LBWG memo 10

Tests of the sagecal subtractor

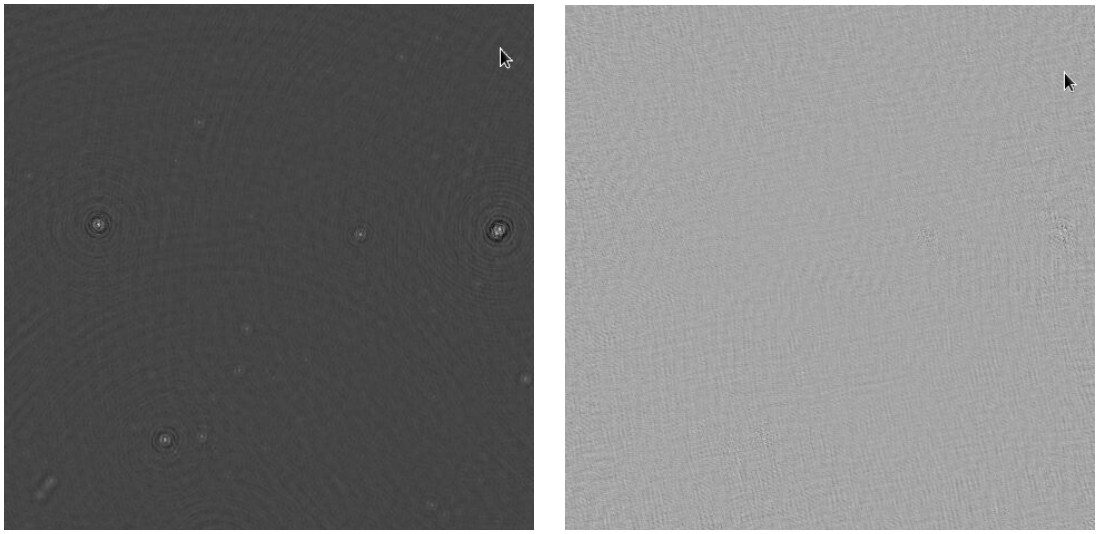
Code by Marco Iacobelli - this documentation/tests, Neal Jackson, 20.7.18

Basis of loop 1. Code now exists (written by Marco Iacobelli) that can serve as the basis for loop 1. It performs the following functions:

- Use `wsclean` to make a wide-field map of one or more subbands
- Run `pybdsf` to identify islands/clusters and make a model, and interrogate the LBCS catalogue to identify expected objects in the field
- Use `sagecal` to subtract as necessary
- Run `wsclean` again to form a residual image.

This memo is rather narrative at the moment and is a work in progress regarding preliminary testing.

In the first test a single subband of the pipeline run, using file `ndppp_prep_target` was used. The figure below shows the central 1 square degree of the P205+55 field - on the left is the image before subtraction, and on the right the post-sagecal map produced with wsclean. Both images have beams of about 3-4 arcseconds.



This looks like a very good start - sources disappear quite satisfactorily (you can just see them in the residual map).

I tried this again and all attempts to reproduce this (same software, same data) then failed for a while. I had several days of this and still do not understand why. My best guess is that I had picked up a phase-calibrated dataset the first time (I don't believe that you would get a figure like this with uncalibrated data including the IBs) and not the others.

"Failed" in this context means that the initial wsclean map had a relatively small beam and did not contain anything which produced islands in pybdsf, which then crashed the script further on. I don't think it was anything to do with the wsclean, because the command was the same for cases which worked and which didn't (you can see the wsclean command in full in the header of images it produces).

Sanity was restored by using the maxuvw-m option in wsclean, which when set to 50km excluded the international baselines and produced good maps again (with larger beams). This supports the phase calibration hypothesis, as these data have been run through prefactor, so the Dutch baselines should have reasonably good direction-independent phase calibration already. `Update: it's just produced a not-well-subtracted image however.`

So far so good. There are questions we should talk about regarding how to integrate this into the pipeline:

1. How many sources get subtracted and how are they chosen? At present this is everything that is found in the initial wsclean map.
2. What models are used for the subtraction? You might get away with using a low-resolution model which has been produced using the NL-only data, because the model probably only has to be accurate for the shorter baselines - but see 3)
3. What is the procedure for producing final maps of the bright sources which are subtracted in loop 1? These might be a substantial fraction of the good maps produced by the pipeline - if they are produced from a sagecal run which does not have a good high-res model to start with they might not be very good.
4. How wide-field a map do you need? It currently uses a radius of 1 degree, but you need more than this for the whole field. Unfortunately wsclean seems to have a very steep dependence on the number of pixels.
5. How long is it going to take? On one subband and one core it takes about half an hour (tests by Marco and me seem to agree). The sagecal step is probably parallelisable, and the wsclean step isn't.
6. How do we incorporate delay calibration into this? It has been argued previously that if we are going to use > 100 subbands, differences in delay across the field are going to matter.