The Astrophysical Journal Supplement Series, 266:23 (17pp), 2023 June https://doi.org/10.3847/1538-4365/acc324

Statistical Recovery of 21 cm Visibilities and Their Power Spectra with Gaussian-constrained Realizations and Gibbs Sampling

Fraser Kennedy[1], Philip Bull[2,3], Michael J. Wilensky[2], Jacob Burba[2], and Samir Choudhuri[4]
[1] Astronomy Unit, Queen Mary University of London, Mile End Road, London, E1 4NS, UK; f.kennedy@qmul.ac.uk
[2] Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester, M13 9PL, UK
[3] Department of Physics and Astronomy, University of Western Cape, Cape Town, 7535, South Africa
[4] Centre for Strings, Gravitation and Cosmology, Department of Physics, Indian Institute of Technology Madras, Chennai, 600036, India
Received 2022 November 8; revised 2023 March 8; accepted 2023 March 8; published 2023 May 23

Abstract

Radio interferometers designed to probe the 21 cm signal from Cosmic Dawn and the Epoch of Reionization must contend with systematic effects that make it difficult to achieve sufficient dynamic range to separate the 21 cm signal from foreground emission and other effects. For instance, the instrument's chromatic response modulates the otherwise spectrally smooth foregrounds, making them difficult to model, while a significant fraction of the data must be excised due to the presence of radio-frequency interference, leaving gaps in the data. Errors in modeling the (modulated and gappy) foregrounds can easily generate spurious contamination of what should otherwise be 21 cm signal-dominated modes. Various approaches have been developed to mitigate these issues by, for example, using nonparametric reconstruction of the foregrounds, in-painting the gaps, and weighting the data to reduce the level of contamination. We present a Bayesian statistical method that combines these approaches, using the coupled techniques of Gaussian-constrained realizations and Gibbs sampling. This provides a way of drawing samples from the joint posterior distribution of the 21 cm signal modes and their power spectrum in the presence of gappy data and an uncertain foreground model in a computationally scalable manner. The data are weighted by an inverse covariance matrix that is estimated as part of the inference, along with a foreground model that can then be marginalized over. We demonstrate the application of this technique on a simulated Hydrogen Epoch of Reionization Array–like delay spectrum analysis, comparing three different approaches for accounting for the foreground components.

Unified Astronomy Thesaurus concepts: Bayesian statistics (1900); Reionization (1383); Interferometry (808)

1. Introduction

The redshifted 21 cm emission line from neutral hydrogen is much anticipated as a sensitive probe of the thermal history of the first billion years or so of cosmic time, particularly the periods when the first stars and galaxies formed (Cosmic Dawn) and then reionized the intergalactic medium (the Epoch of Reionization, or EoR; Madau et al. 1997; Furlanetto et al. 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012). A number of radio interferometers have been designed specifically to measure the statistical fluctuations in this signal, which arises in the approximate redshift range $6 \lesssim z \lesssim 27$, corresponding to frequencies of ∼50–200 MHz. Examples of those searching for the signal from the EoR include the Giant Meterwave Radio Telescope (Swarup et al. 1991; Paciga et al. 2013), the Murchison Wide-field Array (Tingay et al. 2013; Wayth et al. 2018; Trott et al. 2020), the Low Frequency Array (van Haarlem et al. 2013; Patil et al. 2017; Mertens et al. 2020), the Long Wavelength Array (Hallinan 2014; Garsden et al. 2021), the Precision Array to Probe the Epoch of Reionization (Parsons et al. 2010; Kolopanis et al. 2019), and the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017; HERA Collaboration et al. 2022). The usual goal of these experiments is to build up the large sensitivity required to make a statistical detection of the power spectrum of the brightness-temperature fluctuations of the 21 cm line as a function of redshift, by observing for long periods of time with large numbers of receiving elements and baselines. The presence of bright foreground emission places extremely stringent requirements on the fidelity of the instrumental calibration required to detect the 21 cm signal, however (Barry et al. 2016; Patil et al. 2016; Ewall-Wice et al. 2017; Gehlot et al. 2018; Joseph et al. 2018; Byrne et al. 2019; de Gasperin et al. 2019; Mouri Sardarabadi & Koopmans 2019), with other systematic effects such as mode mixing (e.g., Morales et al. 2012), nonredundancy (e.g., Orosz et al. 2019; Joseph et al. 2020; Choudhuri et al. 2021), in-painting artifacts (e.g., Chakraborty et al. 2022; Pagano et al. 2023), reflection and coupling artifacts (e.g., Ewall-Wice et al. 2016; Kern et al. 2019), as well as polarization leakage and the ionosphere (Gehlot et al. 2018; Kariuki Chege et al. 2022) needing to be handled carefully.

The foreground emission is particularly challenging because of the high dynamic range between it and the target 21 cm signal, which is expected to be somewhere in the region of 10^4–10^5 times fainter (in temperature). This in itself is not too problematic if the foreground signal is well segregated from the 21 cm signal when projected onto a suitable basis, such that the foreground emission can be localized to a handful of modes that can then be modeled or filtered out. Since the dominant source of foreground emission is synchrotron radiation, which has a smooth (power-law) frequency spectrum, an effective localization of the foregrounds into only the smoothest components of a harmonic/Fourier basis could reasonably be achieved (Di Matteo et al. 2002; Santos et al. 2005; Ali et al. 2008). This is unfortunately disrupted by the complicated spectral response of the instruments themselves, which

modulate the sky signal, introducing additional spectral features that are generally difficult to model with sufficiently high precision. As a result, the instrumental response (and errors in its calibration) "scatters" foreground emission outside of its intrinsic localization region, with even a small amount of scattering (say, at the 10^{-4} level) able to swamp the 21 cm emission. Even with precise instrumental calibration, a large wedge-shaped region of foreground contamination is introduced into the 2D Fourier space of the interferometric visibilities by instrumental effects (Datta et al. 2010; Morales et al. 2012; Pober et al. 2013; Thyagarajan et al. 2015), and must either be filtered out ("foreground avoidance"; e.g., Liu et al. 2014a, 2014b; Thyagarajan et al. 2015) or modeled and subtracted ("foreground mitigation"; e.g., Di Matteo et al. 2002; Santos et al. 2005; Morales et al. 2006; Bowman et al. 2009; Chapman et al. 2016).

The difficulty of handling the foreground contamination makes it especially important to avoid analysis steps that could introduce further scattering of the foregrounds, for example through coupling of Fourier modes inside and outside the wedge. Unfortunately, a number of other systematic effects can naturally cause scattering of this kind. Spurious artificial radio emission (radio-frequency interference, or RFI) poses a particular challenge to low-frequency experiments, because many transmitters operate inside the experimental band, leaving extremely bright signals in (typically narrow) regions in frequency that effectively render that part of the data unrecoverable during the transmitter's time of operation. These regions must be aggressively identified and masked to prevent substantial contamination of the data (see, e.g., Barry et al. 2019; Li et al. 2019; Mertens et al. 2020; Wilensky et al. 2020). The mask itself then becomes the issue: masking introduces discontinuous jumps in the data, which are problematic for steps of the analysis that rely on harmonic transforms, such as the Fourier transforms that are used during power spectrum estimation (Offringa et al. 2019; Wilensky et al. 2022). When a Fourier transform is applied to a step or spike, ringing artifacts are generated that strongly couple modes inside the foreground wedge to those outside, causing widespread leakage of the foreground emission across the rest of the Fourier space.

Methods must therefore be found that can prevent or suppress the ringing in order to successfully measure the power spectrum. These typically fall into two camps: *inpainting*, which replaces the missing/masked data with a plausible model (e.g., CLEAN, DAYENU, and Gaussian process regression; Parsons & Backer 2009; Offringa et al. 2019; Mertens et al. 2020; Ewall-Wice et al. 2021; Kern & Liu 2021; Chakraborty et al. 2022), and harmonic analysis methods (e.g., least-squares spectral analysis; Trott et al. 2016; Patil et al. 2017; Chakraborty et al. 2022) that account for nonuniformly sampled data. The general goal of these methods is to allow the 21 cm power spectrum to be recovered in an unbiased manner, without ringing artifacts. Beyond this, other desirable features may be sought, such as minimizing the variance of the power spectrum estimates, avoiding strong model dependence, enhancing the interpretability of the power spectrum estimate, computational efficiency, or correct propagation of uncertainty.

As a final complication, we would also like to apply an inverse covariance weighting to the data in order to recover an optimal power spectrum estimate while also accounting for correlations and mode mixing. This requires a high-fidelity estimate of the true covariance matrix of the data, which is generally not available. Due to the inherently high dynamic range, the signal eigenmodes in the data covariance matrix have small eigenvalues and are easily misestimated. Using the empirical covariance matrix measured from the data as an estimate leads to signal loss in the quadratic estimator formalism, since the estimated signal becomes a quartic (rather than quadratic) function of the data (Kolopanis et al. 2019). Simulations rely on empirical sky models that are incomplete.

Solving these problems requires a method that is able to account properly for missing frequency-space data, model the foregrounds (as corrupted by the instrument), and weight the data in an optimal way. In this paper, we present a Bayesian signal recovery method that aims to capture the statistical interactions between the components of the data in order to achieve these aims. Our goal is to estimate the joint posterior distribution of a model that is sufficiently flexible to recover the 21 cm signal and foregrounds without requiring strong model assumptions. From the joint posterior, we can then derive best-fit models of the components, their uncertainties, and any correlations between them.

Flexible models typically require large numbers of parameters. Our method is based on Gibbs sampling (Geman & Geman 1984), which provides a way of sampling from the joint posterior distribution of a model with many parameters by iteratively sampling from a set of more tractable conditional distributions instead. This has been used to good effect in cosmic microwave background (CMB) inference problems such as foreground separation (Wandelt et al. 2004; Eriksen et al. 2008), and has also recently been applied to 21 cm data at lower redshift to estimate power spectra in the presence of masked data (The CHIME Collaboration et al. 2023). We follow the structure of the method in Eriksen et al. (2008), which used a Gibbs sampling scheme to recover the joint posterior distribution of the CMB signal field and covariance, as well as various foreground parameters. In our case, we aim to recover the joint posterior distribution of the (baseline-dependent) EoR 21 cm signal visibilities, their power spectra, and a foreground model in the presence of missing frequency channels in the data. We compare three Gibbs sampling implementations designed to achieve that goal:

1. A "total signal" sampler, which does not differentiate between the EoR signal and the foregrounds, aiming to produce an estimate of the total EoR plus foreground delay spectrum.
2. A sampler which models foregrounds using the eigenmodes of simulated foreground covariance matrices.
3. A sampler which jointly samples the foregrounds along with the signal, conditional on the foreground covariance matrix being known.

We apply these methods to simulations of visibility data with realistic point-source foregrounds, a simple Gaussian EoR signal model, and various patterns of gaps to mimic the effect of RFI flagging. We test these models under different noise conditions, and also examine their robustness to incomplete sky models used to generate the priors/models for the foreground component.

The structure of the paper is as follows. In Section 2, we provide some grounding in Bayesian methods relevant to Gibbs sampling, namely Wiener filtering and the Gaussian-constrained realization (GCR) equation, and detail three examples

of Gibbs sampling implementations that can be used to recover the EoR signal. In Section 3, we describe the visibility simulations we use in the rest of our study. In Section 4, we present the performance of each of the Gibbs sampling implementations on our simulated data under different flagging and noise conditions, and compare the achieved recovery in each case. We conclude in Section 5.

2. Bayesian Recovery of the 21 cm Signal and its Power Spectrum

In this section we construct Gibbs sampling implementations capable of sampling from the full joint posterior distribution of the signal model, its power spectrum, and a set of foreground parameters. We begin by describing our model for visibility data and writing down the posterior distribution for the signal, and then demonstrate some of the key components of the Gibbs sampler using a hierarchy of Bayesian methods of increasing complexity, running from a simple maximum a posteriori (MAP) solution, to sampling from the signal distribution conditional on the signal covariance, to sampling from the full joint posterior distribution with a Gibbs sampler. We then detail three implementations that differ in their handling of the foreground component. Our implementation structures and notation follow those in Eriksen et al. (2008).

2.1. Data Model and Posterior

We model visibility data $\boldsymbol{V}$, which are complex valued, as

$$V_{mn}(\nu, t) = w(\nu, t)[s_{mn}(\nu, t) + n_{mn}(\nu, t)], \tag{1}$$

where the indices m, n label the antennas used to form each visibility; ν, τ, and t label frequency, delay (the Fourier conjugate to frequency), and observation time, respectively; $\boldsymbol{w}$ is a mask vector with values of 1 (unflagged) or 0 (flagged); $\boldsymbol{n}$ is a Gaussian noise component with covariance $\boldsymbol{N} \equiv \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle$; and $\boldsymbol{s}$ is the total (signal+foreground) component in the data space. It is also useful to express the signal component in a discrete Fourier transform (DFT) basis as $\boldsymbol{s} = \boldsymbol{T}\tilde{\boldsymbol{s}}$, where $\boldsymbol{T}$ is a DFT matrix operator[5] and $\tilde{\boldsymbol{s}}$ are coefficients of the total signal +foreground component in the Fourier basis. We use bold symbols to denote vector quantities and upper-case letters to denote matrices. In general, the total signal will be comprised of multiple components such as foregrounds, which we label $\boldsymbol{f}$, the 21 cm signal, labeled $\boldsymbol{e}$, and instrumental systematics (which we neglect here). We assume that these components carry independent information, so that the total signal $\boldsymbol{s}$ and the total signal covariance $\boldsymbol{S}$ can be written

$$\boldsymbol{s} = \boldsymbol{e} + \boldsymbol{f}, \tag{2}$$

$$\boldsymbol{S} \equiv \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle = \langle \boldsymbol{e}\boldsymbol{e}^\dagger \rangle + \langle \boldsymbol{f}\boldsymbol{f}^\dagger \rangle = \boldsymbol{E} + \boldsymbol{F}. \tag{3}$$

We would now like to find a way to estimate the joint posterior distribution of the model, $p(\boldsymbol{s}, \boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{d})$, conditioned on the measured data $\boldsymbol{d}$, across all components of the signal and their covariance, and any other parameters $\boldsymbol{\theta}$ that may be considered in the analysis. The joint posterior contains not just estimates for $\boldsymbol{s}$ and $\boldsymbol{S}$ but also complete information about statistical uncertainties and correlations between parameters. In our case it is a function of a large number of parameters, however, including the values of the signal/foreground visibilities at each frequency channel, time, and baselines, and the elements of each covariance matrix. It would therefore be prohibitively expensive to explore the posterior directly due to the large number of dimensions. In the next sections, we outline three approaches of increasing complexity to handle the Bayesian estimation of the parameters in our model.

2.2. Maximum A Posteriori Solution (Wiener Filter)

As a first step, consider the posterior distribution conditional on known covariance information. We consider the case of a single baseline and drop the antenna labeling indices m, n. Using Bayes' theorem, the signal's posterior distribution conditional on known covariances $\boldsymbol{S}$, $\boldsymbol{N}$ and measured data $\boldsymbol{d}$ is

$$p(\boldsymbol{s}|\boldsymbol{S}, \boldsymbol{N}, \boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{s}, \boldsymbol{S}, \boldsymbol{N})p(\boldsymbol{s}|\boldsymbol{S}). \tag{4}$$

The second right-hand side distribution is a prior term for the signal $\boldsymbol{s}$ given the data-space signal covariance, $\boldsymbol{S}$, which will generally be independent of the data and the noise covariance. We assume our noise to be Gaussian distributed, which leads to the following conditional distribution:[6]

$$p(\boldsymbol{s}|\boldsymbol{S}, \boldsymbol{N}, \boldsymbol{d}) \propto e^{-(\boldsymbol{d}-\boldsymbol{s})^\dagger \boldsymbol{N}^{-1}(\boldsymbol{d}-\boldsymbol{s})} e^{-\boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s}}. \tag{5}$$

Under the assumption of Gaussianity, the maxima of the posterior and log-posterior occur at the same location. To obtain the MAP estimate of the signal $\hat{\boldsymbol{s}}$, the first derivative of the logarithm of $p(\boldsymbol{s}|\boldsymbol{S}, \boldsymbol{N}, \boldsymbol{d})$ can be set to zero,

$$\frac{\partial}{\partial \boldsymbol{s}}\bigg|_{\boldsymbol{s}=\hat{\boldsymbol{s}}} ((\boldsymbol{d} - \boldsymbol{s})^\dagger \boldsymbol{N}^{-1}(\boldsymbol{d} - \boldsymbol{s}) + \boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s}) = 0, \tag{6}$$

to obtain

$$\boldsymbol{d}^\dagger \boldsymbol{N}^{-1} = \hat{\boldsymbol{s}}^\dagger \boldsymbol{N}^{-1} + \hat{\boldsymbol{s}}^\dagger \boldsymbol{S}^{-1}. \tag{7}$$

Using the fact that covariance matrices are Hermitian to remove complex conjugation, this equation can be rearranged to the "generalized Wiener filter" for the signal component $\boldsymbol{s}$:

$$\boldsymbol{s}_{\rm wf} = [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1}\boldsymbol{N}^{-1}\boldsymbol{d}. \tag{8}$$

In essence, the Wiener filter answers the question: "Given an assumption about the signal and noise covariance information, what is the most likely form of the realized signal in this data?" Despite being the MAP solution for the signal component of a data vector, the expectation of the solution $\langle \boldsymbol{s}_{\rm wf} \rangle \neq \boldsymbol{s}$ in general (e.g., Rybicki & Press 1992), though this bias can be partially ameliorated. The covariance of the Wiener filter estimate is also strictly smaller than the covariance of the parent conditional distribution. These biases are discussed in Appendix A. Defining the Wiener filter operation from Equation (8) as

$$\boldsymbol{G} \equiv [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1}\boldsymbol{N}^{-1}, \tag{9}$$

and returning to the conditional Equation (5), the prior and likelihood terms can be combined by completing the square to

[5] $T_{mn} = e^{-2\pi i mn/N_{\rm freq}}$. We use a Fourier convention such that $\boldsymbol{T}^\dagger\boldsymbol{T} = \boldsymbol{I}$, where the † symbol denotes the Hermitian (conjugate) transpose.

[6] Note that these are complex Gaussian distributions, which do not have a factor of 1/2 in the exponent (or a square root of the determinant in the normalization factor) when written in complex vector form (Gallager 2013).

form a single Gaussian:

$$p(\boldsymbol{s}|\boldsymbol{S}, \boldsymbol{N}, \boldsymbol{d}) \propto \exp\left(-(\boldsymbol{s} - \boldsymbol{Gd})^{\dagger}(\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1})(\boldsymbol{s} - \boldsymbol{Gd})\right), \tag{10}$$

where the matrix identity $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{A}[\boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}]\boldsymbol{B}$ has been used, and data-only terms that are constant in the posterior have been absorbed as constants of proportionality. Here we see how the Wiener-filtered data $\boldsymbol{s}_{\rm wf} = \boldsymbol{Gd}$ represents the mean of the distribution, and therefore that the Wiener filter minimizes the expected residual $\langle|\boldsymbol{s} - \boldsymbol{Gd}|\rangle$ between the true signal and the signal estimate.[7] We also see that the conditional distribution has covariance $[\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1}$.

The data contain flagged regions which have been set to zero due to the presence of RFI. Since we do not have access to the information underneath the RFI mask, we set the noise variance in those regions to infinity. In implementation terms, this can be handled by using an amended form of the inverse noise covariance $\tilde{\boldsymbol{N}}^{-1}$:

$$\tilde{\boldsymbol{N}}^{-1} = \boldsymbol{w}\boldsymbol{w}^{\rm T} \circ \boldsymbol{N}^{-1}, \tag{11}$$

where $\boldsymbol{w}$ is the mask vector and $\circ$ denotes element-wise multiplication. Substituting $\tilde{\boldsymbol{N}}^{-1}$ for $\boldsymbol{N}^{-1}$ in Equation (8) amounts to zeroing the contribution from the data inside the mask ($\boldsymbol{N}^{-1}\boldsymbol{d}$ term). The inverse signal covariance term does *not* typically go to zero in these regions; the prior "takes over" signal estimation in lieu of information from the likelihood function. We should clarify that this is *not* simply a matter of filling in the masked regions with draws from the prior: the Wiener filter inside the masked regions is constrained to match up with the solution in the unmasked regions, and close to the mask boundary both the prior and the data contribute to the solution.

2.3. Gaussian-constrained Realizations from the Conditional Distribution of the Epoch of Reionization Signal

Having the MAP solution for the signal component in hand, the next step in the Bayesian hierarchy is accessing the full conditional distribution $p(\boldsymbol{s}|\boldsymbol{S}, \boldsymbol{N}, \boldsymbol{d})$ and generating samples from it. To generate samples from the conditional, we make use of the GCR equation, which we now describe. Since Equation (10) defines a multivariate Gaussian distribution, realizations drawn from the distribution (denoted $\boldsymbol{s}_{\rm cr}$) may be generated by adding random normal (Gaussian) realizations of the signal and noise fluctuation terms to the Wiener filter equation that are scaled correctly by their respective covariances (Eriksen et al. 2008), i.e.,

$$\boldsymbol{s}_{\rm cr} = [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1}[\boldsymbol{N}^{-1}\boldsymbol{d} + \boldsymbol{S}^{-1/2}\boldsymbol{\omega}_0 + \boldsymbol{N}^{-1/2}\boldsymbol{\omega}_1]. \tag{12}$$

The new terms in the right bracket as compared with Equation (8) are independent realizations of zero mean, unit variance Gaussian random vectors, $\boldsymbol{\omega}$, scaled by the noise and signal covariances.[8] With this form, it is straightforward to check that the covariance of a GCR solution is equal to the covariance of the Gaussian distribution in Equation (10). Many realizations of the signal consistent with the given data vector $\boldsymbol{d}$ and covariances $\boldsymbol{S}$, $\boldsymbol{N}$ can be generated by solving Equation (12) repeatedly with different random realizations of the vectors $\boldsymbol{\omega}$, and these realizations trace the full conditional distribution.

Depending on the size of the data and the particular forms of the covariance matrices, solving Equations (8) and (12) may become computationally demanding. A useful technique is to "precondition" the linear system by multiplying through by an easily computable factor that makes the linear operator closer to the identity matrix (which would give a trivial solution to the linear system). In other words, for a linear system $\boldsymbol{Ax} = \boldsymbol{b}$, preconditioned as $\boldsymbol{PAx} = \boldsymbol{Pb}$, the best preconditioning matrix $\boldsymbol{P}$ is one that gives $\boldsymbol{PA} \approx \boldsymbol{I}$, and where computing $\boldsymbol{PA}$ is quite fast. As pointed out by Eriksen et al. (2008), an effective preconditioner for the Wiener filter and GCR equations can be obtained by multiplying through by $\boldsymbol{S}^{1/2}$, enabling much faster convergence in practice. With this preconditioning scheme, Equation (12) becomes

$$[\boldsymbol{I} + \boldsymbol{S}^{1/2}\tilde{\boldsymbol{N}}^{-1}\boldsymbol{S}^{1/2}]\boldsymbol{y}_{\rm cr} = \boldsymbol{S}^{1/2}\tilde{\boldsymbol{N}}^{-1}\boldsymbol{d} + \boldsymbol{\omega}_0 + \boldsymbol{S}^{1/2}\tilde{\boldsymbol{N}}^{-1/2}\boldsymbol{\omega}_1. \tag{13}$$

The $\boldsymbol{y}_{\rm cr}$ solution vector can be obtained using a conjugate gradient solver, from which the solution for the signal component can then be found as $\boldsymbol{s} = \boldsymbol{S}^{1/2}\boldsymbol{y}_{\rm cr}$.

Illustrative results from the GCR solver are shown in Figure 1, with a comparatively large flagged region close to the middle of the band, plus some randomly flagged channels, for a total flag fraction of 15%. This example has been run as part of a full Gibbs scheme (Scheme 2; see Section 2.6 below), which runs for 800 iterations and includes sampling of the EoR and foreground covariance matrices/model parameters. The upper-right panel of Figure 1 shows an essentially seamless in-painting, with no visible discontinuity between the unflagged and in-painted regions. This continuity is due to the solution inside the flagged region being conditioned on the data outside, as well as the frequency structure of the EoR/foreground models and their covariances.

The lower-left panel of Figure 1 shows the difference between the absolute value of the mean (over 800 samples) of the GCR-sampled EoR plus foreground model with the true (input) EoR plus foreground model used in the simulation. The difference is small outside the flagged regions, as expected when fitting an accurate model to data with high signal-to-noise ratio (S/N; for reference, the mean noise rms on the real and imaginary parts of the visibilities is 0.065, whereas the input EoR signal has an rms of 0.52 in the units of Figure 1). It is larger inside the flagged regions, but only shows structure that is nonsmooth with frequency. This is also expected: the data outside the flagged region constrain the possible behaviors of the EoR signal plus foregrounds inside the region, but do not fully specify them, particularly as one moves further from the edge of the region. The EoR signal in particular is allowed to be nonsmooth in frequency, and so the solution inside the flagged region need not be strongly correlated with the solution outside. In any case, there is no clear bias toward either over- or underestimating the combined EoR plus foreground signal inside the flagged region; as we will show later, the power

[7] The Wiener filter yields equivalent estimates to Gaussian process regression (GPR) for a given covariance matrix/kernel (Särkkä & Solin 2013), hence the similarities between the equations here and the ones for GPR methods (e.g., Kern & Liu 2021).

[8] Since the visibility data are complex valued, so too must be the random vectors $\boldsymbol{\omega}$. In order to ensure that they have unit variance, we sum draws for unit variance real and complex parts, then divide this sum by $\sqrt{2}$.

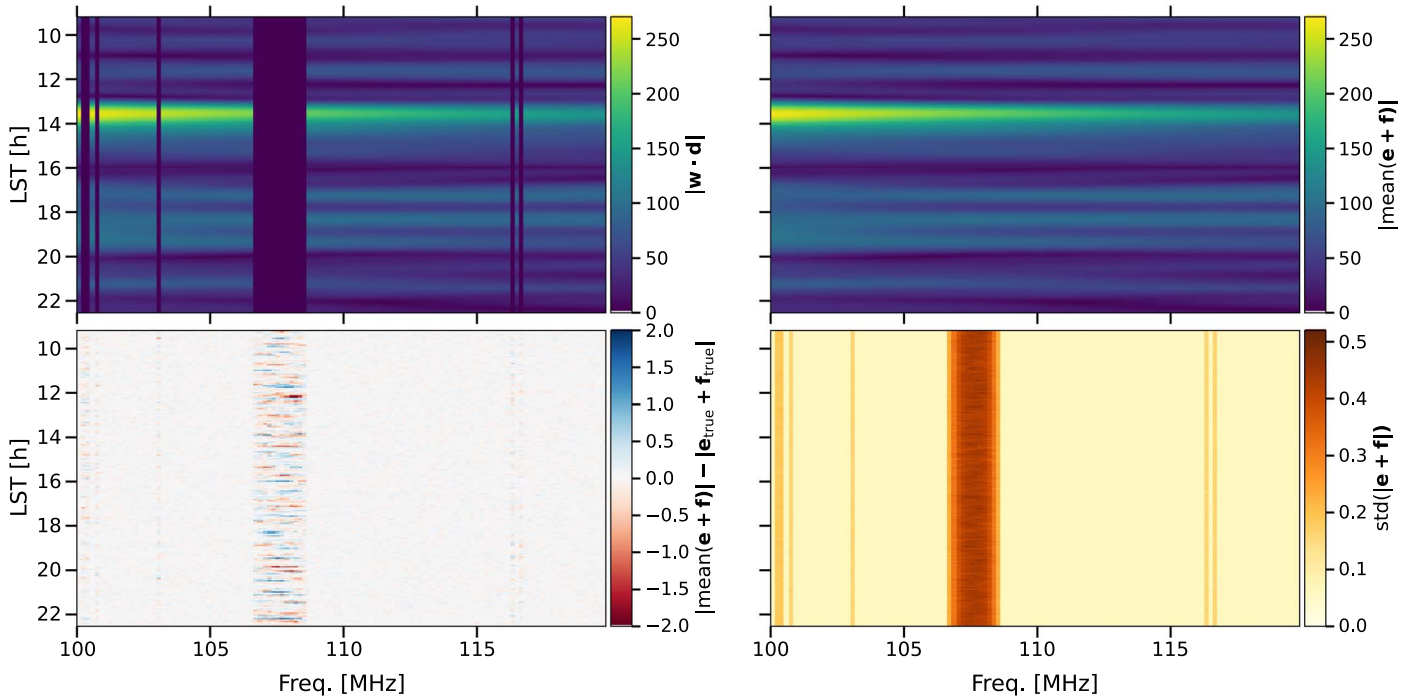

Figure 1. An example set of in-painted visibilities using Scheme 2, for a single 14.6 m E–W baseline. Fifteen percent of the band has been flagged, with a combination of a broad region and several randomly flagged channels. The units of the visibilities are given as janskys. Upper left: amplitude (absolute value) of the "observed" visibilities, including flagged regions. Upper right: amplitude of the mean of the EoR plus foreground realizations from the sampler (i.e., the average over all 800 samples). Lower left: difference between the mean of the EoR plus foreground realizations (upper-right panel) and the true (input) EoR plus foreground model used in the simulation. Lower right: standard deviation of the amplitude of the EoR plus foreground realizations from the sampler (calculated over all 800 samples).

spectrum of the EoR component in particular is recovered in an unbiased manner.

The lower-right panel of Figure 1 shows the variability in the EoR plus foreground realizations, plotted as the standard deviation of the absolute value of these quantities over 800 samples. It can be seen that the variability is lower for narrow flagged regions, as only relatively smaller deviations are allowed in order for the solution to remain consistent with the surrounding data. The variability is larger inside the broad flagged region, but shows a smooth transition from the edge to the center, again due to the solution being constrained more strongly by neighboring data near the edge of the region. Note that the variability of the realizations is also nonzero outside the flagged regions. This is driven mostly by the noise level; neither the EoR nor foreground solutions are completely certain in the unflagged regions, even though the S/N is reasonably large.

2.4. Realizations from the Joint Posterior (Gibbs)

Gibbs sampling is a method for recovering the joint posterior distribution, in our case $p(\boldsymbol{s}, \boldsymbol{S}|\boldsymbol{d})$, via Markov Chain Monte Carlo (MCMC) sampling (Geman & Geman 1984). Under the condition that a given joint probability density is strictly positive across the span of each variable (i.e., that no point in the joint space has zero probability density), then that joint density is specified uniquely by the full set of conditional distributions for all the parameters. Gibbs sampling uses this fact by sampling from each conditional distribution in turn, in the process updating conditioned-on variables with the sample obtained for them at the previous iteration. Since the joint posterior we wish to evaluate is a function of two (vector) quantities, the signal $\boldsymbol{s}$ and the signal covariance $\boldsymbol{S}$, the joint posterior can be evaluated by sampling (indicated by $\leftarrow$) from each of the two conditional distributions iteratively:

$$\boldsymbol{s}_{i+1} \leftarrow p(\boldsymbol{s}_i|\boldsymbol{S}_i, \boldsymbol{N}, \boldsymbol{d}), \tag{14}$$

$$\boldsymbol{S}_{i+1} \leftarrow p(\boldsymbol{S}_i|\boldsymbol{s}_{i+1}). \tag{15}$$

In the above representation, the top line is sampled first, followed by the second line, with the index i running over iterations. The distribution for the signal covariance $\boldsymbol{S}$ is not conditioned on the noise covariance $\boldsymbol{N}$ and data $\boldsymbol{d}$ since all of the relevant information is contained in the current realization of $\boldsymbol{s}$. Equation (14) has the form of a multivariate Gaussian, and so sampling is achieved by making use of the GCR equation, Equation (12). The conditional distribution of the covariance, Equation (15), on the other hand, has the form of a complex inverse Wishart distribution:

$$p(\boldsymbol{S}|\boldsymbol{s}) = \frac{p(\boldsymbol{s}|\boldsymbol{S})p(\boldsymbol{S})}{p(\boldsymbol{s})} \propto \frac{1}{\det(\boldsymbol{S})} \exp\left(-\boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s}\right), \tag{16}$$

where "det (...)" signifies the matrix determinant (recall that there is no square root of the determinant when written in complex vector notation; see footnote 6). Since we solve for full realizations of the signal component at the GCR step (i.e., that have values both inside and outside regions masked due to RFI), the need for an in-painting process that explicitly estimates the missing data is eliminated.

In the next three subsections, we detail a series of Gibbs sampling implementations that treat the foreground component in different ways.

2.5. Scheme 1 (Combined Signal+Foreground Sampler)

The first scheme is a straightforward implementation sampling the total signal from a complex Gaussian distribution ($\mathcal{CN}$) using the GCR equation (i.e., sampling $\boldsymbol{s} = \boldsymbol{e} + \boldsymbol{f}$), followed by sampling the total covariance matrix $\boldsymbol{S}$ from the complex inverse Wishart ($\mathcal{CW}^{-1}$) distribution:

$$\boldsymbol{s}_{i+1} \leftarrow \mathcal{CN}(\boldsymbol{Gd}, [\boldsymbol{S}_i^{-1} + \boldsymbol{N}^{-1}]), \tag{17}$$

$$\boldsymbol{S}_{i+1} \leftarrow \mathcal{CW}^{-1}(\Sigma_{i+1}, \nu_f, N_{\rm freq}), \tag{18}$$

where the sample covariance matrix is

$$\Sigma_{i+1} = \frac{1}{N_{\rm times} - 1} \sum_{t}^{N_{\rm times}} \boldsymbol{s}_{t,i+1} \boldsymbol{s}_{t,i+1}^{\dagger}, \tag{19}$$

the parameter $\nu_f = N_{\rm freq}(N_{\rm freq} + 1)/2$ denotes the number of degrees of freedom in the sample covariance, and $N_{\rm times}$ is the number of local sidereal times (LSTs) in the visibility data. We discuss obtaining samples from a complex inverse Wishart distribution in Appendix B. In each iteration of the Gibbs sampler the GCR equation (Equation (17) above) is solved using visibility data across all LSTs separately, using as a prior the covariance matrix arrived at during the last iteration. The mean covariance used for sampling from the distribution in Equation (18) is the sum over outer products of the obtained GCR solutions from each LST at the previous step. This sampling scheme will arrive at realizations of both the total signal and total data covariance in frequency space.

2.6. Scheme 2 (Joint Sampler with Foreground Templates)

In the second implementation, we make two simplifying assumptions: (i) the delay spectrum of the EoR signal component is diagonal, i.e., all of the statistical information about $\boldsymbol{s}$ is contained by its power spectrum,[9] and (ii) that the foreground covariance matrix $\boldsymbol{F}$ is known, and hence fixed (not sampled). This implementation follows a Gibbs sampling scheme for the CMB signal map and power spectrum (Eriksen et al. 2008). Foreground covariance matrices can be estimated directly from simulations based on past observations (see Section 3), and, as such, the foreground covariance is known to a higher degree of accuracy than the covariance of the EoR field. Fixing this quantity to the values arrived at via simulations significantly reduces the volume of parameter space to be covered by the sampler. Second, the cosmological signal $\boldsymbol{e}$ having a diagonal covariance in delay space (i.e., with variance given by the delay spectrum in each delay bin) reduces the complex inverse Wishart distribution of Equation (18) to a product of inverse-Gamma distributions, one for each delay spectrum bandpower.

As with the implementation in Eriksen et al. (2008), we make use of a set of foreground templates $\boldsymbol{g}_j(\nu)$ with respective amplitudes $\boldsymbol{a}_{\rm fg}$. For our templates, we use the first $N_{\rm pc}$ principal components of our simulation-derived foreground covariance matrix on each baseline such that the (nonsquare) template matrix has dimension $N_{\rm freq} \times N_{\rm pc}$. An example set of principal-component modes is shown in Figure 2. The data model in this

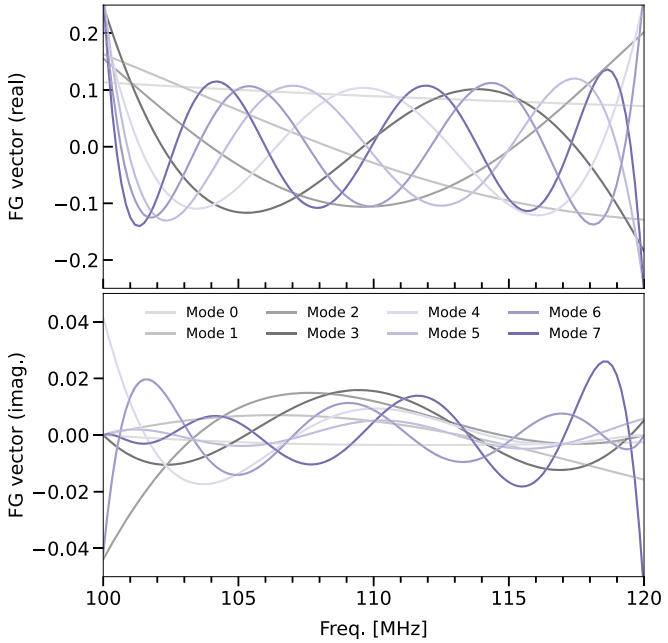

Figure 2. First eight foreground principal components (covariance eigenmodes) derived from the frequency–frequency covariance matrix measured from simulations of a 14.6 m E–W baseline. The upper and lower panels show the real and imaginary parts, respectively.

scheme is given by

$$\boldsymbol{d} = \boldsymbol{e} + \boldsymbol{g}_j \cdot \boldsymbol{a}_{\rm fg} + \boldsymbol{n}, \tag{20}$$

where $\boldsymbol{e}$ denotes the cosmological signal in frequency space. We follow Eriksen et al. (2008) in defining a vector $\boldsymbol{x}$ where the first $N_{\rm freq}$ entries are $\boldsymbol{e}$ and the second $N_{\rm pc}$ entries are $\boldsymbol{a}_{\rm fg}$, and a corresponding response vector $\boldsymbol{u} = (\boldsymbol{1}, \boldsymbol{g}_j)^{\rm T}$, so that the signal and foreground amplitude conditional distribution may be written

$$\begin{aligned} p(\boldsymbol{e}, \boldsymbol{a}_{\rm fg} | \boldsymbol{E}, \boldsymbol{g}_j, \boldsymbol{N}, \boldsymbol{d}) &\propto p(\boldsymbol{d} | \boldsymbol{e}, \boldsymbol{a}_{\rm fg}, \boldsymbol{g}_j, \boldsymbol{N}) p(\boldsymbol{e} | \boldsymbol{E}) \\ &\propto \exp\left(-(\boldsymbol{d} - \boldsymbol{x} \cdot \boldsymbol{u})^{\dagger} \boldsymbol{N}^{-1} (\boldsymbol{d} - \boldsymbol{x} \cdot \boldsymbol{u})\right) \exp\left(-\boldsymbol{s}^{\dagger} \boldsymbol{S}^{-1} \boldsymbol{s}\right) \\ &\propto \exp\left(-(\boldsymbol{x} - \hat{\boldsymbol{x}})^{\dagger} \boldsymbol{A}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{x}})\right), \end{aligned} \tag{21}$$

where the form of the vector $\hat{\boldsymbol{x}}$ and matrix $\boldsymbol{A}$ follow from a completed-square representation of the line above. The symbolic linear system to be solved, $\boldsymbol{Ax} = \boldsymbol{b}$, comparable to Equation (12), takes the explicit form

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{E}^{-1} + \boldsymbol{N}^{-1} & \boldsymbol{N}^{-1} \boldsymbol{g}_j \\ \boldsymbol{g}_j^{\dagger} \boldsymbol{N}^{-1} & \boldsymbol{g}_j^{\dagger} \boldsymbol{N}^{-1} \boldsymbol{g}_j \end{bmatrix} \begin{bmatrix} \boldsymbol{e} \\ \boldsymbol{a}_{\rm fg} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{N}^{-1} \boldsymbol{d} + \boldsymbol{E}^{-1/2} \boldsymbol{\omega}_0 + \boldsymbol{N}^{-1/2} \boldsymbol{\omega}_1 \\ \boldsymbol{g}_j^{\dagger} \boldsymbol{N}^{-1} \boldsymbol{d} + \boldsymbol{g}_j^{\dagger} \boldsymbol{N}^{-1/2} \boldsymbol{\omega}_1 \end{bmatrix}, \end{aligned} \tag{22}$$

where the foreground model can be recovered as $\boldsymbol{f} = \boldsymbol{g}_j \cdot \boldsymbol{a}_{\rm fg}$, and $\boldsymbol{E}$ is the EoR signal covariance in frequency space. The sampling is performed for each LST independently and in parallel, and so we obtain a set of N_t solution vectors $\boldsymbol{e}$ and $\boldsymbol{a}_{\rm fg}$ at each iteration.

The bandpowers of the EoR signal delay spectrum $\boldsymbol{p} = (P(\tau_0), \ldots, P(\tau_N))^{\rm T}$ for delay bins $\tau_0, \ldots, \tau_N$ are related to the frequency-space EoR covariance by $\boldsymbol{E} = \boldsymbol{T}\tilde{\boldsymbol{E}}\boldsymbol{T}^{\dagger}$, where the delay-space covariance $\tilde{\boldsymbol{E}}$ is zero everywhere except on the

[9] In making this assumption we automatically lose sensitivity to any non-Gaussian information in the measured EoR field. In practice, that means that such an assumption is fairly limiting since the EoR field is expected to be significantly non-Gaussian. Testing the extent to which this assumption limits us will require more sophisticated simulations than we employ in this work.

diagonals, where $\mathrm{diag}(\tilde{\boldsymbol{E}}) = \boldsymbol{p}$. Each bandpower is sampled independently from an inverse-Gamma distribution with a scale parameter calculated from the variance (over LST) of the current realization of the corresponding EoR signal delay mode, i.e., the variance for delay mode τ at iteration $i+1$ is

$$\sigma^2_{\tau,i+1} = \frac{1}{N_{\mathrm{times}} - 1} \sum_{t}^{N_{\mathrm{times}}} \tilde{e}^*_{\tau,t,i+1} \tilde{e}_{\tau,t,i+1}, \tag{23}$$

where $\tilde{\boldsymbol{e}}_t = \boldsymbol{T}^\dagger \boldsymbol{e}_t = (\tilde{e}_{\tau_0,t},\ldots,\tilde{e}_{\tau_N,t})^{\mathrm{T}}$. This Gibbs scheme therefore has two steps:

$$\begin{aligned} \boldsymbol{e}_{i+1}, \boldsymbol{a}_{\mathrm{fg,i+1}} &\leftarrow p(\boldsymbol{e}_i, \boldsymbol{a}_{\mathrm{fg,i}} | \boldsymbol{E}_i, \boldsymbol{g}_j, \boldsymbol{N}, \boldsymbol{d}); \\ P(\tau)_{i+1} &\leftarrow \mathrm{Inv-Gamma}(\sigma^2_{\tau,i+1}, \alpha), \end{aligned} \tag{24}$$

where the joint amplitude sampling is carried out for each LST separately, the bandpower sampling is performed separately for each delay mode, and $\alpha = N_{\mathrm{visibilities}} - 1$. In words, the above scheme jointly samples the EoR signal vector (in frequency space) and amplitudes of the foreground templates $\boldsymbol{g}_j$ for each LST independently by solving the linear system of Equation (22). It then calculates the sample variance of each EoR signal delay mode over all available LSTs, and uses this as a scale parameter to draw a sample of each delay spectrum bandpower from an inverse-Gamma distribution, independently for each delay mode. It then finally performs an inverse Fourier transform of a delay-space covariance matrix constructed by putting the delay spectrum bandpower samples along the diagonal (and zeros elsewhere), resulting in the $i+1$th sample of the frequency-space signal covariance, $\boldsymbol{E}_{i+1}$. Due to our enforcement of the property that $\boldsymbol{E}$ be diagonal in delay space, it is necessarily a circulant matrix in frequency space (Messerschmitt 2006).

The delay spectra of the foregrounds and the EoR signal are degenerate where they overlap at low delay. This is due to the foreground modes corresponding to smooth functions in frequency that are representable with only a handful of low-wavenumber Fourier modes. We implement a symmetric prior on a few low-delay bins of $\boldsymbol{p}$ to control the degeneracy. Specifically, we set the result of the delay spectrum (inverse-Gamma) sample $P(\tau)_{i+1}$ to be equal to the true EoR delay spectrum $P_{\mathrm{true}}(\tau)$ inside delays $-100\,\mathrm{ns} < \tau < 100\,\mathrm{ns}$. This prior prevents time-consuming exploration of the degeneracy, which would slow convergence of the chains.

It is important to note that without this highly specific prior, the sampler is unable to disambiguate between the smoothest foreground modes and the EoR signal, meaning that the lowest-delay EoR signal modes will not be recovered correctly. In the case of real data, a different prior, for example a continuity prior, is likely to be more suitable as the true EoR delay spectrum is unknown. One could also use a set of physically motivated models to constrain possible behaviors of the EoR signal at low delay. We do not explore these possibilities further here.

2.7. Scheme 3 (Joint Sampler with Signal and Foreground-constrained Realizations)

This scheme jointly samples from the EoR signal and foreground components $\boldsymbol{e}$ and $\boldsymbol{f}$ at the GCR step using the full foreground covariance matrix $\boldsymbol{F}$, rather than its leading principal components as in Scheme 2. The assumption is maintained that the foreground covariance does not vary.[10] We again define a block vector $\boldsymbol{x}$, where now the first N_{freq} entries are $\boldsymbol{e}$ and the second N_{freq} entries are $\boldsymbol{f}$, and a corresponding response vector $\boldsymbol{u} = (\boldsymbol{1},\ \boldsymbol{1})^{\mathrm{T}}$ such that the data model is $\boldsymbol{d} = \boldsymbol{x} \cdot \boldsymbol{u} + \boldsymbol{n}$. From the joint conditional distribution of the signal and foreground vectors, we obtain

$$\begin{aligned} p(\boldsymbol{e}, \boldsymbol{f} | \boldsymbol{E}, \boldsymbol{F}, \boldsymbol{N}, \boldsymbol{d}) &\propto p(\boldsymbol{d} | \boldsymbol{e}, \boldsymbol{f}, \boldsymbol{N}) p(\boldsymbol{e} | \boldsymbol{E}) p(\boldsymbol{f} | \boldsymbol{F}) \\ &\propto \exp\left(-(\boldsymbol{d} - \boldsymbol{x} \cdot \boldsymbol{u})^\dagger \boldsymbol{N}^{-1} (\boldsymbol{d} - \boldsymbol{x} \cdot \boldsymbol{u})\right) \\ &\quad \times \exp\left(-\boldsymbol{e}^\dagger \boldsymbol{E}^{-1} \boldsymbol{e}\right) \exp\left(-(\boldsymbol{f} - \bar{\boldsymbol{f}})^\dagger \boldsymbol{F}^{-1} (\boldsymbol{f} - \bar{\boldsymbol{f}})\right), \end{aligned} \tag{25}$$

with $\bar{\boldsymbol{f}}$ being the mean vector of the foregrounds on this baseline. Completing the square with the vector $\boldsymbol{x}$, the corresponding linear scheme to solve to sample from the conditional distribution is

$$\begin{pmatrix} \boldsymbol{E}^{-1} + \boldsymbol{N}^{-1} & \boldsymbol{N}^{-1} \\ \boldsymbol{N}^{-1} & \boldsymbol{F}^{-1} + \boldsymbol{N}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{e} \\ \boldsymbol{f} \end{pmatrix} = \boldsymbol{b}, \tag{26}$$

with $\boldsymbol{b}$ given by

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{N}^{-1}\boldsymbol{d} + \boldsymbol{E}^{-1/2}\boldsymbol{\omega}_0 + \boldsymbol{N}^{-1/2}\boldsymbol{\omega}_1 \\ \boldsymbol{N}^{-1}\boldsymbol{d} + \boldsymbol{F}^{-1}\bar{\boldsymbol{f}} + \boldsymbol{F}^{-1/2}\boldsymbol{\omega}_2 + \boldsymbol{N}^{-1/2}\boldsymbol{\omega}_1 \end{pmatrix}. \tag{27}$$

This implementation samples from the foreground and signal components jointly. Since the foreground covariance on the baseline is taken to be known, it is not sampled from. The (baseline-specific) foreground covariance matrix $\boldsymbol{F}$ is a dense matrix in frequency space that needs to be inverted in order to use this implementation. This Gibbs sampling scheme is very similar to Scheme 2:

$$\begin{aligned} \boldsymbol{e}_{i+1}, \boldsymbol{f}_{i+1} &\leftarrow p(\boldsymbol{e}_i, \boldsymbol{f}_i | \boldsymbol{E}_i, \boldsymbol{F}, \boldsymbol{N}, \boldsymbol{d}); \\ P(\tau)_{i+1} &\leftarrow \mathrm{Inv-Gamma}(\sigma^2_{\tau,i+1}, \alpha), \end{aligned} \tag{28}$$

with foreground vectors now being sampled jointly with the EoR signal realizations. As with Scheme 2, the signal realizations are used to form the basis for inverse-Gamma samples of the delay spectrum, which again is used to define a diagonal covariance matrix in delay space. We also implement the same prior on the central delay bins of $\boldsymbol{p}$ as for Scheme 2.

3. Simulations

In this section we describe the steps taken to build model covariance matrices for the foregrounds, and the covariance used as our EoR signal model. We use the simulation methodology described in Choudhuri et al. (2021) in what follows. The simulations include separate sets of simulated visibilities for point sources, diffuse emission (which we have neglected here), and a simple model of the EoR. They cover a bandwidth of 100–120 MHz in 120 channels, and contain 13.4 hr of LST at 40 s integration time per sample, in the LST range 9.2–22.5 hr. Only the pseudo-Stokes I polarization channel is simulated. The `hera_sim`[11] package is used to perform the visibility simulations themselves, with an analytic approximation to the HERA beam (Choudhuri et al. 2021; Fagnoni et al. 2021) used as the primary beam model, which is furthermore assumed to be identical between receivers.

[10] In principle, this scheme could be extended to also sample the foreground covariance matrix $\boldsymbol{F}$, e.g., as another inverse Wishart Gibbs step, but we leave this to future work.

[11] https://github.com/HERA-Team/hera_sim/

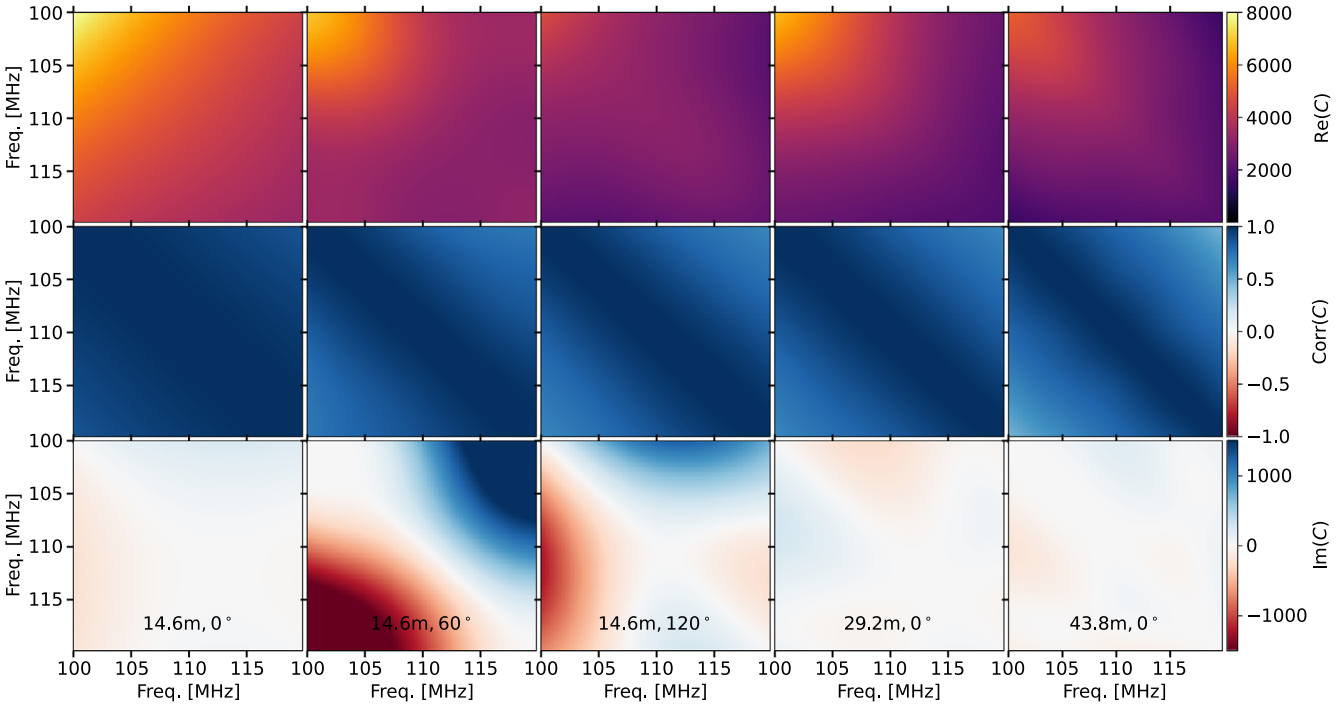

Figure 3. Point-source foreground covariance matrices estimated from averaging over 1200 simulation time samples at a cadence of 40 s in the 100–120 MHz band. Real and imaginary parts (in square janskys) are shown in the top and bottom rows, respectively, while the middle row shows the correlation matrix of the real part, $\rho_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$. The first three columns are for baselines with length 14.6 m, oriented at approximately 0°, 60°, and 120° degrees from the E–W direction, respectively (the shortest redundant baselines). A clear directional dependence is found in the structure and amplitude of the covariance matrix. The last two panels show covariances from baselines of length 29.2 and 43.8 m (both E–W aligned). The real part of each of the covariances shows strong correlations, with an imaginary part that has a smaller variance in general.

To simulate point sources we choose a sky model based on the GaLactic and Extragalactic All-Sky Murchison Wide Field Array (GLEAM) catalog (Hurley-Walker et al. 2017). We also include a few bright sources and Fornax A, which are not present in the catalog (see Table 2 of Hurley-Walker et al. 2017). There are some blank regions in the GLEAM catalog (e.g., north of $+30°$ decl., Galactic latitudes within 10° of the Galactic plane, and a handful of localized areas such as the Magellanic Clouds) with no sources. We fill those gaps with sources taken from other parts of the sky, as described in Choudhuri et al. (2021).

Covariance matrices for simulated point-source foregrounds are estimated per baseline by averaging in the LST (time) direction (ignoring any nonstationarity of the statistics of the foregrounds). For a set of mean-subtracted visibility data $\boldsymbol{V}$ of dimension $N_{\rm freq} \times N_{\rm times}$, the frequency–frequency covariance matrix can be estimated as

$$\boldsymbol{F} = \frac{1}{N_{\rm times} - 1}\boldsymbol{V}\boldsymbol{V}^{\dagger}. \tag{29}$$

In Figure 3, we show examples of point-source foreground covariance matrices formed from different baseline lengths and orientations in the simulated array. The real parts (upper row) show strong correlations, and the imaginary parts (lower row) have structures that change with baseline length and orientation. In the first three columns we show the covariance matrix of point-source foregrounds on baseline vectors of length 14.6 m that form an equilateral triangle. Though the real part takes on a similar structure for each baseline-type triangle, there are notable differences. The first panel shown (14.6 m, 0°) has the largest magnitude (denoting higher variance) and has strong correlations over larger frequency separations than the second panel (14.6 m, 60°). The third panel (14.6 m, 120°) shows a magnitude of the real part that is approaching a factor of 2 smaller than the first panel. The differences in the imaginary part are more striking, with very different correlation structures visible in each.

We also show covariance matrices for longer baselines in the fourth and fifth panels, of length 29.2 m and 43.8 m, respectively, both oriented E–W (0°). These baselines have covariances with shorter correlation lengths than the equivalent 14.6 m baseline, which is expected as longer baselines are more chromatic. The longer baselines also show reduced magnitude of the real part compared to the shorter baselines.

Various point-source covariance calculations and fitting functions exist (e.g., Santos et al. 2005; Murray et al. 2017; Ghosh et al. 2020), and this study of the per-baseline covariances seems to indicate that a one-size-fits-all approach is likely to fall short. The simulations of Choudhuri et al. (2021) include very bright sources such as Cen A, which can significantly increase the total observed power (and therefore the variance) as they transit, for example. This implies that deviations from the assumptions we have made in calculating the covariance—stationarity, and statistical homogeneity/isotropy—can be significant. The orientation of the baselines (and therefore the fringes) on the sky as sources rotate through them at different rates (i.e., at different "fringe rates") will also contribute to the structure of the correlations in the real and imaginary parts. We leave attempts at systematization to future work, and for the results in later sections take the foreground prior covariance $\boldsymbol{F}$ to be the covariance matrix evaluated using

Equation (29) for the baseline that we make use of to generate mock data, a 14.6 m baseline oriented E–W.

For our EoR signal model we use a Gaussian plus a constant offset in delay space. In frequency space, this is

$$\boldsymbol{E}_g(\nu, \nu') = A_s \exp\left(-\frac{1}{2}\frac{(\nu - \nu')^2}{\omega_s^2}\right) + r_s \delta_{\nu\nu'}, \tag{30}$$

where $\delta_{\nu\nu'}$ is a Kronecker delta function, and the parameters take constant values, $A_s = 0.25$, $\omega_s = 0.5$, and $r_s = 0.025$. Realizations from this signal covariance model produce Gaussian fluctuations with a correlation length and amplitude parameterized by ω_s and A_s, respectively. r_s controls the size of a diagonal (in frequency) component that also ensures that $\boldsymbol{E}_g$ is positive definite. This signal model produces a power spectrum that is a Gaussian function around delay zero, plus an offset that is constant in delay.

4. Results

In this section we show results from each sampler over a substantial number of iterations (800) and under different simulated data scenarios, and compare the delay spectrum of the recovered EoR component with the true, input EoR delay spectrum. The primary statistic we consider is the distribution of recovered delay spectra across the set of iterations, which approximates the marginal posterior distribution of the EoR delay spectrum.

First, we show results from runs of each Gibbs sampling scheme under a fiducial testing setup with 5% random flagging applied in the same way to each LST, a S/N of 5 at high delay for each visibility, and the Gibbs sampler chain initialized at the true signal and signal covariance values to avoid a lengthy burn-in period. We then consider three different scenarios: how changes to the flagging fraction affect delay spectrum recovery; how the signal power spectrum recovery is affected by changes to the S/N; and how an incorrectly estimated foreground covariance matrix affects delay spectrum recovery (in this case, one formed from visibility simulations that have no faint sources below 15 Jy).

4.1. Simulation Realizations

We generate our simulated data using the following method. A point-source foreground simulation from a 14.6 m E–W baseline covering 1200 LSTs (40 s spacing) is taken as a base (see Section 3). This simulation is held fixed throughout, i.e., we do not use any other realization of the foreground visibilities than that in the simulated data. On top of it, we add an independent complex Gaussian white noise draw to each time and frequency channel, plus a simple EoR signal component that we generate using 1200 independent complex Gaussian random draws from the signal frequency–frequency covariance matrix (Equation (30)). The shape of the EoR signal power spectrum is not chosen to represent any particular physical model; instead, we use a Gaussian shape with the peak at $\tau = 0$ ns and a width of around 1000 ns, as it provides a simple but nontrivial shape to recover, and allows low-, intermediate-, and high-S/N regimes to be studied in the same power spectrum.

Since each time sample of the EoR signal is drawn separately, the 1200 LSTs each contain an independent realization drawn from the underlying signal power spectrum. This is akin to each time sample being taken approximately one primary beam crossing time apart, such that a different patch of the sky has rotated into the mainlobe of the beam in its entirety. For HERA, this timescale is of order ∼1 hr depending on the observing frequency, and so the choice of independent samples of the EoR signal is clearly an idealization. In reality, observations 40 s apart would see a very similar sky within the primary beam mainlobe, and so the signal realizations would be strongly correlated. Recent HERA analyses (e.g., HERA Collaboration et al. 2022) have performed coherent averaging of visibilities over timescales of a few minutes (with fringe stopping to reduce decoherence), which effectively increases the separation between each (post-averaging) time sample. Coherent averaging over baselines may also be performed to increase S/N. We do not model either of these forms of averaging here, however.

At each iteration the samplers return 1200 EoR signal GCR solutions, one for each LST. We estimate the "empirical" delay spectrum of the samples at each LST separately, by multiplying the Fourier transform of the mean-subtracted, tapered GCR solution by its complex conjugate and then averaging across all LSTs in the iteration to obtain

$$\boldsymbol{P}_{\text{empirical}} = (\hat{\sigma}^2_{\tau_0,i},\ldots,\hat{\sigma}^2_{\tau_N,i})^{\mathrm{T}} \tag{31}$$

for iteration i. We have used $\hat{\sigma}^2_{\tau,i}$ to denote the same quantity as in Equation (23), but with the delay mode $\tilde{e}_\tau$ replaced by an equivalent quantity that was tapered and mean-subtracted before the Fourier transform. Note that $\boldsymbol{P}_{\text{empirical}}$ is different from the delay spectrum estimate $P(\tau)$ that is obtained via sampling by the second step of the Gibbs scheme. In particular, $P(\tau)$ is subject to (inverse-Gamma) sample variance while $\boldsymbol{P}_{\text{empirical}}$ is not, and the calculation of $\boldsymbol{P}_{\text{empirical}}$ includes a tapering operation while $P(\tau)$ does not. We run each sampler for 800 iterations, which is sufficient to achieve good convergence based on visual inspection of the traces of the chains.

4.2. Comparison of the Three Gibbs Schemes

The results from runs of each Gibbs sampling scheme under the conditions described above, for a 5% random channel flagging pattern, are shown in Figure 4 (Scheme 1) and Figure 5 (Schemes 2 and 3). The top panel in each figure compares the true input power spectrum of the foregrounds and EoR signal, the injected noise level (representative of the noise on a visibility for a single time and baseline, i.e., before any time averaging), and the mean recovered EoR power spectrum (or total signal power spectrum, in the case of Figure 4). When power spectra are calculated from samples of the frequency-space visibilities (i.e., in the top and second panels), the visibility data are first mean-subtracted and tapered with a Blackman–Harris window, since the foreground component of the visibilities is generally large and discontinuous at the band edges and so would otherwise cause ringing. This procedure causes a small dip at low delay in all power spectra. The second-from-the-top panels show the fractional residual between the true (input) signal power spectrum and the power spectrum estimated by squaring and averaging the GCR samples, along with 68% confidence intervals estimated from the GCR samples. The lower panels in Figure 5 show the recovered fractional residual between the true input signal power spectrum and the delay spectrum samples obtained at

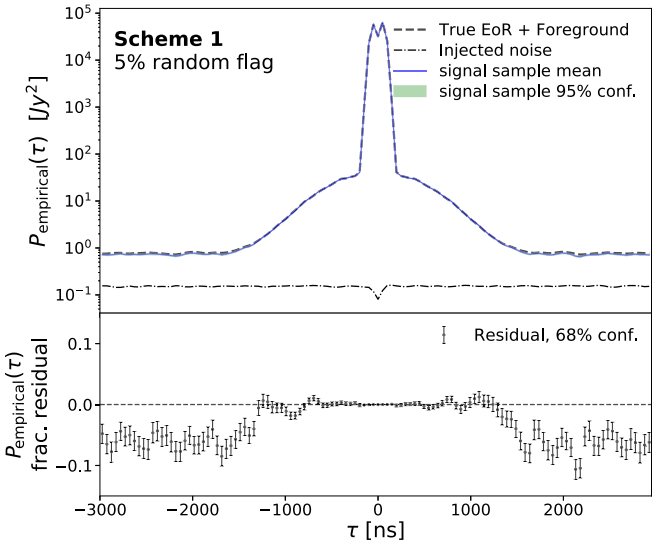

Figure 4. Signal recovery using Scheme 1 (total signal sampler) after 800 iterations. Top panel: true input power spectrum (sum of EoR and foreground components, dashed line) compared with the mean of the power spectrum of the total sky signal estimated from samples of the signal from the GCR step (blue line). Power spectra are calculated after mean-subtraction and tapering with a Blackman–Harris window, which produces a small decrease at $\tau = 0$ ns. The dotted–dashed line shows the noise power for each visibility (i.e., a single time sample; S/N of 5 at high τ); 1200 time samples are combined to measure the total power spectrum. The 95% confidence interval (green shaded region) is too small to see. Bottom panel: fractional residual between the true power spectrum and the recovered power spectrum from the upper panel. The error bars show the 68% confidence region. Outside $|\tau| \gtrsim 1000$ ns, where the S/N is lower, this sampling scheme recovers a distribution that is biased low by around 5%. This bias does not appear when the S/N at high delay is increased.

each iteration from the second (covariance sampling) step of the Gibbs sampling schemes.

Note that the delay spectrum estimate $P(\tau)$ obtained from the second step in the Gibbs scheme is in principle the correct one to use, as it properly takes into account sample variance. The power spectrum estimated from the GCR samples is the "empirical" power spectrum of the particular realization of the EoR field that we see, and does not include sample variance. As such, it is expected to have narrower error bars, which is indeed what we see for Schemes 2 and 3.

In Schemes 2 and 3, shown in Figure 5, where the EoR signal is separated from the foreground component, the samplers exhibit a degeneracy in their solutions for the foreground power spectrum and EoR signal power spectrum where these components overlap in delay. Left unchecked, this degeneracy results in the sampler exploring the degeneracy region very slowly throughout the iterations. For this reason we implement a prior on the delay spectrum (covariance) samples at very low delay in these schemes, setting the delay spectrum samples in the five central delay bins to be equal to their true values, a "hard prior." The region where the residual vanishes at low delay corresponds to this hard prior. This simplistic implementation of the prior is intended to show how these schemes can work when this degeneracy is prevented from being explored by the sampler. Other priors, such as a continuity prior, are likely to be appropriate in practice. We have checked that the results are not sensitive to the choice of hard prior; setting the prior on the delay spectrum in the central five bins to be 50% higher than the true value when carrying out the same run configurations, we found that neither the recovered signal power spectrum mean nor the error bars were affected.

The Scheme 1 "total signal" sampler in Figure 4 recovers a signal power spectrum that is biased low at higher delays, where the S/N is lowest. The reason for the bias is unclear, but we note that when the overall S/N is increased the bias no longer appears. Schemes 2 and 3 recover ostensibly unbiased signal power spectra as seen in their residuals (middle/lower panels), with the width of the residual distribution increasing toward higher delay where the S/N is lowest. Slight biases (dips) are noted for Scheme 2, inside the foreground delay range $|\tau| < 200$ ns, around the edge of the hard prior, where a degeneracy is expected between the signal and foreground amplitudes (see below). Immediately outside of this range, the middle panel shows a 68% confidence region spanning a fractional residual of approximately $\pm 0.5\%$, which increases to $\pm 2.5\%$ at high delay, consistent with the S/N decreasing with delay. Wiggles with correlated error bars are observed in the middle panels for both Schemes 2 and 3, which are expected as a taper has been applied. The delay spectrum samples in the lower panels also appear to be unbiased for both sampling schemes, with larger error bars that show less evolution with delay, as sample variance is the dominant source of uncertainty.

There is little to separate Scheme 2 and Scheme 3 based on their performance in the residual with the true input power spectrum, but Scheme 2, which relies on fitting a small number of foreground templates rather than the entire foreground covariance, reaches full sets of solutions for its GCR step approximately twice as fast as Scheme 3 does. We do however note the small dips in the recovered power spectra (both middle and lower panels) for Scheme 2 just outside the prior-dominated region of Figure 5. We suspect that this is caused by the truncation of the set of foreground modes at the 8th mode. If this is correct, including more modes in the foreground model would allow more of the residual foreground emission at low delay to be absorbed, and this feature would not arise.

Finally, we note that an advantage of the Gibbs sampling approach is that the samples can be used to directly reconstruct the marginal posterior distributions of each parameter (or subset of parameters), without resorting to Gaussian approximations or otherwise. For the particular applications presented here, we did not find any particularly compelling examples of non-Gaussian behavior of the marginal posteriors, however. Visually inspecting the marginal distributions for the foreground amplitude parameters and delay spectrum bandpowers for Scheme 2, for the 10% continuous flag and 10% random flag cases, we found that they were generally consistent with Gaussianity, i.e., we did not note any strong skewness, heavy tails, etc. More quantitative tests for Gaussianity could be performed if desired.

On a related note, the true EoR 21 cm field is expected to have a non-Gaussian component (e.g., due to the formation of ionized bubbles around early sources), but we have modeled it as Gaussian in our analysis. Non-Gaussian features of the field can be captured in this framework, and the power spectrum is still a well-defined quantity that can be measured. Failing to explicitly account for the non-Gaussianity of the field will lead to these features, and therefore the power spectrum bandpowers, being weighted incorrectly, however, and so statistics such as means and uncertainties could be biased. We leave an examination of this issue to future work.

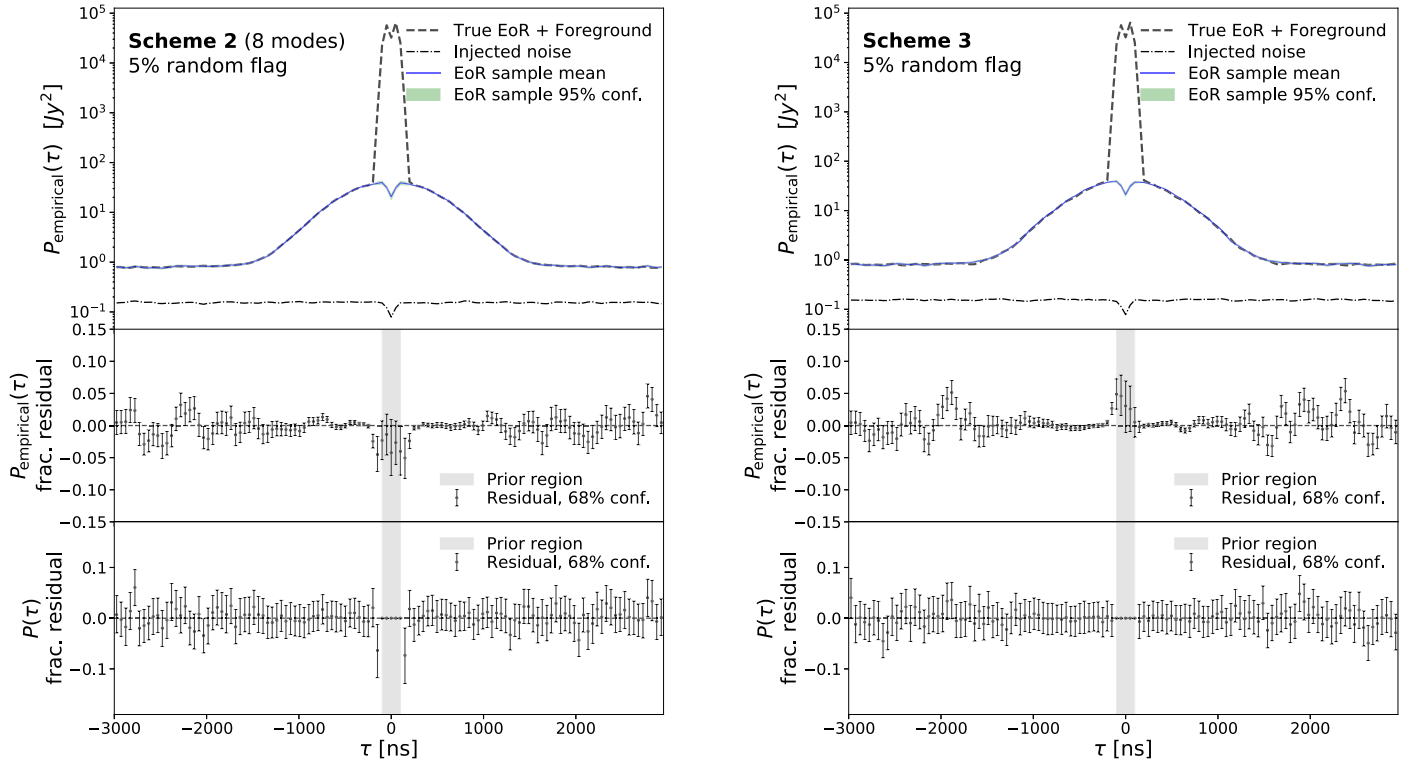

Figure 5. Signal recovery using Scheme 2 (foreground template fitting, left panel) and Scheme 3 (joint foreground sampling, right panel). These runs match the parameters seen in Figure 4: 800 total iterations, S/N at high delay of 5:1. These schemes differentiate between the foreground and EoR signal components and aim to recover the EoR component separately (unlike Scheme 1). Top panel: mean of the EoR power spectrum estimated from the GCR samples of the EoR signal across 800 iterations, with 95% confidence region shown. The injected noise level is shown as a dashed–dotted line. Middle panel: fractional residual between the true (input) EoR power spectrum and the mean of the power spectra derived from the GCR samples from the top panel. Outside the central delays where there is degeneracy between the foregrounds and EoR, the EoR signal recovery is essentially unbiased (within ∼5% at high delay in both cases). Biases inside the degeneracy region are seen to take on slightly different shapes for Scheme 2 and Scheme 3; 68% confidence regions are also shown. Bottom panel: fractional residual between the true delay spectrum and the inverse-Gamma sampled delay spectrum from each iteration, at the covariance sampling step of the sampling schemes. The zero residuals at the central delays are due to the prior that is imposed at low delays on this sampling step. Recovery of the delay spectrum is approximately unbiased when using both Gibbs sampling schemes under this test setup.

We also calculated the covariance matrix for the parameters, finding no strong evidence for substantial covariance between the delay spectrum bandpowers and per-LST foreground amplitudes, even at low delay. This is likely due to a combination of the strong prior on the delay spectrum at low delay and the fact that the foreground amplitudes are estimated for each LST, whereas the bandpowers are estimated in an LST-averaged sense, which will tend to average down any correlations. We did however find covariance between the foreground amplitude parameters themselves. This is not unexpected considering that the foreground eigenmodes were calculated from the LST-averaged frequency–frequency covariance matrix, and so are not necessarily eigenmodes of an equivalent per-LST quantity (fits that use nonorthogonal basis functions will typically result in some covariance between their coefficients). We also noted a weak positive correlation between the $\tau = 0$ ns delay spectrum bandpower and other bandpowers, particularly at high delay, as well as a moderate correlation between the 1 and 2 delay spectrum bandpowers either side of the prior region. The latter is likely a manifestation of the incompleteness of the eight-mode foreground model, as mentioned above.

4.3. Dependence on Flag Fraction

In Figure 6, we use Scheme 2 to assess how the signal power spectrum recovery changes when the flagging fraction is increased. In this study we use a continuous (rather than random) mask increasing up to 10% of the band in length, in order to evaluate a reasonable worst-case scenario. We again add noise such that the S/N takes a value of 5 at high delay, and test a case with no flags, a 5% flag, and a 10% flag. The flags are applied near the center of the frequency range and in the same position at all LSTs.

The recovered signal power spectrum distributions in each of these runs again appear unbiased, even in the 10% flagging case. The width of the error bars increases as the size of the flagged region increases, however, as one would expect from reducing the effective number of data points in the data set. Looking at the middle panels, we see that the error bar width increases significantly at lower delays (e.g., just outside the hard prior region) as the flag fraction is increased. A large continuous flagging region in frequency space increases the overall uncertainty in all Fourier modes, but particularly prevents the lower-delay modes from being measured much more accurately that the higher-delay modes, as they are in the no-flagging case. Despite this, the power spectra sampled by the second Gibbs step remain essentially unchanged, as they are still dominated by sample variance. It is notable that no bias has been introduced into these power spectra either, however; the samplers are successfully marginalizing over the missing signal inside the region in such a way that the Gibbs-sampled power spectrum is recovered correctly. This is in contrast to,

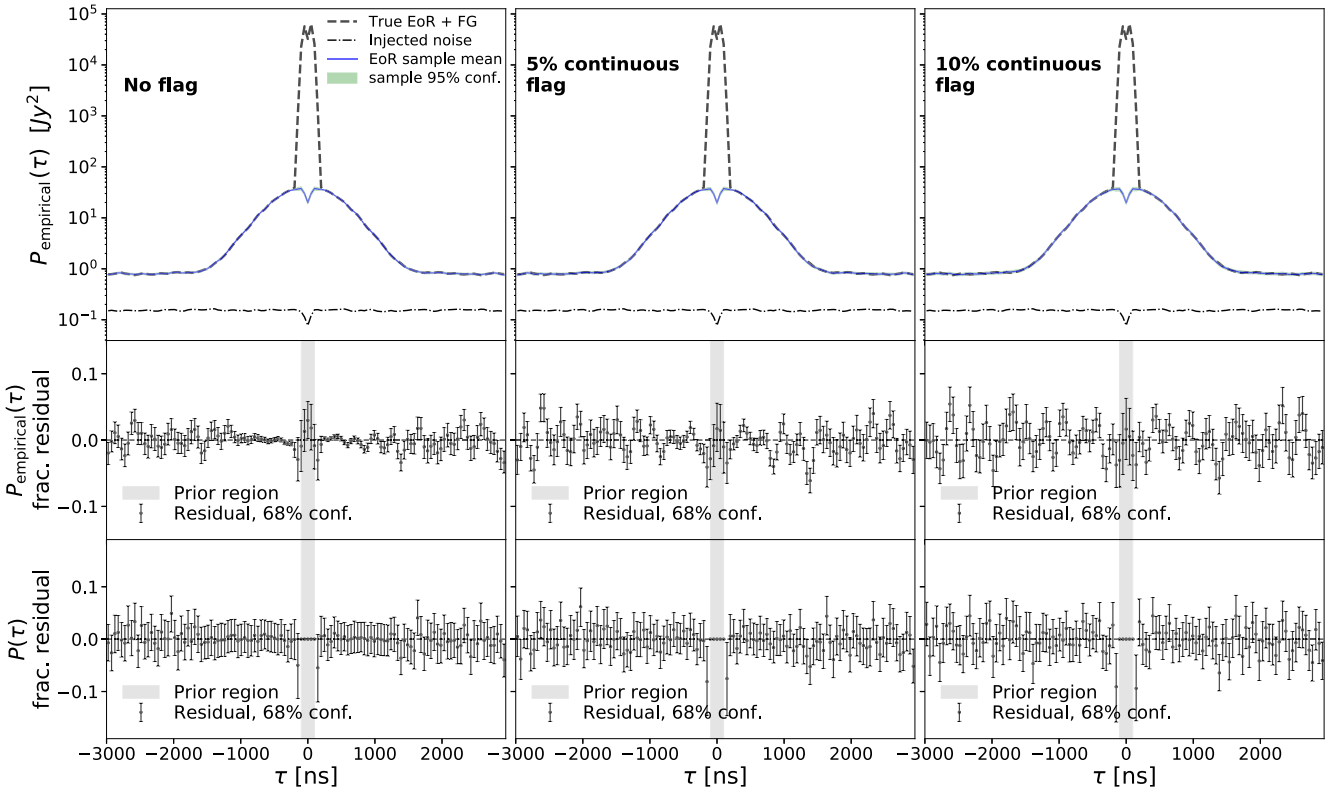

Figure 6. Flagging fraction comparison using the Scheme 2 sampler (foreground templates) for 800 iterations, and with a S/N of 5:1 at high delay in each case. Left column: no flags applied. Centre column: continuous flag through 5% of the band. Right column: continuous flag through 10% of the band. The applied flagging is approximately in the center of the frequency range and is applied to the data at all LSTs. Top panels show the mean $P_{\rm empirical}(\tau)$ and 95% confidence regions compared with the true power spectrum, and the injected noise level. Middle and lower panels show the fractional residual on $P_{\rm empirical}(\tau)$ and $P(\tau)$ samples across all iterations. Both statistics are approximately unbiased outside of the region of signal/foreground degeneracy ($|\tau| < 300$ μs) in all flagging cases, though as the flagging fraction increases more fluctuations are noted in the residual means. Flagging can be seen to increase the width of the $P_{\rm empirical}(\tau)$ distribution, particularly at low delay. The width of the $P(\tau)$ residual does not appear to be altered significantly when increasing the flagging fraction, as this is dominated by sample variance, but the recovery of the "best-fit" delay spectrum (i.e., the central estimate shown with each error bar) becomes noisier.

for example, a Wiener filter method, which would be biased toward zero signal (lower power) due to the flagged regions.

4.4. Dependence on Signal-to-noise Ratio

Figure 7 shows results from runs that use Scheme 3 to examine the effect of changing the S/N of the visibility data, with S/N values of 30, 3, and 0.5 at high delay being compared (reflected by the noise levels in the top panels). The data in these runs are unflagged to simplify the comparison. We change the S/N by rescaling the overall amplitudes of the EoR signal and foreground components, keeping the simulated noise level the same.

The highest-S/N run obtains subpercent error bars on the average signal power spectrum from the GCR step (middle panels). Decreasing the S/N by an order of magnitude (center panels) causes larger fluctuations in the residual, on the order of 5%, to appear in the recovered mean, along with an increase in the size of the error bars (around 4% at high delay). At the lowest S/N of 0.5 (right panels), the size of the fluctuations in the residual approaches 15% at high delay, and an oscillation-like structure is observed, perhaps associated with the taper that has been applied. In the top panel of this run, the 95% confidence interval is much more clearly visible than it has been in any of the previous figures. This interval substantially increases in size once the EoR signal power falls below the noise level. Nevertheless, the recovered power spectra (in both the middle and lower panels) remain unbiased with respect to the true input power spectrum.

4.5. Foreground Covariance from an Incomplete Sky Model

Schemes 2 and 3 both require a model of the frequency–frequency covariance of the foregrounds. Any foreground model based on observations will inevitably be incomplete, as there will be potentially large numbers of faint sources missing from the source catalog used to generate the model. We test the robustness of the Gibbs sampling schemes to this scenario. Figure 8 shows the results from runs using a foreground covariance matrix model that is generated from point-source simulations with a lower source flux limit of 15 Jy, i.e., containing only the brightest sources. (The fluxes and spectral indices of the bright sources are assumed to be known and perfectly calibrated, however.) We carry out runs using Schemes 2 and 3 with a data S/N of 3 at high delay, and find that there is no qualitative difference with runs carried out using the correct foreground covariance matrix.

The "correct" (containing faint sources) foreground covariance matrix and the matrix generated without faint sources differ element wise at subpercent level, with the first eight eigenmodes (as used by the Scheme 2 sampler) being close to identical on inspection. In Figure 9, we show the scalar product of the first eight eigenvectors of the correct foreground covariance matrix with the corresponding set of eigenvectors for the foreground model without faint sources. If the

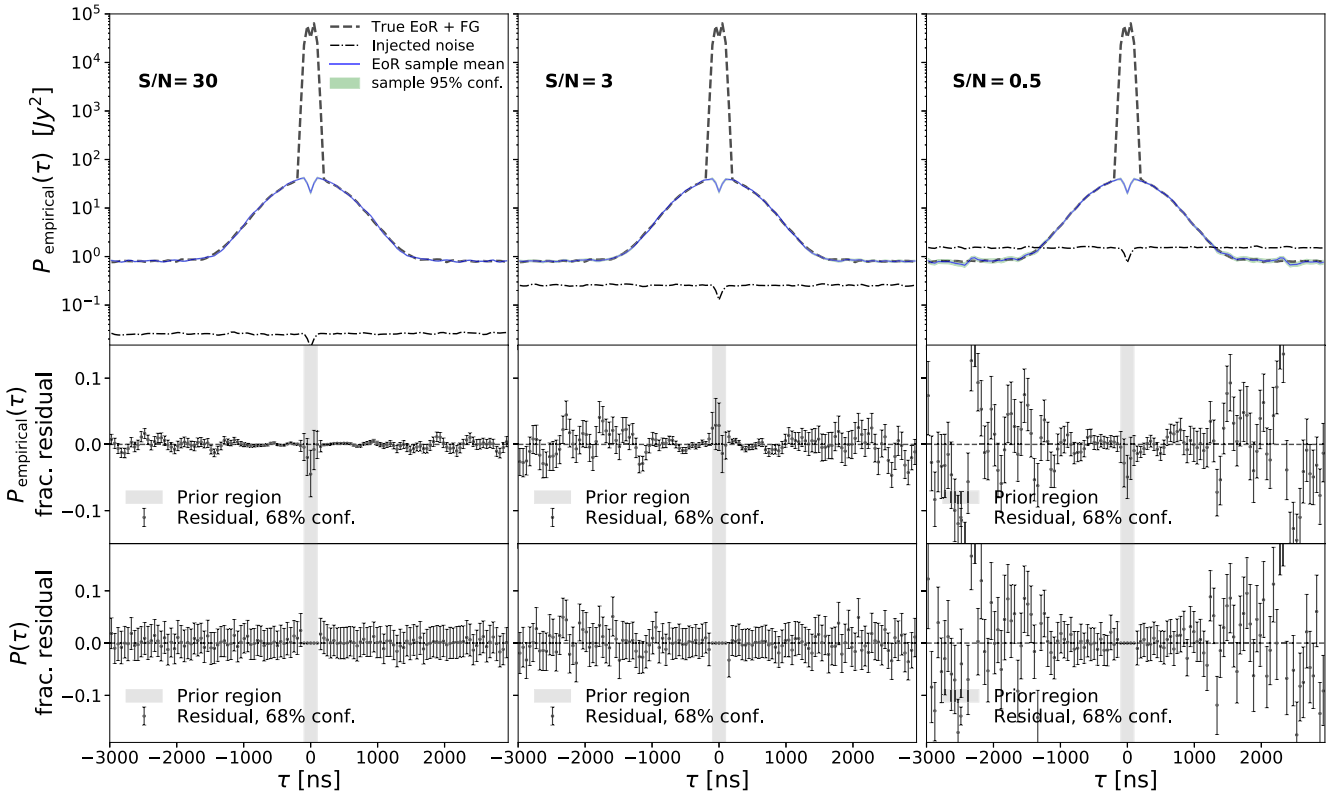

Figure 7. Signal-to-noise ratio comparison using the Scheme 3 sampler (joint foreground sampling) for 800 iterations, and with no flags applied. Left column: S/N 30. Centre column: S/N 3. Right column: S/N 0.5. Top panels show the mean signal GCR power spectrum and 95% confidence regions compared with the true power spectrum, and the injected noise level per LST (the other quantities are averaged over LST). The middle and lower panels show the fractional residual on $P_{\rm empirical}(\tau)$ and $P(\tau)$ across all iterations. As the S/N is decreased, larger fluctuations appear in the residuals of both statistics. The largest fluctuations in the residual appear at high delay, where the signal model is noise-like (flat) and the S/N is lowest.

eigenvectors were identical, the real part of the plot (left panel) would be the identity matrix, and the imaginary part would be all zeros. As it is, the two sets of eigenvectors deviate from orthogonality at around the 1% level or less.

Subject to the caveat that we have assumed a perfect model of the brightest sources, this result is encouraging: the frequency–frequency covariance matrix derived from a sky model containing all point sources, when averaged through the time direction, differs minimally from the same covariance matrix generated without faint sources. This is likely due to a few bright sources making a dominant contribution to the structure of the foreground templates. Depending on the baseline, there may be an ideal number of foreground templates (covariance eigenmodes) to use that capture the bright source contributions well, and change very minimally when the catalog is incomplete. We have not studied the potential for optimization further here, however.

5. Conclusions

Power spectrum estimation for 21 cm EoR experiments is subject to a number of challenges, with particular difficulties arising from the large dynamic range between the EoR signal and foreground emission. Great care must be taken to avoid coupling Fourier modes inside and outside the foreground wedge, as otherwise leakage of the foregrounds into uncontaminated regions of Fourier space will occur. A serious source of leakage is ringing due to sharp edges caused by RFI masking; Fourier analysis becomes ill-posed in the case of a cut (partially flagged) domain, and some way must be found to effectively model the contribution of the missing data to each Fourier mode. A common approach is to "in-paint" a plausible signal into the flagged regions to reduce the discontinuity with the unflagged data. Alternatively, knowing the flagging pattern and the covariance of the data, one could in principle "undo" most of the leakage by weighting the data by an appropriately masked inverse covariance matrix, which would serve to decorrelate the Fourier modes that were correlated by the masking. In both cases, the difficulty is in finding a model of the data and/or covariance matrix that is sufficiently accurate.

In this paper, we have presented an approach that effectively acts as a combined foreground separation, in-painting, and power spectrum estimation method. It is based on constructing a parametric model of the foregrounds, EoR signal, and their respective power spectra/covariance, and estimating the joint posterior distribution of all of these parameters using the combined techniques of Gibbs sampling and GCRs. By exploiting the structure of the model and the Gaussianity of the likelihood, we can write the joint posterior of all the parameters as an iterative sampling scheme across multiple tractable conditional distributions for subspaces of the parameter space. With a Gaussian likelihood and suitable Gaussian or uniform priors, conditional distributions in which the model is linear in the parameters reduce to multivariate Gaussian distributions, which can be sampled from efficiently using GCR, a method based on a simple extension of Wiener filtering. This is possible even for very high dimensional spaces, potentially with hundreds of thousands of parameters. A useful byproduct of GCR (and Wiener filtering for that matter) is their ability to fill-in regions of missing data with

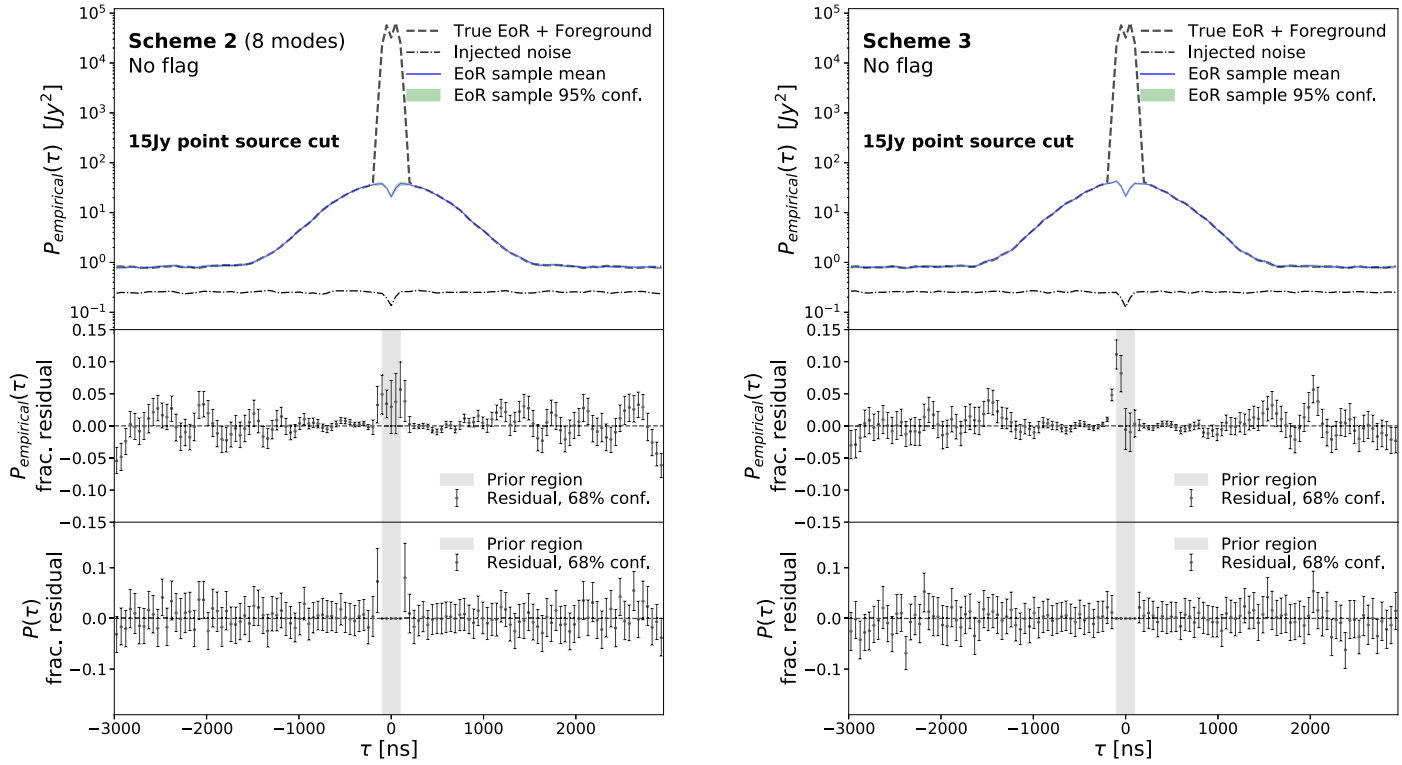

Figure 8. Runs with Schemes 2 (left panel) and 3 (right panel) making use of foreground covariance matrix priors that have faint sources below 15 Jy cut from the catalog. These runs are carried out to test the robustness of the method to faint source loss. The flux cut is set very high in order to examine a reasonable worst-case scenario. Both runs return an unbiased residual, and have very similar error bars to runs where the "correct" covariance matrix is used, e.g., the right panel is directly comparable with the center panels of Figure 7.

plausible realizations of the data model, constrained by the surrounding unmasked data and an estimate of the signal covariance. A statistically well-posed in-painting step is therefore naturally incorporated in this method.

The iterative nature of the method also allows us to start with a relatively poor/simplistic estimate of the signal and foreground covariance matrices, but then converge to much better estimates through repeated sampling. Once convergence is achieved, the other sampling steps (e.g., the GCR steps) are (on average) using the true inverse covariance to weight the data, making it in some sense close to an "optimal" way of estimating the power spectra. This is not quite the same as performing an optimal quadratic estimate of the power spectrum, however, as in that case one is conditioning on a fixed estimate of the data covariance rather than marginalizing over it. The marginalization over the joint posterior of the signal and its covariance should also prevent signal loss occurring (see Kolopanis et al. 2019), as possible correlations between the signal and the covariance are properly taken into account. We have gone some way toward a practical demonstration that this is the case in Section 4, in that we have shown that our methods are largely unbiased under Schemes 2 and 3, with Scheme 1 showing signal loss at a 5% level at high delay. We note that a similar effect can be achieved by marginalizing over (rather than simply optimizing) the kernel hyperparameters in GPR methods (Kern & Liu 2021).

The Gibbs sampling approach permits a variety of modeling choices, particularly in terms of how to parameterize the foregrounds and the EoR covariance matrix/power spectrum. We have presented three Gibbs sampling schemes that sample from the joint posterior distribution of the EoR signal, its covariance, and some representation of the foreground emission. All three schemes make use of per-baseline foreground frequency–frequency covariance matrices estimated from point-source simulations with similar properties (e.g., beams) to a subset of the HERA array.

We tested each of the samplers on a fiducial simulation containing 1200 LSTs for a single 14.6 m E–W baseline, with a Gaussian EoR signal model with a S/N of 5 at high delay and 5% of the frequency band randomly flagged. We found that the "total signal" sampling scheme (Scheme 1), which samples the full signal-plus-foregrounds covariance matrix from a complex inverse Wishart distribution, is biased at the 5% level at higher delays, but that this bias is not present when the S/N in the data is increased. This, along with the fact that sampling from a dense covariance matrix containing both foregrounds and EoR signal (as opposed to the EoR signal only, which can be modeled as diagonal in Fourier space) decreases the efficiency of exploring the joint posterior distribution, leads us to disfavor Scheme 1.

Conversely, we found that foreground template fitting (Scheme 2) and joint foreground sampling (Scheme 3) are able to recover unbiased signal power spectra when the degeneracy between EoR and foreground Fourier amplitudes is broken by a prior on the EoR signal at the lowest delays during the covariance sampling step. This would initially seem to be an overly strong prior assumption; however, we confirmed that using an incorrect hard prior (e.g., a 50% larger fixed value than the true EoR signal at low delay) does not influence EoR recovery at other delays. A different prior, such as a continuity prior, would be more appropriate in practice, and this would be

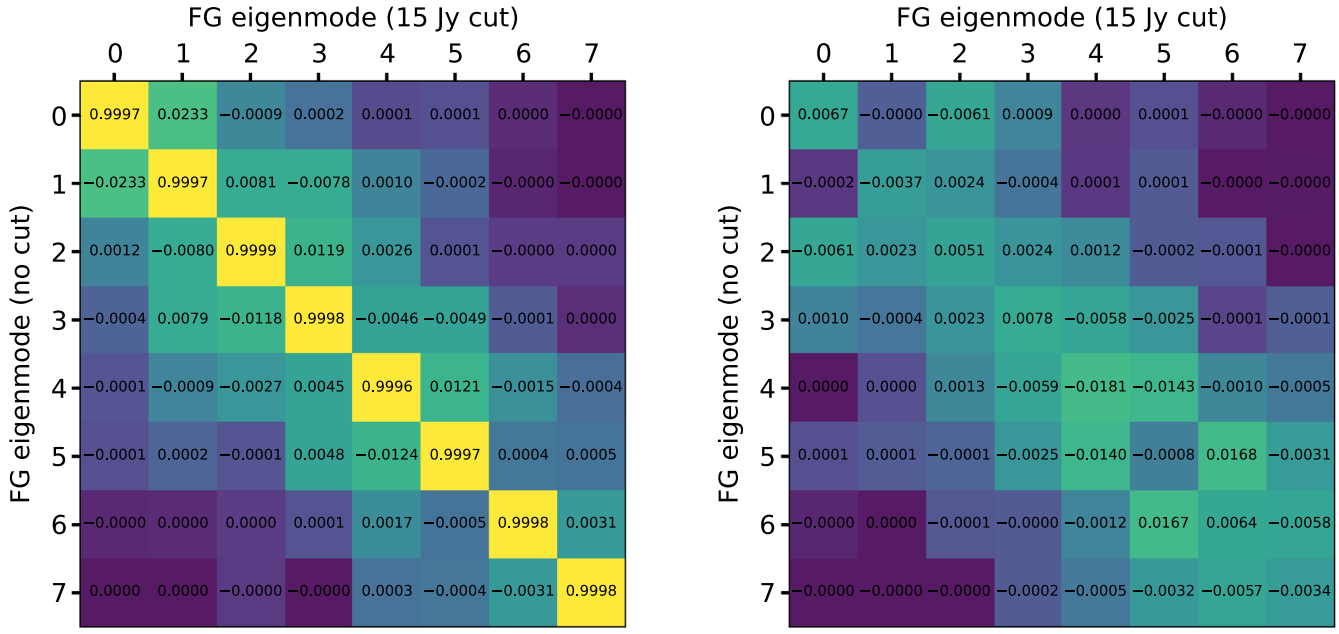

Figure 9. Projection of the first eight foreground eigenmodes for the "correct" foreground covariance matrix (vertical axis) onto the ones for the foreground model with a flux cut at 15 Jy (horizontal axis), calculated as $\boldsymbol{g}_{\rm full} \cdot \boldsymbol{g}^{\dagger}_{\rm faint}$. The real part is shown on the left and the imaginary part on the right. The color scheme is chosen arbitrarily to highlight regions of similar value for the projection.

an important line of inquiry for future work with Gibbs sampling schemes following this structure.

While our analysis has only considered a single 14.6 m, E–W oriented baseline as an example, we anticipate similar behaviors of our method for longer baselines and different orientations, albeit with increasing spectral structure (and therefore more correlations with low-delay 21 cm signal modes) as the baseline length increases and as the orientation changes. The simulation-derived foreground frequency–frequency covariance matrices shown in Figure 3 give some sense of how the correlation structure changes with baseline length, with the 43.8 m baseline showing substantially reduced off-diagonal values compared with the 14.6 m baselines, for instance. Orientations away from the E–W direction also have a reduced level of frequency correlation, including enhanced off-diagonal components in the imaginary part. Studying the sensitivity of the method to these features, and the accuracy with which they are modeled, is left to future work.

We further tested the foreground template and joint foreground samplers under a range of flagging and noise conditions. We found that the width of the error bar on the recovered signal power spectrum increases when the flagging fraction is increased, particularly at lower delay, which is expected as more information has been lost by flagging a larger fraction of the data. We find that a lower S/N of the data also increases the uncertainty on the recovered EoR power spectrum, particularly when it falls below 1, again as expected.

Finally, we tested the robustness of the method to missing faint sources in the point-source sky model used to estimate the foreground frequency–frequency covariance matrix. We found no qualitative difference in the recovery of the EoR power spectrum that was achieved, despite using quite a high flux cut of 15 Jy to model missing faint sources. We attribute this result to the fact that the covariance matrix of point-source foregrounds (and particularly the leading principal components) changes comparatively little when faint sources are removed; the brightest sources appear to dominate its frequency structure. An important caveat is that we have ignored the contribution from diffuse foreground emission, which is likely to play an important role in determining the frequency structure of the foreground component, particularly for the shortest baselines.

Another caveat is that we have assumed that the EoR signal realizations are uncorrelated between time samples. This is an optimistic assumption, as one would only expect them to be independent on timescales of roughly the primary beam crossing time (i.e., long enough for a beam-sized patch of the sky to fully rotate through the primary beam mainlobe). Coherent averaging of visibilities on timescales of a few minutes would help a realistic analysis approach this idealization, but it will still be important to quantify the time–time covariance of the signal and foregrounds in order to avoid overestimating the amount of independent information contained within a given data set. The mathematical formalism required to incorporate time (as well as frequency) correlations should be a straightforward extension to what we have presented here, but the computational cost may be considerably higher.

Both the foreground template and joint foreground fitting schemes appear to be sufficient to recover the EoR power spectrum from individual baselines in an unbiased way, and with similar error properties. In the absence of significant differences in recovery of the EoR, we note that the foreground template sampler (Scheme 2) completes iterations approximately twice as fast as the joint sampler (Scheme 3).

We are grateful to H. Garsden, B. Hazelton, N. Kern, A. Liu, and M. Morales for useful comments and suggestions. This result is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 948764; P.B. and M.J.W.). P.B. and F.K. acknowledge support from STFC grants No. ST/T000341/1 and No. ST/X002624/1.

We acknowledge use of the following software: `HEALPix` (Gorski et al. 2005), `matplotlib` (Hunter 2007), `numpy` (van der Walt et al. 2011), and `scipy` (Virtanen et al. 2020).

Data Availability

The code used to generate the results for this paper is available at https://github.com/fraserlkennedy/Hydra-PSpec-prototype.

Appendix A
Wiener Filter Bias

Despite the Wiener filter being the MAP solution for the combined signal $\boldsymbol{s}$ component of a data vector $\boldsymbol{d} = \boldsymbol{s} + \boldsymbol{n}$, in general it is a biased estimator of $\boldsymbol{s}$ when the mean of the signal component differs from zero. The expectation value of the Wiener filter solution (Equation (8)) is

$$\langle \boldsymbol{s}_{\rm wf} \rangle = [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1} \boldsymbol{N}^{-1} \bar{\boldsymbol{s}}, \tag{A1}$$

where we have defined $\bar{\boldsymbol{s}} \equiv \langle \boldsymbol{s} \rangle$, i.e., the true mean of the signal. Thus, in general $\langle \boldsymbol{s}_{\rm wf} \rangle \neq \boldsymbol{s}$, but the bias vanishes when the mean of the signal $\bar{\boldsymbol{s}} = 0$. A separate bias is also incurred on the variance of the Wiener filter, which impacts two-point statistics like the power spectrum. Taking the covariance of Equation (8), with data $\boldsymbol{d}$ drawn from the appropriate statistical ensemble (i.e., not fixed), we find

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{s}_{\rm wf}) &= \langle \boldsymbol{s}_{\rm wf}\, \boldsymbol{s}_{\rm wf}^{\dagger} \rangle - \langle \boldsymbol{s}_{\rm wf} \rangle \langle \boldsymbol{s}_{\rm wf}^{\dagger} \rangle \\ &= [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-2} \boldsymbol{N}^{-2} [\boldsymbol{S} + \boldsymbol{N} - \bar{\boldsymbol{s}} \bar{\boldsymbol{s}}^{\dagger}], \end{aligned} \tag{A2}$$

where the noise covariance $\boldsymbol{N}$ has been assumed to be diagonal. If the data are instead treated as being fixed, we find $\mathrm{Cov}(\boldsymbol{s}_{\rm wf}) = 0$, which is equivalent to the statement that the Wiener filter with given data and prior assumptions is unique. Comparing with the covariance of the GCR equation solution (Equation (12)), and noting that independent Gaussian random variables are uncorrelated, $\langle \boldsymbol{\omega}_0 \boldsymbol{\omega}_1^{\dagger} \rangle = \langle \boldsymbol{\omega}_0 \boldsymbol{n}^{\dagger} \rangle = \langle \boldsymbol{\omega}_0 \boldsymbol{s}^{\dagger} \rangle = 0$ and $\langle \boldsymbol{\omega}_i \boldsymbol{\omega}_i^{\dagger} \rangle = \boldsymbol{I}$, we obtain

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{s}_{\rm cr}) &= \mathrm{Cov}(\boldsymbol{s}_{\rm wf}) + [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-2} \boldsymbol{N}^{-1} \\ &\quad + [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-2} \boldsymbol{S}^{-1} \\ &= \mathrm{Cov}(\boldsymbol{s}_{\rm wf}) + [\boldsymbol{S}^{-1} + \boldsymbol{N}^{-1}]^{-1}. \end{aligned} \tag{A3}$$

The GCR equation solutions, $\boldsymbol{s}_{\rm cr}$, therefore have strictly greater variance than Wiener filter solutions, $\boldsymbol{s}_{\rm wf}$, and the additional variance has the form expected from the rearrangement shown in Equation (10): a sum of inverses, stemming from the product of Gaussian distributions in the posterior of Equation (4). This difference in variance is expected, as the Wiener filter is a summary statistic (the MAP estimate); summary statistics can typically be measured with lower variance than the variance of the parent distribution.

Appendix B
Sampling from the Complex Inverse Wishart Distribution

Equation (18) requires us to draw samples from a complex inverse Wishart distribution. This appendix briefly describes a practical algorithm to perform the necessary random draws.

Functions to draw random matrices from the real Wishart distribution are available in numerical libraries such as `scipy`. Real inverse Wishart samples can be drawn by noting that the inverse Wishart distribution for a covariance matrix C is closely related to a Wishart distribution where the covariance is replaced by the precision matrix, $P = C^{-1}$ (e.g., Zhang 2021).

A similar connection persists for the complex inverse Wishart distribution, and so to sample from it we only need to be able to sample from a complex Wishart distribution. Unfortunately, a suitable function is not readily available in most standard numerical libraries. A useful observation is that a complex multivariate Gaussian distribution with N parameters can be written as a real multivariate Gaussian with $2N$ parameters by splitting the complex numbers into their real and imaginary parts. Care must be taken to define the blocks of the resulting covariance matrix correctly (e.g., Goodman 1963), as the general complex multivariate Gaussian distribution actually depends on two covariance-like objects. For a vector of complex Gaussian random variates $\boldsymbol{z}$, these are the covariance, $\boldsymbol{C} \propto \boldsymbol{z}\boldsymbol{z}^{\dagger}$ (note the conjugate transpose), and the relation matrix, $\boldsymbol{D} \propto \boldsymbol{z}\boldsymbol{z}^{\rm T}$. In this work we are interested in the "circular" case where $\boldsymbol{D} = 0$, since the relative magnitude of $\boldsymbol{D}$ compared with $\boldsymbol{C}$ for our visibility simulations is below the expected level of noise given the number of samples used to estimate the covariance ($<0.5\%$), and furthermore the posterior distributions we later aim to sample from are dependent on the information contained in $\boldsymbol{C}$ rather than $\boldsymbol{D}$.

With an appropriate splitting of the complex parameters into their real and imaginary parts, and the covariance matrix replaced with the precision matrix as explained above, it is then possible to use the standard real Wishart sampling function to make random draws from the inverse Wishart distribution for the $2N$ real and imaginary parameters. We generate a $2N \times 2N$-sized real block matrix of the form

$$\boldsymbol{C}_{\rm block} = \begin{pmatrix} \boldsymbol{C}_{\rm RR} & \boldsymbol{C}_{\rm IR} \\ \boldsymbol{C}_{\rm RI} & \boldsymbol{C}_{\rm II} \end{pmatrix} \tag{B1}$$

from the complex covariance matrix we wish to sample from using the relations

$$\begin{aligned} \boldsymbol{C}_{\rm RR} &= \frac{1}{2}\Re(\boldsymbol{C}); \quad \boldsymbol{C}_{\rm II} = \frac{1}{2}\Re(\boldsymbol{C}); \\ \boldsymbol{C}_{\rm IR} &= \frac{1}{2}\Im(\boldsymbol{C}); \quad \boldsymbol{C}_{\rm RI} = -\frac{1}{2}\Im(\boldsymbol{C}), \end{aligned} \tag{B2}$$

where the relation matrix $\boldsymbol{D}$ has been assumed to be zero. We then use a `scipy` routine to generate a real-valued inverse Wishart sample using the $\boldsymbol{C}_{\rm block}$ matrix, which carries out a Wishart sample using the precision matrix $\boldsymbol{C}_{\rm block}^{-1}$. This block-form sample contains a random draw of not only the covariance information $\boldsymbol{C}$ but also the relation information $\boldsymbol{D}$, and so it is not consistent with $\boldsymbol{D} = 0$ in general. Because we are only interested in sampling $\boldsymbol{C}$, we calculate this quantity directly from the block-form sample using the relation

$$\boldsymbol{C} = \boldsymbol{C}_{\rm RR} + \boldsymbol{C}_{\rm II} + i(\boldsymbol{C}_{\rm IR} - \boldsymbol{C}_{\rm RI}). \tag{B3}$$

A similar procedure is described in, for example, Wilson et al. (2018) and Zhang (2021).

ORCID iDs

Fraser Kennedy https://orcid.org/0000-0002-5883-6543
Philip Bull https://orcid.org/0000-0001-5668-3101

Michael J. Wilensky https://orcid.org/0000-0001-7716-9312
Jacob Burba https://orcid.org/0000-0002-8465-9341
Samir Choudhuri https://orcid.org/0000-0002-2338-935X

References

Ali, S. S., Bharadwaj, S., & Chengalur, J. N. 2008, MNRAS, 385, 2166
Barry, N., Hazelton, B., Sullivan, I., Morales, M., & Pober, J. 2016, MNRAS, 461, 3135
Barry, N., Wilensky, M., Trott, C.M., et al. 2019, ApJ, 884, 1
Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, ApJ, 695, 183
Byrne, R., Morales, M., Hazelton, B., et al. 2019, ApJ, 875, 70
Chakraborty, A., Datta, A., & Mazumder, A. 2022, ApJ, 929, 104
Chapman, E., Zaroubi, S., & Abdalla, F. B. 2016, MNRAS, 458, 2928
The CHIME Collaboration, Amiri, M., Bandura, K., et al. 2023, ApJ, 947, 16
Choudhuri, S., Bull, P., & Garsden, H. 2021, MNRAS, 506, 2066
Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
de Gasperin, F., Dijkema, T. J., Drabent, A., et al. 2019, A&A, 622, A5
DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001
Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, ApJ, 564, 576
Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, ApJ, 676, 10
Ewall-Wice, A., Dillon, J. S., Liu, A., & Hewitt, J. 2017, MNRAS, 470, 1849
Ewall-Wice, A., Dillon, J. S., Hewitt, J. N., et al. 2016, MNRAS, 460, 4320
Ewall-Wice, A., Kern, N., Dillon, J. S., et al. 2021, MNRAS, 500, 5195
Fagnoni, N., Acedo, E., DeBoer, D. R., et al. 2021, MNRAS, 500, 1232
Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, PhR, 433, 181
Gallager, R. 2013, Stochastic Processes: Theory for Applications (Cambridge: Cambridge Univ. Press),
Garsden, H., Greenhill, L., Bernardi, G., et al. 2021, MNRAS, 506, 5802
Gehlot, B. K., Koopmans, L. V. E., de Bruyn, A. G., et al. 2018, MNRAS, 478, 1484
Geman, S., & Geman, D. 1984, ITPAM, PAMI-6, 721
Ghosh, A., Mertens, F., Bernardi, G., et al. 2020, MNRAS, 495, 2813
Goodman, N. R. 1963, Ann. Math. Stat., 34, 152
Gorski, K. M., Hivon, E., & Banday, A. J. 2005, ApJ, 622, 759
Hallinan, G. 2014, AAS Topical Conf. Series 2, Exascale Radio Astronomy, 102.03
HERA Collaboration, Abdurashidova, Z., Aguirre, J. E., et al. 2022, ApJ, 925, 221
Hunter, J. D. 2007, CSE, 9, 90
Hurley-Walker, N., Callingham, J. R., Hancock, P. J., et al. 2017, MNRAS, 464, 1146
Joseph, R. C., Trott, C. M., & Wayth, R. B. 2018, AJ, 156, 285
Joseph, R. C., Trott, C. M., Wayth, R. B., & Nasirudin, A. 2020, MNRAS, 492, 2017
Kariuki Chege, J., Jordan, C. H., Lynch, C., et al. 2022, arXiv:2207.12090
Kern, N. S., & Liu, A. 2021, MNRAS, 501, 1463
Kern, N. S., Parsons, A. R., Dillon, J. S., et al. 2019, ApJ, 884, 105
Kolopanis, M., Jacobs, D. C., Cheng, C., et al. 2019, ApJ, 883, 133
Li, W., Pober, J. C., Barry, N., et al. 2019, ApJ, 887, 141
Liu, A., Parsons, A. R., & Trott, C. M. 2014a, PhRvD, 90, 023018
Liu, A., Parsons, A. R., & Trott, C. M. 2014b, PhRvD, 90, 023019
Madau, P., Meiksin, A., & Rees, M. J. 1997, ApJ, 475, 429
Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2020, MNRAS, 493, 1662
Messerschmitt, D. 2006, Autocorrelation Matrix Eigenvalues and the Power Spectrum, UCB/EECS-2006-90, EECS Department, Univ. of California, Berkeley, https://www2.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-90.html
Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
Morales, M. F., & Wyithe, J. S. B. 2010, ARA&A, 48, 127
Mouri Sardarabadi, A., & Koopmans, L. V. E. 2019, MNRAS, 483, 5480
Murray, S., Trott, C., & Jordan, C. 2017, in IAU Symp. 333, Peering towards Cosmic Dawn, ed. V. Jelić & T. van der Hulst (Cambridge: Cambridge Univ. Press), 199
Offringa, A. R., Mertens, F., & Koopmans, L. V. E. 2019, MNRAS, 484, 2866
Orosz, N., Dillon, J. S., Ewall-Wice, A., Parsons, A. R., & Thyagarajan, N. 2019, MNRAS, 487, 537
Paciga, G., Albert, J. G., Bandura, K., et al. 2013, MNRAS, 433, 639
Pagano, M., Liu, J., Liu, A., et al. 2023, MNRAS, 520, 5552
Parsons, A. R., & Backer, D. C. 2009, AJ, 138, 219
Parsons, A. R., Backer, D. C., Foster, G. C., et al. 2010, AJ, 139, 1468
Patil, A. H., Yatawatta, S., Zaroubi, S., et al. 2016, MNRAS, 463, 4317
Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65
Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJL, 768, L36
Pritchard, J. R., & Loeb, A. 2012, RPPh, 75, 086901
Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169
Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
Särkkä, S., & Solin, A. 2013, in Image Analysis. SCIA 2013, ed. J. K. Kämäräinen & M. Koskela (Berlin: Springer), 172
Swarup, G., Ananthakrishnan, S., Kapahi, V. K., et al. 1991, CSci, 60, 95
Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015, ApJ, 804, 14
Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, PASA, 30, e007
Trott, C. M., Pindor, B., Procopio, P., et al. 2016, ApJ, 818, 139
Trott, C. M., Jordan, C. H., Midgley, S., et al. 2020, MNRAS, 493, 4711
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, A&A, 556, A2
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, PhRvD, 70, 083511
Wayth, R. B., Tingay, S. J., Trott, C. M., et al. 2018, PASA, 35, e033
Wilensky, M. J., Barry, N., Morales, M. F., Hazelton, B. J., & Byrne, R. 2020, MNRAS, 498, 265
Wilensky, M. J., Hazelton, B. J., & Morales, M. F. 2022, MNRAS, 510, 5023
Wilson, D. K., Pettit, C. L., & Ostashev, V. 2018, Proc. Meet. Acoust., 35, 055005
Zhang, Z. 2021, J. Behav. Data Sci., 1, 119