LBWG memo 8

How do we determine which sources need subtracting?

Neal Jackson, 28.6.18

Motivation. The pipeline has an outer loop which controls the overall strategy of which sources are to be imaged, in what order, and with what phase corrections applied. The outer loop will need to be run twice, because there are two categories of source: those that interfere with the imaging of other sources on spatial scales of interest, and those that don't. The former will need to be processed in the first outer loop, subtracted using Marco's `sagecal` interface from the unaveraged data, and then the loop needs to be run again on the latter class of source.

We need to know as early as possible in the process which sources need to be subtracted. This leads to a possible chicken-and-egg problem if we adopt the approach of imaging all of the sources to determine which need to be subtracted before imaging all of the sources. Ideally we could examine some easily measurable quantity for each source, in order to determine which sources need subtracting before any imaging/selfcal takes place.

Rule of thumb from looking through lots of SPFLG: you can see bright things in the field on 50-km baselines with 50-kHz channels. They go away if you average to several hundred kHz (or 30s-1min in time). You can't see bright things in the field on 200-400km baselines unless they are really close (less than half a degree).

Measurable quantity: S:N in transformed dynamic closure phase spectra.

In principle we can use dynamic closure phase spectra (memo 4) to assess the amount of correlated signal on any source on a particular baseline, or set of baselines. Whether a source is bright enough to need subtracting depends on its correlated flux on the minimum baseline which we aim to use with appreciable weight in the final pipeline maps. In the limit of using all possible baselines, we would be back in the situation of NL-LOFAR for which source subtraction is a much more challenging problem (requiring simultaneous imaging of multiple sources).

On the assumption that the minimum baseline to be given full weight is about 80 km, representing the longer RS-RS baselines, we are interested in structures of order $5''$. Appropriate closure triangles therefore consist of ST001 and one of the more remote RS stations, RS210 or RS509 (both about 70-80 km from Exloo).

Two triangles have been tried: ST001-RS210-RS509, and ST001-RS106-RS210 (RS106 is a remote station but is very close to ST001). An additional problem appears that was not seen at longer baselines: a signal-to-noise floor appears even for sources that are not detected, at a level of about 30 for ST001-RS106-RS210, and about 8 for ST001-RS210-RS509. This floor results from highly organised phases on the short baselines.

Comparison with measure of sources which "need subtracting".

We then need a way of evaluating, in each source, whether it needs subtracting. This is done by making a dirty image, using 100 subbands, of calibrated data on the 43 sources described in memo 4. The dirty image is made with a u-v taper which gives a resolution of approximately $4''$: this dirty image is scrutinised by eye and an assessment is made of the radius from the source at which sidelobes and ripples from the source can be seen above the thermal noise. (Additional statistics - e.g. σ as a function of radius - were tried, but the eye's pattern-recognition capabilities make it a quick way of determining a

source's influence in the field.

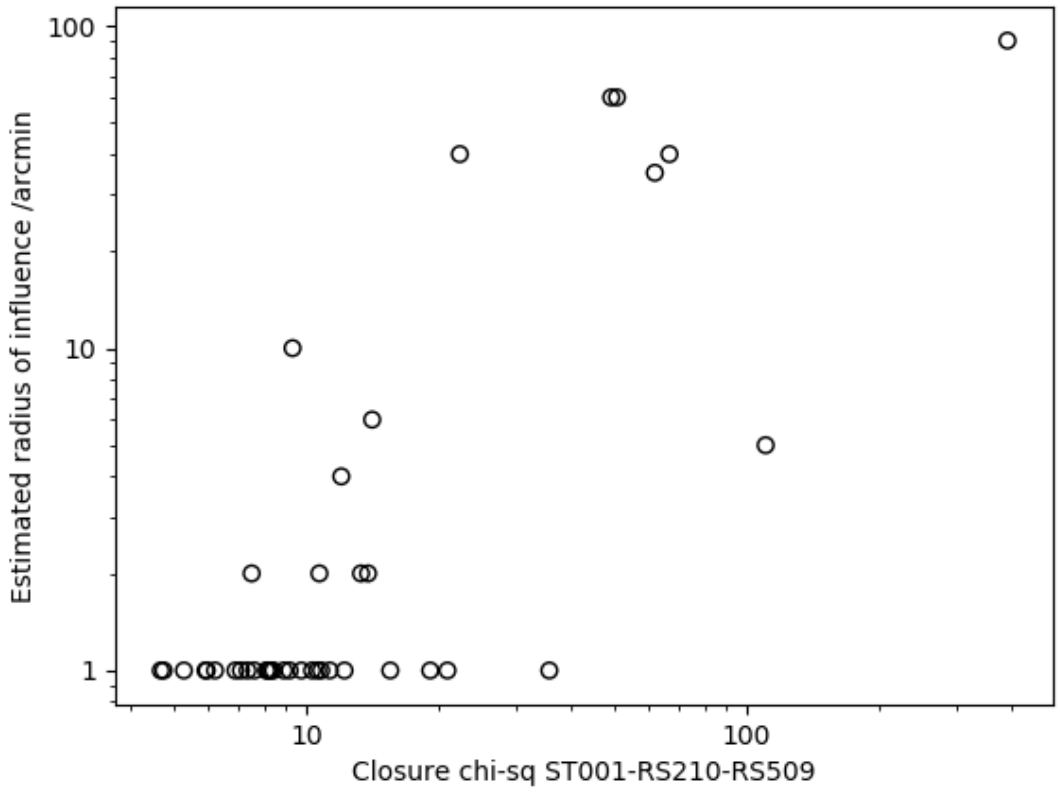

Figure 1: Estimated radius of "sphere of influence" (radius over which artefacts associated with a source can be seen in a 4-arcsec resolution dirty image) against the closure-chi parameter (Memo 4).

Figure 1 shows the two quantities plotted against each other[1]; any source which does not show visible sidelobe structure is reported as a radius of influence of 1 arcminute. There is clearly a correlation between the two quantities, with the brightest LBCS calibrator (1327+5504) showing the highest statistic in both cases.

Any source should be subtracted if its sidelobes impinge significantly on other sources which are being mapped with the long baselines. Since we are likely to be able to see several hundred sources in a field of about 12 square degrees, this implies that any source with a radius of influence of more than about 10′ should be subtracted, which corresponds to a closure scatter statistic of about 20 from Fig. 1. Although the degree of averaging in the final u-v datasets has yet to be decided, it is probable that sources within a few arcminutes will anyway be within the smearing radius and can be subtracted.

Conclusion. The current best recommendation for the criterion for subtracting a source, for an 8-hour Surveys track, is if its noise statistic (memo 4) on the ST001-RS210-RS509 closure triangle exceeds 20. This figure should be reviewed after a few fields have been

[1]The raw data for this plot, including the source names, have been uploaded to this repository as `chisqRS210.txt`.

processed. It implies the modelling and subtraction of about 5-10 bright sources per field in the first iteration of pipeline loop 1, before the other sources are attempted.