Foreground modelling via Gaussian process regression: an application to HERA data

Abhik Ghosh [1,2,3]★ Florent Mertens [4,5]★ Gianni Bernardi, [2,6,7]★ Mário G. Santos, [1,2]
Nicholas S. Kern, [8] Christopher L. Carilli, [9,10] Trienko L. Grobler, [7,11]
Léon V. E. Koopmans, [4] Daniel C. Jacobs, [12] Adrian Liu, [8,13] Aaron R. Parsons, [8]
Miguel F. Morales, [14] James E. Aguirre, [15] Joshua S. Dillon, [8]† Bryna J. Hazelton, [14,16]
Oleg M. Smirnov, [2,7] Bharat K. Gehlot, [4,12] Siyanda Matika, [7] Paul Alexander, [10]
Zaki S. Ali, [8] Adam P. Beardsley, [12]† Roshan K. Benefo, [17] Tashalee S. Billings, [15]
Judd D. Bowman, [12] Richard F. Bradley, [18] Carina Cheng, [8] Paul M. Chichura, [15]
David R. DeBoer, [8] Eloy de Lera Acedo, [10] Aaron Ewall-Wice, [19] Gcobisa Fadana, [2]
Nicolas Fagnoni, [10] Austin F. Fortino, [15] Randall Fritz, [2] Steve R. Furlanetto, [20]
Samavarti Gallardo, [15,21] Brian Glendenning, [9] Deepthi Gorthi, [8] Bradley Greig, [22,23]
Jasper Grobbelaar, [2] Jack Hickish, [8] Alec Josaitis, [10] Austin Julius, [2] Amy S. Igarashi, [17,24]
MacCalvin Kariseb, [2] Saul A. Kohn, [15] Matthew Kolopanis, [12] Telalo Lekalake, [2]
Anita Loots, [2] David MacMahon, [8] Lourence Malan, [2] Cresshim Malgas, [2]
Matthys Maree, [2] Zachary E. Martinot, [15] Nathan Mathison, [2] Eunice Matsetela, [2]
Andrei Mesinger, [25] Abraham R. Neben, [19] Bojan Nikolic, [10] Chuneeta D. Nunhokee, [7,15]
Nipanjana Patra, [8] Samantha Pieterse, [2] Nima Razavi-Ghods, [10] Jon Ringuette, [15]
James Robnett, [9] Kathryn Rosie, [2] Raddwine Sell, [2] Craig Smith, [2] Angelo Syce, [2]
Max Tegmark, [19] Nithyanandan Thyagarajan, [9,12]‡ Peter K. G. Williams [26,27]
and Haoxuan Zheng [19]

Affiliations are listed at the end of the paper

ABSTRACT

The key challenge in the observation of the redshifted 21-cm signal from cosmic reionization is its separation from the much brighter foreground emission. Such separation relies on the different spectral properties of the two components, although, in real life, the foreground intrinsic spectrum is often corrupted by the instrumental response, inducing systematic effects that can further jeopardize the measurement of the 21-cm signal. In this paper, we use Gaussian Process Regression to model both foreground emission and instrumental systematics in ~2 h of data from the Hydrogen Epoch of Reionization Array. We find that a simple co-variance model with three components matches the data well, giving a residual power spectrum with white noise properties. These consist of an 'intrinsic' and instrumentally corrupted component with a coherence scale of 20 and 2.4 MHz, respectively (dominating the line-of-sight power spectrum over scales $k_{\parallel} \leq 0.2\,h\,\mathrm{cMpc}^{-1}$) and a baseline-dependent periodic signal with a

★ E-mail: abhik.physicist@gmail.com (AG); mertens@astro.rug.nl (FM); giannibernardi75@gmail.com (GB)

period of ~ 1 MHz (dominating over $k_\parallel \sim 0.4$–$0.8\,h\,\mathrm{cMpc}^{-1}$), which should be distinguishable from the 21-cm Epoch of Reionization signal whose typical coherence scale is ~ 0.8 MHz.

1 INTRODUCTION

Observations of the redshifted 21-cm signal from neutral Hydrogen hold the promise of revealing the detailed astrophysical processes occurring during the Epoch of Reionization (EoR) and the Cosmic Dawn (CD). The 21-cm signal can provide insights into the formation and evolution of the first structures in the Universe (see e.g. Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010; Pritchard & Loeb 2012; Mellema et al. 2013, for reviews): for example, when the intergalactic medium (IGM) is still largely neutral, it is a sensitive probe of the first sources of Ly α and X-ray radiation (Mesinger & Furlanetto 2007; Santos et al. 2010, 2011; McQuinn & O'Leary 2012; Fialkov, Barkana & Visbal 2014; Fialkov et al. 2017) and, during the subsequent EoR, its large-scale fluctuations map the evolution of the global ionization fraction (Lidz et al. 2008; Bolton et al. 2011). The 21-cm emission gives insights into the nature of formation of the first stars, galaxies, and their impact on the physics of the IGM (Loeb & Furlanetto 2013; Zaroubi 2013).

At present, several experiments are attempting to detect the power spectrum of the 21-cm signal from the EoR (e.g. GMRT,[1] LOFAR,[2] MWA,[3] PAPER[4]) or the sky-averaged 21-cm emission using a single dipole (Bowman & Rogers 2010; Greenhill & Bernardi 2012; Patra et al. 2015; Bernardi et al. 2016). Some of these ongoing efforts have achieved increasingly better upper limits on the 21-cm signal power spectra (Ali et al. 2015; Beardsley et al. 2016; Patil et al. 2017; Barry et al. 2019; Kolopanis et al. 2019; Li, Pober & Barry 2019), showing the way for the second-generation experiments such as the Square Kilometre Array (SKA[5]) and the Hydrogen Epoch of Reionization Array (HERA[6]). Recently, a detection of an absorption profile in the sky-averaged 21-cm signal centred at 78 MHz has been reported (Bowman et al. 2018), although the unexpected depth of the trough is calling for independent confirmations (Fraser et al. 2018) – including interferometric observations (Gehlot et al. 2019).

The main challenge in detecting the faint 21-cm signal is the presence of Galactic and extra-galactic foregrounds that are around 3–4 orders of magnitude stronger (e.g. Bernardi et al. 2009, 2010; Ghosh et al. 2011; Dillon et al. 2014; Parsons et al. 2014). Foregrounds as well as the instrumental response have a highly correlated continuum spectrum and can, in principle, be separate from the 21-cm signal that has structure on smaller frequency scales due to the intrinsic redshift evolution of the IGM (e.g. Bharadwaj & Sethi 2001; Zaldarriaga, Furlanetto & Hernquist 2004; Santos, Cooray & Knox 2005). However, the inherent smoothness of the foreground emission is often compounded by the interferometric response (mode-mixing), including frequency-dependent primary beams, side-lobe from bright, mis-subtracted sources, and ionospheric distortions (Bowman, Morales & Hewitt 2009; Koopmans 2010; Ghosh et al. 2011; Vedantham, Udaya Shankar & Subrahmanyan

2012; Yatawatta et al. 2013; Barry et al. 2016; Vedantham & Koopmans 2016; Patil et al. 2017; Byrne et al. 2019; Gehlot et al. 2019). Polarization leakage due to improper calibration may also add additional spectral structures to the unpolarized cosmological 21-cm window (Geil, Gaensler & Wyithe 2011; Asad et al. 2015; Nunhokee et al. 2017).

The study of foreground properties and their separation from the 21-cm signal have been a very active research area over the years (e.g. Datta, Bowman & Carilli 2010; Liu & Tegmark 2011; Morales et al. 2012; Trott, Wayth & Tingay 2012; Dillon et al. 2014). One strategy is to attempt to 'avoid' foregrounds, i.e. to avoid k modes that are contaminated by foregrounds and to estimate the 21-cm power spectrum using the uncontaminated modes. This assumes that foregrounds are well localized in k-space and the mode-mixing effects can be kept well under control (Thyagarajan et al. 2015). This foreground avoidance method has the disadvantage of considerably reducing the sensitivity of the instrument because of reduction in the number of k-modes that can be probed to characterize the EoR signal (Pober et al. 2014). The second approach involves subtracting the best possible foreground model and, possibly, recover access to the foreground-dominated power spectrum region. One of the possible disadvantages here is the risk of contamination of the cosmological 21-cm signal from the cleaning process. Foreground wedge also corrupts nearly all the redshift space 21-cm signal, making it difficult to extract cosmological information without foreground subtraction (Pober 2015; Jensen et al. 2016). There are also recent efforts to develop a somewhat hybrid analysis where a Galactic and Extragalactic All-sky MWA (GLEAM; Hurley-Walker et al. 2017) catalogue of sources including Pictor A and Fornax A were first subtracted from Precision Array for Probing the Epoch of Reionization data and then the power spectrum was estimated. This is equivalent to an additional visibility-based filtering within the foreground avoidance paradigm (Kerrigan et al. 2018).

Chapman et al. (2016) pointed out that blind foreground removal methods such as Generalized Morphological Component Analysis (GMCA; Chapman et al. 2013) can still model relatively non-smooth foregrounds effectively on short baselines ($k_\perp \lesssim 0.2\,h\,\mathrm{cMpc}^{-1}$), while avoidance suffers some degradation as the frequency-dependent small-scale structure cannot be confined purely in a region at small $k_\parallel$ modes. Several techniques have been proposed to model and remove foreground emission taking advantage of its spectral smoothness, including parametric (Jelić et al. 2008; Bonaldi & Brown 2015) and non-parametric methods (Harker et al. 2009; Chapman et al. 2013; Mertens, Ghosh & Koopmans 2018; Mertens et al. 2020). Both methods have the limitation that they may suppress the 21-cm signal and do not always reach a level of modelling error better than the noise for $k \leq 0.3\,h\,\mathrm{cMpc}^{-1}$, compared to the desired level of the 21-cm signal power spectrum (Mertens et al. 2018). In general, foreground subtraction allows to use virtually all k modes at the risk of contamination of the 21-cm signal, whereas foreground avoidance does not corrupt the cosmological signal within the EoR window, but cannot take advantage of any mode in the foreground wedge.

Recently, a novel, non-parametric (in the signal) method based on Gaussian Process Regression (GPR) has been studied in detail with simulations, where intrinsic smooth foregrounds, mid-scale

---

[1] http://www.gmrt.ncra.tifr.res.in/
[2] Low-Frequency Array, http://www.lofar.org.
[3] Murchison Widefield Array, http://www.mwatelescope.org.
[4] Precision Array to Probe EoR, http://eor.berkeley.edu.
[5] http://www.skatelescope.org
[6] http://reionization.org

frequency fluctuations associated with mode-mixing, Gaussian random noise, and a basic 21-cm signal model, are modelled with Gaussian Process (GP), and subsequently a separation with a precise estimation of the uncertainty was carried out (Mertens et al. 2018; Gehlot et al. 2019). The advantage of this method over previous ones is its implementation in a Bayesian framework that allows to incorporate different physical processes in the form of covariance structure priors (currently spectral and possible spatial implementation in future) on the various components. GPR further allows much better control over the coherence structure (and hence power spectra) of all components rather then be 'blind' for their physical origins, as are GMCA, Independent Component Analysis, or fitting polynomials. Further, it also offers a good way to extract foreground models from the data.

In this paper, we apply GPR to model foregrounds in a ∼100 min long observation with HERA-47. The GPR method was originally developed to be applied to observations with good *uv* coverage, but here we adapted it to work directly to visibilities, without being affected by the sparse HERA *uv* coverage. Foreground modelling helps us to assess the level of contamination of the data and the covariance models that can properly describe foregrounds. It can ultimately guide the foreground cleaning process and help finding the scales that should be safe to use in a foreground avoidance approach. We use the line-of-sight and the delay power spectrum in the $(k_\perp, k_\parallel)$ plane as our metric to characterize the foreground models.

The paper is organized as follows: Section 2 summarizes our observations along with the delay power spectrum estimation procedure; Section 3 describes our technique to calculate the foreground power spectrum using the GPR formalism. Finally, we conclude in Section 4. Cosmological parameters used here are from Planck Collaboration XIII (2016).

2 OBSERVATIONS AND DATA REDUCTION

The HERA (DeBoer et al. 2017) is an ongoing experiment to use the red-shifted 21-cm radiation originating from the cosmological distribution of neutral hydrogen (H I) to study the formation of first stars and black holes from CD ($z \sim 30$) to the full IGM ionization history ($6 \lesssim z \lesssim 12$). In its final configuration, the array will consist of 350 parabolic dishes of ∼14 m diameter, with an effective area of ∼154 m² per antenna, closely packed in a hexagonal split-core (Dillon & Parsons 2016), plus outriggers up to ∼1 km distance. The experiment is optimized for robust power spectrum detection while minimizing foreground contamination (Pober et al. 2014; Thyagarajan et al. 2015; Ewall-Wice et al. 2017)

In this paper, we used data from the deployment of the first 47 dish (HERA–47) array. It covers a frequency range of 100–200 MHz with a channel resolution of ∼97.6 kHz. The results presented in this paper were generated from 10 nights of HERA–47 data, starting on 2017 October 5, using only the 'xx' polarization cross-products. In the paper, we refer to this as stokes 'I'. We selected snapshots of 10 min (see Fig. 1 for the corresponding *uv* coverage) of data close to 21ʰ Local Sidereal Time (LST) over 10 d from the HERA data repository.[7] In total, we used around 100 min of data. We used the PYUVDATA[8] software (Hazelton et al. 2017) to convert the correlator output to the Common Astronomy Software Applications (CASA)[9] Measurement Set format. Antenna 50 was
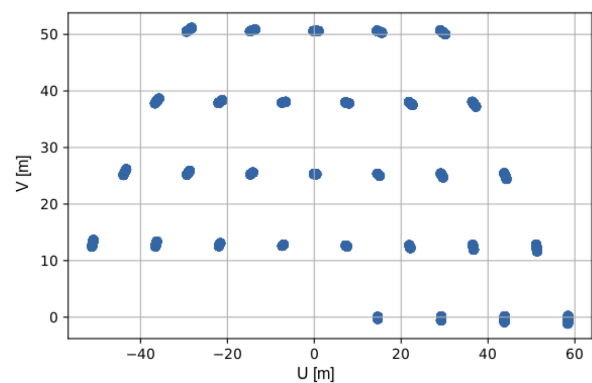
Figure 1. This figure shows the 10 min *uv* coverage around 150 MHz of HERA-47 which are analyzed in this paper.

found to be bad for the initial 7 d and was permanently flagged. We also flagged the band edges as well as the channels that were persistently affected by Radio Frequency Interference, i.e. mostly the following channels: 0–100, 379–387, 510–512, 768–770, 851–852, and 901–1023, where channel '0' corresponds to 100 MHz and channel '1023' to 200 MHz. We then used the CASA task RFLAG to perform further flagging in time and frequency. The threshold for 'timedevscale' and 'freqdevscale' was fixed to the default values of '5' each. This implies that for each channel any visibility will be flagged if the local RMS of its real and imaginary part, is larger than five times (RMS + median deviation) within a sliding time window. Similarly, for each integration time, the real and imaginary parts of the visibilities were flagged if they exceed five times the deviation from the median value across channels.

Calibration was performed using custom CASA pipelines.[10] The starting flux density model included the five brightest point sources within the HERA field of view (GLEAM 2101–2800, GLEAM 2107–2525, GLEAM 2107–2529, GLEAM 2101–2803, and GLEAM 2100–2829), chosen from the MWA GLEAM point source catalogue (Hurley-Walker et al. 2017). Their model flux density was corrected for the HERA primary beam response following the electromagnetic simulations of the HERA feed and dish (Fagnoni & de Lera Acedo 2016), obtaining a flux density estimate for each source at each frequency channel. This sky model was used to solve for three types of antenna gains: antenna-based delay ('K' term in the CASA terminology), followed by a complex gain for all the channels and the whole 10 min interval ('G' term in the CASA terminology) and by a complex bandpass calibration ('B' term in the CASA terminology). Calibration solutions were determined for the snapshot observation on 2017 October 15 and applied to the rest of the nine nights of data, though data from each day was flagged individually. The calibration solutions are used as-is, and are not smoothed across frequency before applying them to the data. This can allow spectral-dependent calibration errors generated by unmodelled sky sources and baseline-dependent systematics to be applied to the data, and can further corrupt the EoR window (Kern et al. 2020b); however, in this work we seek to model these terms through a combination of the foreground mode mixing and periodic kernel discussed next.

Calibrated visibilities were phased to a common right ascension $\alpha = 21^h$ and Fourier transformed into 21° × 21° images using the w-projection algorithm with 128 planes and the multifrequency
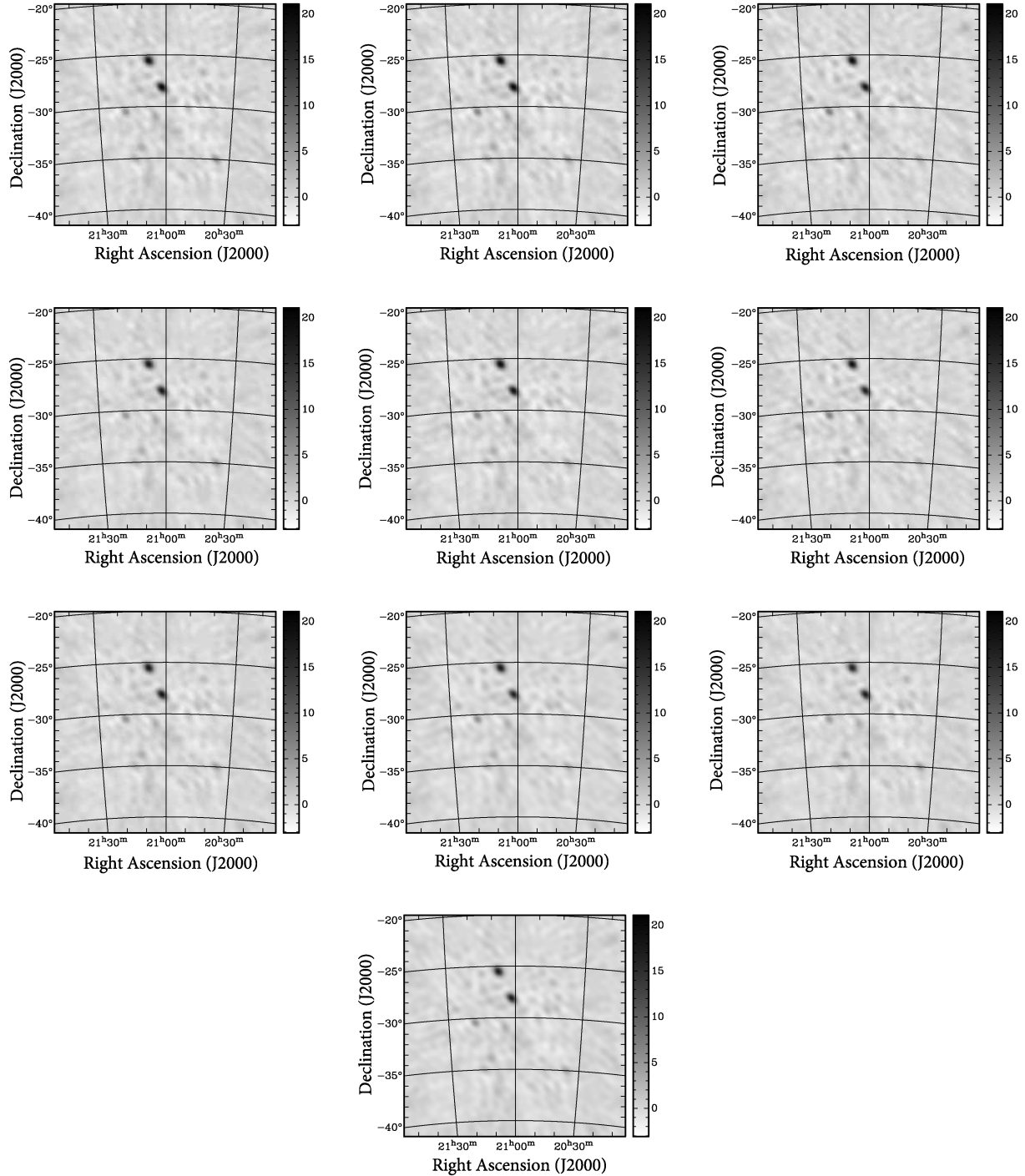
Figure 2. This figure shows 10 snapshot images from HERA-47 at 150 MHz. The flux density scale is in Jy beam^{-1} units.

synthesis algorithm to combine the whole bandwidth together. Uniform weights were used, leading to a 43.2 arcmin × 35.4 arcmin synthesized beam. Each image was conservatively deconvolved down to a threshold of 10 per cent of the image peak using the Cotton–Schwab algorithm implemented in the CASA *CLEAN* task.

Images of the 10 snapshots are shown in Fig. 2. Images at different days are very similar, qualitatively showing good instrumental stability. Image to image variation of the RMS noise in regions of the sky that are mostly empty (away from phase center and void of sources) is between 0.35 and 0.45 Jy beam^{-1}. In these parts of the sky, the primary beam response for the individual fields is considerably lower than the field centre and we expect them to be noise dominated. As the primary beam response slightly changes based on the transit time at the HERA location, we find that the peak flux density of the images varies up to 5–10 per cent over the 10 d (Fig. 3), likely due to time variations of the bandpass and imperfect primary beam corrections across snapshots – the primary beam was computed for the first snapshot but observations took place at slightly different LSTs. This variation essentially sets the
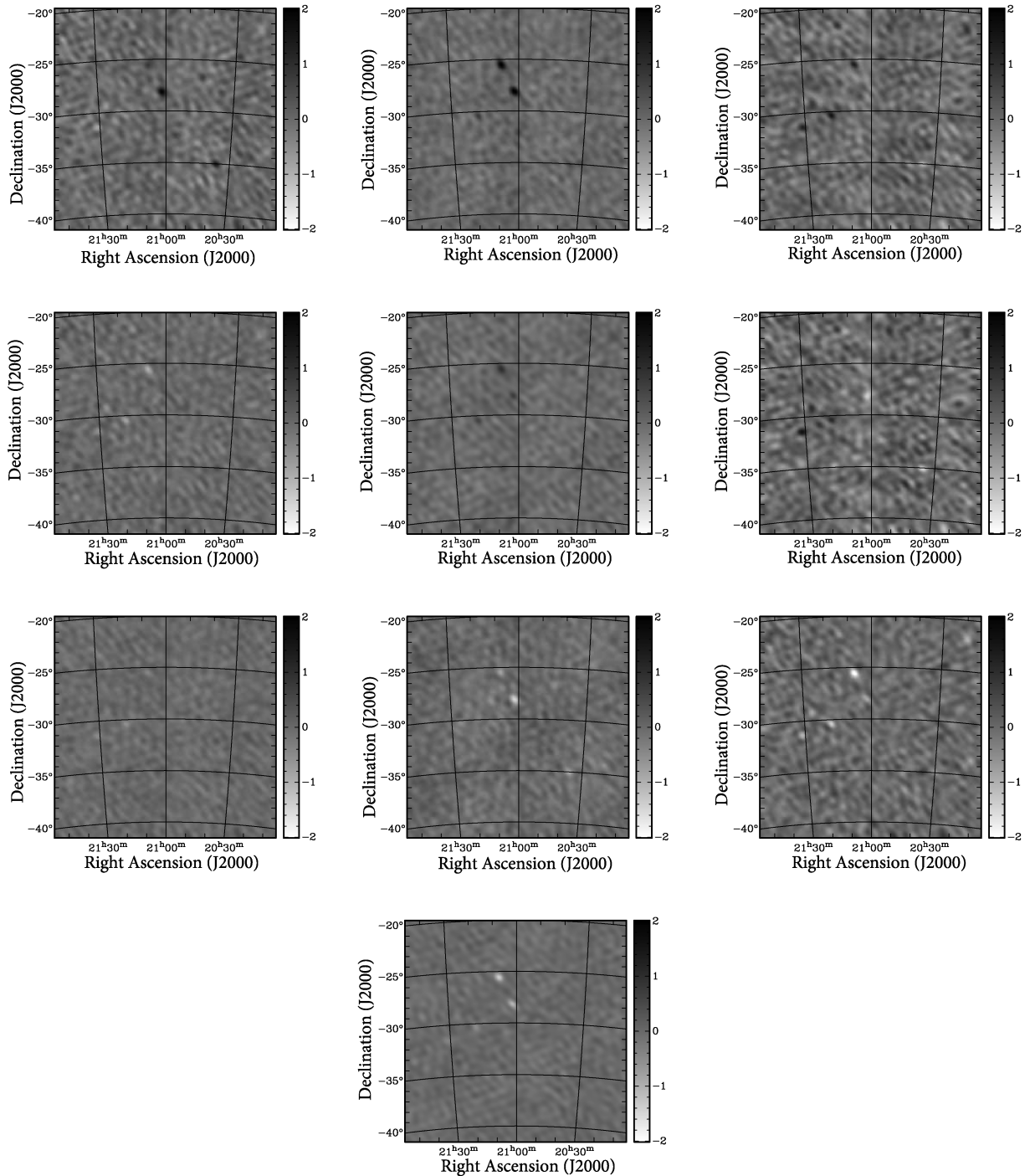
Figure 3. This figure display the 'difference image' where the mean image has been subtracted out. The flux density scale is in Jy beam^{-1} units.

accuracy of our absolute flux density calibration. We also note that Cygnus A is above the horizon at the time when observations were taken. Although ∼70° away from the pointing direction and therefore heavily attenuated by the primary beam, it still appears as a source with ∼7 Jy beam^{-1} peak flux density, possibly affecting the bandpass calibration. We also leave for future work the application of techniques that leverage on the array redundancy to improve calibration (Marthi & Chengalur 2014; Zhang, Liu & Parsons 2018; Grobler, Bernardi & Kenyon 2018; Dillon et al. 2018; 2020).

2.1 LST binning and SEFD evaluation

We bin each night of visibility data in LST. We chose a 2 min bin resolution, such that we can minimize the variation of the primary beam. For each observing night and LST bin, we only average redundant baselines (e.g. baselines of the same length and orientation). This ensures that we are coherently averaging the baselines and not mixing up emissions from the sky as the earth rotates.

We empirically estimate the System Equivalent Flux Density (SEFD) of the different LST combined visibility data sets by taking
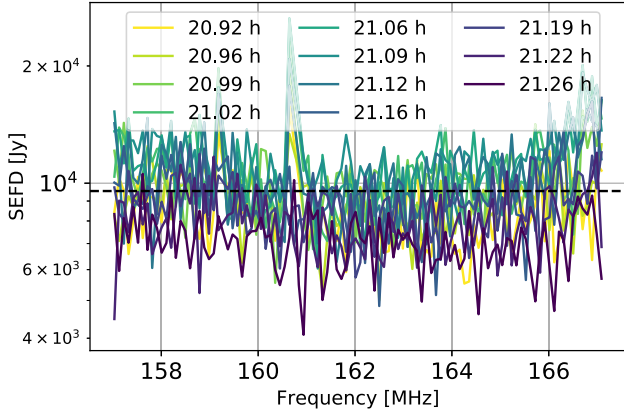
Figure 4. Estimated SEFD as a function of frequency for the different LST bins. The dashed line shows the mean value for all the LST bins and frequency channels.

the difference of two adjacent frequency channels (Patil et al. 2016). This difference can be used to estimate the noise RMS, $\sigma(u, v, \nu)$. For each polarization, we have (Thompson, Moran & Swenson 2017)

$$\sigma(u, v, \nu) = \frac{1}{\sqrt{N_{\mathrm{vis}}(u, v, \nu)}} \frac{\mathrm{SEFD}(\nu)}{\sqrt{\Delta\nu\,\Delta t}}, \tag{1}$$

where $\Delta\nu$ and Δt are the frequency channel width and the integration time, respectively. We use equation (1) to estimate the SEFD for the different LST bins as a function of frequency (Fig. 4). The factor $N_{\mathrm{vis}}(u, v, \nu)$ in equation (1) is the number of redundant visibilities. We find a SEFD $\sim(9.5 \pm 2.4) \times 10^3$ Jy (here, the mean and the uncertainty is estimated from all the LST bins and frequency channels in Fig. 4) for the 157.03–167.09 MHz range that we use for the power spectrum analysis. In temperature units, this is equivalent to $\sim$327 $\pm$ 84 K around a central frequency of 162 MHz. We use a scaling factor of $(10^{-26}\lambda^2)/(2\,k_B\,\Omega_P)$, where Ω_P is the angular area of the primary beam (Parsons et al. 2017), to convert from Jy to K. The estimated SEFD values are consistent with the HERA system temperature derived using differences of visibility spectra for sky-calibrated data for a fixed LST on two consecutive days (Carilli 2017).

Visibilities observed at the same LST time should 'see' the same sky. Assuming that over the 2-min LST bin the change in the primary beam is not significant, all the 2-min averaged visibilities corresponding to similar LST bins are therefore also coherently averaged (after baselines of same length and slope have been averaged). Visibility data sets from different LST bins, on the other hand, correspond to different parts of the sky and therefore cannot be coherently averaged. However, the 21-cm signal power spectra should only depend on baseline length, not time. We can thus incoherently combine them when producing power spectra (e.g. we average the power spectrum from different LST bins). In the following subsection, we describe our power spectrum estimation procedure. We focus our discussion on the line-of-sight and delay power spectrum in the $k_\perp - k_\parallel$ plane or, equivalently, baseline–delay plane.

2.2 Delay power spectrum

Intrinsic flat spectrum sky emission appears as a Dirac delta function in delay space, where the Fourier transform along the frequency axis (delay transform) acts as a one-dimensional, per-baseline 'image'

(Parsons et al. 2012a). Smooth-spectrum foregrounds are bound by the maximum geometric delay that depends upon the baseline length. We investigate such foreground isolation via the 'delay spectrum' $\hat{V}(\mathbf{u}, \tau)$, defined as the inverse Fourier transform of $V(\mathbf{u}, f)$ along the frequency coordinate (Parsons et al. 2012a,b):

$$\hat{V}(\mathbf{u}, \tau) \equiv \int V(\mathbf{u}, f)\,W(f)\,e^{i2\pi f\tau}\,\mathrm{d}f, \tag{2}$$

where $W(f)$ is a spectral window function (Vedantham et al. 2012; Thyagarajan et al. 2013; Choudhuri et al. 2016) and τ represents the signal delay between antenna pairs $\tau = \frac{\mathbf{u}\cdot\hat{s}}{c}$, where $\mathbf{u}$ is the baseline vector towards the direction $\hat{s}$ and c is the speed of light. We finally squared the visibilities, $\hat{V}(\mathbf{u}, \tau)$, to form the delay power spectrum. Unlike an image-based estimator where the upper and lower frequencies incorporate information from baselines of different physical length, the delay power spectrum respects baseline migration, i.e. the same baselines contribute to all frequencies (e.g. Morales et al. 2012). In our analysis, we used baselines with length $|\mathbf{u}| \leq 60$ m and a non-uniform discrete Fourier transform to compute the line-of-sight delay transform of the visibilities in order to take proper account of the flagged frequency channels. We choose a Blackman window function that offers a $\sim$−67 dB side lobe suppression.

For a single baseline, we can estimate the delay power spectrum (e.g. the cylindrical power spectrum) as (Parsons et al. 2012a):

$$P(k_\perp, k_\parallel) \approx \left(\frac{10^{-26}\lambda^2}{2\,k_B}\right)^2 \times \frac{X^2 Y}{\Omega_{PP} B}\left|\hat{V}(\mathbf{u}, \tau)\right|^2, \tag{3}$$

where λ corresponds to the wavelength of the mid-frequency of the band, k_B is the Boltzmann constant, B is the bandwidth, Ω_{PP} is the angular area of the primary beam, and X, Y are the conversion factors from angle and frequency to co-moving scales. As discussed, the power spectrum is averaged over all LST bins. Moreover, we also average over all k modes with the same $k_\perp$, i.e. the modes that have the same baseline length. We used the power-square beam from the HERA beam measurements[11,12](Parsons et al. 2017) to estimate the beam area (equation B10 in Parsons et al. 2014). The power spectrum has units of K^2 [h^{-3} cMpc^3]. Fourier modes ($k_\perp$, $k_\parallel$) are in units of inverse co-moving distance and are given by (e.g. Morales, Bowman & Hewitt 2006; Trott et al. 2012)

$$k_\perp = \frac{2\pi|\mathbf{u}|}{D_M(z)}, \tag{4}$$

$$k_\parallel = \frac{2\pi H_0 \nu_{21} E(z)}{c(1+z)^2}\tau, \tag{5}$$

$$k = \sqrt{k_\perp^2 + k_\parallel^2}, \tag{6}$$

where $D_M(z)$ is the transverse co-moving distance, H_0 is the Hubble constant, ν_{21} is the frequency of the hyperfine transition, and $E(z)$ is the dimensionless Hubble parameter (Hogg 1999).

Fig. 5 shows the delay power spectrum for a 2 min LST binned data (e.g. we consider one LST bin only) as a function of $k_\parallel$, up to $\tau \sim 3.6\,\mu$s, corresponding to $k_\parallel \sim 2.0$ h cMpc^{-1}. We used a 10 MHz bandwidth centred at 162.06 MHz to estimate the delay spectrum. We found most of the foreground power is confined within $k_\parallel \leq 0.2$ h cMpc^{-1} and foreground excess beyond that is largely limited for most baselines, however, there is some signature of a signal with an $\sim$1 MHz period (Kern et al. 2020a), corresponding to $k_\parallel \sim 0.5$ h cMpc^{-1}, for all the baselines considered.

[11]https://github.com/HERA-Team/hera-cst
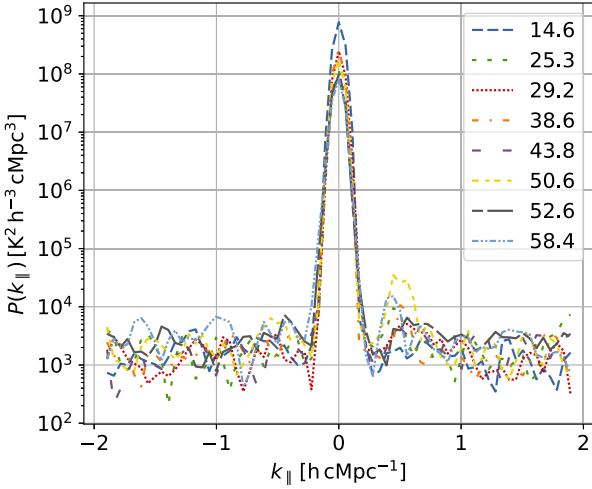[12]http://reionization.org/science/memos/

Figure 5. Power spectrum of a 2 min LST binned data as a function of $k_{\parallel}$ for the baselines included in our analysis. The baselines in units of meters are shown in the legend.
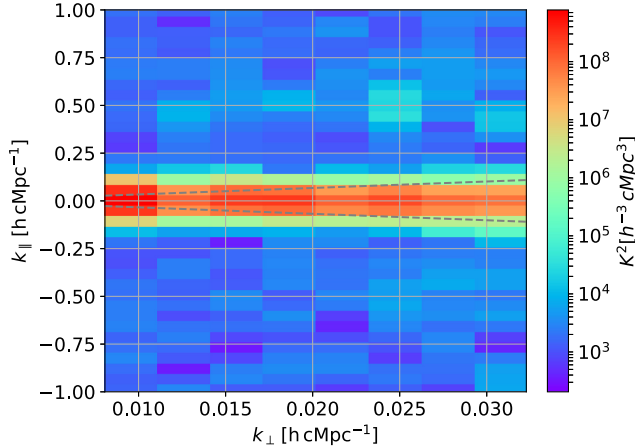

Figure 6. Delay power spectrum in the $(k_{\perp} - k_{\parallel})$ plane for a 2 min LST binned data. The dashed line represents the horizon line corresponding to $\theta_{\max} = 90°$.

Fig. 6 presents the delay power spectra in the $k_{\perp} - k_{\parallel}$ plane related to the same LST bin. We found that the smooth diffuse foreground in the $k_{\perp} - k_{\parallel}$ plane dominates at low $k_{\parallel}$, with most power localized within $k_{\parallel} \leq 0.2 \, \mathrm{h \, cMpc^{-1}}$. The foreground power drops by four to five orders of magnitude in the $k_{\parallel} \geq 0.2 \, \mathrm{h \, cMpc^{-1}}$ region, where the EoR signal is expected to dominate over the foreground emission. We notice some signature of a wedge-like structure in k space (Datta et al. 2010; Morales et al. 2012), although in the current HERA antenna layout we are mostly limited to short baselines and hence the foreground wedge is not clearly visible. This wedge line is defined by (Liu, Parsons & Trott 2014; Dillon et al. 2015)

$$k_{\parallel} = \left[\sin(\theta_{\text{field}}) \frac{H_0 D_M(z) E(z)}{c \, (1+z)} \right] k_{\perp}, \tag{7}$$

where θ_{field} is the angular radius of the field of view. We also show on the figure the wedge line corresponding to the horizon limit ($\theta_{\max} = 90°$).

3 GAUSSIAN PROCESS REGRESSION AND FOREGROUND CHARACTERIZATION

The delay power spectrum results show that the data is mostly dominated by foregrounds. GPR offers a way to model these foregrounds in a maximum likelihood way. In this section, we summarize the GPR formalism (for a detailed review of how the method works see Mertens et al. 2018) and apply it to model foreground components in HERA-47 observations.

In this framework, the different components of 21-cm observations, such as the astrophysical foregrounds, mode-mixing contaminants, and the 21-cm signal, are modelled with a GP. A GP is the joint distribution of a collection of normally distributed random variables (Rasmussen et al. 2005; Gelman et al. 2014). The covariance matrix of this distribution is specified by a covariance function, which defines the covariance between pairs of observed data points (i.e. at different frequencies). The co-variance function ultimately determines the structure that the GP will be able to model (for example, here the smoothness of the foregrounds).

The GPR process requires the choice of the model for the covariance function and a selection of the best-fittingting parameters of such a model (what we call the hyper-parameters). Model selection is done in a Bayesian sense by maximizing the marginal likelihood, also called the evidence, which is the integral of the likelihood over the prior range, given the data. For a fixed model, standard gradient-based optimization or Monte Carlo Markov Chain (MCMC) methods can be adapted to determine the best-fitting parameters of the covariance functions. We note here that currently we model the data only in the frequency axis and no baseline dependence has been introduced in the hyper-parameter optimization with GPR (i.e. there is no dependence on baseline length). This assumption is supported by Fig. 5, where we can see that the power spectrum is similar for different baseline lengths.

In the following equations, d represents the time-averaged visibilities within a given LST bin and we have not explicitly shown the time dependence of the data. Considering an observed data d and a GP co-variance model that includes a foreground term K_f and a residual term (noise and 21-cm signal) K_r, the data co-variance can be expressed as, $K = K_f + K_r$. After GPR, we can retrieve the foreground part of the signal $E(f_{fg})$ that always refers to the total signal except for noise or 21-cm signal through basically a Wiener filter (Wiener 1949):

$$E(f_{fg}) = K_f [K_f + K_r]^{-1} d. \tag{8}$$

In the GPR context, this is referred to as the posterior mean matrix, while

$$\text{cov}(f_{fg}) = K_f - K_f [K_f + K_r]^{-1} K_f \tag{9}$$

is the posterior co-variance matrix.

Assuming that the GP co-variance model is optimal and taking $\langle dd^H \rangle = K_f + K_r$, then $\langle E(f_{fg}) E(f_{fg})^H \rangle = K_f - \text{cov}(f_{fg})$. This highlights that to obtain the expected co-variance model of the foregrounds, K_f, directly from $E(f_{fg})$, we need to un-bias the estimator using $\text{cov}(f_{fg})$. We implement a similar unbiasing for the delay power-spectra of the different foreground components by first taking the delay transform of $E(f_{fg})$ and $\text{cov}(f_{fg})$ and then adding them in the power spectrum domain. We finally normalize by the observed cosmological volume to construct the delay power spectrum $P(k_{\perp}, k_{\parallel})$ in units of K^2 [h^{-3} cMpc^3]. More specifically, we calculate the covariance matrices by fitting the hyper-parameters to all the data, while the posterior mean is obtained for each time-averaged visibility (so, the covariance calculated from $E(f_{fg})$ is not
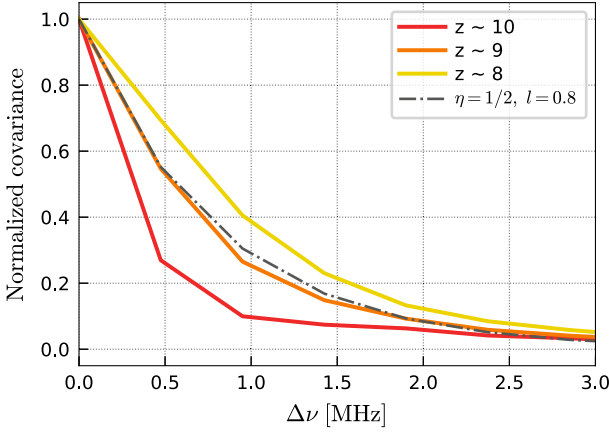
Figure 7. Example of the normalized GP exponential co-variance function with a frequency coherence scale $l_{inj} = 0.8$ MHz (dot–dashed line), compared to the co-variance of a simulated 21-cm EoR signal at different redshifts using 21cmFAST (Mesinger & Furlanetto 2007; Mesinger et al. 2011).

necessarily the same as the initial K_f). In this paper, we consider the power spectrum of the different foreground components. This implies calculating $E(f_{fg})$ for each of the foreground components, where we replace K_f by the optimized co-variance of the corresponding foreground component (while keeping the term in square brackets, $[K_f + K_r]$, the same, since it is the total co-variance).

3.1 Covariance functions

In this section, we review the co-variance functions for the different components of the data. The selection of a co-variance function κ for the 21-cm signal can be chosen by comparison to a range of 21-cm signal simulations. For this analysis, we choose a Matern $\eta = 1/2$ co-variance function with a frequency coherence scale l parameter:

$$\kappa_{Matern}(\nu_p, \nu_q) = \sigma_f^2 \frac{2^{1-\eta}}{\Gamma(\eta)} \left(\frac{\sqrt{2\eta}r}{l}\right)^\eta K_\eta \left(\frac{\sqrt{2\eta}r}{l}\right), \quad (10)$$

where σ_f is the signal variance, $r = |\nu_q - \nu_p|$ and K_η is the modified Bessel function of the second kind. The parameter η controls the smoothness of the resulting function. For $\eta = 1/2$, the Matern kernel is equivalent to an exponential kernel. The choice of this co-variance kernel well matches the co-variance of the EoR signal with 21cmFAST (Mesinger, Furlanetto & Cen 2011, Fig. 7). Following Mertens et al. (2018), we used a uniform prior in the 0.01–1.25 MHz range on the hyper-parameter l.

The intrinsic smooth foregrounds are modelled with a Radial Basis Function (RBF) kernel (also known as the 'squared-exponential' or a 'Gaussian' kernel):

$$\kappa_{RBF}(\nu_p, \nu_q) = \sigma_f^2 \exp\left(-\frac{r^2}{2l^2}\right), \quad (11)$$

where the coherence scale l controls the smoothness of the function, σ_f is the signal variance and the frequency coherence scale was bounded in the 10–200 MHz range. We note that the Matern kernel (equation 10) is a generalization of the RBF kernel, parametrized by an additional parameter η. When η tends to infinity, the kernel becomes equivalent to RBF kernel. Medium-scale fluctuations coming from a combination of the instrumental chromaticity and imperfect calibration (termed as 'mode-mixing' components) are

also modelled by a GP with an RBF covariance function where the characteristic coherence scale l is bounded in the 2–20 MHz range.

3.2 Foreground modelling

Here, we discuss the GPR model foreground components, including the modelling and subsequent removal of the frequency and amplitude, modulated periodic signal with an additional GP co-variance kernel. Again, following Mertens et al. (2018), we modelled the GP co-variance function by decomposing the foreground co-variance as

$$K_{fg} = K_{sky} + K_{mix}, \quad (12)$$

where the 'sky' denotes the intrinsic smooth foreground sky and 'mix' denotes the mode-mixing contaminants that introduce oscillations in frequency mostly caused by the instrument. It is expected that K_{sky} will pick up the frequency dependence of the foreground signal at a given uv point, whereas the mixing component K_{mix} can model relatively rapidly varying foreground components such as the fact that the uv point itself also moves in the uv plane with frequency (Morales et al. 2012) and hence is sensitive to extra angular and frequency scale structures. We remind the reader that here we use the RBF kernel to model the foregrounds and an exponential kernel is used to represent the 21-cm signal co-variance function. To select the optimal mode-mixing co-variance function, we considered the Matern kernel with $\eta = 3/2$ and $5/2$ along with the RBF kernel with a uniform prior in the 2–20 MHz range. We found that the difference in the log-likelihood for the RBF kernel from the Matern 3/2 and the Matern 5/2 kernels are 1717 and 854, respectively (keeping the other covariances fixed). Based on this evidence, we choose to use the RBF kernel to model the mode-mixing component. We optimize the log-marginal-likelihood for the full set of visibilities (real and imaginary part separately) for the six variances and the coherence length scales hyper-parameters (namely, σ_{21}^2, l_{21}, σ_{sky}^2, l_{sky}, σ_{mix}^2, and l_{mix}), assuming the coherence scale is spatially invariant, i.e. the same for each baseline type. The PYTHON package GPY[13] is used to do the optimization using the full set of visibilities. The noise term is modelled with a fixed variance where the covariance matrix describes the variance along the frequency direction. The noise in the real data has both a frequency and a time dependence but here we choose only the frequency axis to approximate the noise variance. We found that the frequency coherence scale of the 'sky' and 'mix' co-variance kernel are about 20 and 2.4 MHz, respectively.

In general, the coherence scale is expected to be dependent on the baseline length, with longer baselines de-correlating faster than shorter baselines. We investigate this effect by implementing a 'per-baseline' GPR approach that allows us to model a coherence scale for different baseline lengths using a smaller data set with an increased number of degrees of freedom. We found that the coherence scale decreases at longer baselines (Fig. 8), from $l \sim 3.2$ MHz for the 14.6 m baseline to $l \sim 2.2$ MHz for $\sim$60 m baseline. In the 'all-baselines' GPR implementation, the optimal coherence scale was about 2.4 MHz, which falls inside the maximum–minimum range of 'per-baseline' GPR approach. These results further agree well with our initial choice of the 2–20 MHz prior range for the medium-scale frequency fluctuations introduced in Section 3.1. The main reason for this behaviour is the limited baseline range used in this analysis over which the foreground co-variance remains similar.
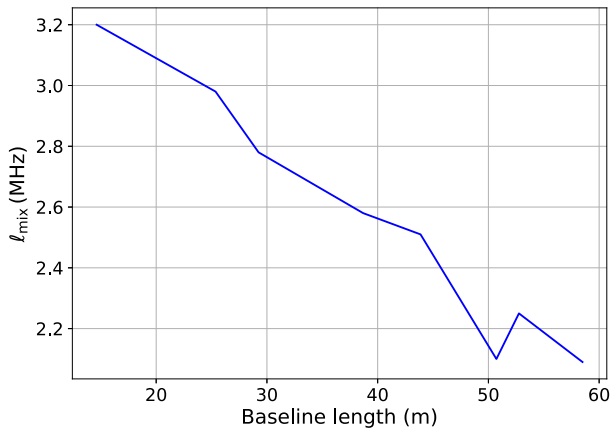
[13]https://sheffieldml.github.io/GPy/

Figure 8. The optimized co-variance scale l_{mix} for the mode-mixing kernel as a function of baseline lengths using a 'per-baseline' GPR technique.

The inclusion of significantly longer baselines would likely require a 'per-baseline' GPR fit as the foreground coherence will change more significantly across the range of baseline lengths. This could be implemented without significantly increasing the number of degrees of freedom by allowing the coherence scale parameters to be a function of the baseline length. We leave this investigation for future work.

We then considered all nights, coherently LST combined data sets for which we estimate different foreground components using GPR. For the GPR foreground modelling and the power spectrum estimation we used the PYTHON package PS_EOR.[14]

Fig. 9 shows the power spectrum and the variance across frequency for the different foreground components that we recover using the GPR technique. Note that in our GPR power spectrum estimation method, the hyper-parameters of the covariance model are optimized using the Bayesian evidence. Doing an MCMC, we then get the posterior distribution of these hyper-parameters, and this is the first source of uncertainty that we use and propagate to the power spectra. The shaded area in Fig. 9 shows the propagated 2σ uncertainty on the hyper-parameters and the uncertainty on the model fit (equation 9) and from the MCMC run (for details see Section 3.2.3) on to the different foreground power spectrum components. An important point to note here is that the uncertainty on the power spectra that we estimate are correct assuming that our assumed co-variance functions are appropriate.

We notice that the 'FG mix' model has a small coherence scale (2–3 MHz) and therefore the variance has a wave-like pattern, but for the intrinsic foregrounds it is mostly smooth across frequency. We detect a 'bump' in the power spectrum around $k_{\parallel} \sim 0.5\,h\,cMpc^{-1}$, corresponding to a $\sim$900 ns delay, indicating the presence of a non-negligible contamination in the data (possibly due to internal signal chain reflections, or a more dominant instrumental cross-talk feature spanning delays of 800–1200 ns that does not look like an EoR signal and can therefore be filtered out, Kern et al. 2020a), which we investigate in more detail in Section 3.2.1.

Fig. 10 shows the correlation along the frequency direction of the different GPR components (intrinsic foreground, mode-mixing foreground, and residuals) as a function of LST difference for all the combinations of LST binned data sets, covering an LST range of $\sim$20.92^h−21.26^h (each cross represents an LST bin difference).

Here, we compute the correlation for every combination of the LST binned visibility data sets for each baseline and finally we average over the baselines to determine the final value. The correlation coefficient is given by

$$\rho_{LST1,LST2}^{cmpt} = \frac{cov\left(fg_{LST1}^{cmpt}, fg_{LST2}^{cmpt}\right)}{\sigma\left(fg_{LST1}^{cmpt}\right)\sigma\left(fg_{LST2}^{cmpt}\right)}, \quad (13)$$

where fg_{LST1}^{cmpt} and fg_{LST2}^{cmpt} are the foreground model components corresponding to the 'sky', 'mix', or the residual at two different LSTs. For large LST differences, the correlation should go down since we are looking at different parts of the sky. From Fig. 10, we see that the intrinsic foreground correlation remains above 80 per cent regardless of the time difference. The correlation coefficient starts to decay only for LST differences >2–4 min (as the sky starts to shift). The mode-mixing de-correlates significantly as a function of LST difference. This typically depends on the coherence scale in the *uv*-plane as a baseline moves through it and is faster for longer baselines. The mode-mixing is also more affected by LST difference de-correlation because it contains fluctuations due to small beam differences mainly further away from the phase centre.

3.2.1 Characteristics of the periodic signal

After foreground removal, the residual power spectrum is dominated by noise and an almost periodic signal that reveals itself by an excess power at $k_{\parallel} \sim 0.5\,h\,cMpc^{-1}$. We find the periodic signal is baseline dependent and it also varies with LST difference. Fig. 11 shows the correlation of the residual visibilities (equation 13) for different baselines as a function of LST difference. Two periodic signals from two different LST times appear to be phase shifted. A closer inspection reveals that the amplitude and periodicity of this signal does not remain stationary but varies with frequency. For example, the residual visibilities for a specific baseline and the fit to the periodic signal is shown in Fig. 12. Similar frequency-dependent complex patterns are also seen for other baselines. This profile can be fitted using the GPR method and a combination of a RBF and Cosine co-variance function, K_{per}, on each baseline individually. The co-variance function κ_{per} for the periodic kernel depends on the characteristic coherence scale l_{per} over which the periodic signal vary, the signal variance σ_{per}^2, and the period p_{per}:

$$\kappa_{per}(\nu_p, \nu_q) = \sigma_{per}^2 \exp\left(-\frac{r^2}{2l_{per}^2}\right)\cos\left(\frac{2\pi r}{p_{per}}\right). \quad (14)$$

We found the main periodicity is $\sim$1 MHz.

Kern et al. (2020a) provides a thorough investigation of such systematic effect, attributing it to a combination of instrumental cross-coupling (e.g. mutual coupling and cross-talk) and cable reflections within the analogue signal chain. Kern et al. (2019) present methods for modelling and removing these systematic terms in the data. In the following section, we show how it can be modelled and subtracted in the GPR formalism.

3.2.2 Filtering the periodic signal with GPR

To model this locally periodic signal with amplitude varying over a certain coherence scale, we introduce an additional kernel in the foreground co-variance model. We use a combination of an RBF and a cosine kernel to model the period in
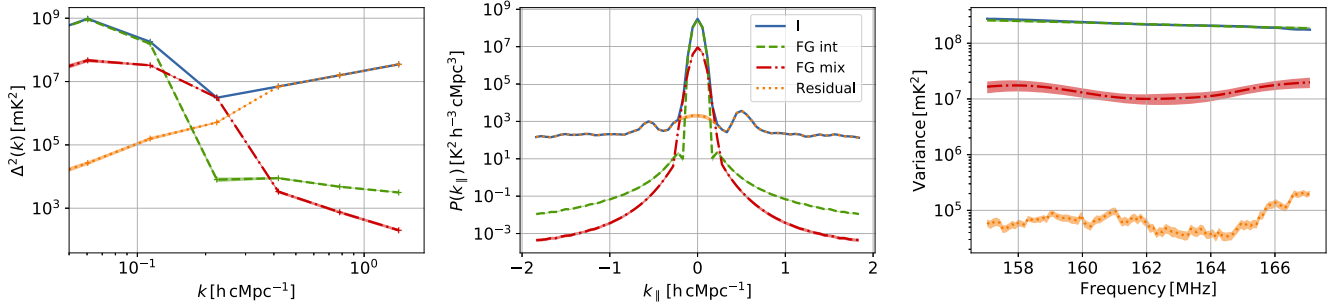
Figure 9. Power spectra and derived foreground components after GPR analysis of the coherently averaged combined data: spherically averaged power spectra (left-hand panel), delay spectra averaged over all baselines (middle panel). The right-hand panel shows the variance of the different foreground components across frequency. The shaded area highlights the uncertainty (2σ) from the GP process (equation 9) and the uncertainty on the model fit from the MCMC run on to the foreground power spectrum components.
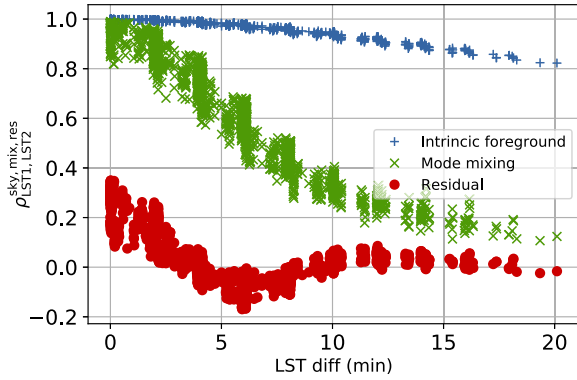
Figure 10. Correlation coefficient $\rho_{LST1,LST2}^{sky,mix,res}$ as a function of LST difference for all the night, LST-binned data. Each marker symbol ('+', 'x' or 'o') represents an LST difference.
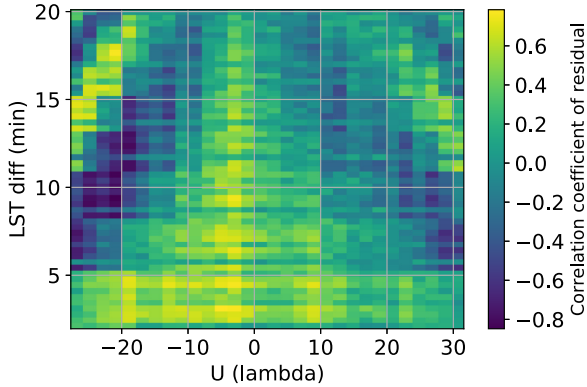
Figure 11. Correlation coefficients of the residual visibility data after GPR as a function of baseline length and LST difference. The plot shows very strong dependence of the periodic signal on baseline length $|u|$ and it also varies with LST difference.

frequency. The updated foreground co-variance function is modelled as

$$K_{fg} = K_{sky} + K_{mix} + K_{per}, \tag{15}$$

where the K_{per} represents the periodic signal contaminant (see equation 14). We used this updated foreground co-variance model in our GP optimization. The GPR estimates of the parameters for the periodic co-variance function are found to be $p_{per} \sim 1$ MHz and $l_{per} \sim 1.2$ MHz, respectively.
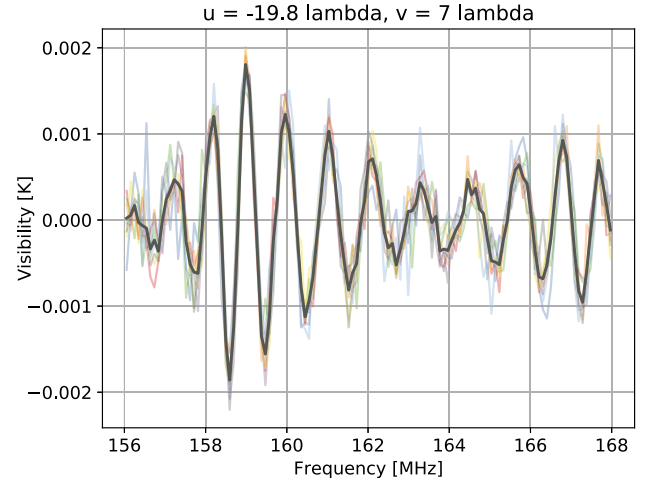

Figure 12. Example of the periodic signal for 38.6 m baseline with coordinates $(u, v) = (-19.8, 7)\lambda$. The real part of the visibility is shown in 'K' units. The different transparent lines correspond to different LST data sets on which we apply a frequency phase offset to align the periodic signal. This signal can be fitted using GPR with a combination of an RBF and Cosine co-variance kernel (the solid line).

Fig. 13 displays the power spectrum of different GPR modelled components including the periodic signal. We notice that the periodic signal peaks around ∼0.4–0.8 h cMpc^{-1}. In the middle panel, we display the GPR model that nicely isolates the periodic signal component around $k_{\parallel} \sim 0.5$ h cMpc^{-1}. In general, we find that the periodic signal is k dependent. It appears at $k \sim 0.17$ h cMpc^{-1}, reaching a ∼10^7 mK^2 peak at $k \sim 0.4$ cMpc^{-1} – approximately six orders of magnitude brighter than the expected 21-cm power spectrum (Mesinger et al. 2011).

The average variance across frequency for the periodic signal component is ∼3.8 × 10^4 mK^2, while the mean variance of the mode-mixing signal is around ∼1.2 × 10^7 mK^2, approximately three orders of magnitude higher. The noise power spectrum shown in Fig. 13 is estimated by splitting the data set in even and odd times with a 10.7 s time separation and taking the difference between the two. At this time resolution, the foregrounds cancel out almost perfectly. We find the residual power spectrum level is close to the estimated noise power spectrum, especially at $|k_{\parallel}| \geq 0.85$ h cMpc^{-1}.

The residual in Fig. 14 reveals that there is still some time correlation left, but overall, we find the residuals have become more uncorrelated and noise-like compared to Fig. 10 where the
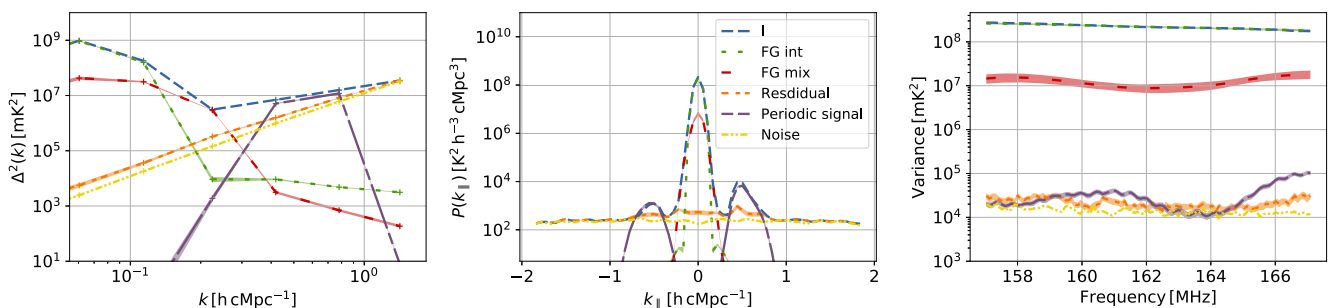
Figure 13. Same as Fig. 9 but now including the periodic signal and the estimated noise power spectrum and variance.
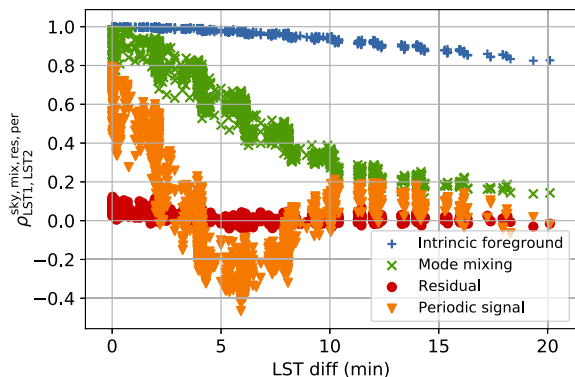
Figure 14. Same as Fig. 10, but here the periodic signal has been modelled with an RBF and cosine kernel and then removed.

GP foreground co-variance was modelled only with a combination of 'sky' and 'mode-mixing' kernels.

3.2.3 *Foreground model hyper-parameter uncertainties*

We sampled the posterior distribution of the foreground model hyper-parameters and characterize their correlation with an MCMC. We used the EMCEE PYTHON package[15] (Foreman-Mackey et al. 2013), which uses an ensemble sampler algorithm based on the affine-invariant sampling algorithm (Goodman & Weare 2010).

Fig. 15 shows the resulting posterior probability distribution of the GP model hyper-parameters. The variance of the EoR kernel, which was modelled with a GP exponential kernel, is found to be un-constrained and low. The data can be well modelled by the 'sky', 'mix', 'per' (periodic) foreground kernels and the noise covariance matrix (modelled with a fixed variance) that contributes a large part of the variance at large $k_\parallel$. We compared the evidence values with and without the EoR co-variance kernel in the GP optimization. We find the evidence remains mostly unchanged and the Bayes factor (Jeffreys 1961) is around ∼0.93 for GP models with and without the EoR co-variance kernel. This essentially confirms that the signal is dominated by a noise-like component once the foregrounds are removed and adding an EoR kernel has an insignificant effect. Overall, the confidence intervals of other kernel hyper-parameters are reasonably well constrained, except the variance of the '21-cm signal' component that is consistent with zero. The significance of the coherence scale of the '21-cm signal' is also reduced given the non-significant variance of this component. Table 1 highlights the parameter estimates and confidence intervals

for the posterior probability distribution of the foreground model hyper-parameters. The estimated median values of the frequency coherence scale of the 'sky' and 'mix' covariance kernel is about 19.4 and 2.4 MHz, respectively, which is close to the GPR optimized values as presented in Section 3.2.

4 DISCUSSION AND CONCLUSIONS

In this paper, we have used a novel foreground separation method, first introduced in Mertens et al. (2018), in order to model foregrounds with the HERA-47 array. The mainstream HERA data analysis takes advantage of the concept of avoiding foregrounds and provides a single-baseline power spectrum estimate: recent data analysis showed evidence of systematic effects that contaminate the EoR window, and motivated the development of strategies to mitigate their impact to the avoidance paradigm (Kern et al. 2020b). An alternative effort to detect the 21-cm signal using closure quantities is actively being pursued (Thyagarajan, Carilli & Nikolic 2018; Carilli et al. 2018; Carilli, Thyagarayan & Kent 2020).

The method presented here uses GPR to model various stochastic foreground components, such as the spectrally smooth intrinsic sky, mode-mixing components generating from the chromatic instrument and imperfect calibration, as well as a 21-cm signal. It therefore bears analogies with the avoidance approach as they both attempt to model and subtract systematic effects in the EoR window, but also more broadly models the foreground emission – which is not within the purpose of the avoidance approach. Foreground modelling may be a necessary step in order to reduce the leakage in the EoR window and access the high signal-to-noise ratio small k modes (e.g. Kerrigan et al. 2018; Ewall-Wice et al. 2020; Lanman, Pober & Kern 2020).

Our analysis included a different co-variance function for each of intrinsic sky, mode-mixing, and 21-cm signal components in the GP modelling. We found that the frequency coherence scale of the 'sky' and 'mix' co-variance kernel are about 20 and 2.4 MHz, respectively. As a comparison, the typical (theoretical) frequency coherence scale for the 21-cm EoR signal is found to be around ∼0.8 MHz, when fitted to the co-variance of a simulated 21-cm EoR template. The foreground power spectrum is shown to be contaminated by an ∼1 MHz periodic signal whose amplitude changes from baseline to baseline. The periodic signal dominates the $0.25 < k < 0.9\,h\,\mathrm{cMpc}^{-1}$ range. We included a combination of RBF and a cosine kernel to model this signal within our GPR method and found a fairly cleaner and flatter residual power spectrum across the $0.05 < k < 1.83\,h\,\mathrm{cMpc}^{-1}$ range. The residual power spectrum is also mostly consistent with the estimated noise power spectrum, especially at high $k_\parallel$ values, whereas residuals are still present in the foreground and periodic signal-dominated region of the pwoer spectrum.
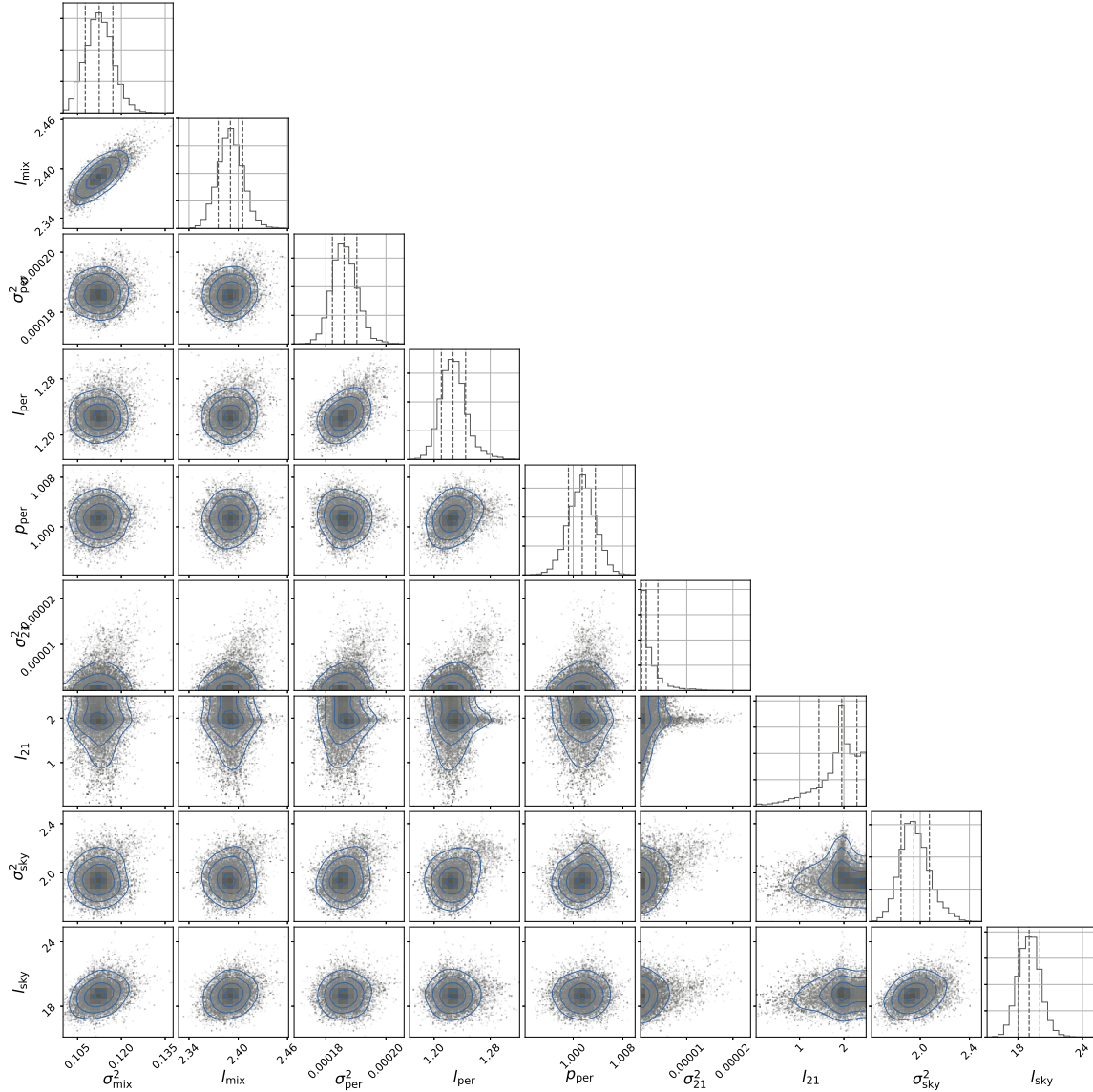
Figure 15. Posterior probability distributions of the GP model hyper-parameters. We show here the coherence scale and strength of the EoR co-variance kernel (l_{21} in MHz and σ_{21} in K^2), the coherence scale and strength of the mode-mixing foreground kernel (l_{mix} in MHz, σ_{mix} in K^2), the intrinsic foreground kernel (l_{sky} in MHz, σ_{sky} in K^2) and the periodic co-variance kernel hyper-parameters (l_{per}, p_{per} in MHz and σ_{per}^2 in K^2). The vertical dashed lines show the first, second, and third (Q1, Q2, and Q3) quantile levels. The diagonal panels represent the marginalized probability distribution of each parameter.

Table 1. Summary of the estimated median and confidence intervals (first and third quantile levels (Q1 and Q3)) of the respective GP model hyper-parameters including the periodic co-variance kernel.

Hyper-parameter	Prior	Estimate
l_{mix} (MHz)	$\mathcal{U}(2, 20)$	$2.40^{+0.02}_{-0.01}$
σ_{mix}^2 (K^2)	$\mathcal{U}(0.1, 0.9)$	$0.115^{+0.007}_{-0.005}$
l_{sky} (MHz)	$\mathcal{U}(10, 200)$	$19.42^{+1.25}_{-1.18}$
σ_{sky}^2 (K^2)	$\mathcal{U}(0.02, 2.5)$	$1.89^{+0.10}_{-0.09}$
l_{per} (MHz)	$\mathcal{U}(1, 5)$	$1.23^{+0.01}_{-0.01}$
p_{per} (MHz)	$\mathcal{U}(0.628, 1.256)$	$0.999^{+0.002}_{-0.002}$
σ_{per}^2 (K^2)	$\mathcal{U}(0.00001, 0.01)$	$0.000183^{+0.000004}_{-0.000003}$

As foreground subtraction is potentially at risk of altering the 21-cm signal, we plan to further explore this approach using more HERA data and test the cleaning with signal injection tests using full-scale HERA simulations. In this paper, we have restricted ourselves to foreground modelling only and left the characterization of residual power spectra to future work, which will include end-to-end signal injection tests.

Finally, we note that the foreground model used in this paper might not be complete, although it seems to be enough at this noise level. In particular, it does not include other foregrounds contaminants such as the instrumental polarization leakage, residual RFIs, and the phase errors caused by the ionosphere or imperfect calibration. We plan to include these additional subtle effects in our GP co-variance modelling. In addition to these, we intend to implement a per-baseline GPR approach where the coherence scale parameters are a function of the baseline length without exploding

the number of degrees of freedom of the GPR fit. This will be relevant for longer HERA baselines where the larger baselines will de-correlate faster compared to the shorter baselines. Also, the present mode-mixing model can be improved by integrating the $k_\perp$ dependence of the foreground wedge. We further plan to include the isotropic nature of the 21-cm signal and its evolution at different redshift bins, which will also ensure a more sensitive and detailed modelling.

ACKNOWLEDGEMENTS

REFERENCES

Ali Z. S. et al., 2015, ApJ, 809, 61
Asad K. M. B. et al., 2015, MNRAS, 451, 3709
Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, MNRAS, 461, 3135
Barry N. et al., 2019, 884, 1
Beardsley A. P. et al., 2016, ApJ, 833, 102,
Bernardi G. et al., 2009, A&A, 500, 965
Bernardi G. et al., 2010, A&A, 522, A67
Bernardi G. et al., 2016, MNRAS, 461, 2847
Bharadwaj S., Sethi S. K., 2001, J Astrophys. Astron., 22, 293
Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, MNRAS, 416, L70
Bonaldi A., Brown M. L., 2015, MNRAS, 447, 1973

Bowman J. D., Rogers A. E. E., 2010, Nature, 468, 796
Bowman J. D., Morales M. F., Hewitt J. N., 2009, ApJ, 695, 183
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Byrne R. et al., 2019, ApJ, 875, 70
Carilli C. L., 2017, HERA Memo 60. http://reionization.org/science/memos/
Carilli C. L., Nikolic B., Thyagarayan N., Gale-Sides K., 2018, Radio Sci., 53, 845
Carilli C. L., Thyagarayan N., Kent J., 2020, ApJS, 247, 67
Chapman E. et al., 2013, MNRAS, 429, 165
Chapman E., Zaroubi S., Abdalla F., Dulwich F., Jelić V., Mort B., 2016, MNRAS, 458, 2928
Choudhuri S., Bharadwaj S., Chatterjee S., Ali S. S., Roy N., Ghosh A., 2016, MNRAS, 463, 4093
Datta A., Bowman J. D., Carilli C. L., 2010, ApJ, 724, 526
DeBoer D. R. et al., 2017, PASP, 129, 045001
Dillon J. S., Parsons A. R., 2016, ApJ, 826, 181
Dillon J. S. et al., 2014, Phys. Rev. D, 89, 023002
Dillon J. S. et al., 2015, Phys. Rev. D, 91, 123011
Dillon J. S. et al., 2018, MNRAS, 477, 5670
Dillon J. S. et al., 2020, preprint (arXiv:2003.08399)
Ewall-Wice A., Dillon J. S., Liu A., Hewitt J., 2017, MNRAS, 470, 1849
Ewall-Wice A. et al., 2020, preprint (arXiv:2004.11397)
Fagnoni N., de Lera Acedo E., 2016, CST Simulation of HERA and Comparison with Measurements, HERA Memo 21
Fialkov A., Barkana R., Visbal E., 2014, Nature, 506, 197
Fialkov A., Cohen A., Barkana R., Silk J., 2017, MNRAS, 464, 3498
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Fraser S. et al., 2018, Phys. Lett. B, 785, 159
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Gehlot B. K. et al., 2019, MNRAS, 488, 4271
Geil P. M., Gaensler B. M., Wyithe J. S. B., 2011, MNRAS, 418, 516
Gelman et al., 2014, Bayesian Data Analysis, 3rd edn. Chapman & Hall, London
Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N., 2011, MNRAS, 418, 2584
Goodman J., Weare J., 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Greenhill L. J., Bernardi G., 2012, preprint (arXiv:1201.1700)
Grobler T. L., Bernardi G., Kenyon J. S., 2018, MNRAS, 476, 2410
Harker G. et al., 2009, MNRAS, 397, 1138
Hazelton B. et al., 2017, J. Open Source Softw., 2, 140
Hogg D. W., 1999, preprint (arXiv:astro-ph/9905116)
Hurley-Walker N. et al., 2017, MNRAS, 464, 1146
Jeffreys H., 1961, Theory of probability, 3rd edn. Clarendon Press, Oxford
Jelić V. et al., 2008, MNRAS, 389, 1319
Jensen H., Majumdar S., Mellema G., Lidz A., Iliev I. T., Dixon K. L., 2016, MNRAS, 456, 66
Kern N. S., Parsons A. R., Dillon J. S., Lanman A. E., Fagnoni N., de Lera Acedo E., 2019, ApJ, 884, 105
Kern N. S. et al., 2020a, ApJ, 888, 70
Kern N. S. et al., 2020b, ApJ, 890, 122
Kerrigan J. R. et al., 2018, ApJ, 864, 131
Kolopanis M. et al., 2019, ApJ, 883, 133
Koopmans L. V. E., 2010, ApJ, 718, 963
Lanman A., Pober J. C., Kern N. S., 2020, MNRAS, 487, 5840
Li W., Pober J. C., Barry N., 2019, ApJ, 887, 14
Lidz A., Zahn O., McQuinn M., Zaldarriaga M., Hernquist L., 2008, ApJ, 680, 962
Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006
Liu A., Parsons A. R., Trott C. M., 2014, Phys. Rev. D, 90, 023018
Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe, by Abraham Loeb and Steven R. Furlanetto. Princeton Univ. Press, Princeton, NJ
Marthi V. R., Chengalur J., 2014, MNRAS, 437, 524
McQuinn M., O'Leary R. M., 2012, ApJ, 760, 3
Mellema G. et al., 2013, Exp. Astron., 36, 235

Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640

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

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

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

Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127

Morales M. F., Bowman J. D., Hewitt J. N., 2006, ApJ, 648, 767

Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137

Nunhokee C. D. et al., 2017, ApJ, 848, 47

Parsons A. R., 2017, HERA Memo 27, Power Spectrum Normalizations for HERA. University of California, Berkeley,

Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012a, ApJ, 753, 81

Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012b, ApJ, 756, 165

Parsons A. R. et al., 2014, ApJ, 788, 106

Patil A. H. et al., 2016, MNRAS, 463, 4317

Patil A. H. et al., 2017, ApJ, 838, 65

Patra N., Subrahmanyan R., Sethi S., Udaya Shankar N., Raghunathan A., 2015, ApJ, 801, 138

Planck Collaboration XIII, 2016, A&A, 594, A13

Pober J. C., 2015, MNRAS, 447, 1705

Pober J. C. et al., 2014, ApJ, 782, 66

Pritchard J. R., Loeb A., 2012, Rep. Prog. Phys., 75, 086901

Rasmussen, Carl Edward, Williams, Christopher K. I., 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning. MIT Press, Cambridge

Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575

Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, MNRAS, 406, 2421

Santos M. G., Silva M. B., Pritchard J. R., Cen R., Cooray A., 2011, A&A, 527, A93

Thompson A. R., Moran J. M., Swenson G. W., 2017, Interferometry and Synthesis in Radio Astronomy, 3rd edn. Springer, Berlin

Thyagarajan N. et al., 2013, ApJ, 776, 6

Thyagarajan N. et al., 2015, ApJ, 804, 14

Thyagarajan N., Carilli C. L., Nikolic B., 2018, Phys. Rev. Lett., 120, 251301

Trott C. M., Wayth R. B., Tingay S. J., 2012, ApJ, 757, 101

Vedantham H. K., Koopmans L. V. E., 2016, MNRAS, 458, 3099

Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176

Wiener N., 1949, Extrapolation and Smoothing of Stationary Time Series: With Engineering Applications. MIT Press, Cambridge

Yatawatta S. et al., 2013, A&A, 550, A136

Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622

Zaroubi S., 2013, The First Galaxies, 396, 45

Zhang Y. G., Liu A., Parsons A. R., 2018, ApJ, 852, 110

[1]*Department of Physics and Astronomy, University of Western Cape, Cape Town 7535, South Africa*

[2]*The South African Radio Astronomy Observatory (SARAO), 2 Fir Street, Black River Park, Observatory, Cape Town 7925, South Africa*

[3]*Department of Physics, Banwarilal Bhalotia College, GT Rd, Ushagram, Asansol, West Bengal 713303, India*

[4]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*

[5]*LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, F-75014 Paris, France*

[6]*INAF - IRA, via P. Gobetti 101, I-40129 Bologna, Italy*

[7]*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*

[8]*Department of Astronomy, University of California, Berkeley, CA 94720, USA*

[9]*National Radio Astronomy Observatory, Socorro, NM 87801, USA*

[10]*Cavendish Astrophysics, University of Cambridge, CB3 0HE Cambridge, UK*

[11]*Department of Mathematical Sciences, Computer Science Division, Stellenbosch University, Private Bag X1, 7602 Matieland, South Africa*

[12]*School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA*

[13]*Department of Physics and McGill Space Institute, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada*

[14]*Department of Physics, University of Washington, Seattle, WA 98105, USA*

[15]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*

[16]*eScience Institute, University of Washington, Seattle, WA 98195, USA*

[17]*Department of Physics and Astronomy, Center for Particle Cosmology, University of Pennsylvania, Philadelphia, PA 19104, USA*

[18]*National Radio Astronomy Observatory, Charlottesville, VA 22903, USA*

[19]*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA*

[20]*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*

[21]*California State University of Los Angeles, 5151 State University Dr, Los Angeles, CA 90032, USA*

[22]*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*

[23]*ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), University of Melbourne, VIC 3010, Australia*

[24]*Department of Astronomy, San Diego State University, San Diego, CA 92182, USA*

[25]*Scuola Normale Superiore, I-56126 Pisa, PI, Italy*

[26]*Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA 02138, USA*

[27]*American Astronomical Society, Washington, DC 20006, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.