



Publication Year	2021
Acceptance in OA	2022-06-14T15:21:12Z
Title	H I intensity mapping with MeerKAT: calibration pipeline for multidish autocorrelation observations
Authors	Wang, Jingying, Santos, Mario G., BULL, PHILIP, Grainge, Keith, Cunnington, Steven, Fonseca, José, Irfan, Melis O., Li, Yichao, Pourtsidou, Alkistis, Soares, Paula S., SPINELLI, MARTA, BERNARDI, GIANNI, Engelbrecht, Brandon
Publisher's version (DOI)	10.1093/mnras/stab1365
Handle	http://hdl.handle.net/20.500.12386/32315
Journal	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
Volume	505

H I intensity mapping with MeerKAT: calibration pipeline for multidish autocorrelation observations

Jingying Wang¹,^{*} Mario G. Santos,^{1,2} Philip Bull^{1,2}, Keith Grainge,⁴ Steven Cunnington³,
José Fonseca^{5,6,3,1} Melis O. Irfan,^{1,3} Yichao Li¹, Alkistis Pourtsidou^{3,1}, Paula S. Soares³,
Marta Spinelli^{7,8}, Gianni Bernardi^{9,10,2} and Brandon Engelbrecht¹

¹Department of Physics and Astronomy, University of the Western Cape, Cape Town 7535, South Africa

²South African Radio Astronomy Observatory (SARAO), 2 Fir Street, Observatory, Cape Town 7925, South Africa

³Astronomy Unit, Queen Mary University of London, Mile End Road, London E1 4NS, UK

⁴Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, UK

⁵Dipartimento di Fisica ‘G. Galilei’, Università degli Studi di Padova, Via Marzolo 8, I-35131 Padova, Italy

⁶INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy

⁷INAF – Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34143 Trieste, Italy

⁸IFPU – Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy

⁹INAF – Istituto di Radioastronomia, Via Gobetti 101, I-40129 Bologna, Italy

¹⁰Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

Accepted 2021 May 8. Received 2021 May 8; in original form 2020 November 30

ABSTRACT

While most purpose-built 21-cm intensity mapping experiments are close-packed interferometer arrays, general-purpose dish arrays should also be capable of measuring the cosmological 21-cm signal. This can be achieved most efficiently if the array is used as a collection of scanning autocorrelation dishes rather than as an interferometer. As a first step towards demonstrating the feasibility of this observing strategy, we show that we are able to successfully calibrate dual-polarization autocorrelation data from 64 MeerKAT dishes in the *L* band (856–1712 MHz, 4096 channels), with 10.5 h of data retained from six nights of observing. We describe our calibration pipeline, which is based on multilevel radio frequency interference flagging, periodic noise diode injection to stabilize gain drifts, and an absolute calibration based on a multicomponent sky model. We show that it is sufficiently accurate to recover maps of diffuse celestial emission and point sources over a $10^\circ \times 30^\circ$ patch of the sky overlapping with the WiggleZ 11-h field. The reconstructed maps have a good level of consistency between per-dish maps and external data sets, with the estimated thermal noise limited to $1.4 \times$ the theoretical noise level (~ 2 mK). The residual maps have rms amplitudes below 0.1 K, corresponding to < 1 per cent of the model temperature. The reconstructed Galactic H I intensity map shows excellent agreement with the Effelsberg–Bonn H I Survey, and the flux of the radio galaxy 4C + 03.18 is recovered to within 3.6 per cent, which demonstrates that the autocorrelation can be successfully calibrated to give the zero-spacing flux and potentially help in the imaging of MeerKAT interferometric data. Our results provide a positive indication towards the feasibility of using MeerKAT and the future Square Kilometre Array to measure the H I intensity mapping signal and probe cosmology on degree scales and above.

Key words: instrumentation: spectrographs – methods: data analysis – methods: statistical – cosmology: observations – large-scale structure of Universe – radio lines: galaxies.

1 INTRODUCTION

Intensity mapping provides a way of measuring the cosmological clustering signal from large numbers of galaxies without having to resolve them individually. Using a spectral line, such as the 21-cm line of neutral hydrogen, it is possible to accurately recover the spatial distribution of brightness temperature fluctuations of the line in three dimensions (as a function of angle and redshift), and therefore – assuming that the neutral hydrogen is a biased tracer of the dark matter distribution – recover familiar clustering signals such as the

baryon acoustic oscillations (BAOs) and redshift space distortions in the late Universe, or the evolving ionization state of the intergalactic medium during Cosmic Dawn and the Epoch of Reionization. This method, known as H I intensity mapping (H I IM), is a promising technique for efficiently mapping the large-scale structure of the Universe and thus delivering precision constraints on cosmological models (Chang et al. 2008; Loeb & Wyithe 2008; Mao et al. 2008; Pritchard & Loeb 2008; Wyithe & Loeb 2008; Wyithe, Loeb & Geil 2008; Peterson et al. 2009; Bagla, Khandai & Datta 2010; Seo et al. 2010; Lidz et al. 2011; Ansari et al. 2012; Battye et al. 2013; Bull et al. 2015; Kovetz et al. 2017).

H I IM experiments differ from traditional spectroscopic galaxy surveys in their ability to recover 3D clustering information on large

* E-mail: astro.jywang@gmail.com

scales with a high survey speed over a very wide redshift range, potentially allowing us to survey extremely large fractions of the total available comoving volume of the cosmos along our past light-cone using this method. While the 21-cm signal itself is much fainter than many foreground sources – by a factor of around 10^4 – 10^5 compared to Galactic synchrotron emission for example – the advent of large radio telescope arrays, consisting of many elements with low-noise, high-bandwidth receivers, has now made it practical to perform large cosmological surveys with sufficient depth for detection within a reasonable observing time.

The majority of radio telescopes currently targeting the cosmological 21-cm signal are large interferometric arrays observing at low frequencies, i.e. at $z > 6$. These include the Low-Frequency Array,¹ the Murchison Widefield Array,² the Precision Array for Probing the Epoch of Reionization,³ the Hydrogen Epoch of Reionization Array,⁴ and the 21 Centimetre Array.⁵ Several arrays targeting intermediate redshifts are also either operating or under construction, including the Canadian Hydrogen Intensity Mapping Experiment⁶ and Tianlai,⁷ and Hydrogen Intensity and Real-Time Analysis experiment.⁸

While these interferometric experiments have tremendous sensitivity in principle, the only positive detections of the cosmological 21-cm signal to date have come from observations by ‘single-dish’ (or autocorrelation) telescopes. In particular, the Green Bank Telescope (GBT)⁹ and Parkes telescope¹⁰ have pioneered the detection of the H I signal in cross-correlation with optical galaxy surveys (Chang et al. 2010; Masui et al. 2013; Wolz et al. 2017; Anderson et al. 2018; Li, Staveley-Smith & Rhee 2021; Wolz et al. 2021). While neither Parkes nor GBT were able to definitively detect the autopower spectrum of the H I signal (but see Switzer et al. 2013, for upper limits), other single-dish IM projects are planned, including the purpose-built BINGO¹¹ instrument, and surveys on existing or planned telescopes such as the Five-hundred-meter Aperture Spherical Telescope,¹² and the Square Kilometre Array (SKA).¹³

In the case of the SKA, an H I IM survey has been proposed that would use SKA1-MID in autocorrelation mode. SKA1-MID is a mid-frequency array of 197 parabolic-dish antennas that will be built in South Africa, and a large 21-cm autocorrelation survey has been ranked as one of its science priority cases (Santos et al. 2015). While an interferometric intensity mapping survey would also be possible with SKA1-MID, this is expected to be insufficiently sensitive to large-scale cosmological features such as the BAO (Bull et al. 2015).

MeerKAT,¹⁴ as a pathfinder (and, ultimately, a component) of SKA1-MID, will play a crucial role in the development of a successful multidish autocorrelation survey method. In this work, we address the most critical calibration and systematics issues that are expected to affect dual-polarization autocorrelation observations with MeerKAT. With SKA1 scheduled to begin full operations in

2028, there is a good window of opportunity for MeerKAT to robustly demonstrate the single-dish survey method and make a significant first detection before the SKA comes online (Santos et al. 2017).

This paper is organized as follows:

Section 2 describes the basic set-up of our MeerKAT pilot survey.

Section 3 gives a detailed overview of our calibration pipeline.

Section 4 explains how the calibrated data are combined into multifrequency maps.

Section 5 studies the characteristics of the calibrated autocorrelation data, and presents a variety of quality and consistency checks on the calibrated data.

Section 6 presents a variety of comparisons with previous observations of diffuse Galactic emission and point sources.

Section 7 presents our conclusions.

Appendix A presents expressions for weighting the flux from point sources.

2 DESCRIPTION OF THE SURVEY DATA

In this section, we will review the basic parameters of our survey.

2.1 Target field

To check our observation and calibration strategy, we proposed a pilot survey through the MeerKAT open time call. The observation was done in the *L* band and targeted a single patch of about 200 deg^2 . The main factors affecting the choice of the sky region were to avoid the strong Galactic emission and to pick an area with good spectroscopic coverage in the redshifts we are observing ($z < 0.5$). This spectroscopic coverage will ultimately allow us to attempt a detection of the cross-correlation power spectrum between the H I IM signal and optical galaxy surveys, which is easier to achieve than a direct autocorrelation detection of the cosmological H I signal because residual foregrounds and other systematics that bias the autocorrelation drop out in cross-correlation with optical surveys as they are uncorrelated (e.g. see Chang et al. 2010; Masui et al. 2013; Wolz et al. 2017; Anderson et al. 2018).

As in Masui et al. (2013), we set our scans to cover the 11-h field of the WiggleZ Dark Energy Survey (Drinkwater et al. 2010), which is a large-scale spectroscopic survey of emission-line galaxies selected from ultraviolet (UV) and optical imaging. The footprint of our observation and the WiggleZ coverage is shown in Fig. 1. Note that this area is also covered by Baryon Oscillation Spectroscopic Survey (Reid et al. 2016). Our observations took place between 2019 February and July, when the WiggleZ 11-h field appeared with high elevation ($> 40^\circ$) during the night. The total usable observation time was about 10.5 h, which corresponds to around 630 h in total (before any flagging) when combined over 60 dishes on average. This can be compared to Masui et al. (2013), which covered the same region with a total area of $\sim 41 \text{ deg}^2$ and 190 h of total integration time, although at lower frequencies (700–900 MHz) and higher angular resolution due to the large ($\sim 100 \text{ m}$) diameter of the GBT dish.

2.2 Scanning strategy

Observations with single dishes typically require a scanning strategy where the dishes are rapidly moved across the sky. The aim is to cover the relevant angular scales within the stability time-scale of the instrument, when the gains are approximately constant (set by the so-called $1/f$ noise knee; e.g. Harper et al. 2018). This required the development of a new observing mode not previously available with MeerKAT. The MeerKAT antennas were set to scan in azimuth at

¹<http://www.lofar.org/>

²<http://www.mwatelescope.org/>

³<http://eor.berkeley.edu/>

⁴<http://reionization.org>

⁵<http://21cma.bao.ac.cn>

⁶<https://chime-experiment.ca/en>

⁷<http://tianlai.bao.ac.cn/>

⁸<https://hirax.ukzn.ac.za/>

⁹<https://greenbankobservatory.org/>

¹⁰<https://www.parkes.atnf.csiro.au/>

¹¹<http://www.bingotelescope.org/en/>

¹²<https://fast.bao.ac.cn/>

¹³<https://www.skatelescope.org/>

¹⁴<https://www.sarao.ac.za/>

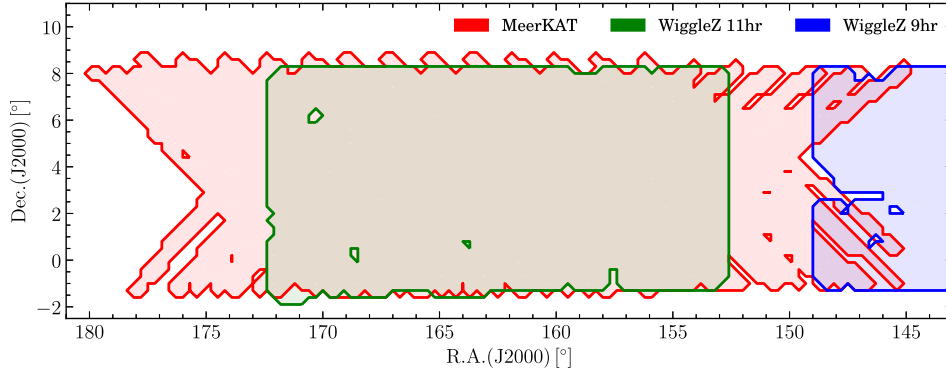


Figure 1. The footprint of the MeerKAT survey field and the WiggleZ survey fields.

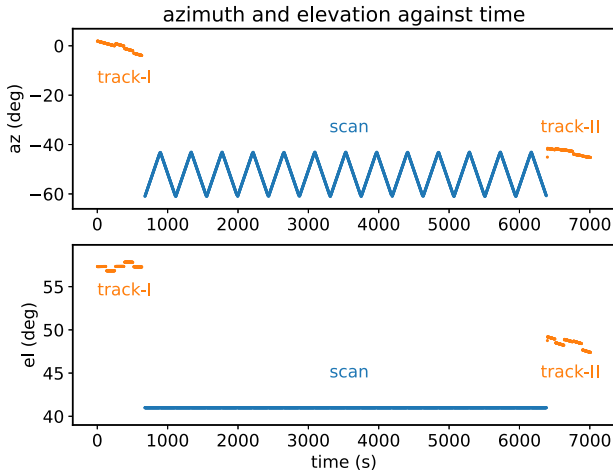


Figure 2. An example of the azimuth and elevation against time during a single scan for MeerKAT dish m000 in obs190225. Blue denotes the scanning of the target field and orange denotes the tracking of the calibrator source 3C 273.

constant elevation (see an example in Fig. 2) to minimize fluctuations of ground spill and airmass.

The telescope scan speed was set to 5 arcmin s^{-1} along azimuth, corresponding to a projected speed on the sky of $5 \cos(el) \text{ arcmin s}^{-1}$ (where el is the elevation angle). This ensures that the telescope pointing does not move significantly compared to the width of the primary beam ($\sim 1 \text{ deg}$) during a single time dump. With a time resolution of 2 s, this gives a scan speed of no more than 10 arcmin per time sample, which is well within the beam size. The system-induced $1/f$ -type variations for the MeerKAT receiver are well under the thermal noise fluctuations over ~ 100 -s time-scales after appropriate frequency filtering is performed (see Li et al. 2020, for details). For a scan speed of 5 arcmin s^{-1} , we can scan about 10 deg in 100 s while retaining gain stability. In order to maintain gain stability over longer time-scales, noise diodes attached to each receiver were fired for 1.8 s once every 20 s during the observation to provide a relative time-ordered data (TOD) calibration reference. The 1.8-s duration was chosen so that we can clearly see when the noise diode overlaps two time dumps, since we are currently unable to synchronize the noise diode fires with the correlator dumps (see Section 3.6).

The dishes are moved back and forth with a slew of 18 deg in each direction, corresponding to an observing time of about 200 s per stripe. At fixed elevation, two scans can be performed of each stripe per night, corresponding to when the field is rising and

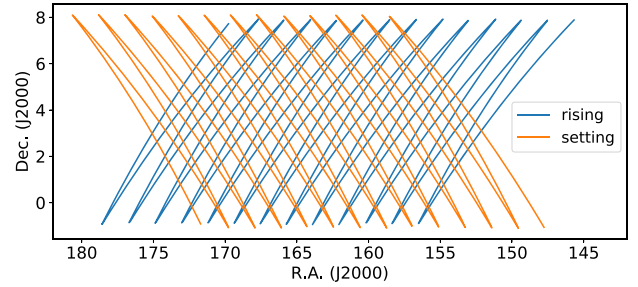


Figure 3. Coverage map of the cross routes for the field rising (blue) and setting (orange) scans in the same night.

Table 1. Specifications of the MeerKAT observations.

Antennas	All 64 MeerKAT dishes
Observation mode	Single-dish
Polarization	Linear (horiz. + vert.) feeds
System temperature	$\sim 16 \text{ K}$
Target field	WiggleZ 11-h field ($10^\circ \times 30^\circ$)
Frequency range	856–1712 MHz
Frequency resolution	0.2 MHz
Number of channels	4096
Time resolution	2 s
Exposure time	1.5 h $\times$ seven scans
Extent of azimuth scan	18 deg
Scan speed (along azimuth)	5 arcmin s^{-1}
Diode injection	Pattern mode (1.8 s per 20 s)

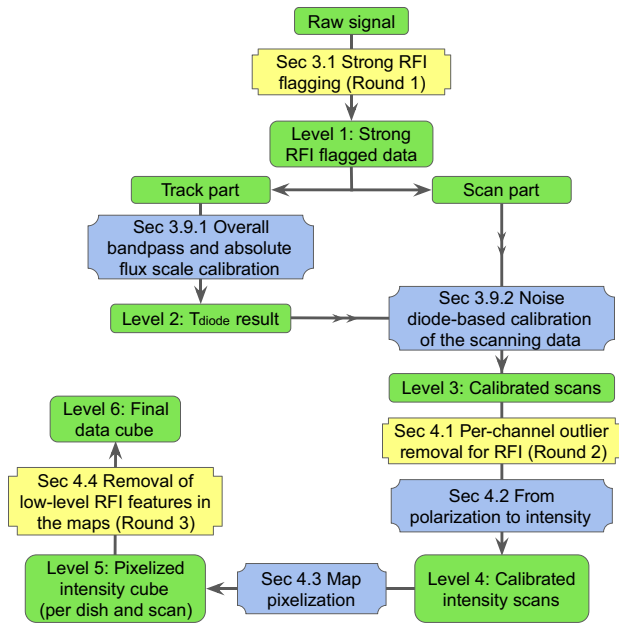
respectively. The two scans will cross each other as shown in Fig. 3, to achieve good sky coverage in the region of overlap. The duration of each set of scans is about 1.5 h, and before and after each scan we spent about 15 min tracking a nearby celestial point source to use as a bandpass calibrator and absolute flux calibrator. The technical details of the observations are listed in Table 1. Not all planned observations were successful due to equipment issues. After basic data quality checks, we were able to retrieve seven blocks of data that were suitable for further data analysis. These blocks are labelled by their observation start Unix time, as listed in Table 2.

3 DATA REDUCTION PIPELINE

The data reduction pipeline for the TOD includes steps for flagging of human-made radio frequency interference (RFI), bandpass and absolute calibration using a known point source, and calibration of receiver gain fluctuations based on interleaved signal injection from

Table 2. Basic information of observation blocks used in this work.

Block ID (Unix timestamp)	Short name (in this paper)	Observation start time (UTC time)	Sunset	az range ($^{\circ}$)	el ($^{\circ}$)	Calibrator source	Motion of field
obs1551037708	obs190224	2019-02-24 19:48:28	02-24 17:09:57	[41.6, 59.6]	42.0	3C 237	Rising
obs1551055211	obs190225	2019-02-25 00:40:11	02-24 17:09:57	[−61.0, −43.0]	41.0	3C 273	Setting
obs1553966342	obs190330	2019-03-30 17:19:02	03-30 16:29:35	[43.7, 61.7]	40.5	Pictor A	Rising
obs1554156377	obs190401	2019-04-01 22:06:17	03-31 16:28:22	[−61.0, −43.0]	41.0	3C 273	Setting
obs1556052116	obs190423	2019-04-23 20:41:56	04-23 16:02:05	[−60.3, −42.3]	41.5	3C 273	Setting
obs1556138397	obs190424	2019-04-24 20:39:57	04-24 16:01:04	[−60.3, −42.3]	41.5	3C 273	Setting
obs1562857793	obs190711	2019-07-11 15:09:53	07-11 15:43:57	[−55.3, −37.3]	43.4	3C 273	Setting

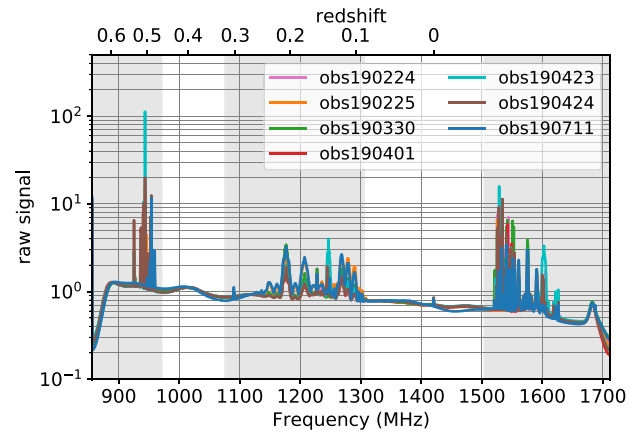
**Figure 4.** Flowchart showing each step in the calibration and map-making pipeline. The numbers of the sections from this paper where each step is described are also shown.

a noise diode (map making is described in Section 4). Fig. 4 shows the flowchart of the pipeline, with an indication of the different stages of calibrated data and a reference to the sections in the paper where they are described.

3.1 Strong RFI flagging (Round 1)

The RFI from modern telecommunications and satellites is an important contaminant of radio observations (Harper & Dickinson 2018). In Fig. 5, we show the time-averaged raw signal in a typical observation. Several RFI groups can be seen clearly. MeerKAT has RFI mitigation systems to prevent RFI before and during the observation (Jonas & MeerKAT Team 2016). However, as even the best RFI mitigation methods cannot completely prevent all RFI (Baan 2010), we must employ methods to reduce the effect of RFI after observation. One method is to flag those parts of the spectrum that are dominated by RFI and not use those frequencies/times in scientific data analysis. In this work, we use the RFI package of Signal Extraction and Emission Kartographer (SEEK; Akeret et al. 2017), which follows the SumThreshold algorithm (Offringa et al. 2010).

We denote the raw input data as $\hat{T}_{\text{raw}}(t, \nu)$, where the hat indicates data in arbitrary correlator units. We group the data in time according to the diode status (t_{nd}^0 , t_{nd}^{1a} , and t_{nd}^{1b} , corresponding to diode off, diode

**Figure 5.** The normalized time-averaged raw signal for m010h, all seven observations. The clusters of bright spike features correspond to RFI-dominated bands. Unshaded regions are the frequency bands for further data analysis.

on/starting, and diode on/stopping; see Section 3.6 for more details), as well as its scan state (whether tracking a calibrator source or scanning across the survey field). The SEEK flagging algorithm is then applied to these six groups of raw data separately, followed by an additional filter that discards any frequency channels or time dumps where > 70 per cent of the data are flagged.

This strong RFI flagging progress is applied to the data in two steps: first, on all of the raw data directly, to flag the strongly RFI-contaminated channels that are occupied by satellite communications, and secondly, focusing only on the target frequency bands, 971–1075 MHz (channels 550–1050) and 1305–1504 MHz (channels 2150–3100). In the second step, the value of $\hat{T}_{\text{raw}}$ at each frequency is normalized by the mean value along the time direction to account for the shape of the bandpass. At the same time, the locations of bright point sources are protected by temporary masks, as otherwise they would show as transient peaks in the time stream as the dishes scan across the sky, potentially leading to them being erroneously flagged.

This two-step SEEK flagging is applied to the per-dish HH and VV raw polarization signals separately. The union of the flags from the two polarizations is then adopted as the primary set of flags. Typically, about 35 per cent of the raw data are flagged in this step. Note that additional rounds of RFI flagging are performed during and after the calibration to prevent contamination from weaker RFI signals that can only be detected after further processing (see Section 4.4).

In the following calibration process, we focus on the 971–1075 MHz (channels 550–1050) and 1305–1504 MHz (channels 2150–3100) bands, considering the RFI distribution and the bandpass flatness. In these frequency ranges, most of the satellite communications are avoided, with only a few strong RFIs and some weak RFIs appearing intermittently.

3.2 Calibration model and strategy

Our autocorrelation calibration strategy is based on propagating an absolute flux and bandpass calibration obtained from tracking observations of bright point celestial point sources through to scanning observations of the survey field by using periodic noise diode fires as a relative calibration reference. We use a Bayesian approach to jointly fit the free parameters in our calibration and sky models, as described in the next section.

Bright celestial sources (e.g. active galactic nucleus and supernova remnants) are commonly taken as bandpass calibrators and absolute flux calibrators for single-dish observations. Despite being resolved in interferometric observations, well-known bright sources such as 3C 273 and Pictor A are sufficiently small compared to the single-dish beam size that we can treat them as point sources. The noise diode injections are taken as stable-in-time calibrators to remove receiver gain drifts, which are otherwise known to limit the sensitivity of single-dish observations. We examine the limitations of some of these assumptions in Section 5.

We construct a model for all of the components that will contribute to the total signal, and fit the free parameters by comparing the model to the TOD using a prescribed likelihood with priors. Our model, in temperature units, is expressed as

$$T_{\text{model}}(t, \nu) = T_{\text{ps}}(t, \nu) + T_{\text{diffuse}}(t, \nu) + T_{\text{el}}(t, \nu) + T_{\text{diode}}(t, \nu) + T_{\text{rec}}(t, \nu), \quad (1)$$

where T_{ps} , T_{diffuse} , and T_{el} are the antenna temperature models of the point source contributions, celestial diffuse component, and elevation-dependent terrestrial emission (atmosphere and ground-spill), respectively, while T_{diode} is the noise diode contribution and T_{rec} is the receiver temperature, i.e. the equivalent temperature of the noise power injected by the receiver. All quantities external to the dish are considered to be already convolved by the primary beam; e.g. $T_{\text{ps}}(t, \nu)$, $T_{\text{diffuse}}(t, \nu)$, and $T_{\text{el}}(t, \nu)$ should already have the beam applied. As will be described later, the models for T_{diffuse} and T_{el} are fixed, while T_{rec} is assumed as a smooth function of time (at each frequency). Therefore, any unknown offset on the fixed models (or indeed any unknown, slowly time changing, contamination) will be absorbed into T_{rec} .

In order to compare with the raw data from the correlator, we need to multiply the signal model by the gain, $g(t, \nu)$, so that

$$\hat{T}_{\text{model}}(t, \nu) = g(t, \nu) T_{\text{model}}(t, \nu), \quad (2)$$

where again the hat indicates a temperature in the (arbitrary) correlator units. The signal and gain models are then fitted to the TOD for each polarization, frequency channel, dish, and observation scan, all of which are treated independently (i.e. we do not consider correlations in parameters across these dimensions). Using the gain solution, we can then obtain the calibrated temperature as follows:

$$T_{\text{cal}}(t, \nu) \equiv \hat{T}_{\text{raw}}(t, \nu) / g(t, \nu). \quad (3)$$

The residual temperature difference between the signal model and calibrated data is then

$$T_{\text{res}}(t, \nu) \equiv T_{\text{cal}}(t, \nu) - T_{\text{model}}(t, \nu). \quad (4)$$

Note that throughout most of the paper we consider each polarization separately, so temperature values should be taken to refer to a single polarization. We will make it explicit when the two polarizations are combined.

3.3 Bayesian model fitting framework

The fitting process is done within a Bayesian framework, where the posterior probability is

$$p_{\text{post}} \propto p_{\text{prior}} \cdot \mathcal{L}(\hat{\mathbf{T}}_{\text{raw}} | \hat{\mathbf{T}}_{\text{model}}), \quad (5)$$

where p_{prior} is the prior probability for all the parameters and $\mathcal{L}(\hat{\mathbf{T}}_{\text{raw}} | \hat{\mathbf{T}}_{\text{model}})$ is the likelihood for the data given the model. $\hat{\mathbf{T}}_{\text{raw}}$ and $\hat{\mathbf{T}}_{\text{model}}$ are vectors representing all the timestamps.

We assume that the difference between the data and our model is noise dominated and that this noise is uncorrelated in time. In that case, for fitting at a certain frequency in a single scan, the joint likelihood can be written as

$$\mathcal{L}(\hat{\mathbf{T}}_{\text{raw}} | \hat{\mathbf{T}}_{\text{model}}) = \prod_{t_i=0}^{t_{\text{max}}} \frac{1}{\sqrt{2\pi\sigma^2(t_i)}} \exp \left\{ -\frac{[\hat{T}_{\text{raw}}(t_i) - \hat{T}_{\text{model}}(t_i)]^2}{2g^2(t_i, \nu)\sigma^2(t_i)} \right\}, \quad (6)$$

where $\sigma(t_i)$ is the expected noise level for which we take the standard radiometer equation (Wilson, Rohlfs & Hüttemeister 2009):

$$\sigma^2(t_i) = \frac{T_{\text{model}}^2}{\delta t \delta \nu}, \quad (7)$$

where δt and $\delta \nu$ account for the time and frequency resolution, respectively.

We assume independent prior probabilities, such that the total prior for the vector of all parameters $\mathbf{p}$, denoted by $p_{\text{prior}}(\mathbf{p})$, is the product of normalized 1D Gaussian distributions with mean $\bar{p}_j$ and standard deviation σ_j for each parameter p_j .

In principle, by simply maximizing the posterior probability p_{post} , we can obtain the optimal $\hat{\mathbf{p}}$. Usually, we take the log of this probability and, ignoring any constant terms (which are irrelevant for the fitting process), we get

$$\ln p_{\text{post}} = - \sum_{t_i=0}^{t_{\text{max}}} w(t_i) \left[\ln \sigma(t_i) + \frac{[\hat{T}_{\text{raw}}(t_i) - \hat{T}_{\text{model}}(t_i)]^2}{2g^2(t_i, \nu)\sigma^2(t_i)} \right] - \sum_j \frac{(p_j - \bar{p}_j)^2}{2\sigma_j^2}, \quad (8)$$

where $w(t_i)$ is an extra weight we apply if we want to change the contribution for a given time t_i . As explained later, in the tracking part of the timestream, we set $w(t_i) = 1$ for the entire integration time, while in the scan part we use $w(t_i) = 1$ for diode-off and $w(t_i) = 4$ for diode-on time samples. For every 10 time samples, two have noise injection; therefore, to equal the relevance of diode-on/off calibration we up-weight the diode-on by a factor of 4. This ensures that the noise diode contribution acts as the main calibrator of the gain drift in the scan part. Note that we cannot take $\sigma(t_i)$ as a constant since it also depends on the model. Finally, we can use a gradient descent (specifically, the modified Powell's method) to find

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} [-\ln p_{\text{post}}(\mathbf{p})]. \quad (9)$$

3.4 Beam model

The large dynamic range in brightness between foreground emission and the 21-cm fluctuations makes it particularly important to understand the beam pattern of the dishes. This is particularly relevant for calibrator point sources, since, as the beam is pointing on the source peak and outskirts (see an example in Fig. 6), we will measure the calibrator source contribution at different positions of the beam.

We adopt a beam model based on astro-holographic observations and electromagnetic simulations for MeerKAT, as described in Asad

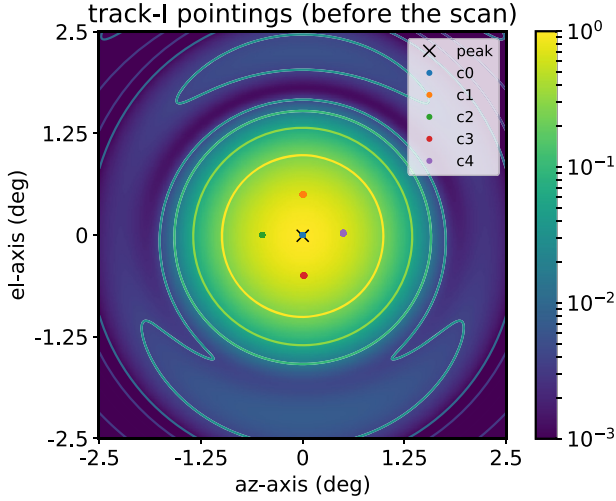


Figure 6. Normalized beam pattern for polarization HH at 1023 MHz, taken from the Asad et al. (2019) model. Note the asymmetry of the beam pattern beyond the main lobe. Also shown are the track-I pointing locations of the calibrator for observation obs190225.

et al. (2019). This model exhibits a degree of asymmetry (see Fig. 6), which is necessary to match multipointing observations of the calibrator sources, and the beam pattern has a non-trivial frequency and polarization dependence. While this level of complexity in the beam model comes at the cost of requiring point source coordinates to be projected to the telescope coordinate system for every time sample, the computational overhead involved is not significant compared with the overall runtime of the calibration pipeline. Nevertheless, in Section 6 we will also make use of a simple frequency-dependent two-dimensional (2D) Gaussian fit to the beam when performing certain comparisons with external data.

3.5 Calibrator point sources

We use a small number of bright point sources as flux and bandpass calibrators, selected for their low polarization fractions, and visibility at high elevations before and after each scan. For the WiggleZ 11-h field, our main calibrators were 3C 273, 3C 237, and Pictor A. We modelled the sources as strictly point-like, with frequency spectra

$$S_{\text{ps}}(\nu) = S_{\text{ps}}^{1.41} \left(\frac{\nu}{1.41 \text{ GHz}} \right)^{\alpha}, \quad (10)$$

where $S_{\text{ps}}^{1.41}$ is the flux density at 1.41 GHz derived from the Parkes catalogue (Wright & Otrupcek 1990), and α is the spectral index, derived using the observed fluxes at 1.41 GHz and 408 MHz. All three sources are bright in the L band (3C 237: 6.6 Jy; 3C 273: 42 Jy; and Pic A: 66 Jy at 1.41 GHz), and have polarization fractions of a few per cent or less (e.g. Perley, Roser & Meisenheimer 1997; Courvoisier 1998). 3C 273 in particular has a rather flat spectrum ($\alpha = -0.219$), which makes it an ideal bandpass calibrator (whereas $\alpha = -0.744$ and -0.683 for Pic A and 3C 237, respectively).

At a given frequency and time, the antenna temperature contribution from a point source to each polarization is

$$T_{\text{ps}}(t, \nu) = P_{\text{B}}(t, \nu) \frac{1}{2k_{\text{B}}} S_{\text{ps}}(\nu) A_{\text{eff}}(\nu), \quad (11)$$

where $P_{\text{B}}(t, \nu)$ is the normalized beam power pattern that depends on the relative displacement of the point source from the beam centre, k_{B} is the Boltzmann constant, and $A_{\text{eff}}(\nu)$ is the effective

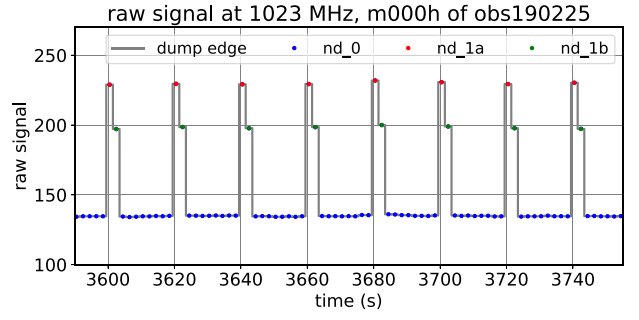


Figure 7. Enlarged plot for the raw signal along time in one frequency channel to show the distribution of nd_0 (time dumps without diode injection), nd_1a (the first dump with diode injection), and nd_1b (the second dump with diode injection).

beam area that is computed by integrating over the 2D beam pattern within 10-deg width.

3.6 Noise diode model

The temperatures of the noise diodes have previously been measured in the lab at several frequencies. We interpolate them to other frequency channels using a radial basis function spline and produce a reference value $T_{\text{diode}}^{\text{ref}}(\nu)$ to be used as a starting point in our fitting. The noise diode signals can be modelled as

$$T_{\text{diode}}(t, \nu) = f_{\text{diode}}(t) \bar{T}_{\text{diode}}(\nu), \quad (12)$$

where $\bar{T}_{\text{diode}}(\nu)$ is the noise diode injected temperature and $f_{\text{diode}}(t)$ is the fraction of time the noise diode is on in each 2-s time sample. The noise diodes are supposed to be very stable and in our calibration process we assume them to be constant in time.

Since at present the noise diode is not synchronized with the data sampling, we cannot guarantee that the injection only happens in a single time dump. This could be problematic if only a small fraction falls in one of the time samples, making it difficult to know when the noise diode was fired. To avoid this, we made the choice of injecting noise with a duration of 0.9 time samples (1.8 s). This way there is always a high chance that a single instance of the injection falls into two neighbouring samples (t_{nd}^{1a} and t_{nd}^{1b} hereafter) with a strong signal in both (Fig. 7). We used an injection period of 20 s as we expect the gains to be stable on this time-scale (Li et al. 2020). Since the asynchronicity between the noise injection and the data sampling is constant throughout one observation, $f_{\text{diode}}(t)$ can be expressed as a periodic function with a period of 20 s:

$$f_{\text{diode}}(t) = \begin{cases} f_{\text{d}}, & \text{if } t \in t_{\text{nd}}^{1a}, \\ 0.9 - f_{\text{d}}, & \text{if } t \in t_{\text{nd}}^{1b}, \\ 0, & \text{if } t \in t_{\text{nd}}^0, \end{cases} \quad (13)$$

where t_{nd}^0 represents the time dumps without noise injection and f_{d} is one of the parameters to be fitted in the calibration (see more details in Section 3.9).

3.7 Diffuse sky emission model

The (large-scale) diffuse celestial components include Galactic emission and the cosmic microwave background (CMB):

$$T_{\text{diffuse}}(t, \nu) = T_{\text{Gal}}(t, \nu) + \bar{T}_{\text{CMB}}, \quad (14)$$

where the Galactic emission $T_{\text{Gal}}(t, \nu)$ is derived from PySM (Thorne et al. 2017) and the CMB emission is set to be a uniform $\bar{T}_{\text{CMB}} =$

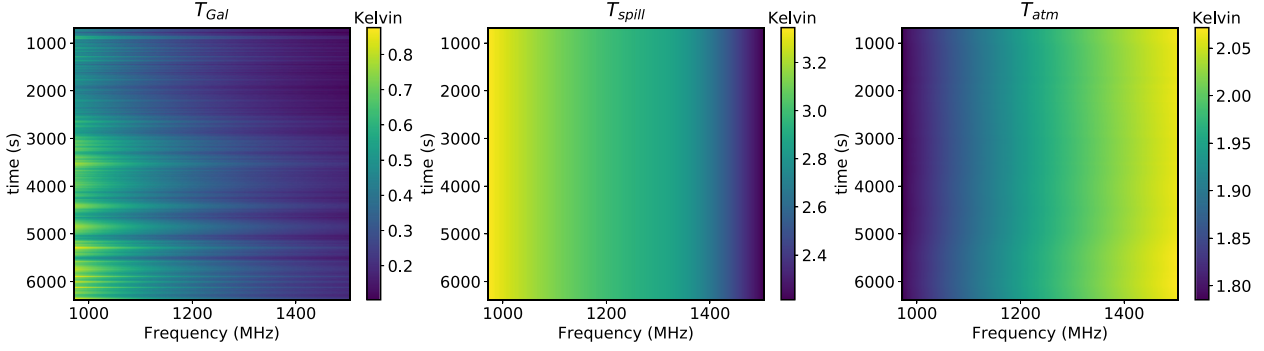


Figure 8. The models of Galactic emission, spillover from the ground and the atmospheric emission, m000h of obs190225 as an example.

2.725 K (Fixsen 2009). We assume these models to be fixed during the calibration process, with no free fitting parameters, but will return to the impact of this assumption in Sections 5 and 6. See Fig. 8 for an example frequency–time (waterfall) plot of the Galactic emission.

Note that we use the $s1$ model from PySM, which is based on the Haslam 408-MHz total intensity map [Remazeilles et al. 2015; (56 ± 1) -arcmin resolution] on large scales, but which has added structure generated as a simple lognormal random field on small scales. This map is used at a resolution of $n_{side} = 64$, corresponding to ~ 55 arcmin pixels. We do not apply a convolution with the MeerKAT beam to this map, instead relying on the limited map resolution to prevent the residuals from being contaminated with the synthetic small-scale structure. We have checked that the differences between the PySM model and Haslam map do not impact the calibration solutions (the effect on the gain is about 0.01 per cent).

3.8 Elevation-dependent terrestrial emission

Elevation-dependent terrestrial emission includes the atmospheric emission and spillover from the ground (i.e. ground pickup):

$$T_{el}(t, \nu) = T_{atm}(t, \nu) + T_{spill}(t, \nu). \quad (15)$$

A spillover model for the MeerKAT telescope has been provided by EMSS Antennas¹⁵ at several frequencies and elevations. We apply a 2D interpolation of their results to obtain $T_{spill}(t, \nu)$. The contribution of atmospheric emission to the receiver system temperature can be modelled as

$$T_{atm}(t, \nu) = T_{atm}^0(t) [1 - e^{-\tau \csc(e\ell)}], \quad (16)$$

where $T_{atm}^0(t) = 1.12 T_{st}(t) - 50$ is the atmospheric temperature related to the temperature measured at the MeerKAT weather station T_{st} in K, $\csc(e\ell)$ is the relative airmass along the direction of the beam pointing as a function of the elevation angle $e\ell$, and τ is the opacity of the atmosphere at zenith, which is related to the atmospheric temperature, air relative humidity, air pressure, telescope height above sea level, and frequency.¹⁶ In contrast to the atmospheric emission that is applied to $T_{atm}^0 \sim 270$ K, we ignore the atmospheric absorption since the transmission factor is very close to unity and will be multiplying temperatures of only a few K.

¹⁵<https://www.emssantennas.com/>

¹⁶See International Telecommunication Union Recommendation ITU-R P.676-9 for more details.

3.9 Calibration model fitting method

We take a two-step approach to the calibration process. The first step is to use the flux model of a strong point source during tracking observations to calibrate the bandpass and absolute flux, as well as the injection power of the noise diode. The second step then uses the noise diode injections to calibrate the TOD during the constant-elevation scanning phase. This is done per frequency, per dish, per polarization, and for each individual scan.

3.9.1 Overall bandpass and absolute flux scale calibration

We start by fitting the model in equations (1) and (2) to the raw autocorrelation signal from the tracking data around the calibrator point source. This is done for each polarization (HH and VV) at each frequency. Since we expect the gain to be smooth in time, we parametrize it as

$$g(t) = \sum_{n=0}^4 a_n P_n(x); \quad x = \frac{2(t - t_{\min})}{t_{\max} - t_{\min}} - 1, \quad (17)$$

where $P_n(x)$ are the Legendre polynomials. We choose this model on the basis that the coefficients of an orthogonal polynomial basis are less likely to be strongly correlated than that of an ordinary polynomial, and thus are easier to handle with numerical fitting techniques (cf. Bradley & Srivastava 1979; Shacham & Brauner 1997; Tian 1998). We checked that the order of the polynomial was sufficient to represent the data; choosing higher orders produced noise-dominated coefficients. The parameters to be fitted at each frequency are then $\{a_n\}$ for the gain function, plus $\bar{T}_{diode}$, T_{rec} , and f_d . We fit the parameters by maximizing equation (8) and setting the weights to $w(t_i) = 1$. For the calibration of the tracking parts of the TOD, we use the catalogue value of the flux of the calibrator point source, the beam model, and the difference of the raw signals between the centre and four outskirt pointings (e.g. see Fig. 6) to get an approximate gain level, which we then use as a starting point for a_0 .

The hyperparameters of the Gaussian priors from equation (8), $(\bar{p}_j, \sigma_j)$, are

$$\begin{aligned} \bar{T}_{diode}(\nu) &\sim (T_{diode}^{ref}(\nu), \sigma_{diode}), \\ T_{rec}(\nu) &\sim (T_{rec}^{ref}(\nu), 0.5T_{rec}^{ref}(\nu)), \end{aligned} \quad (18)$$

where the reference values T_{rec}^{ref} and T_{diode}^{ref} have been previously measured for MeerKAT (see Section 3.6), σ_{diode} is the measured standard deviation of T_{diode}^{ref} across the full frequency band, and we use $0.5T_{rec}^{ref}$ for the standard deviation of the receiver temperature as a conservative prior. Fig. 9 shows the fitting results for two tracking observations, before and after the scan. In this example, the fitted

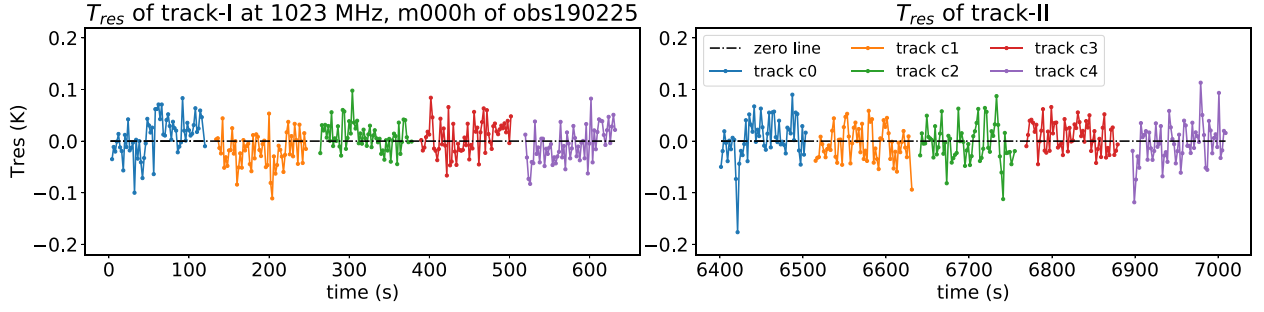


Figure 9. The residual between the observed signal and best-fitting model (equation 4) for the calibration source tracking data, in this case several pointings around 3C 273. The colours denote the different pointings, either directly on source (c0) or offset by ~ 0.5 deg (c1 . . 4) to probe different parts of the beam and surrounding diffuse emission (cf. Fig. 6).

$\bar{T}_{\text{diode}}$ are 20.08 and 20.15 K at 1023 MHz for the two tracks, respectively. So later we use their mean value $\bar{T}_{\text{diode}} = 20.12$ K for the next step in the calibration. The rms of the residuals is below the 1 per cent level of the total temperature.

3.9.2 Noise diode-based calibration of the scanning data

During the constant elevation scans, there is no bright calibrator source in the field of view, and so equation (1) becomes

$$T_{\text{model}}(t, \nu) = T_{\text{diffuse}} + T_{\text{el}} + T_{\text{diode}} + T_{\text{rec}}, \quad (19)$$

where the terms on the right-hand side implicitly depend on t and ν . Given the longer duration of these data, in this case we allow the receiver temperature T_{rec} to depend on time, parametrizing it using Legendre polynomials as we did previously for the gains. The parameters to be fitted are then

$$\begin{aligned} \{a_n\}, \text{ for } g(t) &= \sum_{n=0}^4 a_n P_n(x), \\ \{b_n\}, \text{ for } T_{\text{rec}}(t) &= \sum_{n=0}^3 b_n P_n(x), \\ f_d, \text{ constant number,} \end{aligned} \quad (20)$$

where x was defined in equation (17). We again checked that the order of the polynomials was sufficient to represent the data, with higher order choices returning noise-dominated coefficient values.

For the scan part, the gain is initially set to be $a_0 = 10$, which is close to the typical gain during the tracking phase. In this case, the hyperparameters ($\bar{p}_j, \sigma_j$) for the prior are

$$\bar{T}_{\text{rec}}(\nu) \sim (T_{\text{rec}}^{\text{ref}}(\nu), 0.5T_{\text{rec}}^{\text{ref}}(\nu)), \quad (21)$$

where $\bar{T}_{\text{rec}}(\nu)$ is the average of the receiver temperature along time. Again, we fit the parameters by solving equation (8), except in this case we set the weights to $w(t_i) = 1$ for the times when the noise diode is off and $w(t_i) = 4$ when the diode is on (both t_{nd}^{1a} and t_{nd}^{1b} samples). This accounts for the fact that the noise diodes should be our main calibrators in this part of the timestream.

We repeated this approach for all frequency channels, receivers, and observation blocks independently. Fig. 10 shows an example of the fitted gain and T_{rec} solutions as a function of frequency and time for a single receiver and observation block. While the time variations are smooth, as guaranteed by our polynomial model, there is some variation in the solutions between neighbouring frequency channels, most noticeably for T_{rec} . Smooth overall trends in the frequency direction are also observed, suggesting that a low-order polynomial parametrization in frequency could also be used to reduce the number of degrees of freedom in the calibration model, at the expense of requiring the observations to be fitted in frequency and time simultaneously.

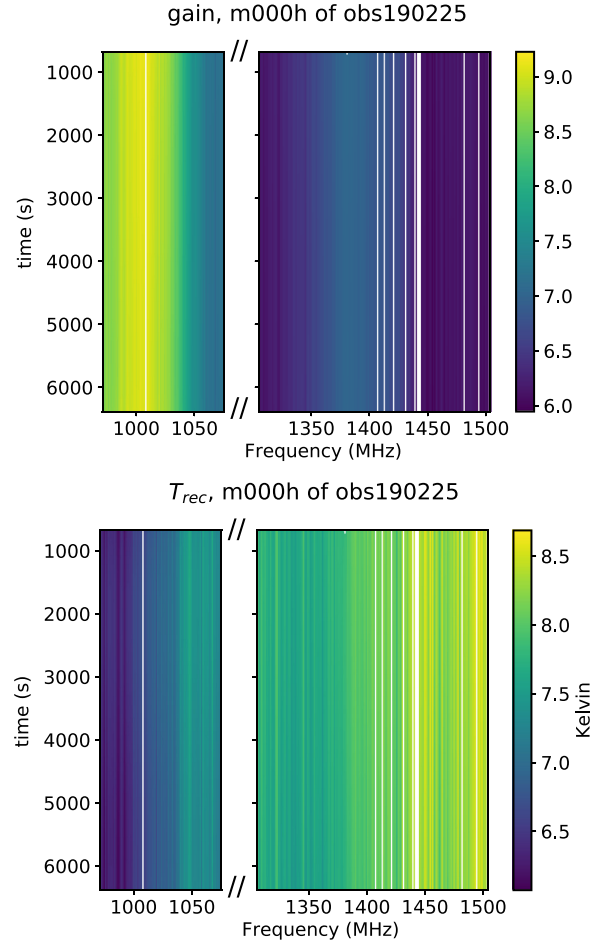


Figure 10. Best-fitting Legendre polynomial models for the time-dependent gain and T_{rec} , using dish m000h and obs190225 as an example. Note the smoothness of the solutions in time, with considerably more variation in the frequency direction (where smoothness was not enforced).

An example of the residual, T_{res} , from the TOD model fit for a single frequency channel is shown in Fig. 11. The calibrated signal and best-fitting model agree well, typically to within ~ 0.15 K for the entirety of the scan, although some structure is visible in the residuals. This is most noticeable from ~ 5000 s onwards in this example, where oscillations on a time-scale of the order of 100 s can be seen. This is comparable to the duration of a single scan (around 200 s).

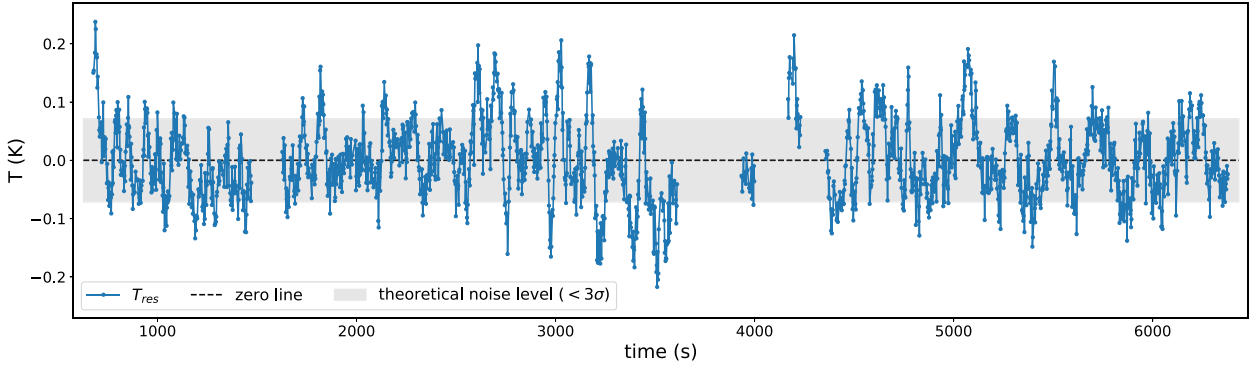


Figure 11. The residual between the observed data and fitting model in a single channel (1023 MHz) as a function of time, using receiver m000h and obs190225 as an example.

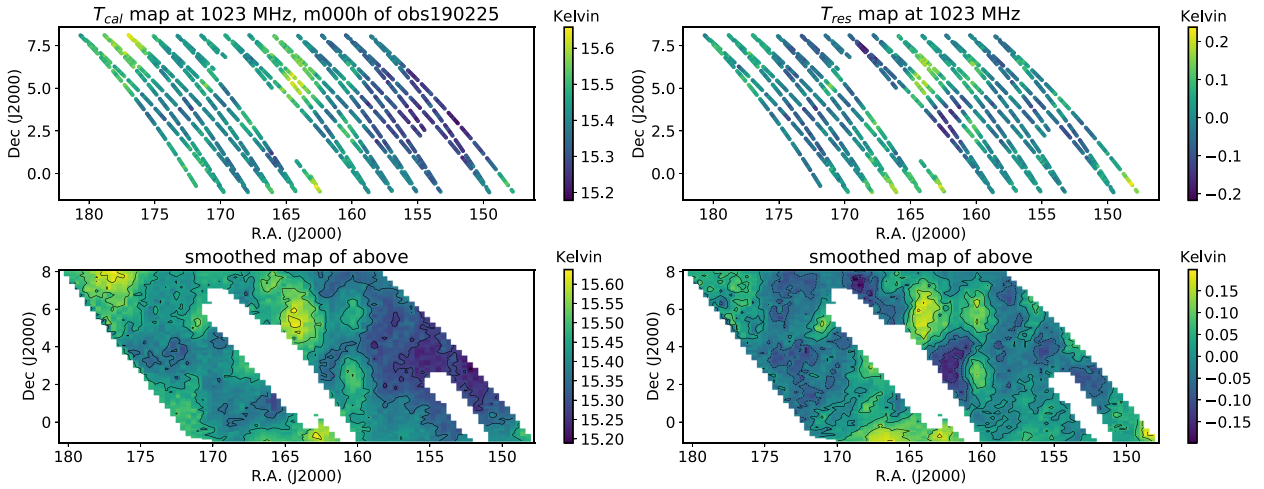


Figure 12. The calibrated data (upper panel) and interpolation maps (lower panel) for the total temperature (left) and residual (right), for a single frequency channel at 1023 MHz, receiver m000h, and observation obs190225.

We can study the structure of the residuals in more detail using Fig. 12, which shows the calibrated data and its residual for the same frequency and receiver. Several of the brighter structures in the residuals are seen to be correlated with features in the calibrated map, suggesting a missing component in our sky model. This is indeed found to be the case: As we will discuss in subsequent sections, much of this structure can be explained by relatively bright (> 1 Jy) point sources that are known to exist in this field, but which were not included in our sky model. We investigate including point sources in equation (19) later, by using the final combined T_{sky} (see details in Section 4.5) to replace the T_{diffuse} model. It results in changes of about 0.01 per cent on the gain map and 0.05 per cent on the T_{sky} map. The results confirm that the unmodelled point sources have a very small impact on the calibration.

4 COMBINED MAPS AND DATA CUBES

In this section, we present our methods for combining the calibrated data from all dishes, scans, and polarizations to form a final set of combined pixelized maps as a function of frequency. The increased depth of the data necessitates additional RFI flagging steps to remove outlying frequency channels and weak RFI features in individual stripes.

4.1 Per-channel outlier removal for RFI (Round 2)

The calibration process is done per frequency channel and per polarization, ignoring any potential correlation between the calibration of different channels. However, the resulting solutions $[T_{\text{cal}}(t, \nu), T_{\text{res}}(t, \nu), T_{\text{rec}}(t, \nu), \text{ and } g(t, \nu)]$ are expected to be relatively smooth in frequency. Despite this, contaminated data may still be present, even after the previous round of RFI flagging, and so we apply a per-channel filter to identify and mask such bad channels. To do so, we use the sigma clipping method (Astropy Collaboration 2013). For each dish and each observation, we apply spline fits to the time averages of $\bar{T}_{\text{cal}}(\nu)$, $\bar{T}_{\text{res}}(\nu)$, $\bar{T}_{\text{rec}}(\nu)$, and $\bar{g}(\nu)$. We then flag any outlier frequency channels present in any of the quantities, for all times and both polarizations. Outliers are defined as channels in which any of the time-averaged quantities deviate from the spline fit by $> 4\sigma$, where σ is defined as the sample standard deviations from the median. We perform the spline fitting and flagging process iteratively, running up to six iterations to make sure that the final result is stable (although in most cases a stable result is attained after one or two iterations).

4.2 From polarization to intensity

So far, all the analysis has been done per polarization (HH and VV). We have calibrated and flagged TOD data for the total measured

temperature $T_{\text{cal}}(t, \nu)$ using equation (3) and for the residual temperature $T_{\text{res}}(t, \nu)$ using equation (4). The calibrated sky temperature (per polarization) is obtained by removing the telescope, ground, and atmospheric contributions to the total signal:

$$T_{\text{sky}}(t, \nu) \equiv T_{\text{cal}}(t, \nu) - T_{\text{el}}(t, \nu) - T_{\text{rec}}(t, \nu). \quad (22)$$

We can then construct equivalent TOD data arrays of total intensity (corresponding to Stokes I), as the mean value of the two calibrated polarization temperatures, HH and VV,

$$T(t, \nu) = \frac{1}{2} (T_{\text{HH}}(t, \nu) + T_{\text{VV}}(t, \nu)). \quad (23)$$

We note that this is a pseudo-stokes I (or temperature T) as it does not correct for the primary beam shape and polarization leakage. The polarization contribution to stokes I is negligible at the current sensitivity level since the telescope leakage is below the 10 per cent level and the Galactic polarization is also around 10 per cent of the intensity. Correction for the beam asymmetries can be relevant in the presence of strong point sources (like we do for the calibrator). For diffuse emission, it should be enough to account for the beam smoothing using a symmetric beam model (see e.g. Matshawule et al. 2020).

4.3 Map pixelization

The intensity maps from different blocks need to be added together to get a better sky coverage and higher signal-to-noise (S/N) ratio (this can be done for Stokes I, while Stokes Q, U, and V would require to compensate for the sky rotation before averaging). Since different scans have different pointings (ra, dec), the intensity maps need to be pixelized first and then joined together in the same coordinates system. We chose the Zenith Equal Area map projection method (Astropy Collaboration 2013) to get the coordinates of each TOD point on a pixel grid, denoted as $(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}})$, with a pixel size of 0.3 deg chosen to be of about one-third of the beam size. We do the pixelization per scan and per dish. The relation between time and pixel is not necessarily one to one given the pixel size.

The final maps combining all scans and dishes, $\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu)$, were created by performing a simple average of all unflagged data points that fell in each pixel,¹⁷

$$\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu) = \frac{1}{N_{\text{count}}} \sum_{i=1}^{N_{\text{count}}} T_i(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu), \quad (24)$$

where N_{count} is the number of unflagged data points per pixel and per frequency channel, with $N_{\text{count}, \text{max}} = 366$ in the best case scenario. The final residual maps were created by applying the same averaging procedure to the individual residual maps. Note that in principle we would have up to $64 \times 7 = 448$ independent scans in the data, but this has not been the case due to malfunctioning dishes during the observation blocks.

4.4 Removal of low-level RFI features in the maps (Round 3)

While for some astronomical observations one can safely ignore low levels of residual RFI, the faintness of the H I IM clustering signal means that further RFI removal is generally required. After

¹⁷Inverse T_{sys}^2 -weighted maps were also generated for testing purposes, but differed from the simple average maps by $\lesssim 5$ mK per pixel; a more detailed treatment of map making is deferred until later work (Doré et al. 2001; Natoli et al. 2001).

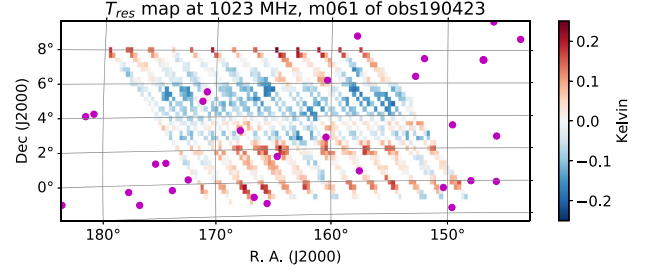


Figure 13. Example of a pixelized intensity map with weak RFI contamination that would be removed by the Round 3 RFI filter. These data are for single scan at 1023 MHz, for dish m061 and obs190423. Magenta dots mark the positions of point sources with flux > 1 Jy at 1.4 GHz.

the first rounds of RFI removal and calibration, some relatively faint line-shaped structures are detectable in maps across relatively broad frequency bands (see Fig. 13 for an example). We believe these features to be caused by geostationary satellites, e.g. due to heavily suppressed spillover of emission outside of their main transmitting band (Harper & Dickinson 2018). The previous rounds of flagging likely missed most of this emission because it is faint in comparison to other RFI sources, and because the spatial coherence of the signal on the sky did not produce as obvious a pattern in the timestream domain.

We therefore apply another round of RFI flagging to remove these features, taking advantage of the characteristic line structures left behind in the residual maps. As seen in Fig. 13, these structures occur at a fixed declination.

Our flagging method proceeds as follows:

- (i) Point sources with flux > 1 Jy in the total combined residual maps, $\bar{T}_{\text{res}}$, are temporarily masked (see the magenta points in Fig. 13). Neighbouring pixels within 1 deg are also masked.
- (ii) For each frequency channel and declination, we calculate the mean and standard deviation of the total combined residual maps along the RA direction.
- (iii) Similar to the first step, we temporarily mask the point sources on the single scan maps (per dish), T_{res} , and calculate the median value along the RA direction for each frequency and declination.
- (iv) Finally, we check whether the median at each scan, frequency, and declination is within 2.5 standard deviations of the mean of the total combined residual maps, and flag the whole scan in that frequency channel if any points exceed this threshold.

We iterate this process three times. The 2.5σ threshold was chosen as it gave good results when comparing with affected scans by eye. The decision to flag the entire scan is aggressive, but reduces the chance that slightly weaker RFI at neighbouring declinations remains unflagged. The result is a significant amount of additional flagging that particularly affects the higher frequency sub-band, as shown in Fig. 14.

4.5 Final maps

Our final output is the combined calibrated temperature maps for all scans and all dishes at all frequencies, after the three rounds of RFI flagging discussed previously. We show the final sky temperature map at 1023 MHz in Fig. 15, along with the residual map for all observations combined, and the PySM diffuse Galactic emission model with the same sky coverage. Much of the diffuse Galactic structure has been recovered on large scales, and the contribution of the bright point sources that were not included in our calibration

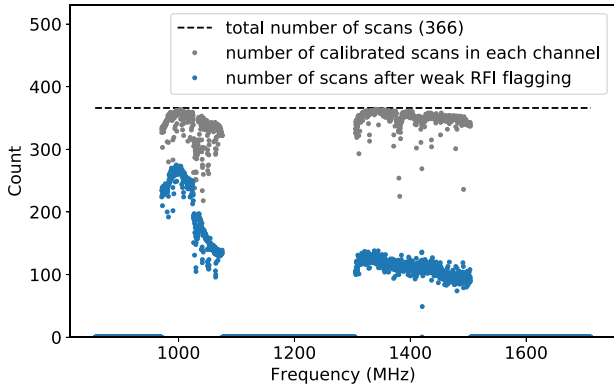


Figure 14. The total number of scans available before (grey) and after (blue) the Round 3 weak RFI flagging. The higher frequency band is particularly badly affected by low-level broadband RFI.

model is seen clearly in the residual in most instances. We use the recovered flux and frequency spectra of these point sources to evaluate the quality of our final calibrated maps in Section 6.3.

In comparing Fig. 15 with the map and residuals before Round 3 (weak RFI) flagging, we found little difference, which suggests that the weak RFI was not a large contribution in comparison with the dominant foreground signals. This is expected – if it was brighter, this kind of RFI would likely have been flagged in previous rounds. This is encouraging if these data are to be used for, e.g. studying Galactic foregrounds, as the effect of residual weak RFI does not seem to be significant. The efficacy of this round of flagging in helping to detect much fainter signals (such as the 21-cm fluctuations) is less clear, however.

Note that the residual map should in principle contain a series of other unmodelled components as well as the bright point sources, including noise, small-scale Galactic structures that are not included in PySM, the cosmological signal (below the noise level here), and other extragalactic emission, i.e. confusion noise. Calibration model errors and unmodelled instrumental artefacts will also contribute to the residuals. In subsequent sections, we will therefore use the properties of the residual maps as a way to check the quality of the data and our calibration solutions. Further analysis should be applied to the full calibrated data, either $\bar{T}_{\text{sky}}$ or $\bar{T}_{\text{cal}}$, however, rather than $\bar{T}_{\text{res}}$. This is because there is a risk of oversubtraction of sky components and therefore signal loss if the residuals are used directly.

5 CHARACTERISTICS OF AUTOCORRELATION DATA

In this section, we study the properties of the autocorrelation data, particularly with respect to the accuracy of the calibration and the effectiveness of the RFI flagging. Towards the end of this section, we also plot the per-dish sky maps to show the variations in data quality between dishes, and then calculate the principal components of the whole data set as a way of demonstrating the correlation structure and dynamic range of the data.

5.1 Gain solutions and bandpass shape

It is instructive to study the structure of the gain solutions, $g(t, \nu)$, that are obtained by the calibration pipeline. The ideal for an intensity mapping experiment is to find solutions that are smoothly varying in time and frequency, as in this case it is less likely that sky signal on the scales of interest will be erroneously absorbed into the gains, leading

to a spurious modulation of the actual cosmological signal. Smooth variations in time are also indicative of the stability of the instrument. We also hope to find gain solutions that are independent of the sky; gains arise from purely instrumental effects, and so solutions that are appreciably correlated with the sky suggest the presence of errors in the sky model for example.

In terms of time stability of the gains, first recall the gain solutions that were plotted in Fig. 10 for scan m000h of obs190225. The gain along time is rather flat, which means the gain does not change too much during the 1.5-h observation. This is representative of the gain solutions we obtained for all observations and receivers.

Next, we calculated the time-averaged mean gain against frequency, $g(\nu) = \langle g(\nu, t) \rangle_t$, also known as the bandpass. The bandpass shape is expected to be stable for each receiver. To check this, we compared the bandpass from the same MeerKAT receiver but different observations, normalizing each bandpass using its mean value (since the absolute bandpass values also incorporate the effects of gain drifts in time). Fig. 16 shows the resulting bandpass for receiver m000h over several days of observations as an example. While there are variations from observation to observation, these are limited to < 10 per cent, and there is good overall consistency in the bandpass shape across all observations.

5.2 RFI statistics

In our analysis, RFI flagging has been performed as a multistep process starting from the masking of the strongest contamination, to a pre-channel outlier removal and finally a map-based cleaning of residual weak RFI sources. In this section, we discuss the fraction of data that have been discarded through these RFI flagging processes.

Globally, strong RFI flagging (i.e. Round 1 described in Section 3.1) varies from a total of ~ 40 per cent to ~ 60 per cent across observations. Focusing on the two target frequency bands, where most of RFI from satellite is avoided, flagging is below ~ 25 per cent, reaching values down to ~ 5 per cent for the best observations. We show in Fig. 17 the fraction of RFI flagged data as a function of frequency for each observation.

We report in Table 3 the fraction of flagged data with respect to the initial data, for Round 1 and for the subsequent RFI flagging rounds. The flagging is computed over time and frequency, considering either the lower band (i.e. 971–1075 MHz, channels 550–1050) or the higher band (i.e. 1305–1504 MHz, channels 2150–3100). Channel filtering (i.e. Round 2 described in Section 4.1) increases the flagged fraction from about +1 per cent to +20 per cent in the lower frequency band, with similar values for the higher frequency band. Map-based weak RFI removal (i.e. Round 3 described Section 4.4) increases further the flagging up to very high percentage, in particular for the higher frequency band.

We show in Fig. 18, separately for each observation, the percentage of antennas, out the 64 MeerKAT dishes, for which we retain data at the end of the map making procedure, as a function of frequency. We recall that flagging is not only due to RFI but also includes other issues such as possible bad functioning of a specific dish.

5.3 Receiver temperature estimate

For a radio receiver, the system temperature (T_{sys}) is a combination of the internal temperature of the system as well as Galactic and atmospheric contributions:

$$T_{\text{sys}}(t, \nu) = \bar{T}_{\text{CMB}} + T_{\text{Gal}}(t, \nu) + T_{\text{el}}(t, \nu) + T_{\text{rec}}(\nu). \quad (25)$$

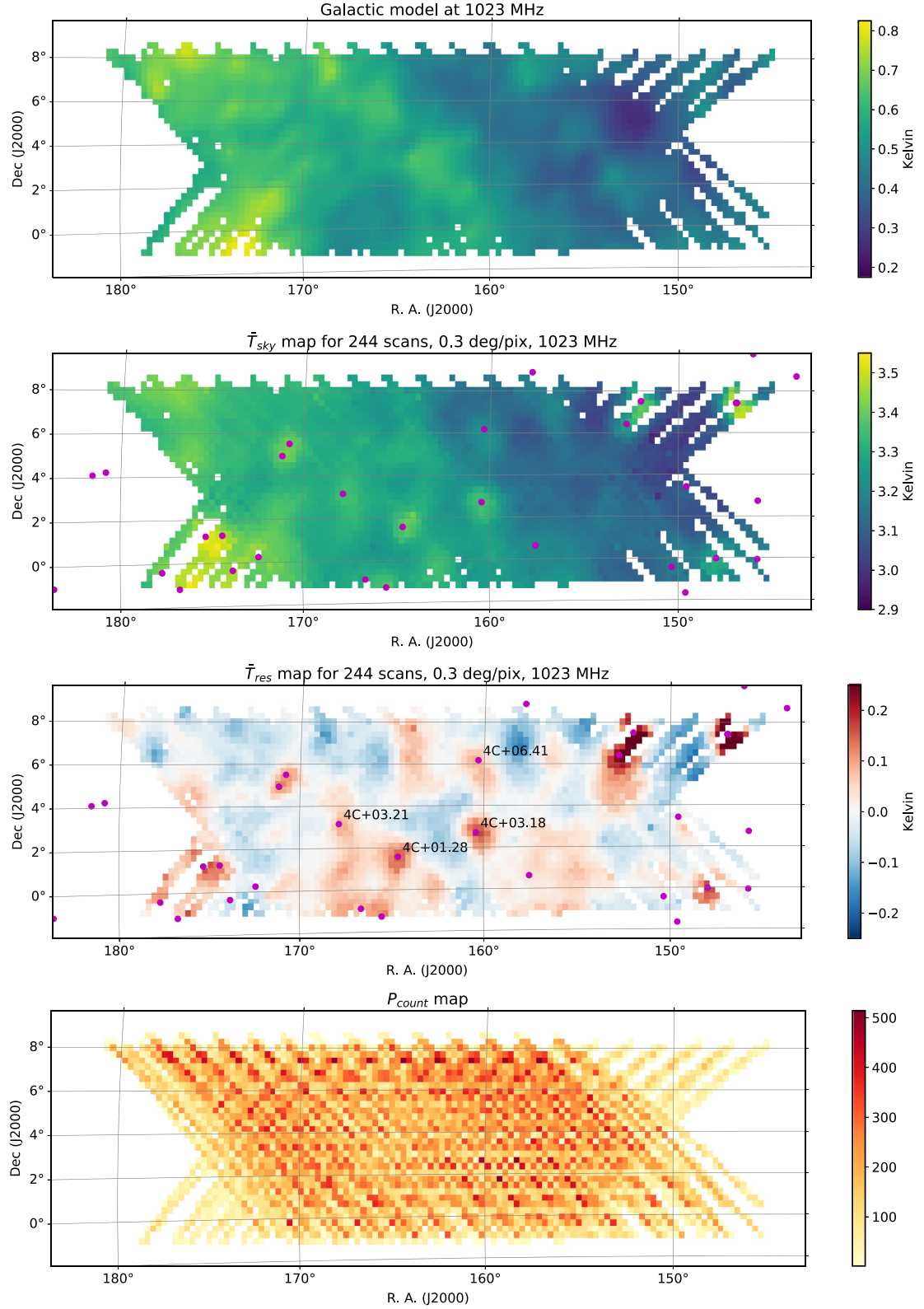


Figure 15. The mean intensity maps at 1023 MHz from the combination of all scans after Round 3 RFI subtraction. The upper panel shows the model of diffuse Galactic emission derived from PySM with the same sky coverage as our data. The middle panels show the sky map and residuals that we obtained. The lower panel shows the total number of time samples (across all dishes and observations) that were combined in each pixel. Magenta dots mark the positions of point sources with flux >1 Jy at 1.4 GHz, which were not included in the sky model used for calibration.

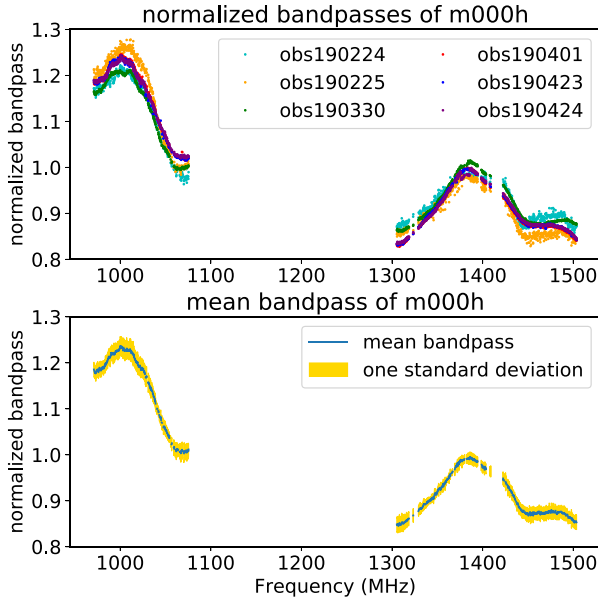


Figure 16. Upper panel: The normalized bandpasses of the m000h receiver for six observations, taken over a period of several months. Lower panel: The mean bandpass for all observations. The estimated standard deviation is a few per cent or less across the entire band.

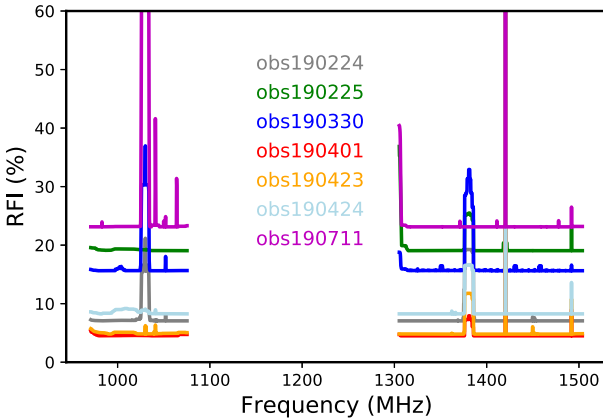


Figure 17. The Round 1 RFI percentage flagged in the data as a function of frequency for the low and high bands of interest. Different colours label the different observations.

Table 3. Percentage of discarded data, out of the initial ones, in each observation for both the lower and higher frequency bands. The three RFI flagging rounds – i.e. strong RFI flagging (R1), channel filtering (R2), and map-based weak RFI identification (R3) – are reported separately to show the progressive identification of contamination in the data.

Block ID	Lower band (per cent)			Higher band (per cent)		
	R1	R2	R3	R1	R2	R3
obs190224	7.9	23.3	56.5	7.9	13.3	80.7
obs190225	19.2	24.5	33.0	19.8	22.9	56.9
obs190330	18.1	19.0	30.2	18.0	18.7	99.9
obs190401	4.6	8.6	81.7	4.9	5.2	60.4
obs190423	6.5	29.4	71.5	6.8	31.8	91.6
obs190424	10.3	13.1	72.5	10.6	10.9	94.5
obs190711	26.8	30.8	32.1	23.4	28.8	39.4

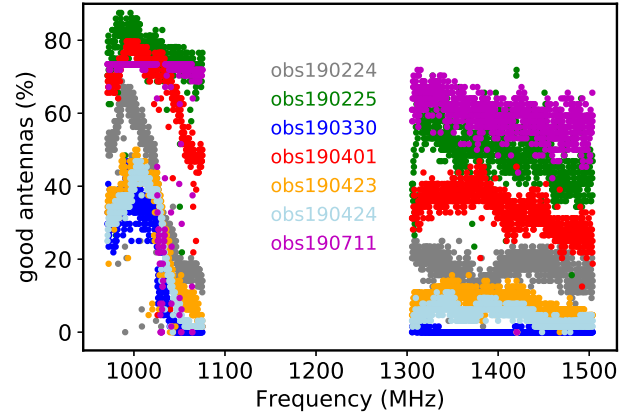


Figure 18. The percentage of retained antennas (out of 64) at the end of Round 3 as a function of frequency for the low and high bands of interest. As in Fig. 17, the different observations are labelled with different colours.

In reality, the receiver temperature T_{rec} can also depend on time as discussed before in the calibration process. However, in this section we only consider a single observational block and data from a small enough region of the sky so as to assume that T_{rec} only varies across frequency.

To provide an absolute calibration of the instrument, where each receiver measurement has no zero-level offset, the receiver and elevation-dependent temperatures would need to be known to a high level of accuracy. This is not necessary for the purpose of H I IM, since we are more concerned with fluctuations, but can be relevant for instance to make maps of the Galactic synchrotron.

The method employed by the MeerKAT single dish reduction pipeline to obtain T_{rec} uses a model for diffuse Galactic emission. This section provides a verification of the pipeline receiver temperatures using a method that does not require a Galactic emission model. Both the pipeline method and this ‘model-free’ method require the subtraction of the ground spillover and atmospheric emission contributions [$T_{\text{el}}(t, \nu)$] and the CMB monopole ($\bar{T}_{\text{CMB}}$) from the full system temperature. Both methods also require RFI contaminated data to have been flagged and removed.

The model-free method is based on the use of temperature–temperature (T – T) plots. T – T plots only require two different frequency observations of the same region of sky; if the dominant emission within said region is correlated across frequency, then a linear regression across the region pixels will display a clear straight-line fit. The gradient of the fit will be dependent on the emission spectral index and the y-intercept will depend on any additive offsets present in each of the two observations (Wehus et al. 2017). As T_{rec} is assumed not to change with location on the sky for this section, it is an additive offset that can be determined using the T – T plot method.

After the estimated ground spillover, atmospheric emission and CMB monopole are removed from the total calibrated temperature and RFI signals are masked out, the remaining contributions to the measured temperature are

$$T'_{\text{cal}}(t, \nu) = T_{\text{rec}}(\nu) + T_{\text{sync}}(t, \nu) + T_{\text{ueg}}(\nu), \quad (26)$$

where $T_{\text{ueg}}(\nu)$ is the unresolved extragalactic point source offset level. We are assuming that the diffuse Galactic emission is purely synchrotron at these frequencies and Galactic latitudes. $T_{\text{ueg}}(\nu)$ is assumed to be constant over all pixels but changes over frequency following a power law with a spectral index of around -2.7 (Battye et al. 2013). In reality, the unresolved point sources will also provide clustering and Poisson contributions to the total temperature that,

unlike the mean temperature contribution, will vary across pixels. These contributions, however, will be minor with respect to T_{sync} .

If we consider two frequencies, ν_1 and ν_2 , the diffuse Galactic synchrotron emission within those two frequency channels is related by a power law:

$$T_{\text{sync}}(t, \nu_2) = T_{\text{sync}}(t, \nu_1) \left(\frac{\nu_2}{\nu_1} \right)^{\beta_{\text{sy}}}, \quad (27)$$

meaning that

$$\begin{aligned} T'_{\text{cal}}(t, \nu_2) &= T_{\text{rec}}(\nu_2) + T_{\text{sync}}(t, \nu_1) \left(\frac{\nu_2}{\nu_1} \right)^{\beta_{\text{sy}}} + T_{\text{ueg}}(\nu_2) \\ &= T_{\text{rec}}(\nu_2) + (T'_{\text{cal}}(t, \nu_1) - T_{\text{rec}}(\nu_1) - T_{\text{ueg}}(\nu_1)) \left(\frac{\nu_2}{\nu_1} \right)^{\beta_{\text{sy}}} \\ &\quad + T_{\text{ueg}}(\nu_2). \end{aligned}$$

If the difference in frequency between ν_2 and ν_1 is less than a few MHz, then we can assume that $T_{\text{rec}}(\nu_2) = T_{\text{rec}}(\nu_1)$. Setting $m = (\nu_2/\nu_1)^{\beta_{\text{sy}}}$ and rephrasing the equation above as a linear regression between the total intensity maps yield

$$T'_{\text{cal}}(t, \nu_2) = m T'_{\text{cal}}(t, \nu_1) - m (T_{\text{rec}}(\nu_1) + T_{\text{ueg}}(\nu_1)) + T_{\text{rec}}(\nu_1) + T_{\text{ueg}}(\nu_2).$$

Plotting $T'_{\text{cal}}(t, \nu_2)$ against $T'_{\text{cal}}(t, \nu_1)$ for several times will produce a straight-line graph of gradient m with a y-intercept (c) of

$$c = -m (T_{\text{rec}}(\nu_1) + T_{\text{ueg}}(\nu_1)) + T_{\text{rec}}(\nu_1) + T_{\text{ueg}}(\nu_2). \quad (28)$$

Therefore, T_{rec} can be obtained from the fitted y-intercept and gradient as follows:

$$T_{\text{rec}}(\nu_1) = \frac{c - T_{\text{ueg}}(\nu_2) + m T_{\text{ueg}}(\nu_1)}{(1 - m)}. \quad (29)$$

To summarize, we assume that: (1) the receiver temperature is constant across the patch of the sky we are analysing (and the same for nearby frequencies); (2) we know the unresolved extragalactic point source offset level up to a certain accuracy (see below), and (3) the time changing quantity is related across frequencies by a spectral index (e.g. the Galactic synchrotron). Deviations from these assumptions will show up as errors in the linear fit. Moreover, any constant offset such as unmodelled ground pickup will also be absorbed into T_{rec} as before.

For the unresolved extragalactic point source offset level, we use the 1.278-GHz source counts under 12 mJy from Mauch et al. (2020) and the recipe provided in Battye et al. (2013) to calculate the point source offset temperature at 1.278 GHz. We then scale this value of 0.22 K at 1.278 GHz across our frequency range using a spectral index of -2.7 . Allowing for 10 per cent errors on the offset temperature at 1.278 GHz and a possible spectral index range of (-2.5) – (-2.9) leads to a 15 per cent error budget for the unresolved extragalactic point source contributions. As T_{rec} is ~ 6 – 10 K, our estimate of the unresolved point source level affects the T_{rec} measurements at the 1 per cent level.

The MeerKAT channels are separated by roughly 0.2 MHz. We calculate T_{rec} using the T–T plot method between channels in intervals of 5, thus providing a separation of 1 MHz. To prevent anomalous measurements skewing the receiver temperature calculation, we consider several pairs of channels at one time. In practice, this is achieved by selecting 10 adjacent frequencies and performing linear regression between pairs of frequencies separated by 1 MHz, e.g. frequency channel 1 and 6 out of the 10, frequency channel 2 and 7, frequency channel 3 and 8, and so on. This yields five estimates of T_{rec} for those 10 adjacent frequencies, the median of which is used as the T_{rec} at middle frequency (channel 5).

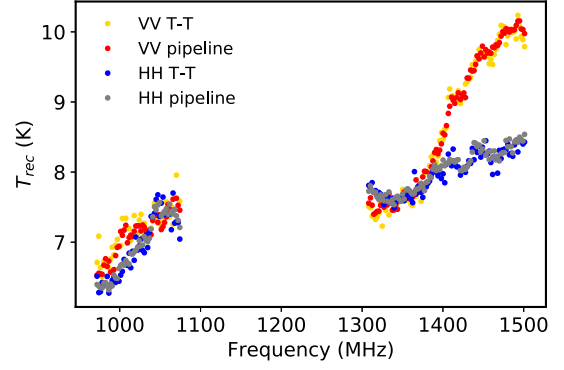


Figure 19. T_{rec} as calculated by the T–T plot and pipeline method for m000h and m000v of obs190225. Note that any unmodelled constant offset in time would be absorbed into T_{rec} (e.g. unmodelled ground pickup).

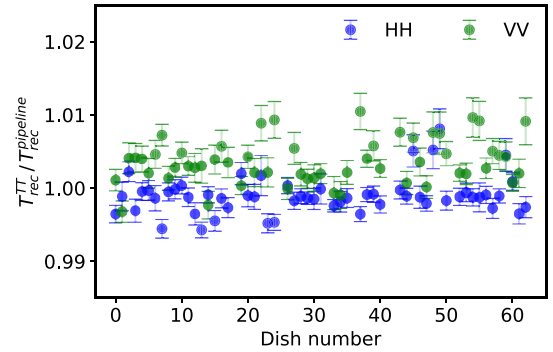


Figure 20. Mean ratio of T_{rec} as calculated by the T–T plot and pipeline method for all the receivers observing during a particular scan (obs190225). Error bars represent the standard deviation on the mean.

We choose to perform the linear regression across a small region of the sky within the observed WiggleZ field: $155^\circ < \text{ra} < 157^\circ$ and $2^\circ < \text{dec} < 4^\circ$. As the synchrotron spectral index is known to vary across the sky, the point of choosing a small region is to keep these variations as low as possible. The region selected is also free from notable point source emission, so as to specifically target diffuse synchrotron emission for the linear regression. Fig. 19 shows T_{rec} for m000h and m000v, during a particular observation block (obs190225) as calculated by both the T–T plot (model-free) method and the pipeline method. For the pipeline, since T_{rec} changes (slowly) with time, we choose the times corresponding to the same pixels on the sky we are using and average the values. The two methods can be seen to be in close agreement with each other. Both methods also show large variations in T_{rec} over frequency.

In Fig. 20, we plot the mean ratio of T_{rec} as calculated by the T–T plot (model-free) and pipeline method for all the receivers observing during obs190225. The difference between these two methods is seen to be consistently below the 1 per cent level.

5.4 Stability of the noise diode

The noise diodes, as the TOD calibrator, need to be a stable source during the observation time-scale. In the calibration, we calibrate the power of the noise diode twice, once in the beginning (track-I) and another at the end (track-II). Here, we calculate the percentage

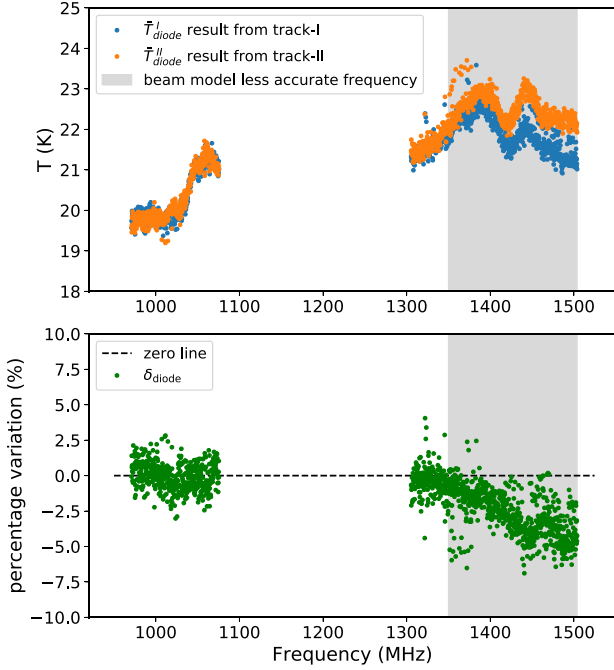


Figure 21. The $\bar{T}_{\text{diode}}$ results against frequency and the corresponding percentage variation (equation 30), for dish m000h of obs190225 as an example. The shaded area marks the frequency range where the beam model is reported to be less accurate.

variation of $\bar{T}_{\text{diode}}$ as

$$\delta_{\text{diode}}(\nu) = \frac{\bar{T}_{\text{diode}}^{\text{I}}(\nu) - \bar{T}_{\text{diode}}^{\text{II}}(\nu)}{\frac{1}{2}(\bar{T}_{\text{diode}}^{\text{I}}(\nu) + \bar{T}_{\text{diode}}^{\text{II}}(\nu))}. \quad (30)$$

In Fig. 21, we show the $\delta_{\text{diode}}(\nu)$ result for m000h of obs190225. The lower frequency band shows $\delta_{\text{diode}} \lesssim 2.5$ per cent, which will be reduced when we calculate the temperature map from HH and VV data and calculate the final map by averaging over 100 scans. The higher frequency band shows a clear bias above 1350 MHz, which may be caused by two factors: (1) As shown in Fig. 14, the RFI pollution is more significant at higher frequency, while the calibrator source is also close to the equator during the track observation; (2) as mentioned in Asad et al. (2019), the beam model is less accurate at higher frequencies (> 1350 MHz), which may cause errors when calibrating the power of the diode noise from the calibrator point source. Further insight into this problem will be provided by an upcoming observation that is far from the equator. In any case, our focus is mostly in the lower frequency band given the cosmology goals.

5.5 Residuals of the calibration

In Section 3, we presented a model (equation 1) to calibrate the sky. From this pipeline, we are left with calibration residuals $T_{\text{res}}(t, \nu_i)$ (equation 4) for both HH and VV polarizations. If the model was a complete description of the observation, such residuals should behave as Gaussian random noise centred at zero. We already know that is not the case, as we did not include the point sources in the scan. However, note that any offset leftover in the scan will be absorbed into the receiver temperature, so that the residuals will still be centred at zero. Any skewness in the histogram would be an indication of the presence of contaminants, while the variance gives an estimate of the level of accuracy of our model.

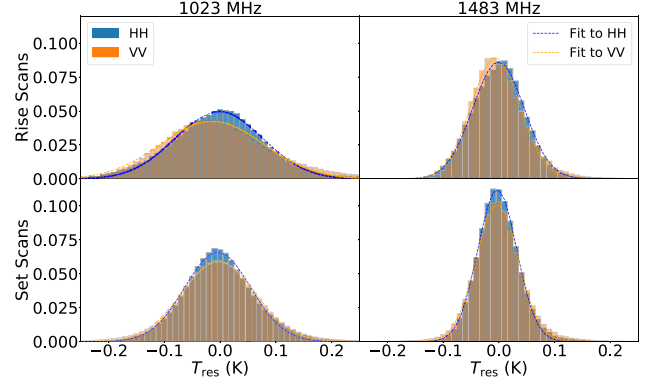


Figure 22. Histograms of level-3 temperature residuals $T_{\text{res}}(t, \nu)$ (equation 4) for both polarizations. Each panel corresponds to a different combination of channel and scan type. The top row corresponds to rising scans while the bottom row to setting scans. Left column exemplifies a lower frequency while the right column a higher frequency of the total range.

In Fig. 22, we present normalized histograms of the residuals for the different polarizations using two example frequency channels and divided into ‘rising’ and ‘setting’ scans. The set scans look well behaved; i.e. both frequency channels have the distribution of the residuals centred near zero and reasonably described by Gaussian curves. The standard deviations of the setting scans are $\sigma_{1023}^{\text{HH}} = 0.06$ K, $\sigma_{1023}^{\text{VV}} = 0.07$ K, $\sigma_{1483}^{\text{HH}} = 0.04$ K, $\sigma_{1483}^{\text{VV}} = 0.05$ K. This is still above the thermal noise, but corresponds to about a 0.3 per cent error in our model, since our T_{sys} is about 16 K. In reality, we have some outliers in the residuals that go up to 0.3 K (see Fig. 15) that is still less than a 2 per cent correction to our model.

The rising scans show an interesting skewness in the distribution of the residuals for the VV polarization. This might indicate the presence of residual RFI in the VV polarization due to some residual omni-directional (vertical) transmitting antennas from terrestrial systems. Geographically speaking, the setting scans look north-west while the rising scans look north-east. As the effect was present in both rising observational nights, it is likely that a faint terrestrial vertically polarized source (GSM cell phone antennas, VHF radio, UHF TV signals) is present in the north-east direction from the MeerKAT site (there is indeed a communication tower 160 km away from the core of the MeerKAT in that direction). This possible RFI is, however, quite low and will get diluted in the final averaged maps. The main point is that corrections to our model that is used for the calibration have rms values below the 1 per cent level that should translate to gain fluctuations at the same level.

5.6 Noise level estimation

It is also useful to measure the noise level from the final maps and check if they are consistent with the theoretical expectation (and dropping as square root of time). Unfortunately, as we have seen, the residuals still have some contamination, in particular from point sources (Fig. 15), so just calculating the variance of such maps would not work. However, using the final data cube, T_{res} , we can estimate the noise level using the difference between four neighbouring channels (so-called ABBA), as

$$\Delta T_i(\nu^*) = \frac{1}{2} (T_{\text{res}}^i(\nu) + T_{\text{res}}^i(\nu + \Delta\nu)) - \frac{1}{2} (T_{\text{res}}^i(\nu - \Delta\nu) + T_{\text{res}}^i(\nu + 2\Delta\nu)), \quad (31)$$

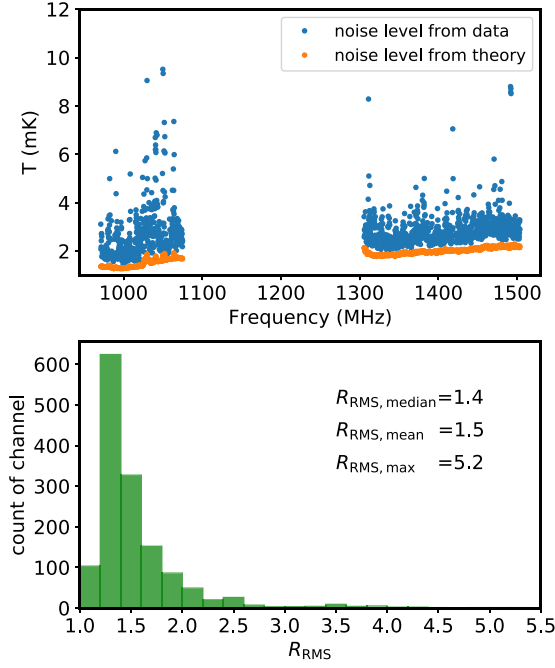


Figure 23. The comparison between noise levels derived from theory equation and from data (upper panel) and the histogram of their ratio $R_{\text{RMS}} = \Delta T_{\text{RMS}}(v^*)/\sigma_{\text{th}}(v^*)$ (lower panel).

where the index i goes over all pixels in the map and $*$ means the combined result from the four channels ($v - \Delta v$, v , $v + \Delta v$, and $v + 2\Delta v$). This will cancel any quantity that is linear in v across these channels and should therefore cancel most, if not all, of the sky contributions. If there was only white noise in the maps, the corresponding variance in each pixel would be

$$\sigma_{\text{th},i}^2(v^*) = \frac{1}{4} \left(\sigma_{\text{th},i}^2(v) + \sigma_{\text{th},i}^2(v + \Delta v) + \sigma_{\text{th},i}^2(v - \Delta v) + \sigma_{\text{th},i}^2(v + 2\Delta v) \right), \quad (32)$$

which would give the variance of a single map $\sigma_{\text{th},i}^2(v)$ if they were the same for the four frequencies. We can calculate this theoretical variance by starting with the standard radiometer equation (Wilson et al. 2009) to get the noise in the initial TOD:

$$\sigma_{\text{th}}^2(t, v) = \frac{T_{\text{sys}}^2(t, v)}{\Delta v \Delta t}, \quad (33)$$

where $\Delta v = 0.2$ MHz is the channel resolution and $\Delta t = 2$ s. For the system temperature, we can use the calibrated data itself for each polarization, $T_{\text{sys}} = T_{\text{cal,HH}}(t, v)$ or $T_{\text{cal,VV}}(t, v)$. In order to get the final data cube, $T_{\text{res}}^i(v)$, the data go through several stages of averaging. We propagate the variance taking into account this averaging in order to get to the final $\sigma_{\text{th},i}^2(v)$.

From the data, our best estimate for the noise variance is the rms over the $\Delta T_i(v^*)$ map. Since the noise is not necessarily uniform across the map, we take a weighted rms

$$\Delta T_{\text{RMS}}^2(v^*) = \frac{N_p}{N_p - 1} \left(\frac{\sum_i w_i \Delta T_i^2}{\sum_i w_i} - \left[\frac{\sum_i w_i \Delta T_i}{\sum_i w_i} \right]^2 \right), \quad (34)$$

where the sums are over the number of pixels in the map, N_p , and we have suppressed the dependence of w_i and ΔT_i on v^* . For the weight, w_i , we used the theoretical expected variance itself, $w_i(v^*) = 1/\sigma_{\text{th},i}^2(v^*)$ (but note that, as explained above, such

theoretical variance is calculated using T_{cal} as a proxy for the system temperature). The theoretical rms that can be compared to $\Delta T_{\text{RMS}}(v^*)$ in equation (34) is then simply

$$\sigma_{\text{th}}^2(v^*) = N_p \left(\sum_i \sigma_{\text{th},i}^{-2}(v^*) \right)^{-1}. \quad (35)$$

In Fig. 23, we show the noise level against frequency and the histogram of ratio $R_{\text{RMS}} = \Delta T_{\text{RMS}}(v^*)/\sigma_{\text{th}}(v^*)$. The median value of R_{RMS} is 1.4, while some channels are noisy and R_{RMS} can reach 5.2. We see that we are reaching a noise level of about 2 mK that is within expectations and promising for our cosmology goals (averaging 100 pixels in the cube would reach an rms of 0.2 mK that is at the level of the H I signal).

5.7 Per-dish maps

In Figs 24 and 25, we show the per-dish maps T_{sky} and T_{res} , respectively, with all available data from the seven observations. These maps show a good consistency between each other and have no visible striping artifacts. We noticed that the different dishes have different ‘clean’ scans left after the weak RFI flagging (Section 4.4). It shows a slight relation to the position of the dish in the MeerKAT array, which denotes that the weak RFI may come from the surrounding facilities or be affected by the terrain.

5.8 Eigendecomposition of calibrated maps

We expect the sky signal in the calibrated maps to be highly correlated across frequency. We can test this assumption by measuring the v , v' covariance of the data given by $\mathbf{C} = \mathbf{X}^T \mathbf{X} / (N_p - 1)$. Here, $\mathbf{X}$ represents the final calibrated maps $\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, v)$ unravelled into single vectors of length N_p pixels at each frequency. The matrix $\mathbf{X}$ therefore has dimensions $N_p \times N_p$. We use data in the frequency range $971 < v < 1025$ MHz least affected by RFI, mask point sources in each map, and then mean centre the data for each channel. The eigendecomposition is then given by $\mathbf{C}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues ordered by descending magnitude, and $\mathbf{V}$ the corresponding eigenvectors. Fig. 26 shows the eigenvalues for the calibrated maps. One aim in calibrating the telescope for H I IM is for the majority of foreground covariance to be contained in just a few dominant modes that can be removed to better isolate the underlying H I signal, which should have a smooth, flat eigenvalue spectrum since it is approximately Gaussian. The fact that the full calibrated data (blue solid line) do not reach a flattened plateau after the first few modes (which we expect to contain foregrounds) suggests that there are frequency-correlated systematics in the data that dominate over the cosmological H I signal.

Splitting the data into two subsections and subtracting one subsection map from the other should leave residuals that are absent of constant astrophysical signal and any systematics that are constant throughout the observation. This is shown by the orange dashed and green dash-dotted lines where we split by time block and observing dishes. As expected, these residuals reduce the most dominant mode significantly that is likely the Galactic synchrotron. However, there remain some reasonably correlated modes suggesting that the data still contain some time-varying and dish-varying systematics that are correlated in frequency. Similar tests were also presented in Switzer et al. (2015) (see their fig. 3). These results indicate hints of systematics in the calibrated data that could be due to instrumental effects such as polarization leakage, chromatic beam effects, calibration issues with the bandpass, $1/f$ noise, or residual

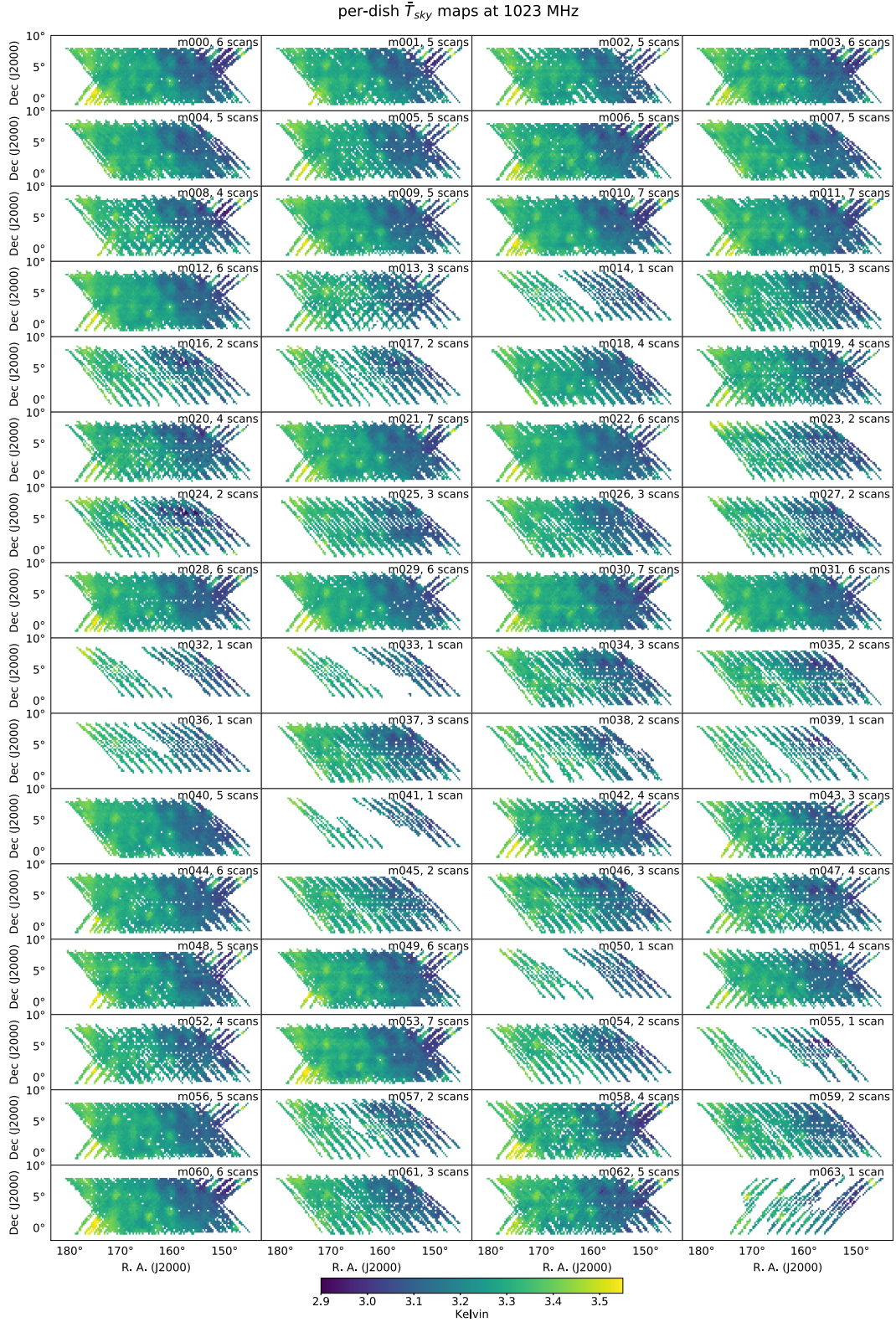


Figure 24. Per-dish $\bar{T}_{\text{sky}}$ maps at 1023 MHz.

RFI. Further investigation into characterizing these systematics will be the aim in future work. For instance, one way to mitigate the chromatic beam effects is to convolve the maps to a common resolution, something that was performed on GBT data. Also, this

analysis was done on calibrated maps, $\bar{T}_{\text{sky}}$, with models already subtracted, while for our future cosmological analysis we might work directly with the total calibrated map, $\bar{T}_{\text{cal}}$ and apply prior filtering on the TOD in order to deal with the $1/f$ noise.

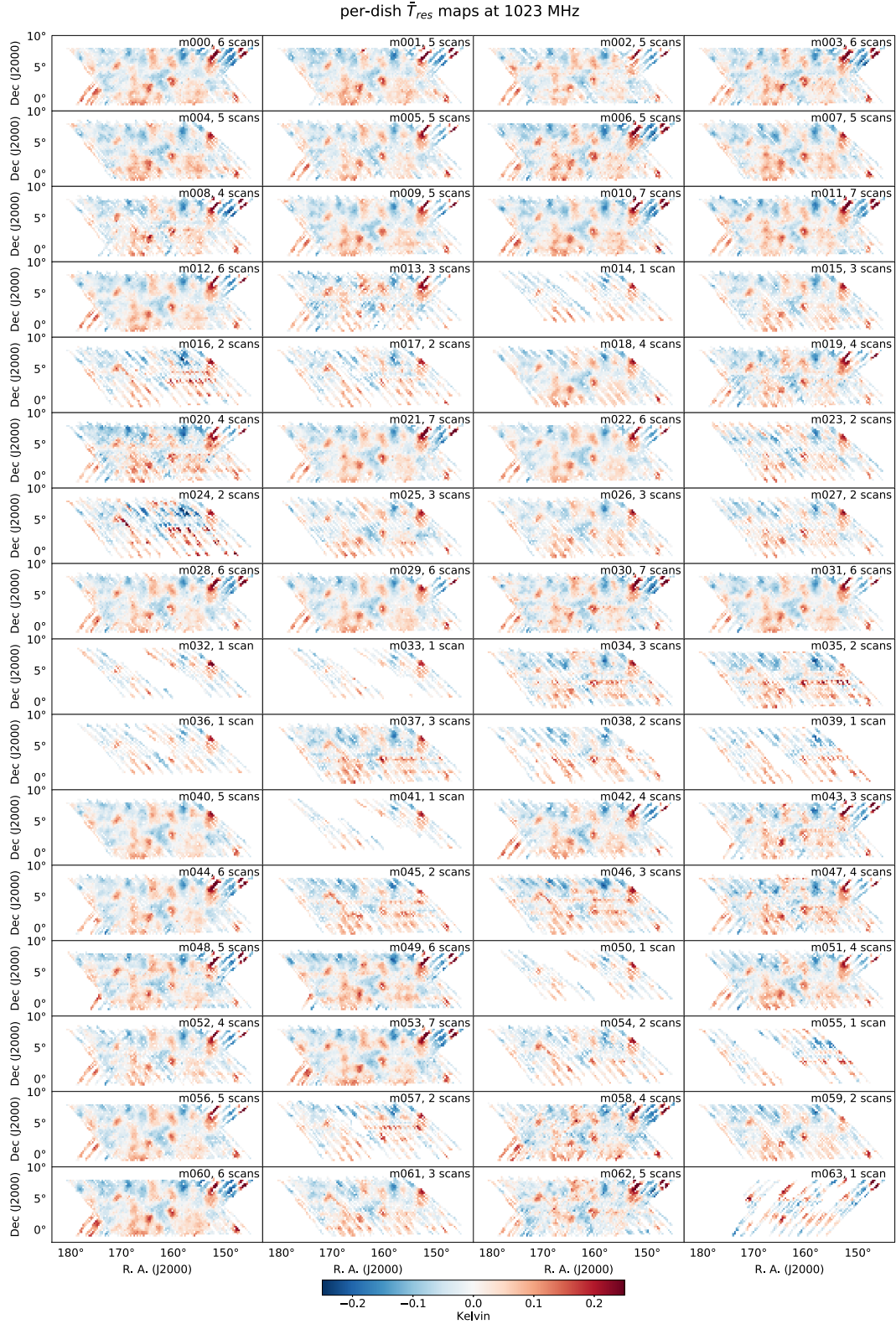


Figure 25. Per-dish $\bar{T}_{\text{res}}$ maps at 1023 MHz.

Fig. 27 shows the mean signal from the reconstructed first five eigenvectors of the v, v' covariance matrix (this can also be compared to fig. 4 in Switzer et al. 2015). The first (top) mode represents the synchrotron signal, which is expected to dominate, and the slope

seen in this mode is consistent with a synchrotron spectral index. An ideal instrument would produce a perfectly smooth slope in this first mode and allow a simple polynomial model to be used to remove this contribution from the data. However, the slight oscillations

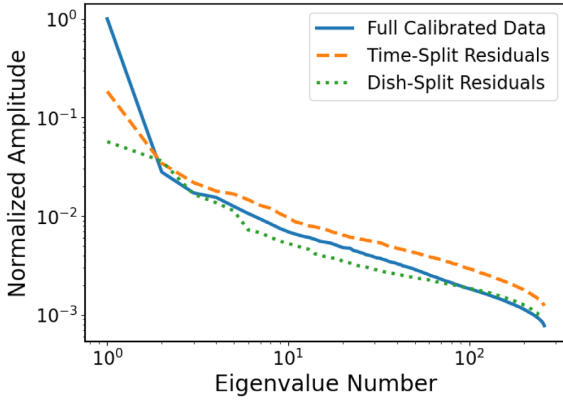


Figure 26. Principal components from the frequency–frequency covariance matrix for the final $\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu)$ calibrated maps, restricted to the $971 < \nu < 1025$ MHz frequency range least affected by RFI. Blue solid line shows the fully calibrated data from all dishes. Also shown are the principal components from the residuals of subtracting two subsections split by time blocks (orange dashed) and dishes (green dotted).

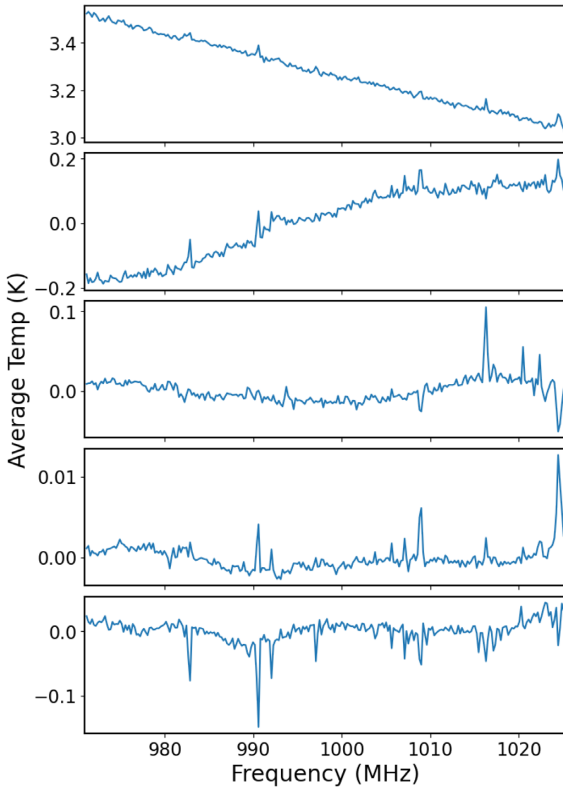


Figure 27. First five eigenvectors (descending in order of magnitude from top to bottom) from the frequency–frequency covariance matrix for the final $\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu)$ calibrated maps, restricted to the $971 < \nu < 1025$ MHz frequency range least affected by RFI.

through frequency motivate the requirement of a more sophisticated foreground removal technique for analysing cosmological H I. The remaining modes are likely caused by instrumental response systematics as discussed in the previous paragraph and identified by Fig. 26. In order to probe the cosmological H I, it is these modes, and further more dominant ones, that would need to be removed or mitigated before a detection is possible.

6 COMPARISON WITH PREVIOUS OBSERVATIONS

6.1 Recovery of diffuse Galactic radio emission

We aim to compare our calibrated sky data to realistic simulations of the diffuse Galactic radio emission, which are extrapolated from real data. Using the Global Sky Model (GSM; de Oliveira-Costa et al. 2008; Zheng et al. 2016), we generate realistic foreground maps at our frequencies of observation. The GSM is a model of diffuse Galactic radio emission, which uses 29 data sky maps to extrapolate this emission from 10 MHz to 5 THz. We select the frequency range of 971–1025 MHz due to containing the least RFI contamination, and generate accurate simulated foregrounds at each frequency channel of observation using the GSM. The output of the GSM extrapolation is a full sky map at a desired frequency, in HEALPIX¹⁸ (Górski et al. 2005; Zonca et al. 2019) format. Since we are only interested in the patch of sky targeted by our survey, we have to convert the GSM output into a flat-sky map that is in the same coordinates as our data. To do this, we first take the angular coordinate of each pixel in our MeerKAT map and then match it to the pixel in the HEALPIX map with the closest angular coordinate. While this is an approximation, we find that it is sufficient in our case since the angular size of the map is not too large, and we recover similar structure in the foreground simulation as is present in the MeerKAT data. In addition, we also mean centre each frequency slice of the MeerKAT and foreground simulation maps, by calculating the mean temperature in each frequency slice and subtracting that mean from all the pixels in that particular slice. Finally, in order to make the simulated foreground maps match the MeerKAT data in resolution, we convolve the foreground simulation maps with a frequency-dependent beam calculated according to the fitted 2D Gaussian beam model (Section 3.4). We also mask point sources in the MeerKAT data, since these are removed in the GSM simulations. The comparison is done using the MeerKAT sky temperature data, equation (22), averaged over all scans and dishes, $\bar{T}_{\text{sky}}(\text{ra}_{\text{pix}}, \text{dec}_{\text{pix}}, \nu)$.

Fig. 28 shows a comparison between the MeerKAT calibrated sky map data at 1023 MHz, and the simulated GSM at the same frequency. It is clear from a visual inspection of these maps that similar structure is present in both. In order to check their agreement more quantitatively, we calculate the pixel correlation between the MeerKAT data cube and the GSM data cube in the frequency range 971–1025 MHz. Since this involved a large amount of data points, we show our results in the form of a 2D histogram in Fig. 29, where the darkest colours indicate the highest concentration of data points. The Pearson correlation coefficient is found to be 0.94, indicating a strong correlation between our data and simulation. In addition, the cluster of points follows the $x = y$ line to a good degree. This demonstrates that the calibration pipeline yields a result that is highly correlated with a realistic simulation of the diffuse Galactic radio emission, both visually and quantitatively.

6.2 Recovery of the Galactic H I map

Another interesting test is to compare our own H I Galactic map to the literature. The 21-cm emission of the Milky Way has different structures to the Galactic synchrotron emission due to their different radiation mechanisms. In Fig. 30, we show the normalized time-

¹⁸<https://healpix.sourceforge.io/>

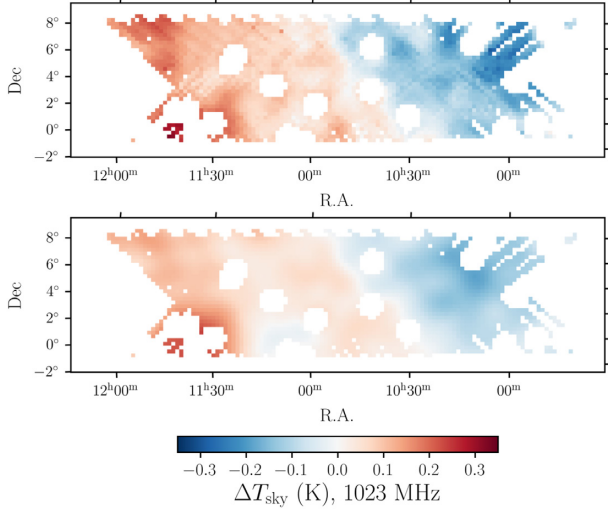


Figure 28. *Top:* MeerKAT sky data at 1023 MHz, mean centred such that the colourbar shows ΔT_{sky} , the deviations from the mean map temperature in K. *Bottom:* The same, but for the foreground simulation generated using GSM at the same frequency and angular coordinates. Point sources with flux > 1 Jy at 1.4 GHz and neighbouring pixels within 1 deg are masked on both maps.

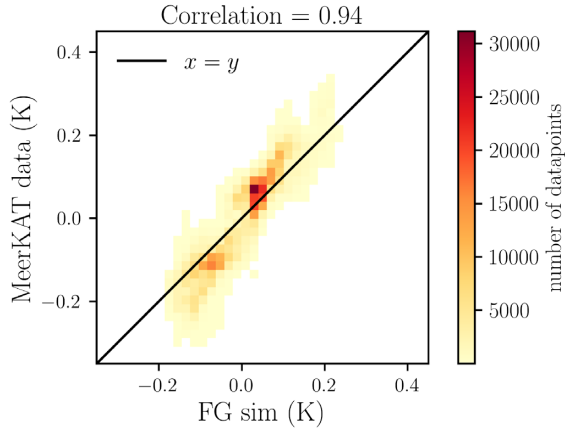


Figure 29. Correlation between the MeerKAT sky data (all dishes and scans) and the simulated foreground temperature fluctuations in the frequency range 971–1025 MHz. Darker colours indicate a higher concentration of data points, and the Pearson correlation coefficient between the two data sets is found to be 0.94.

averaged raw signals of seven observations across frequency. We see that the Galactic 21-cm lines, which correspond to a frequency of 1420.4 MHz, are extended across several frequency channels due to the Doppler shifts. We start by selecting three frequency channels corresponding to $\nu_{\text{GHI}} \in [1420.3, 1420.7]$ MHz, to extract the Galactic H I emission. The channels ch_{GHI} are flagged before our calibration, so we interpolate the gain and T_{cal} from our previous results to these three channels so that

$$T_{\text{GHI}}(t, \nu_{\text{GHI}}) = \hat{T}_{\text{raw}}(t, \nu_{\text{GHI}}) / g(t, \nu_{\text{GHI}}) - T_{\text{cal}}(t, \nu_{\text{GHI}}). \quad (36)$$

This calculation is applied on the RFI flagged HH and VV data. The final Galactic H I intensity map is shown in Fig. 31. We overlay the contours of the HI4PI H I column density map (HI4PI Collaboration 2016), which show a good agreement to our map.

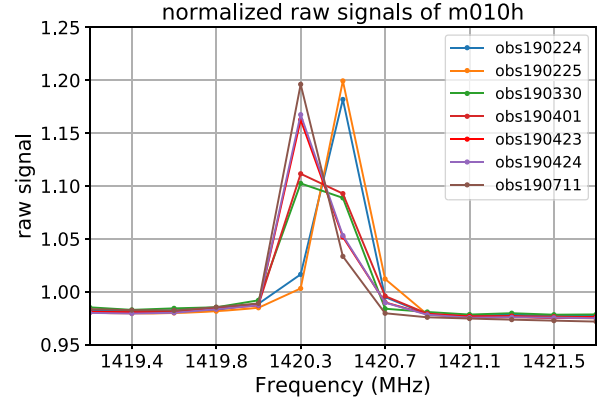


Figure 30. The time-averaged raw signals of seven observations, which are normalized by their mean values. The frequency is in the geocentric rest frame and the Galactic 21-cm lines are extended across several frequency channels due to the time- and pointing-related Doppler shifts.

6.3 Recovered point source fluxes versus catalogue values

We select four radio point sources in the WiggleZ 11-h field, the position of which are marked on the $\bar{T}_{\text{res}}$ map in Fig. 15, and extract their fluxes from the final $\bar{T}_{\text{res}}$ data cube in order to compare to the literature. One issue is that we are scanning over a large area and therefore the signal to noise on each of the point sources is relatively low. To compensate for that, we average the flux over 11 channels $[\text{ch} - 5, \text{ch} + 5]$ and combine the nearby pixels around the point sources ($r_{\text{cut}} = 0.5^\circ$) with appropriate weights.

The frequency-averaged flux on a given pixel (x, y) is

$$\bar{S}(x, y, \text{ch}) = 2k_B \frac{\bar{T}_{\text{res}}(x, y, \text{ch})}{A_{\text{eff}}(\text{ch}) P_B(x, y, \text{ch})} \times 10^{26} \text{ Jy}, \quad (37)$$

where $\bar{T}_{\text{res}}(x, y, \text{ch})$ is the weighted channel mean temperature (see Appendix A1), A_{eff} is the effective area, and P_B is the normalized power pattern amplitude at pixel (x, y) that is displaced from the source being studied with an angle $\theta(x, y)$. Here, we use the fitted 2D Gaussian beam model (Section 3.4)

$$P_B(x, y, \text{ch}) = \exp \left(-\frac{\theta^2(x, y)}{2\sigma_{\text{Beam}}^2} \right). \quad (38)$$

In Fig. 32, we show the $\bar{T}_{\text{res}}(x, y, \text{ch})$ map at a given frequency for the four point sources with the $r_{\text{cut}} = 0.5^\circ$ circle indicated as a white dashed line. In Fig. 33, we show $\bar{S}(x, y, \text{ch})$ against $\theta(x, y)$ for the four point sources. The plot shows the $\theta(x, y)$ range $0-1^\circ$, while in the calculation we only use data with $\theta < 0.5^\circ$. Including more pixels increases the chance of adding contamination from other sources.

We then average $\bar{S}(x, y, \text{ch})$ over the selected pixels with appropriate weights to obtain the mean spectral fluxes and standard errors. See Appendix A2 for more details. The extracted fluxes and the spectrum calculated from Parkes data are compared in Fig. 34. In order to access the quality of the fit, we calculate the normalized root mean square error (NRMSE) for the full band as the percentage error:

$$\text{NRMSE} = \sqrt{\frac{1}{N_{\text{ch}}} \sum_{\text{ch}} \left(\frac{\bar{S}(\text{ch}) - S_{\text{ref}}(\text{ch})}{S_{\text{ref}}(\text{ch})} \right)^2}, \quad (39)$$

where S_{ref} is the reference flux derived from Parkes observation data between 408 MHz and 1.4 GHz (or 2.7 GHz if 1.4 GHz data are

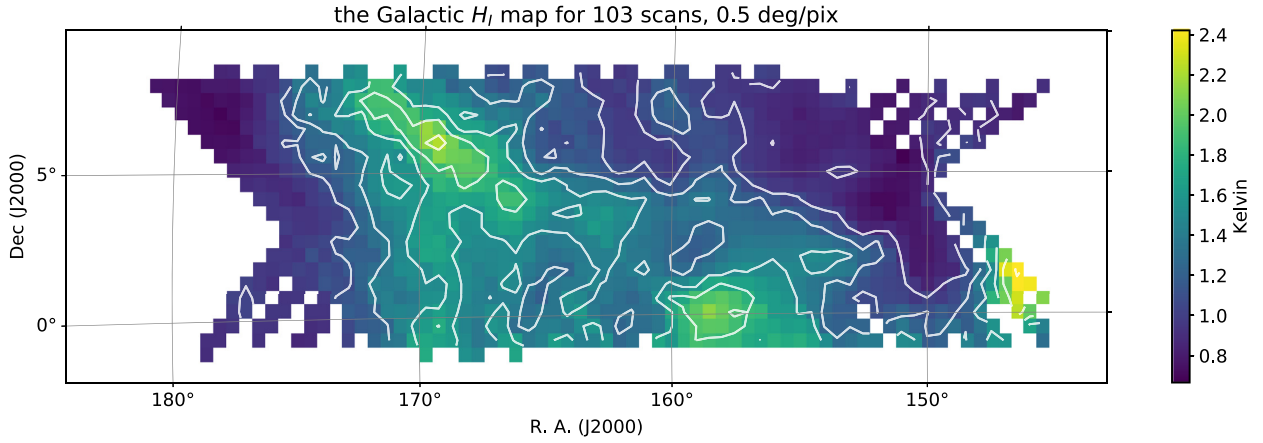


Figure 31. The Galactic H I intensity map that is extracted from frequency band $\nu_{\text{GHI}} \in [1420.3, 1420.7]$ MHz. The white contours are derived from the HI4PI H I column density map (HI4PI Collaboration 2016), which is constructed from the Effelsberg–Bonn H I Survey.

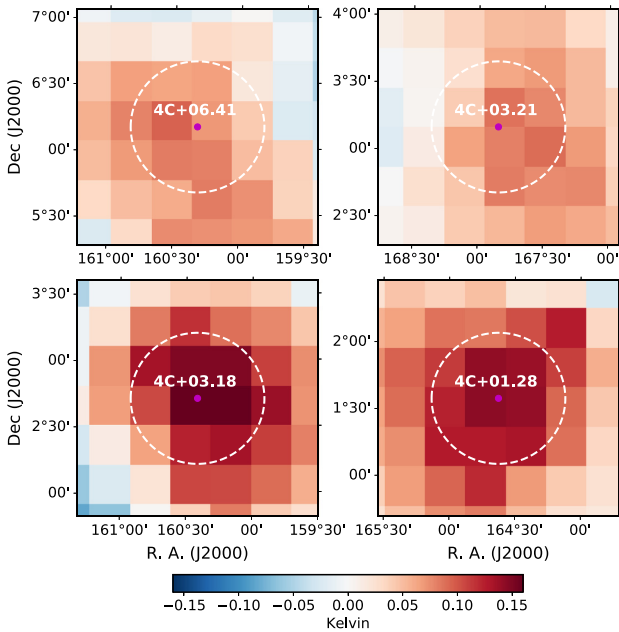


Figure 32. $\bar{T}_{\text{res}}(x, y, \text{ch})$ maps for the four point sources in a binned channel with the centre channel 800. The white dash circles label the positions of $r_{\text{cut}} = 0.5^\circ$.

absent), $\text{ch} \in [550, 600, \dots, 1050]$ and $[2150, 2200, \dots, 3100]$, and $N_{\text{ch}} = 31$ is the total number of fluxes measured across frequency.

The NRMSE values are listed in Fig. 34, which shows strong correlations to the source types and also to the brightness, pixel count, and declination:

(i) 4C + 06.41 and 4C + 01.28 are BL Lac objects (Plotkin et al. 2008; Lefaucheur & Pita 2017). Monitoring by the OVRO 40-m telescope shows that these sources are significantly variable at 15 GHz over decade time-scales (Richards et al. 2011) and we interpret variability to be the cause of the discrepancy between the measured flux spectra and the Parkes predictions.

(ii) 4C + 03.21 is the only one that has no redshift record and is labelled as ‘blank field’ in Parkes Catalog (1990, version 1.01) in these four point sources. The Parkes data at 408 MHz, 1.41 GHz, and 2.7 GHz are not consistent with a single power-law spectrum, so

it is not surprising that the MeerKAT data have a relatively poor fit to the 408 MHz to 1.41 GHz spectral prediction (NRMSE = 0.197).

(iii) 4C + 03.18 is identified as a radio galaxy and its flux might be expected to be stable at low radio frequency. We find an excellent agreement with the Parkes prediction with NRMSE = 0.036.

Our flux estimation also has some other potential risks, such as the observations being done in scanning mode rather than tracking and the $\bar{T}_{\text{res}}$ maps possibly containing Galactic structures that are unmodelled in PySM.

7 CONCLUSIONS

We presented a full data reduction pipeline for single dish data obtained from a pilot survey through the MeerKAT open time call. The observation was done in the *L* band (856–1712 MHz, 4096 channels) and targeted a single patch of about 200 deg² with the aim of testing the single-dish H I IM technique. The scan patch covered the 11-h field of the WiggleZ Dark Energy Survey, which is a large-scale spectroscopic survey of emission-line galaxies selected from UV and optical imaging. The total (usable) observation time was about 10.5 h, which, with about 60 dishes on average, corresponds to 630 h effectively (before any flagging).

The data reduction pipeline for the TOD includes steps for flagging of human-made RFI, bandpass, and absolute calibration using known point sources, and calibration of receiver gain fluctuations based on interleaved signal injection from a noise diode. Fig. 4 shows the flowchart of the pipeline.

We construct a model for all the components that will contribute to the total signal, which includes the calibrator point sources, celestial diffuse components, elevation-related emission (atmospheric emission and ground pickup), noise diode signals, and the receiver temperature. By comparing the model to the TOD using a prescribed likelihood with priors, we fit the free parameters and use the gain solution to obtain the calibrated temperature.

We show that the pipeline is sufficiently accurate to recover maps of diffuse celestial emission and point sources, showing a good correlation with Galactic synchrotron maps in the literature. We also extracted the Galactic H I emission, showing that the reconstructed map has an excellent agreement to the HI4PI H I column density map (HI4PI Collaboration 2016), which is constructed from the Effelsberg–Bonn H I Survey.

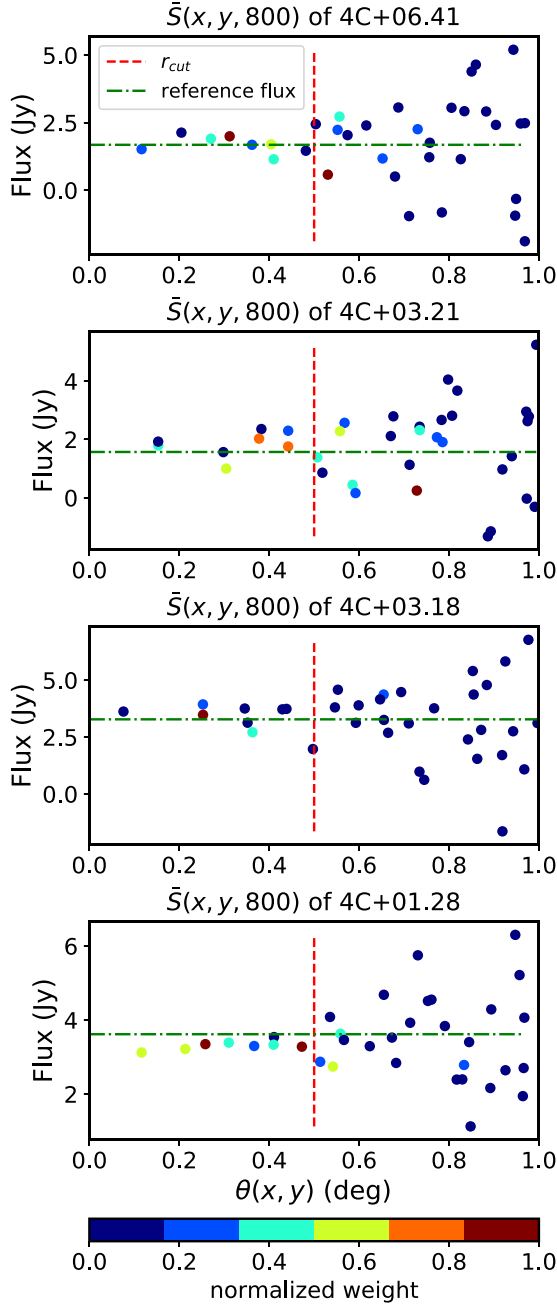


Figure 33. $\tilde{S}(x, y, \text{ch})$ against $\theta(x, y)$ for the four point sources in a binned channel with the centre channel 800 ($\nu = 1023$ MHz). The colours on data show the normalization of weight $w(x, y, \text{ch})$ (equation A3).

The reconstructed maps have no visible striping artefacts and a good level of consistency between per-dish maps and external data sets. We noticed that the different dishes have different ‘clean’ scans left after the three rounds of RFI flagging. This shows a slight relation to the position of the dish in the MeerKAT array, which denotes that the weak RFIs may come from the surrounding artificial facilities or be affected by the terrain.

The residual maps have rms amplitudes below 0.1 K, corresponding to < 1 per cent of the model temperature. This shows that the emission missing in our sky model, such as point sources, has a negligible impact on the calibration. We also estimate the noise level in the final data cube after averaging over all dishes and scans using

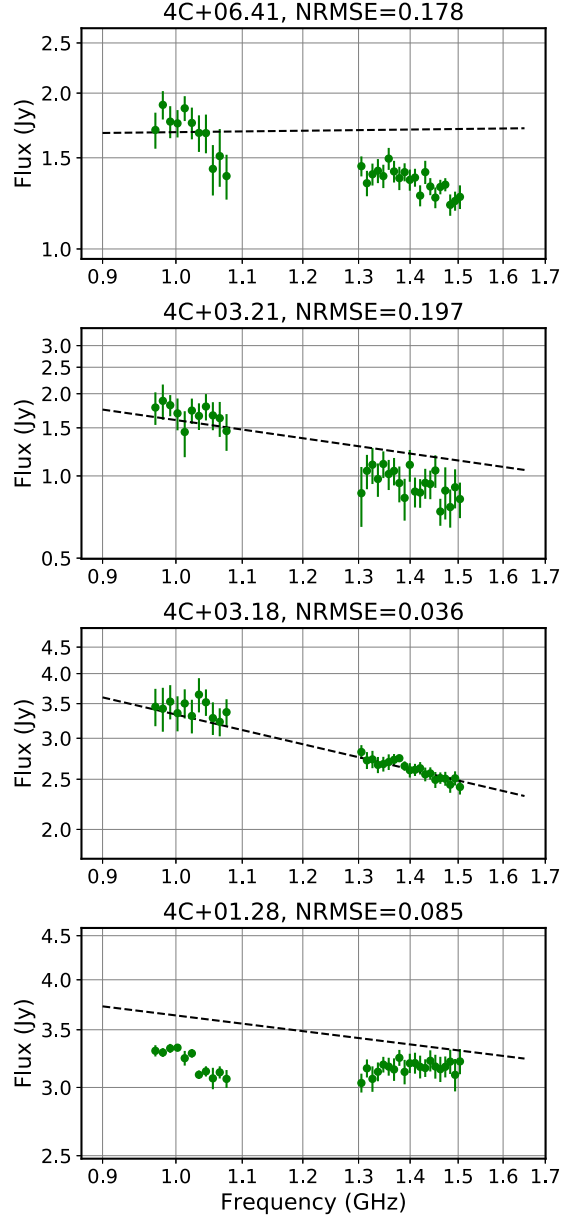


Figure 34. The extracted fluxes of point sources compared with the spectra calculated from the Parkes data (black dashed line).

the difference between four neighbouring channels. The rms is about 2 mK, with a median value of $1.4 \times$ the expected theoretical noise level, although there are deviations in some of the channels.

Finally, we select four radio point sources in the WiggleZ 11-h field and extract their fluxes from the total T_{res} data cube in order to compare to the Parkes results. The flux of the radio galaxy (4C + 03.18) is recovered within 3.6 per cent, while other point sources have reasonably larger errors: 17.8 per cent and 8.5 per cent for two BL Lac objects (4C + 06.41 and 4C + 01.28), and 19.7 per cent for the ‘blank field’ source (4C + 03.21). This also demonstrates that the autocorrelation can be successfully calibrated to give the zero-spacing flux and potentially help in the imaging of MeerKAT interferometric data (Rau, Naik & Braun 2019).

In conclusion, we have shown that it is possible to calibrate the single dish data from the MeerKAT interferometer with high accuracy using sky calibrators and noise diodes. No obvious show stoppers

were found although the RFI contamination remains challenging. This work opens the door to use MeerKAT and the future SKA to measure the H I IM signal and probe cosmology on degree scales and above. In a follow-up paper, we will be using these data to constrain the H I power spectrum and its cross-correlation with galaxy surveys.

ACKNOWLEDGEMENTS

We would like to thank Stuart Harper for several useful discussions during the development of this work. JW, MGS, MI, YL, and BE acknowledge support from the South African Radio Astronomy Observatory and National Research Foundation (grant number 84156). PB acknowledges support from Science and Technology Facilities Council grant ST/T000341/1. SC is supported by STFC grant ST/S000437/1. PS is supported by the STFC (grant number ST/P006760/1) through the DISCnet Centre for Doctoral Training. AP is a UK Research and Innovation Future Leaders Fellow, grant MR/S016066/1, and also acknowledges support by STFC grant ST/S000437/1. JF was supported by the University of Padova under the STARS Grants programme *CoGITO: Cosmology beyond Gaussianity, Inference, Theory and Observations* and by UK Science & Technology Facilities Council (STFC) Consolidated Grant ST/P000592/1. MS acknowledges funding from the INAF PRIN-SKA 2017 project 1.05.01.88.04 (FORECaST). GB acknowledges support from the Ministero degli Affari Esteri della Cooperazione Internazionale-Direzione Generale per la Promozione del Sistema Paese Progetto di Grande Rilevanza ZA18GR02 and the National Research Foundation of South Africa (grant number 113121) as part of the ISARP RADIOSKY2020 Joint Research Scheme. We acknowledge the use of the Ilifu cloud computing facility, through the Inter-University Institute for Data Intensive Astronomy (IDIA). The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation.

DATA AVAILABILITY

The analysis pipeline software is available on request. Data products are also available on request but are subject to an access policy within a 12-month proprietary period.

REFERENCES

Akeret J., Seehars S., Chang C., Monstein C., Amara A., Refregier A., 2017, *Astron. Comput.*, 18, 8
 Anderson C. J. et al., 2018, *MNRAS*, 476, 3382
 Ansari R. et al., 2012, *A&A*, 540, A129
 Asad K. M. B. et al., 2019, *MNRAS*, 502, 2970
 Astropy Collaboration, 2013, *A&A*, 558, A33
 Baan W., 2010, Proc. Sci., RFI Mitigation in Radio Astronomy. SISSA, Trieste, PoS#1
 Bagla J. S., Khandai N., Datta K. K., 2010, *MNRAS*, 407, 567
 Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, *MNRAS*, 434, 1239
 Bradley R. A., Srivastava S. S., 1979, *Am. Stat.*, 33, 11
 Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, *ApJ*, 803, 21
 Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, *Phys. Rev. Lett.*, 100, 091303
 Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, *Nature*, 466, 463
 Courvoisier T. J. L., 1998, *A&AR*, 9, 1

de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, *MNRAS*, 388, 247
 Doré O., Teyssier R., Bouchet F. R., Vibert D., Prunet S., 2001, *A&A*, 374, 358
 Drinkwater M. J. et al., 2010, *MNRAS*, 401, 1429
 Fixsen D. J., 2009, *ApJ*, 707, 916
 Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
 Harper S. E., Dickinson C., 2018, *MNRAS*, 479, 2024
 Harper S. E., Dickinson C., Battye R. A., Roychowdhury S., Browne I. W. A., Ma Y. Z., Olivari L. C., Chen T., 2018, *MNRAS*, 478, 2416
 HI4PI Collaboration, 2016, *A&A*, 594, A116
 Jonas J., MeerKAT Team, 2016, Proc. Sci., MeerKAT Science: On the Pathway to the SKA, PoS(MeerKAT2016). SISSA, Trieste, PoS#1
 Kovetz E. D. et al., 2017, preprint ([arXiv:1709.09066](https://arxiv.org/abs/1709.09066))
 Lefaucheur J., Pita S., 2017, *A&A*, 602, A86
 Li Y., Santos M. G., Grainge K., Harper S., Wang J., 2020, *MNRAS*, 501, 4344
 Li L.-C., Staveley-Smith L., Rhee J., 2021, *Res. Astron. Astrophys.*, 21, 030
 Lidz A., Furlanetto S. R., Oh S. P., Aguirre J., Chang T.-C., Doré O., Pritchard J. R., 2011, *ApJ*, 741, 70
 Loeb A., Wyithe J. S. B., 2008, *Phys. Rev. Lett.*, 100, 161301
 Mao Y., Tegmark M., McQuinn M., Zaldarriaga M., Zahn O., 2008, *Phys. Rev. D*, 78, 023529
 Masui K. W. et al., 2013, *ApJ*, 763, L20
 Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2020, preprint ([arXiv:2011.10815](https://arxiv.org/abs/2011.10815))
 Mauch T. et al., 2020, *ApJ*, 888, 61
 Natoli P., de Gasperi G., Gheller C., Vittorio N., 2001, *A&A*, 372, 346
 Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, *MNRAS*, 405, 155
 Perley R. A., Roser H.-J., Meisenheimer K., 1997, *A&A*, 328, 12
 Peterson J. B. et al., 2009, *Astro2010: The Astronomy and Astrophysics Decadal Survey*. p. 234
 Plotkin R. M., Anderson S. F., Hall P. B., Margon B., Voges W., Schneider D. P., Stinson G., York D. G., 2008, *AJ*, 135, 2453
 Pritchard J. R., Loeb A., 2008, *Phys. Rev. D*, 78, 103511
 Rau U., Naik N., Braun T., 2019, *AJ*, 158, 3
 Reid B. et al., 2016, *MNRAS*, 455, 1553
 Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M. A., Ghosh T., 2015, *MNRAS*, 451, 4311
 Richards J. L. et al., 2011, *ApJS*, 194, 29
 Santos M. et al., 2015, Proc. Sci., Advancing Astrophysics with the Square Kilometre Array, PoS(AASKA14). SISSA, Trieste, PoS#19
 Santos M. G. et al., 2017, preprint ([arXiv:1709.06099](https://arxiv.org/abs/1709.06099))
 Seo H.-J., Dodelson S., Marriner J., McGinnis D., Stebbins A., Stoughton C., Vallinotto A., 2010, *ApJ*, 721, 164
 Shacham M., Brauner N., 1997, *Ind. Eng. Chem. Res.*, 36, 4405
 Switzer E. et al., 2013, *MNRAS*, 434, L46
 Switzer E. R., Chang T.-C., Masui K. W., Pen U.-L., Voytek T. C., 2015, *ApJ*, 815, 51
 Thorne B., Dunkley J., Alonso D., Naess S., 2017, *MNRAS*, 469, 2821
 Tian G.-L., 1998, *Stat. Probab. Lett.*, 38, 289
 Wehus I. K. et al., 2017, *A&A*, 597, A131
 Wilson T. L., Rohlf K., Hüttemeister S., 2009, *Tools of Radio Astronomy*, 5th edn. Springer-Verlag, Berlin
 Wolz L. et al., 2017, *MNRAS*, 464, 4938
 Wolz L. et al., 2021, preprint ([arXiv:2102.04946](https://arxiv.org/abs/2102.04946))
 Wright A., Otrupcek R., 1990, PKSCAT90, HEASARC Online Data Catalog
 Wyithe J. S. B., Loeb A., 2008, *MNRAS*, 383, 606
 Wyithe J. S. B., Loeb A., Geil P. M., 2008, *MNRAS*, 383, 1195
 Zheng H. et al., 2016, *MNRAS*, 464, 3486
 Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *J. Open Source Softw.*, 4, 1298

APPENDIX A: WEIGHTED FLUX FOR POINT SOURCES

A1 Channel-weighted mean map

On combining $2n + 1$ channels, the channel mean temperature is

$$\bar{T}_{\text{res}}(x, y, \text{ch}) = \frac{1}{N(x, y, \text{ch})} \sum_{i=\text{ch}-n}^{\text{ch}+n} T_{\text{res}}(x, y, i) N_{\text{count}}(x, y, i), \quad (\text{A1})$$

where the number of scans and channels contributing to the channel-weighted mean map is

$$N(x, y, \text{ch}) = \sum_{i=\text{ch}-n}^{\text{ch}+n} N_{\text{count}}(x, y, i), \quad (\text{A2})$$

and N_{count} is the number of (unflagged) scans in the pixel and frequency channel. The weight assigned to pixel (x, y) is

$$w(x, y, \text{ch}) = \frac{N(x, y, \text{ch})}{\sigma^2(x, y, \text{ch})} P_{\text{B}}^2(x, y, \text{ch}), \quad (\text{A3})$$

where

$$\sigma^2(x, y, \text{ch}) = \frac{N(x, y, \text{ch})}{N(x, y, \text{ch}) - 1} V(x, y, \text{ch}), \quad (\text{A4})$$

and the variance is

$$\begin{aligned} V(x, y, \text{ch}) &= \frac{1}{N(x, y, \text{ch})} \left(\sum_{i=\text{ch}-n}^{\text{ch}+n} N_{\text{count}}(x, y, i) T_{\text{res}}^2(x, y, i) \right) - \bar{T}_{\text{res}}^2(x, y, \text{ch}). \end{aligned} \quad (\text{A5})$$

Note here that $N > 2n + 1$ should be calculated from equation (A2).

A2 Pixel-weighted flux

The weighted flux of the point source is

$$\bar{S}(\text{ch}) = \frac{\sum_{(x,y)} w(x, y, \text{ch}) \bar{S}(x, y, \text{ch})}{\sum_{(x,y)} w(x, y, \text{ch})}. \quad (\text{A6})$$

Its standard deviation is

$$\sigma_S(\text{ch}) = \sqrt{\frac{\sum_{(x,y)} w(x, y, \text{ch}) (\bar{S}(x, y, \text{ch}) - \bar{S}(\text{ch}))^2}{\sum_{(x,y)} w(x, y, \text{ch})}}, \quad (\text{A7})$$

and the flux error is

$$\epsilon_S(\text{ch}) = \sqrt{\frac{\sum_{(x,y)} w^2(x, y, \text{ch})}{\left(\sum_{(x,y)} w(x, y, \text{ch}) \right)^2}} \sigma_S(\text{ch}). \quad (\text{A8})$$

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.