Open Research Online

LOFAR Facet Calibration

Journal Item

For guidance on citations see FAQs.

oro.open.ac.uk

LOFAR FACET CALIBRATION

R. J. van Weeren[1,25], W. L. Williams[2,3,4], M. J. Hardcastle[4], T. W. Shimwell[2], D. A. Rafferty[5], J. Sabater[6], G. Heald[3,7], S. S. Sridhar[3,7], T. J. Dijkema[3], G. Brunetti[8], M. Brüggen[5], F. Andrade-Santos[1], G. A. Ogrean[1,26], H. J. A. Röttgering[2], W. A. Dawson[9], W. R. Forman[1], F. de Gasperin[2,5], C. Jones[1], G. K. Miley[2], L. Rudnick[10], C. L. Sarazin[11], A. Bonafede[5], P. N. Best[6], L. Bîrzan[5], R. Cassano[8], K. T. Chyży[12], J. H. Croston[13], T. Ensslin[14], C. Ferrari[15], M. Hoeft[16], C. Horellou[17], M. J. Jarvis[18,19], R. P. Kraft[1], M. Mevius[3], H. T. Intema[2,20], S. S. Murray[1,21], E. Orrú[3], R. Pizzo[3], A. Simionescu[22], A. Stroe[2], S. van der Tol[3], and G. J. White[23,24]

[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA; rvanweeren@cfa.harvard.edu
[2] Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands
[3] ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA, Dwingeloo, The Netherlands
[4] School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
[5] Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany
[6] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[7] Kapteyn Astronomical Institute, P.O. Box 800, 9700 AV Groningen, The Netherlands
[8] INAF/Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy
[9] Lawrence Livermore National Lab, 7000 East Avenue, Livermore, CA 94550, USA
[10] Minnesota Institute for Astrophysics, University of Minnesota, 116 Church Street S.E., Minneapolis, MN 55455, USA
[11] Department of Astronomy, University of Virginia, Charlottesville, VA 22904-4325, USA
[12] Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Kraków, Poland
[13] School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK
[14] Max Planck Institute for Astrophysics, Karl-Schwarzschildstr. 1, D-85741 Garching, Germany
[15] Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Blvd de l'Observatoire, CS 34229, F-06304 Nice cedex 4, France
[16] Thüringer Landessternwarte Tautenburg, Sternwarte 5, D-07778, Tautenburg, Germany
[17] Department of Earth and Space Sciences, Chalmers University of Technology, Onsala Space Observatory, SE-43992 Onsala, Sweden
[18] Oxford Astrophysics, Department of Physics, Keble Road, Oxford, OX1 3RH, UK
[19] University of the Western Cape, Bellville 7535, South Africa
[20] National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801-0387, USA
[21] Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[22] Institute of Space and Astronautical Science (ISAS), JAXA, 3-1-1 Yoshinodai, Chuo, Sagamihara, Kanagawa, 252-5210, Japan
[23] Department of Physical Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
[24] RALSpace, The Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0NL, UK

ABSTRACT

LOFAR, the Low-Frequency Array, is a powerful new radio telescope operating between 10 and 240 MHz. LOFAR allows detailed sensitive high-resolution studies of the low-frequency radio sky. At the same time LOFAR also provides excellent short baseline coverage to map diffuse extended emission. However, producing high-quality deep images is challenging due to the presence of direction-dependent calibration errors, caused by imperfect knowledge of the station beam shapes and the ionosphere. Furthermore, the large data volume and presence of station clock errors present additional difficulties. In this paper we present a new calibration scheme, which we name facet calibration, to obtain deep high-resolution LOFAR High Band Antenna images using the Dutch part of the array. This scheme solves and corrects the direction-dependent errors in a number of facets that cover the observed field of view. Facet calibration provides close to thermal noise limited images for a typical 8 hr observing run at ~5″ resolution, meeting the specifications of the LOFAR Tier-1 northern survey.

Key words: techniques: interferometric

1. INTRODUCTION

LOFAR is a powerful new radio telescope operating between 10 and 240 MHz (van Haarlem et al. 2013). It is designed to carry out a range of astrophysical studies in this relatively unexplored part of the radio band. LOFAR consists of antenna dipoles that are grouped into stations. LOFAR stations are located in various countries in Europe, with the majority being in the northeast of the Netherlands.

Two different dipole antenna types are used. The High Band Antennas (HBAs) cover the 110–240 MHz range and the Low Band Antennas (LBAs) cover the 10–90 MHz range. At the stations, the dipoles signals are combined digitally into a phased array. The signals from these stations are then sent via high-speed fiber to a central GPU correlator where they are correlated with those from other stations to form an interferometer. The electronic beam-forming at the station level allows the generation of multiple beams on the sky, which, together with the large field of view (FOV) at these low-frequencies, makes LOFAR an ideal survey instrument. For more details about the instrument we refer the reader to the overview paper by van Haarlem et al. (2013).

One of the major goals of LOFAR is to carry out a survey of the northern sky as part of the LOFAR Surveys Key Science Project (Surveys KSP; Röttgering et al. 2006). For the HBA part of the survey, the aim is to reach a depth of ~0.1 mJy beam^{-1} at a resolution of ~5″ (i.e., the Tier-1 survey depth, Röttgering et al. 2011) across the entire northern sky.

Over the last years it has become clear that, to reach this depth, advanced calibration and processing techniques are needed.

One of the main challenges for the calibration is the ionosphere (e.g., Lonsdale 2005; Intema et al. 2009). The ionosphere results in delay differences between antenna stations causing errors in the phases of the measured visibilities. The amount of phase change is directly related to the free electron column density along a line of sight through the ionosphere and the observing frequency. These ionospheric phase errors thus change with the viewing direction, and if the array is large, as is the case for LOFAR, differ from station to station. The phase errors result in shifting, deformation, and splitting of sources in the image plane, which, if not corrected, cause deconvolution artifacts and an increase in the overall image noise.

Another challenge concerns the complex time-varying station beam shapes. The reason for the time-variation is that the stations have no moving parts. Sources are tracked by adjusting the delays between the dipole elements when the sources move across the sky. In addition, small differences in station beam models and the actual station beam shapes cause errors that need to be corrected for in order to produce high-quality images, in particular for bright sources.

In this work we present a new calibration scheme, facet calibration, that has enabled us to make images that reach the LOFAR HBA Tier-1 survey depth and resolution. Facet calibration was developed to process a LOFAR HBA observation of the "Toothbrush" galaxy cluster and we will use this particular data set to lay out the method. The scientific results on this galaxy cluster are presented in van Weeren et al. (2016).

Facet calibration builds upon the "peeling" technique (e.g., Noordam 2004), where calibration solutions in a discrete number of directions are obtained, similar to other low-frequency calibration schemes (e.g., SPAM, Sagecal, and MeqTrees; Intema et al. 2009; Noordam & Smirnov 2010; Kazemi et al. 2011; Smirnov 2011a).

The layout of this paper is as follows. We start with a description of the observations and general characteristics of a typical LOFAR data set in Section 2. We provide an overview of the direction-independent calibration and processing in Section 4. The corrections for direction-dependent effects (DDE) are described in Section 5. We end with a discussion and conclusions in Sections 6 and 7.

2. LOFAR HBA OBSERVATIONS

Below we describe the setup of the Cycle 0 Toothbrush cluster observations, which will serve as a reference data set to demonstrate the facet calibration scheme.

The Toothbrush cluster was observed on 2013 February 24, mostly during nighttime with the LOFAR HBA stations. Two station beams were formed: one on the target and one on the nearby (8°3 separation) calibrator 3C147. By default, all four correlation products were recorded and the frequency band was divided into subbands, each 195.3125 kHz wide. Each subband was further divided into 64 channels. The integration time was set to 1 s to facilitate the removal of radio frequency interference (RFI). Complete frequency coverage between 112 and 181 MHz was obtained on the Toothbrush field, while

---

Table 1
HBA Observations

Observation IDs	L99083, L99084
Pointing center[a]	$06^h03^m33^s5$, $+42°19'58''5$
Pointing center[b]	$05^h42^m36^s1$, $+49°51'07''0$
Integration time	1 s
Observation date	2013 Feb 24
Total on-source time	10 hr
Used on-source time	8.8 hr
Correlations	XX, XY, YX, YY
Frequency setup[a]	112–181 MHz full coverage
Frequency setup[b]	112–181 MHz, 121 subbands[c]
Used bandwidth on target	120–181 MHz, except 169–171, 177–179 MHz[d]
Bandwidth per subband	195.3125 kHz
Channels per subband	64

Notes.
[a] Toothbrush field.
[b] 3C147.
[c] The subbands are approximately evenly distributed within this frequency range.
[d] These frequency ranges are affected by strong RFI.

---

the calibrator 3C147 was covered with 121 subbands[27], randomly spread between 112 and 181 MHz. An overview of the observations is given in Table 1.

For the observations 13 remote and 21 Dutch core stations were used, giving baselines that range between 68 m and 80 km. The large number of short baselines is important to image diffuse extended emission. The core station layout is different from that of the remote stations, as the latter are split into two sub-stations, each with 24 dipole tiles and a diameter of 30.75 m. The HBA_DUAL_INNER configuration was employed for the remote stations (van Haarlem et al. 2013) meaning that only the inner 24 tiles of a remote station are used. This is done to obtain similar station beam sizes as for the split core stations. Thus, in total $21 \times 2 + 13 = 55$ stations were correlated for this observation. The international ("non-Dutch") stations were not included. For the stations, the half-power beam width (HPBW) is about 3°8 at 150 MHz. The monochromatic uv-coverage for the Toothbrush field observations is displayed in Figure 1.

3. HBA DATA REDUCTION AND CALIBRATION: AN OVERVIEW

As mentioned in Section 1, advanced calibration and processing techniques are needed to obtain deep high-fidelity images at low radio frequencies.

The Toothbrush observation was the first Cycle 0 observation of the LOFAR Galaxy Cluster Working group, which is part of the LOFAR surveys Key Science Project (Röttgering et al. 2006). Therefore, the observation also served as a testbed to develop calibration strategies to reach the required Tier-1 survey depth. Below we provide an brief overview of the calibration method that was developed to reach the required depth and resolution, before we go into more detail in the next sections.

The data reduction and calibration consists of two main components: a non-directional and a directional part. The non-directional part includes the following steps: (1) removal of RFI, (2) bright off-axis source removal, (3) averaging, (4) solving for the calibrator complex gains, (5) "clock-total

---

[27] This number is constrained by the total amount of bandwidth that can be handled by the system.
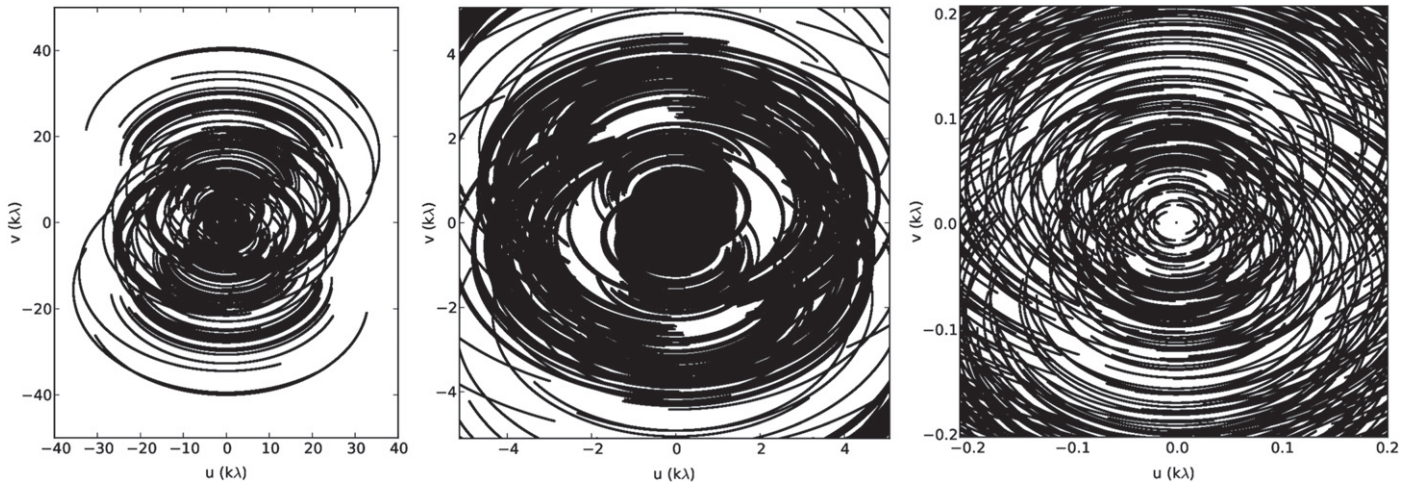
Figure 1. Monochromatic uv-coverage for the Toothbrush field at 150 MHz. The middle and right panels progressively zoom inwards, showing the dense inner uv-coverage. The large fractional bandwidth fills the uv-plane radially (not shown in the figure).

electron content" (clock-TEC) separation on the calibrator, (6) transfer of the amplitudes and clocks from the calibrator to the target field, and (7) amplitude and phase (self)calibration of the target field at medium (20″–30″) resolution. This is then followed by a scheme to obtain direction-dependent corrections to reach near thermal noise limited images using the full resolution offered by the longest "Dutch-LOFAR" baselines of about 10^2 km. Below, the non-directional and directional parts of the calibration are described in more detail.

4. NON-DIRECTIONAL REDUCTION AND CALIBRATION

4.1. RFI Removal

The first step in the reduction of the HBA data consisted of removal of RFI with the `AOFlagger` (Offringa et al. 2010, 2012). The amount of data affected by RFI was typically only a few percent. One malfunctioning core station was flagged entirely. In addition, the first and last three channels of each subband were also flagged, as they were noisy. After flagging, the data were averaged to 5 s and 4 channels per subband to reduce the data size.

4.2. Removal of Sources in the Far Sidelobes

Cas A and Cyg A are sufficiently bright that they can contribute flux through the sidelobes of the station beam. We therefore computed the contribution of Cas A and Cyg A using the station beam model (Hamaker 2011; van Haarlem et al. 2013). Whenever the apparent flux density of these sources exceeded 0.5 Jy for the core-core baselines, we obtained gain solutions toward these two sources using the `BlackBoard Selfcal` (BBS) software (Pandey et al. 2009). In addition, simultaneous gain solutions were obtained toward 3C147 or the Toothbrush Field. We solved for the gains using all four correlations (XX, XY, YX, YY). For 3C147 we took a point source model. For the Toothbrush Field the model was derived from a GMRT 150 MHz image (presented in van Weeren et al. 2012a) with the `PyBDSM` (Mohan & Rafferty 2015) source detection package. The Cas A and Cyg A models came from ∼10″ resolution LOFAR low-band antenna observations. We assumed that all sources are unpolarized. After solving for the gains, Cas A and Cyg A were subtracted from the data with the

appropriate gain solutions. The data were then averaged to 10 s and 2 channels (97.6562 kHz per channel) per subband for the target field and 5 s and 1 channel (195.3125 kHz) for 3C147; no gain corrections were applied. For 3C147, bandwidth smearing is not an issue, as it is located in the phase center and dominates the flux in the field. These data serve as input for the rest of the processing. The target field averaging parameters are a compromise between bandwidth and time smearing and finite computing capabilities.

4.3. Obtaining Calibration Solutions toward the Primary Calibrator

We obtained diagonal (XX and YY) gain solutions toward 3C147 with `BBS`, solving on a timescale of 5 s per subband basis. Besides solving for these parallel-hand gains, we also solved for a rotation angle (β) per station to take into account differential Faraday rotation, which could otherwise affect the parallel-hand amplitudes in the linear correlation basis. The corresponding Jones matrix (F) for the differential Faraday rotation (Smirnov 2011a) is

$$F = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}. \qquad (1)$$

We assume that β is constant within a single subband, which is a valid approximation in the HBA frequency range. The derived rotation angles were small ($\beta \ll 1$ rad), indicating little differential Faraday rotation, which is not unexpected in the HBA frequency range. The LOFAR station beam was applied during the solve so that gain variations caused by the changing station beams are not absorbed into the gain solutions. For the point source model of 3C147 we used the flux-scale of Scaife & Heald (2012), giving a flux density of 66.7 Jy at 150 MHz.

4.4. Clock-TEC Separation

The remote LOFAR stations have their own clocks (to timestamp the data before correlation) that are not perfectly synchronized with the single clock that is used for all the core stations. This causes a strong phase delay across the frequency band (phase $\propto \nu$) for the remote-remote and core-remote baselines. For our observation, very large clock offsets (delays)

were present for some of the remote stations, up to 200 ns (with respect to the core). In addition to these clock offsets, the clocks for the remote stations can drift by ±15 ns over the course of an observation with respect to the core station clock. The large clock offsets need to be corrected (see Section 5), before we can proceed with the directional part of the calibration.

We developed a method to derive the clock values from an observation of a bright calibrator source. In our case this was 3C147, which has a large enough flux contribution on all baselines to obtain high signal-to-noise ratio (S/N) gain solutions on 5 s timescales. We will refer to this method as "Clock-TEC separation." Once the clock values are determined we can correct the target field data for this effect with BBS. Removing the time varying clock is equivalent to applying a frequency and time-dependent phase correction to each station. In the next paragraphs we outline this method.

A complication for low-frequency radio observations is that the phases also vary because of the ionosphere and that these phase variations are not linear across the observed bandwidth. When we are observing far enough from the ionospheric plasma frequency (which is typically located around ~10 MHz) the effect has a frequency dependence of: phase $\propto \nu^{-1}$. The amount of phase change is directly related to the free electron column density along a line of sight through the ionosphere. The unit for TEC is 10^{16} m^{-2}. For an interferometer (and neglecting polarization) only the differences in TEC (and clock) are relevant. The observed phase difference for a baseline can be thus written as

$$\Delta\mathrm{phase}(\nu, t) = 2\pi p_0(t)\nu - \frac{8.448 \cdot 10^9 p_1(t)}{\nu}[\mathrm{rad}], \quad (2)$$

where p_0 is the clock difference and p_1 is the TEC difference (the constant of $8.448 \cdot 10^9$ relates TEC to phase). In principle we could directly solve for the clock and TEC differences on the visibility data, employing the fact that the phase frequency dependences of the clock and ionosphere are different. This requires that there is sufficient frequency coverage and/or signal-to-noise to detect the nonlinear behavior (the p_1/ν term) of the ionosphere in Equation (2). We decided not to directly solve Equation (2) on the visibility data, because this would require including all 121 subbands for the solve. This is computationally expensive and that also made it less practical for testing the clock-TEC separation approach.

Instead of directly solving for clock and TEC on the visibility data, we use the 3C147 phase solutions. We thus have 121 phase solutions along the frequency axis as input to fit clock and TEC difference via Equation (2) for every 5 s time slot. We do this fitting in python using scipy.optimize.leastsq. A difficulty with the this least-squares solve is that Equation (2) has many local minima in χ^2 space and the actual global minimum is not much deeper than neighboring local minima. In other words, our frequency coverage is not wide enough to easily separate the nonlinear behavior of the TEC from the clock, unless good initial guesses are available.

We use a brute force search on a grid to find clock and TEC initial guesses that are close enough to the actual solutions to achieve convergence. As expected, for the core stations the clock differences were close to zero (less than 1 ns for most stations); see Figure 2. The fitting was done separately for the XX and YY phases. For a few stations, we found a small but

constant offset between the XX and YY phases. We determined these offsets by taking the median phase difference between the XX and YY phases over the entire length of the observation for each station.[28] After the offsets were taken out, the resulting clock and TEC values were averaged and smoothed with a running median filter (with a window size of 15 s); see Figures 2 and 3.

Based on the TEC solutions, we decided to flag the first 1.2 hr of the data since the TEC differences were very large and changed rapidly over time. During this time period the phase rate exceeded more than 1 rad per 10 s on the longest baselines. To avoid time decorrelation we discarded the first 1.2 hr of data for the rest of the processing.

4.5. Clock and Amplitude Transfer

The clock values and XX–YY phase offsets were transferred, together with the amplitudes, to the target field data. The amplitudes were also inspected for outliers and smoothed along the time and frequency axis with a running median filter (with a window size of 5 minutes along the time axis and 3 subbands along the frequency axis). After applying these corrections, the resulting target field data were free of clock delays and the visibility amplitudes were in units of jansky. We did not transfer the TEC values because these are direction-dependent and differ for the calibrator and target fields.

For the LOFAR observations discussed in this work, simultaneous calibrator observations were available for the entire length of the observing run. However, this is not required and a short ~10 minutes observation of a primary calibrator source at the start or end of the main observing run is also sufficient to transfer the clock and amplitudes. The reason for this is that the amplitudes are usually stable over the length of an observation, and the clock drifts remain within $\pm\sim15$ ns of the global offsets. The remaining clock drifts are small enough to be corrected for at a later stage (during the self-calibration or the direction-dependent calibration).

### 4.6. Self-calibration of the Target Field Data

We combined the clock and amplitude corrected individual subbands of the target field into groups of 10. The resulting combined data sets thus have about 2 MHz bandwidth and 20 channels. It is assumed that the phase change across 2 MHz bandwidth is small enough to be neglected during the calibration. As can be judged from Figure 4 (green lines) this a reasonable assumption since the phase change due to the TEC is much smaller than 1 rad across a 2 MHz band. The concatenation of 10 subbands also increases the S/N for the calibration by a factor of about $\sqrt{10}$. We performed another round of RFI removal with AOFlagger on the combined 10 subband data sets as this allowed the detection of lower-level RFI. Typically, 95% of the target field data remained after this second round of RFI excision. We discarded the data between 112 and 120 MHz, as signal to noise was relatively poor there compared to the data above 120 MHz.

We phase-calibrated these 10 subband data sets against the GMRT 150 MHz model described in Section 4.2, which has a resolution of $26'' \times 22''$ and an rms noise that varies mostly between 1 and 2 mJy beam^{-1}. The phase calibration was carried out on a 10 s timescale to avoid phase decorrelation by

---

[28] We do not know the origin of these XX–YY phase offsets.
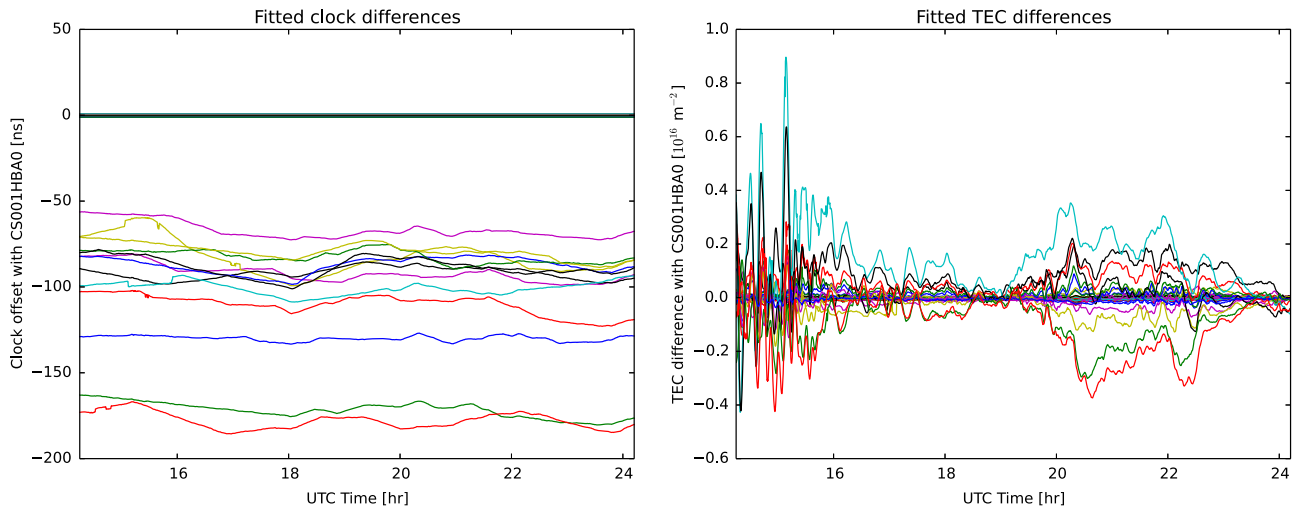
Figure 2. Fitted clock and TEC differences based on the 3C147 phase solutions. The fitting was performed on the solution interval timescale of 5 s. The resulting values were smoothed with a running median filter with a local window size of 15 s. Left: fitted station clock offsets as function of time with respect to the reference (core) station CS001HBA0. The clock values show a bimodal distribution with all the core stations having clock values close to zero, as is expected, since the core stations operate on a single clock (van Haarlem et al. 2013). The remote stations all have large negative clock offsets with respect to the core stations, with the largest offset approaching 200 ns. The fact that all remote stations have negative offsets indicates that it is the core station clock that has quite a large offset with respect to the average remote station clocks. The changes in the slope of the clock drifts for the remote stations clocks are caused by adjustments of the clock rate. The clock rate adjustments are based on a comparison with a global positioning system (GPS) signal. These adjustments are made to prevent the clocks from drifting more than $\sim\pm15$ ns. Right: TEC values for the same stations. Differences of up to 1 TEC unit are observed, which indicates a very active ionosphere, in particular for the first ~1 hr of the observation where the TEC values change very rapidly (more typical values are $\lesssim0.2$ TEC units).
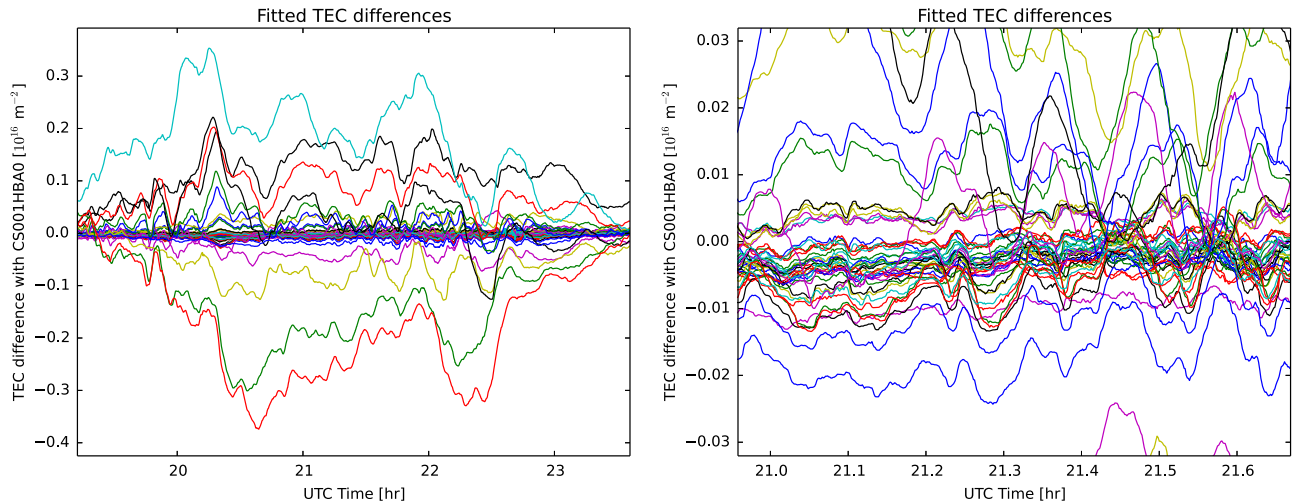

Figure 3. TEC values with respect to reference station CS001HBA0. These two figures provide zooms for Figure 2 (right panel). The "band" of lines with TEC values close to zero represents core stations that see more or less the same ionosphere as the reference station. The large TEC values and rapid changes require corrections on short (10 s) timescales to prevent ionospheric "blurring" of the image.

the ionosphere. A spectral index of -0.8 was assumed to scale the GMRT model to the different frequencies. The S/N in the 10 subband data sets was high enough to obtain good quality phase solutions for all stations, i.e., we could easily track the phase solutions over time. The station beam model was used when computing the sky model. We then imaged these 10 subband data sets with CASA and corrected for the effect of the station beams in the phase center. This does not allow for a full proper beam correction across the FOV when imaging but is enough to allow for subsequent self-calibration cycles.

From these CASA images we created a new sky model with PyBDSM. This model is an apparent sky model, as it is not fully corrected for the station beams (except in the phase center). We carried out an amplitude and phase (self)calibration against this model and again re-imaged the data. All the imaging was done

with an outer uv-range cut of 7 kλ, limiting the resolution to $\sim25''$, a FOV of $\approx13°$, and Briggs (1995) weighting (robust = 0). The resolution and FOV imaged are a compromise between the accuracy of the sky model obtained and processing speed. For the solving step we included all available baselines. W-projection (Cornwell et al. 2005, 2008) was employed to deal with the non-coplanar nature of the array. Furthermore, during the imaging clean masks were used, with the mask derived from a previous imaging run without mask. The clean masks were made with PyBDSM, detecting islands of emission with a $3\sigma_{\rm rms}$ island threshold and a pixel threshold of $5\sigma_{\rm rms}$. A box size of 70×70 pixels was used to compute the locally varying rms across the maps to take into account (calibration) artifacts around strong sources.
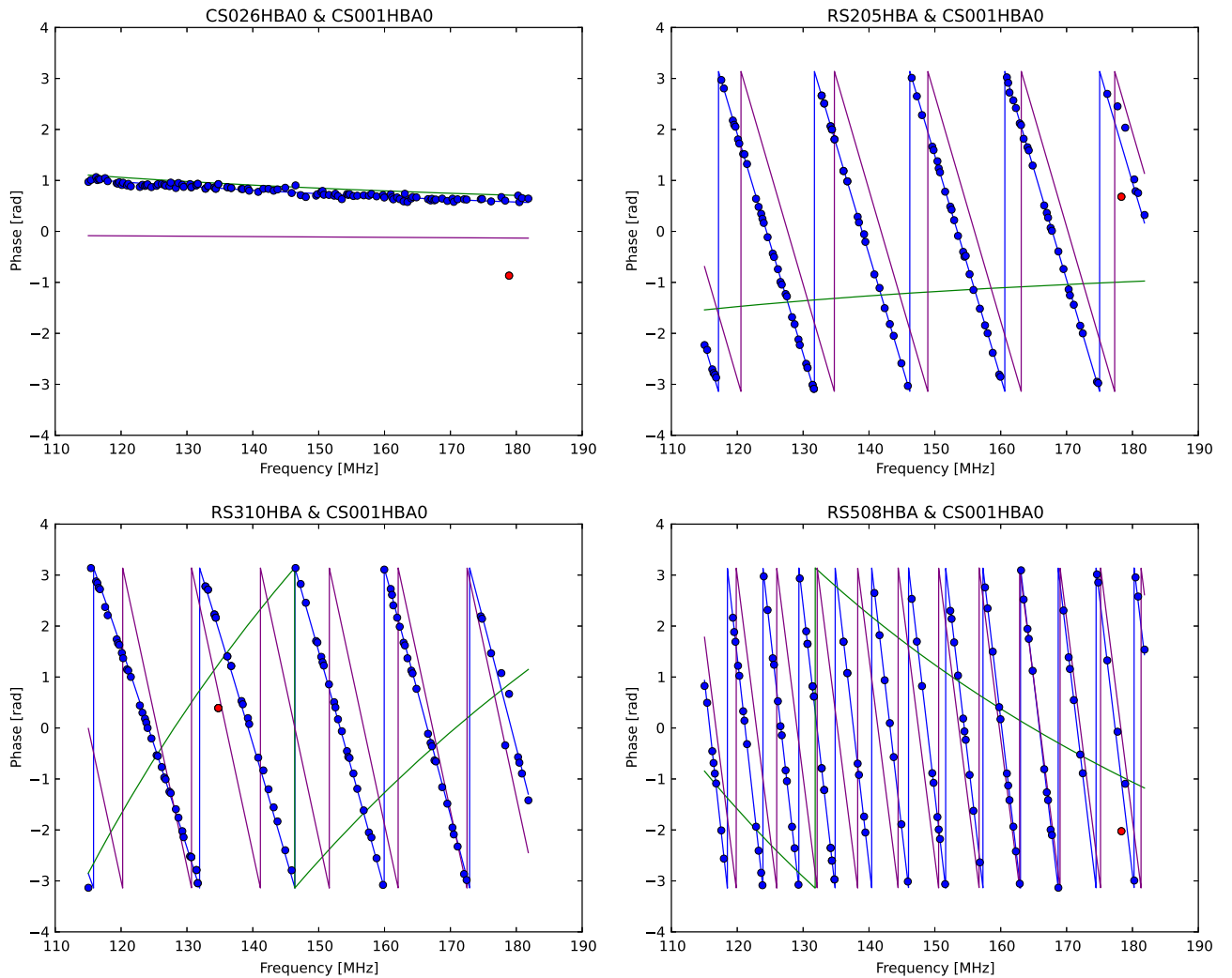
Figure 4. Examples of the Clock-TEC fitting for one core and three remote stations. The distances to the reference station CS001HBA0 are 1.0, 6.4, 51.8, and 37.1 km for CS026HBA0, RS205HBA, RS310HBA, and RS508HBA, respectively. The blue points show the 3C147 phase solutions as a function of frequency at an arbitrary 5 s time slot. Red points are outliers that were iteratively rejected during the fitting. These bad solutions are usually caused by subbands that are strongly affected by RFI. The blue line is the fitted model to the data. The separate contributions of the TEC and clock components to fitted phases are shown with green and purple lines, respectively. From these fits it can be seen that the very large clock offsets for the remote stations cause the majority of the phase changes across the HBA band. For the core station CS026HBA0 the fitted clock contribution is much smaller than for the remote stations. This is expected, as all core stations operate in a single clock. The TEC component for CS026HBA0 is small as well, because the core stations are close enough together that they see more or less the same ionosphere.

The products of the above steps are apparent-flux "medium resolution" images of the sky between 120 and 181 MHz in steps of about 2 MHz (Figure 6). From this process we obtained a total of 29 of these images. Two blocks of 10 subbands were excluded because they were affected by strong RFI; see Table 1.

4.6.1. Accuracy of the Flux-scale Bootstrapping

To assess the accuracy of the transfer of the flux-scale from the calibrator 3C147 to the target field, we extracted catalogs from the GMRT and medium resolution LOFAR image at 150 MHz with PyBDSM. In this case, the HBA image was produced with awimager (Tasse et al. 2013), which fully corrects for the time-varying primary beam. We used the same box size of 70 × 70 pixels to compute the rms map. We cross-matched the catalogs using a matching radius of 5″. Sources with a S/N less than 10 or that were significantly resolved (a ratio of integrated flux over peak flux larger than 2) were excluded.

The ratio of the HBA over GMRT integrated flux densities is shown in Figure 5. The ratio seems to be constant as a function of radial distance to the pointing center. From the medium flux ratio we determine a difference in flux scale of about a factor 0.8 between LOFAR and the GMRT. Such differences are not unexpected given the uncertainties in the calibration of the low frequency flux-scale with the GMRT (which is estimated to be about 10%, Intema et al. 2011) and the complex LOFAR station beams (van Weeren et al. 2012b, 2014). Recently, it has been found that the normalization of the LOFAR HBA beam was not correctly implemented in the models (work is ongoing to fix this issue). When transferring a flux-scale from one pointing to another, differences of the order of what we find are expected (based on work carried out as part of the MSSS survey, Heald et al. 2015). We therefore suspect that most of the difference with the GMRT flux-scale is related to this LOFAR beam model issue. However, at the moment it is not yet possible to calculate what the expected error is for our specific observation. We take a conservative approach here and
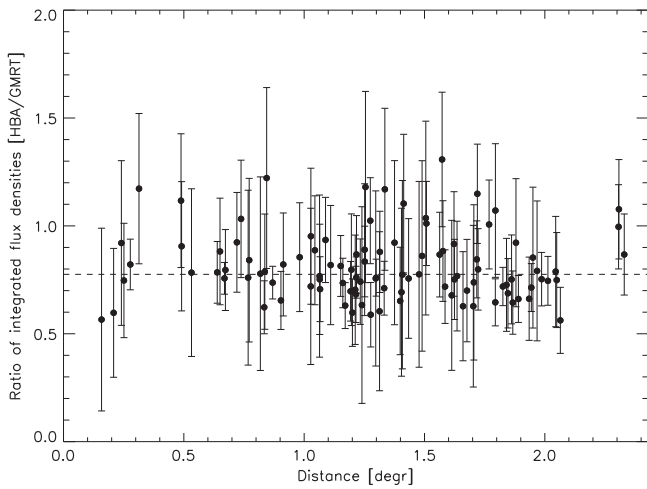
Figure 5. Integrated flux density ratios (LOFAR HBA over GMRT) for sources in the Toothbrush field at 150 MHz as a function of distance to the pointing center. The dashed line indicates the median flux density ratio.

assume that the GMRT flux-scale is correct and use a scaling factor of 1.2 to correct our extracted flux densities from the LOFAR HBA maps.

4.7. Subtracting the Sources from the Data

As a final step, we subtracted the sources from the 10 subband data sets. This was done using a two step approach. First we subtracted the clean components from the medium resolution images (with the corresponding gain solutions). We then re-imaged this clean component subtracted data at a lower ~1'.5 resolution using an outer uv-range cut of 2 kλ, but now with a FOV of ≈33°. This allowed us to detect sources in the first and second sidelobes as well as extended low-surface brightness emission in the field that was not cleaned in the medium resolution imaging. The clean components found in the low-resolution image were again subtracted with the direction-independent self-calibration solutions. The reason for imaging this far out is to (i) remove the contribution of sources in the far sidelobes, so they do not influence the direction-dependent calibration and (ii) search for the presence of "off-axis" sources that are bright enough that they need to be included in the direction-dependent calibration (often 3C sources).

Completing the above steps, we are left with 29 2-MHz wide data sets with all sources subtracted and corresponding sky models. The 29 sky models take care of the frequency dependence of the sky at this point. Note that these data sets do not have the gain solutions applied; rather, the clean components were "corrupted" with the gain solutions and subtracted from the uncorrected data. These "empty" data sets serve as input for the direction-dependent calibration scheme.

5. CORRECTION FOR DDE: "FACET CALIBRATION"

After self-calibration, significant artifacts remain around (bright) sources in LOFAR images; see Figure 6. Also the rms noise levels in the 10 subband medium resolution images are a few mJy beam^{-1}. This is a factor of 5–10 higher than the expected thermal noise with these imaging settings. The increased noise and artifacts are caused by DDEs, namely the station beam and the ionosphere.
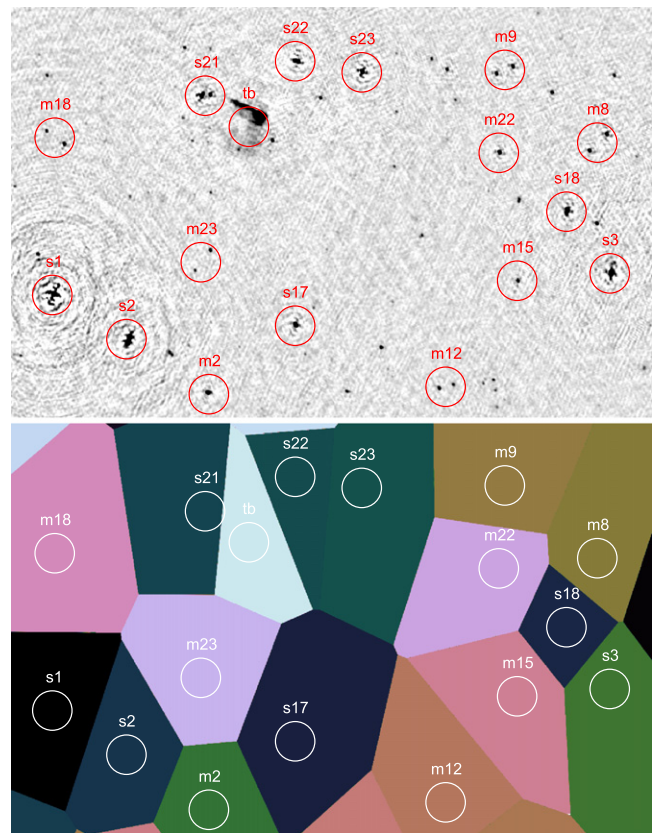
Figure 6. Example of the calibration directions on top of a 150–152 MHz image (26″ resolution) displaying a region around the Toothbrush cluster (top panel). Note that calibration artifacts are visible around the brighter sources, as only self-calibration has been performed and no DDE calibration has been applied; see Section 4.6. The FOV shown measures 2°.1 × 1°.4. The center of each circle defines a direction toward a bright radio source (group). Direction m9 is an example of a "group." Based on these directions the sky is partitioned via a Voronoi tessellation scheme (bottom panel). The different colors associated with the regions are arbitrary and just for visual representation.

Previous work has shown that gain corrections toward ~10^2 directions are needed to correct for these errors (Yatawatta et al. 2013). A concern here is that this direction-dependent calibration involves solving for many parameters. Ideally one should try to keep the number of degrees of freedom (dof) in the calibration small with respect to the number of independent measured visibilities. There is also the requirement of having enough S/N in each direction to enable an accurate calibration. The accuracy and completeness of the sky model to calibrate against, is also important. This is particularly relevant for LOFAR, because no high-resolution models of the sky are available at these frequencies. For a more detailed discussion on these various topics see Intema et al. (2009), Smirnov (2011b), Kazemi & Yatawatta (2013), Kazemi et al. (2013), and Yatawatta (2015). To keep the number of dof small during the calibration and achieve sufficient S/N, we make the following (reasonable) assumptions:

1. the station beams vary slowly over time and frequency (Yatawatta et al. 2013);
2. differential Faraday rotation can be neglected in the HBA band (unless extremely high dynamic range is required), meaning that the ionosphere affects the XX and YY phases in the same way;

3. the frequency dependence of the phase is: phase $\propto \nu^{-1}$ (note that the effects of the clocks have been taken out);

4. no other calibration errors are present besides ionosphere and beam;

5. all the DDE vary smoothly across the FOV (e.g., Cohen & Röttgering 2009).

The last assumption is particularly relevant, as it implies that we can divide up the sky into a number of "isoplanatic patches" (e.g., Schwab 1984).

In the next subsections we outline the direction-dependent calibration scheme, which we will refer to as "facet calibration." The scheme has some similarities to SPAM (Intema et al. 2009; Intema 2014) and Sagecal (Yatawatta et al. 2008; Kazemi et al. 2011), although there are also a few important differences (see Section 6.2). Another technique that was developed to correct for the ionospheric phase errors is "field-based calibration" (Cotton et al. 2004; Cotton 2005). For field-based calibration, snapshot images of bright sources are made and their position offsets are measured. With these measurements, an ionospheric model is fit that is subsequently applied during the imaging to correct for the source movements. However, field-based calibration methods are not suitable for arrays with very long baselines such as LOFAR (Lonsdale 2005; Intema et al. 2009).

Below we discuss this facet calibration scheme in more detail, in particular focusing on the parameters that are solved for during the calibration. A schematic overview of the calibration scheme is given in Figure 7.

5.1. Dividing up the Sky in Facets

A first step in the directional calibration is to divide up the sky into facets. When the sky is divided up into facets we make the assumption that the DDE calibration solutions toward the bright source (group) apply to the facet as a whole. The "center points" of the facets are located on bright sources, or the approximate center of a group of closely separated (less than a few arcmin) bright sources. The number of facets required depends on (i) the specific field, (ii) ionospheric conditions and station beam shapes, (iii) the required dynamic range or noise level, and (iv) the science aim. The considerations for the choice of calibrator directions, which define the facet layout, are very similar to, for example, Sagecal or SPAM. The main consideration is having sufficient flux available for calibration and the complexity of the sources (for example, very extended sources might require multi-scale clean which slows down the deconvolution steps).

The selection of the center points of the facets is done by the user (but see Section 6.5). An apparent flux density of at least ∼0.4 Jy is required to define a center point. This is determined by the need to obtain direction-dependent solutions with sufficient S/N. Center points are selected by visually inspecting the 25″ resolution images from the direction-independent self-calibration. Naturally, the sources which show the strongest calibration artifacts (typically the brightest sources) end up in the user defined list of center points. In the case of a source group, the approximate center position of such a group is taken.

For the Toothbrush field this resulted in a list of 67 center points (i.e., directions), which cover an area of about 1.5–2× the HPBW of the station beam. Fewer directions are defined beyond the HPBW because the number of sources with

an apparent flux density of >0.4 Jy decreases steeply beyond this radius. Two bright outlier sources (3C147 and 3C153), located at radial distances >8° from the pointing center, were also included for the Toothbrush field. This number directions is of the same order as used by Yatawatta et al. (2013). We also included an additional 20 directions with <0.4 Jy of flux density beyond the HPBW.[29]

We then employ a Voronoi tessellation (e.g., Okabe 2000) scheme to make the facets; an example of this is given in Figures 6 and 8. This tessellation scheme assigns each point on the sky to the closest calibrator source (group). The area covered by facets is limited by the maximum image size allowed by the user for a given facet. This is done to prevent facets from growing too large. A consequence of this is that the user must take care to have a reasonably uniform distribution of calibrator directions in order to avoid the appearance of gaps in the final image.

5.2. Adding Back the Bright Source (Group)

The next step in the scheme consists of adding back a bright source or source group (which defines the facet position) to the visibility data. Typically, the source covers an area of a few sq. arcmin, which is much smaller than the size of the facet created via the Voronoi tessellation.[30] The data are then phase rotated to the position of the source that was added back and averaged down to a channel resolution of ∼2 MHz, so each 10 subband block is averaged down to 1 channel. The fact that the source (group) only covers a small area means that after phase rotation we can average much more in frequency without being affected by bandwidth smearing. This averaging step is crucial, as the size of the data is reduced by a factor of 20, which speeds up the subsequent calibration cycles (see the next section). No time averaging is done because we need to correct for the ionospheric phase changes on short timescales.

### 5.3. Self-calibrating a Bright Source or Bright Source Group

To obtain the DDE solutions a self-calibration cycle is performed on the bright source (group) that defines a particular facet. This is similar to the procedure followed by SPAM. At this time all the data (120–181 MHz) are imaged together using the MS-MFS clean algorithm (Rau & Cornwell 2011) as implemented in CASA with nterms = 2. Multi-scale[31] clean (Cornwell 2008) is employed for a few complex extended sources (Figure 10). Clean masking is done as described in Section 4.6. The imaging is carried out with all available baselines and no outer uv-range cut is imposed. To speed up the DDE calibration cycle and obtain a reasonable starting model we apply the direction-independent self-calibration solutions (see Section 4.6) at the first imaging step. After we obtain the starting DDE calibration model the direction-independent self-calibration solutions are discarded.

The first two self-calibration cycles consist of Stokes I phase and TEC calibration on the visibility data using BBS. At this point, instead of having 29 sky models for each of the 10 subband blocks, we now go down to a single clean component model that is valid over the entire 120–181 MHz

---

[29] These were later discarded and were used to determine the limiting flux density (of ∼0.4 Jy) for calibration.

[30] The main reason for not adding all the sources back in a facet is to speed up the (self)calibration cycle that obtains the DDE solutions in this direction.

[31] We use the CASA clean option scales = [0,3,7,25,60,150].
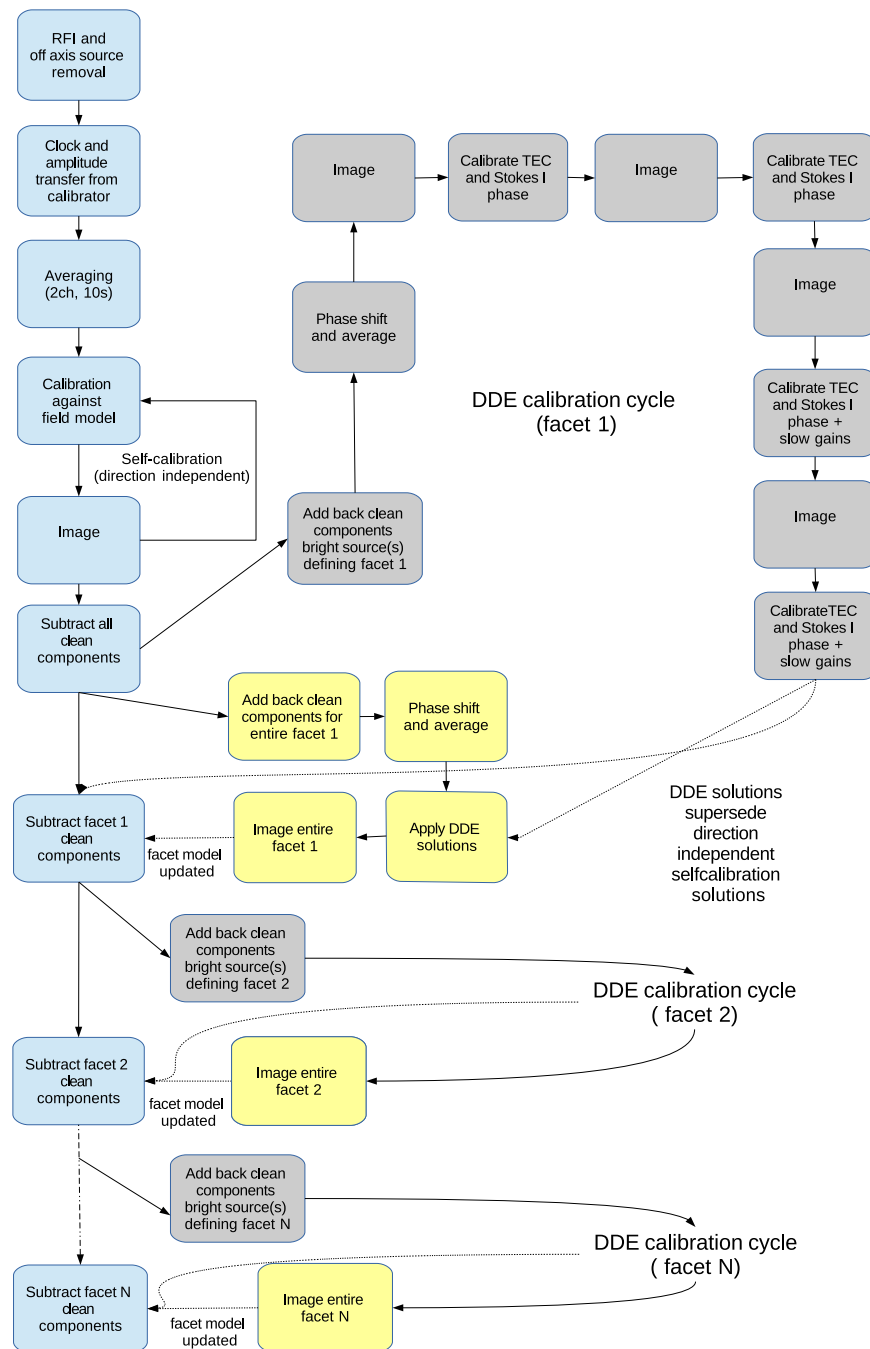
Figure 7. Schematic overview of the HBA calibration scheme employed for this work. Gray colored boxes depict the DDE calibration cycle and yellow boxes depict the imaging of the facets.

band (each clean component having a flux and a spectral index, i.e., corresponding to `nterms = 2` in the `CASA clean` task). Imaging the complete bandwidth at once has the additional advantage of an improved point-spread function (PSF), as the large frequency coverage almost completely fills the uv-plane; in addition it allows cleaning fainter sources/emission that can then be included in the sky model. For most of the directions this phase and TEC calibration is carried out on the shortest timescale of 10 s to correct for the ionosphere. For about a dozen facets we increase this solution interval to 20 s because the available S/N is lower. For the calibration, a single TEC and phase parameter are found per station per 20 MHz bandwidth. We thus solve for only 6 parameters (3 phase and

3 TEC values; see also Section 5.3.1) per antenna for the entire 120–181 MHz band. See Figure 9 (top panel) for an example of the calibration solutions.

After three rounds of this Stokes I phase + TEC self-calibration, we solve for gains on timescales ranging from 5 minutes to 20 minutes, based on the source (group) flux density. The phase + TEC corrections were pre-applied before solving for the XX and YY gains. This "slow gain" calibration is carried out to correct for the slowly varying station beams. The calibration is done independently per 10 subband block because the beam corrections are frequency-dependent. The number of parameters solved for in this step is a factor of ~20 lower than what is typically done with `Sagecal`,
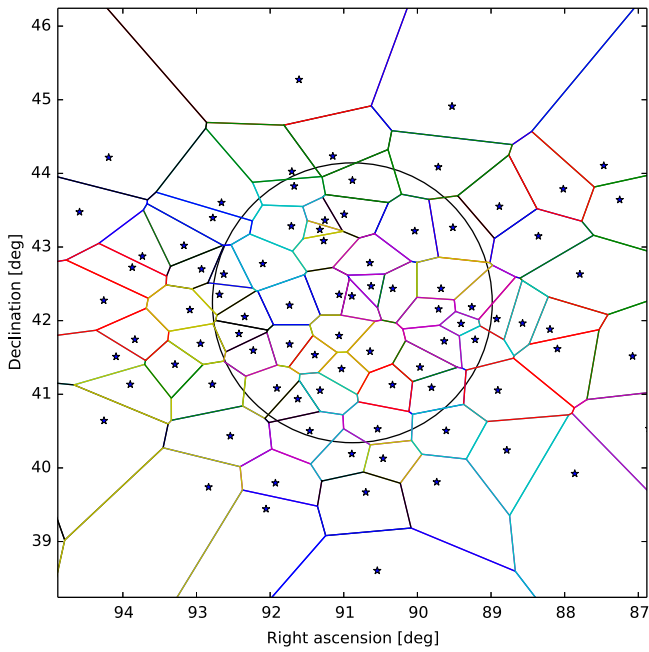
Figure 8. Voronoi tessellation used for the Toothbrush cluster field. The black circle indicates the HPBW of the station beam at 150 MHz. Stars indicate the center points, placed on bright sources (or source groups), that define the tessellation.

which solves for full Jones on a per subband basis on similar timescales.

We run a median window filter on the amplitude solutions to find and replace potential outliers. However, very discrepant amplitude solutions are rare at this stage in the calibration, as no bad data should be present. Bad data should have been identified at earlier steps, i.e., using the calibrator observations or during the direction-independent self-calibration.

The phase components of the "slow gains" are close to zero, i.e., ≪1 rad (Figure 9, middle panel). This is expected because the phase components of the beam variations are small within the main FOV and the ionospheric variations have already been taken out. The slow-gain calibration is followed by another round of Stokes-I phase and TEC calibration and another final round of XX and YY gain calibration. The above scheme basically mimics a joint "short-timescale phase + TEC" and "slow-timescale gain" calibration. The reason for this approach is that BBS cannot jointly solve for parameters on different timescales.

For the final solutions we normalize the global amplitudes to prevent the flux-scale from drifting. In all cases these normalization corrections were very small (a few percent or less). Thus when this self-calibration scheme has finished, we have obtained a set of solutions (Figure 9) for a particular direction. The improvement in image quality, over the previous direction-independent self-calibration (Section 4.6) is very significant after completing this step. An example of the increase in image quality for the calibration scheme described in this section is shown in Figure 10.

The images are not completely free from calibration artifacts: some small scale negative and positive artifacts (at levels of ≲1% of the peak flux) are visible very close to the bright calibrator sources. From tests we noticed that the magnitude of these residuals further decreases when adding additional calibration cycles following the scheme outlined above.
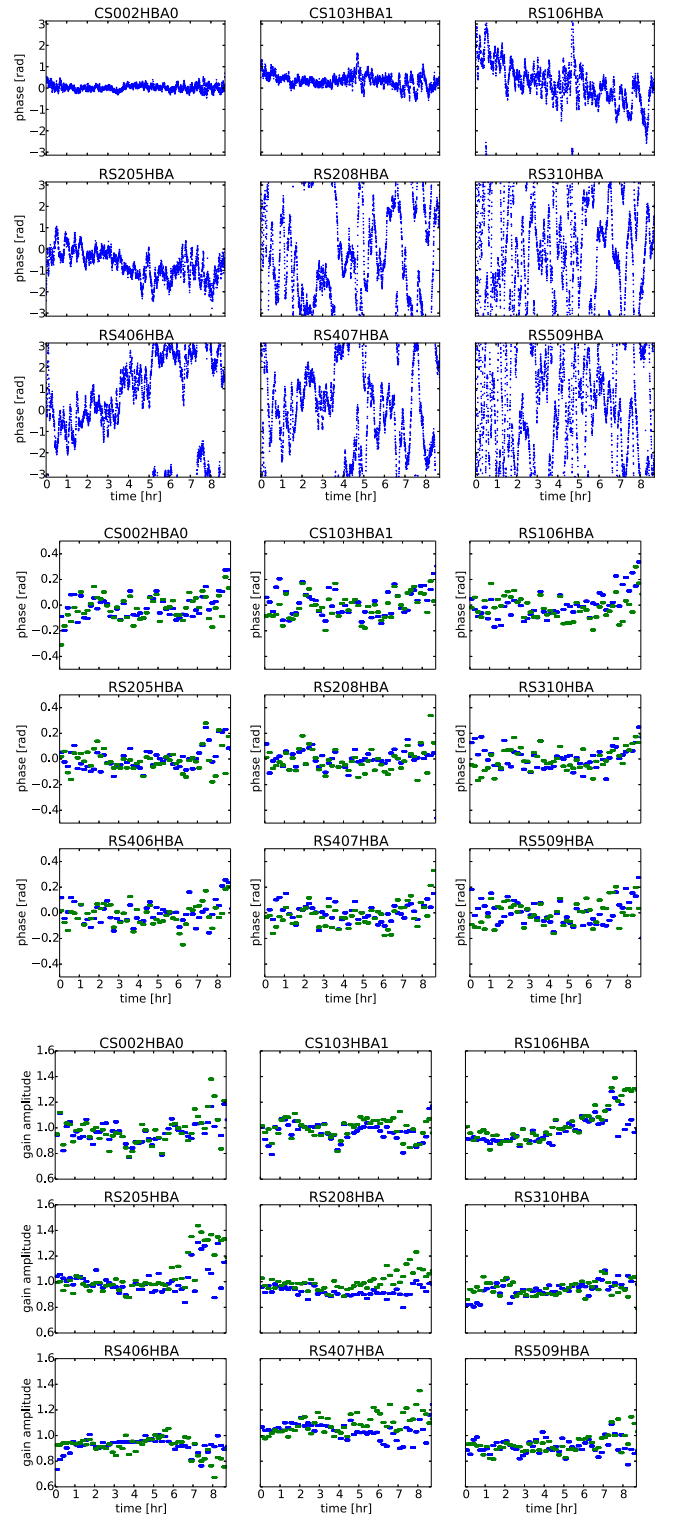
Figure 9. DDE solutions toward direction s21; see Figure 6. Top: the effective Stokes I phase corrections are shown at a frequency of 150 MHz. For the images corresponding to this direction see Figure 10. Because we fit for TEC (and a phase) the actual corrections can be evaluated at an arbitrary frequency. The solutions are obtained on a timescale of 10 s using 20 MHz of bandwidth. Middle: XX (blue) and YY (green) phase solutions for the 150–152 MHz subband block. The solutions are obtained on a timescale of 10 minutes. Phases are always plotted with respect to core station CS001HBA0. Bottom: XX (blue) and YY (green) amplitude solutions corresponding to the middle panel. The distances to CS001HBA0 are 0.37, 2.0, 8.8, 6.4, 27.1, 51.8, 14.3, 20.8, and 55.7 km for CS002HBA0, CS103HBA1, RS106HBA, RS205HBA, RS208HBA, RS310HBA, RS406HBA, RS407HBA, and RS509HBA, respectively.
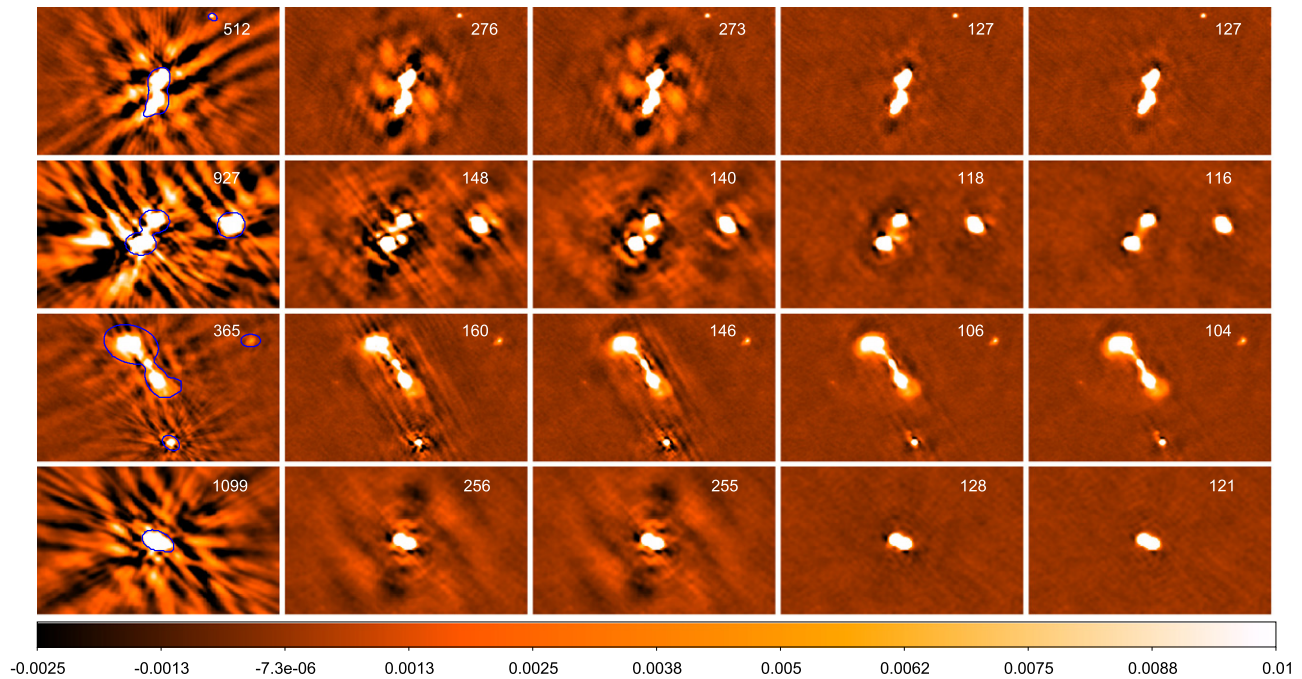
Figure 10. Images showing the incremental improvements during the DDE calibration; see Section 5.3. For reference, the first and second rows of images show direction s2 and s21, respectively (Figure 6). All images are made using the full data set (120–181 MHz, robust = -0.25) and have a resolution of $8'' \times 6''.5$. Note that at this resolution many of the bright DDE calibrator sources are resolved. The first column displays the images made with the (direction-independent) self-calibration solutions; see Section 4.6. The blue contours show the clean mask that was created with PyBDSM for the imaging. The clean mask is updated at each imaging step during the DDE calibration (not shown). The next columns display improvements during the DDE calibration step (see also Figure 9). Second column: first DDE TEC+phase iteration. Third column: second DDE TEC+phase iteration. Fourth column: third DDE TEC+phase iteration and first DDE XX and YY gain (amplitude and phase) iteration. Fifth column: fourth DDE TEC+phase iteration and second DDE XX and YY gain (amplitude and phase) iteration. For all four directions the TEC+phases were solved for on a 10 s timescale. The XX and YY gains were solved for on a 10 minute timescale, except for the source in the top row for which this was 5 minutes. The scale bar at the bottom is in units of Jy beam^{-1}. The images in the first and third rows were cleaned with a multi-scale clean because of extended emission. The rms noise level in each of the images is indicated in the top right corner in units of μJy beam^{-1}.

However, due to computational limitations we decided not to increase the number of calibration cycles. These calibration artifacts could also partially have resulted from the imperfect subtraction of other sources, or indicate calibration errors on shorter timescales or at higher frequency resolutions than what we solve for.

5.3.1. Number of Fitted Parameters versus Number of Measurements

An important consideration is to keep the number of fitted parameters small with respect to the number of measured visibilities to avoid flux loss and overfitting. We define the ratio between the number of measurements and fitted parameters as q. With n stations we have $n(n-1)/2$ complex visibilities per (10 s) time slot per polarization. The solutions are obtained on data that are averaged down to a frequency resolution of 2 MHz. Considering the parallel hand polarizations (XX, and YY), we thus have 86,130 complex measurements in the 120–181 MHz band (29 blocks of 2 MHz with two polarization and $n = 55$).

We fit for $6 \times nm = 22,110$ "phase parameters," where $m = 67$ is the number of directions. The factor of 6 comes from the 3 phase and 3 TEC values we fit per antenna across the 120–181 MHz band. Thus, 86,130 complex measurements are fitted with 22,110 real-valued parameters[32] resulting in $q \approx 7.8$.

For a 10 minute solution interval (60 time slots of 10 s) per 2 MHz block, we also fit for $2 \times nm$ complex gains. In this case $q \approx 24.2$.

Combining both calibrations together, we obtain $q \approx 7.7$. Thus, the number of free model parameters is still significantly smaller than the number of measurements. Ratio q is almost completely determined by the TEC and Stokes I phase corrections.[33]

The above calculation only provides a quick check, as it does not take into account the S/N per measurement, the accuracy of the input calibration model, and the precise functional form that is minimized. However, to first order our value of q seems to be sufficient to prevent overfitting; see Section 6.3.

In our calibration we have traded off the number of measurements with S/N to speed up the calibration, by averaging a factor of 20 in frequency before obtaining the calibration solutions. Without this averaging, keeping a frequency resolution of 2 channels per subband, this would have resulted in a much higher ratio of $q \approx 154$ (see Section 6.5 for options to decrease the number of fitted parameters).

5.4. Adding Back the Other Facet Sources and Imaging

After we obtain a set of DDE corrections for a facet, we add back all other sources in the facet and assume that these fainter sources can be corrected using the same solutions as for the "center" of the facet containing the bright source (group). The

---

[32] The full complex visibility function is minimized during the solve, not only the phase part.

[33] We ignore the fact that for a couple of directions our solutions intervals are longer.

DDE calibration solutions are applied at the original 2 channels per subband frequency resolution. This allows the Stokes I + TEC corrections to be applied on a channel to channel basis.[34] After the corrections are applied the data is averaged by a factor of 5 in frequency and 3 in time. The amount of frequency averaging is less than described in Section 5.3 to avoid bandwidth smearing, because the largest facets have sizes of several tens of arcmin. We then image the facet using MS-MFS clean in CASA (nterms = 2, robust = -0.25) and W-projection (and multi-scale if needed).

Optionally, the facet images can be made with the WSClean imager (Offringa et al. 2014). Note that at this point we do not use the awimager (Tasse et al. 2013), which can correct for the time-varying LOFAR station beams across the FOV during imaging. There are two reasons for this: (1) MS-MFS nterms > 1 imaging was not yet fully implemented and (2) awimager has problems with imaging regions significantly beyond the HPBW of the station beam (the beam response drops close to zero in some regions here and the large beam corrections result in instabilities).

5.5. Subtracting all the Facet Sources

After the imaging of a facet is completed, we have obtained an updated sky model for the region of the sky covering that facet. This sky model is then subtracted from data (which consist of 2 channels per subband at a 10 s time resolution; see Section 5.4), with the corresponding DDE solutions. The new output data that is created should now have the flux from the part of the sky that is covered by the facet correctly removed, while before the sources in the facet were only approximately removed, as they were subtracted with the direction-independent self-calibration solutions, as described in Sections 4.6 and 4.7. As a quality check, we re-image the residual data at low-resolution ($2'$) to verify that the sources in the facet were indeed correctly subtracted and that the magnitude of the residuals has decreased with respect to the direction-independent self-calibration subtraction.

After subtraction, we then proceed with the next direction, and start again with the process described in Section 5.2. This whole process is repeated for all directions. We thus gradually build up a DDE corrected view of the sky; see Figure 7 for the schematic overview of this process. After each direction, the residual visibility data set becomes "emptier" as more and more sources are subtracted with the DDE calibration solutions, instead of the direction-independent self-calibration solutions.

We note that the order in which the facets are treated goes roughly with the brightness of the source (group) that defines a facet. Thus the "worst" facets (with the largest calibration errors in the medium resolution images) are treated first so that at a later point they do not influence the DDE calibration for the facets that have a fainter source (group).

Depending on the science goals not all facets need to be calibrated, as the decrease in overall noise is mostly determined by the facets that contain the brightest ($\gtrsim 1$ Jy) sources (see also Section 6.1). In the end the individual facet images can be combined into an image that covers a larger (or the entire) FOV, which is mostly relevant for survey type science; see Figure 11. This image is then corrected for the primary beam

attenuation by dividing out the primary beam using an image of the beam obtained from awimager.

For survey type science, it is beneficial to re-image all of the facets after the facet calibration with the obtained DDE solutions. The reason for this is that the DDE calibration runs sequentially through the facets. Therefore, the first facet that was produced particularly did not have any other sources subtracted (from other facets) with the DDE calibration solutions. Such re-imaging is less relevant if the target of interest is located within the boundaries of a single facet and that facet is treated as the last one in the list, as was the case for the Toothbrush cluster. In Williams et al. (2016) we will present a facet calibration run where re-imaging was carried out.

In principle the whole DDE scheme could be iterated over; we did not attempt this due to computational limitations. We note, though, that subsequent iterations would offer much smaller improvements than in the first iteration, because the images from the first application of this scheme are already mostly free of calibration artifacts.

## 6. DISCUSSION

6.1. Facet Layout

The amount of sky that needs to be imaged and corrected for DDE depends on the actual science case. We identify two main modes of operation for facet calibration (1) single targets of interest or (2) survey-type science.

In the case of a single target, for example, the observations described in this work, it is actually not essential that the full FOV (within the HPBW) is imaged, or covered with facets. The removal of a limited number of bright sources already significantly improves the image quality for the target of interest. It was found that the improvement in image noise after removal of the >1 Jy sources was relatively small, if the facets that were treated were located far away ($\gtrsim 1°$) from the target of interest.

For survey science, where the main aim is to image as much of the sky as possible, it is important to cover the entire region within the HPBW. If only a few facet directions are chosen, in combination with a maximum image size, only part of the field will be imaged and corrected by the facet calibration scheme, which might not be desirable. Therefore, a sufficient number of calibrator sources is needed that are spread across the region within the HPBW. For example, with less than ~20 directions the assumption that calibration solutions are constant within a facet breaks down. From tests on about a dozen fields, we conclude that in general a sufficient number of calibrator sources is available (always more than 20) to achieve close to thermal noise limited images for a typical 8 hr synthesis run. Results for these fields will be described in upcoming papers.

For fields at low declination (i.e., decl. $\lesssim 10°$), it might be more challenging to find a sufficient number of calibrator sources to correct for DDE across the entire region within the HPBW of the station beam. This is caused by the reduced sensitivity at lower elevations. Therefore this requires a higher integrated flux density limit and thus reduces the number of calibrator sources available. It remains to be determined if this will become an important limitation to carry out facet calibration. Similarly, to achieve much lower noise levels (a few tens of μJy beam^{-1}, or less) higher dynamic range might be required and this could imply that the sky needs to be

---

[34] Recall that in the previous step (Section 5.3), the calibration was carried out on an averaged (phase-shifted) data set with a frequency resolution of 2 MHz.
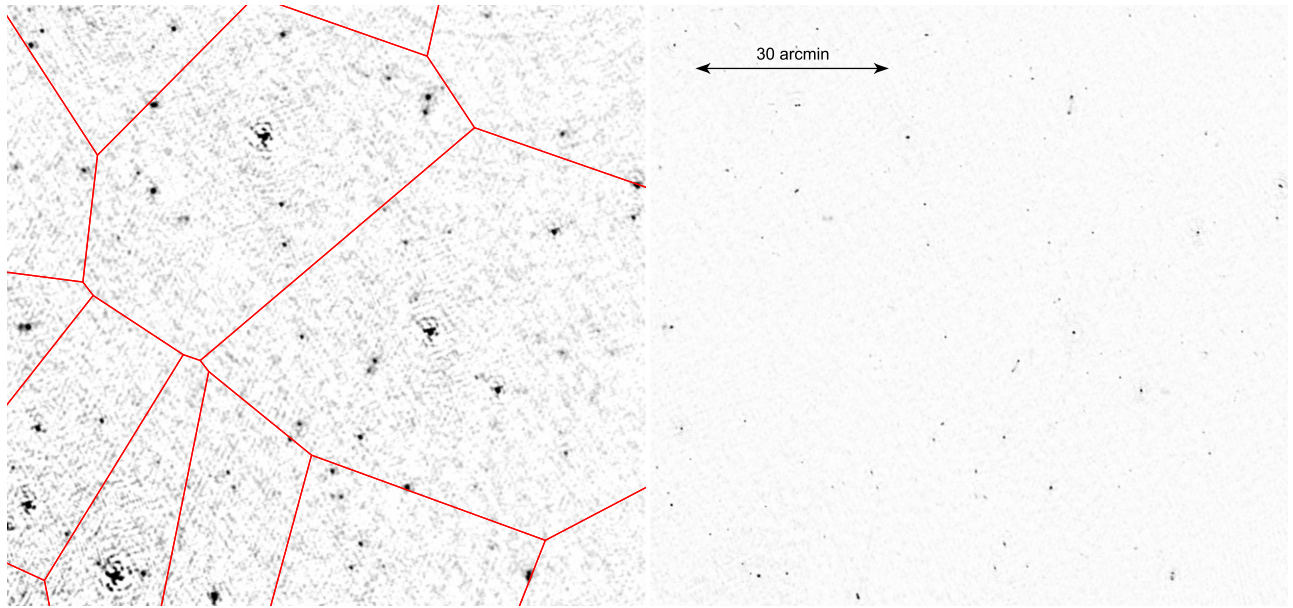
Figure 11. Comparison of a 25″ image at 150 MHz before facet calibration (left) and the high-resolution (8″.0 × 6″.5) full-bandwidth image after facet calibration (right). The rms noise level is close to 0.1 mJy beam^{-1} in the high-resolution image. The images cover the same part of the sky. Individual images of facets were combined to make the high-resolution image. The outline of the facets is shown in red in the left panel. These images are manually scaled in brightness for the purpose of visual comparison.

Table 2
Comparison Facet Calibration, SPAM, and Sagecal

	Facet Calibration	SPAM	Sagecal
Main purpose	Corrected images the sky	Corrected images the sky	Source subtraction
Solving for (scalar) phases	Y	Y	N
Solving for parallel hand gains	Y	Y	Y
Solving for cross hand gains	N[a]	N	Y
Correction and imaging of visibility data with DDE solutions	Y	Y	N[a]
Explicit removal of instrumental effects (clocks)	Y	Y	N
Global phase screen modeling	N[b]	Y	N
Optimized solver	N[c]	N	Y
Solutions obtained for all directions instantaneously	N	N	Y
Works on LOFAR HBA data	Y	N	Y
Solution intervals can vary per direction	Y	Y	Y
Solving for amplitude and phases on different timescales	Y	Y	N

Notes.
[a] Could be implemented and useful for polarization work.
[b] Being attempted.
[c] Stefcal (Salvini & Wijnholds 2014) can be employed for directions with ≳1 Jy of flux density.

divided up into more facets. Due to the computational challenges involved in testing this (see Section 6.4), we leave this as future work.

6.2. Comparison with SPAM and Sagecal

Besides facet calibration, other direction-dependent calibration schemes have been developed to deal with low-frequency radio data. We discuss some of the similarities and differences with the SPAM (Intema et al. 2009) and Sagecal (Yatawatta et al. 2008; Kazemi et al. 2011) packages. SPAM has successfully been applied to a large number of GMRT and VLA data sets. Sagecal has mainly been used for LOFAR HBA observation taken for the Epoch of Reionization (EoR) KSP. In Table 2 we provide a general comparison between these three different calibration schemes.

The main differences between facet calibration and Sagecal are the calibration parameters that are solved for, the underlying solver, and the application of the calibration solutions. Sagecal obtains solutions for all directions at once, while facet calibration obtains solutions for a single direction at each step. Sagecal thus offers significant improvements in speed, also because the underlying solver is different from the "traditional" Levenberg–Marquardt solver that is used in the facet calibration. For more details about the solver employed by Sagecal the reader is referred to the references provided above.

Sagecal only solves for Jones parameters, so complex gain solutions are obtained for all four correlation products for each direction. The facet calibration takes a different approach, since it solves for phases on short timescales and parallel-hand complex gains on longer timescales. The main reason for this

approach is that it enables a correction for the ionosphere on timescales as short as a few seconds, as phases can vary very quickly on the longer LOFAR baselines. Solving for a single phase component per direction per station reduces the number of parameters that are solved for by a factor of 8 compared to full Jones. Solving for all Jones parameters on such short timescales is not feasible, because it would result in over-fitting due to the large number of dof. Facet calibration also directly takes the $1/\nu$ ionospheric frequency dependence into account, further reducing the number of parameters that are solved for in the calibration. Another difference with respect to the facet calibration is that the data are not corrected with the direction-dependent calibration solutions when imaged and deconvolved. In Sagecal, sources are subtracted with the DDE solutions and optionally sources can be restored on an uncorrected residual image.

It is important to note that Sagecal was developed with a different goal in mind, namely the removal of sources from visibility data to detect the EoR, and not to obtain fully corrected images of the sky and correct for the ionosphere on short timescales. However, Sagecal can subtract interfering bright sources that would otherwise decrease the image quality of the main science target in the field, for example.

The facet calibration scheme is more similar to SPAM than Sagecal, as SPAM's main goal is also to produce corrected images of the radio sky. SPAM also attempts to solve for the ionosphere by obtaining phase solutions in a dozen or more directions on short timescales. SPAM does currently not make use of the frequency dependence of the ionosphere, mainly because the GMRT and VLA bandwidths were too narrow to make this practical. A key difference with SPAM is that the facet calibration scheme currently does not obtain a global phase screen from the phase solutions. Attempts to do this using LOFAR data have only been partially successful until now. The production of a global phase screen to model the ionosphere above the array would help to further reduce the number of dof in the calibration. Another difference with facet calibration is that SPAM relies on AIPS (Greisen 2003) and is not well suited to handle LOFAR data. In addition, AIPS does not allow one to solve for more general calibration problems. The underlying solver used in the facet calibration (BBS) is more flexible, allowing, for example, to solve for TEC. In addition, in the facet calibration more emphasis is placed on solving for beam errors, while in SPAM this is less relevant, as it was not meant to work with phased arrays.

6.3. Flux-scale Comparisons After DDE Calibration

To test the effects of the facet calibration on the flux-scale we again compared the integrated LOFAR fluxes to those from the GMRT 150 MHz image (Section 4.6.1). We corrected the LOFAR HBA image with the primary beam as described in Section 5.5. In addition, we corrected for the flux-scale difference found in Section 4.6.

We extracted a source catalog from both images with PyBDSM. For the GMRT map, a box size of 70 pixels was used to compute the locally varying rms noise to take into account (calibration) artifacts around bright sources and variations due to the primary beam attenuation. For the HBA map we took a box size of 250 pixels. We then cross-matched sources in the two catalogs using a matching radius of 2″. A S/N cut of 10 was imposed for all sources. Sources with an integrated flux
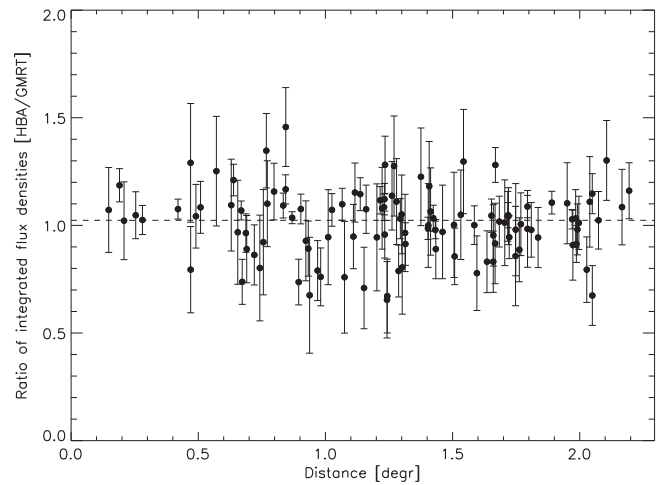
Figure 12. Flux density ratio between GMRT and HBA detected sources as a function of distance to the pointing center.

over peak flux ratio larger than 2 were excluded to avoid very extended and complicated sources.

The resulting flux ratio of LOFAR over GMRT fluxes is shown in Figure 12. We find that the overall flux-ratio between GMRT and LOFAR fluxes remains close to one. The medium of the flux-ratios is 1.02, which indicates the overall flux-scale was not significantly affected by the calibration.

The overall spread in the flux density ratios is somewhat larger than what is expected from the error bars, but part of this extra scatter is related to the source extraction and the formation of sources from the individual Gaussians components (for more information regarding this the reader is referred to the PyBDSM documentation). From the flux-ratio test we conclude that we do not find clear evidence for flux loss during the DDE calibration. In addition, we recover faint sources and source structure in the LOFAR image (van Weeren et al. 2016) that are also seen in higher frequency observations of the field (van Weeren et al. 2012a), but are not included in our initial calibration model derived from the GMRT image at 150 MHz. The above findings are also supported by similar comparisons between images from LOFAR HBA facet calibration runs and GMRT images of the Boötes (Williams et al. 2016) and H-ATLAS fields (M. J. Hardcastle et al. 2016, in preparation). However, we note that more detailed comparisons are needed to fully quantify the effects of the dof in our calibration. To do that, one can, for example, inject fake (faint) sources into the data and check how well they are recovered after facet calibration. Given the processing requirements and amount of effort that will be required to carry out such tests, we leave this for future work.

6.4. Computing Requirements

The amount of computing power required for facet calibration is considerable. For the Toothbrush field, we utilized a single 24 core (two 12-core Xeon E5-2695v2 CPUs) machine with 128 GB RAM. The total amount of pure processing time was 4 months. Therefore, if the facet calibration is to be applied to many more fields, it is desirable to reduce the processing time.

One option that has recently been exploited is to make use of WSClean (Offringa et al. 2014) instead of the CASA imager. This typically reduces the imaging time by a factor of 3 and

decreases the overall processing time by about 30%. In addition, significant speed-up has also been achieved by utilizing `StefCal` for the calibration (Salvini & Wijnholds 2014). Since `StefCal` cannot solve for TEC at the moment, this is currently limited to the brighter ($\gtrsim 1$ Jy) sources because more S/N is required, as the bandwidth needs to be divided into smaller frequency chunks, where the effects of TEC can be neglected. A drawback is that this also increases the number of dof. Another option to reduce processing time is to further parallelize the code. For example, once the brightest sources have been dealt with (and subtracted) several different facets could be imaged and calibrated in parallel.

6.5. Recent Developments and Future Improvements

At the moment the facet calibration needs to be run in a semi-manual way by the user, which means that the method is currently more suited toward individual pointings rather than large surveys for which automated processing is required. To achieve more automated processing, several steps are required.

One is the automatic generation of the list of directions that define the facets. A starting point for the generation of such a list is a source list ordered by integrated flux density. Another aspect concerns the spatial extent of sources, because enough compact flux is needed to obtain calibration solutions for the distant remote LOFAR stations. Finally, a grouping algorithm is required to group closely separated sources.

Another aspect concerns the situation in which facets boundaries can run across (extended) sources; this would preferably be avoided. The order in which the facets are calibrated in is currently also decided by the user, but this is relatively simple to automate. The order could follow the integrated (compact) source flux density. Furthermore, after the brightest sources are removed the order becomes less important.

Another improvement would be to increase the amount of S/N available for the DDE calibration. For example, it is possible that more facets are needed to reach noise levels of a few tens of μJy beam^{-1} (i.e., Surveys KSP Tier-2 and 3 depth). For the purpose of correcting DDEs, more facets are better because corrections in more directions can be applied. On the other hand, depending on the source structure and flux density, as well as the quality of the subtraction from earlier processed facets, at some point the calibration solutions become too noisy for accurate subtraction. A way to further improve the S/N per direction and decrease the number of dof, is to exploit the fact that solutions vary smoothly as a function of time. In addition, the "slow gain solutions" (correcting for the beam) should vary relatively smoothly across the HBA band. This information is currently not used in the facet calibration. Calibration schemes that employ the spatial, frequency, and/or time coherency of the calibration solutions have recently been developed by Tasse (2014), Yatawatta (2015), and Smirnov & Tasse (2015).

Work is ongoing on all the abovementioned aspects. Ultimately, ionospheric phase screens and updated beam models (or amplitude screens) would provide an even larger reduction in the number of dof, but this likely still requires a significant amount of work and study. The same holds for extending the facet calibration to the LBA. In the LBA, the ionospheric effects become more severe and differential Faraday rotation cannot be neglected. A more fundamental limitation is the sensitivity of the LBA stations. LBA stations are about an order of magnitude less sensitive compared to the HBA and it is unclear if solutions in a sufficient number of directions ($\gtrsim 20$) can be obtained in the LBA frequency band.

7. CONCLUSIONS

In this paper we have presented a new calibration scheme to obtain deep high-resolution LOFAR HBA images. We applied this facet calibration scheme to the Toothbrush galaxy cluster field. The scientific results are discussed in van Weeren et al. (2016).

This calibration scheme consists of a direction-independent and a direction-dependent part. For the direction-independent calibration, the LOFAR clock offsets and flux-scale are determined by utilizing the gain solutions from a primary calibrator source. For the direction-dependent calibration, the sky is divided up into facets, with each facet center being defined by a bright source or source group. Calibration solutions for the bright source (group) are obtained, solving for phases on short (~ 10 s) timescales and parallel hand gains on longer timescales (~ 10 minutes). The calibration solutions are applied under the assumption that they are constant across the facet. The updated model of the sky covered by the facet is then subtracted from the data with the solutions obtained. This scheme is repeated for subsequent facets, slowly building up a picture of the full FOV.

University is authorized and regulated by the Financial Conduct Authority.

REFERENCES

Briggs, D. S. 1995, PhD thesis, New Mexico Institute of Mining Technology

Cohen, A. S., & Röttgering, H. J. A. 2009, AJ, 138, 439

Cornwell, T. J. 2008, ISTSP, 2, 793

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 86

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, ISTSP, 2, 647

Cotton, W. D. 2005, in ASP Conf. Ser. 345, From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science, ed. N. Kassim et al. (San Francisco, CA: ASP), 337

Cotton, W. D., Condon, J. J., Perley, R. A., et al. 2004, Proc. SPIE, 5489, 180

Greisen, E. W. 2003, in Information Handling in Astronomy—Historical Vistas, Vol. 285 (Dordrecht: Kluwer Academic Publisher), 109

Hamaker, J. P. 2011, Mathematical-physical Analysis of the Generic Dual-dipole Antenna, Tech. Rep., ASTRON

Heald, G. H., Pizzo, R. F., Orrú, E., et al. 2015, A&A, 582, A123

Intema, H. T. 2014, SPAM: Source Peeling and Atmospheric Modeling, Astrophysics Source Code Library (arXiv:1402.4889)

Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, A&A, 501, 1185

Intema, H. T., van Weeren, R. J., Röttgering, H. J. A., & Lal, D. V. 2011, A&A, 535, A38

Kazemi, S., & Yatawatta, S. 2013, MNRAS, 435, 597

Kazemi, S., Yatawatta, S., Zaroubi, S., et al. 2011, MNRAS, 414, 1656

Kazemi, S., Yatawatta, S., & Zaroubi, S. 2013, MNRAS, 430, 1457

Lonsdale, C. J. 2005, in ASP Conf. Ser. 345, From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science, ed. N. Kassim et al. (San Francisco, CA: ASP), 399

Mohan, N., & Rafferty, D. 2015, ascl:1502.007

Noordam, J. E. 2004, Proc. SPIE, 5489, 817

Noordam, J. E., & Smirnov, O. M. 2010, A&A, 524, A61

Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, MNRAS, 405, 155

Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606

Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, A&A, 539, A95

Okabe, A. (ed.) 2000, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams (2nd ed.; Chichestor: Wiley)

Pandey, V. N., van Zwieten, J. E., de Bruyn, A. G., & Nijboer, R. 2009, in ASP Conf. Ser. 407, The Low-Frequency Radio Universe, ed. D. J. Saikia et al. (San Francisco, CA: ASP), 384

Rau, U., & Cornwell, T. J. 2011, A&A, 532, A71

Röttgering, H., Afonso, J., Barthel, P., et al. 2011, JApA, 32, 557

Röttgering, H. J. A., Braun, R., Barthel, P. D., et al. 2006, (arXiv:astro-ph:0610596)

Salvini, S., & Wijnholds, S. J. 2014, A&A, 571, A97

Scaife, A. M. M., & Heald, G. H. 2012, MNRAS, 423, L30

Schwab, F. R. 1984, AJ, 89, 1076

Smirnov, O. M. 2011a, A&A, 527, A107

Smirnov, O. M. 2011b, A&A, 527, A108

Smirnov, O. M., & Tasse, C. 2015, MNRAS, 449, 2668

Tasse, C. 2014, A&A, 566, A127

Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, A&A, 553, A105

van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2

van Weeren, R. J., Brunetti, G., Brüggen, M., et al. 2016, ApJ, in press (arXiv:1601.06029)

van Weeren, R. J., Röttgering, H. J. A., Intema, H. T., et al. 2012a, A&A, 546, A124

van Weeren, R. J., Röttgering, H. J. A., Rafferty, D. A., et al. 2012b, A&A, 543, A43

van Weeren, R. J., Williams, W. L., Tasse, C., et al. 2014, ApJ, 793, 82

Williams, W. L., van Weeren, R. J., Röttgering, H. J. A., et al. 2016, MNRAS, submitted

Yatawatta, S. 2015, MNRAS, 449, 4506

Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, A&A, 550, A136

Yatawatta, S., Zaroubi, S., de Bruyn, G., Koopmans, L., & Noordam, J. 2008, arXiv:0810.5751