Roadmap for progress with LOFAR imaging in the next 3-4 months
==========

v.1 Neal Jackson 11.01.22 (after imaging BW9 and a meeting with John Conway, George Heald, Neal Jackson, Ronald Nijboer, Roberto
Pizzo, Antonis Polatidis, Bas vd Tol, Reinout van Weeren). Two sections: removal of bright sources (long), AOB (short).
v.2 11.01.26 after comments from above.

1. Removal of bright sources

The most urgent problem facing LOFAR data analysis is the removal of the A-team, and in particular Cas A and Cyg A. This is especially important for LBA observations in which Cas/Cyg appear on essentially all of the core baselines. For the HBA we might be able to get away with simply removing short baselines, and this can be tested once full HBA station calibration is running). Since the advent of the beam model, Cas/Cyg now limit the quality of the maps that can be produced with LOFAR LBA. Some tools now exist which can be used to solve for and remove these sources, but Imaging BW9 (17-21 January) made it clear that we need to quantify how good is the removal and
to compare different methods. A second requirement is to allow the work done by many different people to appear in published form, even before original astronomical results are available for publication.

The possible ways of removing bright sources are:

a) The existing BBS implementation, which solves for direction-dependent gains. Trials have shown that this works, but that removal is typically quite slow. For about 1 hour's data on 1 subband, compressed to 16 channels/subband, this requires about 1 hour on 1 compute node to solve in 2 different directions. The indications are that this scales as somewhere between N^2 and N^3 with the number of directions, so removal of more than CygA/CasA is possible only if the data are highly compressed.

b) As method a), but using a BBS solution on compressed data, allowing it to find and remove phase and amplitude solutions in multiple directions. This needs BBS to have access to decorrelation (D) factors for each averaged time and baseline (i.e. columns of numbers in the compressed data)

c) Observe A-team sources in multiple beams using compressed data, then cross subtract. This requires D-factors and very good beam models, currently 0.1% according to Sarod Yattawatta. (This can be done on uncompressed data but this might only give a relatively modest saving and involve cross-mixing of the target and subtracted source).

d) A more elegant way to deal with the problem is being developed by Bas v.d. Tol in which the observed visibility is written as a sum of source visibilities in different directions, each affected by the appropriate beam/smearing corruptions, and the corrected visibilities are recovered by matrix inversion.

What ideally needs doing (5 sections, commissioning, software, publications, observing, source models. Action items are marked ***. Comm means all commissioners. /d means needs some discussion/opportunity to object.)

---> Commissioning. Priority 1, while awaiting new software development, is to continue the investigation of approach a), recognising that this may only be feasible for small datasets until much more computing power is available. Specifically, what is needed is data (quantitative if possible, i.e. what fraction of corrected data is of acceptable qualiity?) on the quality of correction as a function of:

i) Observed source (distance from A-team .or. which member of A-team
being corrected for others)
ii) Degree of data compression (from modest compression to 1 chan/subband)
iii) Number of directions being corrected for.

It is difficult to cover all of this parameter space in a reasonable time. However for each A-team source (or 3C196/A2256) it should be possible to investigate two to three values of compression and

correction directions. This is important in order to determine
i) The exact time and computing requirements for the operation
ii) A baseline against which to compare approaches b) and c) when available.

Some of this has been done at busy weeks recently. Roberto Pizzo **(***RP)** offered to coordinate this as a continuation of BW9, starting with the next busy Wednesday on February 2. Shortly after this, those responsible for particular objects should send him details of these investigations **(***Comm)**

Priority 2 (shortly priority 1) is to test new software as it becomes available (see below).

---> Software development. Priority 1 is the implementation of some form of smearing correction for subtraction of off-axis sources, if necessary without the correct treatment of flagged data initially. Once this is ready Roberto Pizzo and Neal Jackson for **(***RP/NJ)** can be informed and will then be responsible informing the busy week/busy Wednesday teams and the STWG, respectively, with a view to organising testing. Timescale: 1-2 months?

Equal priority 1 is the development of Bas van der Tol's algorithm **(***BvdT)**. Again Roberto Pizzo and Neal Jackson **(***RP/NJ)** will circulate this further when it is available so that testing can be organised. Timescale: 1-2 months?

Priority 2 is full implementation of the smearing correction with treatment of flagged data. Timescale: 3-4 months?

---> Publication. The idea was discussed at the end of BW9 of aiming towards a technical publication about subtraction of strong sources from low-frequency radio data. This would provide a short-term goal which should be achievable within 6 months. It would also provide recognition in the reasonably short term for those (especially students and postdocs) who have contributed to LOFAR calibration. It would compare the methods outlined above, and possibly present some initial A-team data. It would require coordination with other KSPs who are also working on the problem (e.g. EOR in the case of HBA data). **(***/d)**

---> Observations. Stations RS208 and RS307 appeared not to be functioning well (low amplitudes) during the A-team observations. It would be helpful to repeat these observations with the same setup (4 sources at once) as having only 4 remote stations out of six is problematic for good imaging on the smaller scales. **(***AP)**

A possible solution to data congestion problems on the cluster is to take only 20 subbands per observation, saving a factor of 10 (**(***AP/d)**: it is very rare for more than these to be processed in the current setup. The bands would need to cover the frequency range in order to allow separation between clock problems and ionospheric effects.

---> Source models. George Heald showed the python script produced by Bart Scheers which reads in source models to BBS, and the script of Sarod Yattawatta which implements shapelet decomposition. All are encouraged to use the script for initial source model generation. **(***Comm)**

More generally, any LOFAR image produced that is better than a previous one should be sent to Roberto Pizzo/George Heald for inclusion in the GSM as it develops **(***Comm)**. Roberto will bug commissioners for these after the next busy Wednesday.

GH: An area has been set up for this: /globaldata/gsm.The partition /globaldata has 0.75 TB free space, which should be sufficient ....
The idea here should be that anyone who wishes to submit data for the GSM actually goes through someone
(George/Roberto) for three reasons,
- quality control
- disk space control
- so we know to do something re getting the info in the gsm

2. Other things which emerged during BW9:
---

a) Parset library. Ideally a script could be produced which would take as input
  ---> frequency of observation
  ---> position of source (distance from nearest A=team)
  ---> nature of source (A-team, bright source, faint source, empty field)
and output a "recommended" parset from a library. Commissioners are requested to save their favourite parsets and submit to Roberto Pizzo **(***Comm):** once these are available discussions can be had about scripts. This would be useful for new users; would save time and improve communication for existing users; and be the nucleus of an overall pipeline control system for surveys.

b) clearcal problem. Strictly speaking this is not a showstopper, but the need to remember to clearcal the MS before running BBS, with the penalty for default being the wiping of the corrected-data column, continues to cause problems for even experienced users. The solution to this increases the volume of data, but this is not a short term problem if the suggestion to only keep a fraction of imaging observations is adopted.