

Research Paper

The impact of tandem redundant/sky-based calibration in MWA Phase II data analysis

Zheng Zhang¹ , Jonathan C. Pober¹, Wenyang Li¹, Bryna J. Hazelton^{2,3}, Miguel F. Morales², Cathryn M. Trott^{4,5}, Christopher H. Jordan^{4,5} , Ronniy C. Joseph^{4,5}, Adam Beardsley⁶, Nichole Barry^{5,7}, Ruby Byrne², Steven J. Tingay^{4,5} , Aman Chokshi^{5,7}, Kenji Hasegawa⁸, Daniel C. Jacobs⁶, Adam Lanman¹, Jack L. B. Line^{4,5} , Christene Lynch^{4,5} , Benjamin McKinley^{4,5}, Daniel A. Mitchell^{4,9}, Steven Murray^{4,5}, Bart Pindor^{5,7}, Mahsa Rahimi^{5,7}, Keitaro Takahashi^{10,11}, Randall B. Wayth^{4,5}, Rachel L. Webster^{5,7}, Michael Wilensky², Shintaro Yoshiura⁷ and Qian Zheng¹²

¹Department of Physics, Brown University, Providence, RI 02912, USA, ²Department of Physics, University of Washington, Seattle, WA 98195, USA, ³eScience Institute, University of Washington, Seattle, WA 98195, USA, ⁴International Centre for Radio Astronomy Research, Curtin University, Perth, WA 6845, Australia, ⁵ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO-3D), ⁶School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA, ⁷School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia, ⁸Graduate School of Science, Nagoya University, Nagoya, Japan, ⁹CSIRO Astronomy and Space Science (CASS), PO Box 76, Epping, NSW 1710, Australia, ¹⁰Faculty of Science, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan, ¹¹International Research Organization for Advanced Science and Technology, Kumamoto University, Kumamoto 860-8555, Japan and ¹²Shanghai Astronomical Observatory, Shanghai, China

Abstract

Precise instrumental calibration is of crucial importance to 21-cm cosmology experiments. The Murchison Widefield Array's (MWA) Phase II compact configuration offers us opportunities for both redundant calibration and sky-based calibration algorithms; using the two in tandem is a potential approach to mitigate calibration errors caused by inaccurate sky models. The MWA Epoch of Reionization (EoR) experiment targets three patches of the sky (dubbed EoR0, EoR1, and EoR2) with deep observations. Previous work in Li et al. (2018) and (2019) studied the effect of tandem calibration on the EoR0 field and found that it yielded no significant improvement in the power spectrum (PS) over sky-based calibration alone. In this work, we apply similar techniques to the EoR1 field and find a distinct result: the improvements in the PS from tandem calibration are significant. To understand this result, we analyse both the calibration solutions themselves and the effects on the PS over three nights of EoR1 observations. We conclude that the presence of the bright radio galaxy Fornax A in EoR1 degrades the performance of sky-based calibration, which in turn enables redundant calibration to have a larger impact. These results suggest that redundant calibration can indeed mitigate some level of model incompleteness error.

Keywords: instrumentation: interferometers – methods: data analysis – cosmology: observations – dark ages – reionisation – first stars

(Received 4 July 2020; revised 15 September 2020; accepted 18 September 2020)

1. Introduction

The highly redshifted 21-cm signal of neutral hydrogen gas encodes the perturbation statistics of hyperfine states of neutral hydrogen gas, which trace the conditions of the radiation fields permeating the early universe. Measurements of this signal are therefore a promising path for further constraints on the Epoch of Reionization (EoR) and dark energy. Several experiments aiming to detect the 21-cm EoR signal are either complete or under way, such as the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER; Parsons et al. 2010), the Low-Frequency Array (LOFAR; van Haarlem et al. 2013), the

Murchison Widefield Array (MWA; Tingay et al. 2013a; Bowman et al. 2013), the Giant Metrewave Radio Telescope (GMRT; Gupta et al. 2017), and the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017).

However, the astrophysical foregrounds are four to five orders of magnitude brighter than the faint cosmological signal. But because the foregrounds are spectrally smooth, while the 21-cm signal has significant spectral structure, they can be in principle distinguished after performing a Fourier transform in frequency. The key challenges are precise calibration of the complicated frequency dependence of the instruments, accurate avoidance of radio frequency interference (RFI), and the effects of the ionospheric activity. Significant work in the past decade has led to a number of techniques aimed at overcoming these difficulties: data quality metrics to flag channels contaminated by RFI (Wilensky et al. 2019; Li et al. 2019; Kerrigan et al. 2019; Offringa et al. 2015), metrics to assess ionospheric activity (Jordan et al. 2017; Trott et al. 2018), and a variety of instrumental calibration methods (e.g. Liu et al. 2010; Barry et al. 2019a; Yatawatta 2015).

Author for correspondence: Zheng Zhang, E-mail: zheng_zhang1@alumni.brown.edu
Cite this article: Zhang Z, Pober JC, Li W, Hazelton BJ, Morales MF, Trott CM, Jordan CH, Joseph RC, Beardsley A, Barry N, Byrne R, Tingay SJ, Chokshi A, Hasegawa K, Jacobs DC, Lanman A, Line JLB, Lynch C, McKinley B, Mitchell DA, Murray S, Pindor B, Rahimi M, Takahashi K, Wayth RB, Webster RL, Wilensky M, Yoshiura S and Zheng Q. (2020) The impact of tandem redundant/sky-based calibration in MWA Phase II data analysis. *Publications of the Astronomical Society of Australia* 37, e045, 1–16. <https://doi.org/10.1017/pasa.2020.37>

A leading paradigm of 21-cm analysis is to measure the power spectrum (PS) of the signal in two-dimensional (2D) cylindrical Fourier space (i.e. the $(k_{\perp}, k_{\parallel})$ plane, with Fourier modes perpendicular and parallel to the plane of the sky, respectively^a). In the 2D $(k_{\perp}, k_{\parallel})$ plane, foregrounds are confined to a wedge-shaped region, called the ‘foreground wedge’. The remaining region, namely the ‘EoR window’, is expected to be contaminant-free (Pober *et al.* 2014; Datta, Bowman, & Carilli 2010; Parsons *et al.* 2012; Morales *et al.* 2012; Vedantham, Shankar, & Subrahmanyan 2012; Thyagarajan *et al.* 2013; Trott, Wayth, & Tingay 2012; Liu, Parsons, & Trott 2014).

For perfectly calibrated instruments, signal in the EoR window should only come from 21-cm signal emission in an uncontaminated observation. But in practice, small chromatic instrumental calibration errors introduce spectral structure in astrophysical foregrounds, which contaminates some Fourier modes in the EoR window and thus overwhelms the cosmological PS (Morales *et al.* 2019; Barry *et al.* 2016).

The two most popular methods of interferometric gain calibration are ‘sky-based’ and ‘redundant’, both are subject to systematic errors. Sky-based calibration uses deep foreground catalogues or maps to create model visibilities for calibration, which requires an accurate sky model; errors are introduced if the model does not capture the true sky brightness distribution. In particular, Barry *et al.* (2016) identified a specific class of calibration errors where the instrument’s point spread function interacts with missing sources in the sky model to introduce chromatic errors and couple foreground power into the EoR window. Redundant calibration takes advantage of the fact that any given correlation should depend only on relative positions of two antennas, which reduces a large number of unknowns when calibrating an redundant array. Redundant calibration is, to first order, sky-model-independent but may be contaminated by non-redundancy introduced through beam variation and antenna position errors (Orosz *et al.* 2019; Joseph *et al.* 2019). Joseph, Trott, & Wayth (2018) found that the position offsets introduce a phase bias of the calibration solutions. This phase bias increases with the distance between bright radio sources and the pointing centre, and with the flux density of these sources (Joseph *et al.* 2018). Byrne *et al.* (2019) also showed that calibration errors caused by sky model incompleteness affect redundant calibration, since redundant calibration leaves degenerate parameters that must be constrained via reference to a sky model.

Although the principles and limitations of each calibration method are reasonably well understood, the actual performance of these techniques will be affected by realistic constraints, such as algorithm implementations, array configuration, sky model accuracy, and data contamination from interference. Recent analysis with the HERA telescope shows the potential for redundant calibration to provide some degree of information beyond sky calibration, as well as its value as a data quality metric (Kern *et al.* 2020; Dillon *et al.* 2020). Li *et al.* (2018) used the compact array of Phase II of the MWA (Wayth *et al.* 2018; Beardsley *et al.* 2019), which has both large numbers of redundant baselines and a relatively complete sampling of Fourier space (i.e. UV coverage), to compare redundant calibration with sky-based algorithms in the EoR0 field (see Section 2 for a description of the fields observed

by the MWA EoR project). That work showed that the tandem combination of sky-based calibration and redundant calibration yielded small but non-negligible reductions in foreground contamination in certain modes of the EoR PS. However, Li *et al.* (2019) performed a similar analysis after adding an improved auto-correlation-based technique to the sky-based calibration step and found that the improvements brought about by redundant calibration became negligible. In this paper, we further investigate the performance of tandem redundant plus sky-based calibration in MWA Phase II data analysis in the EoR1 field, where we find a different result than that of Li *et al.* (2019): redundant calibration has a significant positive impact on foreground contamination in the PS even with the auto-correlation-based technique applied to sky-based calibration (as will be shown in Section 4.3). Before presenting the PS results, we compare the calibration solutions derived from both sky and redundant calibration on a 2-min cadence from three nights of EoR1 and one night of EoR0 to investigate differences in the solutions themselves. As in Li *et al.* (2019), we use packages Fast Holographic Deconvolution (FHD) and OMNICAL for state-of-the-art implementations of sky-based and redundant calibration, respectively.

The organisation of this paper is as follows. In Section 2, we describe the compact array of MWA Phase II and the observations analysed in this work. In Section 3, we review the methodology and mathematical framework of sky-based calibration and redundant calibration. In Section 4, we present all our analyses and results, including simulations exploring redundant calibration performance (Section 4.1), inspection of the calibration solutions themselves (Section 4.2), and PS analyses (Section 4.3). We further discuss over these analyses and present our conclusions in Section 5.

2. Observations

The MWA is a low-frequency (80–300 MHz) radio interferometer located at the Murchison Radio-astronomy Observatory in Western Australia, the location of the future low-frequency Square Kilometre Array. In MWA Phase I (2013–2016; Tingay *et al.* 2013b; Bowman *et al.* 2013), 128 antenna tiles^b were arranged in a pseudo-random layout designed for excellent instantaneous UV coverage. The Phase II upgrade (Wayth *et al.* 2018; Beardsley *et al.* 2019) added 128 additional tiles in mid-2016. The correlator is still limited to processing only 128 tiles at a time, so the array is periodically reconfigured between ‘compact’ and ‘extended’ configurations. The MWA Phase II compact array (Figure 1) consists of 72 tiles in a regular hexagonal grid and 56 pseudo-random tiles in a dense core with minimal redundancy. This configuration allows us to perform redundant calibration while retaining good UV coverage.

The individual dipoles in each tile can be time-delayed relative to each other in the beamformer, enabling the array to be ‘pointed’ to a certain patch of the sky. However, the time delays can only be adjusted by discrete amounts (as they are added by inserting additional physical lengths of cable into the signal chain), meaning that the telescope cannot continuously track a patch of sky. Instead, the MWA EoR program observes in a specific direction with the sky drifting overhead for about 30 min and then repoints the tiles to track the field of interest, which is called the ‘drift and shift’ method (Trott 2014). During an observation, the delays are

^aStrictly speaking, the sky is curved and Fourier modes do not provide an orthonormal basis for describing it. However, as long as the fields of view are not too large, the distinction is negligible and is usually not accounted for in the literature; see Liu, Zhang, & Parsons (2016) for a rigorous treatment.

^bEach tile consists of 16 individual dipoles with signals combined in an analog beamformer.

shifted multiple times until the field is too low in the sky to track, so the data in the observation are grouped into ‘pointings’ (each set of tile delays corresponds to one pointing). With the sky transiting from the east, the program begins observing at 5 pointings before zenith and ends observing at 4 pointings after zenith. The zenith is labelled as ‘pointing 0’, thus all the 10 pointings are labelled from pointing -5 to pointing 4 .

The MWA EoR program focuses on three fields with relatively low foreground emission: ‘EoR0’ (RA = $0^{\text{h}}00$, Dec = -27°), ‘EoR1’ (RA = $4^{\text{h}}00$, Dec = -27°), and ‘EoR2’ (RA = $11^{\text{h}}33$, Dec = -10°); see Jacobs et al. (2016) for the motivation for selecting these fields and Trott et al. (2020) for a detailed analysis of the features present in each. Each field fills the main lobe of the MWA primary beam, spanning roughly $\sim 40^{\circ}$ on a side.

In this work, we analyze high band (167–197 MHz) data from three nights observing EoR1 and one night observing EoR0. The most salient difference between these two fields is the presence of Fornax A, a bright, extended source in the EoR1 field; EoR0, in comparison, contains mostly unresolved sources. In this work, we denote the nights as ‘EoR1 night1’ (2016 October 15), ‘EoR1 night2’ (2016 October 17), ‘EoR1 night3’ (2016 November 15), and the ‘EoR0 night’ (2016 November 21), respectively. We explore the four nights using good pointings determined by the window power metric (Beardsley et al. 2016; Li et al. 2019) and the ionospheric metric (Jordan et al. 2017; Trott et al. 2018): for EoR1 night1 and night2, good pointings are $-3, -2, -1, \text{ and } 0$; for EoR1 night3, good pointings are $-3, -2, \text{ and } -1$; for the EoR0 night, good pointings are $-2, -1, 0, 1, \text{ and } 2$.

The MWA uses the AOflogger algorithm^c to flag data contaminated by RFI (Offringa et al. 2015). In this work, we further flag faint RFI not captured by the pipeline using data quality metrics developed by Wilensky et al. (2019) and Li et al. (2019).

3. Calibration formalism

The sky signal is modified along the path from where it was emitted to where it was recorded. In order to measure the cosmological signal, all these modifications must be corrected for. While our RFI flagging and data quality metrics are used to avoid the most corrupted sky signals, the electronic response of the antennas must still be corrected for with a gain calibration algorithm.

In this section, we will briefly introduce two calibration methods used in our analysis: sky-based calibration and redundant calibration, and the code packages and algorithms we use to implement them.

3.1. Measurements of the sky

Suppose antenna i with gain g_i measures a frequency domain signal s_i at a given instant, an approximation of a short time block. s_i is related to the true sky signal h_i and the antenna’s instrumental noise contribution n_i as:

$$s_i = g_i h_i + n_i, \tag{1}$$

which is both an equation of time and frequency.

In this paper, we consider only a single linear polarisation at a time. We assume per-antenna and per-frequency gains. In the MWA, the instrumental gains are assumed to be constant over 2-min observations. Then in a long integration time that $\langle n_i \rangle = 0$, the baseline b_{ij} measures the correlation between signals from antenna i and j ($i \neq j$):

$$v_{ij} = \langle s_i^* s_j \rangle \tag{2a}$$

$$\approx g_i^* g_j \langle h_i^* h_j \rangle + \langle n_i^* n_j \rangle + g_j \langle n_i^* h_j \rangle + g_i^* \langle h_i^* n_j \rangle \tag{2b}$$

$$= g_i^* g_j u_{ij} + n_{ij}, \tag{2c}$$

where $u_{ij} = \langle h_i^* h_j \rangle$ denotes the true visibility and n_{ij} is the residual noise as the synthesis of all noise contributions in Equation (2b). We have applied two assumptions: the noise of different antennas are uncorrelated and the sky signal does not correlate with the noise. Under these assumptions, the last three terms in Equation (2b) all have an expectation value of 0 in the limit that time goes to infinity. However, since our integration length is finite, these three terms contribute an approximately u_i -independent Gaussian random component, which we relabel as a single effective noise term n_{ij} in going to Equation (2c).

Our goal is to solve for the true sky visibilities u_{ij} . Numerous algorithms and strategies have been developed for precise calibration. Most of them can be classified into two approaches: sky-based calibration and redundant calibration. We will briefly introduce both of the two calibration methods in Section 3.2.

3.2. Calibration techniques

In our analysis, we used two calibration packages: one is the FHD^d (Barry et al. 2019a; Sullivan et al. 2012) software pipeline, which is used for sky-based calibration, and the other is OMNICAL^e (Zheng et al. 2014), performing redundant calibration.^f

They are two different methods but both work on maximum-likelihood estimate of per-frequency antenna gains. For any baseline b_{ij} , we assume its noise is Gaussian and mean zero with variance σ_{ij}^2 (Thompson, Moran, & Swenson 2017). Maximising the likelihood function of gains is equivalent to minimising

$$\chi^2 = \sum_{ij} \frac{|v_{ij} - g_i^* g_j u_{ij}|^2}{\sigma_{ij}^2}. \tag{3}$$

Instrumental frequency dependence is implicit in this equation.

3.2.1. FHD: Sky-based calibration

Sky-based calibration works by substituting model visibilities for the true visibilities in Equation (3). The accuracy of the model visibilities determines the accuracy of sky-based calibration. We use GaLactic and Extragalactic All-sky MWA (GLEAM) (Hurley-Walker et al. 2016), an extragalactic point source catalogue with large coverage and high completeness, as our model sky to generate model visibilities. FHD simulates all reliable sources in the catalogue out to typically 1% beam level in the primary lobe and the sidelobes to build a theoretical sky (Barry et al. 2019a; Beardsley et al. 2016). In this work, we exclude baselines shorter than 50 wavelengths (which are sensitive to diffuse Galactic emission absent from GLEAM) to reduce the contamination from differences between the model visibilities and the true visibilities (Barry et al. 2019a).

^d<https://github.com/EoRImaging/FHD>.

^e<https://github.com/jeffzhen/omnical>.

^fSome confusing nomenclature unfortunately exists around this code. The package OMNICAL contains three algorithms: Logcal, Linca1, and Omnical. The Omnical algorithm, first described in Dillon et al. (2020), is a more efficient replacement for Linca1; however, following Li et al. (2018) and (2019), we use only Logcal and Linca1, which are described below. In referring to OMNICAL (all capitalised), we refer to the code package and not the Omnical algorithm, which is not used in this work.

^c<http://aoflogger.sourceforge.net/doc/api/>.

FHD calibration begins with a set of estimated initial solutions for $g_i^*(f)$ and the sky visibilities are substituted with model visibilities m_{ij} . (In practice, the digital gains in the MWA correlator pre-scale the visibilities to roughly the right level—so an initial gain estimate of 1.0 at for all antennas at all frequencies is sufficient for quick convergence of the calibration algorithm.) Then minimising Equation (3) becomes a linear least-squares problem:

$$\chi_{j,\text{sky}}^2(f) = \sum_i \frac{|v_{ij}(f) - g_i^*(f)g_j(f)m_{ij}(f)|^2}{\sigma_{ij}^2(f)}, \quad (4)$$

where m_{ij} is the model visibility. This relation can be solved iteratively to return the values of the g terms that minimise χ^2 (Mitchell et al. 2008; Salvini & Wijnholds 2014).

Sky-based calibration requires an accurate sky model, otherwise the difference between the true sky visibilities and model visibilities will be incorporated into our evaluations of instrumental gains. Barry et al. (2016) demonstrate how gain errors due to sky model incompleteness introduce contamination that prevents recovery of an EoR signal.

3.2.2. OMNICAL: Redundant calibration

If an array is built to have sufficient redundant baselines, the total number of measurements will exceed the sampled points in Fourier space. In this case, our equations are over-determined and we are able to directly solve for both instrumental gains and visibilities (Liu et al. 2010). Pedagogically, we can understand this calibration method as minimising

$$\chi_{\text{red}}^2(f) = \sum_{\alpha} \sum_{\{i,j\}_{\alpha}} \frac{|v_{ij}(f) - g_i^*(f)g_j(f)u_{\alpha}(f)|^2}{\sigma_{ij}^2(f)}, \quad (5)$$

where α indexes types of redundant baselines and $\{i, j\}_{\alpha}$ are sets of antennas that belong to each type of redundant baseline. $\{u_{\alpha}\}$ and $\{g_i\}$ can both be solved for directly.

In this work, redundant calibration incorporates two parts, `Logcal` and `Linca1`. `Logcal` uses logarithms to cast the minimisation of Equation (5) as a linear regression problem. But this method is biased for two reasons: 1. when the signal-to-noise ratio (SNR) is low, it is less accurate to expand the logarithm noise contribution term $\ln(1 + n_{ij}/g_i^*g_ju_{ij})$ as $n_{ij}/g_i^*g_ju_{ij}$, as is done by `Logcal` and 2. `Logcal` applies least-squares estimation to the distributions of noises' magnitudes and phases, which is inaccurate as the magnitudes and phases of noise follow Rayleigh distributions rather than Gaussian distributions.

`Linca1` was developed to mitigate these biases (Liu et al. 2010). In `Linca1`, we Taylor expand Equation (2c):

$$v_{ij} \approx g_i^{0*}g_j^0u_{ij}^0 + g_i^{0*}u_{ij}^0\Delta g_j + g_j^0u_{ij}^0\Delta g_i^* + g_i^{0*}g_j^0\Delta u_{ij}, \quad (6)$$

where Δg_j , Δg_i^* , and Δu_{ij} are equal to $g_j - g_j^0$, $g_i^* - g_i^{0*}$, and $u_{ij} - u_{ij}^0$, respectively. Thus, the measurement equation becomes linear so that we can perform a least-squares fitting: using initial guesses of g_i^0 and u_{ij}^0 , we can solve for all values of $(g_j - g_j^0)$ and $(u_{ij} - u_{ij}^0)$. Then, we use the new estimations for g_i^0 and u_{ij}^0 as inputs and solve the equation iteratively. Note that the initial inputs should be in the minimum containing the true solution. In this work, outputs of `Logcal` are used as the starting point for `Linca1`, so that biases are expected to be mitigated.

Redundant calibration faces another problem, in that there are four intrinsic degeneracy parameters per frequency: the overall amplitude, overall phase, and two phase gradient components (Byrne et al. 2019; Dillon et al. 2018; Liu et al. 2010). These

parameters can take any value without affecting the redundancy of the data; therefore, they cannot be constrained by redundancy-based methods. We constrain these degenerate parameters using the FHD calibration solutions: finding values of the four degeneracy parameters per frequency per polarisation for the whole array, which best fit the sky-based calibration solutions (Li et al. 2018). Li et al. (2018) explored different ways of combining FHD and OMNICAL, each of which used slightly different methods for constraining the degeneracy parameters. We use the method referred to as 'FOCal' (also used in Li et al. 2019), which we now detail in Section 3.2.3.

3.2.3. FHD and OMNICAL in tandem: Redundant plus sky-based calibration

In this paper, we perform redundant plus sky-based calibration using FHD and OMNICAL in tandem. Following Li et al. (2018), we perform FHD sky-based calibration on the raw data first. Then, we perform redundant calibration on FHD-calibrated data, yielding a second set of gains that can be multiplied by the FHD-produced gains to yield a final set of calibration solutions. Before multiplying the OMNICAL and FHD gains together, however, we set the parameters in the redundant calibration degenerate space of the OMNICAL gains to 1 (for the amplitude) and 0 (for the phase terms) so that the FHD values in this subspace are retained after multiplication. This is the process referred to as 'degeneracy projection' in Li et al. (2018). With this approach, we can investigate the performance of each calibration method by looking at both the solutions produced by FHD and the changes to those solutions produced by OMNICAL.

One can expect in our calibration pipeline that based on FHD calibrations, redundant calibration makes further modifications to fit the instrumental redundancy. These modifications on top of FHD calibration results will be small if sky-based calibration works well; based on this expectation, we refer to the gain solutions produced by OMNICAL as multiplicative ' Δ 's' on top of the FHD solutions (i.e. $g_{\text{tandem}} = \Delta_{\text{OMNICAL}} \times g_{\text{FHD}}$). In Section 4, we will look at FHD gain solutions, the redundant calibration gain Δ 's, and the difference in the PS when using only FHD versus FHD plus OMNICAL tandem calibration.

4. Analyses

Our approach for using FHD and OMNICAL in tandem enables us to study the performance of each algorithm relatively straightforwardly. Assuming perfect sky-based calibration with a perfect sky model, the Δ 's produced by redundant calibration should be 1.0 (since the gains are multiplicative). Any deviation from 1 indicates an error—either one in sky-based calibration that is being corrected by redundant calibration or one in redundant calibration due to, for example, non-redundancy in baselines that were assumed to be redundant. Theoretically, the errors in our sky models should vary from field to field on the sky, so that sky-based calibration will behave differently for different observations. However, it is hard to unambiguously evaluate 'effectiveness' in real data, when non-unity values for the redundant calibration Δ 's can result from errors in either calibration algorithm. In this work, we calibrate and process MWA Phase II data from both the EoR1 field (which has not been studied in this fashion before) and the EoR0 field (which was studied in Li et al. 2018 and 2019). By comparing both the calibration solutions from each algorithm and the resultant PS measurements, we can begin to understand how the behaviour of these algorithms varies from field to field.

In this section, we will present three analyses: a simulation exploring redundant calibration performance on noisy but otherwise perfect data (Section 4.1); a study of the calibration solutions from both FHD and OMNICAL and their repeatability from night to night (Section 4.2); and an investigation of the effects applying the redundant calibration solutions have on the PS of our data (Section 4.3).

4.1. Simulation: Redundant calibration performance on noise

Since our measured data have a random noise component, we can never truly expect redundant calibration to return Δ 's of exactly 1.0, even if sky-based calibration is perfect. Here, we present a simple simulation where OMNICAL is run on 'perfect' data with a SNR comparable to the real observations. If we find the redundant calibration Δ 's for the real data are at a level of order the Δ 's we find in this simulation, we might conclude that redundant calibration is purely fitting noise, for example, any changes it wants to make to the FHD calibration are purely random.

To keep the simulation simple, we create a single mock observation with perfect redundancy and no fringes, for example, the signal is real-valued. For the fairness of comparison, the simulated observation is designed to have the same noise level and comparable SNRs with the real observation at each frequency. We characterise the noise level via the standard deviation of the noise and SNR is defined as the ratio between signal amplitude and the noise level. The frequency-dependent noise level has been assumed to be constant over all baselines and over an observation. Thus, we can produce a simulated observation such that the signal $\mu(f)$ and the standard deviation of the noise $\sigma(f)$ are the same over all baselines and times per frequency. Mathematically, the simulated visibility V is

$$V = \mu + n = \text{SNR}(f) \cdot \sigma(f) + N_R(0, \sigma^2(f)) + jN_I(0, \sigma^2(f)), \tag{7}$$

where n denotes the complex noise contribution of the simulated visibility, with standard deviation $\sigma(f)$ in the real and imaginary, and $\mu = \text{SNR}(f) \cdot \sigma(f)$. Both the real and imaginary noise components are randomly sampled as mean zero Gaussian $N(0, \sigma^2(f))$.

We directly calculate the per frequency noise levels of the real observation and use them as $\sigma(f)$ in the simulated observation. The $\text{SNR}(f)$ is also empirically calculated from the real observation. The noise level is calculated per frequency for the entire observation using the even/odd differencing method described in Barry et al. (2019a): visibilities separated by 2 s in time are subtracted from each other to remove the slowly changing sky signal. The residuals are then assumed to be noise-dominated, with a per-frequency standard deviation (taken across time and baseline) that reflects the underlying noise in the data.

The 'signal' level of the real data is calculated from the distribution of the visibility amplitudes for a single observation. For a set of visibilities with constant sky signal magnitude μ and noise level σ , the distribution of visibility amplitudes Z is uniquely determined, even though the visibilities themselves rotate through the complex plane as a function of time. The analytic expression for the mean value of visibility amplitudes can be shown to be:

$$\begin{aligned} \langle Z \rangle &= \int_0^\infty Z \cdot P(Z) dZ = \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma^2} \int_0^\infty Z^2 e^{-\frac{Z^2}{2\sigma^2}} \mathbf{I}\left[0, \frac{Z\mu}{\sigma^2}\right] dZ \\ &= \sqrt{\frac{\pi}{8}} \sigma^{-1} e^{-\frac{\mu^2}{4\sigma^2}} \left(2\sigma^2 + \mu^2 \right) \mathbf{I}\left[0, \frac{\mu^2}{4\sigma^2}\right] + \mu^2 \mathbf{I}\left[1, \frac{\mu^2}{4\sigma^2}\right], \tag{8} \end{aligned}$$

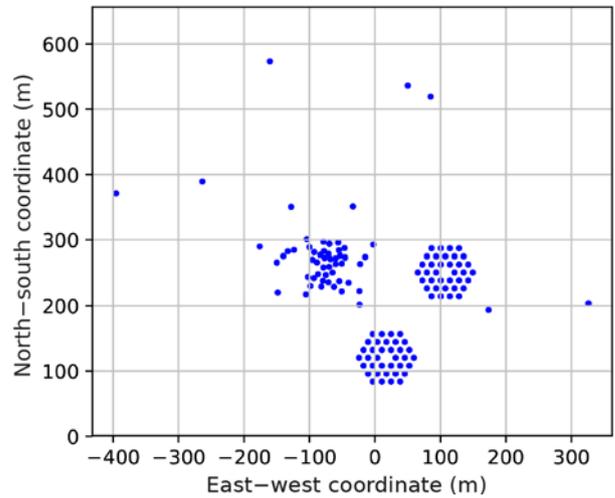


Figure 1. The MWA Phase II compact array layout.

where $\mathbf{I}[\nu, z]$ denotes modified Bessel function of the first kind:

$$\mathbf{I}[\nu, z] = \sum_{k=0}^{\infty} \frac{1}{\Gamma(k + \nu + 1)k!} \left(\frac{z}{2}\right)^{2k + \nu}. \tag{9}$$

This relation lets us calculate a value of μ to use in our simple visibility model (Equation (7)) and $\langle Z \rangle$ of the real observation. For the real observation, we used in comparison, $\langle Z \rangle \approx 2.707\sigma$, giving us $\mu \approx 2.495\sigma$.

We calculated the mean SNR of the real observation over all frequencies and then produce a series $\{\text{SNR}(f)\}$ whose mean value is equal to the real observation's such that $\text{SNR}(f)$ is proportional to frequency to the -1.59 , which is derived from fitting a power law distribution to the empirically measured frequency dependence of SNR. The simulated observation with known $\text{SNR}(f)$ is directly input into OMNICAL, so that the Δ values should be mean 1 with any deviations purely caused by noise.

We compare Δ 's of a real EoR1 observation (Observation ID: 1163254256) with redundant calibration solutions of the corresponding simulated observation in Figure 2, which shows root mean square (RMS) of redundant calibration solution magnitudes/phases. The RMS is taken across the set of antennas. Clearly, redundant calibration is making changes to the real data at a level significantly larger than might be expected were the data perfectly redundant up to the noise level.

We also investigate the effect of SNR on the result of the simulations. We find that as long as the SNR is held fixed (e.g. increases to the signal strength are offset by increases to the noise), the Δ 's produced by redundant calibration are at the same level. Increasing the SNR does reduce the magnitude of the Δ 's, indicating that redundant calibration is indeed affected by noisy data. However, as indicated in Figure 2, the RMS of the Δ 's for the real data are larger than would be expected from SNR alone.

From these results, we can conclude that OMNICAL is responding to real non-redundancy in the data that is above the noise level. Our simulation tells us the expected scale of the Δ 's for a given SNR and perfect redundancy; in the real data, we see Δ 's a factor of $\sim 3 - 5$ larger in RMS. This analysis does not, however, tell us the source of the non-redundancy. It might be that the FHD solutions are not completely correct and that OMNICAL is bringing the gains closer to their true values. However, it might be that the instrument itself is non-redundant and OMNICAL is introducing further errors into the gains (since it assumes perfect redundancy).

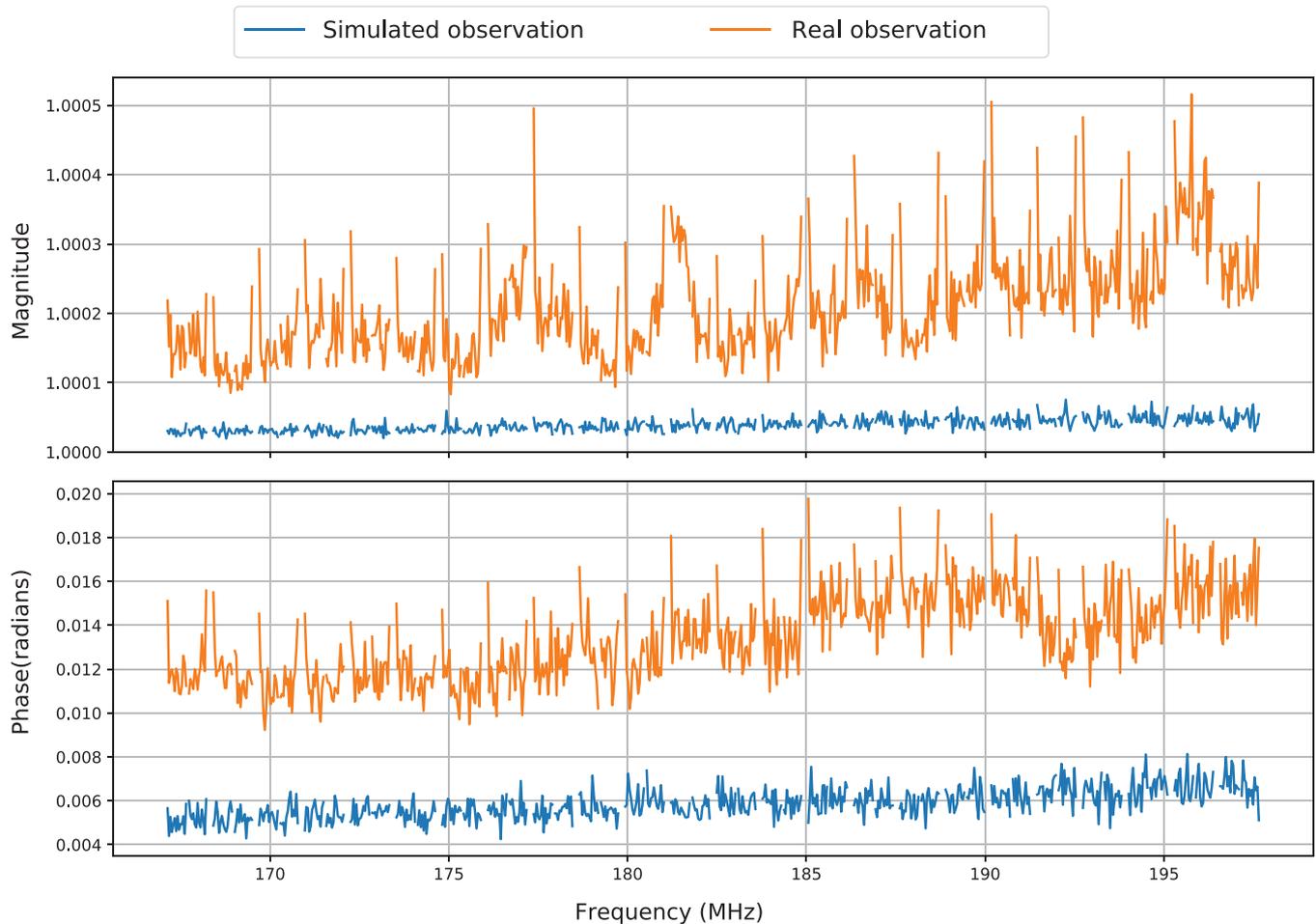


Figure 2. The RMS of redundant calibration solution magnitudes/phases of a real/simulated observation. Orange lines represent the real EoR1 observation and blue lines correspond to the simulated observation, which is constructed to have the same mean noise as the real observation per frequency and the same total SNR averaged over frequencies. In addition, we set the SNR of the simulated observation proportional to frequency to -1.59 which is similar to the real one. Upper: RMS of redundant calibration solution magnitudes; bottom: RMS of redundant calibration solution phases. We can see that redundant calibration makes changes to the real data beyond just noise.

We now continue our analysis of the gain solutions produced by FHD and OMNICAL to see whether the addition of OMNICAL does in fact lead to improvements in our calibration and subsequent analysis.

4.2. Comparison of calibration solutions

To study the performance of FHD and OMNICAL when applied to actual data, we compare the calibration solutions among the four nights. We perform two distinct analyses: a detailed investigation of the spectral behaviour of four sets (one per pointing that passes the cuts described in Section 2) of calibration solutions from 2-min Local Sidereal Time (LST)-matched data sets (Section 4.2.1) and a study of the repeatability of the calibration solutions across all observations in our data set (Section 4.2.2).

4.2.1. Spectral behaviour of the calibration solutions

Within each pointing, we select one observation per EoR1 night such that the sampled EoR1 observations are at well-matched LSTs. Similar LSTs will result in similar model visibilities generated by FHD. Thus, comparing LST-matched EoR1 observations across nights will test the consistency and repeatability of the calibration techniques.

We also select one observation per pointing from the EoR0 night and compare Δ 's with sampled EoR1 observations to explore redundant calibration performance at different EoR fields and a possible sky model dependence. There is obviously no LST accordance between the EoR0 night and EoR1 nights, but the observations chosen are representative.

We begin by first looking at the FHD sky model calibration solutions for each of our three selected EoR1 observations. We normalise the magnitudes of per-frequency FHD solutions for sampled EoR1 observations to mean 1 and then average over antennas to study the typical spectral structure. Figure 3 shows the magnitude of FHD solutions for sampled observations on the EoR1 nights. For each set of observations with similar LSTs, the solutions are largely in agreement with only small discrepancies over nights. In general, EoR1 nights 1 and 2 agree closely, with night 3 showing a steeper spectral slope.

Comparison of redundant calibration solutions are carried out in a similar way. We look at the spectra of the RMS of the magnitudes of the Δ 's over antennas for each sampled observation. Figure 4 illustrates that $|\Delta|$'s are similarly consistent over all the EoR1 nights with repeatable frequency structure. The grey curve shows the results from an EoR0 observation, which has different

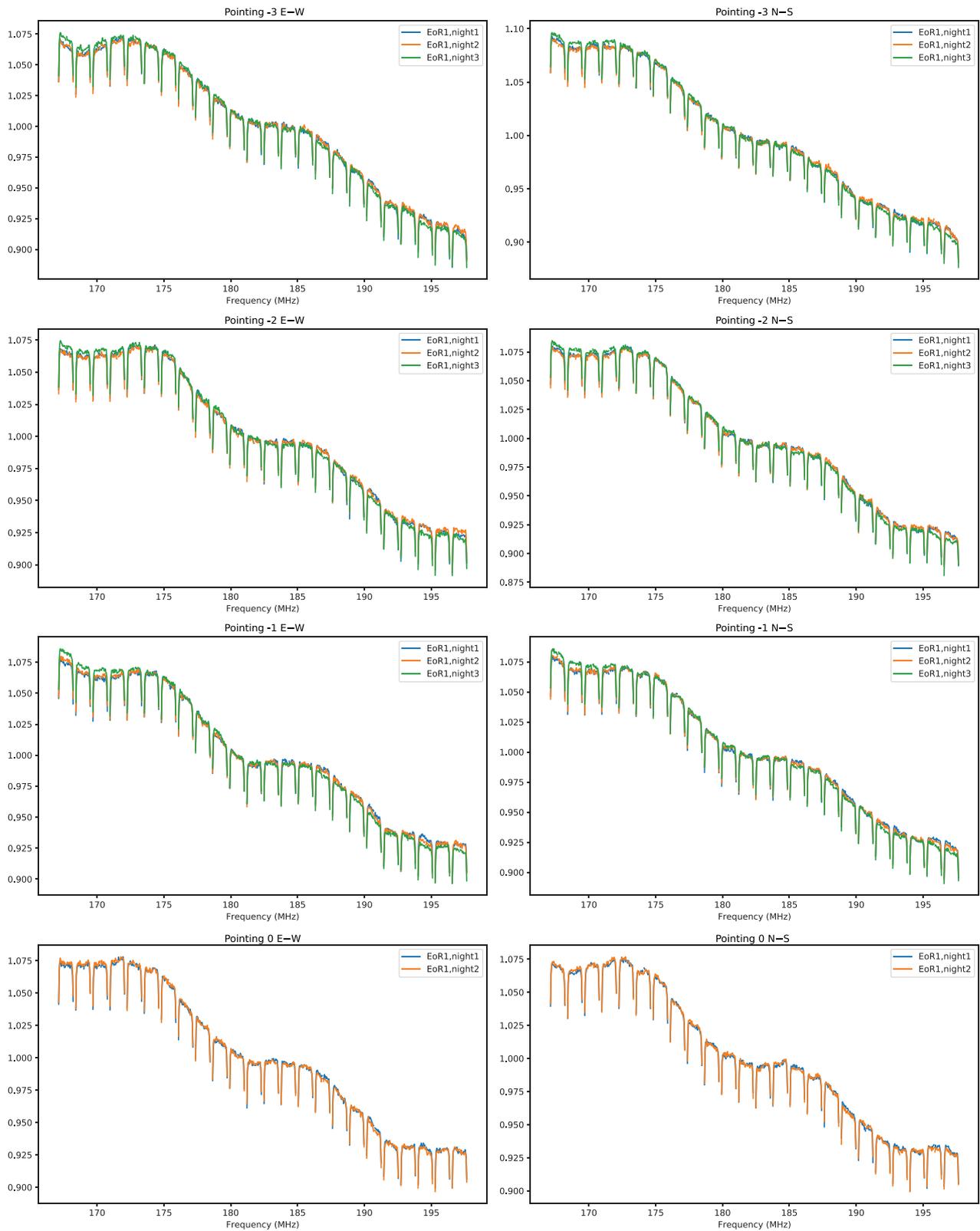


Figure 3. The spectra of FHD solution magnitudes (averaged over antennas and normalised to mean one) of observations with well-matched LSTs over our three EoR1 nights. Blue, orange, and green represent EoR1 night1, night2, and night3, respectively. Two columns correspond to two polarisations and four rows correspond to different pointings. Note that EoR1 night3 pointing 0 is flagged due to ionospheric contamination. The periodic 1.28 MHz scalloping is due to the two-stage Fourier transform used by the MWA correlator. For each 2-min LST-matched data set, the solutions are largely in agreement with only small discrepancies over nights. In general, EoR1 night1 and night2 agree closely, with night3 showing a steeper spectral slope.

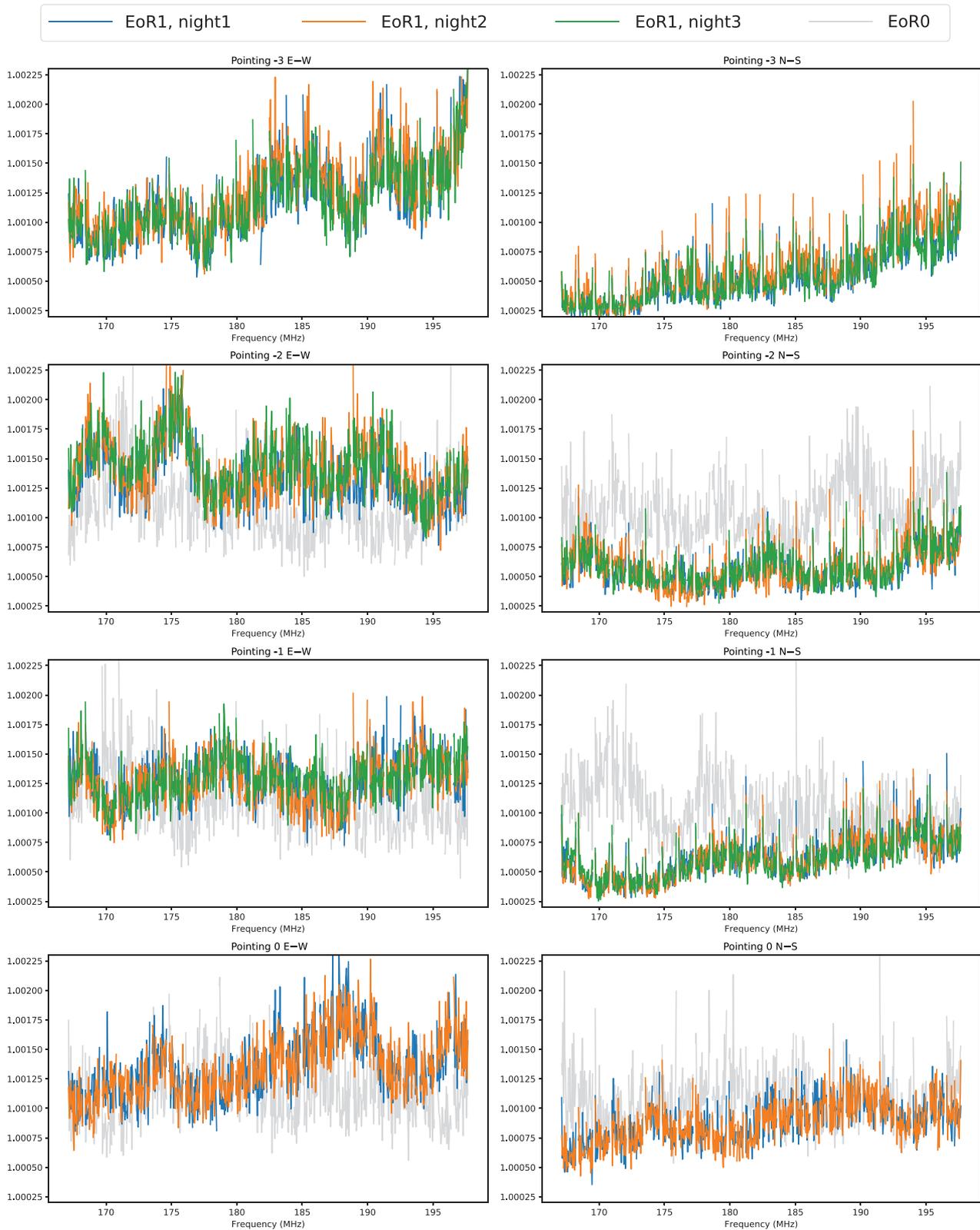


Figure 4. The RMS of magnitudes of Δ 's for observations sampled per pointing per polarisation and per night. EoR1 observations are sampled to have well-matched LSTs over EoR1 nights. We also sampled typical EoR0 observations at pointing -2 , -1 , and 0 . Four different colours denote different nights (see legend at the top of the figure for details). The columns represent the two linear polarisations and each of the four rows correspond to different pointings. We find that $|\Delta|$'s are consistent at any pointing over all the EoR1 nights with repeatable frequency structure, and the results from EoR0 observations have different spectral behaviour. It also shows $|\Delta|$'s for EoR1 observations have bigger values at E-W than at N-S (see further discussion in Section 5).

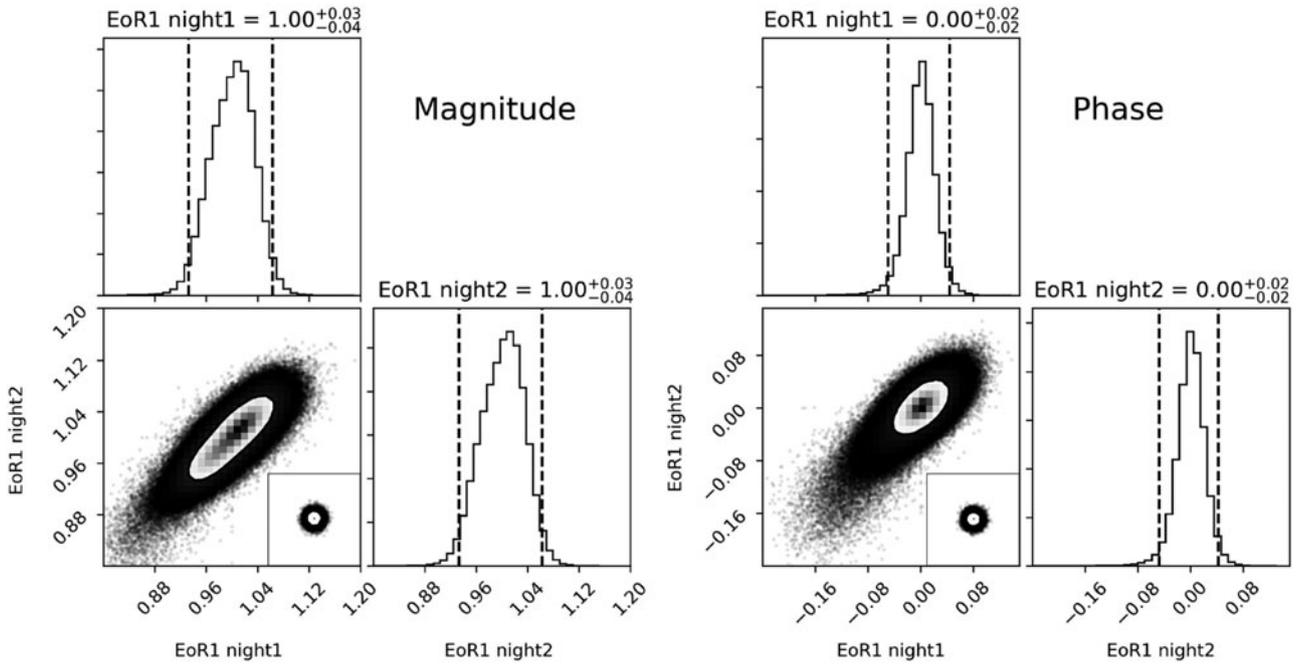


Figure 5. Histograms of redundant calibration solutions for EoR1 night1 and night2. Left: 1 and 2D histograms of the magnitudes of redundant calibration solutions for EoR1 night1 and night2. The subplot in the lower left panel is the 2D histogram of the magnitudes of redundant calibration solutions for two *simulated* data sets with common SNR. Right: 1 and 2D histograms of the phases (in radians) of redundant calibration solutions for EoR1 night1 and night2. The subplot in the lower left panel is the 2D histogram of the phases of redundant calibration solutions for two *simulated* data sets with common SNR. For 1D histograms, the area between the vertical dashed lines contains 95% of the samples; for 2D histograms, a contour (where the plot transitions from a point-by-point scatter plot to a pixelated density plot) contains $1 - e^{-2} \approx 84\%$ of the samples. Comparison with the simulated data shows that there is a scatter larger than would be expected from noise, EoR1 night1 and night2 are overall significantly correlated.

spectral behaviour, but no significant difference in overall magnitudes (recall that -3 is a flagged pointing for EoR0 due to Galactic contamination). A comparison of the East–West (E–W) polarisation and the North–South (N–S) polarisation shows that $|\Delta|$'s for EoR1 observations have bigger values at E–W than at N–S. We will return to this point in Section 5, but note that the polarisation dependence of redundant calibration solutions is significantly weaker in the EoR0 observations.

This result suggests that the spectral structure of our calibration solutions are quite repeatable for the same LST on different nights, but there is a moderate dependence on the pointing. While not shown, we have repeated this analysis for a number of other 2-min LST-matched data and we find the results to be very similar. For redundant calibration solutions, we find a significant difference between the EoR1 and EoR0 fields.

4.2.2. Repeatability of redundant calibration solutions

To make a more robust claim about the night-to-night repeatability of redundant calibration, we pair all the LST-matched redundant calibration Δ 's between two EoR1 nights (up to 10^{10} pairs in total, with one pair for each frequency, antenna, and time sample of the data set) and study the correlation between the two nights. Figure 5 shows the results of this study. The top row of two three-panel 'triangle' plots show the 1D and 2D histograms of all the redundant calibration Δ 's in magnitude (left) and phase (right). Within each triangle, the 1D histograms show the Δ 's from night1 (top left) and night2 (bottom right); the 2D contour/scatter plot (bottom left) shows the values from night1 plotted against the values from night2. For 1D histograms, the area between the vertical dashed lines contains 95% of the samples; for 2D histograms, a contour (where the plot transitions

from a point-by-point scatter plot to a pixelated density plot) contains $1 - e^{-2} \approx 84\%$ of the samples.

The elliptical shape of the 2D histograms, centred on the one-to-one line, suggest that the solutions found on EoR1 night1 and EoR1 night2 are highly correlated. To better assess the significance of this correlation, we perform a similar analysis using two sets of simulated data with common SNR. Since the simulated data (described in Section 4.1) are effectively perfect up to the noise, we expect no correlation in the Δ 's between the two nights (i.e. any change redundant calibration wants to make to the data set is merely a result of fitting noise). The results of this simulation are shown in the bottom two triangle plots in Figure 5. As expected, the 2D histograms are completely circular—meaning there is no correlation in the Δ 's between the two simulated nights. The width of these circles also gives us an estimate of the scatter we might expect in perfectly correlated data (because the simulations have an SNR matched to the data). Comparison between real and simulated data shows that the two EoR1 nights are not perfectly correlated up to the noise; rather, the scatter in these plots (effectively, the minor axis of the ellipses) is larger than would be expected from noise alone. However, the overall correlation is quite significant: the redundant calibration solutions for the two nights of data have Pearson's correlation coefficients of 0.826 (magnitude) and 0.651 (phase), confirming our intuition that redundant calibration is producing largely repeatable results from night to night.

4.3. Comparison of power spectra

Li et al. (2018) and (2019) evaluated tandem calibration techniques by quantifying the reduction of power in the EoR window (assumed to be due to residual foreground contamination). We

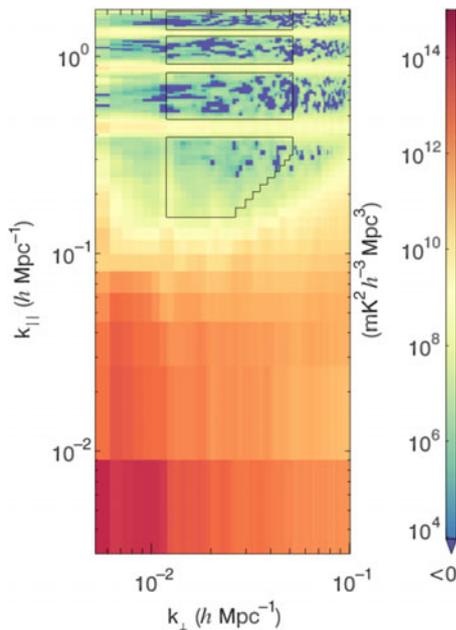


Figure 6. An example for 2D PS plot to highlight modes that will be used for 1D power in k in Figures 7, 8, and 9. To avoid coarse band contamination, we discard 5 k_{\parallel} bins around the centre of each coarse band mode. Low k_{\parallel} values are inside ‘the wedge’ and are therefore not included. In the k_{\perp} direction, we keep modes between a lower bound of 12λ and an upper bound of 50λ (Li et al. 2019). The cutting is performed in the 3D power spectrum and averaged to one-dimension.

apply that formalism in our analyses. To illustrate the difference between the two calibration pipelines, we subtract the PS of data with both FHD and redundant calibration applied from that of the FHD-only applied data in 3D k space. We then select regions of k space where both foregrounds and the periodic contamination from the 1.28-MHz band structures are minimised and bin in spherical annuli to make 1D PS difference plots versus k . We plot the 1D results as $\Delta^2(k) = \frac{k^3}{2\pi^2} P(k)$ to facilitate comparison with the literature. We perform this exercise for both the N–S polarisation and the E–W polarisation and for all four nights. We produce PS using the package Error Propagated Power Spectrum with Interleaved Observed Noise (eppsi⁸; Jacobs et al. 2016; Barry et al. 2019a). In 1D PS difference plots, solid lines represent positive values, that is, the power at that k mode is reduced by the application of redundant calibration to FHD calibrated data, while dashed lines represent negative values. In the EoR window, the former are regarded as ‘improvements’—redundant calibration has lowered the contamination in the region expected to be foreground-free. In Figure 6, we illustrate the modes we select in a 2D PS (although recall the differencing and mode selection takes place in 3D k space). For consistency, we use the same k space selections as used in Li et al. (2019). The selection gives us a range of $|k|$ modes to look at, although we draw particular attention to the modes in 1D PS difference plots where k is about $0.15 \sim 0.20 h\text{Mpc}^{-1}$; these modes have both the largest sensitivity to the EoR signal and come from regions of 2D k space closest to ‘the wedge’, where calibration errors can most easily cause contamination (Morales et al. 2019).

Figure 7 shows the 1D PS difference for each pointing of the EoR1 nights. For each PS plot, we integrate all observations within

the pointing not flagged by our quality metrics. For the E–W polarisation (blue lines), we can see repeatable improvements up to 10^4 mK^2 for all pointings. In contrast, improvements in the N–S polarisation (orange lines) only appear at the zenith pointing (the fourth row). At the off-zenith pointings (first three rows), we see no significant improvements of PS in the N–S polarisation.

A 1D pointing-by-pointing PS difference comparison of EoR0 and EoR1 is shown in Figures 8, and 9 shows PS differences after integrating all good pointings for each night. The improvements in the E–W polarisation are seen for all three EoR1 nights, whereas no significant improvements are seen for EoR0 in any pointing or in the integrated PS, consistent with the results of Li et al. (2019).

In general, when we do see consistent improvements, they are at levels of $\sim 10^3$ to 10^4 mK^2 . Given the recent measurements of Barry et al. (2019b), Li et al. (2019), and Trott et al. (2020), which use the MWA to achieve lowest limits of $\Delta^2 < 3.9 \times 10^3 \text{ mK}^2$, $\Delta^2 < 2.39 \times 10^3 \text{ mK}^2$, and $\Delta^2 < 1.8 \times 10^3 \text{ mK}^2$, respectively, we can see improvements of this scale can be significant.

5. Discussion

In Section 4, we presented analyses looking at both calibration solutions and 1D PS differences. In this section, we will further discuss two key aspects of our results: the comparison of the EoR1 and EoR0 results and the polarisation dependence of redundant calibration’s improvements in the PS of EoR1.

5.1. Comparison of EoR1 and EoR0

The noise simulations (Section 4.1) and general repeatability of both the gain solutions (Section 4.2) and the PS improvements (Section 4.3) across the three nights of EoR1 give us significant confidence that redundant calibration is indeed constraining real information about the telescope. It is then interesting to ask why we see no significant PS improvements in our EoR0 analysis (consistent with the results of Li et al. 2019).

The biggest difference between the two fields is the distribution of flux density on the sky. EoR0 has no significant bright or extended sources (Carroll et al. 2016) and as such is well modelled by the GLEAM catalogue. For EoR1, our sky model is likely not as accurate as the EoR0 due to the presence of Fornax A (see Figure 10) near the centre of the field. GLEAM does not include a Fornax A model (Hurley-Walker et al. 2016) itself; ours is produced directly from Phase I EoR observations using the techniques of Carroll et al. (2016). Figure 11 shows the residual images of EoR1 after our sky model has been subtracted; we return to the polarisation properties of this image in our discussion in Section 5.2, but we generally see a negative feature at the position of Fornax A in the lower right of the image—suggesting our model has an excess of flux density compared to the data. However, we emphasise that our Fornax A model is by no means ‘bad’; the post-subtraction residuals seen in Figure 11 are at the percent level compared to the source itself. Line et al. (2020) recently produced a new model for Fornax A using shapelets (which, at present, cannot be used by the FHD code), but when subtracted from MWA data, the residuals are at comparable level to those seen here (c.f. their Figure 9).

The fact that Fornax A is so difficult to model leads to interesting conclusions about the role of redundant calibration in the EoR1 field. If FHD produces calibration solutions with

⁸<https://github.com/EoRImaging/eppsi>.

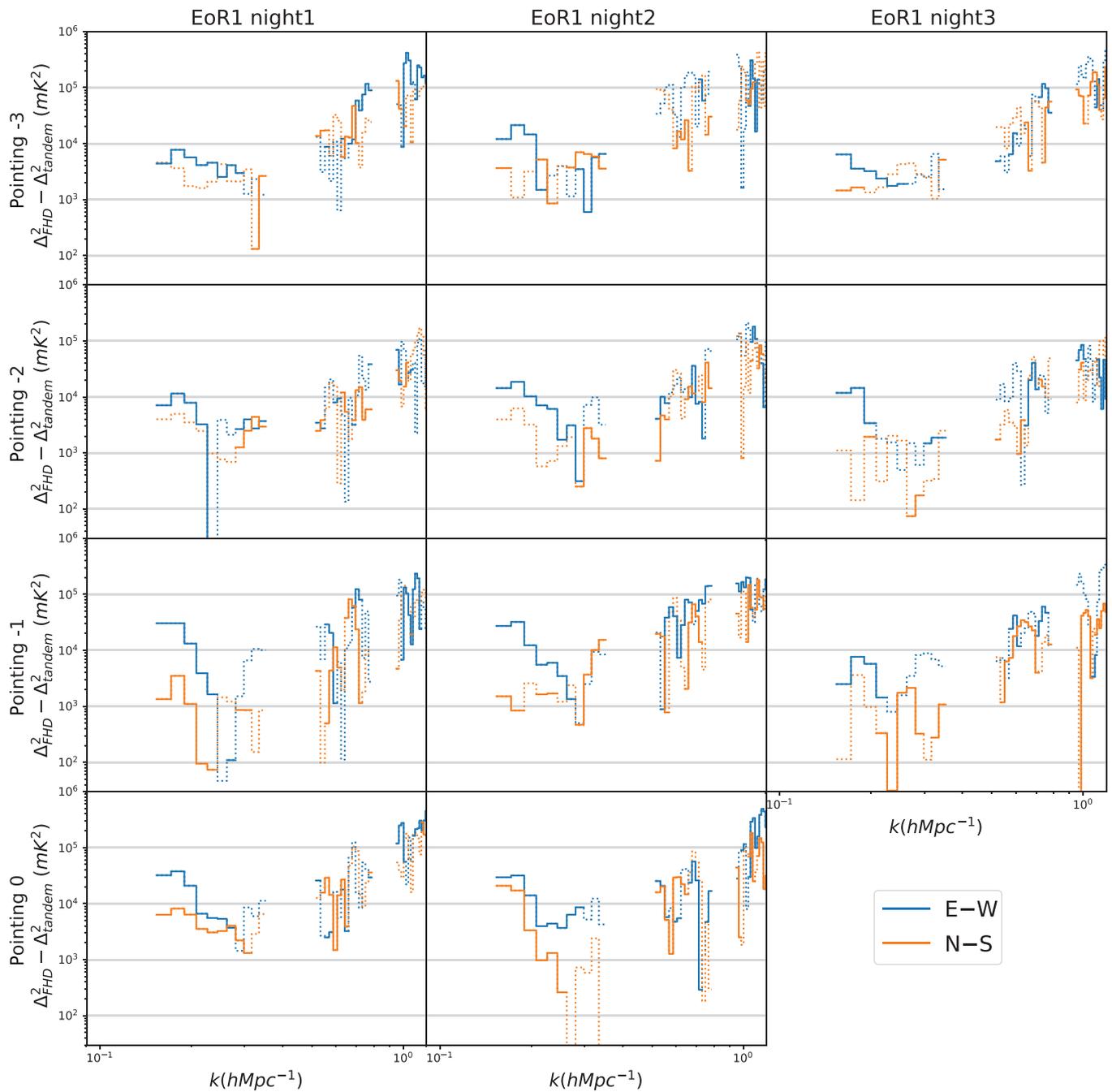


Figure 7. 1D PS difference pointing by pointing for the three EoR1 nights. Solid lines represent k modes where redundant calibration has reduced the overall power, while dotted lines imply a negative reduction in power (i.e. redundant calibration has introduced additional contamination). The three columns correspond to the three nights and the four rows correspond to different pointings. Blue represents the E–W polarisation while orange is N–S. Recall that the night3 zenith pointing is flagged due to ionospheric contamination. For the highest sensitivity modes ($\sim 0.15 - 0.20 \text{ hMpc}^{-1}$), we can see repeatable improvements up to 10^4 mK^2 for all pointings in the E–W polarisation (blue lines), while improvements in the N–S polarisation (orange lines) only appear at the zenith pointing (the fourth row).

frequency-dependent errors driven by sky model incompleteness (as suggested in Barry et al. 2016), then we will expect sky-based calibration to perform relatively worse on EoR1; redundant calibration, to the extent that it is sky-model-independent, will therefore have a more significant effect. Figure 4 indeed shows that redundant calibration solutions behave differently between EoR1 and EoR0, consistent with the idea that the model errors introduced in FHD are significantly different. In fact, even for

the observations at the same pointing but separated by several minutes, redundant calibration solutions can have significantly different spectra, which implies that FHD calibration is quite sensitive to LST. Our results therefore suggest that redundant calibration techniques can still play a role mitigating sky model errors in calibration—especially for fields with difficult, extended sources—even though sky model errors will still enter through the degenerate parameters (Byrne et al. 2019).

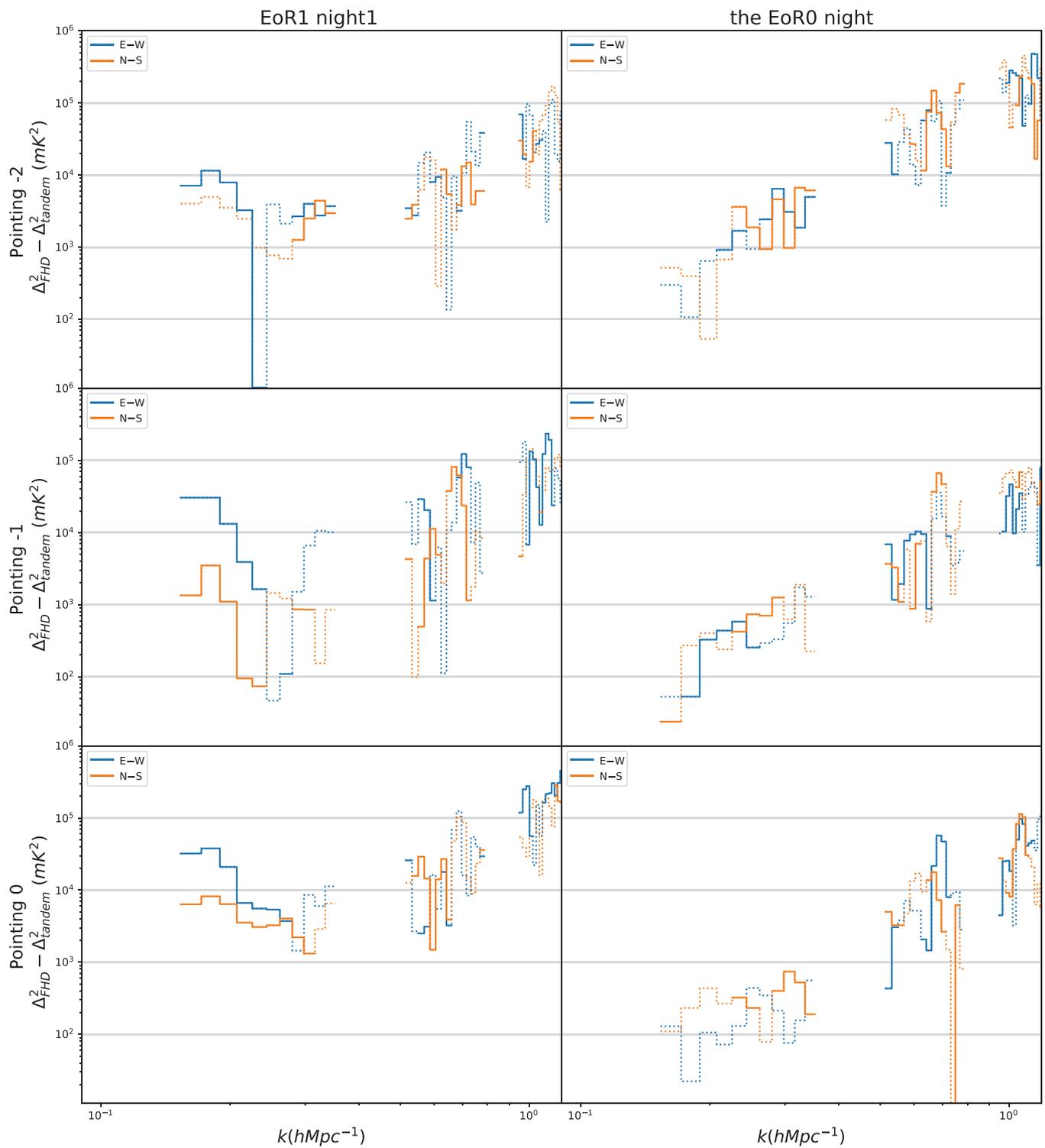


Figure 8. 1D PS difference for EoR1 night1 (left) and the EoR0 night (right). Solid lines represent modes where redundant calibration has reduced the power, while dotted lines imply an increase. Blue represents the E–W polarisation while orange is N–S. The improvements on PS in the E–W polarisation are seen for any pointing at all three EoR1 nights, whereas no significant improvements are seen for any pointing of the EoR0 night. No significant improvements in the N–S polarisation.

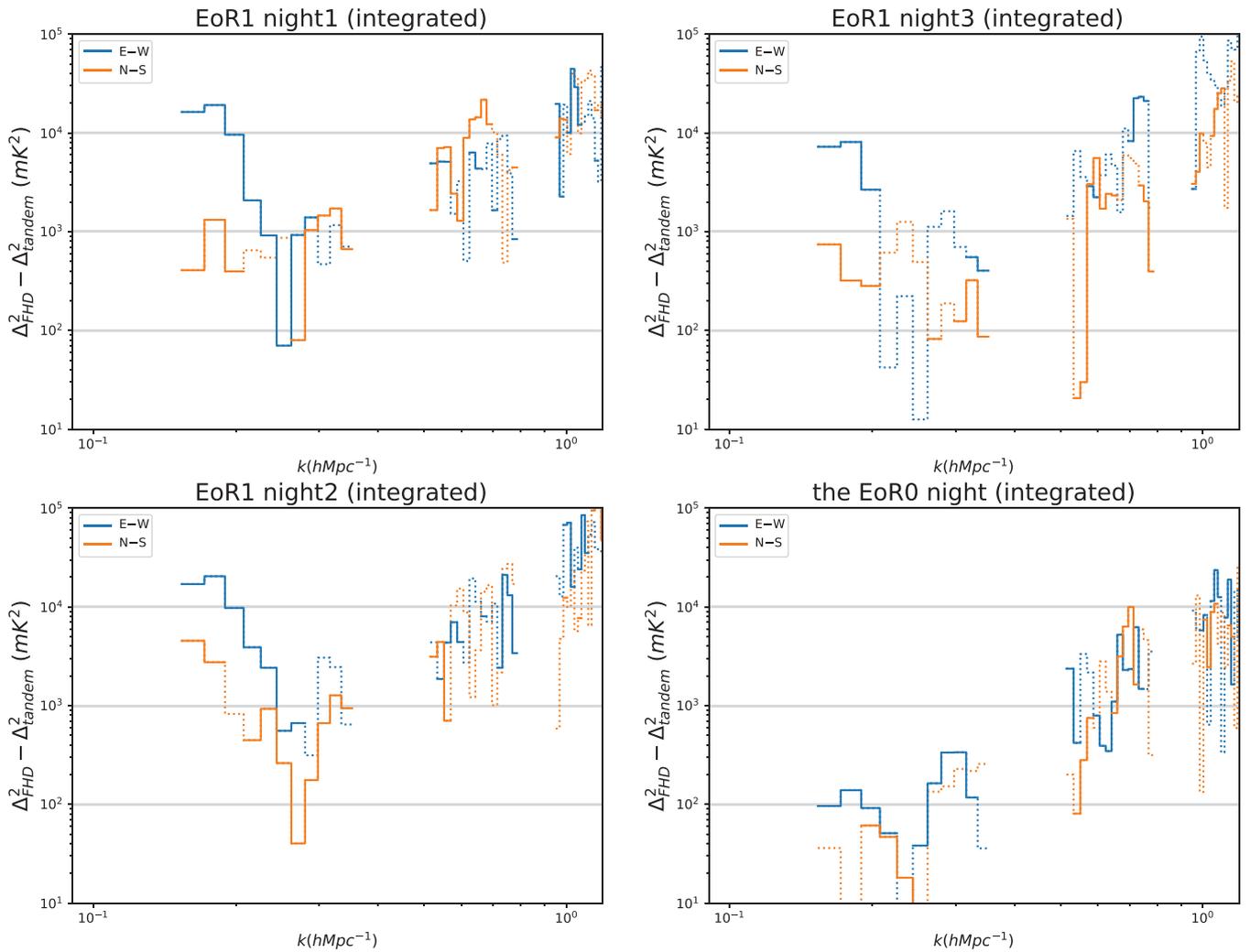


Figure 9. 1D PS difference of the full night for all nights. Solid lines represent a reduction of power from redundant calibration, while dotted lines imply an increase. Upper left: full night residual PS at EoR1 night1; bottom left: full night residual PS at EoR1 night2; upper right: full night residual PS at EoR1 night3; bottom right: full night residual PS at the EoR0 night. Blue represents polarisation E-W while orange represents polarisation N-S. The improvements in the integrated, E-W polarisation PS are seen for all three EoR1 nights, whereas no significant improvements are seen for EoR0. Neither field shows significant improvements in the N-S polarisation.

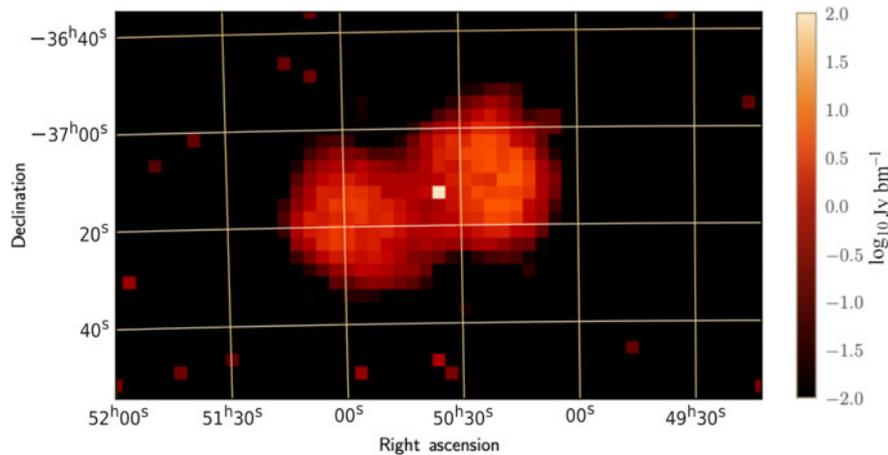


Figure 10. The extended source model used for Fornax A produced using the techniques of Carroll et al. 2016. This image is at 150 MHz and has a peak flux density of ~ 313 Jy. A total of 1925 components are used for this model.

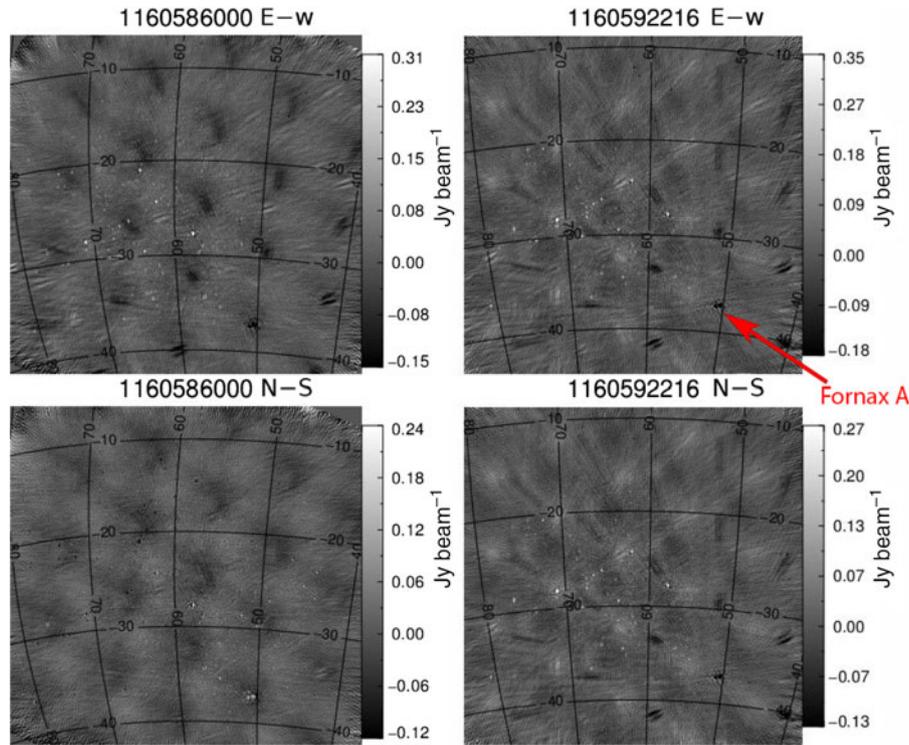


Figure 11. Residual images for both the E–W and N–S polarisation of a zenith observation (observation id: 1160592216) and an off-zenith observation (observation id: 1160586000). Upper left: residual image for E–W off-zenith; bottom left: residual image for N–S off-zenith; upper right: residual image for E–W zenith; bottom right: residual image for N–S zenith.

5.2. Polarisation dependence of redundant calibration performance

In Section 4.3, we saw that the PS improvements from redundant calibration in the EoR1 data were strongly polarisation-dependent: for the E–W polarisation, we saw substantive and consistent improvement at all pointings; for the N–S polarisation, there were no improvements at off-zenith pointings and, while there was improvement at the zenith pointing, it was less than that of the E–W polarisation.

To explain the strong polarisation dependence in redundant calibration’s impact on the PS, we picked one zenith observation and one off-zenith observation of EoR1 night1 and investigated their FHD products. In particular, we looked at the residual images for both polarisations, which are generated from calibrated, residual visibilities, as shown in Figure 11. As mentioned in Section 5.1, Fornax A is located near the centre of EoR1. Looking at residual images for both polarisations from an EoR1 zenith observation (observation ID 1160592216, right-hand column) leads us to conclude that our Fornax A model contains too much flux density, leading to the negative hole at RA 50°, Dec –37°. This error clearly produces many other artefacts in the image, including a strong set of negative features associated with sidelobes of Fornax A and a general miscalibration and undersubtraction of the numerous other points sources in the field.

This pattern is seen in both polarisations of the zenith pointing—both of which are improved by redundant calibration. We also see the oversubtraction of Fornax A and undersubtraction of other sources in the E–W residual image from an EoR1 off-zenith observation (observation ID 1160586000, left-hand column). This is consistent with the fact that redundant calibration

improves the E–W PS of an off-zenith observation. However, the N–S residual image looks quite different. Instead, we see an undersubtraction of Fornax A and oversubtraction of many point sources in the upper left of the image. The sidelobes of Fornax A are also significantly mitigated.

The most likely explanation for this effect is an error in our model of the tile primary beam. For any given pointing, the primary beam differs between the two linear polarisations. At zenith, this is simply because the dipoles for detecting one polarisation are rotated 90° from each other (and the beam pattern of a dipole—and therefore our tiles as well—is not symmetric under a 90° rotation). For pointings off-zenith, the difference between the two polarisations can be even more complicated, as the projection of the tile towards the pointing centre can make the effective spacing between one polarisation’s dipoles appear foreshortened by a different amount than the other. In general, our beam models are best for zenith but become less accurate for other pointings.

The key impact for this work is that the beam model error made at the position of a specific source (e.g. Fornax A) will *not* be the same in the two polarisations. What we find is that, in the N–S off-zenith pointing, a fortuitous combination of error in the model of Fornax A itself and the polarisation-dependent beam model error reduces the total level of sky model error. In the E–W polarisation and the N–S zenith pointing, the model error in Fornax A is the dominant source of error in our calibration solutions; such an error can be mitigated by redundant calibration, explaining why we see PS improvements with OMNICAL in all the observations except the N–S off-zenith pointings. This explanation is consistent with Figure 4, where the $|\Delta|$ ’s at the N–S polarisation are smaller than those at the E–W polarisation, particularly for off-zenith pointings.

6. Conclusions

In this paper, we studied the performance of tandem redundant and sky-based calibration in MWA Phase II data analysis in the EoR1 field. Using FHD and OMNICAL in tandem, we calibrated and processed three nights of EoR1 observations, as well as one night of EoR for comparison with Li et al. (2019). We performed a simulation to study redundant calibration's performance on fitting noise. We also analyzed calibration solutions and improvements in PS. Redundant calibration solutions were found to have night-to-night consistency and resulted in repeatable, substantive improvements in the PS of the EoR1 field. We also saw the strong polarisation dependence of redundant calibration's impact on the PS. Our principal findings are as follows:

Redundant calibration gives non-negligible gain solutions.

Our simulation shows that redundant calibration produces changes to the FHD sky-based gain solutions that are larger than would be expected if it were simply fitting to noise.

Redundant calibration's performance is repeatable. In our night-to-night comparisons, redundant calibration solutions and PS improvements are repeatable over the same range of LSTs.

Redundant calibration brings about substantive improvements in EoR1. Generally, redundant calibration makes bigger improvements in the EoR1 PS than in EoR0 PS, including changes up to 10^4 mK² in the E–W polarisation.

Redundant calibration brings about larger improvements when the sky and beam model of FHD is worse. Both the explanation for the bigger improvements in EoR1 PS than in EoR0 PS and our discussion for why we see the strong polarisation dependence in redundant calibration's impact on PS tell us that when sky model incompleteness errors most affect sky-based calibration, redundant calibration can result in larger improvements in the PS.

These findings suggest that redundant calibration can continue to play an important role in 21-cm cosmology analysis going forward. While the results from EoR0 suggest that with a good enough sky model for sky-based calibration, redundant calibration becomes ineffective, we note that the limits from Barry et al. (2019b), Li et al. (2019), and Trott et al. (2020) are all still one to two orders of magnitude above any plausible EoR signal models. Therefore, all we can conclude is that the errors that can be mitigated by redundant calibration are not the dominant source of error in those works. The present work shows that redundant calibration can indeed mitigate some level of sky-based calibration error and so it may continue to yield PS improvements once the current leading sources of systematic errors are mitigated. To end on a note of caution, however, we should be prepared for the possibility that true non-redundancy in the array may become a real problem for redundant calibration and that, once lower PS levels are reached, the inclusion of redundant calibration algorithms may introduce more errors than they mitigate. Research into new approaches that can balance sky-based and redundancy-based errors (e.g. Sievers 2017 and Byrne et al. 2020) is therefore of particular interest.

Acknowledgements. ZZ, JCP, and WL would like to acknowledge the support from NSF grant #1613040. BJH and MFM would like to acknowledge the support from NSF grants #1506024, #1613855, and #1643011. WL would like to acknowledge the Galkin Foundation Fellowship. KT is partially supported by Grand-in-Aid from the Ministry of Education, Culture, Sports, and Science and Technology (MEXT) of Japan, No. 15H05896, 16H05999 and 17H01110, and Bilateral Joint Research Projects of JSPS. This scientific

work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. We acknowledge the Pawsey Supercomputing Centre which is supported by the Western Australian and Australian Governments. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. CMT is supported by an ARC Future Fellowship under grant FT180100321. The International Centre for Radio Astronomy Research (ICRAR) is a Joint Venture of Curtin University and The University of Western Australia, funded by the Western Australian State government. The MWA Phase II upgrade project was supported by Australian Research Council LIEF grant LE160100031 and the Dunlap Institute for Astronomy and Astrophysics at the University of Toronto. This research was conducted using computation resources and services at the Center for Computation and Visualization, Brown University.

References

- Barry N., Beardsley A. P., Byrne R., Hazelton B., Morales M. F., Pober J. C., & Sullivan I. 2019a, *PASA*, 36, e026
- Barry, N., et al. 2019b, *ApJ*, 884, 1
- Barry, N., Hazelton, B., Sullivan, I., Morales, M. F., & Pober, J. C. 2016, *MNRAS*, 461, 3135
- Beardsley, A. P., et al. 2016, *ApJ*, 833, 102
- Beardsley, A. P., et al. 2019, arXiv e-prints, p. [arXiv:1910.02895](https://arxiv.org/abs/1910.02895)
- Bowman, J. D., et al. 2013, *PASA*, 30, e031
- Byrne, R., et al. 2019, *ApJ*, 875, 70
- Byrne, R., Morales, M. F., Hazelton, B., & Wilensky, M. 2020, arXiv e-prints, p. [arXiv:2004.08463](https://arxiv.org/abs/2004.08463)
- Carroll, P. A., et al. 2016, *MNRAS*, 461, 4151
- Datta, A., Bowman, J. D., & Carilli, C. L. 2010, *ApJ*, 724, 526
- DeBoer, D. R., et al. 2017, *PASP*, 129, 045001
- Dillon, J. S., et al. 2018, *MNRAS*, 477, 5670
- Dillon, J. S., et al. 2020, arXiv e-prints, p. [arXiv:2003.08399](https://arxiv.org/abs/2003.08399)
- Gupta, Y., et al. 2017, *CS*, 113, 707
- Hurley-Walker, N., et al. 2016, *MNRAS*, 464, 1146
- Jacobs, D. C., et al. 2016, *ApJ*, 825, 114
- Jordan, C. H., et al. 2017, *MNRAS*, 471, 3974
- Joseph, R. C., Trott, C. M., & Wayth, R. B. 2018, *AJ*, 156, 285
- Joseph, R. C., Trott, C. M., Wayth, R. B., & Nasirudin, A. 2019, *MNRAS*, 492, 2017
- Kern, N. S., et al. 2020, *ApJ*, 890, 122
- Kerrigan, J., et al. 2019, *MNRAS*, 488, 2605
- Li, W., et al. 2018, *ApJ*, 863, 170
- Li, W., Pober, J. C., Berry, N., & Hazelton, B. J. 2019, in American Astronomical Society Meeting Abstracts #233. p. 225.06
- Line, J. L. B., et al. 2020, arXiv e-prints, p. [arXiv:2005.09316](https://arxiv.org/abs/2005.09316)
- Liu, A., Parsons, A. R., & Trott, C. M. 2014, *PhRvD*, 90, 023018
- Liu, A., Tegmark, M., Morrison, S., Lutomirski, A., & Zaldarriaga, M. 2010, *MNRAS*, 408, 1029
- Liu, A., Zhang, Y., & Parsons, A. R. 2016, *ApJ*, 833, 242
- Mitchell, D. A., et al., 2008, *IEEE JSTSP*, 2, 707
- Morales, M. F., Beardsley, A., Pober, J., Barry, N., Hazelton, B., Jacobs, D., & Sullivan, I. 2019, *MNRAS*, 483, 2207
- Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, *ApJ*, 752, 137
- Offringa, A. R., et al. 2015, *PASA*, 32, e008
- Orosz, N., Dillon, J. S., Ewall-Wice, A., Parsons, A. R., & Thyagarajan, N. 2019, *MNRAS*, 487, 537
- 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
- Pober, J. C., et al. 2014, *ApJ*, 782, 66
- Salvini, S., & Wijnholds, S. J. 2014, 2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS), 1–4

- Sievers, J. L. 2017, arXiv e-prints, p. [arXiv:1701.01860](https://arxiv.org/abs/1701.01860)
- Sullivan, I. S., et al. 2012, *ApJ*, 759, 17
- Thompson, A. R., Moran, J. M., & Swenson George, W. J. 2017, *Interferometry and Synthesis in Radio Astronomy* (3rd edn.), doi: [10.1007/978-3-319-44431-4](https://doi.org/10.1007/978-3-319-44431-4).
- Thyagarajan, N., et al. 2013, *ApJ*, 776, 6
- Tingay, S. J., et al. 2013a, *PASA*, 30, e007
- Tingay, S. J., et al. 2013b, *AJ*, 146, 103
- Trott, C. M. 2014, *PASA*, 31, e026
- Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, *ApJ*, 757, 101
- Trott, C. M., et al. 2018, *ApJ*, 867, 15
- Trott, C. M., et al. 2020, *MNRAS*, 493, 4711
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., & de Bruyn, A. G. 2013, *AA*, 556, 1
- Vedantham, H., Shankar, N. U., & Subrahmanyan, R. 2012, *ApJ*, 745, 176
- Wayth, R. B., et al. 2018, *PASA*, 35, e033
- Wilensky, M. J., Morales, M. F., Hazelton, B. J., Barry, N., Byrne, R., & Roy, S. 2019, *PASP*, 131, 114507
- Yatawatta, S. 2015, *MNRAS*, 449, 4506
- Zheng, H., et al. 2014, *MNRAS*, 445, 1084