



## H2020 Grant Agreement No. 730562 – RadioNet

<u>PROJECT TITLE:</u>	Advanced Radio Astronomy in Europe
<u>STARTING DATE</u>	01/01/2017
<u>DURATION:</u>	48 months
<u>CALL IDENTIFIER:</u>	H2020-INFRAIA-2016-1
<u>TOPIC:</u>	INFRAIA-01-2016-2017 Integrating Activities for Advanced Communities



### Deliverable 7.6

#### Final implementation of advanced calibration algorithms

Due date of deliverable: 2020-02-29

2020-12-07

Actual submission date:

Leading Partner: CHALMERS TEKNISKA HOEGSKOLA AB (OSO)

Document name: Final implementation of advanced algorithms  
 Type: OTHER  
 WP: RINGS - WP7.5  
 Version date: 2020-12-07  
 Authors (Institutes): T. D. Carozzi (Onsala Space Observation, Chalmers)

## Dissemination Level

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

## Index

1 Introduction.....	3
2 The Calibration Problem.....	3
3 Proposed Calibration Method.....	4
4 Results.....	6
5 Software Description.....	8

## Abstract

This document covers the deliverable D7.6 of WP7.5. It deals with an advanced calibration algorithm for radio interferometry data. In particular, an algorithm for robust self-calibration that applies to full-polarization visibility data and implements direction dependent effects (DDE) corrections has been implemented. The main part of the algorithm uses singular value decomposition (SVD) to extract source components directly from visibility (UV) data. Calibration solutions for interferometer gains are then made on these extracted SVD components. This avoids the traditional self-calibration approach of extracting clean components via imaging the data, which can introduce a corrupted image model. In this way, the extraction improves the signal-to-noise ratio, avoids some DDEs and makes the calibration more robust. It can be applied to both polarization channels and so is applicable to full-polarization calibration of visibility data. The algorithm has been implemented in a python module called `RobVisCal` which utilizes `casacore` functionality, and has been made publicly available.

COVID-19 affects the Deliverable D7.6, to which the Art.51 applies as follows:

Due to travel restrictions the advanced calibration algorithms will be slightly de-scoped. The requirement of developing direction dependent calibration and self-calibration will be fulfilled, but focusing on LOFAR LBA.

## 1 Introduction

Radio interferometry differs from optical imaging in that it does not produce images directly. Interferometry images are based on the correlation of signals from multiple receptors. Its directly measured data is called visibilities, and to produce images it typically must Fourier transform these visibilities. However, visibilities can be corrupted by incomplete knowledge of the receptor gains. *Calibration* is the process of trying to determine these gains. One important calibration technique is *self-calibration* or *self-cal* as it is known.

Traditional self-cal is done by making a clean image, using so-called CLEAN deconvolution, and using this as a model image, via an inverse Fourier transform, to fit to the measured visibilities, producing gain solutions. This procedure does not work well in low signal-to-noise ratio situations as it is not very robust. Partly, this is due to images being under-determined: images typically contain more degrees-of-freedom than the measured visibilities, which is just the number of receptors squared.

In this document, we will detail an algorithm that attempts to resolve the limitations of traditional self-cal by doing calibration directly on visibilities rather than the images. The process of obtaining model visibilities is done by SVD, rather than CLEAN images converted to visibilities.

## 2 The Calibration Problem

The core problem of radio interferometric calibration can be stated as follows. An *interferometer* consists of a number of antennas distributed over what is called an array configuration. The wave form (voltages) signal from these antenna are correlated in pairs. Together these powers are called *visibilities*. The visibilities are related to the sky *image* through a Fourier transform. A complication is that the measured visibilities are in practice corrupted in numerous ways, but fortunately many of these effects can be modeled by introducing a multiplicative factor called *gain*. Gains ultimately are the amplification of the received signal, but can also include things such as the atmospheric influence or Faraday rotation on the impinging radio waves.

The main relationship involved in gain calibration is modeled as the  $n \times n$  matrix equation

$$V = GRG^H + N \quad \text{eq. 1}$$

where  $n$  is the number of antennas,  $V$  are the measured visibilities,  $G$  are the gains,  $R$  are the true visibilities and  $N$  is the noise. The problem is to determine  $G$ . A standard approach is to assume a model for the true visibilities  $R$  and then find a  $G$  solution that to the equation. Since the noise is not known, an exact solution is not possible. A reasonable solution can anyways be found if, instead of the exact equation, one uses a fitting or minimization equation such as

$$\min \|V - GRG^H\| \quad \text{eq. 2}$$

where the  $\|\cdot\|$  brackets are some matrix norm. Most commonly the norm used is the L2 norm, meaning the components are squared and summed up. This is known as the least-squared estimate. There are various algorithms for finding a solution to this L2 minimization problem, but this mainly affects the computational efficiency rather than the solution itself. In what follows the problem formulation given here, and its solution, we will call the standard gain calibration algorithm or more succinctly: *LS calibration*. See [1] for more background.

### 3 Proposed Calibration Method

The standard LS calibration is an effective way of estimating gains. However, it suffers from the fact that it relies on having a model for the visibilities, which is of course not well known. Here we detail a different approach that avoids the need for an explicit model. The idea exploits that the calibration of radio telescopes is often done on a field-of-view (FoV) that consists of fairly strong, unresolved, point-like sources, and that these sources are statistically independent from other components in the FoV.

Mathematically, a typical calibration scenario is

$$R = ASA^H + R_b \quad \text{eq. 3}$$

where  $R$  is, as before, the uncorrupted visibilities,  $A$  is the array response vector,  $S$  is a low-rank diagonal matrix of source fluxes and  $R_b$  is the rest, i.e. background, of the visibilities. Now, putting eq. 3 into eq. 1 we get

$$V = GASA^H G^H + R_r$$

where we have put  $R_r = R_b + N$ . The singular value decomposition of  $V$ ,  $V = U \sigma U^H$  and separating out only a few of the largest singular values leads to

$$U_1 \sigma_1 U_1^H + V_r = GASA^H G^H + R_r \quad \text{eq. 4}$$

where one can identify  $V_r = R_r$ ,  $\sigma_1 = S$  and finally

$$U_1 = g \circ A \quad \text{eq. 5}$$

Here, the  $G$  matrix multiplication has been converted into an element-wise product of its diagonal components. Since the columns of  $A$  have a known form (they are the array response vector to a direction), and  $U_1$  is readily determined from the measured visibilities, eq. 5 provides a way of solving for the gains  $g$ .

The main tool is therefore the SVD of the visibilities. Since calibration is often done in short snapshot images and the dimensionality is just the number of antennas, the computation is not that

intensive. Alternatively, the approach can be seen as a principle component analysis (PCA) or also a Karhunen-Loeve transform. These two methods have been used in radio astronomy in the past [2], but only for imaging, not for calibration.

To test the proposed calibration scheme, we produced a simulation. The simulation consists of images with several point sources with user defined positions, fluxes and additive noise. From these images, visibilities were synthesized and gridded onto antenna positions. The positions are the same as those for the LOFAR LBA configuration of the Swedish LOFAR station. Likewise the frequencies were the same per the LBA: 10 to 100 MHz. Discretization was not necessary as LOFAR has enough bit range for most observing scenarios. The visibilities thus obtained are regarded as the true visibilities. Finally, these true visibilities are corrupted by multiplying them with randomly generated gains, one complex value for each antenna, resulting in the measured visibilities.

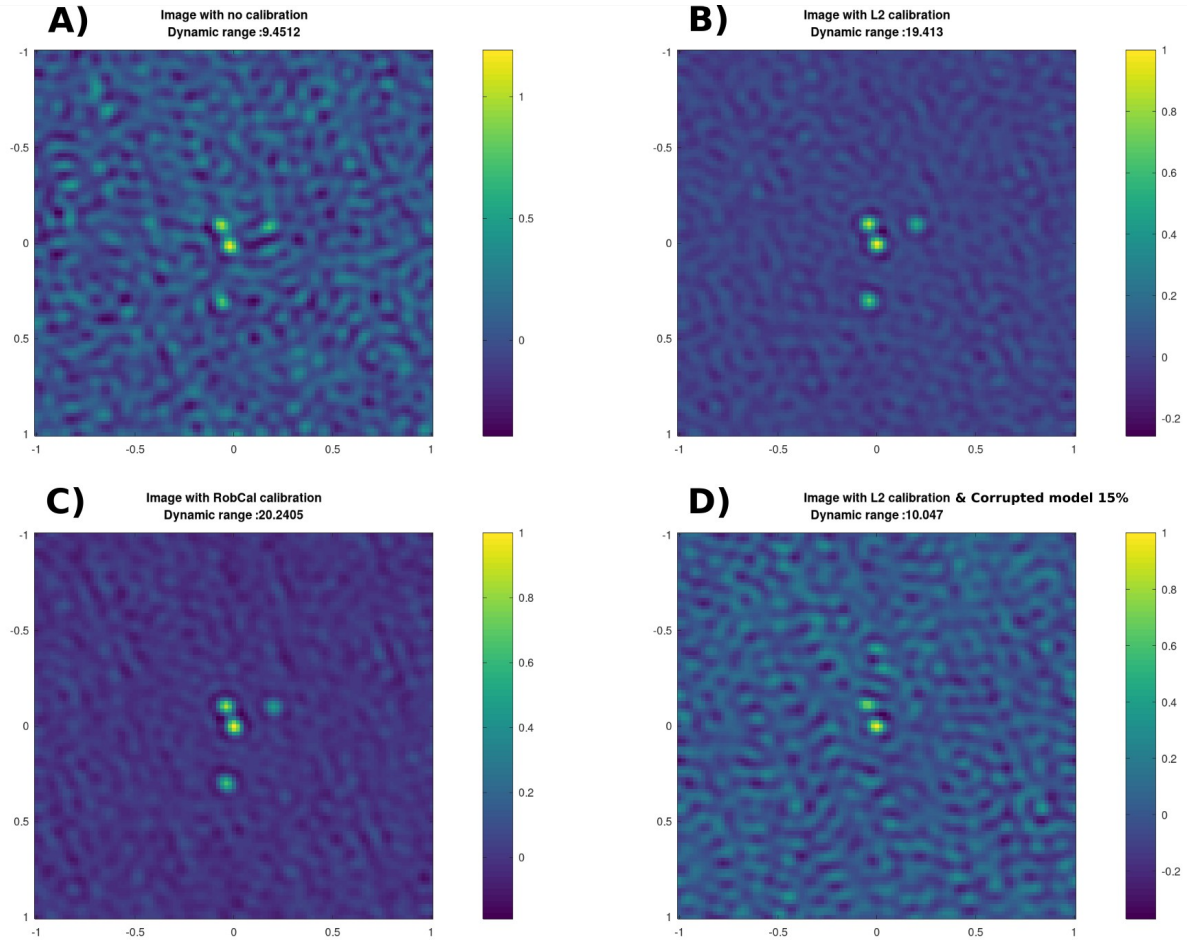
Subsequently, the proposed SVD gain calibration was performed on the measured visibilities  $V$  as follows. SVD was applied to  $V$  and the singular vector corresponding to the largest singular value was selected. This singular vector should represent the voltage response due to the strongest source in the field. The response vector is then phased-up (pointed) towards the source direction; either it is a source calibrator with a known position or the direction is determined from the vector itself. After this phase center shift the components of array response vector are all just ones,  $A=\text{vec}(1)$ . Thus, from eq. 5 the similarly source centered singular vector is equal to the gains. Using this as an estimate for the gains, they are applied to the original measured visibilities  $V$  to produce calibrated visibilities. Finally, the calibrated visibilities are imaged, using just a simple Fourier transform, known as a dirty image. The calibrated image can be compared to the original uncalibrated image, and with the image from standard LS calibration. For the LS calibration, the initially given source image was used as the model.

The results of the simulation are shown as images in Figure 1. As one can see, there are 4 sources and the image has been phased up to the strongest source. The dynamic range is intentionally low to enhance the differences and to show how the proposed calibration dubbed *RobVisCal* copes in such conditions. In the uncalibrated, directly measured image A), only 3 sources can be seen clearly, and the dynamic range is about 10. For the calibrated images B) and C), all 4 sources can be seen and the dynamic range has improved by a factor 2. In B) standard LS calibration has been applied using a perfect model, that is, the model was the true image. While in C) the proposed *RobVisCal* was applied and no model was needed (the direction to the strongest source was determined from the data itself). Finally, in D) the LS calibration was also applied, but this time an imperfect (corrupted true image) model was used resulting in an image that was not much better than the original, uncalibrated image, A).

One of the reasons why *RobVisCal* is able to calibrate in these low signal-to-noise conditions, and despite multiple other sources in the field in addition to the calibrator source, is because the SVD provides an efficient component extraction. Although not shown here, inspection of the extracted visibilities based on the singular vectors of the full visibility matrix clearly show each source separately. Since *RobVisCal* uses these extracted visibility components to do the calibration it is not affected as much by noise or multiple additional sources. This provides the necessary robustness.

The simulation shown here does not include DDEs for the sake of clarity in demonstrating how the basic *RobVisCal* performs. Adding DDEs to the calibration is however not difficult since one only needs to add them as a multiplicative (corrupting) factor to the gains  $g$  in the equations above. In particular, DDEs relating to beam gain patterns, that is the fact that an antenna's gain over the field-of-view is not uniform, has been implemented. This has been done using the *dreamBeam* software package [3], which was part of another deliverable of the RINGS project. It provides antenna beam models for various telescopes such as LOFAR. It is fully polarimetric and produces

Jones matrices (polarimeter gains) based on real, observational parameters. The beam models provided by `dreamBeam` were tested and found to be satisfactorily accurate compared to real data, see [4]. These Jones matrices can be applied to the visibilities inversely to take into account this type of DDE. `dreamBeam` can also model other types of DDEs such as ionospheric effects.



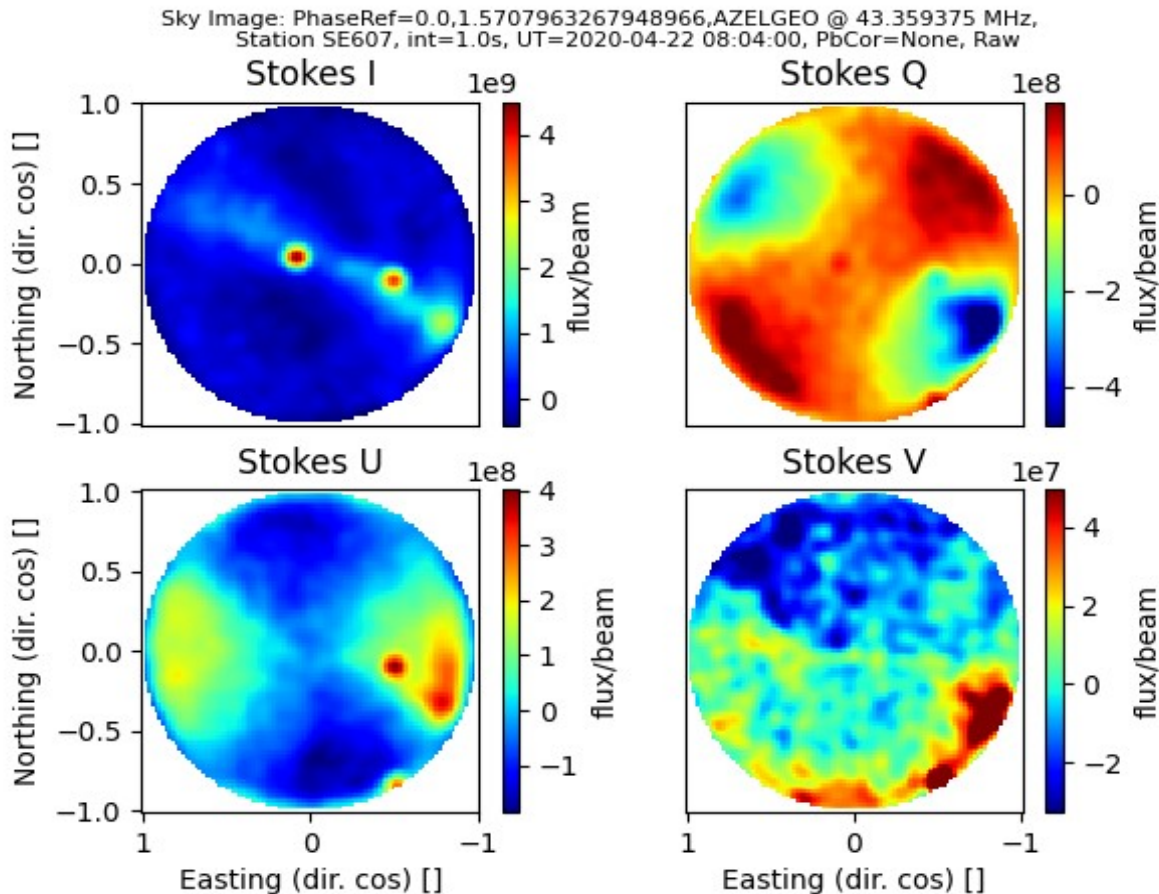
**Figure 1: Simulated images of four sources in the sky based on various types of calibration: A) Uncalibrated image, B) standard LS calibration with perfect model (model=true image), C) proposed RobCal calibration, D) LS calibration with imperfect model (image corrupted by 15%).**

Thus we find that RobVisCal is comparable to LS calibration with the caveat that the model used for LS is good. However, if the model is not that good, say with a distortion of more than 15%, then RobVisCal outperforms LS calibration. So as expected, RobVisCal is robust against ill-determined models.

## 4 Results

Having seen that simulations give promising results for RobVisCal, we turn to real data. The real data is polarized, so the RobVisCal method mentioned previously in an implicitly scalar context had to be generalized to work with the two autocorrelated polarization channels. For LOFAR this means the XX and the YY polarization channels, and specifically, RobVisCal was applied to each separately.

The real data comes from the Swedish LOFAR LBA, taken on 22 April 2020. The LBA sees the entire sky and consists of 96 dual-linear-polarized antennas and has a maximum baseline length of about 50 m. A full-Stokes image of the raw, uncalibrated data is shown in Figure 2 at a frequency of 43.4 MHz. The Stokes I (total flux) image shows Cassiopeia-A close to the center, Cygnus-A to the west of it and the galactic plane running through them. The flux levels are given in arbitrary units as no absolute flux calibration was applied. The Stokes Q and U show strong instrumental DDEs (dipole pattern).

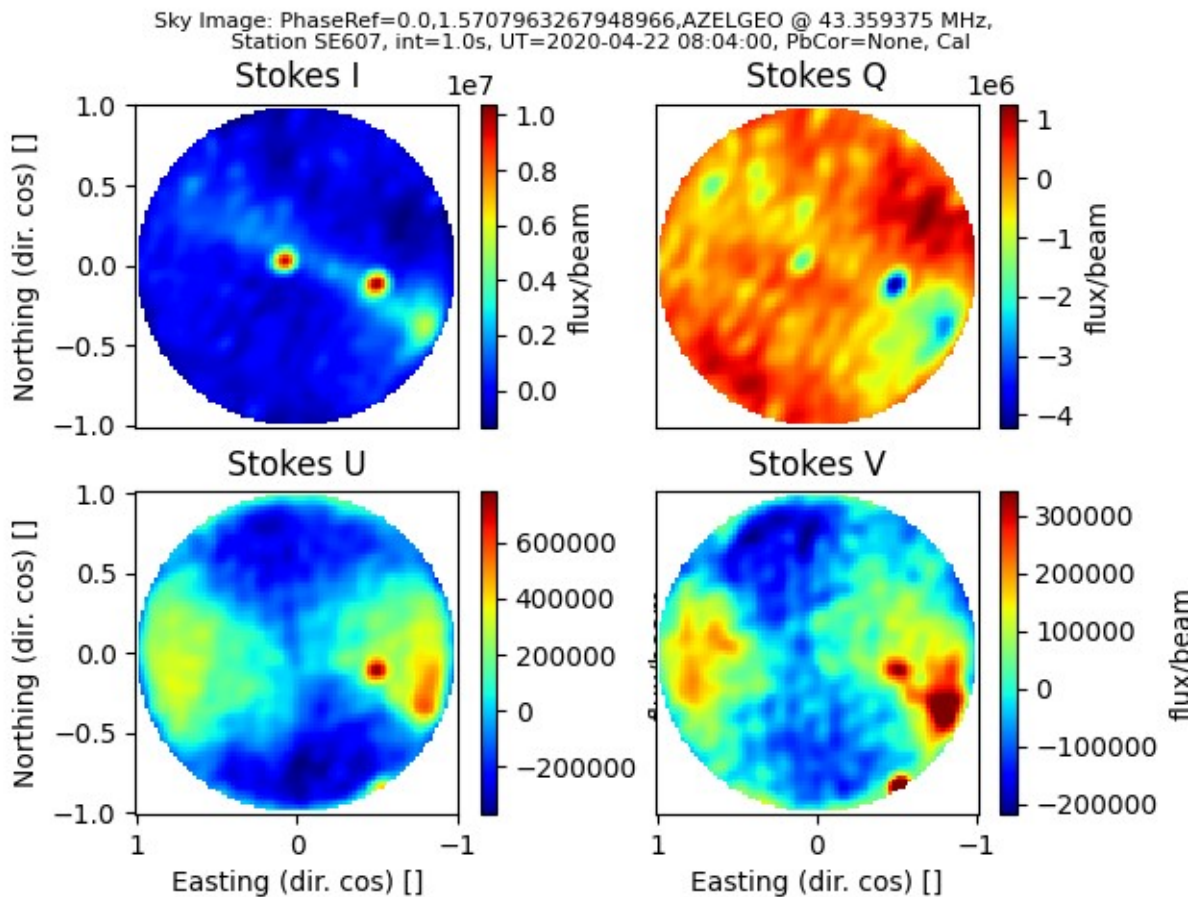


**Figure 2: Raw, uncalibrated, allsky Stokes images from the Swedish LOFAR LBA taken on 22 April 2020. Flux levels are in arbitrary units.**

After applying RobVisCal to the uncalibrated data and imaging it, the result can be seen in Figure 3. It used the same imaging tool as in the uncalibrated case, the only difference was the calibration. The dynamic range improvement in Stokes I is modest, but is still an improvement. In fact one source, Taurus-A, which is not visible in the uncalibrated case, can be seen in the RobVisCal case close to the local horizon in the North-East corner. The U and V images do not show much improvement, but this is to be expected as RobVisCal cannot be directly applied to cross-correlated polarization channels since there is no source signal there. This, in turn, is because the sources are not polarized. If there had been, RobVisCal could be applied (generalization of the hitherto discussed algorithm), so it is applicable to full-polarization. Indeed the Stokes Q image, which is based on the auto-correlated polarizations, shows a marked reduction in the instrumental polarization.

There is still a lot of instrumental polarization left as can be seen in the Stokes U and V images of Figure 3. This is clearly the effect of the dipole-like LBA antennas. Although not shown here, after applying the beam patterns from `dreamBeam` further reductions in the instrumental polarization can be made.

The results indicate that the proposed RobVisCal works with real data, in that it provides good calibration, improving dynamic range. Its performance is comparable to standard LS calibration, but again, it does not require a sky model. Sky models for low frequencies do exist, but most are for Stokes I only, and are extrapolated from observations at higher frequencies. Alternatively one would have to use cleaned images and thus self-cal.



**Figure 3:** Same as previous Figure but with the proposed calibration method applied.

## 5 Software Description

The calibration algorithm proposed here has been implemented as python3 module called `robviscal.py`.

Its dependencies are as follows. The linear algebra involved, such as the SVD, uses the `numpy` package, while the parts that involve radio astronomy specific functionality uses the `python-casacore` [5], which is a python implementation of the `casacore` software suite supported and developed by this project. For the DDEs, beam pattern models are provided by `dreamBeam`, which is part of deliverable D7.4 and described in [4].

`RobVisCal` takes fully polarized visibilities, implemented as `numpy` arrays, and the antenna configuration as inputs and generates gain estimates. No model visibilities are necessary. The module is part of a larger package called `iLiSA`, and is publicly available through for instance `github` [6]. It handles LOFAR raw data and can produce images. Detailed documentation of `RobVisCal` is included in the package.



The computational cost of RobVisCal is almost entirely down to the SVD. The size of the visibilities, in the case of LOFAR LBA, is  $2 \times 96$  squared, and one such SVD is no problem for computers nowadays both with regards to memory or time. All the results presented here were done on a laptop and were done in under 1 second.

## References

- 1: Salvini, Stefano and Wijnholds, S., "Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications" , *Astronomy & Astrophysics*, vol. 571, p. A97, , 2014.
- 2: Wijnholds, S. and van der Veen, A., "*Data driven model based least squares image reconstruction for radio astronomy*", *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.
- 3: T. D. Carozzi, "dreamBeam" , <https://github.com/2baOrNot2ba/dreamBeam>, 2020.
- 4: Creaner, O and Carozzi, T. D., "beamModelTester: software framework for testing radio telescope beams" , *Astronomy and Computing*, vol. 28, p. 100311, <https://doi.org/10.1016/j.ascom.2019.100311>, 2019.
- 5: CASACORE team, "python-casacore" , <https://github.com/casacore/python-casacore>, 2020.
- 6: Carozzi, T. D., "iLiSA" , <https://github.com/2baOrNot2ba/iLiSA>, 2020.

© Copyright 2020 RadioNet

This document has been produced within the scope of the RadioNet Project. The utilization and release of this document is subject to the conditions of the contract within the Horizon2020 programme, contract no. 730562.