An Optimal Latent Space for Amortised Bayesian Uncertainty Quantification and Evidence Evaluation in Interferometric Visibilities

Jacob L. Tutt[1,2★] Peter H. Sims[1,2] Jacob Burba[3] and Eloy de Lera Acedo[1,2]
[1]Astrophysics Group, Cavendish Laboratory, J. J. Thomson Avenue, Cambridge, CB3 0HE, UK
[2]Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, UK
[3]Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester M13 9PL, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Recovering the 21-cm power spectrum from interferometric visibilities requires a faint cosmological signal to be distinguished from foreground emission many orders of magnitude brighter, after the chromatic instrumental response has mapped both into highly degenerate structures. Physically motivated Bayesian forward models provide a principled framework for joint inference and rigorous uncertainty quantification, but can require 10^4–10^6 parameters, rendering posterior exploration and Bayesian-evidence evaluation computationally prohibitive. We first extend BayesEoR beyond its existing tangent-plane foreground treatment by parameterising wide-field emission directly in the spherical image domain. We then derive a Karhunen–Loève basis that preserves the noise-relevant prior-predictive foreground uncertainty by construction, reducing the dense system required for each marginal-likelihood evaluation. The retained rank therefore directly quantifies the noise-relevant dimensionality of foreground uncertainty in visibility space. Extending the model across the final 10° in zenith angle to the horizon increases this dimensionality by a factor of 4.63, demonstrating the disproportionate contribution of near-horizon emission and the need to propagate its uncertainty in cosmological inference. For full-field analysis, the method reduces the joint coefficient space from 306, 163 to just 9, 551, enabling unbiased power-spectrum recovery and Bayesian-evidence evaluation in just 0.32 GPU-hours, giving an estimated computational saving of $\sim$ 830 relative to an unreduced GPU-resident implementation. Throughout, we highlight the importance of Bayesian validation, allowing robust cosmological recovery to be distinguished from apparently plausible but spurious detections caused by model misspecification.

Key words: methods: data analysis – methods: statistical – techniques: interferometric – dark ages, reionization, first stars – cosmology: observations

1 INTRODUCTION

Between the emission of the Cosmic Microwave Background (CMB, $z \sim 1100$; Planck Collaboration et al. 2020) and the earliest observed galaxies ($z \sim 14.4$; Naidu et al. 2026) a vast interval of cosmic history remains largely unconstrained by direct observations. The evolution of the Universe across this interval encompassed some of its most dramatic transformations and is conventionally described in terms of three overlapping epochs: the Cosmic Dark Ages (DA, $z \sim 1100$–30), during which the first structures formed through the gravitational collapse of primordial density perturbations (Bernardeau et al. 2002); the Cosmic Dawn (CD, $z \sim 30$–10), when the first stars, galaxies and black holes emerged and began to heat the surrounding gas (Barkana & Loeb 2001); and finally the Epoch of Reionisation (EoR, $z \sim 10$–6), during which radiation from these early sources progressively ionised the neutral intergalactic medium (IGM; McQuinn 2016). Although the James Webb Space Telescope (JWST) has begun to reveal the brightest individual galaxies emerging towards the end of this interval, inferring their population-level properties and the physics governing their evolution from this sparse sample remains statistically challenging (for a review, see Stark et al. 2026).

For nearly half a century, the redshifted 21-cm hyperfine transition of neutral hydrogen has promised precisely such a probe (Hogan & Rees 1979; Scott & Rees 1990; Madau et al. 1997). Through its sensitivity to the density, ionisation state and thermal history of the IGM, the 21-cm signal probes a uniquely broad range of physics, from its matter content and expansion history during the Dark Ages (Bevins 2026) to the Lyα coupling, X-ray heating and reionisation driven by the first luminous sources (for reviews, see Furlanetto et al. 2006; Pritchard & Loeb 2012; Barkana 2016).

Radio interferometry aims to provide access to the three-dimensional fluctuations in the 21-cm brightness-temperature field, initially through statistical summaries and eventually through tomographic imaging with the Square Kilometre Array (SKA; Koopmans et al. 2015). Towards this goal, current-generation experiments such as HERA (DeBoer et al. 2017), LOFAR (van Haarlem et al. 2013), the MWA (Tingay et al. 2013) and NenuFAR (Zarka et al. 2015) continue to advance towards a statistical detection, with recent observations placing increasingly stringent upper limits on the 21-cm power spectrum across a broad range of redshifts and spatial scales (HERA Collaboration et al. 2026; Mertens et al. 2025; Trott et al.

★ E-mail: jlt67@cam.ac.uk

© 2026 The Authors

2025; Munshi et al. 2025). Even in the absence of a detection, these limits already constrain the thermal and ionisation state of the IGM and the radiative properties of the galaxies that heated and ionised it (HERA Collaboration et al. 2023; Greig et al. 2021). Combining them with complementary probes, including global 21-cm experiments (see Bowman et al. 2018; Singh et al. 2022; de Lera Acedo et al. 2022; Bull et al. 2025; Bye et al. 2026) and observations across the electromagnetic spectrum, further sharpens our picture of the characteristic host halo masses, star-formation efficiencies and emission properties of the first galaxies (Bevins et al. 2024; Pochinda et al. 2024; Dhandha et al. 2025; Sims et al. 2025b).

Direct detection nevertheless remains a formidable challenge, primarily due to the presence of astrophysical foregrounds that are approximately 10^4–10^5 times brighter than the expected cosmological signal (Shaver et al. 1999). In an idealised case, component separation could simply be achieved through a Fourier decomposition along the line-of-sight (frequency) axis, in which spectrally smooth foreground emission would be concentrated within a small number of low-$k_\parallel$ modes, whereas the more rapidly decorrelating 21-cm signal would be distributed over a broader range of spectral modes (for a review, see Liu & Shaw 2020). In practice, however, chromaticity in both the baseline response and the primary beam redistributes angular structure in the foreground sky into spectral structure within the measured visibilities (Morales et al. 2012; Shaw et al. 2014; Liu et al. 2014a), resulting in a characteristic wedge-shaped region of foreground contamination in cylindrical Fourier space, $(k_\perp, k_\parallel)$ (Datta et al. 2010; Parsons et al. 2012; Liu et al. 2014b). This idealised separation is further eroded by wide-field sensitivity to diffuse emission near the horizon (Thyagarajan et al. 2015), signal-chain reflections (Kern et al. 2019; O'Hara et al. 2024) and mutual coupling between array elements (Rath et al. 2025; O'Hara et al. 2025).

Traditionally, methods developed to mitigate this contamination can be broadly divided into two approaches (Chapman et al. 2016). The first, foreground avoidance, restricts the analysis to combinations of the data expected to contain a comparatively small foreground contribution. This can be achieved through hard cuts or filters informed by the expected sky and instrumental response (Parsons et al. 2012; Thyagarajan et al. 2015; Rath et al. 2025), or through statistically optimised decompositions that retain modes with high expected signal-to-foreground ratios (Shaw et al. 2014). Because the contaminated subspace is imperfectly known, these methods often adopt conservative selection thresholds, discarding cosmological information and reducing sensitivity (Pober et al. 2014; Chapman et al. 2016).

The second, foreground removal, seeks to recover cosmological information from contaminated modes by explicitly modelling the foreground contribution. Approaches range from smooth parametric functions fitted along individual lines of sight (Liu et al. 2009) to Gaussian Process Regression (GPR). Rather than explicitly modelling their physical origin, GPR, a non-parametric Bayesian framework, separates the components using phenomenological covariance kernels chosen to represent their expected spectral coherence (Mertens et al. 2018). The recovery therefore depends on those adopted kernels, whose misspecification can cause signal loss or bias the recovered power spectrum (Kern & Liu 2021; Liu et al. 2026).

Given the challenges of cosmological signal extraction, physically motivated Bayesian forward modelling is emerging as one of the most statistically rigorous approaches across the field (Anstey et al. 2021; Tutt et al. 2026a). By propagating parameterised sky and instrument models through the signal chain to the measured data, these approaches capture component uncertainties and degeneracies within a joint posterior. For interferometers, such models now encompass inference of the 21-cm signal and foregrounds (Sims et al. 2016; Sims et al. 2019; Sims & Pober 2019; Burba et al. 2023), calibration (Sims et al. 2022a,b), primary-beam (Glasscock et al. 2024; Wilensky et al. 2024, 2025; Kern 2025), and signal-chain systematics (Murphy et al. 2024; Dutta et al. 2026).

Combining these components into unified end-to-end frameworks faces two scaling problems: large data volumes and high-dimensional models make likelihood evaluation expensive, while the dimensionality and degeneracy of the resulting posterior impede its exploration. Efforts across the field therefore tackle these challenges from two complementary directions: reducing model dimensionality and accelerating likelihood evaluation and posterior sampling.

On the modelling side, Kern (2025) developed a mathematical framework in which adapted angular and spectral bases, together with delay filtering, provide a compact representation of the sky-to-visibility mapping, with the corresponding trade-off that cosmological information in the filtered low-$k_\parallel$ modes is not retained. Alternatively, Kennedy et al. (2023) derived compact bases from simulated visibility-frequency covariances, although the resulting coefficients lack direct physical interpretation and the basis, conditioned on a single sky realisation and beam model, is not guaranteed to span the full range of plausible foreground responses (Burba et al. 2024).

On the computational side, radio-interferometric inference is particularly well matched to high-throughput hardware accelerators such as graphics processing units (GPUs), with forward models dominated by repeated linear-algebra operations that can be evaluated concurrently across baselines, frequencies and observing times. Kern (2025) and Tutt et al. (2026c) have already demonstrated these computational gains in interferometric and global 21-cm inference, respectively. These hardware gains are complemented by modern programming frameworks (Bradbury et al. 2018; Paszke et al. 2019), which enable automatic differentiation through the forward model, allowing its direct integration with gradient-informed samplers such as Hamiltonian Monte Carlo (Duane et al. 1987; Neal 2011; Kern 2025). Alternatively, Gibbs sampling enables efficient exploration of high-dimensional posteriors by successively sampling from simpler conditional distributions (Kennedy et al. 2023; Burba et al. 2024; Dutta et al. 2026).

Beyond exploring the joint posterior, such frameworks can, in principle, provide access to the Bayesian evidence, an often underutilised measure of the overall support afforded to a model by the observed data. Evidence ratios between competing models allow the statistical preference for a signal detection to be quantified and structure attributed to the signal to be tested against alternative foreground or instrumental models (Sims & Pober 2020), while Bayesian validation frameworks rigorously assess susceptibility to biased recovery or spurious detections (Sims et al. 2025a).

Access to the Bayesian evidence nevertheless comes at a substantial additional computational cost. For Nested Sampling (Skilling 2006), a commonly employed evidence-evaluation algorithm, this cost scales as $O(N^3_{\rm Dim})$ with sampled dimensionality (Handley et al. 2015), making its implementation impractical beyond $\sim 10^2$–10^3 parameters (Lovick et al. 2026). One route around this limitation is analytical marginalisation, through which linear model components can be integrated out exactly, reducing the space explored numerically while propagating their uncertainty into the marginal likelihood for the remaining parameters (Luger et al. 2017; Greengard et al. 2022). Although highly effective, applying analytical marginalisation to such large-scale Bayesian linear inverse problems shifts the computational bottleneck from posterior exploration to likelihood evaluation, which requires a dense linear solve scaling as $O(N^3_{\rm Marg})$, with $N_{\rm Marg}$ set by either the marginalised parameter dimension or the data dimension.

In this work, motivated by the wide-field foreground-modelling requirements identified by Burba et al. (2023, hereafter B23), we present an extension of BayesEoR (Sims et al. 2024) that renders cosmological posterior sampling and evidence evaluation with physically motivated, wide-field forward models computationally tractable. Propagating the foreground prior through the physical sky-to-visibility mapping defines its prior-predictive visibility covariance, whose noise-informed Karhunen–Loève decomposition yields a statistically optimal, observation-specific latent representation. The truncation retains orthogonal modes anchored to their physical sky and instrumental origins while discarding those whose visibility-space contributions fall below the thermal-noise level. Distributed GPU acceleration with JAXMG (Tutt & Wiersema 2026) enables this basis to be constructed once before inference, with its cost amortised over all subsequent likelihood evaluations, after which analytical marginalisation propagates the same noise-relevant foreground uncertainty as the original physical representation, but at substantially reduced computational cost. Together, in reducing dimensionality of full-sky inference by orders of magnitude, these developments render joint posterior and Bayesian-evidence evaluation practical.

We demonstrate the framework using simulated HERA-like measurements, performing joint foreground and 21-cm analyses to verify that the reduced representation preserves the information required for accurate power-spectrum recovery. The resulting posterior and Bayesian-evidence calculations provide an end-to-end validation of the theoretical construction and enable direct tests of foreground-model completeness.

Section 2 describes the simulated HERA-like observations, introducing the instrumental configuration, sky components and their mapping into interferometric visibilities. Section 3 develops the joint foreground and cosmological forward model, introducing the mathematical formalism underpinning the statistically optimal latent representation and its incorporation into analytical marginalisation. Section 4 examines how foreground and cosmological model extent affect power-spectrum recovery, demonstrates the role of Bayesian validation in identifying misspecification, and quantifies the computational gains. We then discuss extensions to instrumental uncertainty, learned compression and real observations, before concluding in Section 5.

2 INTERFEROMETRIC OBSERVATIONS

We first outline the simulated HERA-like observations used to validate the framework, summarised in Figure 1. Sections 2.1 and 2.2 describe the instrumental configuration and adopted sky realisation, before Section 2.3 introduces the full-fidelity mapping of the sky through the instrument into simulated visibilities. We adopt an observational configuration closely matched to B23, enabling a controlled comparison with that study.

2.1 Instrument Model

The adopted instrumental response comprises two components: the array and observing configuration (Section 2.1.1), which sets the baseline, time and frequency sampling, and the primary beam (Section 2.1.2), which determines the directional, frequency-dependent weighting of the sky.

2.1.1 Array and Observational Setup

We adopt a HERA-like array in the Karoo Desert, South Africa (DeBoer et al. 2017), comprising 37 antennas arranged on a perfectly redundant hexagonal grid with a centre-to-centre spacing of 14.6 m. We hereafter work in the topocentric frame of the instrument, in which antennas at positions $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ define the physical baseline $\boldsymbol{b}_{ij}$ and its corresponding interferometric (u, v, w) coordinate, $\boldsymbol{u}_{ij}(\nu)$, as

$$\boldsymbol{b}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j, \qquad \boldsymbol{u}_{ij}(\nu) = \frac{\nu}{c}\boldsymbol{b}_{ij} = \left(u_{ij}, v_{ij}, w_{ij}\right), \tag{1}$$

Longer baselines therefore probe progressively finer angular structure and those sharing the same separation and orientation sample the same angular Fourier mode on the sky and are hence termed redundant. Under the idealised instrumental assumptions adopted here, they therefore differ only through independent thermal-noise realisations and can consequently be coherently averaged. We retain the same 30 unique cross-correlation groups satisfying $|\boldsymbol{b}_{ij}| \leq 40$ m adopted by B23 to limit computational cost, although overcoming the dimensionality bottleneck that necessitated this restriction is a central motivation of the framework developed here.

We consider an unphased zenith drift scan observation comprising 47 integrations of 21.5 s, corresponding to a total duration of $\sim$ 16.9 min. The observation contains 39 frequency channels spanning 157.2–160.9 MHz, each with a width of 97.7 kHz. Together with the 30 unique baseline groups, this produces 54,990 complex visibilities.

We additionally inject thermal noise at the level expected after 1,000 h of integration on the same field, modelled as a homogeneous, zero-mean Gaussian noise with covariance

$$\mathbf{N} \equiv \left\langle \boldsymbol{n}\boldsymbol{n}^{\dagger} \right\rangle = \sigma_N^2 \mathbf{I}, \qquad \sigma_N = 6.464 \times 10^{-3}\ \text{mK sr}, \tag{2}$$

where σ_N is the per-visibility noise standard deviation.

2.1.2 Primary Beam Model

We assign each antenna an identical scalar Stokes I primary-beam power response, $P(\hat{\boldsymbol{s}}, \nu)$, modelled using a frequency-dependent Airy pattern corresponding to a circular aperture of diameter $D = 14.2$ m. This provides a first-order approximation to the response of HERA's circular dishes (Neben et al. 2016) with the chromatic sidelobe structure required to fairly study the coupling of wide-field foreground emission into the measured visibilities.

Together, the longest retained baseline and the Fourier support of the primary beam determine the finest angular structure to which the measured visibilities are sensitive. For the adopted configuration, the nominal Nyquist requirement of an angular resolution finer than $\sim 1.1^\circ$ is satisfied by a HEALPix grid with $N_{\text{side}} = 64$. However, adequate angular sampling does not necessarily ensure sufficient accuracy in the discretised visibility integral, as demonstrated by Kittiwisit et al. (2025). We find that $N_{\text{side}} = 128$, corresponding to a characteristic pixel scale of $\sim 0.46^\circ$, is required for reliable power-spectrum recovery in the configuration considered here and therefore adopt this resolution for the forward model.

In this work, we assume a perfectly calibrated instrument. Ionospheric effects, radio-frequency interference and additional signal-chain systematics are neglected, with their incorporation left to future extensions of the framework.

2.2 Sky Model

The simulated high-fidelity sky is represented by the total-intensity brightness $I(\hat{\boldsymbol{s}}, \nu, t)$, combining an astrophysical foreground model,

$I_{\rm FG}(\hat{\boldsymbol{s}}, \nu, t)$, comprising diffuse emission, an extensive population of point sources and a catalogue of bright A-team sources (Section 2.2.1) with a cosmological 21-cm brightness field, $I_{21}(\hat{\boldsymbol{s}}, \nu, t)$ (Section 2.2.2). At each time, a sky direction is represented in the topocentric frame by the unit vector $\hat{\boldsymbol{s}}(t) = (l(t), m(t), n(t))$, where $n = (1 - l^2 - m^2)^{1/2}$ over the visible hemisphere. Consistent with the beam model, we simulate an unpolarised (Stokes I) sky, although polarised emission could also be simulated and modelled in this framework. The observation is centred on HERA Field 1, which passes through zenith at the centre of the observation, corresponding to (RA, Dec) = (2 h, −30.72°).

In order to assess how the angular extent of the foreground sky affects model dimensionality and susceptibility to misspecification, we generate sky realisations with angular diameters $\theta_{\rm FG} = 20°, 40°, \ldots, 180°$. Given the drift-scan observation, each realisation contains all foreground emission passing within $\theta_{\rm FG}/2$ of zenith over the duration of the observation. The simulated sky components are generated at HEALPix $N_{\rm side} = 256$, thereby providing a stress test of the $N_{\rm side} = 128$ grid adopted by the forward model.

2.2.1 Foreground Emission

The diffuse sky is generated from the 2008 Global Sky Model (GSM2008; de Oliveira-Costa et al. 2008) using PYGDSM (Price 2016), which we select over its 2016 counterpart (Zheng et al. 2017). This choice is motivated by recent absolutely calibrated measurements from single-antenna radiometers that identify offset and scale discrepancies in GSM2016 and indicate that GSM2008 provides a more accurate absolute temperature scale across the observing band considered here (McKay et al. 2026; Anstey et al., in prep.).

To introduce realistic compact-source contributions into the simulated sky, we first generate a statistical population of point sources from the 154-MHz differential source counts measured by Franzen et al. (2019) using the GaLactic and Extragalactic All-sky MWA (GLEAM) survey.[1] Sources are distributed isotropically, with reference flux densities, S_{154}, drawn from these counts and spectra given by

$$S_\nu = S_{154}\left(\frac{\nu}{154\,\mathrm{MHz}}\right)^{\alpha}, \qquad \alpha \sim \mathcal{N}(-0.8, 0.05^2). \tag{3}$$

Additionally, the sky's brightest radio sources are represented individually using the A-team catalogue, which provides their positions, reference flux densities and spectral indices (Hurley-Walker et al. 2017; McKinley et al. 2015). The resulting 180° drift-scan field contains 400,125 statistically generated GLEAM-like sources and four A-team sources: 3C 444, Pictor A, the Crab Nebula and Fornax A.

2.2.2 Cosmological 21-cm Signal

We simulate the cosmological 21-cm contribution as a three-dimensional Gaussian brightness-temperature field, sampled through an independent full-sky map at each frequency channel. Adopting uncorrelated pixel amplitudes with standard deviation $\sigma_{21} = 6.4$ mK gives the flat dimensional power spectrum $P_{21}(k) = \sigma_{21}^2 V_{\rm vox}$, where the comoving voxel volume is

$$V_{\rm vox} = \Omega_{\rm pix}\Delta\nu\, D_{\rm M}^2(z_{\rm c})\frac{c(1+z_{\rm c})^2}{H_0\nu_{21}E(z_{\rm c})}, \qquad \Omega_{\rm pix} = \frac{4\pi}{12N_{\rm side}^2}. \tag{4}$$

[1] This implementation follows the HERA VALIDATION-SIM prescription; see https://github.com/HERA-Team/validation-sim.

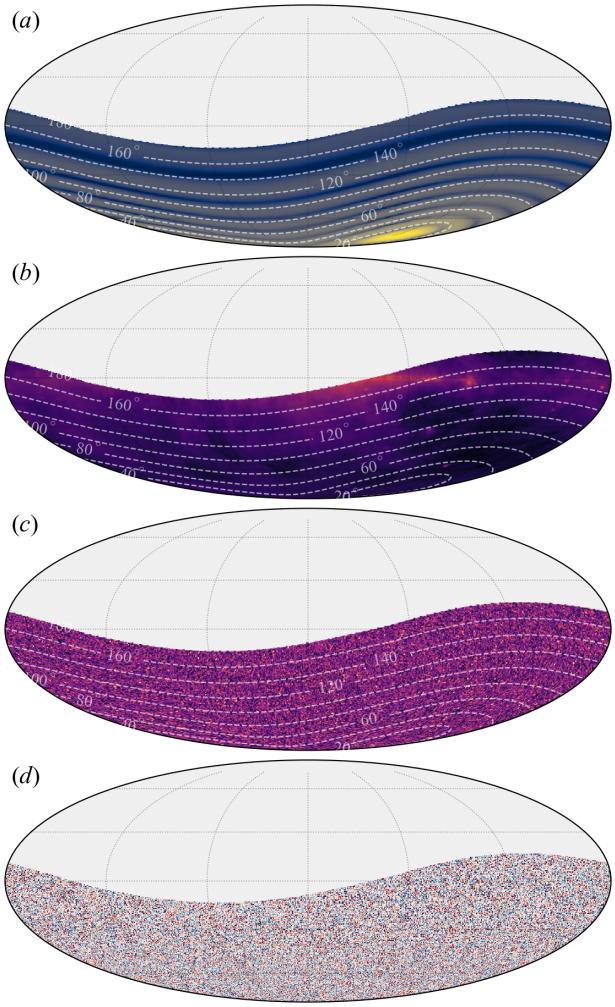


Figure 1. Overview of the components of the full-fidelity simulated drift-scan visibilities, shown in Galactic coordinates (l, b) at the central frequency of 159.082 MHz: (a) the time-integrated response of the 14.2-m HERA-like Airy beam, (b) diffuse foreground emission modelled using GSM2008, (c) the statistically generated extragalactic source population and catalogued A-team sources, and (d) the injected cosmological 21-cm realisation. White dashed contours in panels (a)–(c) mark the discrete foreground extents, $\theta_{\rm FG} = 20°, 40°, \ldots, 180°$, adopted for the suite of simulations and forward models to assess the effect of sky extent on model dimensionality and susceptibility to misspecification.

Here, $D_{\rm M}$ denotes the transverse comoving distance and $E(z) = H(z)/H_0$. Adopting the cosmological parameters from Planck Collaboration et al. (2020) and evaluating at the central frequency, $\nu_{\rm c} = 159.082$ MHz ($z_{\rm c} = 7.929$), gives $V_{\rm vox} = 2.159 \times 10^3$ Mpc^3. While $V_{\rm vox}$ is redshift and therefore frequency dependent, across the relatively narrow observing band considered here it varies by only $\sim 2.5\%$; we therefore adopt its central-frequency value throughout. The adopted Gaussian field thus corresponds to $P_{21} = 8.84 \times 10^4$ mK^2 Mpc^3 and the dimensionless power spectrum

$$\Delta_{21}^2(k) = \frac{k^3}{2\pi^2}P_{21}(k) = 4.48\left(\frac{k}{0.1\,\mathrm{Mpc}^{-1}}\right)^3 \mathrm{mK}^2. \tag{5}$$

2.3 Wide-Field Interferometric Measurement Equation

To motivate the wide-field forward model developed in Section 3, we first set out the standard interferometric measurement equation, making explicit the full mapping from sky emission to observed visibilities which for the simulations used in this work is evaluated using PYUVSIM (Lanman et al. 2019).

At its most basic level, an interferometric visibility is formed by cross-correlating the complex voltage vectors $\boldsymbol{v}_i(\nu,t)$ and $\boldsymbol{v}_j(\nu,t)$ measured by antennas i and j, respectively, and averaging the result within each integration and frequency channel to give the visibility coherency matrix,

$$\mathbf{V}_{ij}(\nu,t) = \left\langle \boldsymbol{v}_i(\nu,t)\boldsymbol{v}_j^{\dagger}(\nu,t) \right\rangle. \tag{6}$$

The power of radio interferometry lies in the relationship between these voltage cross-correlations and the angular brightness distribution of the sky, formalised by the van Cittert–Zernike theorem (van Cittert 1934; Zernike 1938) and its generalisation to the full spherical geometry of a wide field of view (Carozzi & Woan 2009). Under the scalar Stokes I assumptions adopted here, this relationship is expressed by the wide-field measurement equation

$$V_{ij}(\nu,t) = \int_{\Omega_{\mathrm{vis}}(t)} P(\hat{\boldsymbol{s}},\nu) I(\hat{\boldsymbol{s}},\nu,t) \exp\left[-2\pi i \boldsymbol{u}_{ij}(\nu)\cdot\hat{\boldsymbol{s}}\right] \mathrm{d}\Omega, \tag{7}$$

Here, the primary-beam power response $P(\hat{\boldsymbol{s}},\nu)$ weights the direction-, frequency- and time-dependent sky brightness $I(\hat{\boldsymbol{s}},\nu,t)$, as defined in Sections 2.1 and 2.2, while the geometric phase $\boldsymbol{u}_{ij}\cdot\hat{\boldsymbol{s}} = u_{ij}l + v_{ij}m + w_{ij}n$ relates the length and orientation of each baseline to the angular structure it samples across the celestial sphere.

To provide intuition for the linear operators adopted by BAYESEOR throughout the remainder of this work, we express the same relation for a discretised sky of N_{pix} equal-area pixels,

$$V_{ij}(\nu,t) = \sum_{p=1}^{N_{\mathrm{pix}}} F^{-1}_{ij,p}(\nu,t) P(\hat{\boldsymbol{s}}_p(t),\nu) s_p(\nu), \tag{8}$$

where $s_p(\nu)$ is the brightness of fixed celestial pixel p and $\hat{\boldsymbol{s}}_p(t)$ is its direction in the topocentric frame at time t. The remaining term, $F^{-1}_{ij,p}(\nu,t)$, is the full-sky fringe operator, defined as

$$F^{-1}_{ij,p}(\nu,t) = \Delta\Omega \exp\left[-2\pi i \boldsymbol{u}_{ij}(\nu)\cdot\hat{\boldsymbol{s}}_p(t)\right] \tag{9}$$

where $\Delta\Omega$ is the solid angle of each pixel. The discretised measurement equation can therefore be written compactly as

$$\boldsymbol{m}_{\nu t} = \mathbf{F}^{-1}_{\nu t}\mathbf{P}_{\nu t}\boldsymbol{s}_{\nu} \equiv \mathbf{R}_{\nu t}\boldsymbol{s}_{\nu}, \tag{10}$$

where $\boldsymbol{m}_{\nu t}$ is the model visibility vector and $\mathbf{R}_{\nu t} \equiv \mathbf{F}^{-1}_{\nu t}\mathbf{P}_{\nu t}$ is the combined sky-to-visibility response. Stacking over frequency and time gives $\boldsymbol{d} = \mathbf{R}\boldsymbol{s} + \boldsymbol{n}$, where $\boldsymbol{n}$ is the simulated thermal-noise realisation defined in Section 2.1.1. This linear model forms the basis of the inference framework developed in Section 3.

3 INTERFEROMETRIC FORWARD MODELS

In this section, we develop the joint Bayesian framework used to infer the cosmological 21-cm signal in the presence of wide-field foreground emission, with the complete forward-modelling and inference workflow summarised in Figure 2. We first introduce the motivation and statistical foundations of physically motivated Bayesian forward modelling in Section 3.1. Section 3.2 then presents the wide-field foreground model and the construction of the Karhunen–Loève latent representation introduced in this work, together with a discussion of its computational benefits. We outline the existing BAYESEOR model adopted for the cosmological 21-cm brightness field in Section 3.3, before combining both contributions within the analytically marginalised likelihood used for posterior sampling and Bayesian-evidence evaluation in Section 3.4. Finally, Section 3.5 outlines the Bayesian validation framework used to assess the resulting foreground and 21-cm inferences.

3.1 Bayesian Forward Modelling

At its core, physically motivated forward modelling constructs a generative model, $\mathcal{M}$, that transforms a parameterised description of the physical system, $\boldsymbol{\theta}$, into a prediction for the observable data, $\boldsymbol{m}(\boldsymbol{\theta})$. Given this mapping, Bayesian inference uses the observed data to update a prior distribution, $\pi(\boldsymbol{\theta} \mid \mathcal{M})$, describing our initial knowledge, or lack thereof, of the physically realisable parameter values. The resulting joint posterior, $P(\boldsymbol{\theta} \mid \boldsymbol{d}, \mathcal{M})$, is given by Bayes' theorem:

$$P(\boldsymbol{\theta} \mid \boldsymbol{d}, \mathcal{M}) = \frac{\mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta} \mid \mathcal{M})}{\mathcal{Z}(\boldsymbol{d} \mid \mathcal{M})}, \tag{11}$$

where the likelihood, $\mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M})$, quantifies the probability of obtaining the data for a given parameter realisation under the assumed noise model. For the Gaussian noise model adopted in Section 2.1.1, the likelihood for the complex visibility data takes the form

$$\mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M}) = \frac{1}{\pi^{N_{\mathrm{d}}} \det \mathbf{N}} \exp\left[-(\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\theta}))^{\dagger} \mathbf{N}^{-1} (\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{\theta}))\right], \tag{12}$$

where N_{d} is the number of complex visibilities and $\mathbf{N}$ is their noise covariance. Finally, $\mathcal{Z}(\boldsymbol{d} \mid \mathcal{M})$ is the Bayesian evidence, which quantifies the overall support afforded to a model by the observed data:

$$\mathcal{Z}(\boldsymbol{d} \mid \mathcal{M}) = \int \mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta} \mid \mathcal{M}) \mathrm{d}\boldsymbol{\theta}. \tag{13}$$

Although the measurement operator introduced in Section 2.3 maps a discretised sky to the measured visibilities, assigning an independent parameter to every pixel and frequency channel produces a prohibitively high-dimensional inverse problem with no intrinsic distinction between foreground and cosmological structure. We therefore seek physics-informed compact parameterisations that preserve each contribution's physical mapping into the data while enabling joint, tractable inference of their structure and degeneracies.

3.2 Wide-field Foreground Models

3.2.1 Beyond Flat Sky Approximations

Previous implementations of BAYESEOR have parameterised the intrinsic foreground structure on a rectilinear uv grid, corresponding to Fourier modes defined on a tangent plane. The resulting modes are mapped into the image domain through a two-dimensional discrete Fourier transform ($\mathbf{F}'$), weighted by the primary-beam operator ($\mathbf{P}$), and finally propagated to the sampled visibilities through a non-uniform Fourier operator ($\mathbf{F}^{-1}$) (Sims et al. 2019; Sims & Pober 2019; Burba et al. 2023). This construction is, however, only mathematically valid under the flat-sky approximation, for which the field of view is sufficiently narrow that the non-coplanar phase term can be neglected, $(\sqrt{1-l^2-m^2}-1)w \simeq 0$.

To address the wide-field modelling requirements identified by

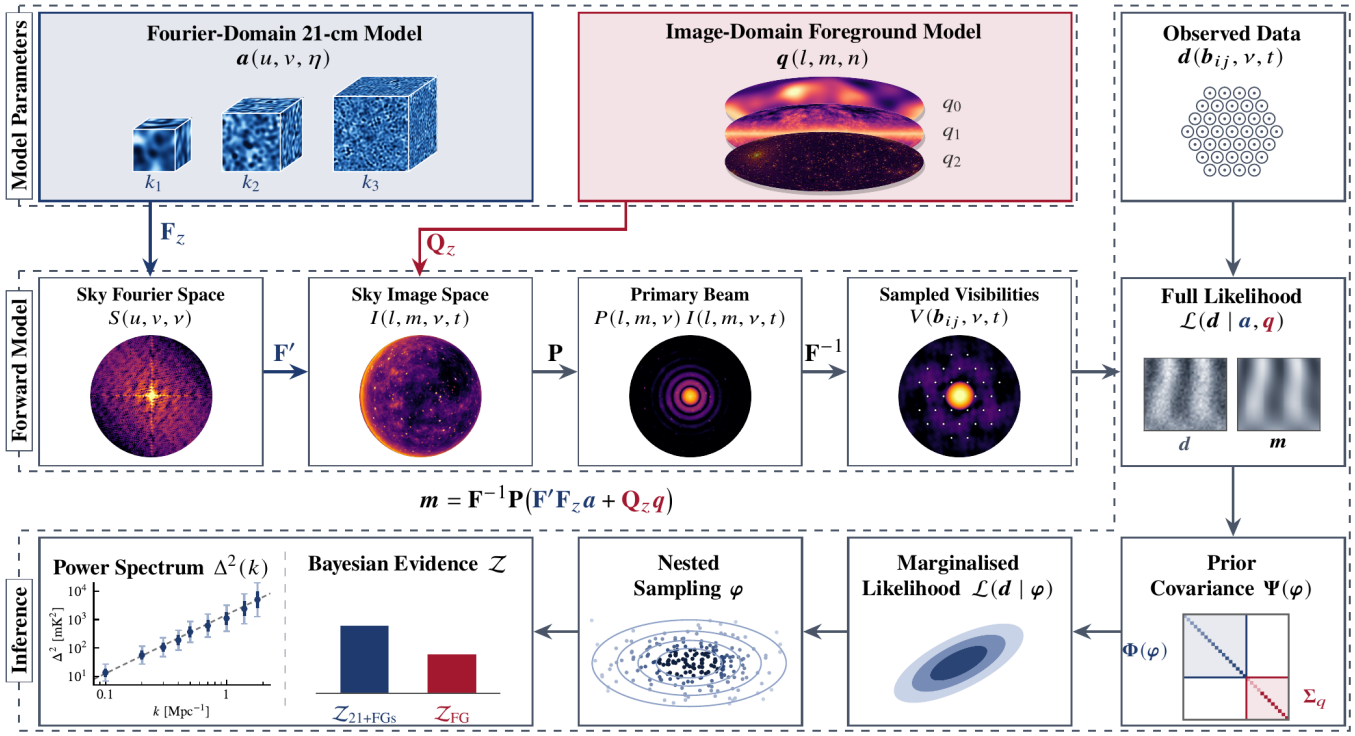


Figure 2. Schematic of BayesEoR's joint forward model and inference workflow. The Fourier-domain 21-cm coefficients, $\boldsymbol{a}$, and image-domain foreground coefficients, $\boldsymbol{q}$, are transformed into a shared sky model, weighted by the primary beam and mapped to the sampled visibilities for comparison with the observed data. Their covariance priors enable analytical marginalisation of the linear coefficients, leaving only the 21-cm power-spectrum amplitudes, $\boldsymbol{\varphi}$, to be explored with Nested Sampling, which enables simultaneous posterior power-spectrum exploration and the Bayesian evidence evaluation. The illustrative Fourier-domain cubes were generated from a 21cmSPACE simulation (Cohen et al. 2018).

(B23), we instead parameterise the foreground brightness directly, pixel by pixel, on a spherical image-domain grid. The primary beam and full three-dimensional fringe response can then be applied directly to the physical foreground sky, removing the initial tangent-plane transform, $\mathbf{F}'$, and simplifying the original BayesEoR foreground operator, $\mathbf{T}_{\rm FG}$.

To inform the model of the contrast required to distinguish foreground emission from the more rapidly decorrelating cosmological 21-cm signal, we exploit the smooth spectral structure of low-frequency foreground emission. At the frequencies relevant to the DA, CD and EoR, the foreground sky is dominated by Galactic diffuse and extragalactic synchrotron emission (Rybicki & Lightman 1979; Santos et al. 2005; Robins et al. 2026), both of which vary smoothly across the relatively narrow observing band. Following Sims & Pober (2019) and B23, we therefore model the foreground brightness in pixel p as

$$s_{{\rm FG},p}(\nu; \boldsymbol{q}_p) = q_{0,p} + q_{1,p}\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm GDSE}} + q_{2,p}\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm EGS}}, \tag{14}$$

Here, ν_0 is the reference frequency, while $\beta_{\rm GDSE} = -2.63$ and $\beta_{\rm EGS} = -2.82$ are spectral indices for Galactic diffuse synchrotron emission and extragalactic radio sources, respectively (Mozdzen et al. 2017; Lane et al. 2014). The vector $\boldsymbol{q}_p = (q_{0,p}, q_{1,p}, q_{2,p})^{\mathsf{T}}$ contains the corresponding basis amplitudes in pixel p. Substituting it into the discretised wide-field measurement equation, the contribution of pixel p to the foreground visibility measured by baseline ij at frequency ν and time t is

$$m^{(p)}_{{\rm FG},ij}(\nu, t; \boldsymbol{q}_p) = \Delta\Omega \exp\left[-2\pi i \boldsymbol{u}_{ij}(\nu)\cdot \hat{\boldsymbol{s}}_p(t)\right] \times P(\hat{\boldsymbol{s}}_p(t), \nu) s_{{\rm FG},p}(\nu; \boldsymbol{q}_p), \tag{15}$$

The pixel solid angle, primary-beam response and full-sky fringe term are defined in Section 2.3. Since both the spectral function and its instrumental mapping are linear in the coefficients, summing over all pixels and stacking the resulting visibilities across baselines, frequencies and times gives the single linear model

$$\boldsymbol{m}_{\rm FG}(\boldsymbol{q}) = \mathbf{F}^{-1}\mathbf{P}\mathbf{Q}_z\boldsymbol{q} \equiv \mathbf{T}_{\rm FG}\boldsymbol{q}. \tag{16}$$

Here, $\boldsymbol{q}$ concatenates the amplitudes of each spectral component across all $N_{\rm pix}$ image-domain pixels, while $\mathbf{Q}_z$ applies the spectral basis of Equation 14, such that the full foreground operator has dimensions

$$\mathbf{T}_{\rm FG} \in \mathbb{C}^{N_{\rm d}\times N_{\rm FG,param}}, \qquad N_{\rm FG,param} \equiv 3N_{\rm pix}, \qquad N_{\rm d} = N_{\rm bl}N_\nu N_t. \tag{17}$$

Following the discussion in Section 2.1.2, the foreground model is evaluated on a HEALPix grid with $N_{\rm side} = 128$. Mirroring the simulated foreground skies described in Section 2.2, we construct corresponding models with angular diameters $\theta_{\rm FG} = 20°, 40°, \ldots, 180°$. As shown in the lower panel of Figure 4, the resulting dimensionality reaches $N_{\rm FG,param} = 300,767$ foreground parameters for the full 180° drift-scan field. We additionally adopt a broad Gaussian prior,

$$\pi(\boldsymbol{q}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_q), \qquad \boldsymbol{\Sigma}_q = {\rm diag}\left(\sigma_0^2\mathbf{I}, \sigma_1^2\mathbf{I}, \sigma_2^2\mathbf{I}\right), \tag{18}$$

with standard deviations

$$(\sigma_0, \sigma_1, \sigma_2) = (8.7,\ 2.5 \times 10^3,\ 1.4 \times 10^3)\ \mathrm{K}. \tag{19}$$

We highlight that the prior adopted in this work is spatially uninformative and encodes no knowledge of the physical morphology of the diffuse foreground emission or the locations of bright extragalactic sources. Instead, its width is set by the overall variation of each foreground component across the sky, with this variation treated as uncertainty in every pixel amplitude. This deliberately conservative choice provides a stringent test of the method but is not required by the KL construction, as discussed further in Sections 3.2.2 and 3.2.3. Assigning the same prior uncertainty to every pixel does, however, make the resulting basis and latent compression independent of the drift-scan field centre, allowing them to be reused across observations with the same instrumental configuration, modelled foreground extent, temporal and frequency sampling, and noise weighting.

3.2.2 *Karhunen–Loève Latent Space*

Given that the foreground amplitudes are analytically marginalised rather than sampled directly, the inference depends on the uncertainty they induce in visibility space rather than their individual values. As formalised in Section 3.4, this dependence is encoded through the prior-predictive visibility covariance

$$\mathbf{C}_{\mathrm{FG}} = \mathbf{T}_{\mathrm{FG}} \boldsymbol{\Sigma}_q \mathbf{T}_{\mathrm{FG}}^{\dagger}. \tag{20}$$

Consequently, the original foreground parameterisation need not be retained explicitly during inference; only a basis spanning the foreground covariance must be preserved (Figure 3), with the minimum number of modes set by the covariance rank:

$$r_{\mathrm{FG}} \equiv \mathrm{rank}(\mathbf{C}_{\mathrm{FG}}) = \mathrm{rank}\left(\mathbf{T}_{\mathrm{FG}} \boldsymbol{\Sigma}_q^{1/2}\right) \leq \min(N_{\mathrm{d}}, N_{\mathrm{FG,param}}). \tag{21}$$

An exact latent representation therefore requires no more than $\min(N_{\mathrm{d}}, N_{\mathrm{FG,param}})$ independent modes. Its likelihood-evaluation cost is consequently bounded by that of related covariance-based methods and is typically reduced further by degeneracies in the physical mapping. The resulting computational gains are discussed in Section 3.2.4.

These degeneracies are, however, difficult to identify a priori because the chromatic primary beam and fringe response produce a complex mapping from angular and spectral foreground structure across the sky into visibilities sampled over baselines, frequencies and observing times. Compact representations are therefore commonly learned from forward-model simulations. In interferometric analyses, PCA modes have been derived from the frequency covariance of a single simulated visibility realisation (Kennedy et al. 2023; Burba et al. 2024). Related global 21-cm methods instead construct bases from ensembles of foreground and instrumental realisations (Tauscher et al. 2018; Hibbard et al. 2020; Saxena et al. 2023).

Alternatively, deep generative models such as variational autoencoders (Kingma & Welling 2019) have been previously utilised to learn nonlinear representations, such as simulation-informed 21-cm covariance kernels for GPR (Acharya et al. 2024). Although potentially compact, the ability of both PCA- and VAE-derived representations to span the full foreground uncertainty in visibility space depends on the simulations used in their construction and cannot be quantified analytically. VAEs introduce further dependence on the adopted architecture and optimisation, while their nonlinear mappings do not preserve the structure required for exact analytical marginalisation.

We therefore address this large-scale Bayesian linear inverse problem using a prior- and noise-informed Karhunen–Loève decomposition, which provides a statistically optimal linear representation of the foreground covariance accessible to the observation (Karhunen 1947; Loève 1945; Tegmark et al. 1997; Spantini et al. 2015). We first weight the physical foreground mapping by its prior covariance and noise-whiten its visibility-space response. The KL basis can then either be constructed directly through the SVD of the resulting operator,

$$\mathbf{A} \equiv \mathbf{N}^{-1/2} \mathbf{T}_{\mathrm{FG}} \boldsymbol{\Sigma}_q^{1/2} = \mathbf{U}\mathbf{S}\mathbf{V}^{\dagger}, \tag{22}$$

where each column of $\mathbf{V}$ describes an orthonormal combination of foreground coefficients scaled by their prior standard deviations, while the corresponding column of $\mathbf{U}$ describes its noise-whitened visibility-space response. The singular values in $\mathbf{S}$ quantify their foreground amplitudes relative to the thermal noise. Alternatively, this can be achieved through an eigendecomposition of the Hermitian positive-semidefinite prior-predictive covariance in noise-whitened visibility space,

$$\mathbf{C}_{\mathrm{FG},w} \equiv \mathbf{N}^{-1/2} \mathbf{C}_{\mathrm{FG}} \mathbf{N}^{-1/2} = \mathbf{A}\mathbf{A}^{\dagger} = \mathbf{U}'\boldsymbol{\Lambda}'\mathbf{U}'^{\dagger}, \tag{23}$$

where the columns of $\mathbf{U}'$ are the covariance eigenmodes and the non-negative diagonal elements λ_i' of $\boldsymbol{\Lambda}'$ give their prior-predictive variance relative to the noise.

The mathematical equivalence of the operator and covariance formulations, and hence their visibility-space bases, can be trivially demonstrated by:

$$\begin{aligned} \mathbf{C}_{\mathrm{FG},w} &= \mathbf{A}\mathbf{A}^{\dagger} \\ &= \mathbf{U}\mathbf{S}\mathbf{V}^{\dagger}\mathbf{V}\mathbf{S}\mathbf{U}^{\dagger} \\ &= \mathbf{U}\mathbf{S}^2\mathbf{U}^{\dagger}. \end{aligned} \tag{24}$$

Comparing this expression with Equation 23 gives $\mathbf{U}' = \mathbf{U}$ and $\boldsymbol{\Lambda}' = \mathbf{S}^2$, confirming that both constructions recover the same visibility-space basis. We hereafter drop the primes and denote the common basis and variance matrix by $\mathbf{U}$ and $\boldsymbol{\Lambda}$. The preferred formulation is determined by the dimensions and associated requirements of keeping the matrix being decomposed in memory. The operator formulation ($N_{\mathrm{d}} \times N_{\mathrm{FG,param}}$) is generally preferred when $N_{\mathrm{FG,param}} < N_{\mathrm{d}}$, whereas the covariance formulation ($N_{\mathrm{d}} \times N_{\mathrm{d}}$) is preferred when $N_{\mathrm{FG,param}} > N_{\mathrm{d}}$.

Regardless of the formulation adopted, the noise-whitened eigenvalues allow the statistical effect of truncating the basis to be quantified directly (Zhou et al. 2018). With the eigenvalues ordered from largest to smallest, retaining only the first r modes discards a total prior-predictive foreground variance $\sum_{i>r} \lambda_i$ in noise-whitened visibility space. We therefore choose the smallest retained rank satisfying

$$\epsilon_{\mathrm{RMS}}(r) = \left[\frac{1}{N_{\mathrm{d}}} \sum_{i>r} \lambda_i\right]^{1/2} \leq f, \tag{25}$$

Here, $N_{\mathrm{retained}} \equiv r$, while f specifies the maximum discarded RMS as a fraction of the thermal-noise RMS. We adopt $f = 0.1$ throughout, with the resulting dependence of the discarded RMS on retained rank shown in the upper panel of Figure 4.

Having fixed r through this criterion, the retained modes define the reduced foreground operator

$$\mathbf{T}_{\mathrm{FG},r} = \mathbf{N}^{1/2} \mathbf{U}_r \boldsymbol{\Lambda}_r^{1/2}, \tag{26}$$

whose latent coefficients have a standard-normal prior. The truncation criterion guarantees that the retained foreground covariance,

$\mathbf{T}_{\mathrm{FG},r}\mathbf{T}_{\mathrm{FG},r}^{\dagger} = \mathbf{C}_{\mathrm{FG}}^{(r)}$, captures the full prior-predictive foreground uncertainty to no greater than f times the thermal-noise level. As illustrated in Figure 3, the physically motivated parameterisation and the standard-normal latent representation therefore provide two routes to the same noise-relevant prior-predictive covariance, such that the reduced operator can reliably replace the original foreground model within the analytically marginalised likelihood. This sensitivity-based truncation can reduce N_{retained} below the algebraic upper bound established in Equation 21, providing the additional computational gains discussed in Section 3.2.4.

We highlight that the spatially uniform foreground prior and homogeneous noise adopted here are not requirements of the construction. Any non-uniform or correlated prior structure, inhomogeneous noise or visibility coverage modified by missing or flagged data can be incorporated directly into the decomposition. The resulting basis therefore can remain truly informed by both degeneracies in the physical mapping and the sensitivity of each measured visibility.

The application of KL decompositions to 21-cm foreground mitigation is itself not new, with previous work using assumed signal and foreground covariances to construct modes for foreground filtering (Shaw et al. 2014). Here, however, both the retained modes and their variances are derived directly from the full prior-predictive covariance generated by the physically motivated sky-to-visibility mapping. This allows foreground uncertainty rooted in its physical sky and instrumental origins to be propagated into posterior inference of the 21-cm signal and Bayesian-evidence evaluation through a statistically optimal latent representation.

3.2.3 Wide-Field Foreground Compression

Figure 4 shows the statistically optimal reduction in foreground dimensionality achieved across the fields of view considered here, conditioned on the physically informed, observation-specific sky-to-visibility mapping and the adopted foreground prior and noise covariance. For the largest $\theta_{\mathrm{FG}} = 180^{\circ}$ model, the representation collapses from $N_{\mathrm{FG,param}} = 300,767$ parameters to $N_{\mathrm{retained}} = 4,155$ modes, a decrease of $\sim 98.6\%$, with the corresponding savings in likelihood-evaluation time and memory discussed in Section 3.2.4. We highlight that this compression is achieved despite the broad, spatially uninformative foreground prior adopted here. A more informative construction could introduce pixel-dependent uncertainties informed by existing knowledge of the diffuse foreground morphology and bright-source locations, analogous to the informed sky priors used by BayesCal (Sims et al. 2022a). By suppressing variance along physically implausible directions, such a prior would be expected to move additional visibility-space modes below the noise-relative truncation threshold and reduce N_{retained} further, an extension left to future work.

A notable feature of Figure 4 is the sharp increase in retained rank from 898 to 4,155 modes as θ_{FG} is extended from 160° to 180°, far exceeding the gradual growth observed across the smaller fields of view. The horizon response is conventionally characterised through the power it introduces: baseline foreshortening enhances sensitivity to large-scale diffuse emission near the horizon, which the chromatic primary beam and frequency-dependent fringe phase then map to the edge of the foreground wedge (Thyagarajan et al. 2015; Lanman et al. 2020). The retained rank instead quantifies the number of independent visibility-space modes required to represent uncertain horizon emission at the adopted noise-relative accuracy. With the same prior uncertainty assigned to every pixel, its growth reflects the additional structure introduced by the sky-to-visibility mapping, conditioned on the adopted prior and noise level, rather than the brightness of a particular sky realisation. The sharp increase therefore indicates that horizon emission requires many additional such modes, which a model truncated below the horizon cannot span, as demonstrated in Section 4.1. Although the horizon response is well established through simulations, to our knowledge this is the first direct statistical quantification of this additional complexity, further motivating rigorous propagation of wide-field foreground uncertainty through cosmological inference.

3.2.4 Comparison with Gaussian Process Regression

To place the computational advantages of the latent representation in context, we compare its requirements with those of analytical marginalisation over the original basis and GPR, another widely used approach to foreground modelling in 21-cm cosmology. All three ultimately incorporate foreground uncertainty through a covariance contribution to the likelihood, although GPR specifies this covariance phenomenologically rather than deriving it from the physical sky-to-visibility mapping. Figure 3 illustrates how this mapping generates covariance across frequency, baseline and observing time. GPR implementations typically adopt stationary frequency kernels applied independently to each visibility spectrum (Mertens et al. 2018), which do not capture this full structure and can therefore lead to signal loss, biased power-spectrum recovery or spurious detections (Kern & Liu 2021; Liu et al. 2026).

Irrespective of how the covariance is constructed, likelihood evaluation in each case is dominated by the factorisation and solution of a dense Hermitian positive-definite system. Analytical marginalisation can be evaluated either in parameter space, through an $N_{\mathrm{FG,param}} \times N_{\mathrm{FG,param}}$ system, or in data space, through the $N_{\mathrm{d}} \times N_{\mathrm{d}}$ prior-predictive covariance. The computationally preferable formulation therefore has dimension $\min(N_{\mathrm{FG,param}}, N_{\mathrm{d}})$, while a conventional dense GPR likelihood is evaluated through its $N_{\mathrm{d}} \times N_{\mathrm{d}}$ data covariance. Since dense Cholesky factorisation scales as $O(N^3)$, the bound $N_{\mathrm{retained}} \leq \min(N_{\mathrm{FG,param}}, N_{\mathrm{d}})$ guarantees that the KL marginalised likelihood requires no larger a factorisation, with degeneracies in the physical mapping and the statistically controlled truncation generally providing significant computational advantages.

For the full 180° configuration considered here, $N_{\mathrm{FG,param}} = 300,767 > N_{\mathrm{d}} = 54,990$, such that analytical marginalisation over the original foreground basis is most efficiently evaluated in data space and requires the same system dimension as a dense GPR likelihood. The KL representation instead retains $N_{\mathrm{retained}} = 4,155$ modes. The corresponding `complex128` matrices alone require approximately 48.4 GB and 276 MB, respectively, while their cubic scaling implies a nominal reduction in factorisation time of approximately 2.3×10^3.

It is important to acknowledge that these gains require the KL basis to be constructed before inference. For the largest configuration considered here, we construct this basis through the covariance formulation, for which forming and eigendecomposing the dense $N_{\mathrm{d}} \times N_{\mathrm{d}}$ covariance takes only ~ 80 s on four H100 GPUs using the JAXMG `syevd` routine, the scaling of which is characterised in Tutt & Wiersema (2026). This modest one-time cost is rapidly amortised over the repeated likelihood evaluations required for posterior sampling and Bayesian-evidence evaluation.

3.3 21-cm Brightness Field Models

Having defined the image-domain foreground model in Section 3.2, we now describe the parameterisation adopted for the cosmological

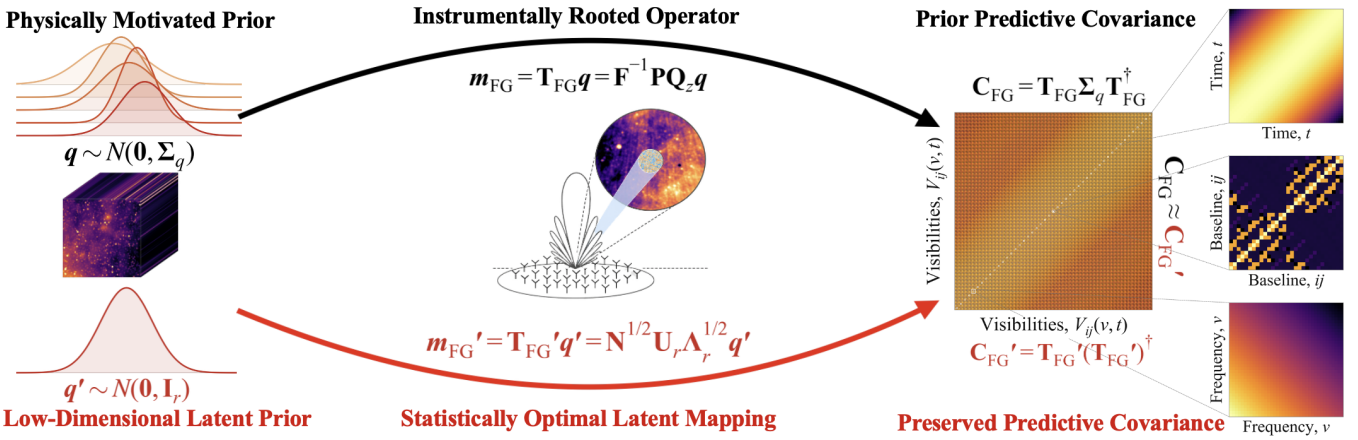


Figure 3. Comparison of the original foreground model and its reduced latent representation, showing how the same noise-relevant prior-predictive covariance is recovered through two different parameterisations. The original model (black) propagates physically motivated foreground priors through the interferometric response, while the reduced model (red) maps a standard-normal latent prior into the corresponding visibility-space uncertainty. Primes denote the reduced latent-space quantities, $\mathbf{T}_{\mathrm{FG},r}$ and $\boldsymbol{q}_r$, defined in Equation 26. The insets isolate the covariance across frequency, baseline and time, with the remaining coordinates held fixed in each case.

21-cm brightness contribution. Although several physically motivated bases are possible, BayesEoR represents this field on a rectilinear three-dimensional Fourier grid (Sims et al. 2019; Sims & Pober 2019; Burba et al. 2023). This representation exposes the line-of-sight spectral structure that distinguishes the 21-cm signal from smooth foreground emission, while providing a direct correspondence between the model coefficients and the cosmological power spectrum we seek to recover.

The complex Fourier amplitudes, $\boldsymbol{a}$, are defined in (u, v, η) coordinates, where u and v describe transverse angular Fourier modes and η is the Fourier dual of observing frequency, related to cosmological Fourier modes by

$$\begin{aligned} k_x &= \frac{2\pi u}{D_{\mathrm{M}}(z_{\mathrm{c}})}, \\ k_y &= \frac{2\pi v}{D_{\mathrm{M}}(z_{\mathrm{c}})}, \\ k_z &= \frac{2\pi H_0 \nu_{21} E(z_{\mathrm{c}})}{c(1+z_{\mathrm{c}})^2}\eta. \end{aligned} \tag{27}$$

Where this conversion is evaluated at the central redshift, z_{c}, with Δu and Δv set by the reciprocal angular dimensions of the 21-cm model field of view and $\Delta\eta = B^{-1}$ by the observing bandwidth, B.

The mapping from the Fourier coefficients to the model visibilities can similarly be understood as a sequence of linear operations: first, $\mathbf{F}_z$ transforms the line-of-sight modes from η to frequency, before $\mathbf{F}'$ maps the transverse (u, v) modes into the image domain. The resulting brightness field is then weighted by the primary-beam operator, $\mathbf{P}$, and propagated to the sampled visibilities by $\mathbf{F}^{-1}$ through the full three-dimensional interferometric response.

For a single coefficient, a_r, located at (u_r, v_r, η_r), its contribution to the visibility measured by baseline ij at frequency ν and time t is thus,

$$\begin{aligned} m_{21,ij}^{(r)}(\nu, t; a_r) = \sum_{p=1}^{N_{\mathrm{pix},21}} & \Delta\Omega \exp\left[-2\pi i \boldsymbol{u}_{ij}(\nu)\cdot\hat{\boldsymbol{s}}_p(t)\right] \\ & \times P(\hat{\boldsymbol{s}}_p(t), \nu) \\ & \times \Delta u\,\Delta v \exp\left[2\pi i(u_r l_p + v_r m_p)\right] \\ & \times \Delta\eta \exp\left[2\pi i \eta_r(\nu - \nu_{\mathrm{c}})\right] a_r. \end{aligned} \tag{28}$$

Table 1. Dimensions of the cosmological 21-cm models considered in this work as a function of model field of view. Each uses $N_\eta = 39$ line-of-sight modes.

θ_{21}	N_u	N_v	$N_{21,\mathrm{param}}$
5°	9	7	2, 356
10°	13	11	5, 396
15°	17	15	9, 652
20°	21	19	15, 124
25°	25	23	21, 812

Summing over the 21-cm coefficients and stacking the resulting visibilities across baselines, frequencies and times gives the linear model

$$\boldsymbol{m}_{21}(\boldsymbol{a}) = \mathbf{F}^{-1}\mathbf{P}\mathbf{F}'\mathbf{F}_z\boldsymbol{a} \equiv \mathbf{T}_{21}\boldsymbol{a}. \tag{29}$$

The total number of independent 21-cm parameters is therefore set by the dimensions of the transverse and line-of-sight Fourier grids. After excluding the transverse monopole and the $\eta = 0$ mode, this becomes

$$N_{21,\mathrm{param}} = (N_u N_v - 1)(N_\eta - 1). \tag{30}$$

Whereas bright foreground emission remains significant through the primary-beam sidelobes, sensitivity to the much fainter cosmological 21-cm signal is concentrated near the phase centre, allowing its model to cover a smaller field. We test this approximation using full angular diameters $\theta_{21} = 5°, 10°, \ldots, 25°$, defined over the drift scan in the same manner as the foreground fields in Section 2.2. The drift scan extends these regions in right ascension at fixed declination, giving $N_u > N_v$. The corresponding model dimensions are listed in Table 1.

Given that a statistical detection of the 21-cm signal concerns the power within each k-space bin rather than a particular brightness-field realisation, we seek to infer the distribution of the Fourier amplitudes rather than their individual values. The cosmological coefficients, $\boldsymbol{a}$, are therefore analytically marginalised rather than sampled directly, while the power-spectrum amplitudes defining their prior are retained as sampled hyperparameters. Assuming that the 21-cm brightness field is statistically homogeneous and isotropic, the prior

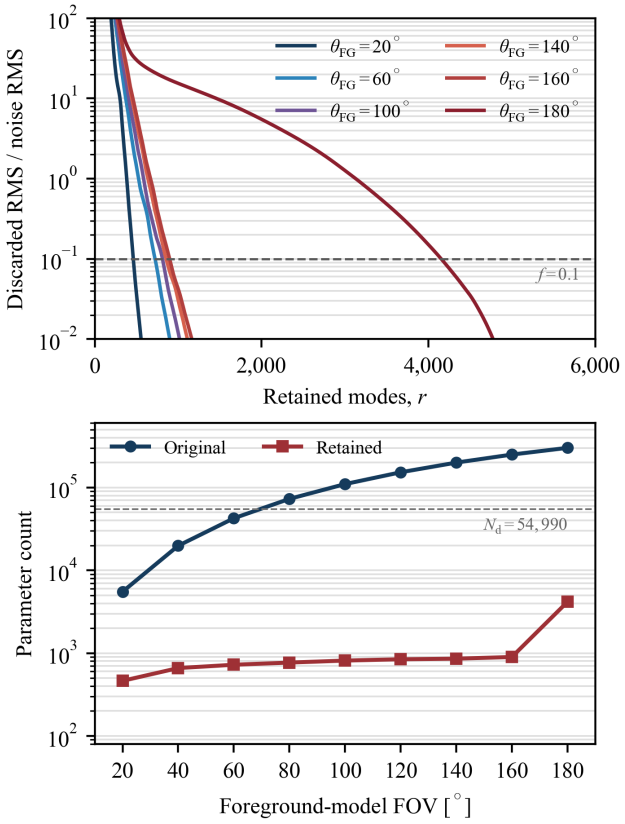


Figure 4. Statistical control of foreground compression through the noise-relative KL truncation. The upper panel shows the discarded foreground RMS relative to the thermal-noise RMS as a function of retained rank for six representative fields of view, with the dashed line marking the adopted threshold, $f = 0.1$. The lower panel shows the resulting dimensionality of the original physically rooted pixel basis (blue) and statistically optimal Karhunen–Loève representation (red) across the foreground-model fields of view. Its dashed line denotes the dimensionality of the measured visibility vector, $N_d = 54,990$.

variance of each coefficient depends only on

$$k_i = \sqrt{k_{x,i}^2 + k_{y,i}^2 + k_{z,i}^2}. \tag{31}$$

The modes are therefore grouped into spherical k-space bins, with each bin, s, assigned a power-spectrum amplitude, $\varphi_s \equiv \Delta^2_{21}(k_s)$, that sets the prior variance of its Fourier coefficients. Denoting the bin containing mode i by $s(i)$ and taking distinct Fourier modes to be uncorrelated, the prior covariance of $\boldsymbol{a}$ is

$$\left\langle a_i a_j^* \right\rangle = [\boldsymbol{\Phi}(\boldsymbol{\varphi})]_{ij} = \varphi_{s(i)}\delta_{ij}. \tag{32}$$

For sampling, we express the power in each bin through the logarithmic amplitude

$$\rho_s \equiv \log_{10}\left(\frac{\varphi_s}{\mathrm{mK}^2}\right), \tag{33}$$

to which we assign the independent uniform priors listed in Table 2.

For the configurations considered here, the 21-cm model does not represent a significant computational bottleneck, and we therefore retain its full physically motivated basis during inference. Future analyses incorporating longer baselines will require finer transverse Fourier sampling and substantially larger 21-cm parameter spaces,

Table 2. Sampled 21-cm power-spectrum parameters and their adopted prior ranges. Each k_s denotes the mean $|k|$ of the Fourier modes assigned to bin s.

Parameter	k_s [Mpc^{-1}]	Prior on ρ_s
φ_0	0.1	$\mathcal{U}(-0.9, 3.1)$
φ_1	0.2	$\mathcal{U}(0.0, 4.0)$
φ_2	0.3	$\mathcal{U}(0.5, 4.5)$
φ_3	0.4	$\mathcal{U}(0.9, 4.9)$
φ_4	0.5	$\mathcal{U}(1.3, 5.3)$
φ_5	0.7	$\mathcal{U}(1.7, 5.7)$
φ_6	1.0	$\mathcal{U}(2.1, 6.1)$
φ_7	1.4	$\mathcal{U}(2.5, 6.5)$
φ_8	1.8	$\mathcal{U}(2.8, 6.8)$

for which the KL decomposition can be applied independently within each k-space bin. This would compress the prior-predictive 21-cm covariance while retaining the power-spectrum amplitudes, $\boldsymbol{\varphi}$, as sampled hyperparameters, providing a statistically optimal latent representation of all sky components within the joint inference. The numerical implementation of this approach is left for future work.

3.4 Joint Inference

Having defined the foreground and 21-cm contributions independently, their combined propagation to the measured visibilities within the joint model, $\mathcal{M}_{21+\mathrm{FGs}}$, can simply be expressed as the sum of the two linear models,

$$\boldsymbol{m} = \mathbf{T}_{21}\boldsymbol{a} + \mathbf{T}_{\mathrm{FG}}\boldsymbol{q} = \mathbf{F}^{-1}\mathbf{P}\left(\mathbf{F}'\mathbf{F}_z\boldsymbol{a} + \mathbf{Q}_z\boldsymbol{q}\right). \tag{34}$$

For the purposes of inference, the foreground contribution can be represented using either the original operator and coefficients, $(\mathbf{T}_{\mathrm{FG}}, \boldsymbol{q})$, or their compressed KL counterparts, $(\mathbf{T}_{\mathrm{FG},r}, \boldsymbol{q}_r)$. Denoting either pair by $(\mathbf{T}'_{\mathrm{FG}}, \boldsymbol{q}')$, the joint forward model can then be written as a single linear operator over the concatenated coefficient vector, $\boldsymbol{b}'$:

$$\boldsymbol{m} = \mathbf{T}_{21}\boldsymbol{a} + \mathbf{T}'_{\mathrm{FG}}\boldsymbol{q}' \equiv \underbrace{\begin{bmatrix}\mathbf{T}_{21} & \mathbf{T}'_{\mathrm{FG}}\end{bmatrix}}_{\mathbf{T}'}\underbrace{\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{q}'\end{pmatrix}}_{\boldsymbol{b}'}. \tag{35}$$

The mathematical form of the analytically marginalised likelihood remains unchanged between these representations, which differ only in dimensionality and in the foreground block of the block-diagonal Gaussian prior covariance,

$$\boldsymbol{\Psi}'(\boldsymbol{\varphi}) = \begin{pmatrix}\boldsymbol{\Phi}(\boldsymbol{\varphi}) & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}'_{\mathrm{FG}}\end{pmatrix}, \tag{36}$$

For the original representation, $\boldsymbol{\Sigma}'_{\mathrm{FG}} = \boldsymbol{\Sigma}_q$, as defined in Section 3.2.1, whereas the retained coefficients in the compressed KL representation have a standard-normal prior, $\boldsymbol{\Sigma}'_{\mathrm{FG}} = \mathbf{I}_{N_{\mathrm{retained}}}$, by construction through Equation 26. The 21-cm prior covariance, $\boldsymbol{\Phi}(\boldsymbol{\varphi})$, is common to both and remains defined by the power-spectrum amplitudes, $\boldsymbol{\varphi}$.

The foundation of this approach is the marginalisation of the individual foreground and 21-cm coefficients, $\boldsymbol{b}'$, under their joint Gaussian prior with covariance $\boldsymbol{\Psi}'(\boldsymbol{\varphi})$. Integrating over these coefficients propagates their uncertainty and degeneracies into the inference without requiring their individual values to be sampled:

$$\mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\varphi}, \mathcal{M}_{21+\mathrm{FGs}}) = \int \mathcal{L}(\boldsymbol{d} \mid \boldsymbol{b}', \mathcal{M}_{21+\mathrm{FGs}}) \times \pi(\boldsymbol{b}' \mid \boldsymbol{\varphi}, \mathcal{M}_{21+\mathrm{FGs}})\mathrm{d}\boldsymbol{b}'. \tag{37}$$

Because the forward model depends linearly on $\boldsymbol{b}'$ and both the prior and visibility likelihood are Gaussian, this integral can be evaluated analytically (Luger et al. 2017; Greengard et al. 2022; Sims et al. 2019; Sims & Pober 2019; Burba et al. 2023) through:

$$\mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\varphi}, \mathcal{M}_{21+\mathrm{FGs}}) = \frac{\exp\left[-\boldsymbol{d}^\dagger \mathbf{N}^{-1}\boldsymbol{d} + \overline{\boldsymbol{d}}^\dagger \boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi})\overline{\boldsymbol{d}}\right]}{\pi^{N_\mathrm{d}} \det \mathbf{N} \det \boldsymbol{\Psi}'(\boldsymbol{\varphi}) \det \boldsymbol{\Sigma}(\boldsymbol{\varphi})}, \tag{38}$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\varphi}) = \mathbf{T}'^\dagger \mathbf{N}^{-1}\mathbf{T}' + [\boldsymbol{\Psi}'(\boldsymbol{\varphi})]^{-1}, \qquad \overline{\boldsymbol{d}} = \mathbf{T}'^\dagger \mathbf{N}^{-1}\boldsymbol{d}. \tag{39}$$

Although the foreground and 21-cm coefficients no longer enter the sampled parameter space, each marginal-likelihood evaluation is dominated by the dense linear system

$$\boldsymbol{\Sigma}(\boldsymbol{\varphi})\widehat{\boldsymbol{b}}'(\boldsymbol{\varphi}) = \overline{\boldsymbol{d}}, \tag{40}$$

The factorisation required to evaluate its determinant and the associated quadratic form in Equation 38 constitutes the principal computational cost. We use the MultiNest implementation (Feroz et al. 2009) of Nested Sampling (Skilling 2006) to explore the posterior over the logarithmic power-spectrum amplitudes, $\boldsymbol{\rho}$, adopting the priors listed in Table 2, while simultaneously evaluating the Bayesian evidence,

$$\mathcal{Z}_{21+\mathrm{FGs}} = \int \mathcal{L}(\boldsymbol{d} \mid \boldsymbol{\rho}, \mathcal{M}_{21+\mathrm{FGs}})\pi(\boldsymbol{\rho} \mid \mathcal{M}_{21+\mathrm{FGs}})\mathrm{d}\boldsymbol{\rho}. \tag{41}$$

Throughout this work, we also evaluate a foreground-only model, $\mathcal{M}_{\mathrm{FG}}$, both to investigate the foreground models developed here in isolation and to provide the null hypothesis against which the joint analyses are assessed within the Bayesian validation framework described in Section 3.5. Marginalising over its foreground coefficients leaves no remaining hyperparameters to be sampled, such that the evidence is given directly by the fully marginalised likelihood,

$$\mathcal{Z}_{\mathrm{FG}} = \int \mathcal{L}(\boldsymbol{d} \mid \boldsymbol{q}', \mathcal{M}_{\mathrm{FG}})\pi(\boldsymbol{q}' \mid \mathcal{M}_{\mathrm{FG}})\mathrm{d}\boldsymbol{q}'. \tag{42}$$

Applying the same mathematical construction used to obtain Equation 38 gives

$$\mathcal{Z}_{\mathrm{FG}} = \frac{\exp\left[-\boldsymbol{d}^\dagger \mathbf{N}^{-1}\boldsymbol{d} + \overline{\boldsymbol{d}}_{\mathrm{FG}}^\dagger \boldsymbol{\Sigma}_{\mathrm{FG}}^{-1}\overline{\boldsymbol{d}}_{\mathrm{FG}}\right]}{\pi^{N_\mathrm{d}} \det \mathbf{N} \det \boldsymbol{\Sigma}'_{\mathrm{FG}} \det \boldsymbol{\Sigma}_{\mathrm{FG}}}, \tag{43}$$

where

$$\boldsymbol{\Sigma}_{\mathrm{FG}} = \mathbf{T}'^\dagger_{\mathrm{FG}} \mathbf{N}^{-1}\mathbf{T}'_{\mathrm{FG}} + \left(\boldsymbol{\Sigma}'_{\mathrm{FG}}\right)^{-1}, \qquad \overline{\boldsymbol{d}}_{\mathrm{FG}} = \mathbf{T}'^\dagger_{\mathrm{FG}} \mathbf{N}^{-1}\boldsymbol{d}. \tag{44}$$

Although no posterior sampling is required for this model, the maximum-a-posteriori foreground coefficients and corresponding model visibilities can still be recovered as

$$\widehat{\boldsymbol{q}}' = \boldsymbol{\Sigma}_{\mathrm{FG}}^{-1}\overline{\boldsymbol{d}}_{\mathrm{FG}}, \qquad \widehat{\boldsymbol{m}}_{\mathrm{FG}} = \mathbf{T}'_{\mathrm{FG}}\widehat{\boldsymbol{q}}'. \tag{45}$$

3.5 Bayesian Validation

We finally briefly summarise the Bayesian validation framework adopted in this work to assess the wide-field foreground model developed in Section 3.2 alongside the 21-cm model described in Section 3.3. Such validation is essential for a trusted detection in 21-cm cosmology, where a faint cosmological signal must be distinguished from far brighter and structurally degenerate foreground and instrumental contributions. Model misspecification can otherwise leave residual structure that is absorbed by the cosmological component, producing spurious detections. For a more detailed discussion of the BaNTER framework and its application to 21-cm inference, we refer the reader to Sims et al. (2025a), Liu et al. (2026) and Tutt et al. (2026c).

The framework combines three complementary checks: statistical support for a signal, susceptibility to spurious detection and consistency of the fitted residuals with the assumed thermal-noise model. The first two use Bayesian model comparison, for which, assuming equal prior model probabilities, the relative support for $\mathcal{M}_1$ over $\mathcal{M}_2$ is quantified by the logarithmic Bayes factor

$$\ln B_{12}(\boldsymbol{d}) = \ln\left(\frac{\mathcal{Z}(\boldsymbol{d} \mid \mathcal{M}_1)}{\mathcal{Z}(\boldsymbol{d} \mid \mathcal{M}_2)}\right), \tag{46}$$

where positive values favour $\mathcal{M}_1$ and negative values favour $\mathcal{M}_2$ (Kass & Raftery 1995). We denote the simulated foreground-plus-21-cm and foreground-only datasets by $\boldsymbol{d}_{21+\mathrm{FGs}}$ and $\boldsymbol{d}_{\mathrm{FG}}$, respectively.

3.5.1 *Statistical Detection*

Firstly, the statistical preference for a 21-cm signal is evaluated on $\boldsymbol{d}_{21+\mathrm{FGs}}$ through the detection Bayes factor

$$\ln B_{\mathrm{det}} = \ln\left(\frac{\mathcal{Z}(\boldsymbol{d}_{21+\mathrm{FGs}} \mid \mathcal{M}_{21+\mathrm{FGs}})}{\mathcal{Z}(\boldsymbol{d}_{21+\mathrm{FGs}} \mid \mathcal{M}_{\mathrm{FG}})}\right), \tag{47}$$

such that $\ln B_{\mathrm{det}} > 0$ favours the joint foreground and 21-cm model over the foreground-only hypothesis.

3.5.2 *Null-test Validation*

Secondly, the susceptibility to spurious detection is assessed by repeating this comparison on the foreground-only validation data, $\boldsymbol{d}_{\mathrm{FG}}$, for which the validation Bayes factor is

$$\ln B_{\mathrm{val}} = \ln\left(\frac{\mathcal{Z}(\boldsymbol{d}_{\mathrm{FG}} \mid \mathcal{M}_{21+\mathrm{FGs}})}{\mathcal{Z}(\boldsymbol{d}_{\mathrm{FG}} \mid \mathcal{M}_{\mathrm{FG}})}\right). \tag{48}$$

Here, $\ln B_{\mathrm{val}} < 0$ is required to favour the foreground-only model; $\ln B_{\mathrm{val}} \geq 0$ instead indicates that the 21-cm component can absorb residual foreground structure and therefore identifies a susceptibility to spurious detection.

3.5.3 *Goodness of Fit*

Evidence ratios quantify the relative support between competing models but do not establish whether the preferred model provides an adequate description of the data. We therefore also test whether the total residual is consistent with the assumed thermal-noise model. Because the linear coefficients are analytically marginalised, we reconstruct their conditional maximum-a-posteriori values for each equally weighted posterior sample, $\boldsymbol{\varphi}_j$. Using the quantities defined in Equation 39, these are given by

$$\widehat{\boldsymbol{b}}'_j = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_j)\overline{\boldsymbol{d}}, \tag{49}$$

where $\boldsymbol{d}$ denotes the dataset associated with the fit. For each reconstructed coefficient vector, we calculate the whitened residual statistic

$$\chi^2_j = \left(\boldsymbol{d} - \mathbf{T}'\widehat{\boldsymbol{b}}'_j\right)^\dagger \mathbf{N}^{-1}\left(\boldsymbol{d} - \mathbf{T}'\widehat{\boldsymbol{b}}'_j\right). \tag{50}$$

This produces a distribution of residual statistics across the posterior, which we summarise through its median,

$$\widetilde{\chi}^2_{\mathrm{res}} = \mathrm{median}_j\, \chi^2_j. \tag{51}$$

To test whether the residuals are consistent with the adopted noise model, we compare this statistic with the corresponding noise-only

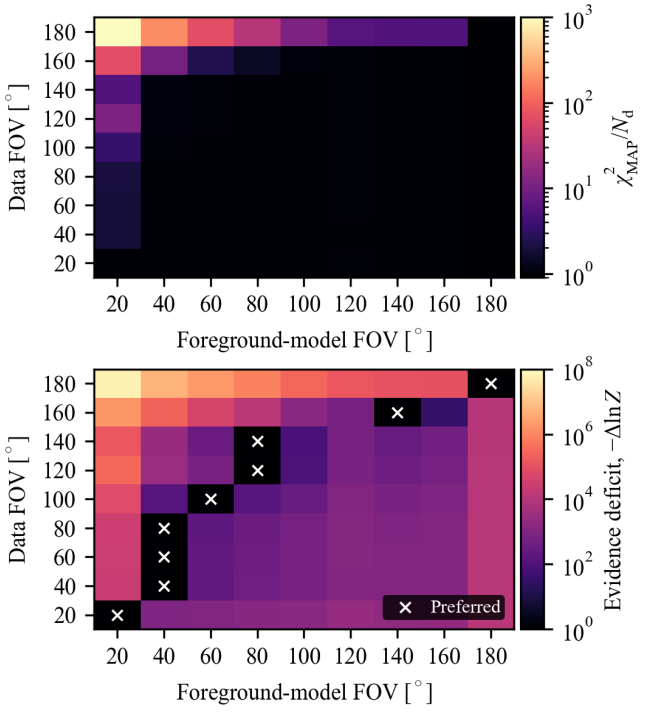


Figure 5. Cross-comparison of the foreground-only simulations and models as a function of their respective fields of view. The upper panel shows the noise-weighted residual statistic for the maximum-a-posteriori foreground reconstruction, $\chi^2_{\rm MAP}/N_{\rm d}$, while the lower panel shows the Bayesian-evidence ratio of each model relative to the best-supported model for each simulated dataset. Each row is therefore normalised independently, with the white cross marking its preferred model.

distribution, which for $N_{\rm d}$ independent complex Gaussian noise samples follows $\chi^2_{\rm noise} \sim \mathrm{Gamma}(N_{\rm d}, 1)$. We therefore estimate the upper-tail probability

$$q = P\left(\chi^2_{\rm noise} \geq \widetilde{\chi}^2_{\rm res}\right) \tag{52}$$

from 1,000 draws of this distribution and require $q \geq q_{\rm threshold} = 0.01$.

Overall, a validated detection must simultaneously satisfy all three criteria: the signal-injected data must favour the joint model, the foreground-only validation data must favour the foreground-only model and the fitted residuals must be consistent with the assumed thermal-noise model.

4 RESULTS AND DISCUSSION

We now evaluate the performance of the statistically optimal KL foreground basis across both foreground-only and joint cosmological analyses. Building on B23, we begin in Section 4.1 by cross-comparing the foreground simulations and models introduced in Sections 2 and 3.2, isolating the effects of sky extent and model misspecification. We then introduce the cosmological component in Section 4.2, presenting the recovered 21-cm power spectrum together with the corresponding Bayesian validation results. Having established a validated recovery, Section 4.3 examines its sensitivity to the adopted 21-cm model field. Section 4.4 then quantifies the computational gains enabled by the reduced basis. Finally, Section 4.5 sets out a path towards broader systematic studies and application to real observations.

4.1 Foreground-only Inference

We begin by examining how the angular extent of the foreground model affects inference in the absence of a cosmological component. For the foreground-only model, $\mathcal{M}_{\rm FG}$, marginalising the foreground coefficients leaves no remaining parameters to sample, allowing both the Bayesian evidence and maximum-a-posteriori reconstruction to be evaluated analytically, as described in Section 3.4. Figure 5 presents these metrics across all pairings of simulated and modelled foreground fields of view (FOV), showing the residual statistic in the upper panel and the log-evidence ratio between each model and the best-supported model for that dataset in the lower panel. The diagonal of each panel therefore represents matching simulated and modelled fields of view; above it, the foreground simulation extends beyond the model coverage, whereas below it the model encompasses a wider field than is present in the data.

Before discussing the results, we first establish the *a priori* expectation for these two diagnostics. Given that the models form a nested sequence, each field of view encompasses all smaller fields. The statistically controlled truncation should therefore allow the matched and larger models to yield residuals consistent with the thermal noise. In contrast, a smaller model should fail wherever emission beyond its angular extent introduces structure in the visibilities outside the space spanned by its basis. At the same time, the Bayesian evidence balances this goodness of fit against the expansion of the prior volume through the Occam penalty (MacKay 1992). We therefore expect the evidence to peak along the diagonal, with smaller models disfavoured through poorer fits and larger models through unnecessary complexity.

Figure 5 confirms that all models whose fields of view match or exceed the simulated foreground extent produce residuals consistent with the assumed noise, while the evidence disfavours expansions of the modelled field of view, and hence of the foreground basis, that do not produce a corresponding improvement in fit. More notably, several smaller-field models also provide noise-consistent reconstructions. This preference is largely confined to $\theta_{\rm FG} \sim 60°$–$140°$, where increasing the foreground extent introduces weakly beam-weighted emission without yet adding the distinct visibility structure associated with the horizon. Although the KL truncation removes visibility-space structure that is not statistically required, the conservative $f = 0.1$ threshold preserves the foreground covariance to below the thermal-noise level, leaving each retained basis marginally broader than required. This additional flexibility allows some smaller-field bases to span the weak visibility structure contributed by emission beyond their nominal angular extent and, because they achieve this with fewer retained modes, they are consequently preferred by the Bayesian evidence.

Crucially, the preference for smaller models does not persist once the foreground reaches the horizon, establishing the need to model the full visible sky. For the 180° dataset, the 160° model fails decisively, yielding $\chi^2_{\rm MAP}/N_{\rm d} = 5.23$, whereas the matched 180° model produces noise-consistent residuals and is overwhelmingly favoured by an evidence ratio of $\ln(\mathcal{Z}_{180}/\mathcal{Z}_{160}) = 1.06 \times 10^5$. This transition coincides with the sharp rise in the number of required modes identified in Figure 4, indicating that the preference is driven by the distinct visibility structure introduced by horizon emission.

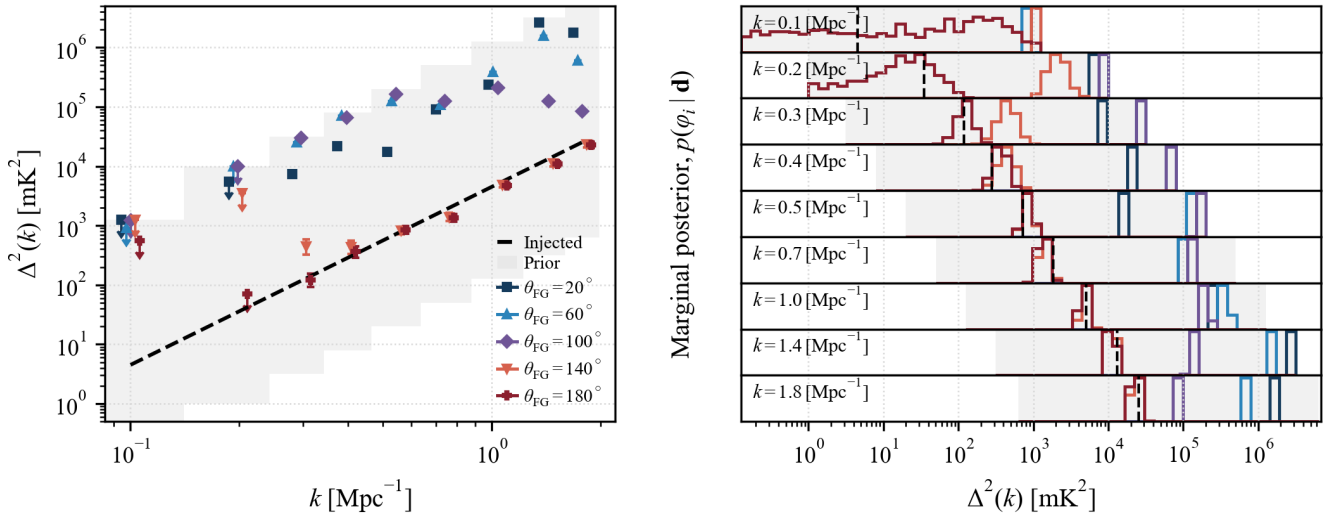


Figure 6. Recovery of the cosmological 21-cm power spectrum as a function of foreground-model field of view, θ_{FG}, with the cosmological-model extent fixed at $\theta_{21} = 25°$. Five of the nine foreground extents are shown for clarity. The left panel presents the recovered power-spectrum amplitudes, reporting central 68 per cent credible intervals for constrained bins and 95 per cent upper limits for the two lowest-k bins; the dashed line shows the injected spectrum. The right-hand panels show the corresponding marginal posterior distributions within each spherical k-space bin, with shaded regions denoting the adopted prior support.

θ_{FG} [°]	20	40	60	80	100	120	140	160	180
$\ln B_{val} < 0$	×	×	×	×	×	×	×	×	✓
$\ln B_{det} > 0$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$q \geq 0.01$	×	×	×	×	×	×	×	×	✓

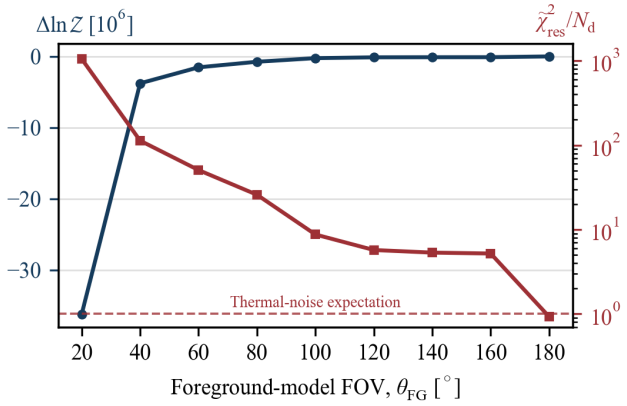


Figure 7. Bayesian validation diagnostics for the joint foreground and cosmological 21-cm inference as a function of foreground-model field of view, θ_{FG}, with $\theta_{21} = 25°$. The upper panel summarises the foreground-only validation, signal-detection and residual-consistency outcomes defined in Section 3.5. The lower panel shows the joint-model log evidence relative to the best-supported configuration (navy; left axis) and the median noise-weighted residual statistic (red; right axis), with the dashed red line marking the thermal-noise expectation.

4.2 Joint Inference

Having established the performance of the foreground model in isolation, we now test whether the parameterised cosmological and KL foreground representations can simultaneously span the structure present in full-fidelity wide-field simulations while remaining sufficiently informed for their contributions to be separated. We assess this through joint recovery of the cosmological 21-cm power spectrum, with the resulting posterior distributions and corresponding validation metrics shown in Figures 6 and 7, respectively. To isolate the effects of foreground misspecification, we analyse paired datasets sharing the same full-sky foreground realisation, with one containing foreground emission alone and the other additionally containing the cosmological signal described in Section 2.2.2. We analyse each dataset using the complete range of foreground-model fields while fixing the cosmological-model field at $\theta_{21} = 25°$. The effect of cosmological-model field of view is then studied independently in Section 4.3.

Given the results established in Sections 3.2.3 and 4.1, it is perhaps unsurprising that every model with $\theta_{FG} < 180°$ fails to recover the injected 21-cm power spectrum accurately. More importantly, however, the correspondence between the biased posteriors in Figure 6 and the validation metrics in Figure 7 provides a clear demonstration of the need for rigorous, evidence-based validation in 21-cm cosmology. In each misspecified model, unmodelled foreground structure is absorbed by the cosmological component, producing significantly biased power-spectrum constraints and a statistical preference for a cosmological signal despite not being present in the data, $\ln B_{val} \geq 0$. Together, these results demonstrate the susceptibility of 21-cm inference to spurious detection under foreground misspecification. Reassuringly, the full-field model recovers the injected spectrum and satisfies all three criteria, thereby providing end-to-end validation of the joint inference framework developed in this work.

For the validated full-field model, we obtain direct detections of the injected power across $k = 0.3$–$1.8\,\mathrm{Mpc}^{-1}$, while placing a stringent upper limit at $k = 0.2\,\mathrm{Mpc}^{-1}$ and a less constraining upper limit at $k = 0.1\,\mathrm{Mpc}^{-1}$. The reduced sensitivity in these two lowest-k bins exposes how the chromatic primary beam and fringe response couple angular and spectral foreground structure into the measured visibilities, creating complex degeneracies with the cosmological signal. Although more informative foreground priors could further tighten these constraints, recovering statistically well-founded upper limits

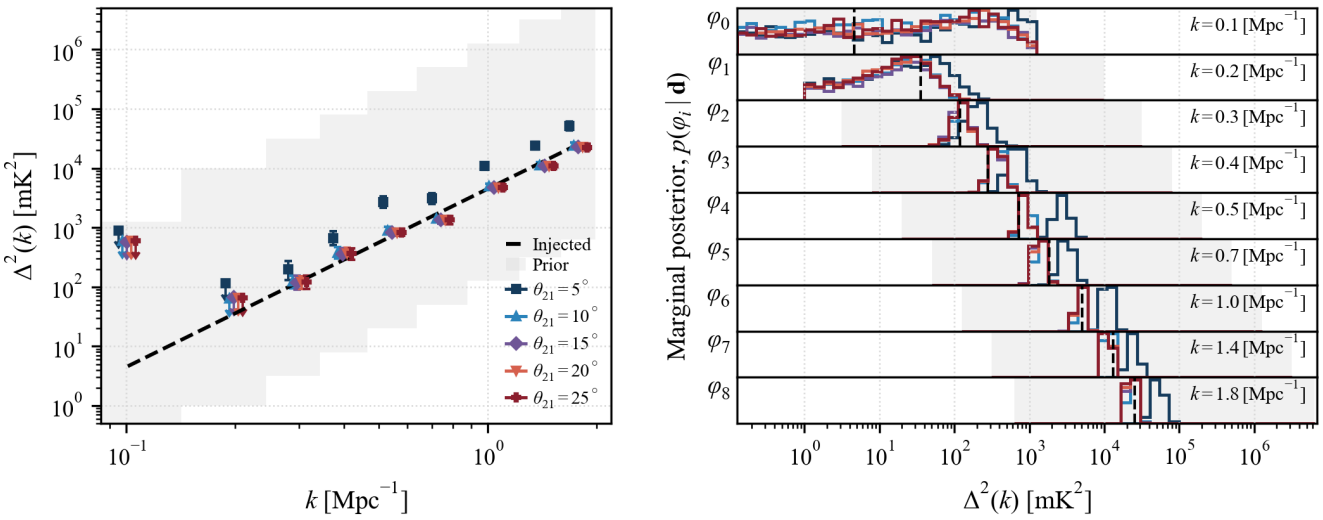


Figure 8. Recovery of the cosmological 21-cm power spectrum as a function of cosmological-model field of view, θ_{21}, with the foreground-model extent fixed at $\theta_{\rm FG} = 180^\circ$. The left panel presents the recovered power-spectrum amplitudes, reporting central 68 per cent credible intervals for constrained bins and 95 per cent upper limits for the two lowest-k bins; the dashed line shows the injected spectrum. The right-hand panels show the corresponding marginal posterior distributions within each spherical k-space bin, with shaded regions denoting the adopted prior support.

rather than spuriously attributing unresolved foreground structure to the 21-cm signal demonstrates the strength of the physically motivated joint modelling and analytical marginalisation developed here.

4.3 Effect of the 21-cm Model Field of View

Having validated the wide-field foreground model using a conservatively large cosmological field of view, we next investigate the extent to which the 21-cm model field of view can be reduced without biasing the recovered power spectrum. We repeat the joint inference with a fixed $\theta_{\rm FG} = 180^\circ$ foreground model while varying θ_{21} from 5° to 25°. Figure 8 shows that the recovered power spectra are statistically indistinguishable for $\theta_{21} = 10^\circ$–25° and accurately recover the injected signal, whereas the 5° model produces biased constraints. The Bayesian evidence in Table 3 reveals the same threshold, assigning substantially less support to the 5° model while remaining largely unchanged across the larger fields, therefore further supporting its evaluation through providing a direct diagnostic of misspecification when the resulting posterior might otherwise appear credible.

For the observational configuration considered here, a cosmological-model field of $\theta_{21} = 10^\circ$ is therefore sufficient, reducing the number of marginalised 21-cm coefficients to just 5,396. More generally, the minimum field required will depend on the amplitude of the cosmological signal, the instrumental response and the observational noise level. Nevertheless, at the central frequency, this field contains 92.3 per cent of the total beam-squared response over the visible sky, providing a useful indication of the angular coverage required for instruments with different primary beams. This angular restriction therefore provides one route to controlling cosmological-model dimensionality; for future longer-baseline analyses requiring finer transverse Fourier sampling, it could be combined with the within-k-bin KL compression noted in Section 3.3.

Table 3. Relative Bayesian evidence for the joint foreground and cosmological 21-cm inference as a function of cosmological-model field of view, θ_{21}, with the foreground-model extent fixed at $\theta_{\rm FG} = 180^\circ$. Values are relative to the best-supported configuration, $\theta_{21} = 25^\circ$, such that $\Delta \ln \mathcal{Z}_{21+\rm FGs} = 0$ denotes the preferred model.

θ_{21} [°]	$\Delta \ln \mathcal{Z}_{21+\rm FGs}$
5	−326.83
10	−2.06
15	−0.50
20	−0.43
25	0.00

4.4 End-to-End Computational Performance

Overall, the framework developed in this work provides a computationally scalable route to physically motivated Bayesian inference of the cosmological 21-cm power spectrum in the presence of wide-field foreground contamination. For the $\theta_{\rm FG} = 180^\circ$ and $\theta_{21} = 10^\circ$ configuration validated here, the KL representation collapses the original 300,767 foreground parameters to 4,155 retained modes. Combined with the 5,396 cosmological coefficients, the joint linear model totals just 9,551 parameters, all of which are analytically marginalised, leaving only the nine power-spectrum amplitudes in the sampled parameter space.

Within the marginalised likelihood of Equation 39, the principal computational cost is the factorisation and solution of the $N_{\rm marg} \times N_{\rm marg}$ system $\boldsymbol{\Sigma}(\boldsymbol{\varphi})$. For the configuration considered here, this corresponds to a $9,551 \times 9,551$ `complex128` matrix occupying 1.46 GB, with each likelihood evaluation taking $\sim$ 47 ms. Using 225 live points, MultiNest required 17,403 likelihood evaluations, completing posterior sampling and Bayesian-evidence evaluation in 13.6 min on a single H100 GPU.

Without the KL reduction, the same marginal likelihood can be evaluated through the $N_{\rm d} \times N_{\rm d}$ data-space covariance, a $54,990 \times 54,990$ `complex128` matrix occupying 48.4 GB. The unreduced im-

plementation has an estimated working-space memory requirement of ∼ 539 GB, requiring six 96-GB H100 GPUs. At an assumed ∼ 9 s per likelihood evaluation, completing the same inference would require ∼ 261 GPU-hours. Including the one-time KL construction described in Section 3.2.4, the reduced analysis requires only ∼ 0.32 GPU-hours in total, giving an estimated computational saving of ∼ 830 relative to this implementation. Under these assumptions, the upfront construction is amortised after only ∼ 6 likelihood evaluations.

4.5 Future Directions

Having established a scalable framework for the joint recovery and validation of the 21-cm power spectrum, we now consider how it can be extended beyond the analysis presented here. These directions encompass quantifying the impact of instrumental uncertainties and identifying those requiring explicit modelling (Section 4.5.1), exploring likelihood-free approaches to compressing and marginalising physically motivated nuisance models (Section 4.5.2), and adapting the framework for application to real interferometric observations (Section 4.5.3). Although motivated by 21-cm cosmology, the challenge of propagating nuisance uncertainty is shared across a broad range of science cases using interferometric visibility data.

4.5.1 Instrumental Systematics and Validation

While this work has focused on foreground-model misspecification under the assumption of a perfectly calibrated instrument and known primary beam, a natural extension is to apply the same validation framework to datasets containing controlled misspecification throughout the instrumental signal chain. Relevant effects include calibration errors, primary-beam perturbations, cable reflections and mutual coupling, all of which can introduce spectral structure that is degenerate with the cosmological signal (Sims et al. 2022a,b; O'Hara et al. 2024; Murphy et al. 2024; Rath et al. 2025; O'Hara et al. 2025).

Through a systematic exploration of how these effects propagate through to the cosmological recovery in end-to-end inference, the framework can establish the instrument- and observation-specific tolerances required for reliable signal recovery, which is a primary motivation of the ValSKA validation toolkit.[2] Where a contribution cannot reliably be known to the required accuracy, as is the case for the foreground structure motivating this work, the same analysis can identify which uncertainties must instead be explicitly parameterised within the physical forward model and marginalised during inference (Sims et al. 2022a; Wilensky et al. 2024, 2025; Kern 2025).

4.5.2 Learned Compression and Simulation-Based Inference

The KL construction developed here shows that the prior-predictive uncertainty of a high-dimensional, physically motivated nuisance model can be compressed into a statistically optimal visibility-space basis. The resulting reduction renders repeated evaluation of the marginalised likelihood computationally tractable.

Simulation-based inference (SBI) offers a complementary route to inference with high-dimensional nuisance models (Cranmer et al. 2020). Although the visibility likelihood can be written explicitly, the challenge demonstrated here lies not in defining it but in marginalising it over a large nuisance-parameter space. SBI instead learns the posterior distribution of the parameters of interest directly from end-to-end realisations drawn from the full prior-predictive model, marginalising implicitly over all nuisance contributions (Alsing & Wandelt 2019; Miller et al. 2021; Bhardwaj et al. 2023; Alvey et al. 2024; Saxena et al. 2024). For interferometric data, this requires an informative low-dimensional summary of the visibilities to be learned jointly with the inference network (Charnock et al. 2018; Makinen et al. 2021; Lehman et al. 2024; Tutt et al. 2026b), a task that may benefit from recent, more flexible SBI architectures (Gloeckler et al. 2024), with the simulation and training costs amortised across subsequent analyses. Developments in neural evidence estimation (Jeffrey & Wandelt 2024; Gessey-Jones & Handley 2024; Leeney et al. 2026) and misspecification tests for SBI (Anau Montel et al. 2024) would further allow such approaches to be incorporated within the Bayesian validation framework adopted here.

4.5.3 Application to Real Data

[TODO: Peter: Going to let you fill this in]

5 CONCLUSIONS

In this work, we have presented a scalable extension of BayesEoR that enables statistically rigorous recovery and validation of the cosmological 21-cm power spectrum in the presence of wide-field foreground contamination. The principal methodological developments are:

- **A wide-field foreground model:** We replace the tangent-plane uv parameterisation used by previous implementations with a spherical image-domain representation in physical brightness-temperature units, allowing foreground emission across the full visible sky to be propagated through the primary beam and three-dimensional fringe response and thereby accurately capturing wide-field effects at the horizon.
- **A statistically optimal latent representation:** A prior- and noise-informed Karhunen–Loève decomposition compresses the physically generated prior-predictive foreground covariance into its noise-relevant visibility-space modes. Its exact rank is bounded by $r_{\mathrm{FG}} \leq \min(N_{\mathrm{d}}, N_{\mathrm{FG,param}})$, with the noise-relative truncation reducing the retained dimensionality further while providing direct statistical control over the omitted foreground uncertainty.
- **Scalable joint inference and validation:** Incorporating the reduced representation within the analytically marginalised likelihood propagates the same physically motivated foreground uncertainty while allowing a far smaller system to be solved. This enables efficient posterior recovery of the 21-cm power spectrum alongside the Bayesian-evidence evaluation required for rigorous end-to-end validation.

We first show that the constructed basis' retained rank provides a direct statistical measure of the dimensionality of foreground uncertainty induced in visibility space. Expanding the model across intermediate fields of view, where the additional emission is strongly downweighted by the primary beam, produces only modest growth in the retained rank. Extending the model to the horizon ($160° - 180°$), however, increases the retained rank from 898 to 4,155 modes, demonstrating the substantial additional structure introduced by the horizon response. Although this effect is well established through simulations, to our knowledge this is the first direct statistical quantification of its noise-relevant visibility-space complexity.

The importance of accurately modelling this additional horizon

[2] https://github.com/uksrc/ValSKA

structure becomes clear in the joint inference, for which every foreground model restricted to $\theta_{\rm FG} < 180^\circ$ produces biased power-spectrum constraints and spuriously favours a 21-cm component when applied to foreground-only validation data. By contrast, the full-field model both recovers the injected spectrum and passes all three validation criteria. This agreement between the accuracy of the recovered power spectrum and the Bayesian validation outcome demonstrates how the evidence can distinguish a robust recovery from an apparently plausible but ultimately spurious detection.

Having established the required foreground extent, we then investigate how far the cosmological model can be reduced without compromising recovery. For the configuration considered here, a field of $\theta_{21} = 10^\circ$, containing 92.3 per cent of the total beam-squared response over the visible sky at the central frequency, is sufficient to provide detections across $0.3 \leq k \leq 1.8\,\mathrm{Mpc}^{-1}$ and informative upper limits on the two largest-scale bins. These upper limits reflect the genuine degeneracy between foreground and 21-cm structure on the largest scales after coupling through the instrumental response. They therefore caution against interpreting constraints on these modes without explicitly propagating the physical foreground–instrument coupling, since unresolved foreground structure may otherwise be attributed to the cosmological signal.

For this validated configuration, the KL construction reduces the full 300, 767-parameter foreground representation to 4, 155 retained modes. Together with the 5, 396 cosmological coefficients, the joint inference therefore marginalises over 9, 551 parameters rather than 306, 163. This allows the complete posterior and Bayesian evidence to be evaluated in 13.6 min on a single H100 GPU, corresponding to an estimated computational saving of ~ 830 relative to the unreduced GPU-resident implementation once construction of the KL basis is included.

Overall, this work establishes that the physical fidelity required for full-sky foreground modelling need not come at prohibitive computational cost. By combining statistically optimal compression with rigorous Bayesian validation, the framework provides a scalable foundation for robust cosmological inference from next-generation interferometric observations.

ACKNOWLEDGEMENTS

The authors thank Oscar S. D. O'Hara for insightful discussions during the preparation of this manuscript. The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI (McIntosh-Smith et al. 2024) is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023]. JT is supported by the Harding Distinguished Postgraduate Scholars Programme (HDPSP) and the Science and Technology Facilities Council (STFC) DTP Studentship. EdLA acknowledges the support of STFC via an Ernest Rutherford Fellowship.

DATA AVAILABILITY

The data that support the findings of this study are available from the first author upon reasonable request.

REFERENCES

Acharya A., et al., 2024, Monthly Notices of the Royal Astronomical Society, 527, 7835
Alsing J., Wandelt B., 2019, Monthly Notices of the Royal Astronomical Society, 488, 5093
Alvey J., Bhardwaj U., Domcke V., Pieroni M., Weniger C., 2024, Physical Review D, 109, 083008
Anau Montel N., Alvey J., Weniger C., 2024, arXiv e-prints
Anstey D., de Lera Acedo E., Handley W., 2021, Monthly Notices of the Royal Astronomical Society, 506, 2041
Barkana R., 2016, Phys. Rep., 645, 1
Barkana R., Loeb A., 2001, Physics Reports, 349, 125
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bevins H. T. J., 2026, The Astrophysical Journal Letters, 997, L35
Bevins H. T. J., Heimersheim S., Abril-Cabezas I., Fialkov A., de Lera Acedo E., Handley W., Singh S., Barkana R., 2024, MNRAS, 527, 813
Bhardwaj U., Alvey J., Miller B. K., Nissanke S., Weniger C., 2023, arXiv e-prints
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67–70
Bradbury J., et al., 2018, JAX: composable transformations of Python+NumPy programs, https://github.com/jax-ml/jax
Bull P., et al., 2025, RAS Techniques and Instruments, 4, rzaf046
Burba J., Sims P. H., Pober J. C., 2023, MNRAS, 520, 4443
Burba J., Bull P., Wilensky M. J., Kennedy F., Garsden H., Glasscock K. A., 2024, Monthly Notices of the Royal Astronomical Society, 535, 793
Bye C. H., et al., 2026, The Electromagnetically Isolated Global Signal Estimation Platform (EIGSEP) (arXiv:2602.02661), https://arxiv.org/abs/2602.02661
Carozzi T. D., Woan G., 2009, Monthly Notices of the Royal Astronomical Society, 395, 1558
Chapman E., Zaroubi S., Abdalla F. B., Dulwich F., Jelić V., Mort B., 2016, Monthly Notices of the Royal Astronomical Society, 458, 2928
Charnock T., Lavaux G., Wandelt B. D., 2018, Physical Review D, 97, 083004
Cohen A., Fialkov A., Barkana R., 2018, Monthly Notices of the Royal Astronomical Society, 478, 2193
Cranmer K., Brehmer J., Louppe G., 2020, Proceedings of the National Academy of Sciences, 117, 30055
Datta A., Bowman J. D., Carilli C. L., 2010, The Astrophysical Journal, 724, 526
DeBoer D. R., et al., 2017, PASP, 129, 045001
Dhandha J., et al., 2025, MNRAS, 542, 2292
Duane S., Kennedy A. D., Pendleton B. J., Roweth D., 1987, Physics Letters B, 195, 216
Dutta S., Bull P., Burba J., Wilensky M. J., Zhang Z., Nasirudin A., 2026, RAS Techniques and Instruments, 5, rzag074
Feroz F., Hobson M. P., Bridges M., 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601
Franzen T. M. O., Vernstrom T., Jackson C. A., Hurley-Walker N., Ekers R. D., Heald G., Seymour N., White S. V., 2019, Publications of the Astronomical Society of Australia, 36, e004
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Physics Reports, 433, 181
Gessey-Jones T., Handley W. J., 2024, Phys. Rev. D, 109, 123541
Glasscock K. A., Bull P., Burba J., Garsden H., Wilensky M. J., 2024, RAS Techniques and Instruments, 3, 607
Gloeckler M., Deistler M., Weilbach C. D., Wood F., Macke J. H., 2024, in Salakhutdinov R., Kolter Z., Heller K., Weller A., Oliver N., Scarlett J., Berkenkamp F., eds, Proceedings of Machine Learning Research Vol. 235, Proceedings of the 41st International Conference on Machine Learning. PMLR, pp 15735–15766 (arXiv:2404.09636), https://proceedings.mlr.press/v235/gloeckler24a.html
Greengard P., Gelman A., Vehtari A., 2022, Computational Statistics, 37, 701
Greig B., et al., 2021, MNRAS, 501, 1
HERA Collaboration et al., 2023, ApJ, 945, 124
HERA Collaboration et al., 2026, ApJ, 998, 33
Handley W. J., Hobson M. P., Lasenby A. N., 2015, Monthly Notices of the Royal Astronomical Society, 453, 4384
Hibbard J. J., Tauscher K., Rapetti D., Burns J. O., 2020, The Astrophysical Journal, 905, 113

Hogan C. J., Rees M. J., 1979, Monthly Notices of the Royal Astronomical Society, 188, 791
Hurley-Walker N., Callingham J. R., Hancock P. J., Franzen T. M. O., Hindson L., et al., 2017, MNRAS, 464, 1146
Jeffrey N., Wandelt B. D., 2024, Machine Learning: Science and Technology, 5, 015008
Karhunen K., 1947, Annales Academiae Scientiarum Fennicae. Series A I. Mathematica-Physica, pp 1–79
Kass R. E., Raftery A. E., 1995, J. Am. Statist. Assoc., 90, 773
Kennedy F., Bull P., Wilensky M. J., Burba J., Choudhuri S., 2023, The Astrophysical Journal Supplement Series, 266, 23
Kern N., 2025, Monthly Notices of the Royal Astronomical Society, 541, 687
Kern N. S., Liu A., 2021, Monthly Notices of the Royal Astronomical Society, 501, 1463
Kern N. S., Parsons A. R., Dillon J. S., Lanman A. E., Fagnoni N., de Lera Acedo E., 2019, The Astrophysical Journal, 884, 105
Kingma D. P., Welling M., 2019, Foundations and Trends in Machine Learning, 12, 307
Kittiwisit P., Murray S. G., Garsden H., et al., 2025, RAS Techniques and Instruments, 4, rzaf001
Koopmans L. V. E., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 001 (arXiv:1505.07568), doi:10.22323/1.215.0001
Lane W. M., Cotton W. D., van Velzen S., Clarke T. E., Kassim N. E., Helmboldt J. F., Lazio T. J. W., Cohen A. S., 2014, Monthly Notices of the Royal Astronomical Society, 440, 327
Lanman A. E., Hazelton B. J., Jacobs D. C., Kolopanis M. J., Pober J. C., Aguirre J. E., Thyagarajan N., 2019, Journal of Open Source Software, 4, 1234
Lanman A. E., Pober J. C., Kern N. S., de Lera Acedo E., DeBoer D. R., Fagnoni N., 2020, Monthly Notices of the Royal Astronomical Society, 494, 3712
Leeney S. A. K., Gessey-Jones T., Handley W. J., de Lera Acedo E., Bevins H. T. J., Tutt J. L., 2026, Conditional Neural Bayes Ratio Estimation for Experimental Design Optimisation (arXiv:2603.26489), https://arxiv.org/abs/2603.26489
Lehman K., Krippendorf S., Weller J., Dolag K., 2024, Learning Optimal and Interpretable Summary Statistics of Galaxy Catalogs with SBI (arXiv:2411.08957)
Liu A., Shaw J. R., 2020, Publications of the Astronomical Society of the Pacific, 132, 062001
Liu A., Tegmark M., Bowman J. D., Hewitt J. N., Zaldarriaga M., 2009, Monthly Notices of the Royal Astronomical Society, 398, 401
Liu A., Parsons A. R., Trott C. M., 2014a, Physical Review D, 90, 023018
Liu A., Parsons A. R., Trott C. M., 2014b, Physical Review D, 90, 023019
Liu Y., de Lera Acedo E., Sims P. H., 2026, Monthly Notices of the Royal Astronomical Society, 549, stag878
Loève M., 1945, Comptes Rendus de l'Académie des Sciences, 220, 380
Lovick T., Yallup D., Handley W., 2026, Automatic Laplace Collapsed Sampling: Scalable Marginalisation of Latent Parameters via Automatic Differentiation (arXiv:2603.26644), https://arxiv.org/abs/2603.26644
Luger R., Foreman-Mackey D., Hogg D. W., 2017, Research Notes of the American Astronomical Society, 1, 7
MacKay D. J. C., 1992, Neural Computation, 4, 415
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Makinen T. L., Charnock T., Alsing J., Wandelt B. D., 2021, Journal of Cosmology and Astroparticle Physics, 2021, 049
McIntosh-Smith S., Alam S. R., Woods C. J., 2024, in CUG 2024 Proceedings. pp 44–54 (arXiv:2410.11199), https://cug.org/proceedings/cug2024_proceedings/includes/files/pap130s2-file1.pdf
McKay L., Subrahmanyan R., Chippendale A. P., Bolli P., Kyriakou G., Dunning A., Ekers R., 2026, Nature Astronomy, 10, 1398
McKinley B., Yang R., López-Caniego M., Briggs F., Hurley-Walker N., et al., 2015, MNRAS, 446, 3478
McQuinn M., 2016, Annual Review of Astronomy and Astrophysics, 54, 313
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, Monthly Notices of the Royal Astronomical Society, 478, 3640
Mertens F. G., et al., 2025, A&A, 698, A186
Miller B. K., Cole A., Forré P., Louppe G., Weniger C., 2021, in Ranzato M., Beygelzimer A., Dauphin Y., Liang P. S., Wortman Vaughan J., eds, Advances in Neural Information Processing Systems Vol. 34, Advances in Neural Information Processing Systems. Curran Associates, Inc., pp 129–143 (arXiv:2107.01214), https://proceedings.neurips.cc/paper_files/paper/2021/file/01632f7b7a127233fa1188bd6c2e42e1-Paper.pdf
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, The Astrophysical Journal, 752, 137
Mozdzen T. J., Bowman J. D., Monsalve R. A., Rogers A. E. E., 2017, Monthly Notices of the Royal Astronomical Society, 464, 4995
Munshi S., et al., 2025, Monthly Notices of the Royal Astronomical Society, 542, 2785
Murphy G. G., Bull P., Santos M. G., et al., 2024, Monthly Notices of the Royal Astronomical Society, 534, 2653
Naidu R. P., Oesch P. A., Brammer G., et al., 2026, The Open Journal of Astrophysics, 9
Neal R. M., 2011, in Brooks S., Gelman A., Jones G. L., Meng X.-L., eds, , Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC, Boca Raton, FL, pp 113–162, doi:10.1201/b10905-6, https://doi.org/10.1201/b10905-6
Neben A. R., Bradley R. F., Hewitt J. N., DeBoer D. R., Parsons A. R., Aguirre J. E., Ali Z. S., et al., 2016, The Astrophysical Journal, 826, 199
O'Hara O. S. D., Dulwich F., de Lera Acedo E., Dhandha J., Gessey-Jones T., Anstey D., Fialkov A., 2024, Monthly Notices of the Royal Astronomical Society, 533, 2876
O'Hara O. S. D., et al., 2025, Monthly Notices of the Royal Astronomical Society, 538, 31
Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, The Astrophysical Journal, 756, 165
Paszke A., et al., 2019, in Advances in Neural Information Processing Systems. Curran Associates, Inc., pp 8024–8035, https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf
Planck Collaboration et al., 2020, A&A, 641, A6
Pober J. C., et al., 2014, The Astrophysical Journal, 782, 66
Pochinda S., et al., 2024, MNRAS, 531, 1113
Price D. C., 2016, PyGDSM: Python interface to Global Diffuse Sky Models, Astrophysics Source Code Library, record ascl:1603.013
Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901
Rath E., Pascua R., Josaitis A. T., et al., 2025, Monthly Notices of the Royal Astronomical Society, 541, 1125
Robins D., Anstey D., Bevins H., de Lera Acedo E., Irfan M. O., 2026, Monthly Notices of the Royal Astronomical Society, 552, stag1804
Rybicki G. B., Lightman A. P., 1979, Radiative Processes in Astrophysics. Wiley-Interscience, New York
Santos M. G., Cooray A., Knox L., 2005, The Astrophysical Journal, 625, 575
Saxena A., Meerburg P. D., de Lera Acedo E., Handley W., Koopmans L. V. E., 2023, Monthly Notices of the Royal Astronomical Society, 522, 1022–1032
Saxena A., Meerburg P. D., Weniger C., de Lera Acedo E., Handley W., 2024, RAS Techniques and Instruments, 3, 724
Scott D., Rees M. J., 1990, Monthly Notices of the Royal Astronomical Society, 247, 510
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Shaw J. R., Sigurdson K., Pen U.-L., Stebbins A., Sitwell M., 2014, The Astrophysical Journal, 781, 57
Sims P. H., Pober J. C., 2019, MNRAS, 488, 2904
Sims P. H., Pober J. C., 2020, Monthly Notices of the Royal Astronomical Society, 492, 22
Sims P. H., Lentati L., Alexander P., Carilli C. L., 2016, Monthly Notices of the Royal Astronomical Society, 462, 3069–3093
Sims P. H., Lentati L., Pober J. C., Carilli C., Hobson M. P., Alexander P., Sutter P. M., 2019, MNRAS, 484, 4152
Sims P. H., Pober J. C., Sievers J. L., 2022a, Monthly Notices of the Royal Astronomical Society, 517, 910
Sims P. H., Pober J. C., Sievers J. L., 2022b, Monthly Notices of the Royal Astronomical Society, 517, 935–961
Sims P. H., Burba J., Pober J. C., 2024, Journal of Open Source Software, 9, 6667

Sims P. H., et al., 2025a, Monthly Notices of the Royal Astronomical Society, 541, 2262–2281
Sims P. H., et al., 2025b, MNRAS, 544, 3856
Singh S., et al., 2022, Nature Astronomy, 6, 607
Skilling J., 2006, Bayesian Analysis, 1, 833
Spantini A., Solonen A., Cui T., Martin J., Tenorio L., Marzouk Y. M., 2015, SIAM Journal on Scientific Computing, 37, A2451
Stark D. P., Topping M. W., Endsley R., Tang M., 2026, Encyclopedia of Astrophysics, pp 453–499
Tauscher K., Rapetti D., Burns J. O., Switzer E., 2018, The Astrophysical Journal, 853, 187
Tegmark M., Taylor A. N., Heavens A. F., 1997, The Astrophysical Journal, 480, 22
Thyagarajan N., Jacobs D. C., Bowman J. D., Barry N., Beardsley A. P., Bernardi G., Briggs F., et al., 2015, The Astrophysical Journal, 804, 14
Tingay S. J., et al., 2013, Publ. Astron. Soc. Australia, 30, e007
Trott C. M., et al., 2025, ApJ, 991, 211
Tutt J., Wiersema R., 2026, JAXMg: A multi-GPU linear solver in JAX (arXiv:2601.14466), doi:10.48550/arXiv.2601.14466, https://arxiv.org/abs/2601.14466
Tutt J. L., Bevins H. T. J., Alvey J., Cumner J., Anstey D. J., Fialkov A., de Lera Acedo E., 2026b, A Cross-Domain Simulator and Benchmark for Uncertainty Quantification in Global 21-cm Inference, in prep.
Tutt J. L., Anstey D. J., Cumner J., Bevins H. T. J., Cavillot J., de Lera Acedo E., 2026a, Towards end-to-end Bayesian forward models in global 21-cm cosmology: surrogate modelling and marginalisation of beam uncertainty (arXiv:2608.18962), doi:10.48550/arXiv.2608.18962, https://arxiv.org/abs/2608.18962
Tutt J. L., Sims P. H., Pattison J. H. N., Anstey D. J., Leeney S. A. K., de Lera Acedo E., 2026c, Monthly Notices of the Royal Astronomical Society, 550, stag1101
Wilensky M. J., et al., 2024, RAS Techniques and Instruments, 3, 400
Wilensky M. J., Bull P., Fagnoni N., 2025, RAS Techniques and Instruments, 4, rzaf042
Zarka P., Tagger M., Denis L., Girard J. N., Konovalenko A., et al., 2015, in 2015 International Conference on Antenna Theory and Techniques (ICATT). IEEE, Kharkiv, Ukraine, pp 1–6 (arXiv:hal-01196457), doi:10.1109/ICATT.2015.7136773, https://hal.science/hal-01196457
Zernike F., 1938, Physica, 5, 785
Zheng H., et al., 2017, MNRAS, 464, 3486
Zhou Q., Liu W., Li J., Marzouk Y. M., 2018, Inverse Problems, 34, 095001
de Lera Acedo E., et al., 2022, Nature Astronomy, 6, 984–998
de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, MNRAS, 388, 247
van Cittert P. H., 1934, Physica, 1, 201
van Haarlem M. P., et al., 2013, A&A, 556, A2