Systematic biases in low-frequency radio interferometric data due to calibration: the LOFAR-EoR case

Article (Published Version)

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **463,** 4317–4330 (2016) doi:10.1093/mnras/stw2277
Advance Access publication 2016 September 9

Systematic biases in low-frequency radio interferometric data due to calibration: the LOFAR-EoR case

Ajinkya H. Patil,[1]★ Sarod Yatawatta,[1,2] Saleem Zaroubi,[1,3] Léon V. E. Koopmans,[1] A. G. de Bruyn,[1,2] Vibor Jelić,[1,2,4] Benedetta Ciardi,[5] Ilian T. Iliev,[6] Maaijke Mevius,[1,2] Vishambhar N. Pandey[1,2] and Bharat K. Gehlot[1]

[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands*
[2]*ASTRON, PO Box 2, NL-7990AA Dwingeloo, the Netherlands*
[3]*Department of Natural Sciences, the Open University of Israel, 1 University Road, PO Box 808, Ra'anana 4353701, Israel*
[4]*Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*
[5]*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85748 Garching bei München, Germany*
[6]*Astronomy Centre, Department of Physics and Astronomy, Pevensey II Building, University of Sussex, Falmer, Brighton BL1 9QH, UK*

Accepted 2016 September 8. Received 2016 September 6; in original form 2016 May 15

ABSTRACT

The redshifted 21 cm line of neutral hydrogen is a promising probe of the epoch of reionization (EoR). However, its detection requires a thorough understanding and control of the systematic errors. We study two systematic biases observed in the Low-Frequency Array-EoR residual data after calibration and subtraction of bright discrete foreground sources. The first effect is a suppression in the diffuse foregrounds, which could potentially mean a suppression of the 21 cm signal. The second effect is an excess of noise beyond the thermal noise. The excess noise shows fluctuations on small frequency scales, and hence it cannot be easily removed by foreground removal or avoidance methods. Our analysis suggests that sidelobes of residual sources due to the chromatic point spread function (PSF) and ionospheric scintillation cannot be the dominant causes of the excess noise. Rather, both the suppression of diffuse foregrounds and the excess noise can occur due to calibration with an incomplete sky model containing predominantly bright discrete sources. The levels of the suppression and excess noise depend on the relative flux of sources which are not included in the model with respect to the flux of modelled sources. We predict that the excess noise will reduce with more observation time in the same way as the thermal noise does. We also discuss possible solutions such as using only long baselines to calibrate the interferometric gain solutions as well as simultaneous multifrequency calibration along with their benefits and shortcomings.

Key words: methods: data analysis – techniques: interferometric – dark ages, reionization, first stars.

1 INTRODUCTION

The first stars and galaxies formed towards the end of cosmic dark ages and their energetic radiation is thought to have ionized matter in the Universe. The epoch of reionization (EoR) is the era in which matter in the intergalactic medium was transformed from being neutral to ionized. The EoR carries a wealth of information about structure formation and the first astrophysical objects in the Universe.

As hydrogen is the most abundant element in the Universe, the 21 cm transition line of neutral hydrogen is a promising probe of the EoR. The evolution of neutral hydrogen through cosmic time can be studied by observing the 21 cm line at different redshifts. The EoR is expected to have occurred between redshifts 6 and 12 (Hinshaw et al. 2013; Planck Collaboration XLVII 2016), which correspond to observational frequencies of 120 to 200 MHz for the redshifted 21 cm transition line. Therefore, several experiments are aiming at observing the EoR with low-frequency radio telescopes including Giant Meterwave Radio Telescope (Paciga et al. 2013), Low-Frequency Array (LOFAR; van Haarlem et al. 2013), Murchison Widefield Array (MWA; Bowman et al. 2013; Tingay et al. 2013; Dillon et al. 2015; Trott et al. 2016), the Donald C. Backer

★ E-mail: patil@astro.rug.nl

Precision Array for Probing the Epoch of Reionization (Parsons et al. 2010; Ali et al. 2015), the Hydrogen Epoch of Reionization Array (DeBoer 2016), the Square Kilometer Array (Mellema et al. 2013; Koopmans et al. 2015).

The contamination due to the Galactic and extragalactic foreground emission is one of the primary challenges in detecting the cosmic redshifted 21 cm emission from neutral hydrogen (hereafter referred as the 21 cm signal). The astrophysical foregrounds are either discrete sources such as radio galaxies and clusters or diffuse synchrotron and free–free emissions from our Galaxy (Shaver et al. 1999; Di Matteo et al. 2002; Oh & Mack 2003; Cooray & Furlanetto 2004; Di Matteo, Ciardi & Miniati 2004). These foregrounds are several orders of magnitude brighter than the expected 21 cm signal. Therefore, an accurate removal of the foregrounds while avoiding possible systematic errors is crucial for the success of EoR experiments. In this paper, we present some systematic biases observed in the residual LOFAR-EoR data after calibration and subtraction of bright discrete foreground sources, investigate their origins and discuss possible solutions.

Two important systematic biases observed in the LOFAR-EoR data after calibration and foreground subtraction are (i) a suppression of diffuse, polarized foregrounds and (ii) an excess of noise. Diffuse foregrounds appear both in total and polarized intensity (Jelić et al. 2014, 2015), but they are difficult to detect in total intensity (Stokes I) in presence of numerous bright discrete sources. Diffuse foregrounds are dominant in polarized intensity, because only few discrete foreground sources show polarized emission. We observe a suppression in the polarized diffuse foregrounds while subtracting discrete foreground sources. Diffuse foregrounds appear predominantly on large angular scales, which are also the most promising scales for a detection of the 21 cm signal (Zaroubi et al. 2012; Chapman et al. 2013; Patil et al. 2014). Although one aims to detect the 21 cm signal in total intensity, a suppression of the diffuse polarized foregrounds could suggest a suppression of the 21 cm signal as well. The second systematic effect is an excess of noise beyond the thermal noise. The excess noise not only reduces sensitivity, but also causes an obstacle in the foreground removal. Several foreground removal or avoidance algorithms separate the foregrounds based on their spectral smoothness [see Chapman et al. (2015) for a review of foreground removal methods]. The excess noise introduces additional random variations along frequency in the data, and hence it makes removal of foregrounds inefficient. We investigate three potential sources of the excess noise: the chromatic nature of the point spread function (PSF), ionospheric scintillation, and calibration artefacts.

The response of a radio interferometer needs to be calibrated in order to correct for variations in electronics and the ionosphere. A bright compact source with known flux is needed to calibrate the gains of interferometric elements. However, few such calibrator sources are known at low radio frequencies, and it is possible that none of them might be located within the field view of an observation. One can instead use self-calibration in such cases. In self-calibration, a model of bright sources in the sky is constructed, and it is used to calibrate the gains of interferometric elements (Schwab 1980; Cornwell & Wilkinson 1981). The sky model and the gain solutions are improved in an iterative manner. The traditional self-calibration obtains one gain solution for each interferometric element. However, this may not be sufficient for the new generation of telescopes with wide field of views, where the gain might change as a function of direction. Direction-dependent self-calibration is then used where the gain solutions in multiple directions are obtained (van der Tol, Jeffs & van der Veen 2007; Wijnholds & van der Veen 2009). Some EoR projects use direction-dependent self-calibration for the calibration and subtraction of bright sources (Mitchell et al. 2008; Yatawatta et al. 2013). Nearby sources can be clustered together to get one solution in the respective direction (Kazemi, Yatawatta & Zaroubi 2013a).

The sky model in self-calibration is often imperfect due to errors in flux, position or morphology of the modelled sources (Datta, Bhatnagar & Carilli 2009; Datta, Bowman & Carilli 2010). The sky model is also incomplete, because it contains only bright discrete sources and excludes faint discrete sources and diffuse emissions. Some artefacts of calibration with an incomplete sky model have been well known. These include generation of spurious source components and suppression of real components (Wilkinson, Conway & Biretta 1988). Grobler et al. (2014) and Wijnholds, Grobler & Smirnov (2016) considered a simple case of one bright and one faint source and provided an analytical description of how spurious sources can be generated when the faint source is excluded in the model for calibration. However, real data are more complex with many discrete sources and diffuse foregrounds. Therefore, in this paper, we rely on simulations to study effects of model incompleteness. In a similar study, Barry et al. (2016) found that excluding faint discrete sources in a sky model leads to contamination of foreground-free power-spectrum modes. In this paper, we also consider effects of diffuse foregrounds and show the contamination in the observed data. Recently, some solutions to artefacts of the calibration with an incomplete sky model have been discussed in literature. Simulations by Nunhokee (2016) showed that using a longer time interval for the calibration reduces suppression of unmodelled sources. However, increasing solution time interval in reality would limit time-scales on which ionospheric effects can be removed. Barry et al. (2016) and Ewall-Wice et al. (2016) used multifrequency calibration for MWA by modelling the instrument response with low-order polynomials. They were limited by intrinsic spectral structures in the instrument such as cable reflections. We will discuss some new solutions in this paper along with their advantages and shortcomings.

An alternative to self-calibration is redundancy calibration which does not require a priory model of the sky (Noordam & de Bruyn 1982; Wieringa 1992). Therefore, the discussion in this paper does not apply to redundancy calibration. However, redundant arrays use a hybrid approach consisting of redundancy calibration followed by a sky model based calibration to resolve degeneracies of the former (Zheng et al. 2014; Ali et al. 2015).

The paper is organized as follows: in Section 2, we briefly describe the data analysis pipeline for the LOFAR-EoR project. In Section 3, we discuss systematic biases observed in the calibrated data, namely, an excess noise and suppression of the diffuse foregrounds. Detailed properties of the excess noise and possible sources of its origin are discussed in Section 4. In Section 5, we show with the help of simulations that the above two systematic biases could be artefacts of calibration with an incomplete sky model. We discuss some possible solutions to the systematic biases in Section 6, before concluding in Section 7.

2 OBSERVATIONS AND DATA PROCESSING

The data used in this paper were observed with LOFAR during observing cycle 0 (2013 February–November) and cycle 1 (2013 November–2014 May). We concentrate on the primary target field of the LOFAR-EoR experiment centred on the North Celestial Pole (NCP). The NCP field was observed with 55 LOFAR High Band Antenna stations in the Netherlands, providing baselines from 68 m

Table 1. Observational details of the data used in this paper.

Telescope	LOFAR High Band Antenna
Observational period:	
LOFAR cycle 0	February–November 2013
LOFAR cycle 1	November 2013–May 2014
Duration of an observation	6–16 h (season-dependent)
Frequency range	115–174 MHz
Field of view at 150 MHz	$3.^{\circ}8$ (FWHM)
Polarization	Linear X–Y
Longest baseline:	
LOFAR core	3.5 km
LOFAR Dutch array	121 km
Collecting area (zenith):	
LOFAR core	512 m^2 × 48 stations
Dutch remote stations	1024 m^2 × 16 stations
Time, frequency resolution:	
Raw data	2 s, 3 kHz
After RFI flagging	2 s, 12 kHz
After calibration	10 s, 183 kHz

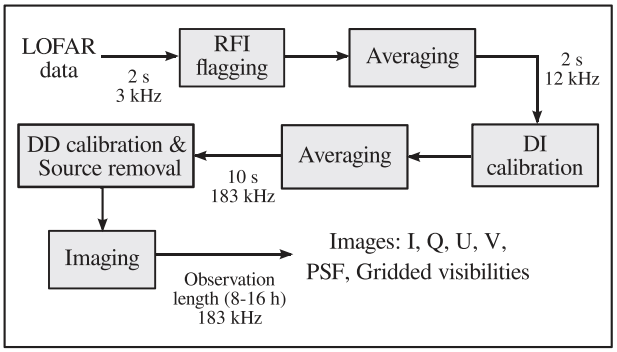

Figure 1. Block diagram of the data reduction pipeline. Time and frequency resolutions at different stages are noted. DI and DD refer to direction independent and direction dependent, respectively. The final output is a set of images of Stokes parameters (I, Q, U, V), the PSF and gridded visibilities.

to 121 km, and operating in the frequency range 115–189 MHz. However, we use the data only up to 174 MHz in this paper, because the 174–189 MHz part of the bandwidth is corrupted by radio frequency interference (RFI). The frequency range 115–174 MHz corresponds to redshifts 7 to 11.35 for the 21 cm line of neutral hydrogen. Visibilities, i.e. correlations of voltages from pairs of antennas, were recorded with 2 s time resolution. The total bandwidth was divided into 195 kHz sub-bands. Each sub-band consisted of 64 channels, thereby providing a frequency resolution of 3 kHz. We observed only during night time to avoid contamination due to the solar emission and minimize ionospheric phase errors. The duration of an observation varied between 6 and 16 h depending on the season at the time of observation. The observational details are summarized in Table 1. For more information about LOFAR capabilities, the reader is referred to van Haarlem et al. (2013). Different steps in the processing of the observed data are summarized in the following subsections. Please see Fig. 1 for a block diagram of the data reduction pipeline.

2.1 Pre-processing

The first step in our data processing is to discard that part of the data which is affected by RFI. The RFI mitigation is performed by the software AOFLAGGER (Offringa et al. 2010; Offringa, van de Gronde & Roerdink 2012) at the highest time and frequency resolution available to minimize information loss. Two frequency channels on either edge of every sub-band are discarded to avoid edge-effects of the polyphase filter. This reduces the bandwidth of each frequency sub-band to 183 kHz. The remaining data is then averaged to 12 kHz, 2 s resolution to reduce its volume for further processing.

2.2 Direction-independent calibration

Usually, a bright source with known flux can be used to calibrate the gain of each interferometric element. However, the region within the field of view at the NCP contains not one dominant source but rather many sources with comparable fluxes e.g. NVSS 7011732+89284 with 7.2 Jy,[1] 3C61.1 with 1 to 11 Jy depending on frequency and several sources with 1 Jy apparent flux at 150 MHz. Therefore, we use 300 sources spread over the area of 10 × 10 deg^2 to calibrate the average station gains over the field of view in the direction of the NCP. We use the Black Board Selfcal package (Pandey et al. 2009) to obtain and apply the calibrated gain solutions for every 10 s time interval and 183 kHz bandwidth. Each station gain is described by a 2 × 2 Jones matrix for two orthogonal linear polarizations.

2.3 Source subtraction

Supernova remnants and radio galaxies and clusters are the discrete foreground sources observed at low radio frequencies. The brightest sources in the NCP field are about six orders of magnitude brighter than the expected 21 cm signal. Therefore, we need to remove the foreground sources with a very high accuracy to reach the required sensitivity for a signal detection. Foreground sources can be subtracted by self-calibration. However, station gains obtained towards the centre of the field or the average gains over the field of view are not good enough for the entire field of view of LOFAR. Varying primary beam shapes and ionospheric effects cause direction-dependent effects (Lonsdale 2005; Koopmans 2010; Vedantham & Koopmans 2015), which require obtaining gains towards multiple directions in which sources are to be removed. This is called a direction-dependent calibration. We use SAGECAL (Yatawatta et al. 2009; Kazemi et al. 2011; Kazemi & Yatawatta 2013; Kazemi, Yatawatta & Zaroubi 2013b) to calibrate the station gains in multiple directions and ultimately subtract sources. SAGECAL takes a sky model containing positions, fluxes and morphologies of a set of known sources as an input. It solves for the station gains in the direction of these sources by minimizing the difference between the observed data and predicted visibilities for the sky model multiplied with the estimated station gains (please see the appendix for a mathematical description of the calibration). Finally, the sources are removed by subtracting their predicted visibilities multiplied with the obtained gain solutions.

It is important to note that the station gain solutions are only used to subtract the modelled sources, they are not applied to the residual data. The residual data still remains affected by direction-dependent errors (DDEs). DDEs are not relevant for the cosmic signal itself, because only a small central region around the pointing centre will be used for an analysis of the cosmic signal where the sensitivity is

[1] The flux of NVSS 7011732+89284 was earlier thought to be 5.3 Jy (Yatawatta et al. 2013) and was used to set the absolute flux scale. It was assumed that the source has a constant spectrum from 100 to 300 MHz. However, recent observations with LOFAR have revealed that the spectrum of the sources rises and falls within this frequency range with the correct flux of 7.2 Jy at 150 MHz (de Bruyn et al., in preparation).

highest due to the primary beam response. However, DDEs affect foregrounds that are further away from the pointing centre and hence their sidelobes in the central region of interest. The primary beam and ionospheric effects causing these errors are expected to vary smoothly with frequency. Therefore, the residual foregrounds can be removed in a second step of foreground removal using algorithms that separate spectrally smooth foregrounds from the thermal noise and the cosmic signal (Chapman et al. 2015).

In order to reduce the data volume, we average data to 10 s and 183 kHz resolution before source subtraction. However, effects of frequency and time smearing are taken into account while predicting visibilities for the sky model. The sky model is regularly updated as we reach better sensitivities by subtracting sources and observing more data. We refer the reader to Yatawatta et al. (2013) for more details about the calibration and source subtraction in the LOFAR-EoR NCP field.

2.4 Imaging

Residual visibilities obtained after source subtraction are imaged using the software package EXCON (Yatawatta 2014). We attempt to maintain the spectral smoothness of foregrounds by using uniform weighting and only the densely sampled part of the uv plane, i.e. baselines between 30 and 800 wavelengths (Patil et al. 2014). Separate images are made for each 183 kHz wide sub-band.

3 SYSTEMATIC BIASES IN THE DATA

As a first step towards the detection of the 21 cm signal, we would like to measure the variance (Patil 2014; Patil et al. 2014) and the power spectrum (Harker et al. 2010; Chapman et al. 2013) of the differential brightness temperature of the 21 cm emission as a function of redshift. Simulations in Patil et al. (2014) show that the 21 cm signal variance can be detected with a 4σ significance in 600 h if all systematic errors can be controlled. However, we identify two systematic biases in the residual data after calibration and subtraction of bright discrete foreground sources, namely, an excess of noise and a suppression in diffuse foregrounds. These two problems are described in the following subsections.

3.1 The excess noise

An accurate determination of the statistical properties of the thermal noise such as its standard deviation and power spectrum is important. The expected standard deviation (σ) of the thermal noise in a visibility can be calculated from the system equivalent flux density (SEFD) as

$$\sigma = \frac{\mathrm{SEFD}}{\sqrt{2\Delta\nu\Delta t}}, \quad (1)$$

where $\Delta\nu$ and Δt are integration frequency bandwidth and time, respectively. The SEFD depends on the elevation of an observation. The expected SEFD of the LOFAR High Band Array towards the NCP is about 4100 Jy, as derived from the empirical SEFD towards the zenith (de Bruyn et al., in preparation). For 10 s and 180 kHz integration, the noise per visibility should be 2.16 Jy. About 7×10^6 visibilities are observed over 12 h of observation. Therefore, the thermal noise in an image made with such an observation should be about 580 μJy. In reality, the noise in an image depends on several factors such as the fraction of the data flagged due to RFI, weights given to different visibilities during imaging, the Galactic background in the direction of observation, calibration artefacts. A more detailed discussion about noise properties will follow in de Bruyn et al. (in preparation).

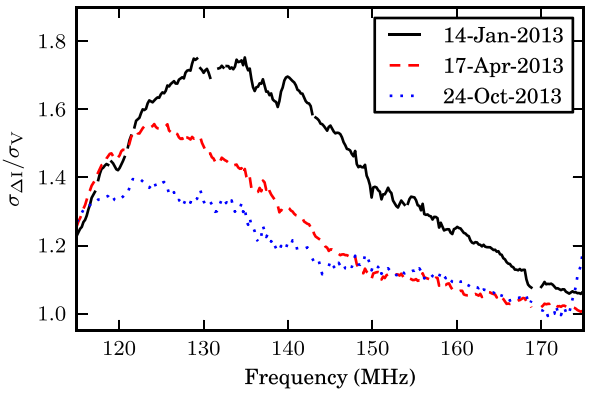

Figure 2. The ratio of the rms of differential Stokes I images ($\sigma_{\Delta I}$) to those of Stokes V images (σ_V), as a function of frequency for three observations. Consecutive sub-bands 195 kHz apart are used for the difference. The ratio is always greater than unity, implying there is an excess of noise in Stokes I as compared to the thermal noise dominated Stokes V. Sub-bands containing strong RFIs have been removed.

The actual thermal noise in an observation can be determined using the circular polarization data, i.e. Stokes V parameter. Most radio sources in the sky do not show circular polarization. Therefore, the Stokes V images are expected to be thermal noise dominated. There can be a small leakage of the total intensity, i.e. Stokes I into Stokes V. Such leakage occurs because of the different projections of the two orthogonal dipoles towards the same direction in the sky. However, the polarization leakage for modelled sources is removed during the calibration and source removal. Furthermore, Asad et al. (2015) have shown that the Stokes I to Stokes V leakage is less than 0.003 per cent. Therefore, the Stokes V images provide good estimates of the noise properties. The root mean square (rms) of the Stokes V noise in our data is about 0.9 mJy for a 13 h and 195 kHz (one sub-band) integration at 150 MHz in uniform-weighted images of 3 arcmin resolution.

Another way to estimate the noise properties directly from the Stokes I parameter is to take the difference between two Stokes I images separated by a small frequency interval. All other signals from the sky, e.g. foregrounds and cosmological signal, should almost be the same in the two channels. The PSF changes by only 0.1 per cent over 0.2 MHz.[2] Hence, the difference between two consecutive frequency channels should be dominated by the thermal noise, especially after the brightest discrete foreground sources have been subtracted. In principle, the noise properties obtained from the differential Stokes I images should be very close to those obtained from Stokes V. However, we find the Stokes I differential noise to be higher, as shown in Fig. 2 where we plot the ratio of their rms values for three different nights of observations. We call this additional noise in the Stokes I images the 'excess noise'. The excess noise could originate from the following sources:

(i) Convolution of residual sources with the chromatic PSF;
(ii) ionospheric scintillation;
(iii) calibration and foreground removal artefacts.

[2] We measure the chromatic variation of the PSF by constructing images of the PSF towards the pointing centre. The rms of the difference between 10° images of the PSF separated by 0.2 MHz in frequency is about 0.001, where each PSF image is normalized to have the maximum value of unity.

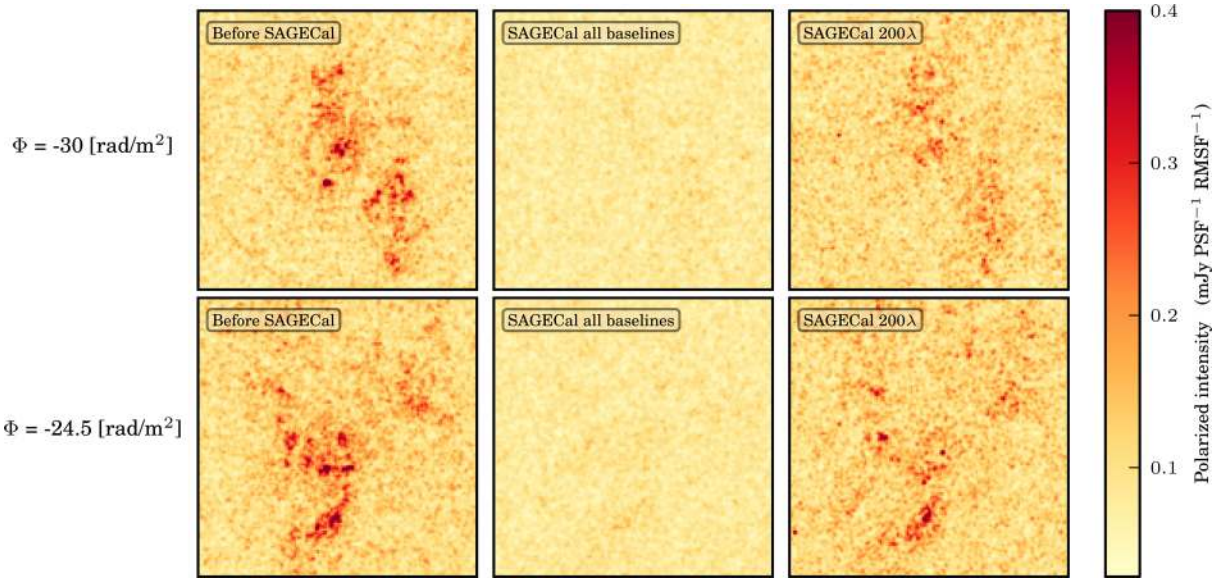

Figure 3. Suppression of the diffuse foregrounds: uniform weighted, 4° polarized intensity maps for the following cases: (i) before subtraction of discrete sources (first column), (ii) after source subtraction using SAGECAL (second column) and (iii) after subtracting sources using baselines only longer than 200 wavelengths in calibration (third column). The top and bottom rows correspond to Faraday depths of −30 and −24.5 rad m^{-2}. The diffuse foregrounds are suppressed during the source subtraction because they are not included in the sky model. They can partially be recovered by excluding short baselines in calibration, but this results into an enhanced noise. The bright discrete sources present in the first column have been removed by SAGECAL in other columns.

We perform several tests and simulations to study properties and causes of the excess noise. The potential sources, i.e. a chromatic PSF and ionospheric scintillation will be discussed in Section 4, whereas calibration artefacts will be discussed in Section 5.

The excess noise cannot be removed by the foregrounds fitting algorithms which are used to remove faint sources and the diffuse foregrounds after subtracting the bright sources. Most of these algorithms separate the foregrounds from the 21 cm signal based on their smooth frequency spectra [Chapman et al. (2015) and references therein]. The excess noise is uncorrelated even on small frequency separations of 0.2 MHz, and hence it cannot be easily removed by standard foreground removal methods that expect spectrally smooth foregrounds.

3.2 Suppression of the diffuse foregrounds

The second systematic effect that we observe in the data is a suppression of the diffuse foregrounds, which occurs in the process of removal of discrete sources. Synchrotron and free–free emissions from our own Galaxy constitute the diffuse foregrounds. These diffuse foregrounds are difficult to model and computationally expensive to include in the sky model for the direction-dependent calibration in SAGECAL. We remove them at a later stage based on their presumed smooth frequency spectra (Harker et al. 2009; Chapman et al. 2012, 2013). Therefore, our sky model for SAGECAL contains only discrete sources, whereas the observed data contains also the diffuse foregrounds in total intensity as well as the linear polarization (Jelić et al. 2014, 2015). A consequence of the difference between the true sky and the calibration sky model could be to suppress structures that are not part of the model, absorbing them in gains applied to the restricted calibration sky model and potentially lead to excess power elsewhere in the image or on different spatial or frequency scales.

The suppression of the diffuse foregrounds is not easy to notice in Stokes I images because they are dominated by bright discrete sources and confusion noise. However, the suppression of the polarized diffuse foregrounds can be easily seen, because not many discrete sources are polarized. The first two columns in Fig. 3 show the diffuse foregrounds in polarized intensity before and after the source subtraction, and the suppression in the latter case is self-evident. We show polarized intensity maps at two Faraday depths (Φ) of −30 and −24.5 rad m^{-2} obtained by rotation measure synthesis (Brentjens & de Bruyn 2005). The diffuse foregrounds appear on large angular scales where a detection of the 21 cm signal is also most promising (Zaroubi et al. 2012; Chapman et al. 2013; Patil et al. 2014). Therefore, our concern is that a suppression in the diffuse foregrounds could mean a suppression of the 21 cm signal as well. A solution for mitigating the suppression of the diffuse foregrounds and the 21 cm signal is to exclude short baselines in the calibration. One can use only baselines longer than a certain baseline length and still obtain the gain solutions for all stations. Previously, Jelić et al. (2015) have used only baselines longer than 800 wavelengths in the calibration to minimize the suppression of the diffuse foregrounds. We use baselines longer than 200 wavelengths to obtain station gains but subtract the sky model sources on all baselines. As shown in the third column in Fig. 3, this reduces the suppression of the diffuse foregrounds. One should note that the first and the third columns in Fig. 3 do not look exactly the same because the bright, largely instrumentally polarized, point sources present in the left-hand panels have been subtracted using SAGECAL in the right-hand panels.

4 PROPERTIES OF THE EXCESS NOISE

We performed several tests with an aim of investigating properties and ultimately the origin of the excess noise. Results of these tests are presented in this section.

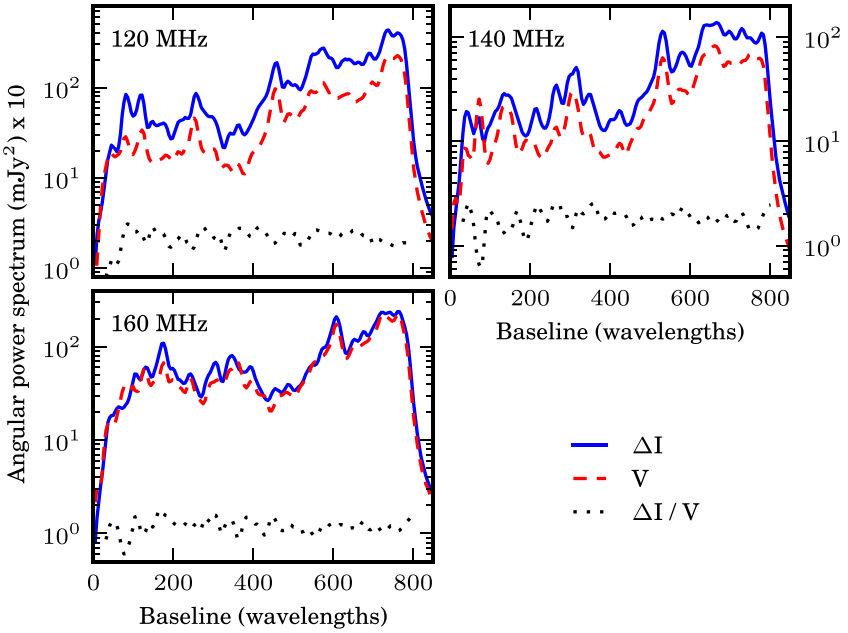

Figure 4. Angular power spectrum of the excess and thermal noise for the observation on 2013 April 17. The ratio of the two remains constant irrespective of the baseline length. The power spectra have been multiplied by 10 for the convenience of plotting their ratio in the same plot.

4.1 Angular power spectrum

The angular power spectrum can be a useful tool in identifying causes of the excess noise. One should expect higher power on smaller angular scales if either sidelobes of sources due to the chromatic PSF or ionospheric scintillation is the dominant cause of the excess noise. Sidelobes of unsubtracted sources are not perfectly subtracted in a sub-band difference due to the chromatic nature of the PSF (Morales et al. 2012; Parsons et al. 2012; Vedantham, Udaya Shankar & Subrahmanyan 2012). The PSF is chromatic because the uv coordinate or the spatial frequency u corresponding to a baseline scales with frequency f as

$$u = \frac{bf}{c}, \tag{2}$$

where b is the physical length of the baseline and c is the speed of light. The rate of change of the uv coordinate with frequency, i.e.

$$\frac{\mathrm{d}u}{\mathrm{d}f} = \frac{b}{c}, \tag{3}$$

is larger at longer baselines. Therefore, we expect the power spectrum of the excess noise to increase with the baseline length, if a chromatic PSF were the dominant cause of the noise. Similarly, ionospheric scintillation noise shows more power on longer baselines (Vedantham & Koopmans 2015, 2016).

We compute the azimuthally averaged angular power spectrum of the excess noise by Fourier transforming the differential Stokes I images and then squaring their magnitude. In Fig. 4, we show the power spectrum of the excess noise as a function of baseline length for the observation on 2013 April 17. We also show the power spectrum of the thermal noise from Stokes V. The ratio of the power spectrum of the excess noise to that of the thermal noise remains constant as a function of the baseline length. Therefore, we conclude that sidelobes of the unsubtracted sources and ionospheric scintillation are unlikely to be the dominant sources of the excess noise. This is in agreement with Vedantham & Koopmans (2016) where it is shown that scintillation noise is confined to the wedge-like structure in the two-dimensional power spectrum similar to smooth spectral foregrounds.

4.2 Contribution due to the chromatic PSF

The analysis presented in Section 4.1 suggests that sidelobes of unsubtracted sources is unlikely a dominant cause of the excess noise. However, we would like to study the chromatic nature of sidelobes in more detail and quantify its contribution to the excess noise in this subsection.

The observed Stokes I signal in a frequency sub-band can be expressed as

$$i_1 = s_1 * p_1 + n_{i1}, \tag{4}$$

where s_1 is the original signal from the sky, p_1 is the PSF, n_{i1} is the thermal noise in Stokes I, and $*$ denotes a convolution operation. Taking a Fourier transform,

$$I_1 = S_1 \times P_1 + N_{i1}, \tag{5}$$

where a capital letter denotes the Fourier transform of the respective quantity in equation (1). For Stokes V,

$$V_1 = N_{v1}, \tag{6}$$

as we assume that the Stokes V contains only the thermal noise. Similarly, for a consecutive sub-band,

$$I_2 = S_2 \times P_2 + N_{i2}, \tag{7}$$

$$V_2 = N_{v2}. \tag{8}$$

For a 195 kHz separation between two consecutive sub-bands, we assume that the signal from the sky does not change, i.e.

$$S = S_1 \approx S_2. \tag{9}$$

The difference between the two sub-bands then becomes

$$\mathrm{d}I = I_1 - I_2 = S\,\mathrm{d}P + N_{i1} - N_{i2}, \tag{10}$$

where $dP = P_1 - P_2$. We can compute the power spectrum of the differential Stokes I as

$$\langle |dI|^2 \rangle = |S|^2 |dP|^2 + \langle |N_{i1}|^2 \rangle + \langle |N_{i2}|^2 \rangle . \quad (11)$$

Equation (10) follows from equation (9) because the thermal noise realizations at different sub-bands do not correlate. Similarly, for Stokes V,

$$\langle |dV|^2 \rangle = \langle |V_1 - V_2|^2 \rangle = \langle |N_{v1}|^2 \rangle + \langle |N_{v2}|^2 \rangle . \quad (12)$$

The noise in Stokes I and V should be statistically identical, implying $\langle |N_{i1}|^2 \rangle = \langle |N_{v1}|^2 \rangle$ and $\langle |N_{i2}|^2 \rangle = \langle |N_{v2}|^2 \rangle$. Therefore, subtracting equation (11) from equation (10),

$$\langle |dI|^2 - |dV|^2 \rangle = |S|^2 |dP|^2 , \quad (13)$$

where the power spectrum of the signal from the sky $|S|^2$ is obtained using

$$\frac{|I_1|^2 - |V_1|^2}{|P_1|^2} = \frac{|S|^2 |P_1|^2 + |N_{i1}|^2 - |N_{v1}|^2}{|P_1|^2} = |S|^2 . \quad (14)$$

The left-hand side of equation (13) is the power spectrum of the observed excess noise. Whereas the right-hand side is the contribution of sidelobes of sources due to the chromatic PSF. Equation (13) implies that in an ideal case, where the sky signal does not change in consecutive sub-bands, nor other effects contribute such as the ionosphere or imperfect calibration, the excess noise should be same as the differential sidelobe noise. We compute the power spectra of Stokes I, V and the PSF using uniform weighted images produced by EXCON. The PSF images are produced by replacing all visibility data points by unity. We use the PSF at the centre of the field in this test, assuming that the PSF does not vary significantly towards different directions. Fig. 5 shows the observed total excess noise and estimated contribution of the sidelobe noise, i.e. the right-hand side of equation (13), computed before and after the direction-dependent calibration and source subtraction. The differential sidelobes amount to the total observed excess noise before source subtraction. However, it is only a small fraction of the excess noise after source subtraction. This suggests that the excess noise might have been introduced in the data during the source subtraction, and we will discuss this in detail in Section 5.

4.3 Correlation with the ionospheric scintillation

As discussed in Section 4.1, the angular power spectrum of the excess noise suggests that ionospheric scintillation is also unlikely to be a dominant cause of the excess noise. However, in this subsection, we further study any possible correlation of the excess noise with the ionospheric conditions in more detail. The ionosphere introduces stochastic phase fluctuations in the low-frequency radio signals. Vedantham & Koopmans (2015, 2016) have studied the scintillation noise due to ionospheric diffraction of discrete sources in the case of wide-field interferometry. We expect the ionospheric scintillation noise to be higher when the diffractive scale is shorter (Vedantham & Koopmans 2015, 2016).

We briefly discuss here how we compute the diffractive scales from the data, but a more detailed description can be found in Mevius et al. (2016). For each baseline, we compute the time series of the phase difference between the direction-independent gain solutions of the pair of stations forming the baseline. We then compute the structure function which is the variance of the time series of the phase difference as a function of the baseline length. The structure function is fit to a power law, and it is expected to have a power-law index of 5/3 for a Kolmogorov-type turbulence. The diffractive

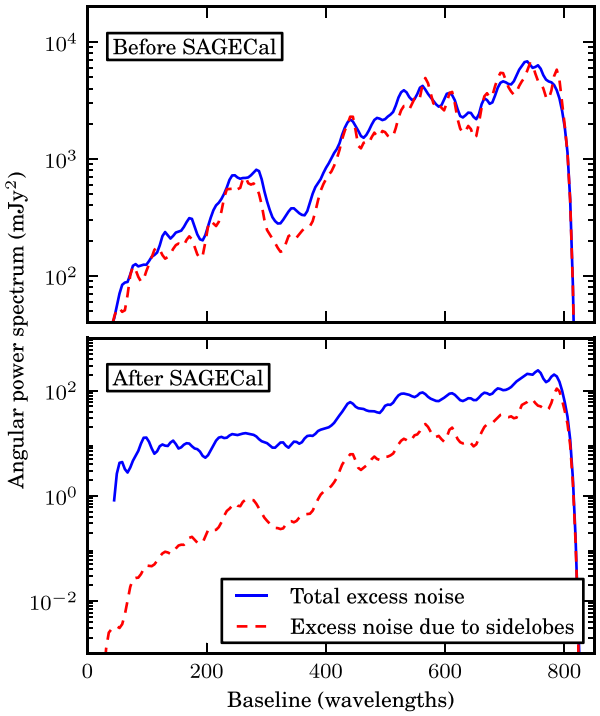

Figure 5. Comparison of the total observed differential excess noise in differential Stokes I images with the differential sidelobe noise due to the chromatic PSF. Top panel: differential sidelobes account for the total excess noise before the direction-dependent (DD) calibration and source subtraction with SAGECAL. Bottom panel: the total excess noise is much higher than the contribution due to the differential sidelobes after the DD calibration.

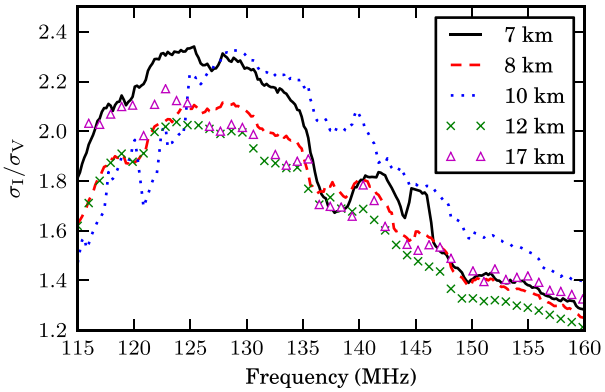

Figure 6. The ratio of the rms of SAGECAL residuals in Stokes I to Stokes V as a function of frequency for different diffractive scales in the ionosphere observed on different nights. The diffractive scales are mentioned at 150 MHz. The shorter the diffractive scale, the higher the ionospheric scintillation noise. However, the noise in the data does not show an obvious anticorrelation with the diffractive scale.

scale is the baseline length at which the phase variance is 1 rad^2. In Fig. 6, we show the ratio of Stokes I to Stokes V rms for different nights of observations with different diffractive scales. We do not find any obvious anticorrelation between the excess noise and the diffractive scale in the ionosphere. This again confirms our conclusion based on the angular power spectrum of the excess noise that the ionosphere is unlikely to be the dominant cause of the excess noise.

We should note that we have seen an anticorrelation between the ionospheric diffractive scale and the noise before the

direction-dependent calibration and source subtraction in our other target field towards 3C196 which contains brighter sources (Mevius et al. 2016). This effect might be difficult to see in the NCP field which does not contain bright sources. Furthermore, the travelling ionospheric disturbances are prominent on time-scales of few minutes, and their effect is likely removed from the NCP data during the direction-dependent calibration.

5 SIMULATIONS

In this section, we test whether the direction dependent calibration can introduce an excess noise using simulations of the calibration and source subtraction process, where effects of the chromatic PSF and ionosphere are eliminated. The simulated mock data sets contain discrete sources, diffuse foregrounds and thermal noise. SAGECAL is then used to obtain station gains and subtract the discrete sources. The steps involved in the simulations are as follows.

(i) 25 sources with brightest apparent (i.e. observed) fluxes are selected from the NCP sky model and their visibilities are predicted. The NCP sky model is constructed from the observed data and contains sources within a radius of 20° around the NCP. The selected brightest 25 sources are located within a radius of 7° from the NCP, and their flux densities range from 5 to 0.24 Jy.

(ii) We predict the Stokes I visibilities of the simulated diffuse foregrounds from Jelić et al. (2008, 2010) multiplied with a time-averaged primary beam of LOFAR. The rms flux density of these diffuse foregrounds is normalized to 5 mJy/PSF, i.e. 7 K of brightness temperature. We do not know the brightness temperature of the diffuse foregrounds in the NCP field in total intensity, but we have assumed it to be 10 times the brightness temperature of the observed polarized diffuse foregrounds in the field.

(iii) The thermal noise of rms 1.5 Jy per visibility is simulated at the resolution of 10 s, 183 kHz at 135 MHz. This results into an rms noise of 0.83 mJy per sub-band image for a 13 h long observation, which is comparable to the observed noise in Stokes V images in the data.

(iv) Visibilities of the discrete sources, diffuse foregrounds and the thermal noise are added to form a mock data set.

(v) SAGECAL is used to calibrate the station gains and remove discrete sources from the simulated data. We cluster the simulated 25 sources in 21 directions for which the station gain solutions are obtained. We keep the number of directions small so that the calibration remains an overdetermined system.[3]

While predicting visibilities for discrete sources, we increase their fluxes by 5 per cent. This is equivalent to station gains being higher than their expected values. This way, we ensure that the actual values of gain solutions in the calibration are not the same as the initial values used in calibration iterations. Such absolute scaling of fluxes does not affect the end result. However, if we were to vary relative fluxes of sources grouped within a cluster that would affect the common solution for that group of sources. In the following subsections, we present the results of different tests performed with the simulations.

[3] Radio interferometric calibration can be considered to be an equivalent of the factor analysis technique, as described in Sardarabadi (2016). For P interferometric elements, the maximum number of directions in which the gain solutions can be obtained, is given by $P - \sqrt{P}$ (chapter 4, Sardarabadi 2016). Therefore, in the case of 64 LOFAR stations in the Netherlands, one can solve for maximum 56 directions in an instantaneous monochromatic snapshot. We use 5 to 20 min time intervals in SAGECAL which provide more constrains.

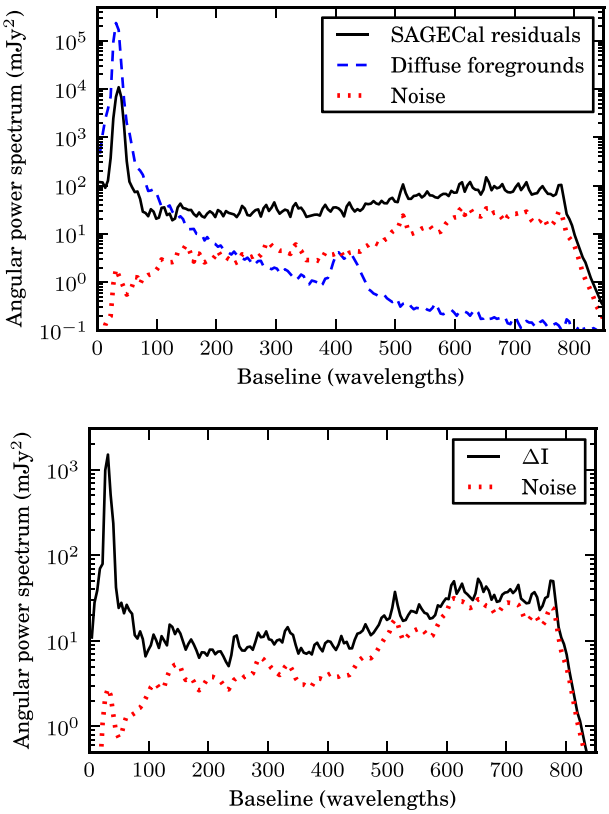

Figure 7. Results from multiple noise realizations of one frequency sub-band. Top panel: angular power spectra of the input diffuse foregrounds, thermal noise and SAGECAL residuals after source subtraction. The diffuse foregrounds are suppressed at short baselines in residuals, whereas long baselines show excess power above the thermal noise. Bottom panel: differential residuals (ΔI) between different noise realizations, which are higher than the thermal noise. The simulated data contains 25 discrete sources (5–0.24 Jy), the diffuse foregrounds (7 K) and the thermal noise (0.83 mJy/PSF).

5.1 Different noise realizations of one sub-band

Here, we simulate multiple realizations of the mock data for one frequency sub-band at 135 MHz. Different realizations contain the same discrete and diffuse foregrounds but different realizations of the thermal noise. The advantage of this test is that we exclude effects of the chromatic PSF in this analysis. Ideally, we expect the discrete sources to get perfectly subtracted and the diffuse foregrounds with the thermal noise to be left as residuals. However, as shown in the top panel of Fig. 7, we find an excess of power in the residuals at baselines longer than 200 wavelengths, i.e. the discrete sources are not perfectly subtracted. Additionally, the power at short baselines is suppressed, i.e. the diffuse foregrounds are partially removed during the source subtraction. As the diffuse foregrounds remain the same in different data realizations, we expect the difference between the residuals of different realizations to be consistent with the thermal noise. However, as shown in the bottom panel of Fig. 7, we see an excess of flux in the differential residuals of different realizations. The power spectrum of the differential residuals resembles thermal noise only at baselines longer than 100 wavelengths. At shorter baselines, the diffuse foregrounds affect the power spectrum of the residuals.

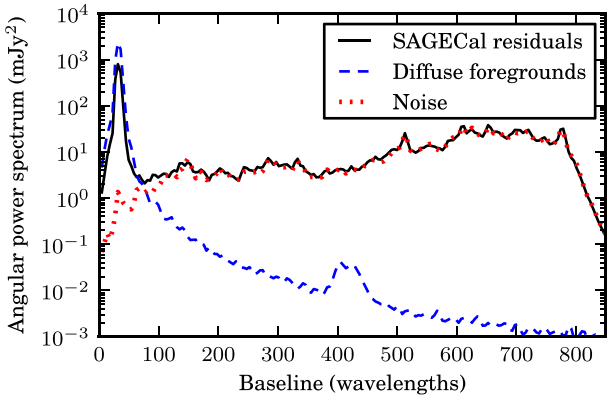

Figure 8. Simulation results, same as Fig. 7, except here the brightness of the diffuse foregrounds is reduced by 10 times. The foreground suppression is reduced, and the excess noise has disappeared as compared to Fig. 7, showing that these systematic effects are functions of the unmodelled flux due to the diffuse foregrounds.

We find that both the suppression of the diffuse foregrounds and the excess noise depend on the brightness of the diffuse foregrounds which are not part of the sky model. In Fig. 8, we show the results when the intensity of the diffuse foregrounds is reduced by a factor 10 to have an rms of 0.7 K. The suppression of foregrounds is reduced, and the residuals reach the thermal noise at long baselines. This test shows that both the foreground suppression and the excess noise problems occur when the sky model used in self-calibration and source subtraction is incomplete. Additionally, the intensity of these problems depends on the missing flux in the model. Barry et al. (2016) suggested unmodelled foregrounds convolved with a chromatic PSF as the source of variations in calibration solutions and an excess noise. However, as evident from this test, unmodelled flux in itself could be sufficient to cause variations in calibration solutions.

5.2 Multiple SAGECAL runs on the same realization of simulation

In order to understand the interplay between unmodelled flux and the thermal noise, we study results of multiple calibration runs on the same realization of the thermal noise in this subsection. Different calibration runs on the same data may not find the exact same gain solutions due to any randomization implemented in the calibration algorithm. In every expectation maximization step in SAGECAL, the order in which the station gains in different directions are solved, is randomized to reduce the systematic errors in the solver. However, the final solution in every run of SAGECAL is expected to reach the global minimum in the likelihood space. Differences between the residuals of different calibration runs on the same data should then be near zero. We find that this is not the case. For a simulation containing discrete sources in the flux range 5–0.24 Jy and diffuse foregrounds of rms brightness temperature 0.7 K, the differential noise is 10 per cent of the thermal noise. The level of this excess noise depends on the relative fluxes of the discrete sources and the diffuse foregrounds as summarized in Table 2. As shown in Fig. 9, the power spectrum of the differential noise resembles that of the thermal noise just as observed in the real data, unless the unmodelled flux dominates on certain baselines which was the case in Fig. 7. This test provides a possible explanation for the excess noise. We think that the unmodelled flux due to the diffuse foregrounds alters the likelihood function of calibration parameters in such a way that

Table 2. The differential noise (ΔI) in residuals of multiple SAGECAL runs on the same realization of the simulated data for different levels of discrete and diffuse foregrounds. The diffuse foregrounds are mentioned in flux densities of rms/PSF and in rms brightness temperature in parentheses. The differential noise in residuals is mentioned as a percentage of the thermal noise.

Discrete sources	Diffuse foregrounds	ΔI/Noise
5 to 0.24 Jy	5 mJy (7 K)	130 per cent
5 to 0.24 Jy	0.5 mJy (0.7 K)	10 per cent
0.5 to 0.24 Jy	0.5 mJy (0.7 K)	25 per cent

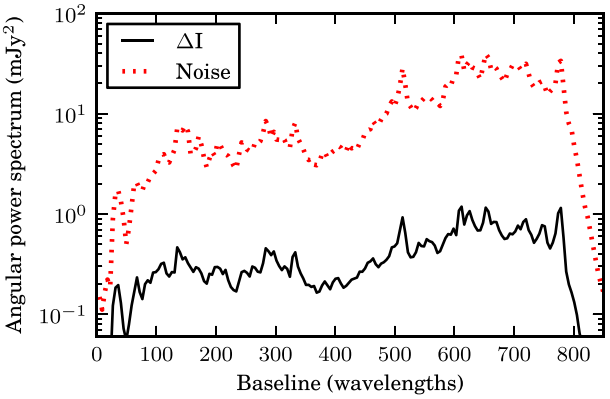

Figure 9. Results from multiple SAGECAL runs on one realization of the simulation. The difference between residuals of different runs (ΔI) is 10 per cent of the thermal noise, and it has the same power spectrum as the thermal noise.

the maximum-likelihood (ML) condition becomes degenerate, i.e. multiple sets of calibration parameters satisfy the condition. The calibration could find any one of these sets of parameters as the gain solution in a run. If the obtained solution is different than the true ML solution, it will lead to residuals in source subtraction containing excess power beyond the thermal noise. However, the difference between the residuals of any two solutions would have the same statistical properties as the thermal noise, because both solutions satisfy the ML condition of the altered likelihood function. This hypothesis could in principle be verified by sampling the likelihood space of calibration parameters. However, this is computationally very expensive for our parameter space of high dimensions (21 directions × 64 stations × 2 polarization components).

In reality, we will not calibrate hundreds of hours of LOFAR-EoR data multiple times, because the direction dependent calibration is a computationally expensive process. However, an important implication and prediction of the above explanation of the excess noise is that the excess noise produced as an artefact of the calibration with an incomplete sky model should not correlate among different calibration runs or even different data sets observed on different nights. As shown in Fig. 10, we indeed find that differential Stokes I images of pairs of observations show only about 10 per cent correlation. All observations used here are from LOFAR cycle 0 and are calibrated with the same sky model. We compute the correlation coefficient between two observations as

$$C_{12} = \frac{\langle \Delta i_1 \times \Delta i_2 \rangle}{\sqrt{(\langle \Delta i_1^2 \rangle - \langle \Delta v_1^2 \rangle)(\langle \Delta i_2^2 \rangle - \langle \Delta v_2^2 \rangle)}}, \quad (15)$$

where Δi_k, Δv_k are differential Stokes I and Stokes V images of the kth observation, respectively. Differential images are obtained by subtracting consecutive frequency sub-bands. We subtract

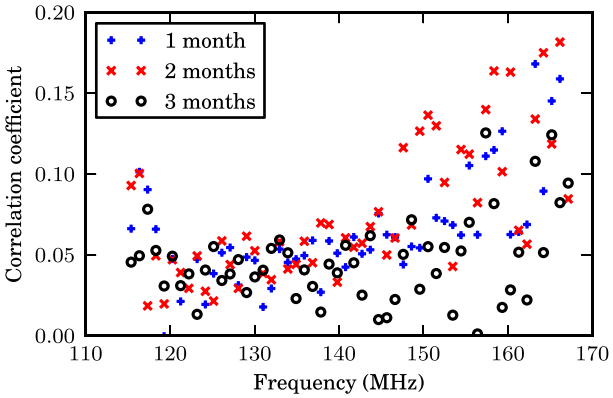

Figure 10. Correlation coefficient between differential Stokes I images of pairs of observations separated by 1, 2 and 3 months.

respective variances of Stokes V images in the denominator to correct the correlation coefficient for the fact that two observations contain different realizations of the thermal noise. The small positive correlation coefficients observed in Fig. 10 could be due to chromatic sidelobes of residual sources in Stokes I images. If the excess noise introduced as an artefact of the calibration is uncorrelated among different observations, its rms will reduce with the square root of the total observation time as we integrate more data. However, we must note that a part of decorrelation observed in Fig. 10 is due to the rotation of foreground sources with respect to the NCP. Although images are made such that a source appears at the same position in images from different observations, the source actually gets convolved with different PSFs at different times of a year depending on its position on the sky. Therefore, sidelobes of sources should partially decorrelate among different observations.

6 POSSIBLE SOLUTIONS TO THE FOREGROUND SUPPRESSION AND EXCESS NOISE

The simulations presented in Section 5 have shown a clear evidence that the excess noise and suppression of the diffuse foregrounds occur because of an incomplete model in self-calibration. We now discuss two possible solutions to mitigate these systematic errors.

6.1 Excluding short baselines from calibration

The diffuse foregrounds are not part of the sky model, but they are dominant only on short baselines. Their brightness is negligible at baselines longer than 200 wavelengths as compared to the discrete sources in total intensity in the NCP field. We can use baselines only longer than 200 wavelengths to obtain gain solutions for all stations and then subtract sources on all baselines. In such a case, the diffuse foregrounds would affect the self-calibration at a much reduced level. In Fig. 11, we compare SAGECAL residuals when all and only long baselines are used for the calibration of the 25 brightest sources in the NCP field in the presence of 7 K diffuse foregrounds. In the former case, the suppression of the diffuse foregrounds and residuals of the discrete sources is evident. Both of these issues are mitigated in the latter case. The top panel of Fig. 12 shows the same phenomenon in the form of angular power spectra. When the short baselines are excluded in the calibration, the diffuse foregrounds remain untouched in the residuals at short baselines. Additionally, there is no excess noise at long baselines because the discrete sources are perfectly removed. Excluding short baselines,

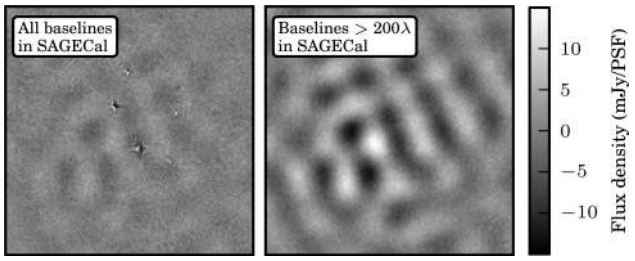

Figure 11. Comparison of SAGECAL residuals (uniform weighted, 10° images) when all baselines are used for calibration (left-hand panel) and only baselines longer than 200 wavelengths are used (right-hand panel). When all baselines are used, the sky model is incomplete due to the missing diffuse foregrounds. As a result, the diffuse foregrounds are suppressed and the discrete sources are imperfectly subtracted. Excluding short baselines in the calibration resolves both of these issues.

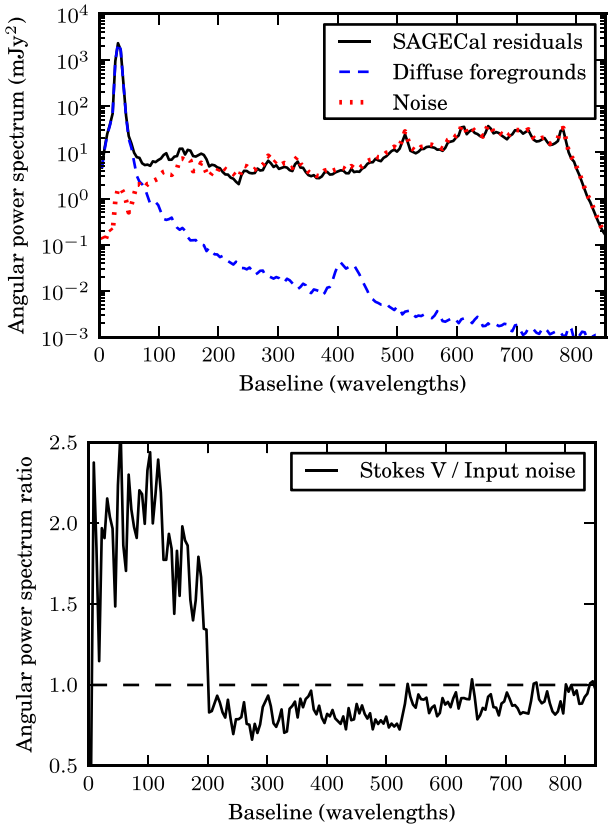

Figure 12. Excluding baselines shorter than 200 wavelengths in calibration. Top panel: the SAGECAL residuals contain the diffuse foregrounds without any suppression. Additionally, the residuals reach the thermal noise at longer baselines implying perfect removal of the discrete sources and no excess noise. Bottom panel: the ratio of the power spectrum of the thermal noise after source removal (Stokes V) to that of the input noise. The noise is enhanced by a factor of 2 on baselines which were excluded from calibration.

however, has a severe disadvantage as it enhances the noise on the excluded baselines. In the bottom panel of Fig. 12, we plot the ratio of the power spectrum of the noise (Stokes V) after source subtraction to that of the input noise. The noise on the excluded baselines is boosted by a factor of 2 in power. A mathematical derivation of this phenomenon is given in the appendix for interested readers. The enhancement of noise implies a loss in sensitivity on short baselines, i.e. large angular scales, which otherwise would have been

most promising for a detection of the 21 cm signal (Zaroubi et al. 2012; Patil et al. 2014). As evident in the lower panel of Fig. 12, the thermal noise is suppressed by 10 per cent on long baselines which are used for the calibration. However, this suppression would not affect any further analysis because these long baselines will only be used for the calibration but not for a detection of the 21 cm signal.

6.2 Simultaneous multifrequency calibration

The calibration is often performed on one frequency sub-band at a time due to computing and memory constraints. This gives the station-gain solutions a partial freedom to vary independently at different sub-bands, producing an excess noise which is uncorrelated along frequency, as also shown by Barry et al. (2016). As seen in our data as well as simulations, the power spectrum of this excess noise is similar to that of the thermal noise. Trott & Wayth (2016) also reached to the same conclusion in the context of bandpass calibration.

The primary beam as well as any ionospheric effects vary smoothly with frequency. Therefore, a parametric calibration can be obtained for a large bandwidth instead of independent gain solutions at each sub-band. Barry et al. (2016) suggested fitting a low-order polynomial to gain solutions along frequency or averaging calibration solutions of multiple interferometric elements. Alternatively, Yatawatta (2015b) have proposed a regularization which enforces smoothness on the calibration solutions to a degree depending on the chosen value of the regularization parameter. As a result, the errors on the station gains are reduced, although the theoretical limit based on the thermal noise cannot be reached due to the model incompleteness. We also believe that a simultaneous multifrequency calibration should reduce the suppression in the diffuse foregrounds. Nunhokee (2016) have shown that using longer time intervals for the calibration reduces suppression of unmodelled flux. The unmodelled flux due to the diffuse foregrounds changes significantly from 115 to 170 MHz. Therefore, the suppression should be reduced if the entire or a significant fraction of the bandwidth is simultaneously used to constrain the calibration solutions. We leave a more detailed analysis of the multifrequency calibration for future work.

7 CONCLUSIONS

The LOFAR EoR project aims to detect the redshifted 21 cm emission from neutral hydrogen from redshift 6 to 11. It is crucial to control the systematic errors for a signal detection, because the foregrounds are several orders of magnitude brighter than the expected signal. In this paper, we have studied two systematic biases observed in the residual LOFAR-EoR data after calibration and subtraction of bright discrete foreground sources: (i) a suppression in the diffuse emission and (ii) excess of noise beyond the thermal component. These biases occur because of the direction-dependent calibration with an incomplete sky model, and they are potential obstacles in a signal detection for the following reasons.

(i) Both the diffuse foregrounds and the 21 cm signal are easiest to detect on large angular scales, and the suppression of the former might imply a suppression of the 21 cm signal as well.

(ii) The excess noise implies a loss in sensitivity and an additional bias in a measurement of the power spectrum of the 21 cm signal. Furthermore, the excess noise would not be removed by the foreground removal methods which remove spectrally smooth signals.

The differential noise between two closely spaced frequency bins after removing the bright sources from the data is higher than the thermal noise. We call this additional noise: 'excess noise'. We have performed tests to study properties of the excess noise and identify its causes. The angular power spectrum of the excess noise resembles that of the thermal noise, i.e. it shows the same power on all baselines. The chromatic PSF and ionospheric scintillation would have shown increasing power with the baseline length. We have estimated that the contribution of sidelobes of the unsubtracted sources due to the chromatic PSF is only a small fraction of the excess noise. The excess noise in different observations does not show any obvious correlations with the diffractive scales in the ionosphere on respective nights. Therefore, we establish that the chromatic PSF and ionosphere scintillation cannot be the dominant causes of the excess noise.

We use simulated data sets to study the systematic errors that could be produced by the calibration and source subtraction algorithms. Just like the real data, the discrete sources are removed by modelling them, calibrating the LOFAR station gains in their directions and then subtracting the sources. The calibration minimizes the difference between the data and the model by adjusting the station gains. In this process, the diffuse foregrounds are suppressed, because they are not part of the model. This also results in imperfect removal of the discrete sources. The source residuals are partially uncorrelated in multiple noise realizations of the simulated data. This could explain the excess noise in the difference between two frequency bins in the actual data which contain uncorrelated realizations of the thermal noise. The angular power spectrum of the excess noise resembles that of the thermal noise in the simulations, just as it does in the actual data, and its magnitude depends on the amount of flux that is included in the sky model relative to the amount of flux that is excluded in the model.

We find that multiple randomized calibration runs of one data set lead to different realizations of the excess noise. Although not yet proven, our interpretation of this finding is that unmodelled flux alters the likelihood function of calibration parameters such that the ML condition becomes degenerate for multiple parameter values. An important implication of this interpretation is that the excess noise among different observations should be uncorrelated, which we verify from our observations. Therefore, although calibration with an incomplete model introduces extra residuals in the data, these residuals will reduce as the square root of the total observation time as we average multiple observations.

We discuss two possible solutions to the observed systematic biases, i.e. the foreground suppression and the excess noise. First, short baselines where the diffuse foregrounds are dominant, can be excluded from the calibration. This ensures that the diffuse foregrounds and the 21 cm signal are not suppressed. However, it enhances the noise on the excluded baselines, implying a poor sensitivity on large angular scales where a detection of the 21 cm signal otherwise would have been most promising. Secondly, we believe a better solution would be to use multifrequency constraints to enforce spectral smoothness on the calibration parameters. Our future efforts are going to be focused on that front (Yatawatta 2015b).

ACKNOWLEDGEMENTS

We thank the anonymous reviewer for their helpful comments, which improved the content of this paper. AHP and SZ thank the Lady Davis Foundation and The Netherlands Organization for Scientific Research (NWO) VICI grant for the financial support. LVEK and BKG acknowledge the financial support from the European

Research Council under ERC-Starting Grant FIRSTLIGHT – 258942. AGdB, SY, MM and VNP acknowledge support by the ERC for project 339743 (LOFARCORE). VJ acknowledges the NWO for the financial support under VENI grant – 639.041.336. ITI was supported by the Science and Technology Facilities Council [grant number ST/L000652/1]. The LOFAR was designed and constructed by ASTRON, the Netherlands Institute for Radio Astronomy, and has facilities in several countries, which are owned by various parties (each with their own funding sources) and are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

REFERENCES

Ali Z. S. et al., 2015, ApJ, 809, 61
Asad K. M. B. et al., 2015, MNRAS, 451, 3709
Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135
Boonstra A., van der Veen A., 2003, IEEE Trans. Signal Process., 51, 25
Bowman J. D. et al., 2013, PASA, 30, e031
Brentjens M. A., de Bruyn A. G., 2005, A&AS, 441, 1217
Chapman E. et al., 2012, MNRAS, 423, 2518
Chapman E. et al., 2013, MNRAS, 429, 165
Chapman E. et al., 2015, Proc. Sci., Cosmic Dawn and Epoch of Reionization Foreground Removal with the SKA. SISSA, Trieste. PoS#005
Cook R. D., Weisberg S., 1982, Residuals and Influence in Regression, Monographs on Statistics and Applied Probability. Chapman and Hall, New York
Cooray A., Furlanetto S. R., 2004, ApJ, 606, L5
Cornwell T. J., Wilkinson P. N., 1981, MNRAS, 196, 1067
Datta A., Bhatnagar S., Carilli C. L., 2009, ApJ, 703, 1851
Datta A., Bowman J. D., Carilli C. L., 2010, ApJ, 724, 526
DeBoer D. R. et al., 2016, preprint (arXiv:1606.07473)
Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
Di Matteo T., Ciardi B., Miniati F., 2004, MNRAS, 355, 1053
Dillon J. S. et al., 2015, Phys. Rev. D, 91, 123011
Ewall-Wice A. et al., 2016, MNRAS, 460, 4320
Grobler T. L., Nunhokee C. D., Smirnov O. M., van Zyl A. J., de Bruyn A. G., 2014, MNRAS, 439, 4030
Hamaker J. P., Bregman J. D., Sault R. J., 1996, A&A, 117, 137
Harker G. et al., 2009, MNRAS, 397, 1138
Harker G. et al., 2010, MNRAS, 405, 2492
Hinshaw G. et al., 2013, ApJS, 208, 19
Jelić V. et al., 2008, MNRAS, 389, 1319
Jelić V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, MNRAS, 409, 1647
Jelić V. et al., 2014, A&AS, 568, A101
Jelić V. et al., 2015, A&AS, 583, A137
Kazemi S., Yatawatta S., 2013, MNRAS, 435, 597
Kazemi S., Yatawatta S., Zaroubi S., Lampropoulos P., de Bruyn A. G., Koopmans L. V. E., Noordam J., 2011, MNRAS, 414, 1656
Kazemi S., Yatawatta S., Zaroubi S., 2012, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Inc., NY, p. 2533
Kazemi S., Yatawatta S., Zaroubi S., 2013a, MNRAS, 430, 1457
Kazemi S., Yatawatta S., Zaroubi S., 2013b, MNRAS, 434, 3130
Koopmans L. V. E., 2010, ApJ, 718, 963
Koopmans L. et al., 2015, Proc. Sci., The Cosmic Dawn and Epoch of Reionisation with SKA. SISSA, Trieste. PoS#001
Laurent R. T. S., Cook R. D., 1992, Journal of the American Statistical Association, 87, p. 985
Laurent R. T. S., Cook R. D., 1993, Biometrika, 80, 99
Lonsdale C. J., 2005, in Kassim N., Perez M., Junor W., Henning P., eds, ASP Conf. Ser. Vol. 345, From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science. Astron. Soc. Pac., San Francisco, p. 399
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mevius M. et al., 2016, Radio Sci., 51, 927
Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, IEEE J. Sel. Top. Signal Process., 2, 707
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137
Neugebauer S. P., 1996, Master's thesis, Virginia Polytechnic Institute and State University
Noordam J. E., de Bruyn A. G., 1982, Nature, 299, 597
Nunhokee C. D., 2016, Master's thesis, Rhodes University
Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, MNRAS, 405, 155
Offringa A. R., van de Gronde J. J., Roerdink J. B. T. M., 2012, A&AS, 539, A95
Oh S. P., Mack K. J., 2003, MNRAS, 346, 871
Paciga G. et al., 2013, MNRAS, 433, 639
Pandey V. N., van Zwieten J. E., de Bruyn A. G., Nijboer R., 2009, in Saikia D. J., Green D. A., Gupta Y., Venturi T., eds, ASP Conf. Ser. Vol. 407, The Low-Frequency Radio Universe. Astron. Soc. Pac., San Francisco, p. 384
Parsons A. R. et al., 2010, AJ, 139, 1468
Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, ApJ, 756, 165
Patil A. H., 2014, XXXIth URSI General Assembly and Scientific Symposium (URSI GASS). IEEE, Inc., NY
Patil A. H. et al., 2014, MNRAS, 443, 1113
Planck Collaboration XLVII, 2016, A&A, preprint (arXiv:1605.03507)
Ross W. H., 1987, Can. J. Stat., 15, 91
Sardarabadi A. M., 2016, PhD thesis, Technische Universiteit Delft
Schwab F. R., 1980, in Rhodes W. T., ed., Proc. SPIE Conf. Ser. Vol. 231, International Optical Computing Conference I. SPIE, Bellingham, p. 18
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&AS, 345, 380
Thompson A. R., Moran J. M., Swenson G. W., 2007, Interferometry and Synthesis in Radio Astronomy. Wiley, New York
Tingay S. J. et al., 2013, PASA, 30, 7
Trott C. M., Wayth R. B., 2016, PASA, 33, e019
Trott C. M. et al., 2016, ApJ, 818, 139
van der Tol S., Jeffs B. D., van der Veen A.-J., 2007, IEEE Trans. Signal Process., 55, 4497
van der Veen A. J., Leshem A., Boonstra A. J., 2005, in Peter J. H., ed., Array Signal Processing for Radio Astronomy. Springer-Verlag, Netherlands, p. 231
van Haarlem M. P. et al., 2013, A&AS, 556, A2
Vedantham H. K., Koopmans L. V. E., 2015, MNRAS, 453, 925
Vedantham H. K., Koopmans L. V. E., 2016, MNRAS, 458, 3099
Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176
Wieringa M. H., 1992, Exp. Astron., 2, 203
Wijnholds S. J., van der Veen A.-J., 2009, IEEE Trans. Signal Process., 57, 3512
Wijnholds S. J., Grobler T. L., Smirnov O. M., 2016, MNRAS, 457, 2331
Wilkinson P. N., Conway J., Biretta J., 1988, in Reid M. J., Moran J. M., eds, Proc. IAU Symp. 129, The Impact of VLBI on Astrophysics and Geophysics. Reidel, Dordrecht, p. 509
Yatawatta S., 2014, XXXIth URSI General Assembly and Scientific Symposium (URSI GASS). IEEE, Inc., NY
Yatawatta S., 2015a, 1st URSI Atlantic Radio Science Conference (URSI AT-RASC). IEEE, Inc., NY, p. 1
Yatawatta S., 2015b, MNRAS, 449, 4506
Yatawatta S., Zaroubi S., de Bruyn G., Koopmans L., Noordam J., 2009, IEEE 13th Digit. Signal Process. Workshop and 5th IEEE Signal Process. Educ. Workshop. IEEE, Inc., NY, p. 150
Yatawatta S. et al., 2013, A&AS, 550, A136
Zaroubi S. et al., 2012, MNRAS, 425, 2964
Zheng H. et al., 2014, MNRAS, 445, 1084
Zmuidzinas J., 2003, J. Opt. Soc. Am. A, 20, 218

APPENDIX A: LEVERAGE AS A DIAGNOSTIC IN CALIBRATION

In this appendix, we provide a mathematical proof of the enhancement of noise on baselines which are excluded in calibration. We use Leverage, a well-known concept in regression analysis, to study the performance of calibration. Leverage (Cook & Weisberg 1982) can be loosely described as the change in the predicted value based on the data model used, due to the change in the data used for estimating the calibration parameters. In non-linear regression, Jacobian Leverage (Laurent & Cook 1992, 1993) is widely used (Neugebauer 1996). Here, we apply it to study calibration. We adopt a case deletion model in regression (Ross 1987) to study the situation where only a subset of baselines (or data points) are used for calibration (Yatawatta 2015a).

A1 Radio interferometric calibration

Here, we give a brief overview of the data model used in radio interferometric calibration (Hamaker, Bregman & Sault 1996; Thompson, Moran & Swenson 2007). In interferometry, the correlated signal from pth and qth stations, $\mathbf{V}_{pq}$ is given by

$$\mathbf{V}_{pq} = \sum_{i=1}^{K} \mathbf{J}_{pi} \mathbf{C}_{pqi} \mathbf{J}_{qi}^{H} + \mathbf{N}_{pq}, \tag{A1}$$

where $\mathbf{J}_{pi}$ and $\mathbf{J}_{qi}$ are the Jones matrices describing errors along the direction of source i at stations p and q, respectively. The matrices represent the effects of the propagation medium, the beam shape and the receiver. There are K sources in the sky model and the noise matrix is given as $\mathbf{N}_{pq}$. The contribution from the ith source on baseline pq is given by the coherency matrix $\mathbf{C}_{pqi}$. We estimate the Jones matrices $\mathbf{J}_{pi}$ for $p \in [1, R]$ and $i \in [1, K]$, during calibration and calculate the residuals by subtracting the predicted model (multiplied with the estimated Jones matrices) from the data. The vectorized form of (A1), $\boldsymbol{v}_{pq} = \mathrm{vec}(\mathbf{V}_{pq})$ can be written as

$$\boldsymbol{v}_{pq} = \sum_{i=1}^{K} \mathbf{J}_{qi}^{\star} \otimes \mathbf{J}_{pi} \mathrm{vec}(\mathbf{C}_{pqi}) + \boldsymbol{n}_{pq} \tag{A2}$$

where $\boldsymbol{n}_{pq} = \mathrm{vec}(\mathbf{N}_{pq})$. Depending on the time and frequency interval within which calibration solutions are obtained, we can stack up all cross-correlations within that interval as

$$\boldsymbol{d} = \left[\mathrm{real}\left(\boldsymbol{v}_{12}^{T}\right)\ \mathrm{imag}\left(\boldsymbol{v}_{12}^{T}\right)\ \mathrm{real}\left(\boldsymbol{v}_{13}^{T}\right) \ldots\ldots \mathrm{imag}\left(\boldsymbol{v}_{(R-1)R}^{T}\right)\right]^{T}, \tag{A3}$$

where $\boldsymbol{d}$ is a vector of size $N \times 1$ of real data points. Thereafter, we have the data model

$$\boldsymbol{d} = \sum_{i=1}^{K} \boldsymbol{s}_i(\boldsymbol{\theta}) + \boldsymbol{n}, \tag{A4}$$

where $\boldsymbol{\theta}$ is the real parameter vector (size $M \times 1$) that is estimated by calibration. The contribution of the ith known source on all data points is given by $\boldsymbol{s}_i(\boldsymbol{\theta})$ (size $N \times 1$). The noise vector is given by $\boldsymbol{n}$ (size $N \times 1$). The parameters $\boldsymbol{\theta}$ are the elements of $\mathbf{J}_{pi}$-s, with real and imaginary parts considered separately.

The ML estimate of $\boldsymbol{\theta}$ under zero mean, white Gaussian noise is obtained by minimizing the least-squares cost

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \left\| \boldsymbol{d} - \sum_{i=1}^{K} \boldsymbol{s}_i(\boldsymbol{\theta}) \right\|^2 \tag{A5}$$

as done in current calibration approaches (Boonstra & van der Veen 2003; van der Veen, Leshem & Boonstra 2005; Kazemi et al. 2011) and this is improved by using a weighted least-squares estimator to account for errors in the sky model (Kazemi & Yatawatta 2013). The Cramer–Rao lower bound is used to find a lower bound to the variance of $\widehat{\boldsymbol{\theta}}$ (Zmuidzinas 2003; van der Tol, Jeffs & van der Veen 2007; Wijnholds & van der Veen 2009; Kazemi, Yatawatta & Zaroubi 2012). However, relating this lower bound to the residual $\boldsymbol{d} - \sum_{i=1}^{K} \boldsymbol{s}_i(\widehat{\boldsymbol{\theta}})$ is not simple. Instead, we propose Leverage to quantify errors on the residuals.

A2 Leverage

Consider a non-linear regression model

$$\boldsymbol{y} = \boldsymbol{m}(\boldsymbol{\theta}) + \boldsymbol{n}, \tag{A6}$$

where $\boldsymbol{y}$ is a $N \times 1$ data vector, $\boldsymbol{n}$ is the $N \times 1$ noise vector, and $\boldsymbol{m}(\boldsymbol{\theta})$ is a non-linear function of the $M \times 1$ parameter vector $\boldsymbol{\theta}$. The residual vector $\boldsymbol{r}(\boldsymbol{\theta})$ is given by

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{y} - \boldsymbol{m}(\boldsymbol{\theta}). \tag{A7}$$

The estimated value of $\boldsymbol{\theta}$ using (weighted) least squares is given by $\widehat{\boldsymbol{\theta}}$ and the predicted value based on the estimated parameters is given by $\widehat{\boldsymbol{y}} = \boldsymbol{m}(\widehat{\boldsymbol{\theta}})$. Now consider perturbing the data by $b\boldsymbol{f}$ where $\boldsymbol{f}$ ($N \times 1$) is any arbitrary vector and b is a real scalar. Let us call the perturbed data as $\boldsymbol{y}_b$ and the estimated value of $\boldsymbol{\theta}$ using the perturbed data as $\widehat{\boldsymbol{\theta}}_b$. The predicted value using $\widehat{\boldsymbol{\theta}}_b$ is denoted by $\widehat{\boldsymbol{y}}_b$. We define the leverage vector as (Laurent & Cook 1992)

$$\boldsymbol{g} \triangleq \lim_{b \to 0} \frac{1}{b} \left(\widehat{\boldsymbol{y}}_b - \widehat{\boldsymbol{y}} \right), \tag{A8}$$

and for (weighted) least-squares estimation, we define Jacobain leverage as (Laurent & Cook 1993)

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) \triangleq \boldsymbol{\eta}_{\theta} \left(\boldsymbol{\eta}_{\theta}^{T} \boldsymbol{\eta}_{\theta} - \sum_{i=1}^{N} \boldsymbol{r}^{i} \left((\boldsymbol{\eta}^{i})_{\theta\theta} \right) \right)^{-1} \boldsymbol{\eta}_{\theta}^{T},$$

$$\boldsymbol{\eta}_{\theta} = \frac{\partial}{\partial \boldsymbol{\theta}^{T}} \boldsymbol{m}(\boldsymbol{\theta}),$$

$$(\boldsymbol{\eta}^{i})_{\theta\theta} = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{T}} \boldsymbol{m}^{i}(\boldsymbol{\theta}), \tag{A9}$$

where $\boldsymbol{r}^{i}$ is the ith element in $\boldsymbol{r}(\boldsymbol{\theta})$ and $\boldsymbol{m}^{i}(\boldsymbol{\theta})$ is the ith element in $\boldsymbol{m}(\boldsymbol{\theta})$. We see that $\boldsymbol{\eta}_{\theta}$ is a matrix of size $N \times M$ and $(\boldsymbol{\eta}^{i})_{\theta\theta}$ is a matrix of size $M \times M$. Once we have $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ ($N \times N$) matrix, and also the estimated parameters $\widehat{\boldsymbol{\theta}}$, given any arbitrary vector $\boldsymbol{f}$ ($N \times 1$), we can find $\boldsymbol{g} = \boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}})\boldsymbol{f}$ (Laurent & Cook 1992).

Now consider the case when the model is a summation of L non-linear functions, and that each function depends only on a subset of parameters (also called as partially separable), i.e.

$$\boldsymbol{m}(\boldsymbol{\theta}) = \sum_{i=1}^{L} \boldsymbol{h}_i(\boldsymbol{\theta}_i) \tag{A10}$$

with $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^{T}, \boldsymbol{\theta}_2^{T}, \ldots]^{T}$. Also assume that we are only interested in finding the diagonal values of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$. In this case, applying (A9) to (A10) yields

$$\boldsymbol{\eta}_{\theta} = \left[\boldsymbol{\eta}_1 \quad \boldsymbol{\eta}_1 \quad \ldots \quad \boldsymbol{\eta}_L \right], \tag{A11}$$

where $\boldsymbol{\eta}_i = \frac{\partial}{\partial \boldsymbol{\theta}_i^T} \boldsymbol{h}_i(\boldsymbol{\theta}_i)$ and

$$(\boldsymbol{\eta}_i^j)_{\theta\theta} = \begin{bmatrix} \mathbf{H}_1^j & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^j & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_L^j \end{bmatrix}, \tag{A12}$$

where

$$\mathbf{H}_i^j = \frac{\partial^2}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_i^T} \boldsymbol{h}_i^j(\boldsymbol{\theta}_i) \tag{A13}$$

with $\boldsymbol{h}_i^j(\boldsymbol{\theta}_i)$ being the jth element of $\boldsymbol{h}_i(\boldsymbol{\theta}_i)$. Substituting (A11) and (A12) to (A9) and only considering the block diagonal entries, we get

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}) = \sum_{j=1}^{L} \boldsymbol{\eta}_j \left(\boldsymbol{\eta}_j^T \boldsymbol{\eta}_j - \sum_{i=1}^{N} \boldsymbol{r}^i \mathbf{H}_j^i \right)^{-1} \boldsymbol{\eta}_i^T, \tag{A14}$$

which can be used to get the diagonal entries of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$.

A3 Calibration with excluded data

We consider the general case where a subset of data (baselines) are excluded during calibration. Consider $\mathcal{J}$ to be the set of indices of excluded data points in (A4). Assume the total ignored data points to be R, $0 \leq R < N$. Following Ross (1987), we modify (A4) as

$$\boldsymbol{d} = \sum_{i=1}^{K} \boldsymbol{s}_i(\boldsymbol{\theta}) + \mathbf{D}\boldsymbol{\gamma} + \boldsymbol{n}, \tag{A15}$$

where $\mathbf{D}$ ($N \times R$) is a matrix whose ith column has 1 at the $\mathcal{J}^i$-th location and the rest of the entries in the column are 0. We introduce an additional parameter vector $\boldsymbol{\gamma}$ ($R \times 1$) into the data model. Normally $\boldsymbol{\gamma}$ is called the cross-validatory residual. The effect of these slack variables is to nullify the constraints introduced by the data points indexed by the set $\mathcal{J}$. If $\boldsymbol{\theta}_r = [\boldsymbol{\theta}^T \boldsymbol{\gamma}^T]^T$ are the augmented parameters ($M + R$), calibration gives us

$$\widehat{\boldsymbol{\theta}}_r = \underset{\boldsymbol{\theta}, \boldsymbol{\gamma}}{\operatorname{argmin}} \left\| \mathbf{d} - \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) - \mathbf{D}\boldsymbol{\gamma} \right\|^2 \tag{A16}$$

even though we do not explicitly solve for $\boldsymbol{\gamma}$. Therefore, the calibration with excluded data (A16) estimates $M + R$ parameters using N constraints, while calibration with all data (A5) estimates M parameters using N constraints. In both cases, the useful set of parameters is still $\boldsymbol{\theta}$ of size M.

Now we apply (A14) for the data model in (A15), where we have $L = K + 2$, with K non-linear functions $\boldsymbol{s}_j(\boldsymbol{\theta}_j)$ (parameters $\boldsymbol{\theta}_j$), one linear function $\mathbf{D}\boldsymbol{\gamma}$ (parameters $\boldsymbol{\gamma}$) and noise $\boldsymbol{n}$.

(i) $\boldsymbol{s}_j(\boldsymbol{\theta}_j)$: the values for $\boldsymbol{\eta}_j$ and $\mathbf{H}_j^i$ for each j can be calculated using (A2), and since this is quadratic, both $\boldsymbol{\eta}_j$ and $\mathbf{H}_j^i$ are non-zero, but they are sparse.

(ii) $\mathbf{D}\boldsymbol{\gamma}$: since this is linear in $\boldsymbol{\gamma}$, $\boldsymbol{\eta}_j = \mathbf{D}$ and $\mathbf{H}_j^i = 0$.

(iii) $\boldsymbol{n}$: for noise, we do not have any parametrization, and therefore, we assume both $\boldsymbol{\eta}_j$ and $\mathbf{H}_j^i$ to be matrices with random entries. We notice the following for the computation of the leverage:

Considering the aforementioned three cases, we see that (i) and (iii) are always present, regardless of calibration using the full data set $R = 0$ or a subset of baselines ($R > 0$). In other words, (i) and (iii) contribute to (A14) in both cases. Moreover, the contribution (iii) is not dependent on $\boldsymbol{\theta}$ and therefore is uniform if the noise $\boldsymbol{n}$ is uniformly distributed. The interesting case is (ii), when $R > 0$. The contribution to (A14) can be written as

$$\boldsymbol{\Gamma}_d = \mathbf{D} \left(\mathbf{D}^T \mathbf{D} - \mathbf{0} \right)^{-1} \mathbf{D}^T = \mathbf{D}\mathbf{D}^T = \widetilde{\mathbf{I}}, \tag{A17}$$

where $\widetilde{\mathbf{I}}$ is a diagonal matrix with 1-s at the locations given by $\mathcal{J}$ and the rest of the entries 0. To sum up: if the ith diagonal entry of $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ calculated with the estimate $\widehat{\boldsymbol{\theta}}$ using the full data set is $\Gamma^{ii}(\widehat{\boldsymbol{\theta}})$, then this value changes to $\Gamma^{ii}(\widehat{\boldsymbol{\theta}}) + 1$ for the case where the ith data point is excluded during calibration. The excluded baselines have an increase in leverage by 1. Therefore, the error in the residuals is enhanced on the excluded baselines. The only way to minimize this error is to minimize the variance of estimated parameters, $\widehat{\boldsymbol{\theta}}$, or in other words, find the global minimum point in the parameter space.

This paper has been typeset from a TEX/LATEX file prepared by the author.