CAMBRIDGE
UNIVERSITY PRESS

Research Article

Verifying the Australian MWA EoR pipeline I: 21-cm sky model and correlated measurement density

J.L.B. Line[1,2] , C.M. Trott[1,2] , J.H. Cook[1,2] , B. Greig[3,2] , N. Barry[1,2] , and C.H. Jordan[1,2]

[1]International Centre for Radio Astronomy Research, Curtin University, Perth, WA, Australia, [2]ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO-3D), Canberra, Australia and [3]School of Physics, University of Melbourne, Parkville, VIC, Australia

Abstract

We present the first of two papers dedicated to verifying the Australian Epoch of Reionisation pipeline (`AusEoRPipe`) through simulation. The `AusEoRPipe` aims to disentangle 21-cm radiation emitted by gas surrounding the very first stars from contaminating foreground astrophysical sources and has been in the development for close to a decade. In this paper, we build an accurate 21-cm sky model that can be used by the `WODEN` simulation software to create visibilities containing a predictable 21-cm signal. We verify that the power spectrum (PS) estimator `CHIPS` can recover this signal in the absence of foregrounds. We also investigate how measurements in Fourier-space are correlated and how their gridded density affects the PS. We measure and fit for this effect using Gaussian-noise simulations of the Murchison Widefield Array (MWA) phase I layout. We find a gridding density correction factor of 2.651 appropriate for integrations equal to or greater than 30 minutes of data, which contain observations with multiple primary beam pointings and LSTs. Paper II of this series will use the results of this paper to test the `AusEoRPipe` in the presence of foregrounds and instrumental effects.

Keywords: Astronomy data analysis; reionisation; GPU computing

1. Introduction

Constraining when and how the first stars were formed has been a goal of astronomers for decades. By measuring cosmically redshifted 21-cm radiation from the hydrogen gas surrounding those first stars, one can conceivably map the effects of the ionising radiation coming from them, to infer their properties as a function of redshift. Unfortunately, the detection of 21-cm radiation from the Epoch of Reionisation (EoR) (redshifted to the 50–250 MHz frequency range) is impeded by a myriad of astrophysical sources, including Active Galactic Nuclei, Radio Galaxies, Supernova Remnants, and the diffuse synchrotron radiation emitted from the Milky Way. These foregrounds drown out the 21-cm signal (Furlanetto, Peng Oh, & Briggs 2006) and must either be subtracted from the data or avoided. One approach that somewhat naturally separates foregrounds from the signal is via measuring the power spectrum (PS). The PS can be used to garner spatial information of the effects of reionisation by the first stars and infer their properties. Furthermore, Fourier transforming over frequency to obtain a PS results in the foregrounds and signal manifesting in different areas of the PS given their differing spectral behaviours. Recent upper limits on the EoR have come from: Hydrogen Epoch of Reionisation Array (HERA, DeBoer et al. 2017); Low-Frequency Array (LOFAR, van Haarlem et al. 2013); The Murchison Widefield Array (MWA, Tingay et al. 2013;

Wayth et al. 2018); New Extension in Nançay Upgrading LOFAR (NenuFar, Zarka et al. 2012).

There are a number of challenges which must be overcome to make the statistical detection of the 21-cm signal. As well as nullifying the aforementioned foregrounds (e.g. Cook, Trott, & Line 2022; Acharya et al. 2023), one must deal with ionospheric refraction (Edler, de Gasperin, & Rafferty 2021; Chege et al. 2022, e.g.), terrestrial interference (e.g. Offringa, Mertens, & Koopmans 2019; Wilensky et al. 2023), and a myriad of instrumental calibration effects (e.g. Kern et al. 2020; Chokshi et al. 2021; Mevius et al. 2022; Kolopanis et al. 2023), all of which cause limiting systematics. These challenges have forced a decade of not only hardware but also software development, to calibrate and treat the data with extreme precision to uncover the underlying signal. Recent best-effort PS limits include Mertens et al. (2020), Trott et al. (2020), Rahimi et al. (2021), HERA Collaboration et al. (2023), Munshi et al. (2024).

Given that measuring the 21-cm PS is as much a software as it is an observational experiment, great care must be taken to understand what biases and systematics the software pipeline imparts upon the data. Given the limiting systematics listed above, a natural route is via simulation, where one can control which systematics are injected into a data set, with a ground 'truth' to be recovered. In short, if you inject a known 21-cm signal into a data set, process through your pipeline and recover what you injected, you gain confidence in your pipeline. This approach has been taken by a number of authors, including Barry et al. (2019a) using MWA data and Aguirre et al. (2022) with HERA data. To date, the Australian EoR pipeline (`AusEoRPipe`), which is designed to process MWA data, has not been tested from end to end. This is the

Corresponding author: J.L.B. Line; Email: jack.l.b.line@gmail.com

Cite this article: Line JLB, Trott CM, Cook JH, Greig B, Barry N and Jordan CH. (2024) Verifying the Australian MWA EoR pipeline I: 21-cm sky model and correlated measurement density. *Publications of the Astronomical Society of Australia* **41**, e067, 1–11. https://doi.org/10.1017/pasa.2024.31

goal of this series of papers. This paper aims to test the gridding and PS estimation aspects of the pipeline; the following paper in this series will focus on calibration.

The native output of an interferometer is a visibility, a sampling of the Fourier transform of the visible sky. There are two main ways to simulate a 21-cm signal in visibility space. One can either make an analytic model that produces similar statistics in a PS, to directly generate in visibility space, or can create an image based model and derive visibilities from that. The former is far less computationally expensive, but the latter locks in projection and geometrical effects inherent to true interferometric observations. Having an image-based 21-cm model locked to celestial coordinates allows for simulated observations at various hour angles, to probe the effects of changing instrumental primary beam patterns and visibility sampling.

Simulating an image-based 21-cm signal for the MWA is non-trivial however, given the sky coverage of the instrument. The full-width half-maximum of the main lobe of the primary beam of the MWA can be greater than 20×20 square degrees and 50×50 square degrees down to the 1% beam level. Simulating a realistic EoR volume capable of covering this sky-area is computationally demanding. Fortunately, Greig et al. (2022) generated a simulated EoR volume large enough, which we use in this work and detail in Section 3. Given the resolution of the instrument is of order an arcminute, the number of pixels required to represent 21-cm lightcones from this EoR volume is $\mathcal{O}(10^7)$. Down-sampling this model can affect the resultant statistics of the signal. For this reason, we use the WODEN simulation package (Line 2022), capable of ingesting sky models with millions of components, and producing realistic simulated MWA visibilities. As WODEN is GPU-accelerated, these simulations can be run in a reasonable time frame (see Section 4.1 for details). Importantly, WODEN is also capable of simulating both the 21-cm signal and the foregrounds at the expected power levels, which can be many orders of magnitude apart. This allows us to test signal recovery in the presence of foregrounds without boosting the amplitude of the 21-cm signal.

This paper is organised as follows. In Section 2 we briefly overview the AusEoRPipe and outline what simulated data are necessary to test them. In Section 3 we detail the construction of a 21-cm sky model compatible with WODEN. In Section 4 we verify the steps taken to project the 21-cm model into celestial coordinates, by applying them to a sky model of Gaussian noise. We also investigate the effects of gridding visibilities when generating power spectra and fit for the effect this has on estimating the resultant power. In Section 5, we test the 21-cm model and how well the AusEoRPipe can recover the expected PS. We discuss our results in Section 6 and conclude in Section 7.

2. Pipeline overview

The AusEoRPipe is designed to take raw data from the MWA correlator, calibrate for instrumental and ionospheric effects, subtract foregrounds, and then estimate a PS. Here, we list the different software packages that make up the AusEoRPipe and suggest simulations needed to test them. This section is intended to give the reader a broad overview of the AusEoRPipe and the overall goals of this series of papers, to give this paper context. At the end of this Section, we detail which parts of the work are covered in this paper.

Birli[a] is designed to take raw MWA data and preprocess it, including geometric correction, averaging, and radio frequency interference (RFI) flagging (via AOFlagger Offringa 2010 or SINSS Wilensky et al. 2019). Birli has been written to replace Cotter (Offringa et al. 2015), to have one unifying preprocessing package that works with both legacy and new MWA data (see Morrison et al. 2023, for details of the old and new MWA correlators). Given Birli has been tested to reproduce results out of Cotter, which has been tested and used within a number of MWA pipelines, we choose not to test Birli via simulation. To do so however, one would need to simulate raw MWA data, which for the original correlator would mean producing visibilities with no phase tracking, adding a frequency-dependent bandpass from the MWA polyphase band filter, and add in realistic RFI.

hyperdrive[b] (Jordan et al. submitted), takes either raw or preprocessed data, calibrates it, and then subtracts foregrounds. Built to replace the RTS (Mitchell et al. 2008), hyperdrive also creates model visibilities to calibrate against via an image-based sky model. It initially performs a direction-independent (DI) calibration step, that derives a gain for each MWA receiving element (tile). All frequencies are calibrated independently. After DI calibration, foregrounds can be subtracted. The Fourier transform along frequency to create the PS is expected to separate the 21-cm signal from the foregrounds, as foreground emission is supposed to be spectrally smooth, whereas the 21-cm signal has spectral structure. Incomplete sampling of the u,v plane however causes mode mixing between the two, as well as other frequency-related instrumental effects. A detection therefore relies on the foregrounds being subtracted without injecting false spectral structure, as to preserve the underlying 21-cm signal. To test this functionality, a simulation needs to contain both a 21-cm signal and foregrounds, as well as frequency-dependent instrumental gain errors. One can also add frequency-dependent effects like cable reflections (e.g. Ewall-Wice et al. 2016) and the bandpass to investigate how calibrating each frequency channel independently performs. These effects are direction-independent and so can be added to visibilities post simulation (functionality which is in development in WODEN). Further simulations including RFI and ionospheric refraction would test how robust the calibration is to environmental influences. These effects are direction-dependent and so must be adding during the calculations of the visibilities, which WODEN is currently incapable of.

CHIPS (Trott et al. 2016) takes the calibrated and subtracted visibilities and grids them into a spectral cube using an optimised Blackman-harris kernel. It then Fourier-transforms along frequency, and then either cylindrically averages to create a two-dimensional PS (2D PS) or spherically averages to create a one-dimensional PS (1D PS). Integrations of tens to hundreds of hours of data into a 1D PS have set the limits that have been released by the AusEoRPipe to date. The minimal test for CHIPS is to simulate a set of visibilities with a known 21-cm signal and see if CHIPS can recover it. Beyond that, taking in calibrated visibilities out of hyperdrive, derived from simulations with realistic instrumental effects and foregrounds, allows us to probe the effects of calibration on the resultant PS.

WODEN,[c] as previously mentioned, simulates MWA visibilities from a sky based model. It can use the same primary beam model

---

as `hyperdrive`, the FEE beam (Sokolowski et al. 2017) (via the `mwa_hyperbeam`[d] package) and creates visibilities that can be fed directly into either `hyperdrive` or `CHIPS`. WODEN calculates the interferometric measurement equation for every component in the sky model. Each component can either be a point source (a dirac-delta function upon the sky), an elliptical Gaussian, or a Shapelet model (see Line et al. 2020, for more detail on Shapelets). The measurement equation encodes baseline and sky projection effects inherently, making the simulator accurate across the large field of view of the MWA. The caveat being that the sky model must be broken into the small components listed here, meaning representations of large-scale structure such as the galactic diffuse emission need millions of components. Aside from the primary beam model, the WODEN code base is entirely independent to `hyperdrive`. Although they use the same methodology to generate visibilities, this redundancy in code is by design; any internal bug that may cancel out when generating visibilities and calibrating them all with the same code base is avoided.

Along with the aforementioned packages, extensive quality metrics are generated at various stages (Nunhokee et al. *submitted*). These metrics are used to cull any observations that are deemed unusable due to insurmountable instrumental effects, ionospheric conditions, RFI events, contamination by bright sources in the primary beam sidelobes, and other effects. The simulations detailed above could be used to check whether these metrics are able to catch poor data and check whether the limits set for each metric still allow for a detection of the 21-cm signal.

In this paper, we ignore calibration entirely, and focus on generating a realistic 21-cm sky model to simulate through WODEN and directly input into `CHIPS`. As visibilities are additive, once we are able to generate accurate 21-cm simulations, we can add a variety of foregrounds and instrumental effects to test the calibration and subtraction step, without the need to rerun the more expensive 21-cm simulations. We leave testing `hyperdrive` to the second paper in this series.

2.1 MWA EoR observing

The MWA is a low-frequency radio interferometer, with receiving elements consisting of 4×4 grids of dual linear-polarisation bow-tie antennas. These 'tiles' are electronically steered through beam-forming. To estimate the 21-cm PS, the MWA EoR collaboration have identified a number of fields with lower sky temperatures, which have now been observed for hundreds of hours (e.g. Trott et al. 2020), in an attempt to average over thermal noise. Given the electronic beamforming, this has resulted in a drift-and-shift observational campaign, where the target field drifts through a number of pointings, with scheduling keeping the field centre as close to the primary beam centre as possible (see Jacobs et al. 2016, for more details).

In this paper we focus on the EoR0 field, centred at RA, DEC = $0^h, -27°$, observed between frequencies of 167–198 MHz (known as the high-band). These parameters have consistently yielded the best MWA limits (e.g. Barry et al. 2019b; Trott et al. 2020; Chege et al. 2022). We also focus on five pointings, labelled -2 through +2 (c.f. Beardsley et al. 2016), where -2 means two pointings before the zenith 0 pointing, and +2 meaning two pointings after 0 pointing. There is an approximate 15° shift between pointings, all along
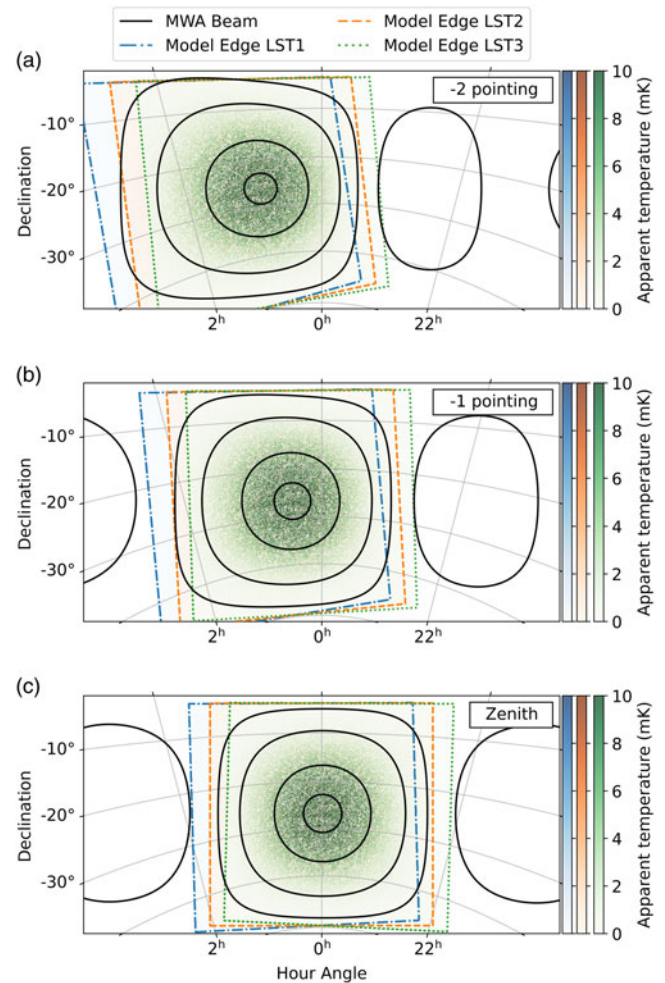
Figure 1. Demonstration of the MWA primary beam and its interaction with the 21-cm sky model, via the: (a) -2 pointing; (b) -1 pointing; (c) Zenith pointing. Each plot is locked to the observer in HA/Dec. The solid black lines represent the instrumental Stokes I primary beam pattern contoured at 1%, 10%, 50%, and 90% power levels. The coloured images show the apparent 21-cm sky model after attenuation by the primary beam at three different LSTs, with the corresponding colour lines marking the edges of the full 21-cm sky model. Both the beam and sky model are shown at 167 MHz, where the primary beam is largest for a high-band observation. Note that the +1, +2, pointings as described in Section 5 are simply westward reflections of the -1, -2 pointings, so aren't reproduced here for brevity.

the meridian. Examples are shown in Fig. 1. MWA EoR data are typically taken in two minute chunks known as an observation.

3. 21-cm sky-model

We use the 21-cm lightcone detailed in Greig et al. (2022), which is derived from a simulated EoR cube with sides of length 7.5 Gpc. This cube was generated using a simplified version of `21CMFACST` (Mesinger&Furlanetto 2007; Mesinger, Furlanetto, & Cen 2011) and was spectrally sampled at 80 kHz to match typical MWA EoR analysis parameters. This methodology generates a volume that projects to a sky coverage of $\sim 50 \times 50$ square degrees, which is essential given the footprint of the MWA primary beam. A number of steps are necessary to translate the lightcone box into a WODEN sky model. The box is a collection of 2D projected 21 cm intensity distributions, each at a unique redshift. These cartesian

projections are taken by slicing a full 3D simulation volume at regular redshift intervals, as the simulation is evolved with time. This leaves us with a number of transformations needed to allow WODEN to ingest the model:

- Each 2D 21-cm slice has a pixel resolution Δx in cMpc; each slice therefore has a different angular resolution, as this quantity is redshift-dependent. To run efficiently, WODEN needs a grid of pixels constantly sampled in RA/DEC, necessitating a redshift-dependent angular interpolation.
- The box is regularly sampled in redshift (Δz), which results in irregular sampling in frequency ($\Delta \nu$). Interferometric data is sampled regularly in frequency, necessitating a second interpolation over frequency.
- The box is in units of mK. WODEN ingests units of Jansky (Jy), which is an integrated flux density. The conversion from mK to Jy/sterrad is straightforward, however, the physical volume of each pixel changes as a function of frequency. Along with the angular extent, the pixel volume is determined by Δz. Given the interpolation over angle and frequency, this volume change must be taken into account, as it effects the resultant variance of the 21-cm map. This effect is noted in the script make_flat_spectrum_eor.py[e] in the Python package pyradiosky.

In the following, we use the cosmology assumed by Greig et al. (2022), a ΛCDM cosmology with: $H_0 = 68$; $\Omega_M = 0.31$; $\Omega_\Lambda = 0.69$; $\Omega_b = 0.048$. We use astropy (Astropy Collaboration et al. 2013, 2018, 2022) for all cosmological calculations.

We choose to interpolate all slices to the finest angular resolution in the box to retain as much angular structure as possible. We call this the reference redshift, $z_{ref} = 7.5693$. Given $\Delta \theta = \Delta x / \mathcal{D}(z)$, where $\mathcal{D}$ is the comoving distance, we interpolate to an angular resolution of ~ 27 arcsec. We found interpolating the Cartesian slice directly into a TAN FITS projection (a gnomic projection; see Calabretta & Greisen 2002, for details) centred at RA, DEC $= 0^h, -27^\circ$ returned the expected PS. We experimented with bilinear and bicubic interpolation, but found significant signal loss in the resultant PS. It's possible some variant of Gaussian Process Regression may be more effective, however for the purposes of this work we found a simple nearest-neighbour approximation was sufficient. Similarly, we applied a nearest-neighbour interpolation along the frequency axis, as again we saw signal loss in the final PS when applying linear or cubic interpolation. Along lines of sight, the 21-cm signal rapidly fluctuates between positive and zero, and so these kinds of interpolation tend to create false signal. Once these interpolations have been applied, we then scale for the change in volume of the pixel ($V_{pix}(z)$) and its effect on the variance. We do this for each redshift by calculating via the differential comoving volume and multiplying by Δz. We then scale each redshift slice by a factor C_{pix} given by

$$C_{pix} = \sqrt{\frac{V_{pix}(z)}{V_{pix}(z_{ref})}}, \qquad (1)$$

with the square root ensuring the variance is scaled by the pixel volume.

Once we have interpolated, scaled, and transformed each slice in Jy/sterrad, we convert each pixel into a point source with units of Jy by multiplying by the pixel solid angle. Ideally, we would the find some way to tile these maps to give an all-sky 21-cm sky model. However, given we cannot interpolate the box without signal loss, it is computationally infeasible. At the lowest frequencies, the number of point sources necessary would approach 200 million. However, given the extreme volume of the simulated 21-cm box, without tiling, the sky model already covers the main lobe of the MWA primary beam down to the ~1% power level for pointings of interest (see Fig. 1).

To investigate whether the projection steps detailed in this Section induce any form of signal loss or bias, we apply them to a purely Gaussian noise simulation in Section 4. We report of the results of the 21-cm sky model in Section 5.

4. White Gaussian noise sky simulation

The PS $P(k)$ of a White Gaussian noise distribution $\mathcal{N}$ of mean $\mu = 0$ and variance σ^2 is proportionate to the variance, that is, $P(k) \propto \sigma^2$. By creating a sky model of purely White Gaussian noise, we can therefore predict an output $P(k)$ and check the normalisation of the AusEoRPipe. Given we have outlined a method to project a 21-cm box from units of cMpc in Section 3, we can simply repeat the entire process, starting from a box of purely White Gaussian noise (from hereon in referred to as the Gaussian noise simulation). We choose to set $\sigma^2 = 16\,\text{mK}^2$ as this yields a power comparable to those expected from the 21-cm signal. We typically report the PS in units of $\text{mK}^2\,\text{Mpc}^3\,h^{-3}$, where h is the reduced Hubble constant; we use $h = 0.68$ in this work. We can therefore predict our expected noise PS P_N via

$$P_N(k) = \sigma^2 V_{pix}(z_{ref})h^3 = 8.923\,\text{mK}^2\text{Mpc}^3 h^{-3} \qquad (2)$$

We use WODEN[f] to simulate a single zenith pointing observation with an 8 s time cadence and 80 kHz frequency resolution, using a frequency-interpolated version of the FEE beam. We use the convention that Stokes I = (XX + YY)/2, where XX and YY are the two linear polarisations. Given we are using simulated data with isotropic and unpolarised sky models, there is little difference between XX and YY, and therefore the power spectra shown are effectively Stokes I. All PS shown in this paper are from the north-south aligned polarisation. We run CHIPS on the simulated visibilities, to produce the 2D PS as shown in Fig. 2. All CHIPS plots were generated using the Python package CHIPS_wrappers.[g]

Fig. 2 demonstrates a known effect of estimating the PS from gridded visibilities, where the power inferred is lower where the density of gridded visibilities is higher, as observed by Barry et al. (2019a, see Appendix A). Note we have not corrected for this in Fig. 2, which is normally done by multiplying by a factor two (as found by Barry et al. 2019a), to explicitly show the effect.

Visibilities are gridding using a kernel which is designed to minimise aliasing and match the area of the MWA primary beam (Trott et al. 2016). Given the array layout, and the need to match

---

[e]https://github.com/RadioAstronomySoftwareGroup/pyradiosky/blob/main/scripts/make_flat_spectrum_eor.py.

[f]For this paper we use the docker v2.0 container of WODEN, installed via docker pull docker://jlbline/woden-2.0.

[g]https://github.com/JLBLine/CHIPS_wrappers.
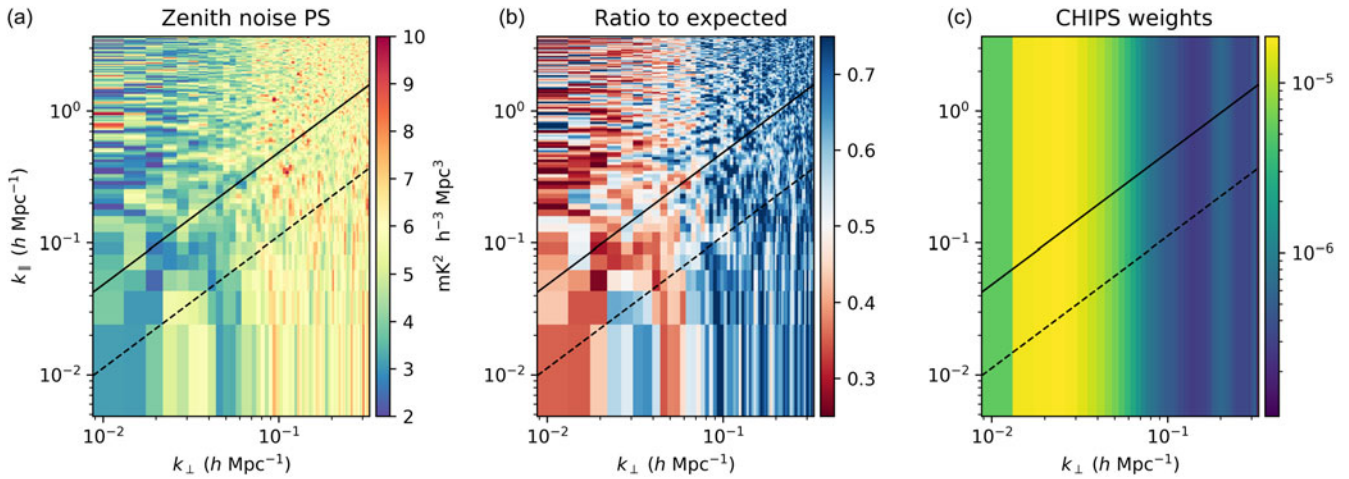
Figure 2. Data from a single zenith observation, without any correction for gridding density, where: (a) shows the 2D PS; (b) the ratio to the expected value; *c)* the CHIPS gridding weights. $k_\perp$ modes are derived from the instrument sampling in the visibility *u,v* plane, averaged down to one dimension, and $k_\parallel$ from the fourier transform of the visibilities along frequency. The CHIPS gridding weights therefore show the *u,v* gridding density averaged into one dimension.
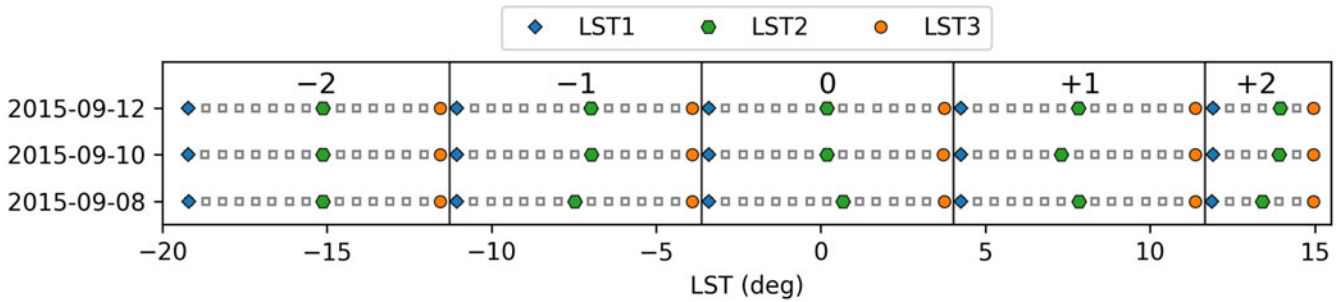
Figure 3. Observations simulated in this paper, based on two real nights of MWA EoR0 observing. The blue diamonds show the simulated LST1 subset, green hexagons the LST2 subset, and orange circles the LST3 subset. Each hollow square shows a different two minute snapshot which was not simulated but exists in the real data set. Dividing lines and labels show the changes from pointings -2 through +2. Note observational constraints mean there are less +2 pointings.

the kernel to the primary beam, the kernel has an unavoidable footprint large enough to overlap with neighbouring gridded visibilities. Depending on the covariance between neighbouring visibilities, and the density of the gridded visibilities, the power estimated will therefore vary as a function of *u,v* gridding location. Using simulations, Barry et al. (2019a) found a factor of two correction was sufficient as a normalisation factor for modes of interest. This estimate was made using the FHD/εppsilon pipeline and a model of the MWA primary beam as a gridding kernel, whereas CHIPS uses a Blackman-Harris kernel. This estimate was also made using 2 s resolution data, rather than 8 s. Given these differences, we investigate the gridding density correction factor in Section 4.1 using CHIPS.

4.1 Gridding density correction factor

The gridding density is directly affected by the *u,v* coverage, which in turn is dictated by the array layout and phase centre. The covariance of neighbouring *u,v* points will therefore be affected by LST (as the phase centre is always set to EoR0 field centre). The primary beam pointing will also have an effect as this changes the amplitude and spectral behaviour of the visibilities. To thoroughly investigate the gridding density we therefore simulate a number of observations spanning realistic ranges of LST and pointings. To do so, we take the first (subset LST1), central (subset LST2), and final observation (subset LST3) across five different pointings (−2 to +2) from three real nights of MWA phase I EoR observing (see Fig. 3). For the LST2 subset, we select the observation where the primary beam pointing is closest to the EoR0 field centre, to maximise the beam coverage over the sky model coverage. As the visibilities are normally averaged to 8 s, 80 kHz we simulate at this cadence to save compute time. Each observation takes a total of ∼32 GPU hours. For 45 simulated observations this is a total of 60 GPU days. We use the Pawsey Garrawarla cluster[h] and split each simulation across 24 GPUs, meaning these simulations take less than 3 days real-time.

By running CHIPS on various combinations of these observations, we can investigate both primary beam pointing and LST effects. All integrations over multiple observations are done coherently, that is, all observations are gridded to the same *u,v* grid, which is then used to estimate the PS. We take the median of the ratio of the observed PS to the expected value as a function

[h]https://pawsey.org.au/systems/garrawarla/.
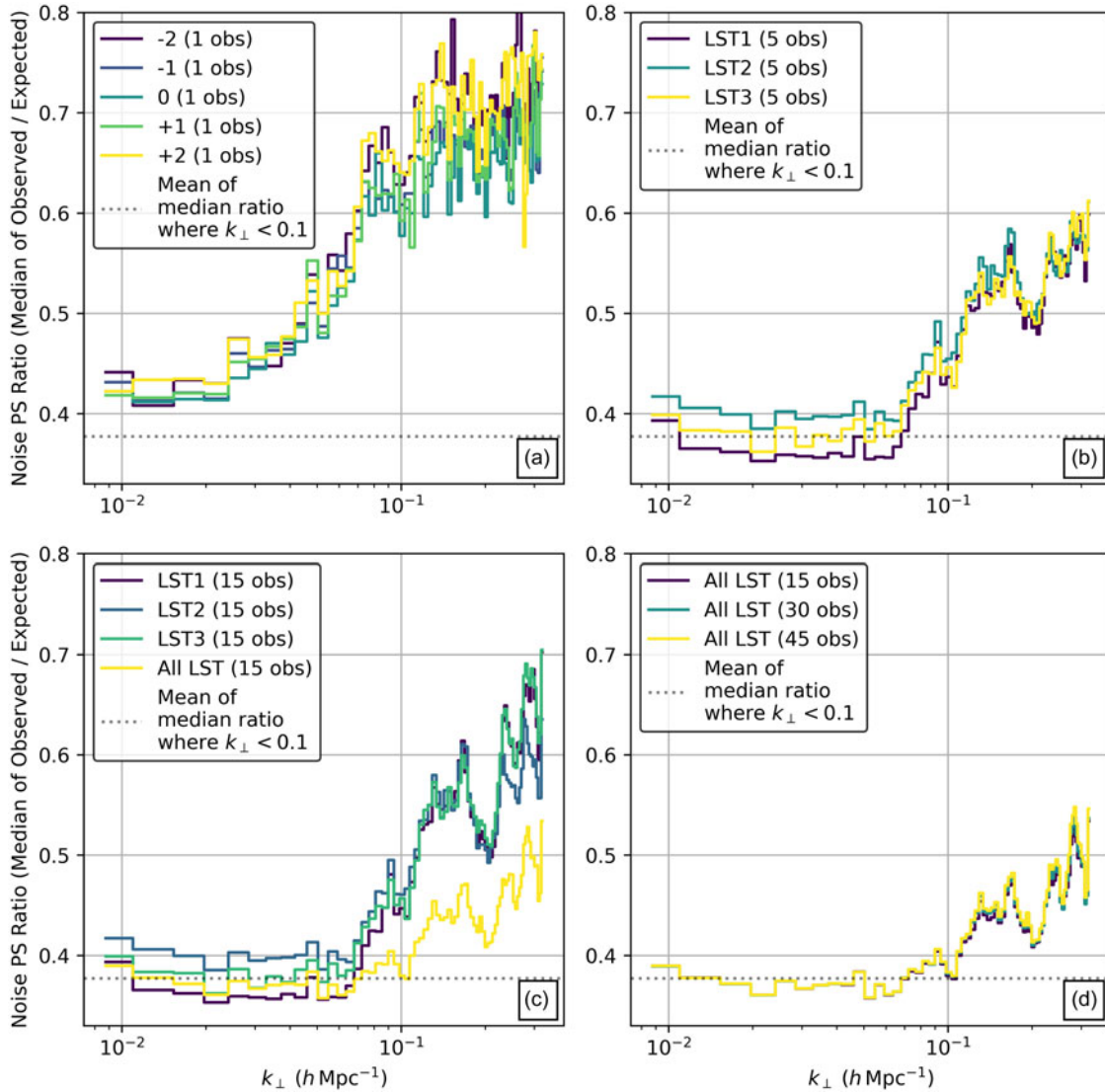
Figure 4. The median ratio of the recovered noise PS to expected value, as a function of $k_\perp$, for: (a) the five different pointings, each for a single observation; (b) the three different LST subsets integrated over 5 observations; (c) the three different LST subsets integrated over 15 observations; (d) all three LST subsets integrated together over 15, 30, and 45 observations. The horizontal dashed line shows the mean of the median ratios for $k_\perp < 0.1$ as measured from the bottom right plot. This dashed line Any dataset which is a multiple of five observations has an even split across the five pointings.

of $k_\perp$ (i.e. take the median along the y-axis of the middle panel in Fig. 2) to see the effects of gridding, as shown in Fig. 4. We take the median, rather than the mean, as the distribution of ratios along each $k_\perp$ bin display significant skew.

Fig. 4a shows that for a single observation, the estimated power is consistent across pointings, with lower power estimated a low $k_\perp$ (higher gridding density). This shows the beam volume correction applied internally in CHIPS returns a consistent power level across the five pointings. Fig. 4b shows that the power recovered from the three LST subsets over 5 observations is roughly consistent, but the power recovered does vary somewhat. This recovered power depends on the exact u,v coverage, and the way the primary beam interacts with the sky model. Fig. 4c shows that when integrating over 15 observations for each LST subset, the recovered power at low $k_\perp$ is consistent with the power recovered when integrating over 5 observations. However, the power recovered at

higher $k_\perp$ is actually higher than at when integrating over 5 observations. This highlights the fact that the power estimated comes from a combination of not just the amount of data gridded, but the covariance between the neighbouring gridded visibilities. This is further highlighted by showing that integrating the three LST subsets together across 15 observations yields significantly lower power at high $k_\perp$. High $k_\perp$ is derived from longer baselines, where the drift of a baseline u,v coordinate over time is greatest. The LST1 and LST3 groups are also closely separated in time, yielding close gridding locations, but overlap as the beam pointing changes. The amplitude and the spectral behaviour of the primary beam therefore sharply changes on the sky, reducing the covariance between neighbouring gridded visibilities. This results in the lower power recovered. Fig. 4d shows that when integrating more and more data with the same combination of LST coverage and pointings, there is no change in power at low $k_\perp$.
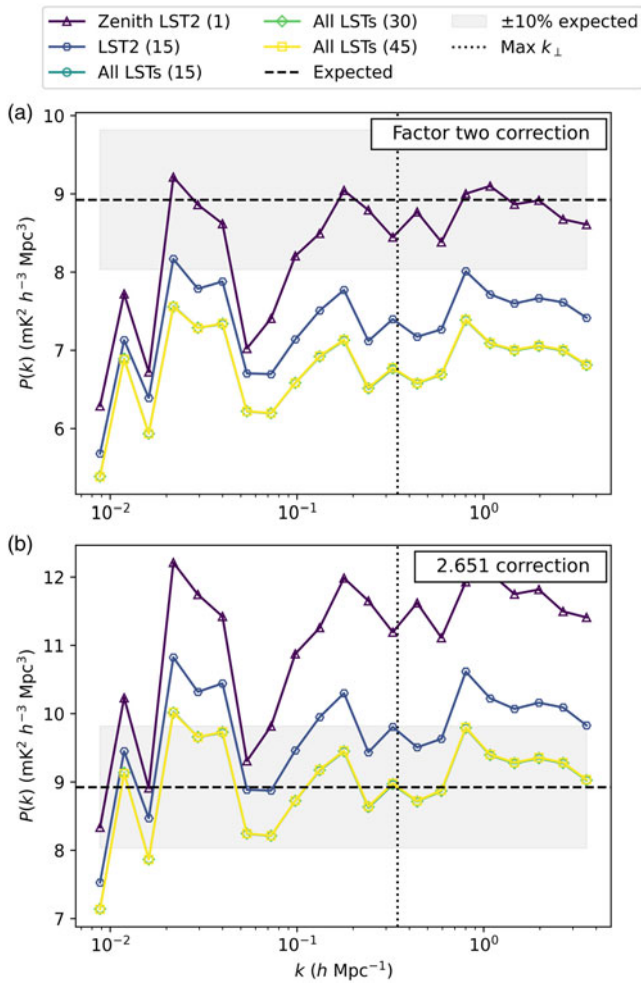
Figure 5. 1D PS for various integrations of the Gaussian noise simulation, where (a) a factor of two was used to normalise for gridding density and (b) a factor of 2.5845 was used. The vertical dashed line shows the maximum $k_\perp$-mode.

Given the stability of the recovered power when combining all LST subsets with an integration of 15 observations or more, one could fit a functional form to derive a correction. We investigate this approach in Appendix A and find little to no difference in using a fit when compared to just applying a scalar normalisation factor. PS upper limits are derived from integrating hours of data to overcome thermal noise, and modes where $k_\perp > 0.1\,h\,\mathrm{Mpc}^{-1}$ are typically cut when using the AusEoRPipe. To obtain a normalisation factor for the gridding density we therefore simply take the mean of the median observed ratios for the integrations over all LSTs and pointings where $k_\perp < 0.1\,h\,\mathrm{Mpc}^{-1}$. This value is plotted as the horizontal grey dashed line in Fig. 4. Inverting this gives a normalisation factor of 2.651. We apply this normalisation and create a number of 1D PS for various integrations, shown in Fig. 5, and compare the outcomes to the previous correction factor of two. This shows that the new normalisation factor causes too much power to be recovered for a single observation, and when only using the LST2 subset, but recovers the correct power to within 10% for most k modes when integrating over all LSTs and pointings. The factor two correction (coincidentally) does well for a single observation, but does not recover the correct power for higher integrations.

5. Recovered 21-cm signal

To test the 21-cm model detailed in Section 3, we simulate 30 observations, matching the first two nights detailed in Fig. 3, with the same parameters as used for the Gaussian noise simulation (detailed in Section 4). We leave out simulating the third night purely to save on computational resources. To qualitatively assess the accuracy of the model through WODEN and into CHIPS, we compare a CHIPS 2D PS to one derived directly from the lightcone box, shown in Fig. 6. We produce the 2D PS directly from the light-cone using the OSIRIS package[i] (Cook, Trott, & Line 2022), where no primary beam or instrument sampling was applied, meaning the entire u,v plane was sampled. We see that the CHIPS PS is broadly consistent with the OSIRIS PS. Noticeably, it seems there is a potential signal loss at high $k_\parallel$ in the CHIPS PS (the darker blue region at the top of Fig. 6b). There is also excess power seen along high $k_\perp$, around the dashed black line, which is expected as the instrument baseline sampling causes mode mixing, moving power up from lower to higher $k_\parallel$.

To test the gridding normalisation, we apply it to various integrations and produce 1D PS, shown in Fig. 7. We compare them to an expected signal, again derived directly from the 21-cm lightcone box. Excellent agreement is shown, and similarly to the Gaussian noise simulation, the gridding density normalisation estimates too much power for short integrations, but renders consistent results for longer integrations with the same mix of LSTs and pointings.

In Fig. 8 we show the ratio of the final recovered 21-cm signal to the expected value, for a 60-min integration. While good agreement is shown, there does seem to be signal loss around $0.1 < k < 1$, which was not seen for the Gaussian noise simulation. This is likely due to the frequency interpolation of the original lightcone as detailed in Section 3. For reference, Fig. 8 shows the maximum $k_\perp$ mode that went into the average from 2D to 1D. k modes above this are binned from only high $k_\parallel$ modes, which is where signals with the most spectral structure manifest. As this area is where the signal loss is found, it is likely the frequency interpolation is causing this. As the Gaussian noise sky model has no correlation over frequency, it should not suffer from this interpolation, which is why it doesn't display the same signal loss.

Fig. 8 also shows the effect of performing the 'wedge-cut', where modes normally dominated by foreground sources are removed. While broadly consistent, the signal loss seems to worsen after the cuts. This is likely due to the reduced number of bins going into the average from 2D to 1D. As can be seen from Fig. 6c, the horizon cut (marked by the solid black line) removes more samples for smaller $k_\parallel$ modes. However, lower sampling is not the only possible reason. The manifestation of the primary beam, as well as the exact u,v distribution, could both cause a systematic bias at specific spots in the 2D PS, which could contribute here.

6. Discussion

The Gaussian noise simulation results in Section 4 confirm that gridding visibilities without accounting for their full covariance from overlapping footprints requires a normalisation factor. They reveal however that this normalisation factor depends on an interplay between the array layout, the LST range of the integrated
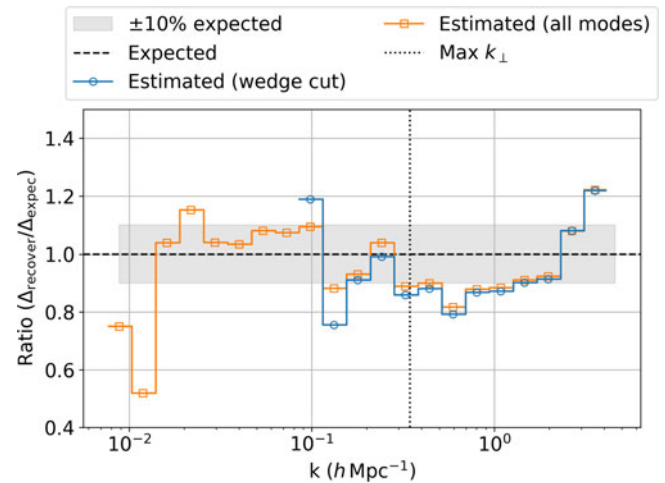
---

[i]https://github.com/JaidenCook/OSIRIS.
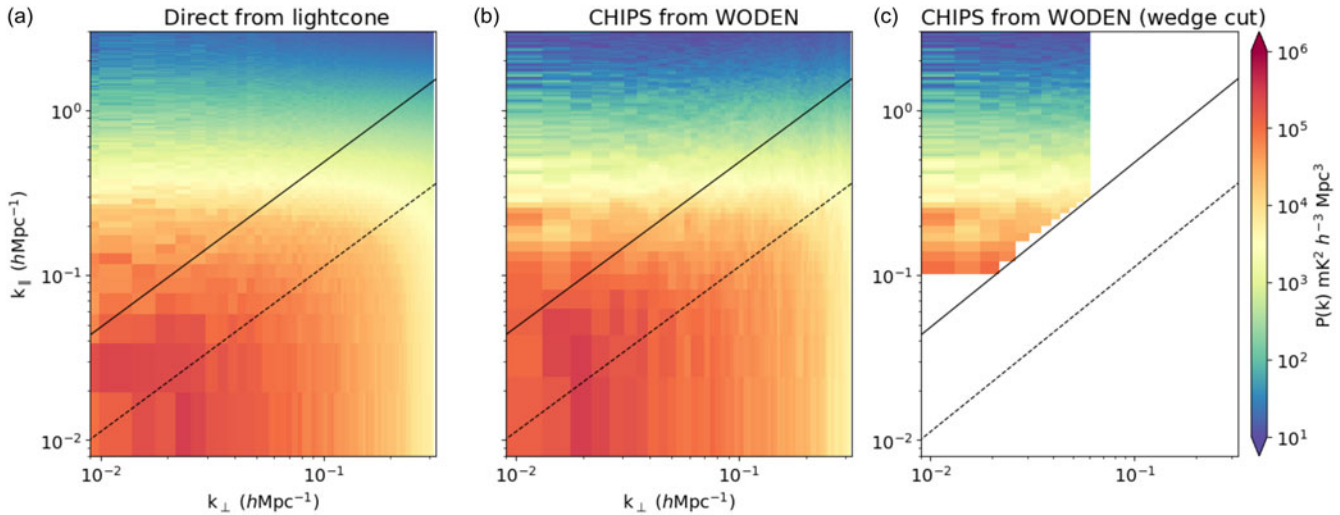
Figure 6. 2D PS from the 21-cm simulation where: (a) produced using OSIRIS directly from the lightcone box from Greig et al. (2022); (b) a CHIPS 2D PS made from integrating over 30 simulated observations; (c) the same PS from (b) but with standard cuts made to remove foreground contamination. The solid black line indicates an estimate of the horizon, and the dashed black line the full-width half-max of the MWA primary beam.
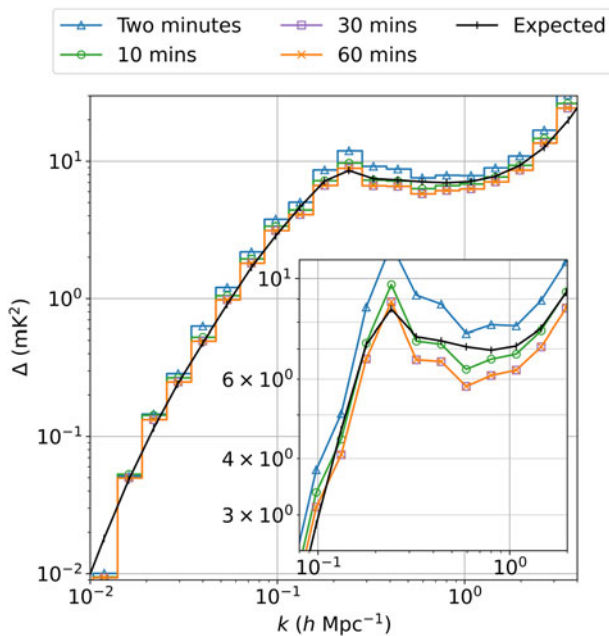
Figure 7. Recovered 21-cm signal from various integration lengths of simulated observations, when normalised with a factor 2.5845. The panel bottom right is included to illustrate the normalisation effects between $0.1 < k < 1.0\,h\,\mathrm{Mpc}^{-1}$. The two minute data set is for a zenith pointing; the 10 minute data set is for the LST2 subset; all other data sets are an even split between the five pointings and LST subsets as described in Section 5.

Figure 8. Ratios of the recovered 21-cm signal to the expected value, for a 60 min integration. The orange with crosses line shows the ratio when using the factor two gridding density correction, and the blue with circles when using the fitted correction. The vertical dashed line shows the maximum $k_\perp$-mode, meaning any k-mode above this was derived purely from the Fourier transform of the visibilities along frequency.

observations, as well as the primary beam pointings, as these all effect the covariance between the visibilities. To truly correct for this effect, one would need to calculate this covariance and propagate it through the gridding and PS estimation steps (see Liu, Parsons, & Trott 2014, for more information on calculating covariances in k-space). We instead simply measure and fit for the effect, given we can estimate the expected power from the Gaussian noise simulation. It could be argued that simply taking the median of

the ratios where $k_\perp < 0.1\,h\,\mathrm{Mpc}^{-1}$ from Fig. 4d is too simplistic, given we can fit a functional form as shown in Appendix A. We choose to err on the side of simplicity however, given the single scaling factor we derive recovers the correct Gaussian-noise power to within 10% for most k-modes. This normalisation factor comes with a set of caveats, which require further investigation to be understood:

- We have only tested the MWA phase I layout. Given the covariance between u,v may well be different for different array layouts, we cannot be sure this correction factor would work with the MWA phase II compact layout, for example

- Upon initial investigation, 2 s resolution data was seen to require a smaller normalisation factor for the same integration length. It is possible that gridding four more times more data will eventually reach the same plateau as 8 s data, but it might change the level of covariance in the gridded visibilities. Testing this will require four times the computational resources

- It is possible that the gridding normalisation will change for a different frequency range, as the rate of change of the u,v points with time will be different. To test this with the method outlined in this paper, another 21-cm lightcone box will need to be generated to match the required frequency coverage

- We attribute the signal loss seen in Fig. 8 to the frequency interpolation of the original 21-cm lightcone box. However, there could be other sources of signal loss, which include the primary beam shape altering the signal distribution, and a dependence on details of the gridding within a $k_\perp$ bin. To remove frequency interpolation and check this signal loss issue, the original lightcone box would need to be regenerated with constant frequency sampling. This is non-trivial, as it would require `21CMFAST` simulations with adaptable voxel sizes. Alternatively, one could run the noise sky simulation test using a type of noise with a spectral structure (e.g. Brownian noise). As long as the shape of the noise could be easily predicted in PS space, one could test for signal loss

It should be noted that this method depends on not only the Gaussian noise model being correctly scaled, but that our prediction for the power level is correct. However, given the normalisation also yields the expected power level for the 21-cm model, with that prediction made directly from the lightcone, this gives us confidence that our prediction and scaling are correct. We further note that we ran the `WODEN` simulations in so-called `float` mode, which sets some of the internal precision to single instead of double precision, with an appreciable speed up (see Line 2022, for more details). We ran a number of comparisons to fully `double` simulations all the way through to the PS and saw no change in the results.

It is worth considering that gridding density is not necessarily the only effect that could be causing a loss in signal. The model itself, the `WODEN` simulator, and `CHIPS` are potentially sources of signal loss. However, given that the Gaussian noise sky model is as simple a model as possible, `WODEN` has been heavily unit tested,[j] and we are fitting a `CHIPS`-specific normalisation factor, we are confident the normalisation derived here is applicable to real data put through the `AusEoRPipe`.

7. Conclusion

We have projected a 21-cm lightcone into an celestial coordinate-locked sky model for use with the `WODEN` simulator. The model is stored in a FITS table format easily adaptable to other simulation packages and is available from the PASA Datastore. In testing the validity of this model, we generated an equivalent White Gaussian noise model. We used this to test the effects of gridding visibilities using a kernel to produce power spectra, through simulated

---

[j]See https://woden.readthedocs.io/en/latest/testing/cmake_testing.html for details on testing.

---

observations. We confirmed the known effect that power estimated from gridded visibilities is lower where the gridding density is higher. We found the effect to be dependent on a combination of array layout, primary beam pointing, and LST of integrated observations. We also observed that the reduction in power estimated plateaus after integrating more than 30 minutes of data. We found a single scalar normalisation factor of 2.651 was sufficient to recover the expected power to within 10% for most k-modes in a 1D PS. We then generated simulated observations of the 21-cm sky model and verified that `CHIPS` was able to recover the expected signal. We found that frequency interpolation of the original lightcone box causes some signal loss at higher k, but with the 2.651 normalisation factor, the recovered power is again close to within 10% of the expect signal. We conclude that this sky model is sufficient to accurately test the `AusEoRPipe` and can be used in conjunction with foreground models and simulated instrumental effects to test calibration.

References

Acharya, A., et al. 2023, MNRAS
Aguirre, J. E., et al. 2022, ApJ, 924, 85
Astropy Collaboration, et al. 2018. AJ, 156, 123
Astropy Collaboration, et al. 2022, ApJ, 935, 167
Astropy Collaboration, et al. 2013, A&A, 558, A33
Barry, N., Beardsley, A. P., Byrne, R., Hazelton, B., Morales, M. F., Pober, J. C., & Sullivan, I. 2019, PASA, 36, e026
Barry, N., et al. 2019, ApJ, 884, 1
Beardsley, A. P., et al. 2016, ApJ, 833, 102
Calabretta, M. R., & Greisen, E. W. 2002, A&A, 395, 1077
Chege, J. K., Jordan, C. H., Lynch, C., Trott, C. M., Line, J. L. B., Pindor, B., & Yoshiura, S. 2022, PASA, 39, e047
Chokshi, A., Line, J. L. B., Barry, N., Ung, D., Kenney, D., McPhail, A., Williams, A., & Webster, R. L. 2021, MNRAS, 502, 1990
Cook, J. H., Trott, C. M., & Line, J. L. B. 2022, MNRAS, 514, 790
DeBoer, D. R., et al. 2017, PASP, 129, 045001. ISSN: 0004-6280
Edler, H. W., de Gasperin, F., & Rafferty, D. 2021, A&A, 652, A37
Ewall-Wice, A., et al. 2016, MNRAS, 460, 4320

Furlanetto, S. R., Peng Oh, S., & Briggs, F. H. 2006, PhR, 433, 181. ISSN: 0370-1573

Gooch, R. 1996, in Astronomical Data Analysis Software and Systems V, Astronomical Society of the Pacific Conference Series, ed. G. H. Jacoby, & J. Barnes, 101, 80

Greig, B., Wyithe, J. S. B., Murray, S. G., Mutch, S. J., & Trott, C. M. 2022, MNRAS, 516, 5588

HERA Collaboration, et al. 2023, ApJ, 945, 124

Jacobs, D. C., et al. 2016, ApJ, 825, 114. ISSN: 1538-4357

Joye, W. A., & Mandel, E. 2003, in Astronomical Data Analysis Software and Systems XII, Astronomical Society of the Pacific Conference Series, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook, 295, 489

Kern, N. S., et al. 2020, ApJ, 888, 70

Kolopanis, M., Pober, J. C., Jacobs, D. C., & McGraw, S. 2023, MNRAS, 521, 5120

Line, J. L. B., et al. 2020, PASA, 37, e027

Line, J. 2022, JOSS, 7, 3676

Liu, A., Parsons, A. R., & Trott, C. M. 2014, PhRvD, 90, 023019

Mertens, F. G., et al. 2020, MNRAS, 493, 1662

Mesinger, A., & Furlanetto, S. 2007, ApJ, 669, 663

Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955

Mevius, M., et al. 2022, MNRAS, 509, 3693

Mitchell, D. A., Greenhill, L. J., Wayth, R. B., Sault, R. J., Lonsdale, C. J., Cappallo, R. J., Morales, M. F., & Ord, S. M. 2008, ISTSP, 2, 707. ISSN: 1932-4553. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4703504

Morrison, I. S., et al. 2023, PASA, 40, e019

Munshi, S., et al. 2024, A&A, 681, A62

Offringa, A. R. 2010, Astrophysics Source Code Library, record ascl:1010.017, October. ascl: 1010.017

Offringa, A. R., Mertens, F., & Koopmans, L. V. E. 2019, MNRAS, 484, 2866

Offringa, A. R., et al. 2015, PASA, 32, e008. ISSN: 1323-3580. http://adsabs.harvard.edu/abs/2015PASA...32...15Y%20http://www.journals.cambridge.org/abstract_S1323358015000077

Rahimi, M., et al. 2021, MNRAS, 508, 5954

Sokolowski, M., et al. 2017, PASA, 34, e062. ISSN: 1323-3580. http://arxiv.org/abs/1710.07478%20http://arxiv.org/abs/1710.07478%20https://www.cambridge.org/core/product/identifier/S1323358017000546/type/journal_article

Tingay, S. J., et al. 2013, PASA, 30, e007. ISSN: 1323-3580. http://arxiv.org/abs/1206.6945%20http://www.journalscambridge.org/abstract_S1323358012000070

Trott, C. M., et al. 2020, MNRAS, 493, 4711

Trott, C. M., et al. 2016, ApJ, 752, 34

van Haarlem, M. P., et al. 2013, A&A, 556, A2. ISSN: 0004-6361

Waskom, M. L. 2021, JOSS, 6, 3021

Wayth, RANDALL B., et al. 2018, PASA, 35, e033. ISSN: 1323-3580. https://arxiv.org/abs/1809.06466%20http://arxiv.org/abs/1809.06466%20https://www.cambridge.org/core/product/identifier/S1323358018000371/type/journal_article

Wilensky, M. J., Morales, M. F., Hazelton, B. J., Barry, N., Byrne, R., & Roy, S. 2019, PASP, 131, 114507

Wilensky, M. J., et al. 2023, ApJ, 957, 78

Zarka, P., Girard, J. N., Tagger, M., & Denis, L. 2012, in Sf2a-2012: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics, ed. S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, & H. Wozniak, 687

Appendix A. Fitting for gridding density correction factor

As noted in Section 4.1, the power estimated from the Gaussian noise simulations is consistent for integrations of 15, 30, and 45 observations (when using the same combination of LSTs and pointings). We can therefore correct for gridding density by fitting the observed ratio of measured to expected power as a function of
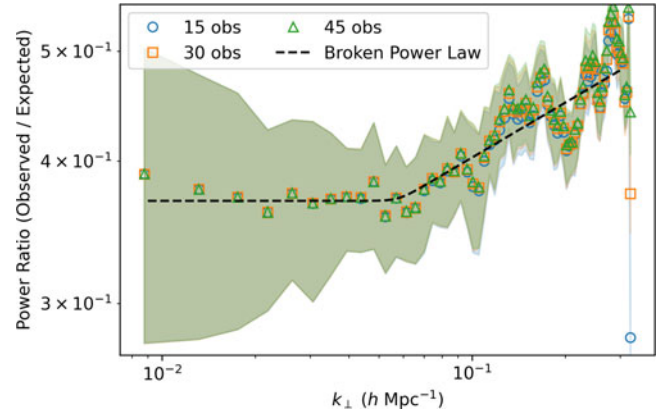
Figure A1. Fitting the median recovered ratio of the noise simulation as a function of $k_\perp$. Circles, squares, and triangles show the median ratio for integrations of 15, 30, and 45 observations, respectively. The shaded regions are bound by the median absolute deviation. The dashed black line shows the fitted broken power-law.

$k_\perp$. We fit a broken power-law (see Fig. A1) as defined by[k]

$$A \frac{k_\perp}{k_{\text{break}}}^{-\alpha_1} \left[0.5 \left(1 + \left(\frac{k_\perp}{k_{\text{break}}} \right)^{\frac{1}{\Delta}} \right) \right]^{(\alpha_1 - \alpha_2)\Delta}, \qquad \text{(A1)}$$

where A is an amplitude factor, k_{break} is the $k_\perp$ value where the power-law breaks, α_1 is the power-law index for $k_\perp < k_{\text{break}}$, α_2 is the power-law index for $k_\perp > k_{\text{break}}$, and Δ is the sharpness of the break. We set $\alpha_1 = 0$ as the gridding density plateaus at low $k_\perp$, and set $A = 0.37728$, the median of the ratios where $k_\perp < 0.1\,h\,\text{Mpc}^{-1}$ from Fig. 4d. We obtain α_2 by performing a least squares fit with a single power-law model $Ck_\perp^{\alpha_2}$, where C is some constant, on ratios where $k_\perp > 0.07\,h\,\text{Mpc}^{-1}$. We can then find k_{break} by finding where the power-law is equal to the amplitude A:

$$A = Ck_{\text{break}}^{\alpha_2}, \qquad \text{(A2)}$$

$$k_{\text{break}} = 10^{\frac{\log_{10}(A) - C}{\alpha_2}}. \qquad \text{(A3)}$$

We fit this function to the median value of the ratios for each $k_\perp$ bin, as we found the values in each bin to have non-zero skew. Taking the median should therefore be a better representation of the data rather than the mean. We summarise the fitted parameters in Table A1. To create a normalisation function, we simply take the inverse of this broken power-law.

We compare how well this fit predicts the observed ratios in Fig. A2. Here, we plot the ratios as a function of the gridding weights, rather than $k_\perp$. We use the seaborn (Waskom 2021) kdeplot[l] function to generate two-dimensional kernel density estimates (KDE) to show the distribution of observed ratios in blue. We also plot the KDE of the ratios predicted by the fitted broken power-law in orange. We see that the fitted function predicts values close to the median value at high gridding weights. As data with a high gridding weight is naturally up-weighted, these data points contribute more to the estimated power in the 1D PS. The net result of this is shown in Fig. A3, where we compare the recovered 1D PS from the Gaussian noise simulation when correcting with the single scalar normalisation factor, and the fitted

---

[k]See https://gist.github.com/cgobat/12595d4e242576d4d84b1b682476317d for more information on the functional form.

[l]https://seaborn.pydata.org/generated/seaborn.kdeplot.html.

Table A1. Parameters for broken power-law fit.

Parameter	Value
a_1	0.0
Δ	0.01
A	0.377280447218
a_2	-0.161193559625
k_{break}	0.064354113566

Figure A3. The recovered 1D PS from the Gaussian noise simulation when correcting with the single scalar normalisation factor, and the fitted functional form as shown in Fig. A1.
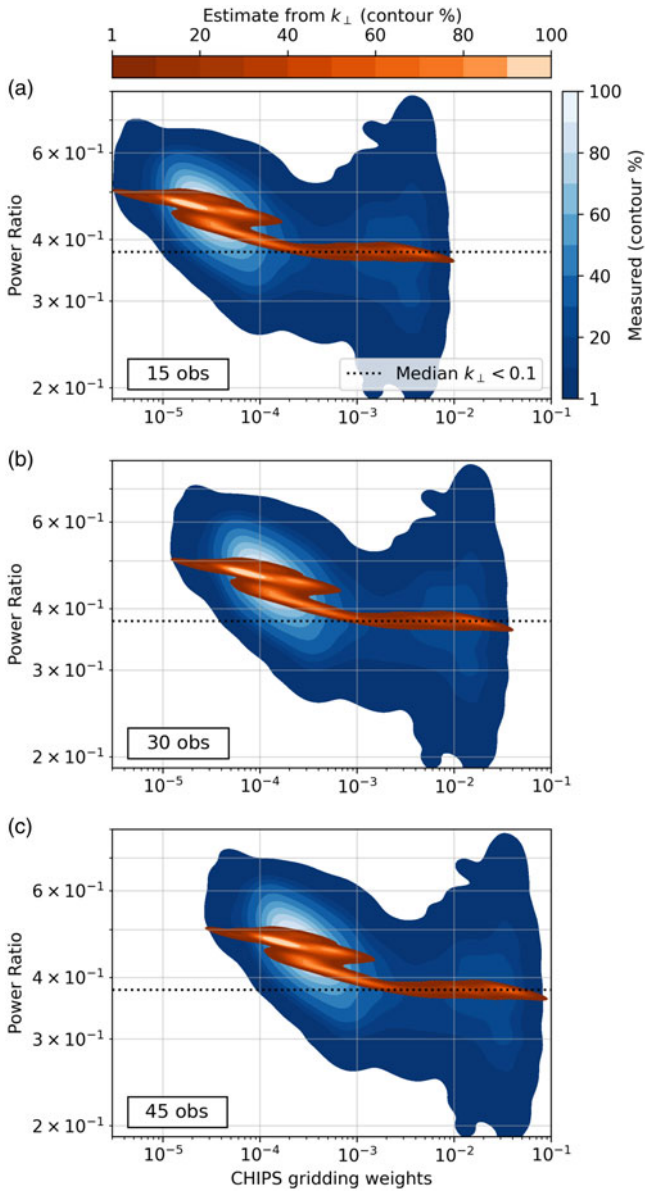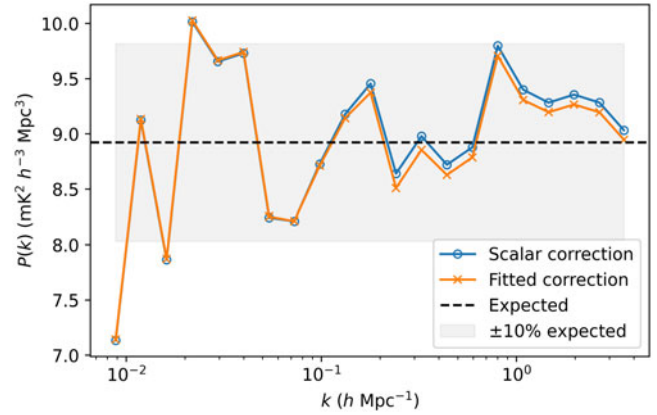

Figure A2. Kernel density estimates of the recovered ratio of the noise simulation as a function of CHIPS gridding weights (blue), and those predicted by fitted broken power-law (orange) for: (a) the 15 observation integration of all LSTS; (b) 30 observations; (c) 45 observations.

functional form. No difference is seen at low k, with minimal difference seen at high k. We conclude that as the functional form provides little difference to any limits, the scalar normalisation factor is sufficient for correcting for gridding density.