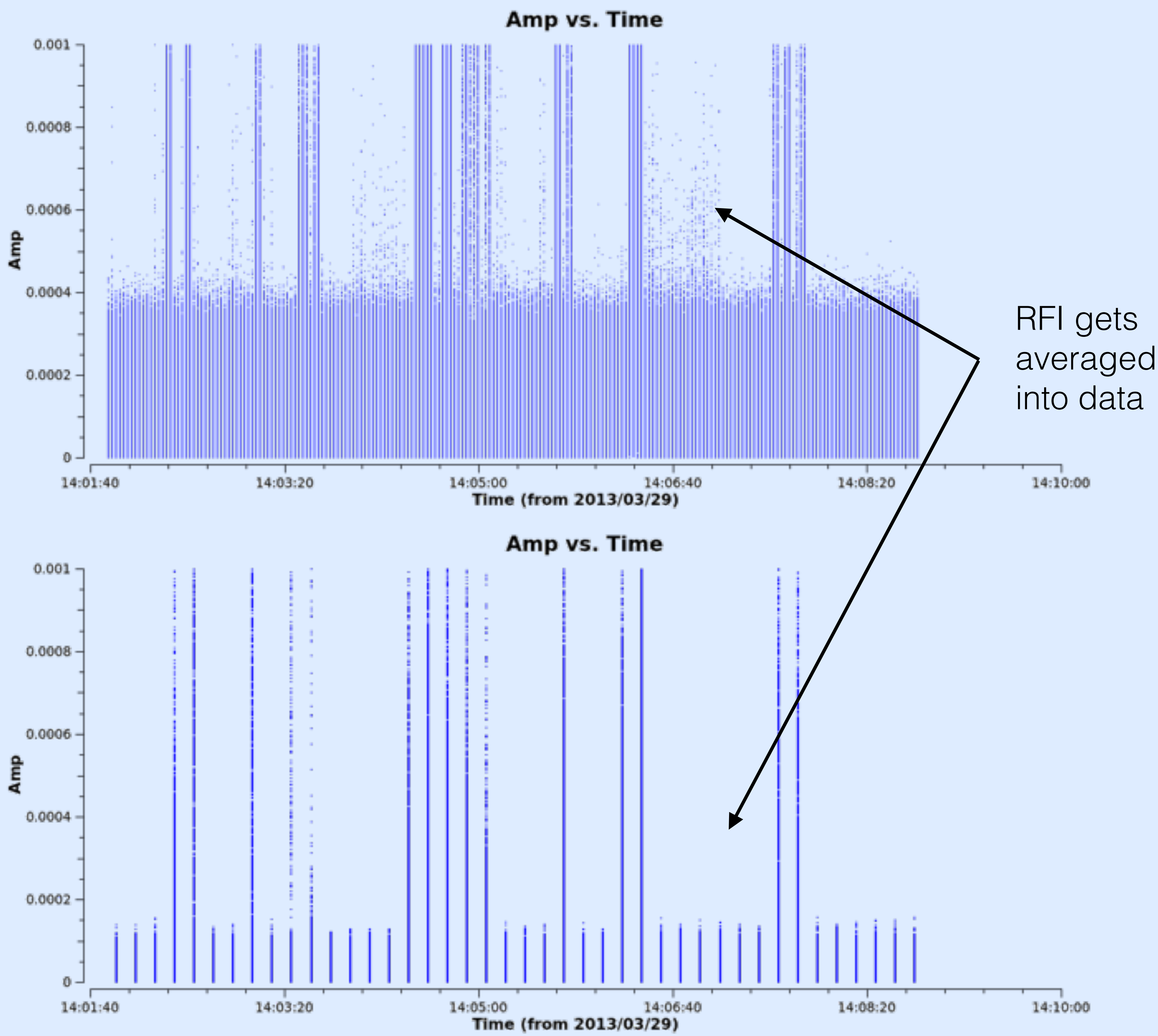
# Frequency allocations, interfering signals




Source: frequentiespectrumkaart 2005

Most RFI near LOFAR is narrowband and/or short duration.

# Data should be flagged at high resolution



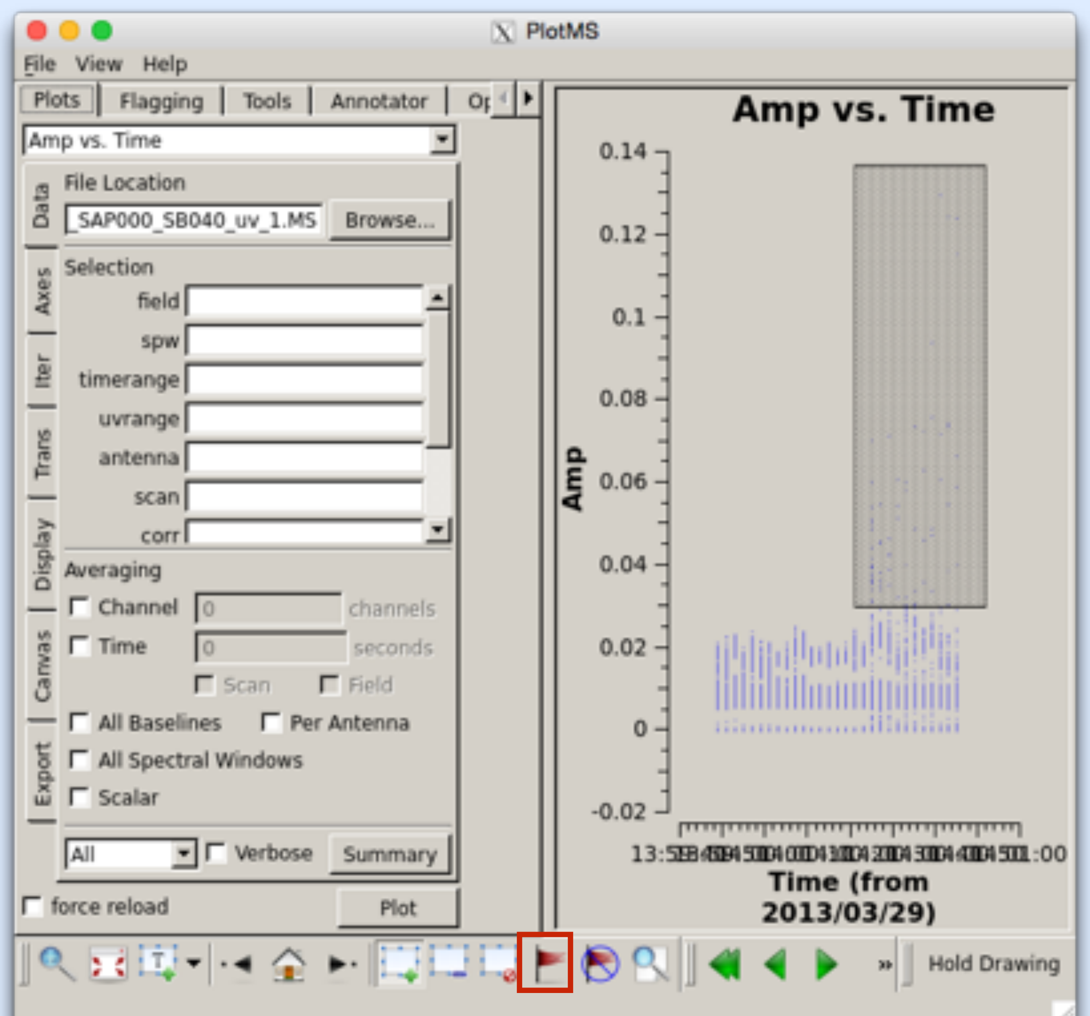
# Three methods of flagging

ASTRON

## 1. Manual Flagging

Inspect data, select visibilities to flag

Can be done with `casaplotms`




## 2. Semi-automatic flagging

For example:

- Flag all autocorrelations
- Flag all signal stronger than 100 Jy
- Flag the first channel

Can be done with `DPPP`, step `preflagger`

## 3. Automatic flagging

For example using AOFlagger

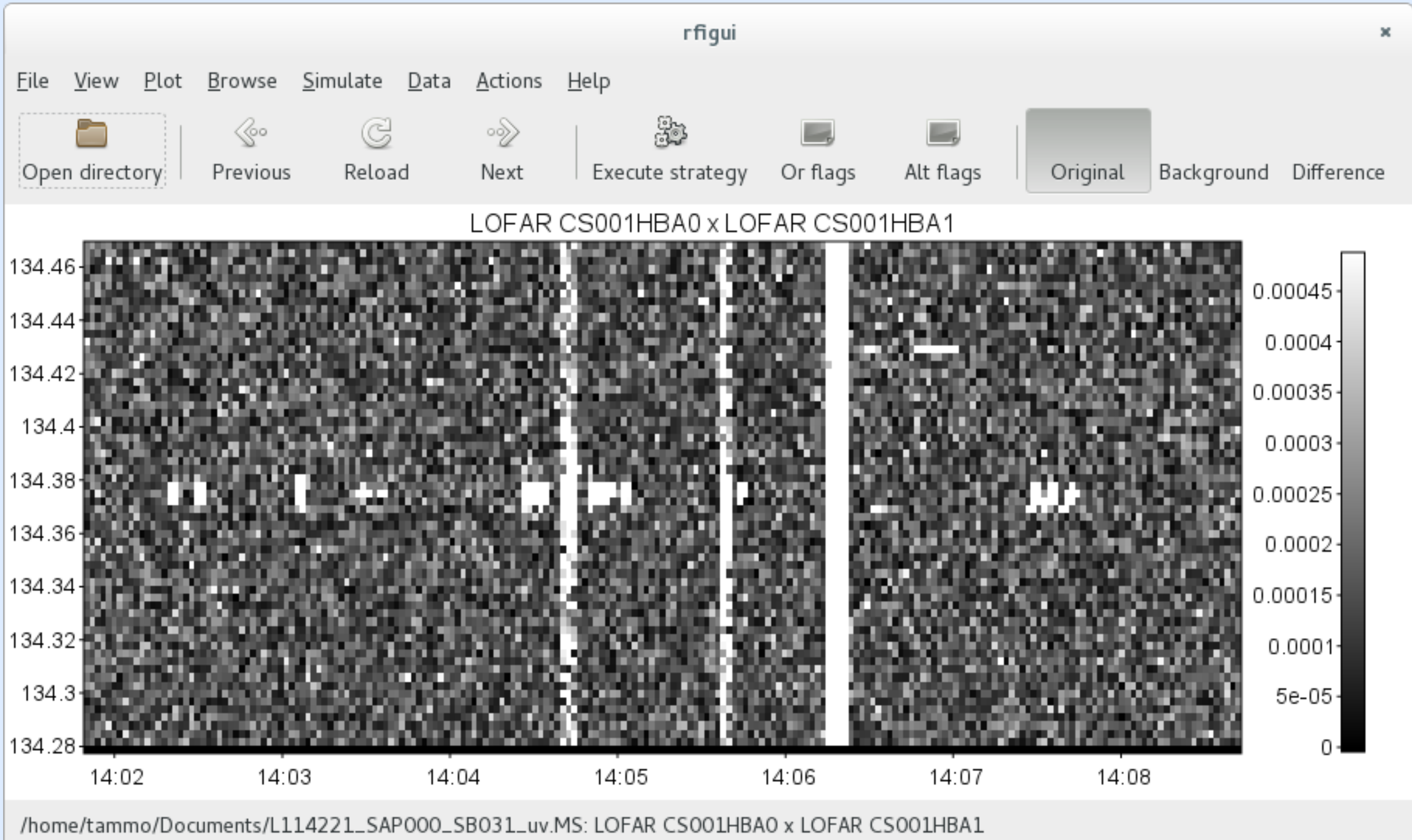
Flags based on time-freq statistics (per bl).
Performs best on long time ranges!

Can be called from `DPPP`, step `aoflagger`

Interactive counterpart: `rfigui / aoflagger`

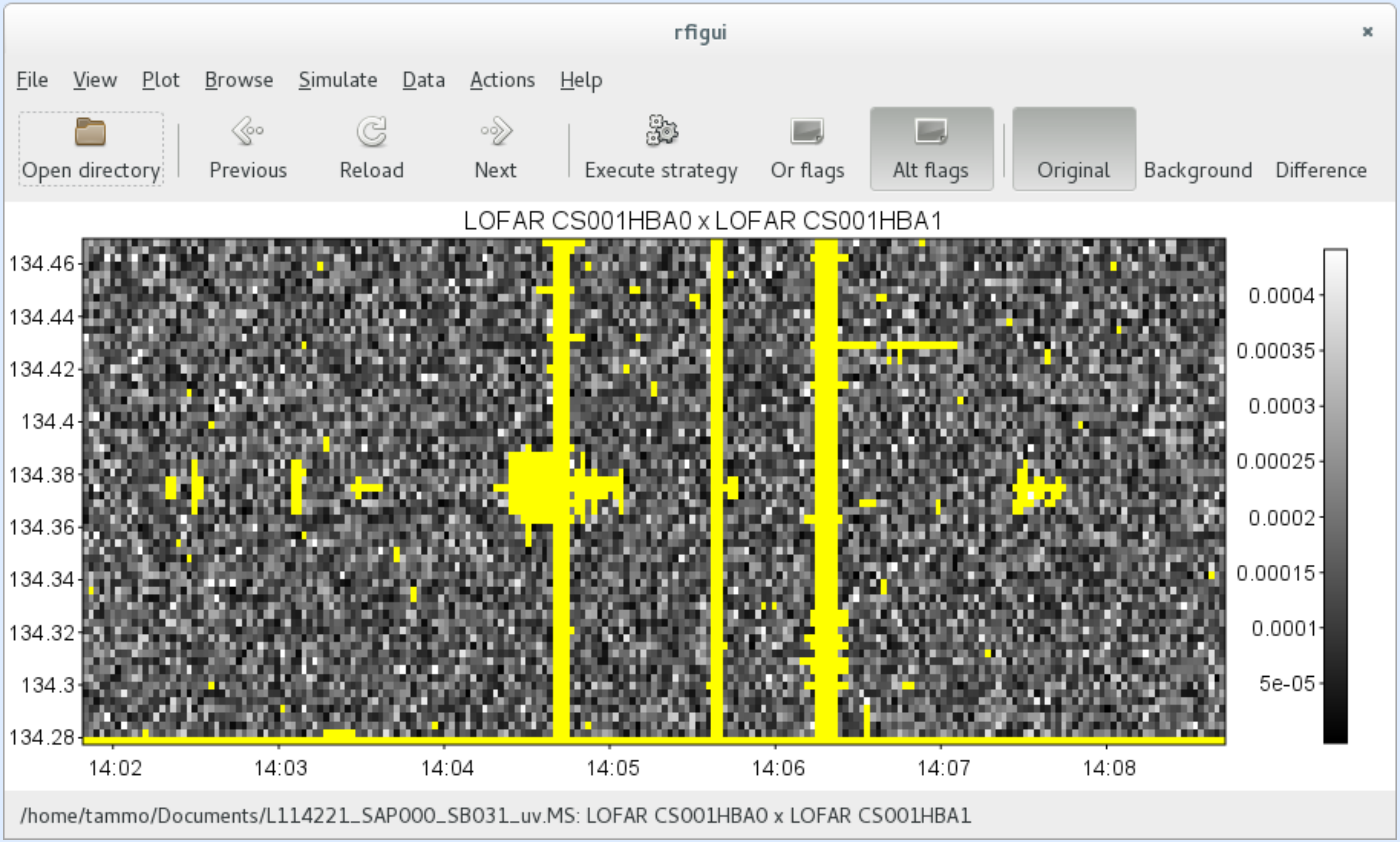
# AOFlagger, rfigui

AOFlagger (André Offringa) flags data based on statistics:



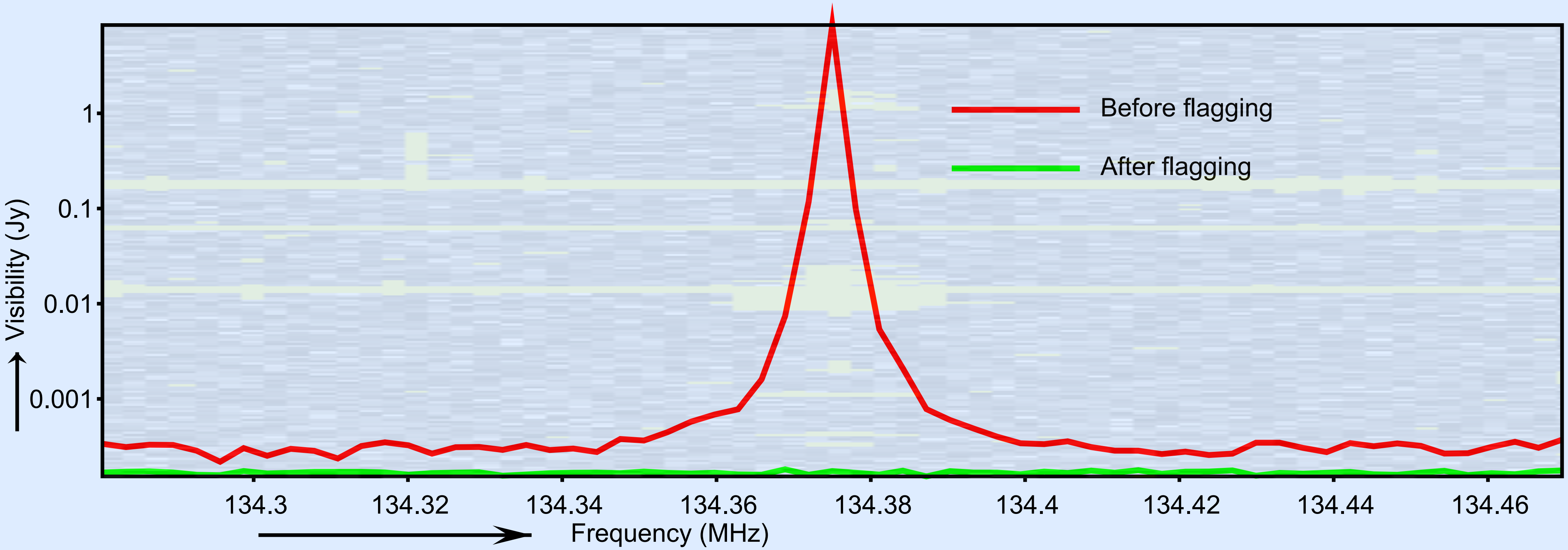
# AOFlagger, rfigui

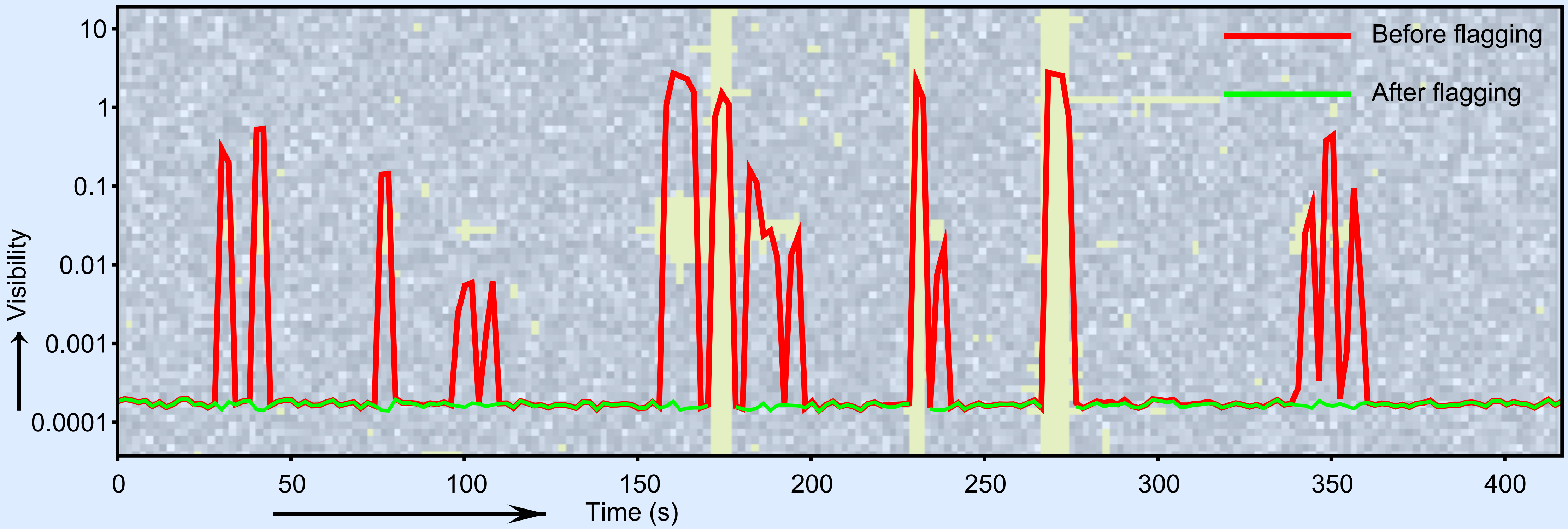
AOFlagger (André Offringa) flags data based on statistics:



# AOFlagger, rfigui

AOFlagger (André Offringa) flags data based on statistics:



# A flagging strategy

1. Flag data with preflagger, flag misbehaving stations.
2. AOFlagger on data on high resolution.
3. Inspect results, maybe some manual flagging.
4. Demix and average data.
5. Run AOFlagger on averaged data.
6. (optional) Inspect results before calibration
7. Calibrate
8. (optional) Run AOFlagger again

# Direction-independent calibration

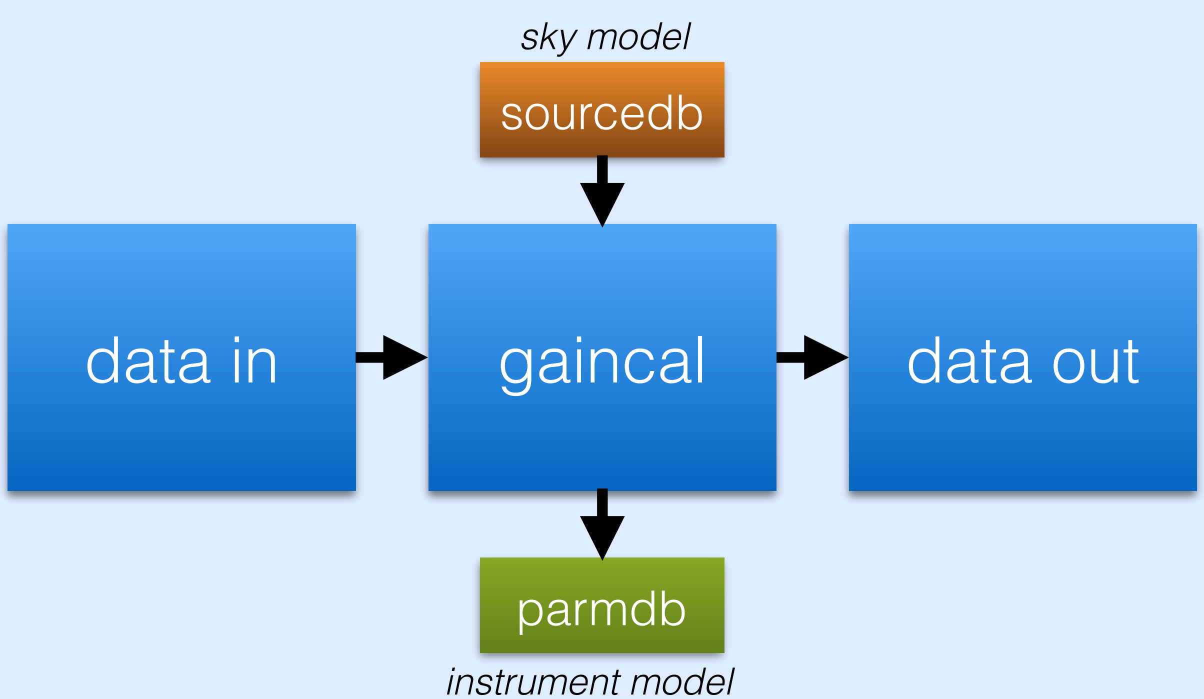
ASTRON

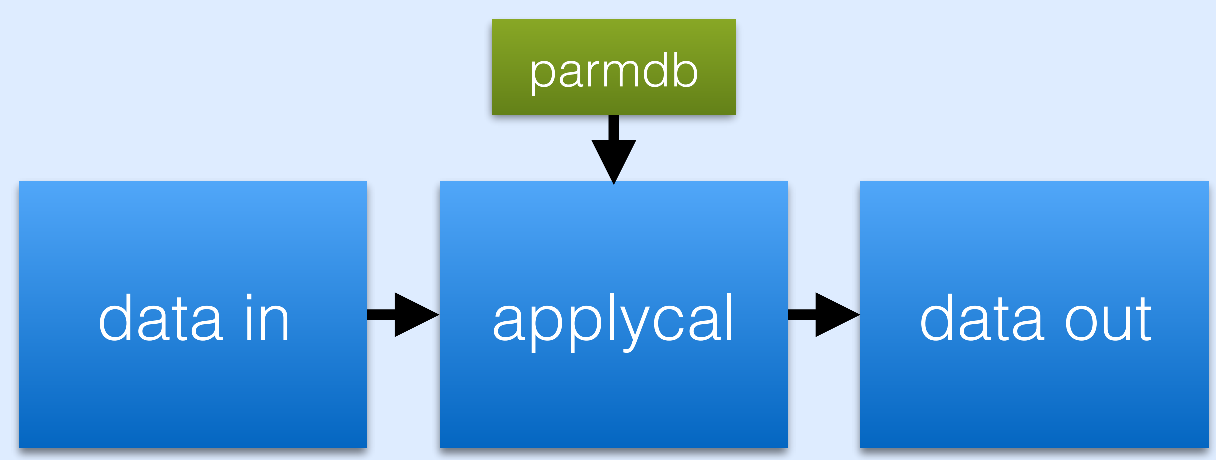
*Real presentation tomorrow by Ger de Bruyn.*

The signal you have measured has been altered by ‘unwanted’ station-, time and frequency dependent effects.

These effects are not known accurately beforehand, so let’s fit them afterwards, by fitting the data to a model sky.

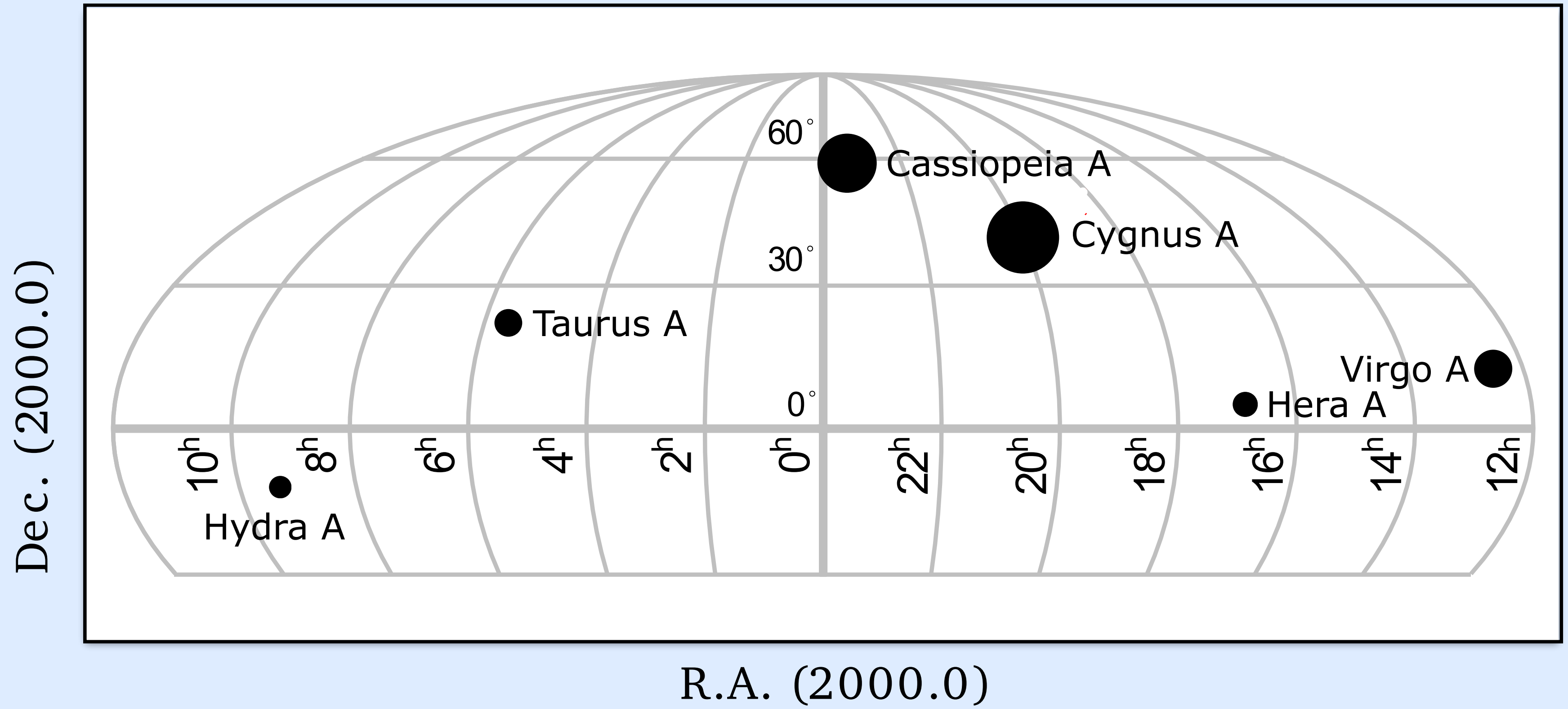
Direction independent calibration in DPPP:



# Demixing

Sky at low frequencies is dominated by a few sources, together called **A-team** sources.



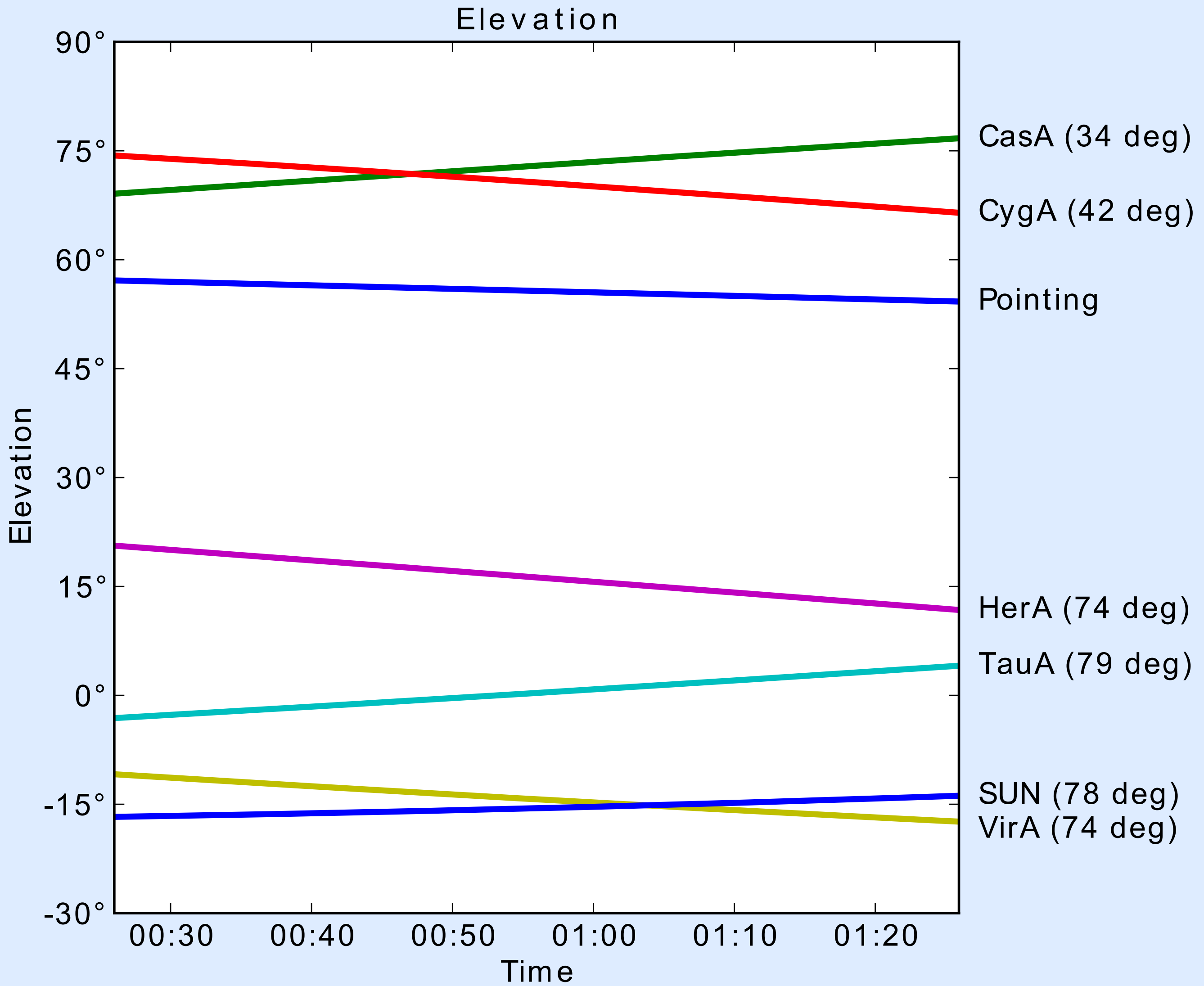

If A-team source is affecting signal, its signal needs to be subtracted. To subtract, the data must be calibrated against a model of the A-team source.

Time and frequency resolution needs to be such that signal from A-team sources is not too much affected by time and frequency smearing.

# Demixing: is your data affected by A-team?

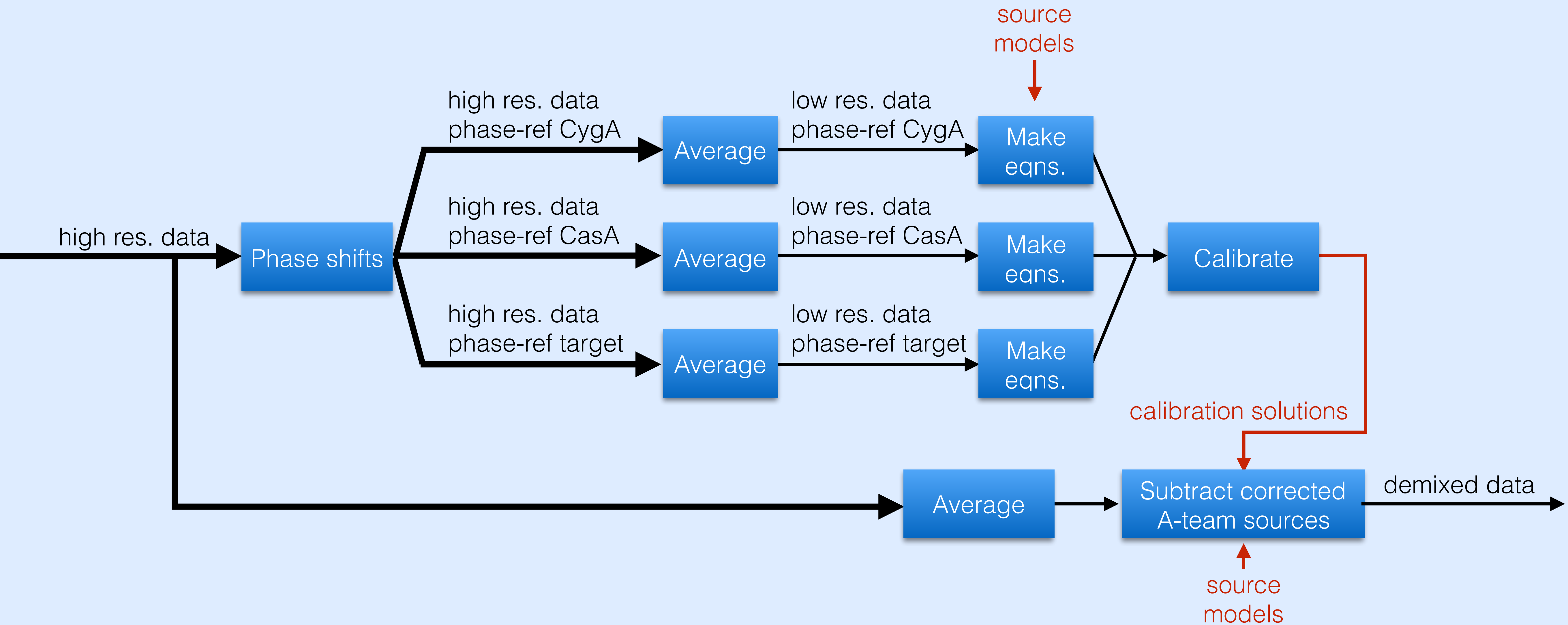
**LBA**: yes, your data is affected by CygA and CasA and perhaps more

**HBA**: your data might be affected by A-team:
- If target is within 30° separation of A-team source
- If A-team elevation is high during observation



# How demixing works

- Demixing: subtracting calibrated model visibilities of bright sources
- Calibration is expensive: should be done on averaged data.
- Data can only be averaged near phase center.
- Idea: phase shift (high-res) data to the bright source, then average.



# Conclusion

Steps performed on raw correlated data:

1. Add weights
2. Flag and remove RFI (radio frequency interference)
3. Remove ("demix") contribution of off-axis bright sources
4. Average / compress the data to lower time and freq. resolution

Compression is critical to decreasing processing time, which is necessary given the data volumes.

Details can be found in the cookbook and DPPP documentation.