Netherlands Institute for Radio Astronomy

AST(RON





Raw data quality assessment

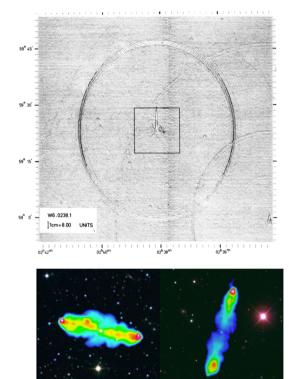


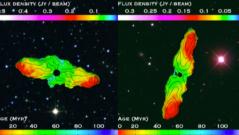




What is data quality?

- 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

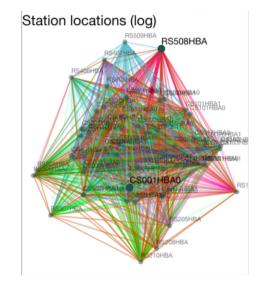


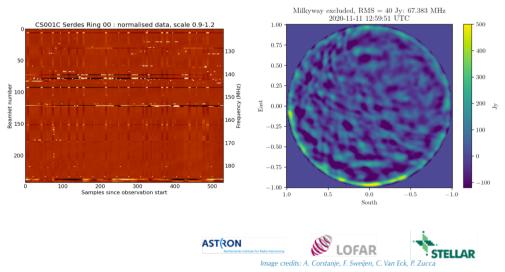


Data quality assessment

- 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?









Interactive LOFAR map



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

V. Moss





How do we assess a data set?

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







Ż

Inspection plots

Summary page

LOFAR inspection plots

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

SAS II	Campaign	Target	DynSpec	Compl C	ompl*	AntennaSet	Band	Start	End	Clock	Subb	Parset
L655840	LT10_010	P031P036REF	<u>BST</u>	<u>0%</u>	<u>N.A.</u> 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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>66%</u>	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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	<u>N.A.</u> 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	<u>BST</u>	<u>100%</u>	<u>N.A.</u> HBA	_DUAL	HBA_110_190	2018-05-30 19:30:00	2018-05-30 19:35:00	200	244	parset
L655778	LT10_010	P005P359REF	<u>BST</u>	<u>100%</u>	<u>39%</u> 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	<u>BST</u>	<u>100%</u>	<u>100%</u> 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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	<u>N.A.</u> 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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	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	<u>BST</u>	<u>100%</u>	<u>N.A.</u> HBA	_DUAL	HBA_110_190	2018-05-30 05:24:00	2018-05-30 06:24:00	200	400	parset
L655756	LT10_001	CasA	<u>BST</u>	<u>66%</u>	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







L655836

Projects

SAP000 SB004 108

Cobalt ERRO	<u>R log</u> +		Correlator errors						
Max file sizes ((MB):								
Correlated data Beamformed data									
All data sets a	re there	←	——— Missing data sets on CEP4						
Input loss repo	ort								
CS013HBA0: 10 CS013HBA1: 10 CS028HBA0: 0.0 CS028HBA1: 0.0 RS503HBA1: 20	0.0000% 0022% 0010%		— Station flagging						
Station Dynamic Spectra ←──── BST plots									
Name	Subband	Freq.							
	(ID)	(MHz)							
SAP000 SB000	104	120.312							
SAP000 SB001	105	120.508	← Visibility plots, per subband						
SAP000 SB002	106	120.703							
SAP000 SB003	107	120.898							

121.094

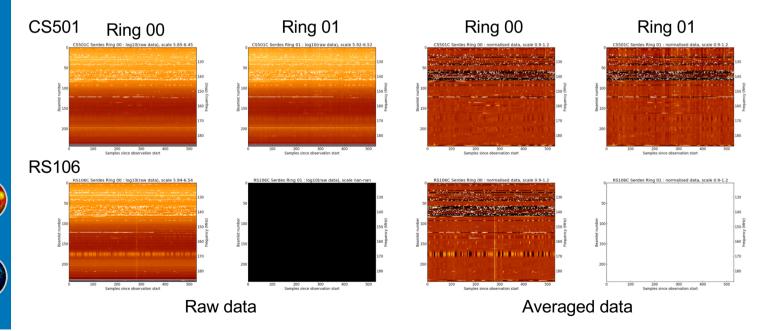




BST plots

- 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-technicalinformation/data-quality-inspection/data-qu





Dynamic spectra examples R5509C Serdes Ring 00 : normalised data, scale 0.위1488e2 0.35 0.30 프로 0.25 - 2

0.20

0.10

130 H 140 U

150 J 150 J 160 J 170 J 180 J 180 J

130 H 140 N

150 S 160 S

170 180 L

Ereau 50

0.15



0.2

0.4

a 0.6

0.8

50

2 100

칠 150

و 200 م

- 250

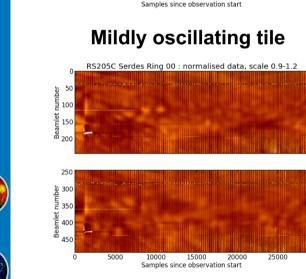
ਕੂ 300

2 350

ট 400

up 450

5000



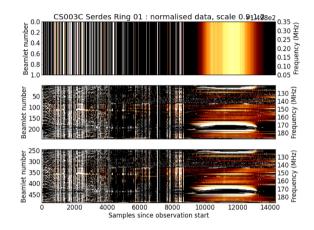
10000

15000

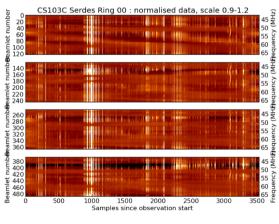
20000

25000

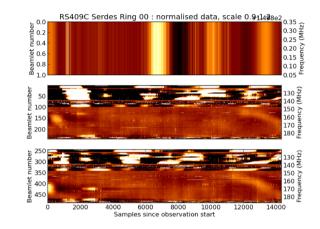
Electric fence



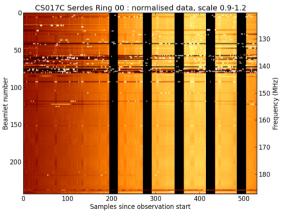
Really oscillating tile



Strong local interference



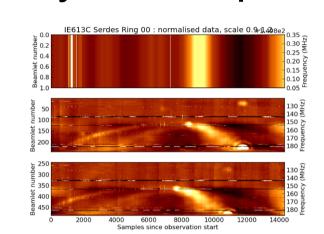
Broadband RFI



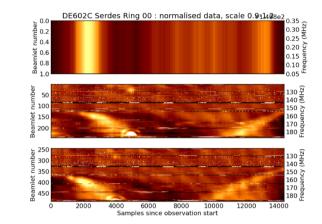
Power supply issue



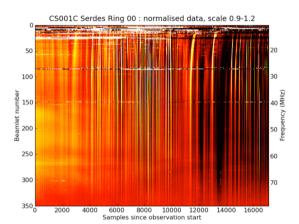
Dynamic spectra examples



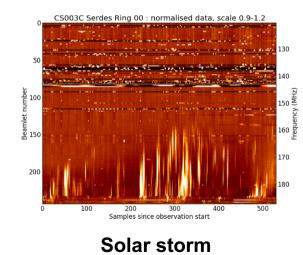
Bright sidelobes

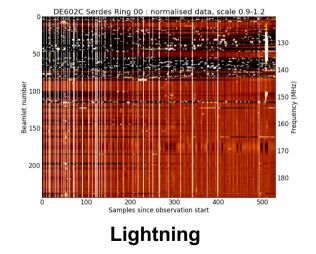






lonospheric scintillation











Visibility plots

9

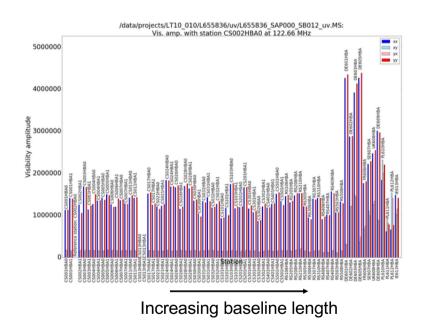
				12_vv MS: abs(vis) with C5002HBA0 at 12			
and the second second second second	vw44	and an and the second second second		5003-840		Second States of the second second	ł
mallating leaders in a property and		genelister polynolauteres	windy with a	Source and the second s	where where	-	ł
man way and a star	week core	mallowshimmed	-	sources	monute	Sounday	ł
Westerna and and and and a star	www.	mainlymphillenderser	-dy mille	5003 mar - Hay may + 1 & may + +++ mar +	- sphrase	Second and a second and a second	
Selaware	GHL	MA)		Sources	way a shifter	SILVERS MARKEN STRANGE	ł
session and the second second	with car		manute	source for the second s	-	Mar front Marriagues	ł
weeter werden and the second		-	waythink		Marcylink	-	ł
washing a second and a second second	WIN CHA			5032-040 1993-1110-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-09-00-0 1993-1993-1993-1993-1993-1993-1993-1993	way to whether	second shall and the second	ł
The support of the second second	WA	and the start and and a second		and the second second		Silinas Antelinen yr Antelinen	Į
and the state of the second second				3202-040 AT-140 (141) - 10 A.P 44-07-07	may to make		ł
and the second and		-		5362-640	munum	EARLING CONTRACTOR	
multon vienary very	C40		morenalts	SSE ANT ANT AND	ways, yuku	SSOLARA	ļ
and the second second second second second	wylithy Marco		wareness	there was not been and the new second s	where the second	1220 Mary all 10 10 10 10 10 10 10 10 10 10 10 10 10	ł
the superior and the second	-	an a		and the second second		succession and the second second	1
Kattan				5	Dettoppendieren		ļ
and a construction of the second		an and a state of the state of	Anna Annala	and a state of the second states and a state of the second states and the second states	water or years	xxx and become and the sport of	4
anonen a handi bahashashashashashashashashashashashashas	max	2 Jour martingly	hout	an alon bergitandaria	north/th		ķ

655836 SAP000_SB012

<u>Subbands</u> Projects

· What do I see here?

Mean visibility amplitude









Interferometric examples



CS007HBA0	C\$007F		C5011H	
	- Anna and the second states	ani Selata da Alakar, que antegra d'Arcetado Antegra de Arcetado		
CS013HBA0 Coolgengy (MAR-spacing) (AT March (New March Indeterming water from the contract (New March			C5017H	
CS021HBA0		BA1	C5024H	
CS026HBA0		BA1	CS028H	Kalena arta ana a
CS030HBA0		BA1	C5031H	
CS032HBA0		BA1 adapted of the product of the second		
C5103H B40		BA1		
CS301HBA0		BA1 ####Katilin-GapCommunity#Colouring#Colouring#Colouring#Colouring#Colouring#Colouring#Colouring#Colouring#Colour	C\$302H	
CS401+BA0		BA1 In Relations and a static st		
PS106HBA		HBA	P5208	
PS305HBA				
R5406HBA	RS407	IBA		BA Chandeland and a chandle and
PS508RBA	PESO9	IBA	DE601	iaa An taraa taraa ay ahaa ahaa ahaa ahaa ahaa ahaa ah
DEGO3RBA	DE604	ны Мактория и рабок станицата спорта с	DE605	BA
SEGOTHEA	UKSOB	HBA	DE609	IBA De Malen por de la construction de De la construction de la construction
R.6118BA	R.612	BA	E6138	BA

and a superior of the superior

and the second second second second second second

Bright calibrator (3C196)

C5001H840	and some and the source of the
CS003HE	1846 add for additional difference
CS005HBAD	unal Maturia distribut
C5007H8A0	Anter
CS013H	Alleidam Milling Milling
CS021HEAD	Hennesteinennikilun
CS026H	the Martin Martin
CS030HRAC	helsehelsennstanlittensel
C5032HBAQ	10th march marker willy
CS103HB40	South Bried Stration of the
CS301H	. challabarten hille damathile d
C5401H8A0	and a Minimum Merch
R5106	an hear she will be seen a built of some
R530548	na na salan katalan katala
RS406HBA	mundering while the stand
RSSOBEBA	
DE 603	
SE607HBA	W STATES AND
R.611	Junited whether Burning als

 CODE
 CODE

 CODE</t

Faint target field

14 004/1811216/uv/1811216 SAP001 SB011 uv MS: abs(vis) with CS002HBA0 at 122 27 MH





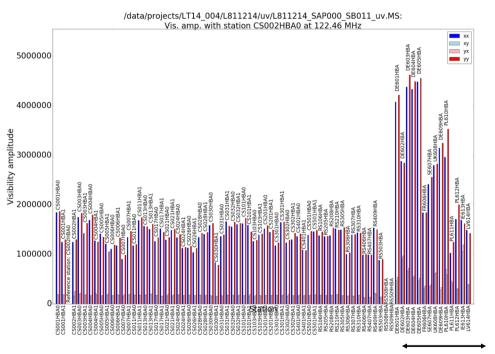
Interferometric examples

Bright calibrator (3C196)

<figure>

Increasing baseline length

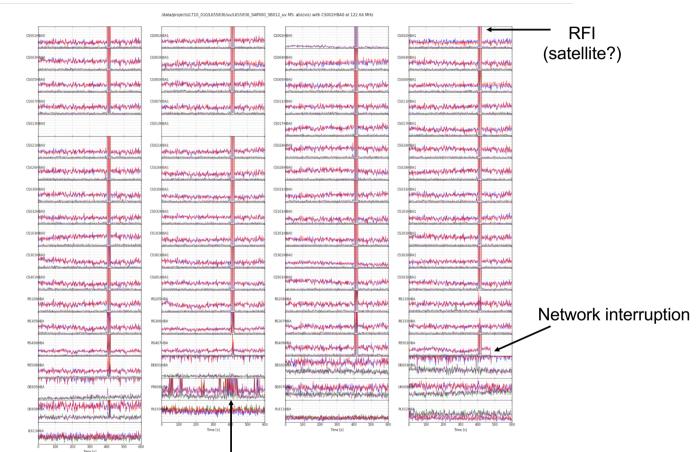
International stations lower signal, source resolved on longest baselines **Bright calibrator (3C48)**



International stations higher signal due to larger collecting aera



Interferometric examples



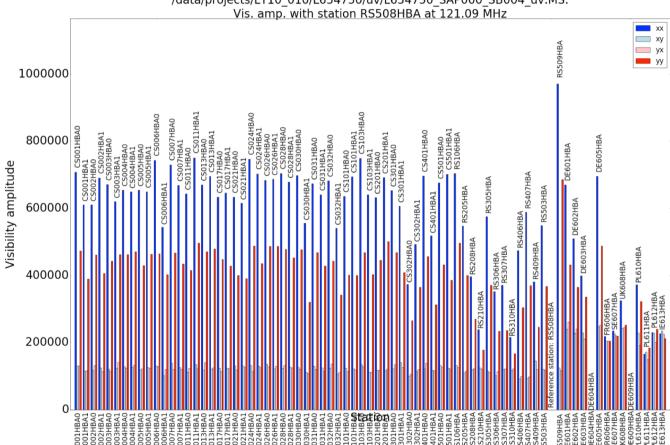
Check neighbouring sub-bands / BST plots to check if general problem or interference in this band







Differential Faraday rotation



/data/projects/LT10_010/L654750/uv/L654750_SAP000_SB004_uv.MS: Vis. amp. with station RS508HBA at 121.09 MHz





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)

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

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://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.







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)
 - $\,\circ\,$ 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 reobservation from the moment the data are made available at the LTA
 - \circ 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.







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 <u>ASTRON helpdesk</u>
 O Provide project code and SAS ID of observation you are interested in



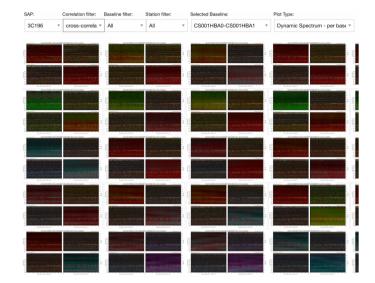


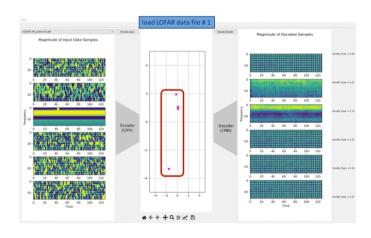




The future of inspection

- 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: <u>https://proxy.lofar.eu/qa</u>





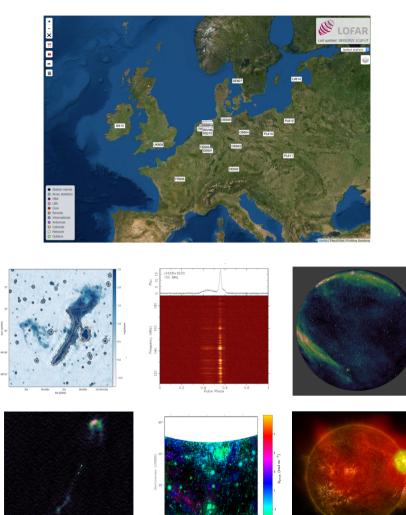
Courtesy of M. Mesarcik, A. Boonstra





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+

