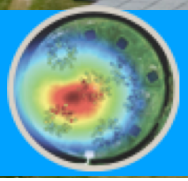
ASTRON
Netherlands Institute for Radio Astronomy

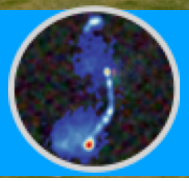
STELLAR

LOFAR

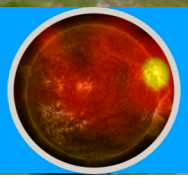
Raw data quality assessment

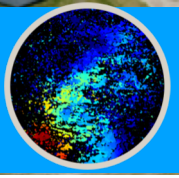
Tom Franzen | ASTRON





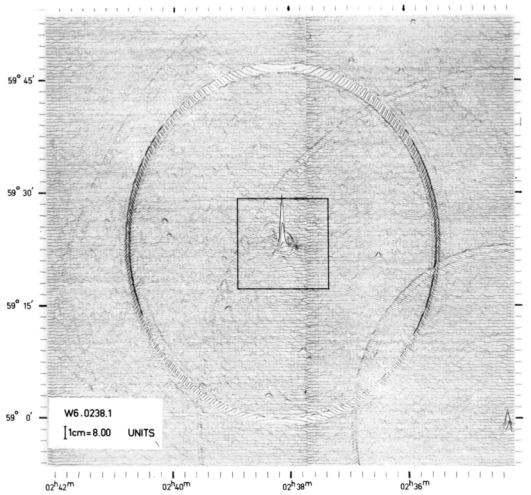
6th LOFAR Data School

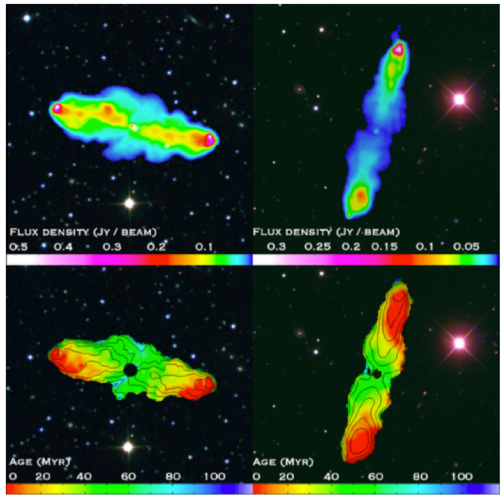




What is data quality?

- Data quality checks involve the assessment of a data set (which in our case is astronomical) to determine how well it meets requirements
- LOFAR/astronomy case: how affected is the data set by anything that could potentially impact the science goals in a negative way?
- Data sets in astronomy have become incredibly complex in the last decade: size, baselines, bandwidth, sensitivity, spectral/time resolution
- We conduct data quality assessments as part of our role in SDC Operations – **but we need users too!**



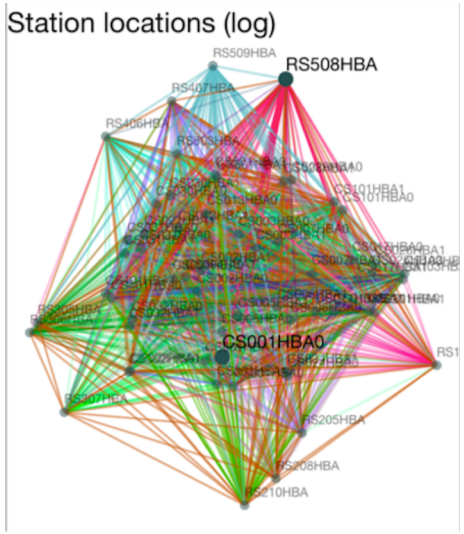


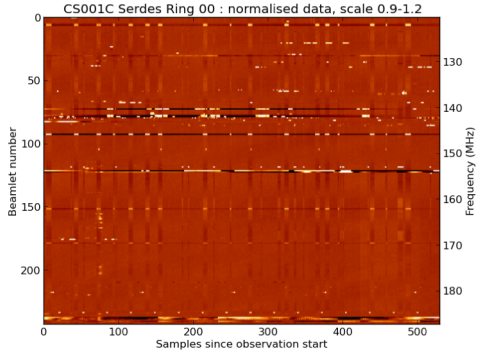
Images: Ekers+1973, Harwood+2016

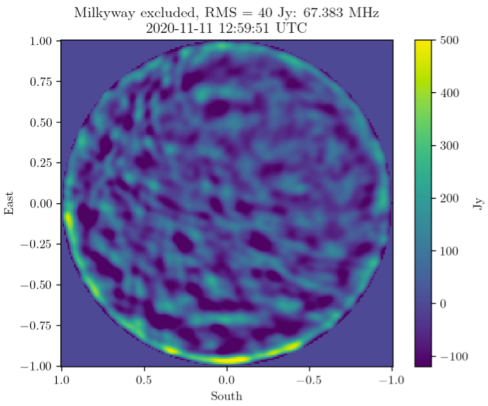
Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Data quality assessment

- LOFAR is a complex network
  - 38 Dutch stations (24 core, 14 remote)
  - 14 international stations
- What is the status of the array?
  e.g. individual stations, cabinets
- What are the properties of the data?
  e.g. data loss, interference
- How is the environment around stations?
  e.g. sources of RFI
- Given this info, was the observation successful?







ASTRON Netherlands Institute for Radio Astronomy · LOFAR · STELLAR

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Interactive LOFAR map

https://www.astron.nl/lofartools/lofarmap.html

V. Moss

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

How do we assess a data set?

- Problems with the data can be identified at 3 different levels:
  - Single antenna
  - **Station**
  - **Baseline**
- Check for known issues at single antenna level using Real Time Station Monitoring (RTSM) at https://proxy.lofar.eu/rtsm/HTML/
  - Comparison of antenna response function with the other antennas
  - Listing per observation and daily summary per station
- Check data quality at station and baseline level using **LOFAR inspection plots** at https://proxy.lofar.eu/inspect/HTML/index.html

ASTRON LOFAR STELLAR
Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Inspection plots

Summary page

LOFAR inspection plots

Last modified: Thu May 31 09:09:33 2018 UTC Full list Ascii table

SAS ID	Campaign	Target	DynSpec	Compl	Compl*	AntennaSet	Band	Start	End	Clock	Subb	Parset
L655840	LT10_010	P031P036REF	BST	0%	N.A.	HBA_DUAL_INNER	HBA_110_190	2018-05-31 07:30:01	2018-05-31 11:30:01	200	487	parset
L655836	LT10_010	3C48	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-31 07:19:01	2018-05-31 07:29:01	200	243	parset
L655850	LT10_001	CasA	BST	66%	N.A.	LBA_OUTER	LBA_10_90	2018-05-31 05:35:00	2018-05-31 06:55:00	200	480	parset
L654622	LC10_001	3C48	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-31 05:23:40	2018-05-31 05:33:40	200	243	parset
L654616	LC10_001	P265P271REF	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-30 21:22:40	2018-05-31 05:22:40	200	487	parset
L654612	LC10_001	3C295	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-30 21:11:40	2018-05-30 21:21:40	200	243	parset
L655784	LT10_010	3C295	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-30 20:56:46	2018-05-30 21:06:46	200	243	parset
L655856	FE_monitoring	B0809+74	BST	100%	N.A.	LBA_OUTER	LBA_10_90	2018-05-30 19:47:00	2018-05-30 19:57:00	200	244	parset
L655854	FE_monitoring	B0809+74	BST	100%	N.A.	LBA_INNER	LBA_10_90	2018-05-30 19:36:00	2018-05-30 19:46:00	200	244	parset
L655852	FE_monitoring	B0809+74	BST	100%	N.A.	HBA_DUAL	HBA_110_190	2018-05-30 19:30:00	2018-05-30 19:35:00	200	244	parset
L655778	LT10_010	P005P359REF	BST	100%	39%	HBA_DUAL_INNER	HBA_110_190	2018-05-30 16:55:46	2018-05-30 20:55:46	200	487	parset
L655774	LT10_010	3C196	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2018-05-30 16:44:46	2018-05-30 16:54:46	200	243	parset
L655814	FE_monitoring	B0809+74	BST	100%	N.A.	LBA_OUTER	LBA_10_90	2018-05-30 16:32:00	2018-05-30 16:42:00	200	244	parset
L655812	FE_monitoring	B0809+74	BST	100%	N.A.	LBA_INNER	LBA_10_90	2018-05-30 16:21:00	2018-05-30 16:31:00	200	244	parset
L655810	FE_monitoring	B0809+74	BST	100%	N.A.	HBA_DUAL	HBA_110_190	2018-05-30 16:15:00	2018-05-30 16:20:00	200	244	parset
L655570	LC10_020	3C76.1	BST	100%	11%	LBA_OUTER	LBA_30_90	2018-05-30 13:05:00	2018-05-30 14:05:00	200	480	parset
L655508	LC10_020	3C49	BST	100%	100%	LBA_OUTER	LBA_30_90	2018-05-30 12:04:00	2018-05-30 13:04:00	200	480	parset
L655446	LC10_020	3C217	BST	100%	100%	LBA_OUTER	LBA_30_90	2018-05-30 11:03:00	2018-05-30 12:03:00	200	480	parset
L655384	LC10_020	3C172	BST	100%	100%	LBA_OUTER	LBA_30_90	2018-05-30 10:02:00	2018-05-30 11:02:00	200	480	parset
L655322	LC10_020	3C123	BST	100%	100%	LBA_OUTER	LBA_30_90	2018-05-30 09:01:00	2018-05-30 10:01:00	200	480	parset
L655260	LC10_020	3C436	BST	100%	100%	LBA_OUTER	LBA_30_90	2018-05-30 08:00:00	2018-05-30 09:00:00	200	480	parset
L654754	LC10_015	J2252+24	BST	100%	N.A.	HBA_DUAL	HBA_110_190	2018-05-30 05:24:00	2018-05-30 06:24:00	200	400	parset
L655756	LT10_001	CasA	BST	66%	N.A.	LBA_OUTER	LBA_10_90	2018-05-30 04:15:00	2018-05-30 05:20:00	200	480	parset

https://proxy.lofar.eu/inspect/HTML/index.html

ASTRON LOFAR STELLAR

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

L655836

- Projects

Cobalt ERROR log ← Correlator errors

Max file sizes (MB):

Correlated data : 3360
Beamformed data : 0

All data sets are there ← Missing data sets on CEP4

Input loss report

CS013HBA0: 100.0000%
CS013HBA1: 100.0000% ← Station flagging
CS028HBA0: 0.0022%
CS028HBA1: 0.0010%
RS503HBA: 20.5472%

Station Dynamic Spectra ← BST plots

Name	Subband (ID)	Freq. (MHz)
SAP000_SB000	104	120.312
SAP000_SB001	105	120.508
SAP000_SB002	106	120.703
SAP000_SB003	107	120.898
SAP000_SB004	108	121.094

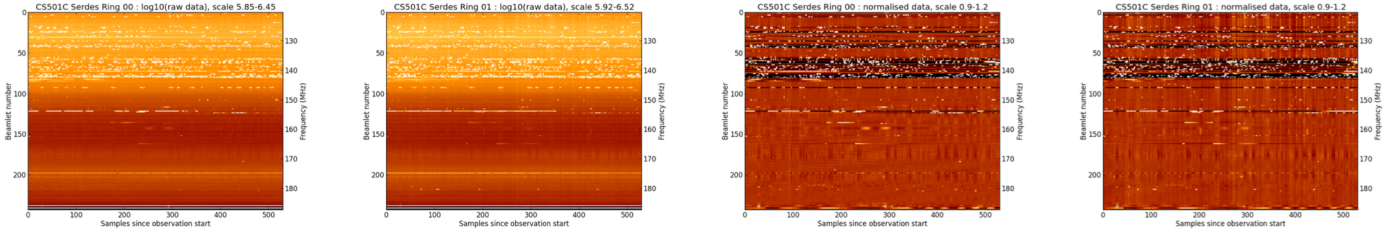
← Visibility plots, per subband

ASTRON LOFAR STELLAR
Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

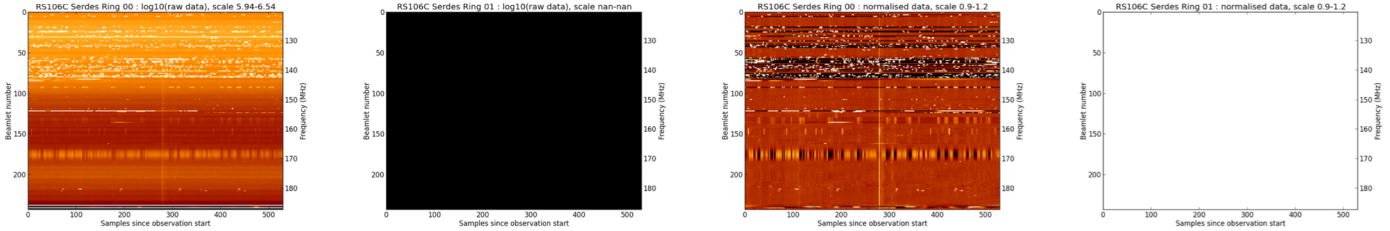
BST plots

- Beamlet statistic / station dynamic spectrum
- One sample every second
- Plots on left are raw data, plots on right are averaged
- Documentation: https://old.astron.nl/radio-observatory/observing-capabilities/depth-technical-information/data-quality-inspection/data-qu

CS501 Ring 00 Ring 01 Ring 00 Ring 01

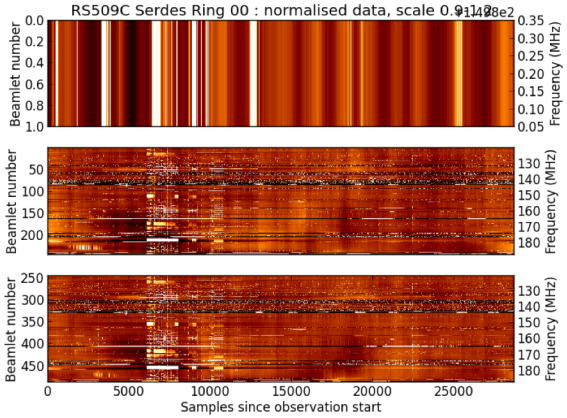


RS106

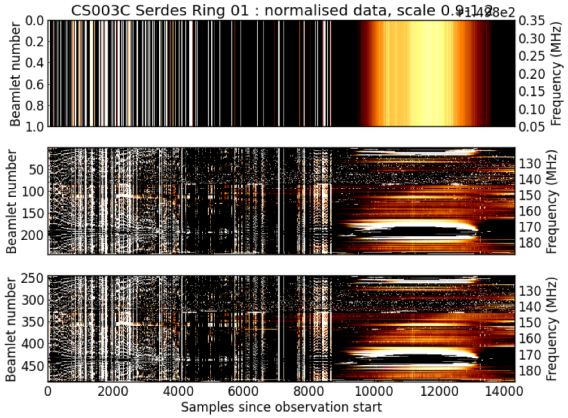


Raw data Averaged data

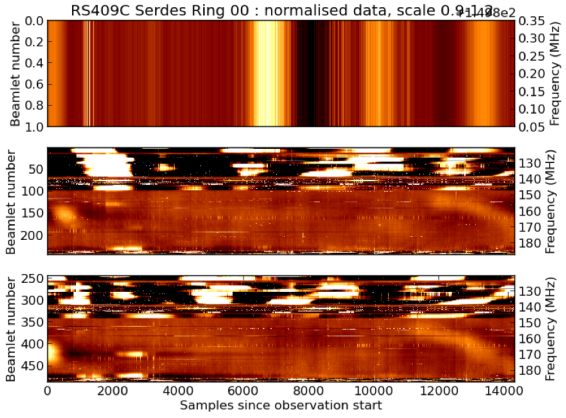
Dynamic spectra examples



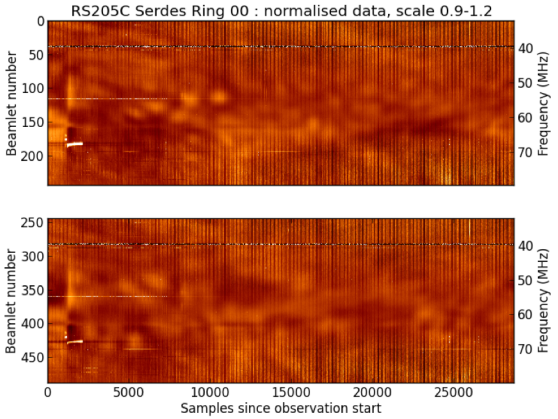
Mildly oscillating tile



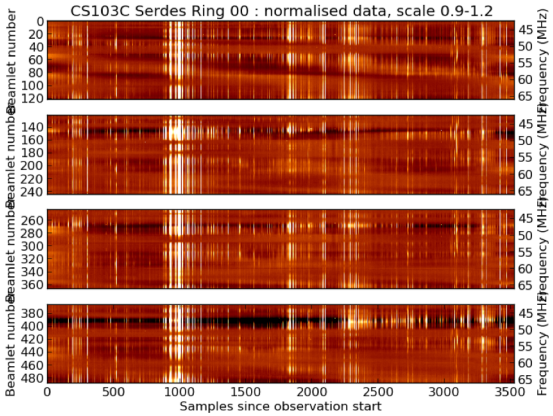
Really oscillating tile



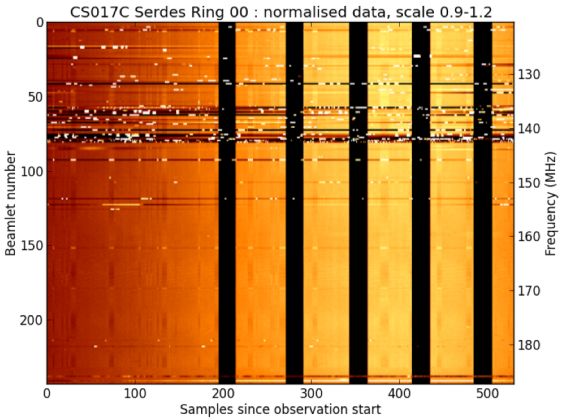
Broadband RFI



Electric fence

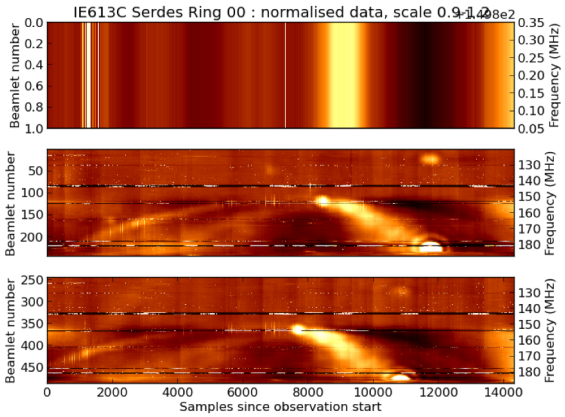


Strong local interference

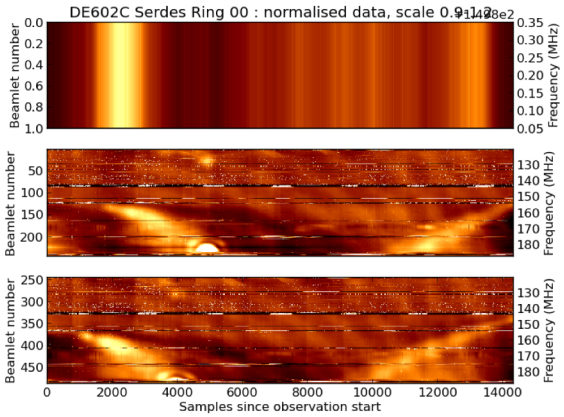


Power supply issue

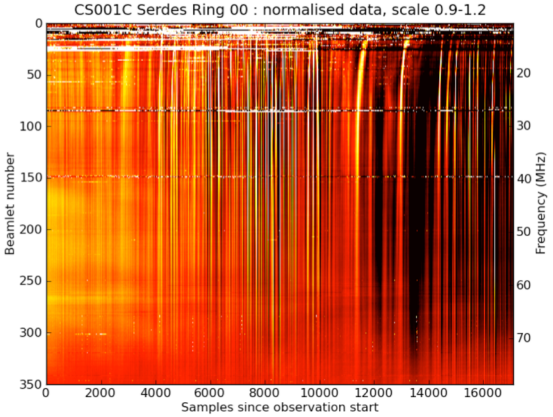
Dynamic spectra examples



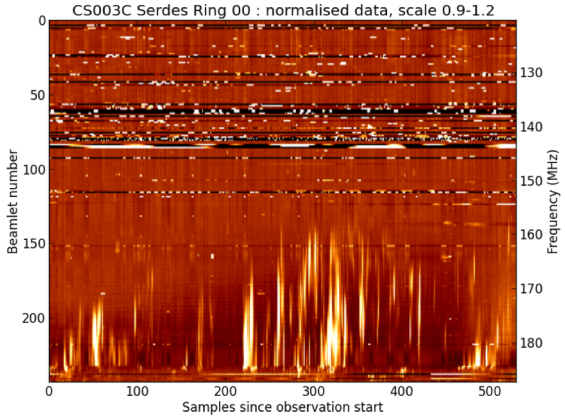
Bright sidelobes



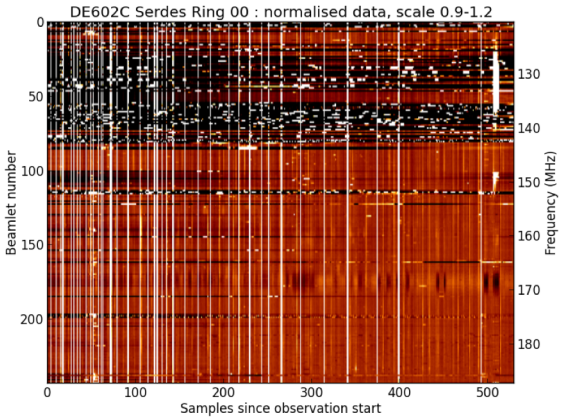
Interfering structure



Ionospheric scintillation



Solar storm



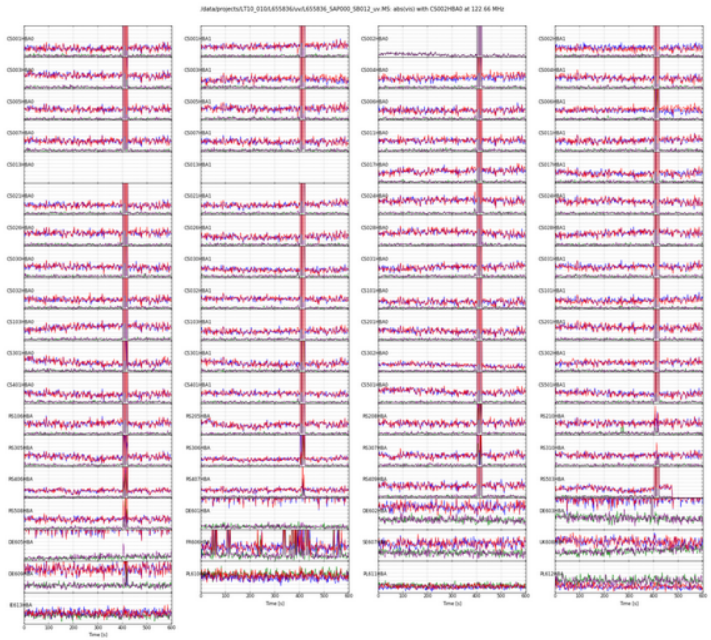
Lightning

Visibility plots

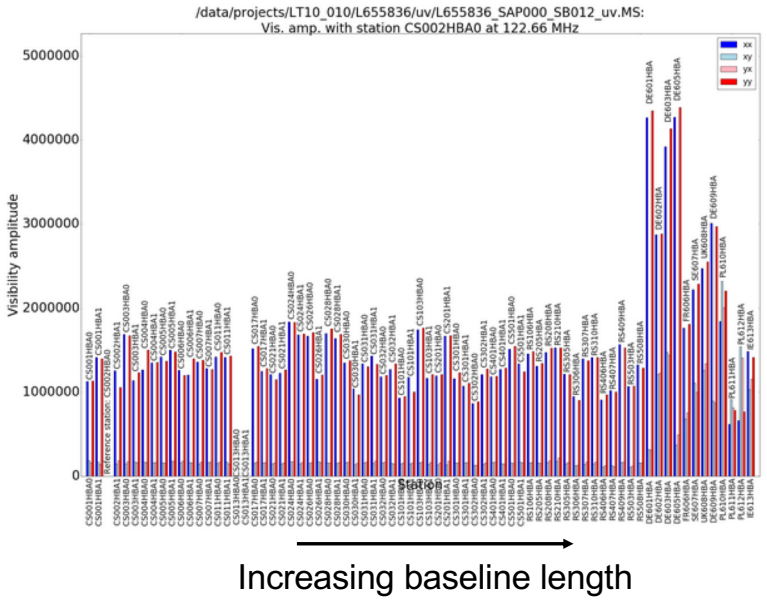
655836 SAP000_SB012

- Subbands Projects
- What do I see here?

Visibility amplitude versus time



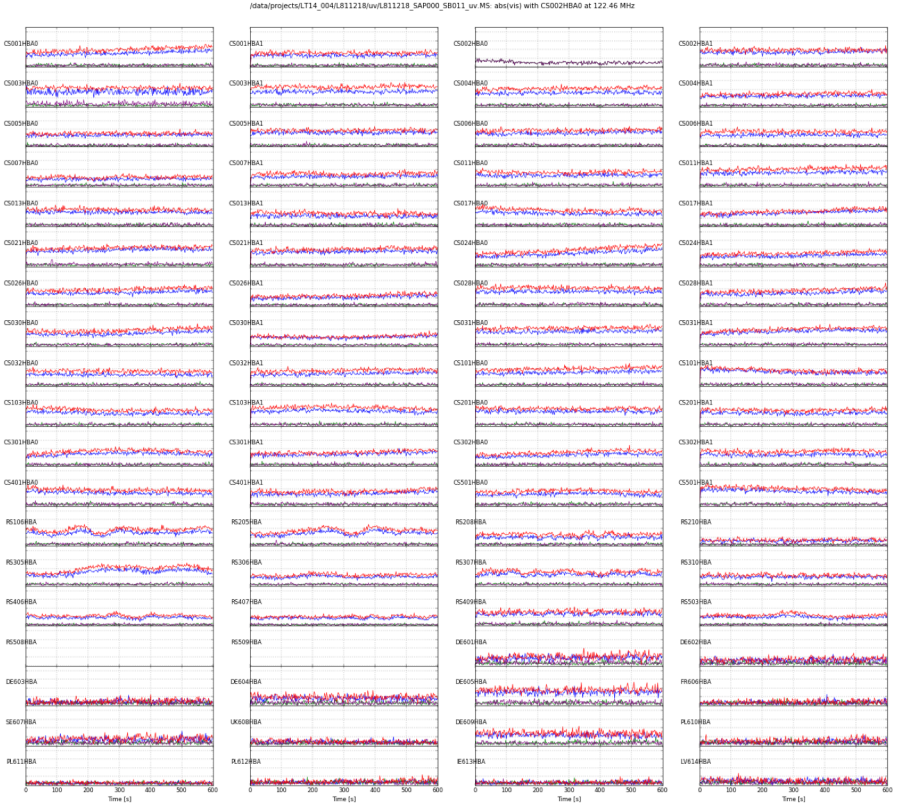
Mean visibility amplitude



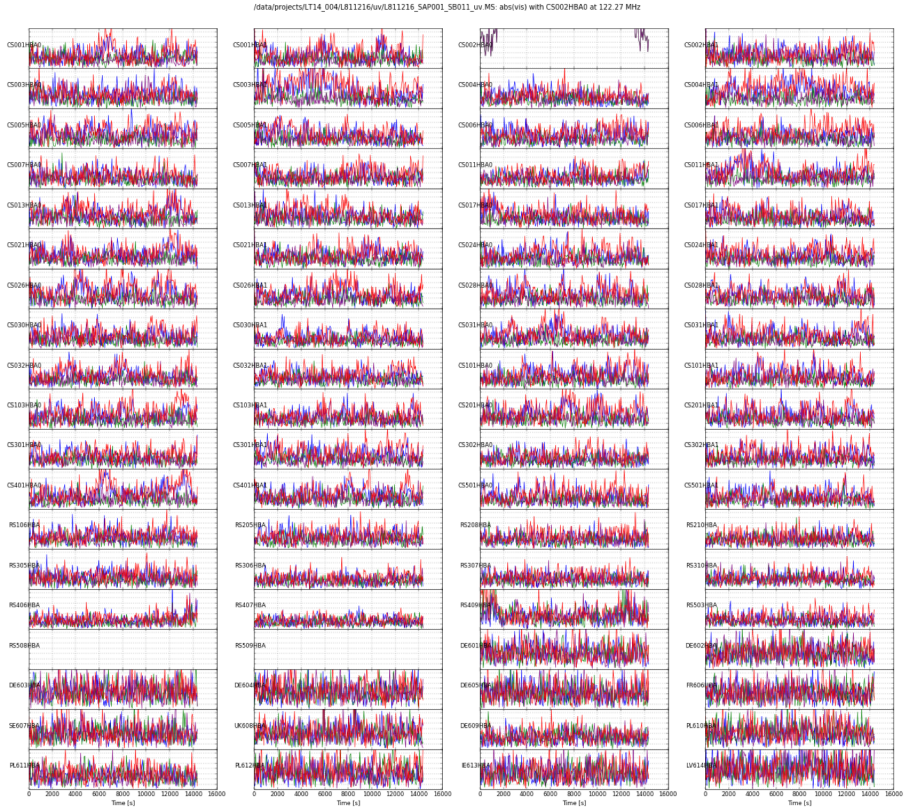
Reference station: CS002HBA0

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Interferometric examples



Bright calibrator (3C196)

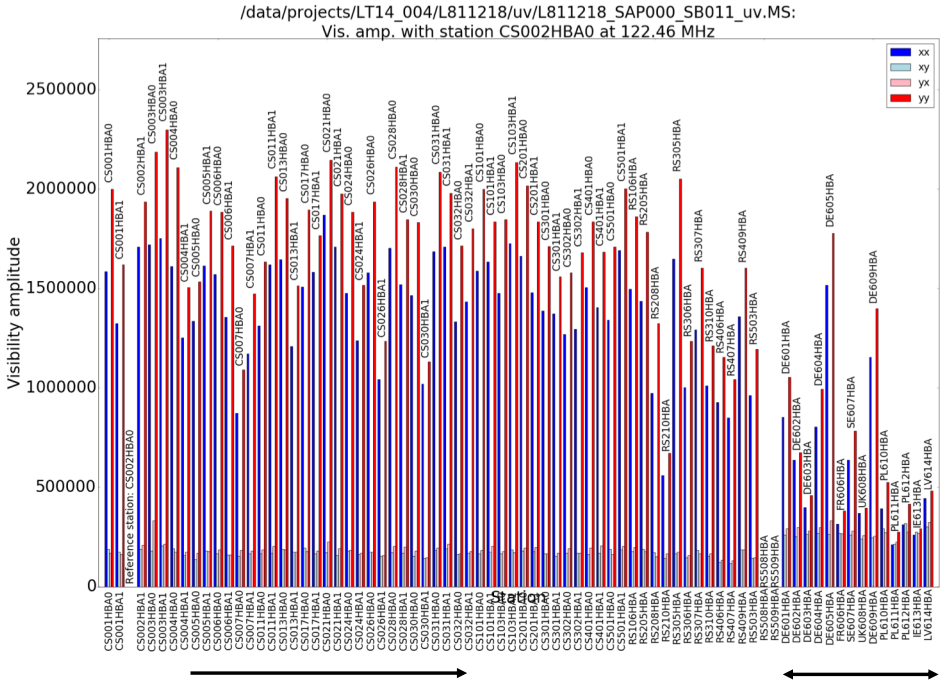


Faint target field

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Interferometric examples

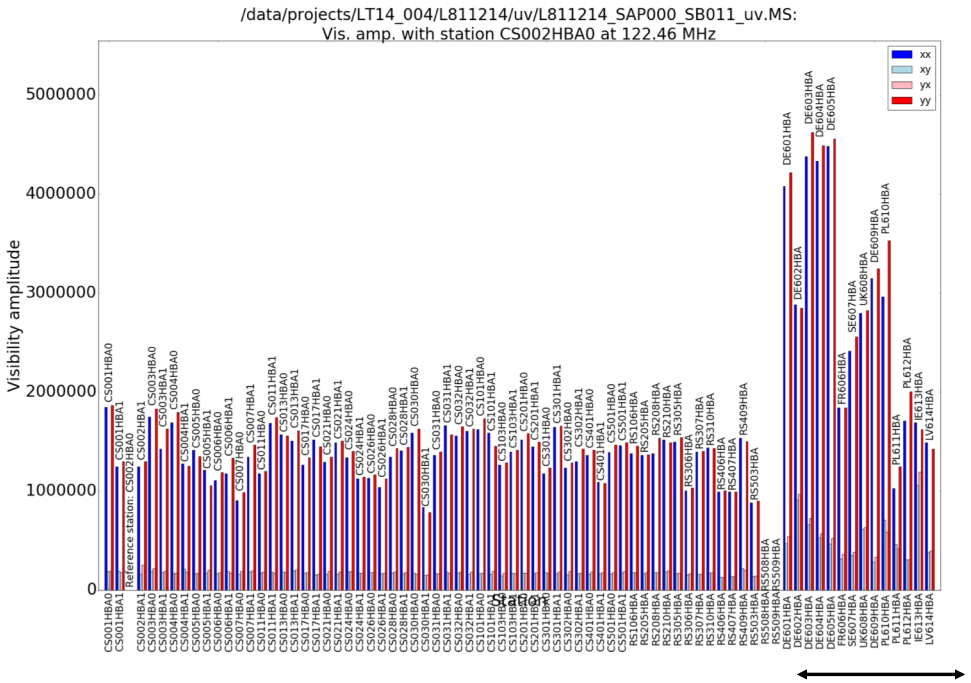
Bright calibrator (3C196)



Increasing baseline length

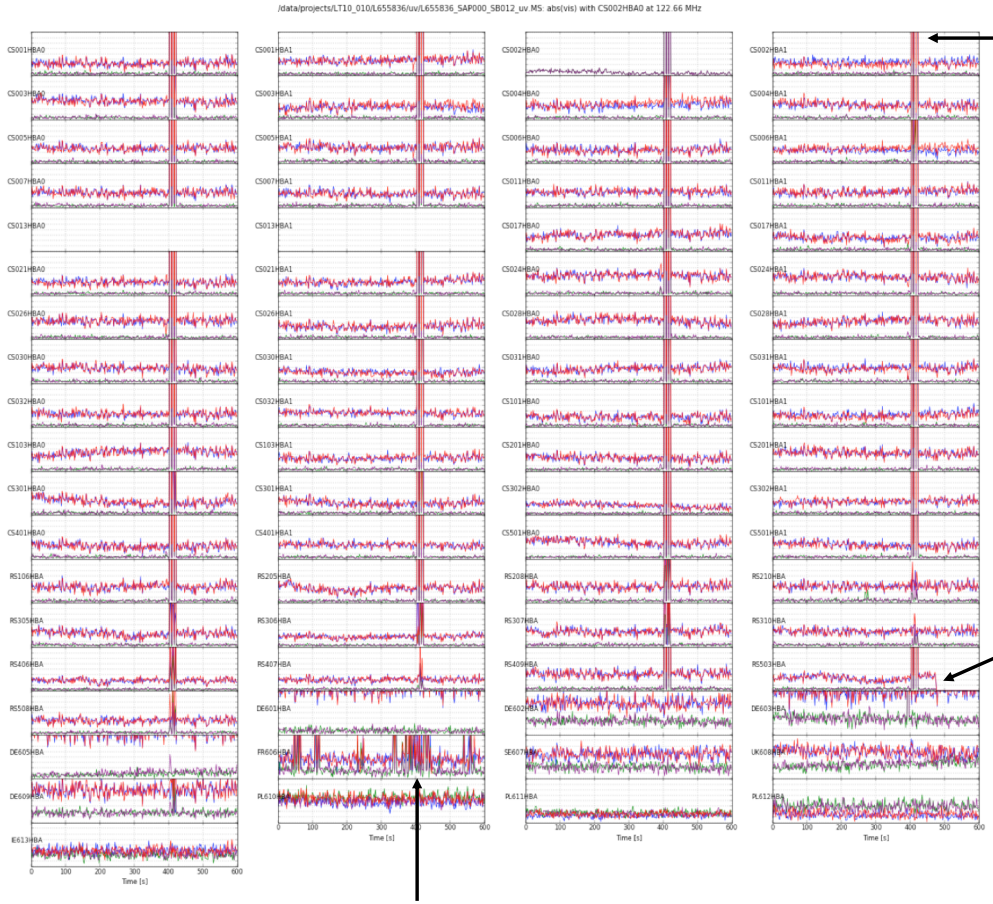
International stations - lower signal, source resolved on longest baselines

Bright calibrator (3C48)



International stations - higher signal due to larger collecting aera

Interferometric examples



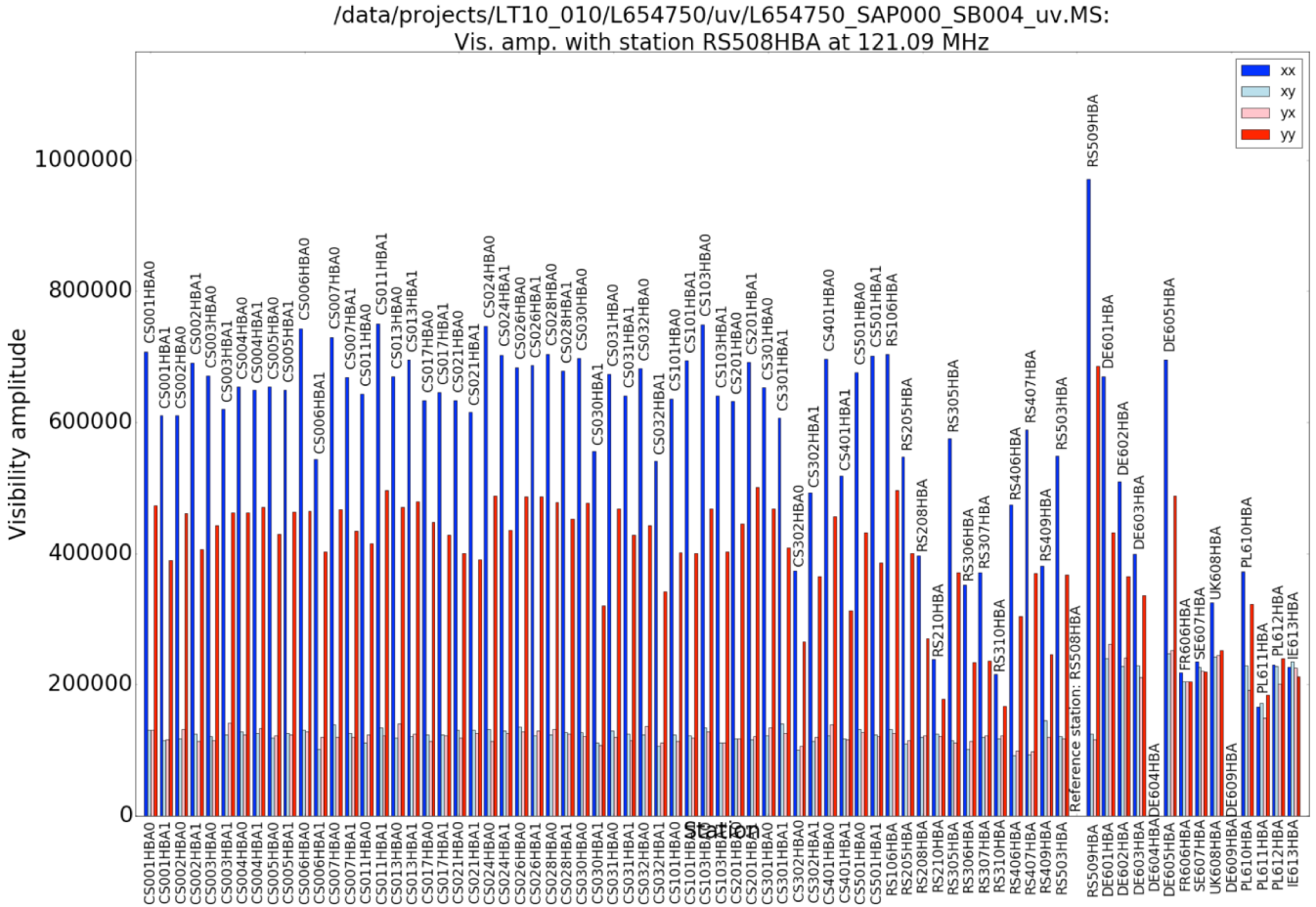
RFI
(satellite?)

Network interruption

Check neighbouring sub-bands / BST plots to check
if general problem or interference in this band

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Differential Faraday rotation



Data quality report to PI

Subject: observations L803642/46/52 successful

Dear Colleague,

The following message contains information regarding a LOFAR Cycle 14 project for which you are listed as the contact author. Please forward this information to your collaborators.

We would like to inform you that observations related to your LOFAR Cycle 14 project have been perfomed and are considered successful. Please find detailed information below.

General notes: any fundamental remarks

Observations: (details of performed observations)

L803652	LC14_019	3C295	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2021-01-10 06:48:41	2021-01-10 06:58:41	200	243	parset
L803646	LC14_019	P161P156REF	BST	98%	99%	HBA_DUAL_INNER	HBA_110_190	2021-01-09 22:47:41	2021-01-10 06:47:41	200	487	parset
L803642	LC14_019	3C147	BST	100%	100%	HBA_DUAL_INNER	HBA_110_190	2021-01-09 22:36:41	2021-01-09 22:46:41	200	243	parset

Performance of the system: any issues to report with stations

Data recording: any part of the requested data missing?

Data processing: status of processing jobs

Archiving: 'has started', 'is scheduled', ...

Remarks: Please analyse the validation plots at https://proxy.lofar.eu/inspect/HTML/ within 24 hours after this notification and submit a support request at https://support.astron.nl/sdchelpdesk in case you need to report problems about their quality. After this time window has passed, we will assume that your judgement is that the observation was successful and we will complete the actions described above to support your run.

From the moment the data are made available to you at the LTA you have four weeks to check their quality and to report any problems to SDCO. After this time window has passed, no requests for re-observation will be considered.

Actions: if you need any further clarification, please do not hesitate to contact us through the SDC-helpdesk at https://support.astron.nl/sdchelpdesk , specifying your project code in the subject.

ASTRON LOFAR STELLAR

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

Observation report: policies about failures

- Policies to determine if an observation is successful:
  old.astron.nl / Radio Observatory / LOFAR policies

- Most relevant are:
  - An observation will be considered failed if **> 5%** of the data are missing on disk
  - Processing will be considered failed if **> 5%** of the processed data are missing (w.r.t. the raw data)
  - In other cases, observations may be considered failed on a case-by-case basis
  - The only raw data inspection available to users prior to data reduction is via the **inspection plots**
  - Users have 4 weeks to check the data quality and, if necessary, request for re-observation from the moment the data are made available at the LTA
  - Failed observations may be repeated only once if the observing schedule allows it
  - All ‘priority A’ Cycle projects (with the exception of ToO projects) that cannot be completed by the end of the Cycle they refer to will remain active only during the following semester and they will be observed then with second priority with respect to the new Cycle projects.

ASTRON LOFAR STELLAR
Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

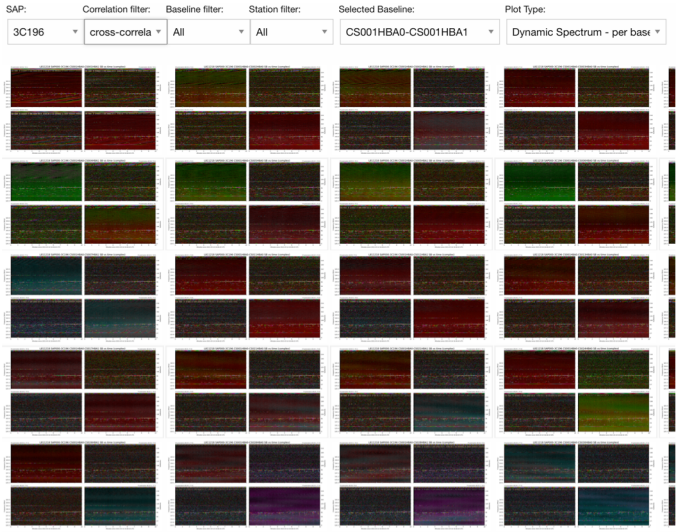
Archive of LOFAR inspection plots

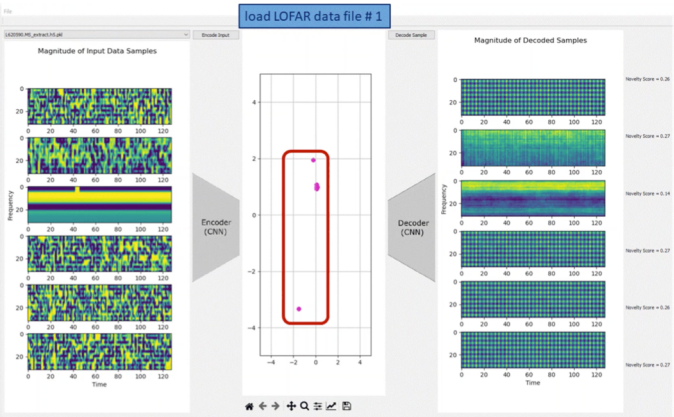
- LOFAR inspection plots routinely generated since Aug. 2017
- Only remain online for 3 weeks from date of observation
- After that, they are compressed and transferred to offline storage
- If you wish to access inspection plots before downloading observation or pipeline on LTA, contact SDCO staff via ASTRON helpdesk
  - Provide project code and SAS ID of observation you are interested in

Image credits: A. Corstanje, F. Sweijen, C. Van Eck, P. Zucca

The future of inspection

- New, more efficient QA system is being developed at ASTRON
- Generates dynamic spectra **per baseline** (1000s of plots per observation!)
- Uses machine learning to group baselines into clusters and automatically identify wide range of anomalies in data
- ADDER inspection plots: https://proxy.lofar.eu/qa

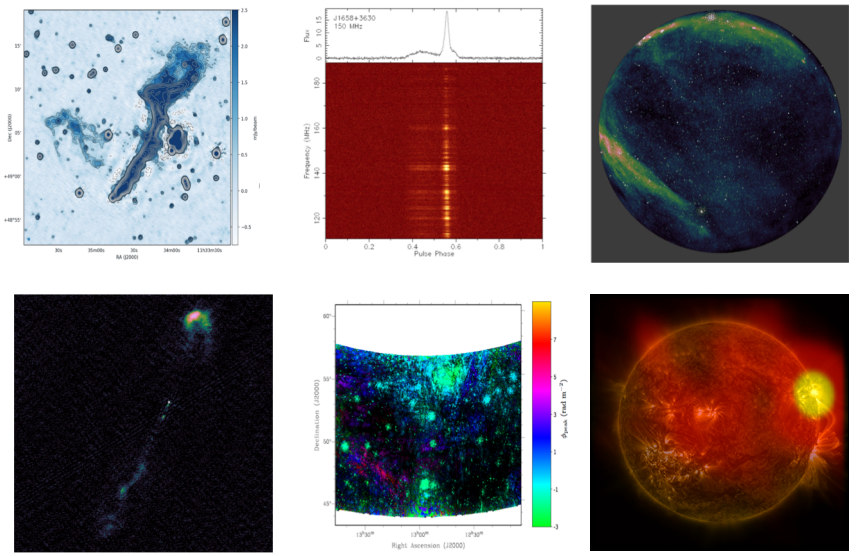




Courtesy of M. Mesarcik, A. Boonstra

Summary

- Data quality only makes sense in the context of the **requirements** that data sets need to meet
- **SDCO** assesses data sets based on a series of characteristics defining **generic data quality**
- We have a **policy** which helps us to determine whether a data set is **successful** or not
- We are exploring machine learning to speed up identification of problems
- As a **LOFAR user**, it is your **responsibility** to help determine whether the data quality is sufficient to reach your science goals



Images: Shimwell+, Tan+, Shulevski+ Sweijen+, Van Eck+, Zucca+