HERA117
Received: 21 Oct 2022

Wavelets and Multiresolution Analysis in Radio Interferometry

Matyas Molnar and Bojan Nikolic

Astrophysics Group, Cavendish Laboratory, University of Cambridge

20th October 2022

Abstract

Wavelet transforms are an efficient and flexible way of visualising and interpreting the spectrum of a signal that is changing over time (or in our case frequency). Such changes are present both in the cosmological H I signal (due to light-cone effects) and in the noise (due e.g. to radio-frequency interference (RFI)). In this memorandum, we briefly review the theory of wavelet transforms and apply them to HERA data as a tool of visualising and identifying RFI.

1 Introduction

The Fourier transform (FT) (and hence power spectrum (PS)) works very well in transforming a signal from its time domain to its frequency domain when the frequency spectrum is stationary and does not evolve in time. However, in the case of a dynamic signal (whose spectral features are not constant with time), the performance of the FT will be degraded; this is the case for most of the signals seen in nature. A non-stationary alternative to the FT is, therefore, required for the spectral decomposition of dynamic signals. In particular, for 21 cm cosmology, we compute PS over considerable frequency bandwidths: the Universe can change over such periods, known as the light-cone effect (Barkana & Loeb, 2006), so a multiscale analysis may be able to better constrain the statistics of the observed era. Wavelet transforms (WTs) offer just this capability by being able to localize the 21 cm signal in both frequency and delay, and the corresponding power provides an unbiased estimator that can fully characterize the 21 cm in all spatial dimensions (since we now include the line-of-sight axis). We introduce wavelets and their transforms in §3.

Wavelets can also be used as both an additional diagnostic and outlier detection tool. If erroneous modes exist in a given signal, it is beneficial to simultaneously locate the fault to a point in the signal and Fourier domain, which is not possible with standard spectral decomposition. We present an error detection technique based on wavelets in §4 and identify contaminated times and baselines that were used in the final H1C_IDR2.2 PS results presented in The HERA Collaboration et al. (2022).

Power spectral estimates with wavelets are also computed for the 14 m and 29 m east-west (EW) baselines in §5 as a new potential method to evaluate spectral power that is not yet established in 21 cm cosmology.

We note that throughout this memorandum, wavelet and Fourier theory are introduced by considering a time signal whose frequency dual is in the frequency domain. These methods are, however, applied to radio interferometric visibilities that are functions of frequency and that are transformed into the delay domain.

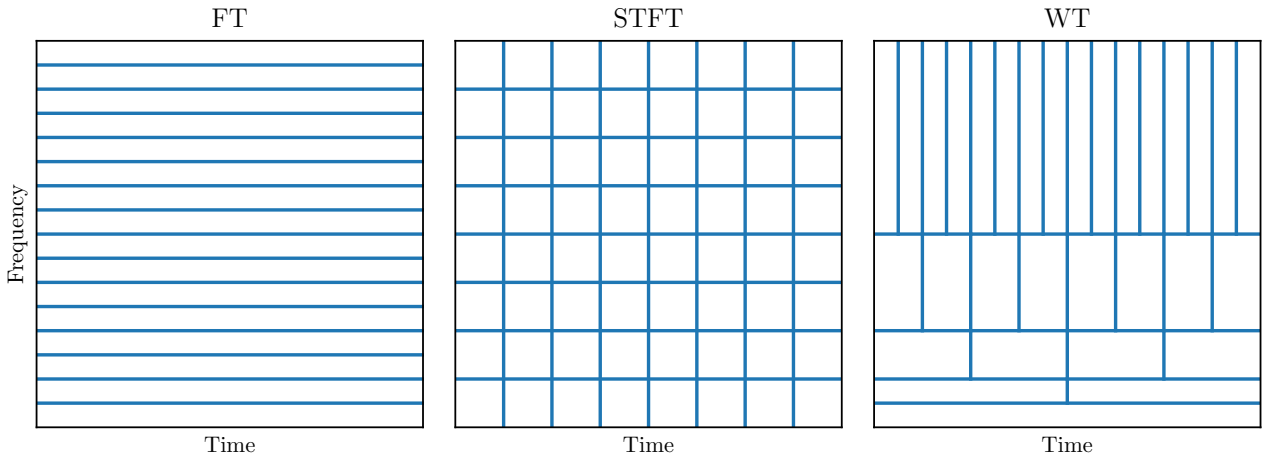


Figure 1: Time-frequency tiling illustration for the FT, STFT and WT. In signal (time) space, the tiling would be purely vertical with full temporal resolution and no spectral information.

2 Short-time Fourier transform

A natural and commonly used method in time-frequency analysis is to divide the full time signal into shorter segments of equal length and to apply the FT locally on those segments. This operation is called the short-time Fourier transform (STFT), and is done by multiplying the signal $x(t)$ with a window function w that is nonzero for the concerned period before transforming to frequency space. The continuous formulation is given by

$$X(\tau, f) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-2\pi i f t} \, \mathrm{d}t \tag{1}$$

with the discrete case by

$$X_{m,k} = \sum_{n=0}^{N-1} x_n w_{n-m} e^{-2\pi i \frac{kn}{N}} \tag{2}$$

We look at the squared magnitude of the STFT $|X|^2$ to obtain the localized PS of the signal, which we can plot for different time localizations to get a spectrogram.

A caveat of the STFT is that it has fixed resolution, as determined by the width of the windowing function, with narrow windows giving good temporal resolution but poor spectral resolution and vice versa. In general, we wish to have good temporal resolution at high frequencies while also having good spectral resolution at low frequencies; this is where the WT comes in.

3 Wavelet transform

The WT builds on the STFT by having increased temporal resolution at higher frequencies at the sacrifice of spectral information, with the opposite occurring at lower frequencies. We, thus, sample the dual time-frequency space at different resolutions to get a more complete and often more helpful representation of the signal. This principle is illustrated in Fig. 1, where the time and frequency resolutions for the different transforms introduced in this memorandum are shown.

Wavelets are a powerful tool with many applications and have been used with great success in other fields of astrophysics. Their usage in 21 cm cosmology for Epoch of Reionization (EoR) detection has, however, been limited. Trott (2016) studied wavelets and their potential in detecting the 21 cm signal, and showed that this novel approach has improved estimation performance compared to the standard Fourier PS that is commonplace in the literature. As EoR experiments advance by growing and improving their sensitivity and resolution, wavelets and their transforms will become increasingly

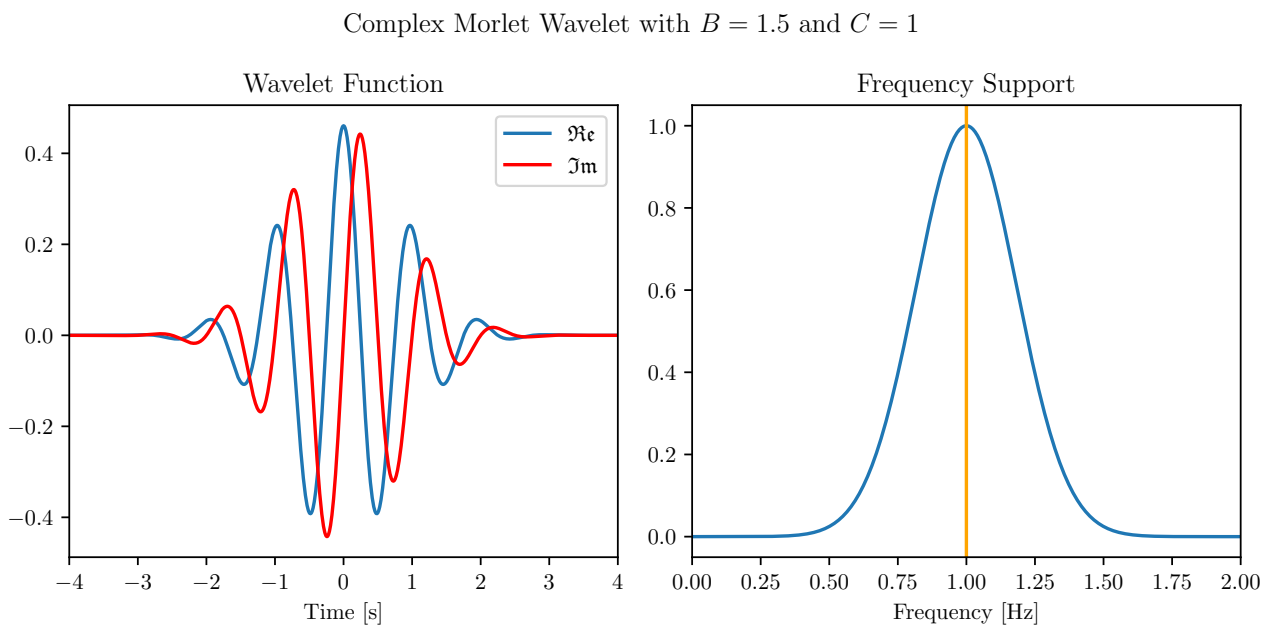


Figure 2: Complex Morlet wavelet $\psi_{\mathrm{cM}}(t; B = 1.5, C = 1)$ time function (left) and frequency support (i.e. FT, right). Only the amplitude of the FT is shown; the $\mathfrak{Re}$ and $\mathfrak{Im}$ components in frequency look much like the wavelet function on the left.

important since they are able to retain location (frequency/redshift) information. Wavelets will likely become a key tool to statistically probe the 21 cm signal.

The basic principles of wavelet analysis are described in the rest of this section. We refer to Mallat (2009); Debnath & Shah (2015) for comprehensive overviews of wavelets and their properties.

We use the continuous wavelet transform (CWT) as the tool to represent the signal; it is defined as

$$\gamma(\tau, a) = \langle x(t), \psi_{\tau,a}(t) \rangle = \int_{-\infty}^{\infty} x(t) \psi^*_{\tau,a}(t)\, \mathrm{d}t \tag{3}$$

where $\psi_{\tau,a}$ are a set of basis functions called wavelets, which are generated from a single wavelet ψ called the *mother* wavelet:

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \tag{4}$$

where a is the scale factor and τ is the translation factor.

A commonly used wavelet is the complex Morlet (also called Gabor) wavelet, which is given by

$$\psi_{\mathrm{cM}}(t; B, C) = \frac{1}{\sqrt{\pi B}} e^{-\frac{t^2}{B}} e^{i2\pi C t} \tag{5}$$

where B is the bandwidth and C is the centre frequency. The former, B, is the inverse of the variance of the Morlet wavelet in the frequency domain; therefore, increasing B reduces the variance in frequency, resulting in a higher energy concentration around the central frequency. The Morlet wavelet is composed of the product of a complex carrier (i.e. the standard Fourier mode) with a Gaussian envelope (with an energy normalization factor in front). See Fig. 2 for plots of the Morlet wavelet function and its corresponding FT for $B = 1.5$ and $C = 1$. The CWT with $\psi_{\mathrm{cM}}(t; B = 1, C = 1)$ is a minor modification of the STFT with Gaussian window, and is also known as a Gabor transform (Gabor, 1946).

In Fig. 3 we also depict how the scaling and translation of wavelets affect their spreads in frequency

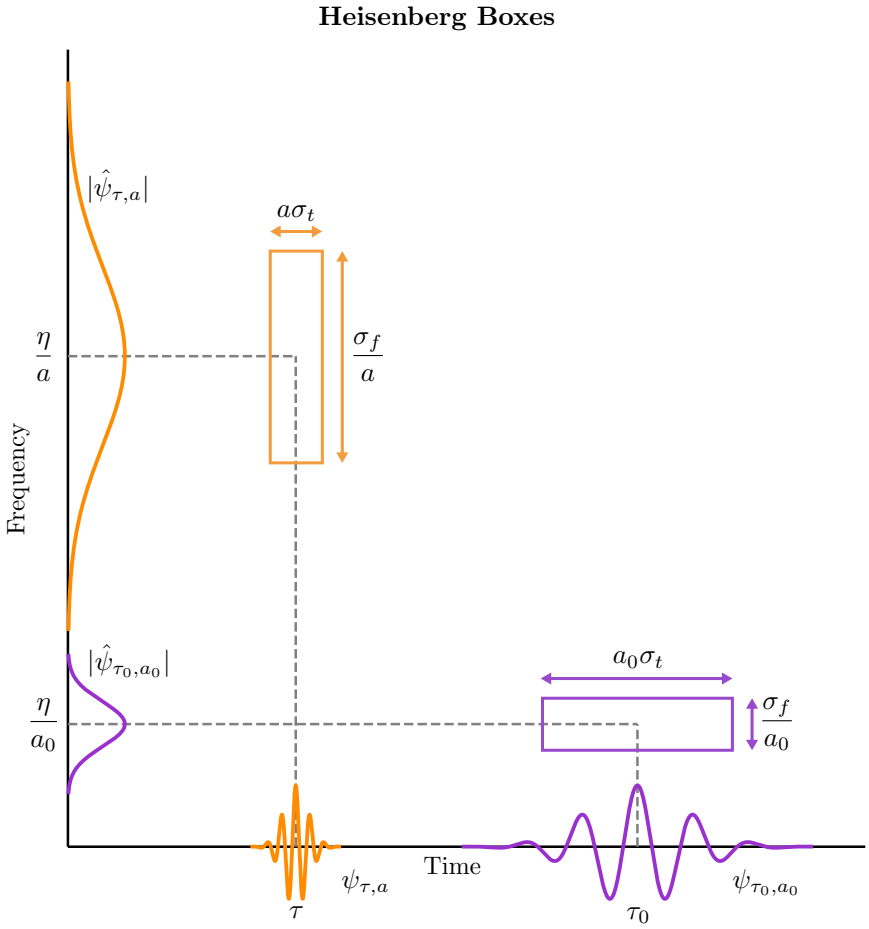


Figure 3: Heisenberg boxes of two Morlet wavelets with different translation and scaling factors, showing the time-frequency spread of the given wavelets. We can see that a scaling a in time corresponds to a dilation in frequency by $1/a$. The wavelet functions and their FTs are normalized to unity amplitude, and the lengths and widths of the boxes are not to scale with the wavelet functions and their FTs; this is done for pictorial purposes. As per the uncertainty principle, the minimum area of such a Heisenberg box is $1/4\pi$, which is achieved with Morlet wavelets (see Sections C and D).

and time.

The CWT for a finite sequence of equally spaced samples x_n (not to be confused with the discrete wavelet transform (DWT), see §E) is given by

$$\gamma_{m,a} = \sum_{n=0}^{N-1} x_n \psi^*_{\{m,a\},n} \tag{6}$$

with Eqs. (4) and (5) similarly discretized.

There are a plethora of wavelets to choose from, each with its own characteristics and benefits. In general, it is advantageous to select a wavelet that is similar to the signal under investigation. The Morlet wavelet $\psi_{\rm cM}(t; B = 1.5, C = 1)$ is a good all-purpose wavelet with similar variance in both time and frequency (slightly better frequency resolution); we use this wavelet for the work in the rest of this section. See §D for further information on how Morlet wavelets are constructed. A wavelet with similar characteristics to the signal under investigation will improve results. We defer the choice of wavelet to study the 21 cm cosmological signal to future work.

We note that the continuous wavelets used for CWT analysis do not form an orthogonal basis, and the wavelet coefficients $\gamma_{m,a}$ are highly redundant. Thus, this representation of the signal is inefficient; it

is, however, more than adequate if we wish to analyse and better understand the spectral structure of a signal, especially through visual means. For optimal signal decomposition and eventual reconstruction, discrete wavelets are used.

We call the square of the modulus of the CWT the wavelet power spectrum (WPS):

$$\mathrm{WPS}_{m,a} = |\gamma_{m,a}|^2 \tag{7}$$

which is analogous to the standard PS, but has the additional dimension indexed by m (or τ for the continuous case) that locates the spectral power in the signal domain. We predominantly use this measure to spectrally describe the signal using wavelets.

Let us not forget that in obtaining a time-frequency wavelet representation of a signal, the temporal and spectral information gained are constrained by the uncertainty principle (see §C).

When dealing with interferometric visibilities, the signal lives in the frequency domain and its Fourier dual in the delay domain. The CWT and WPS therefore, have units [Jy Hz] and [$\mathrm{Jy}^2\,\mathrm{Hz}^2$] (like the FT and PS), unless stated otherwise. We use the PyWavelets[1] (Lee et al., 2019) package for WT computations.

In Fig. 4, we show the WPS for sample Hydrogen Epoch of Reionization Array (HERA) pseudo-Stokes I visibilities[2] on a frequency-delay heatmap, which we call a *scaleogram*. The visibility signal and its corresponding FT are also plotted on the sides of the scaleogram.

4 Error detection

We look at the final H1C_IDR2.2 pI visibilities, which are the end products of the the analysis and preprocessing pipelines. As a first pass, scaleogram cubes with dimensions (scales, frequencies, times, baselines) were created. The napari[3] (Sofroniew et al., 2022) multidimensional image viewing software was used to browse through the scaleograms for each time and baseline combination, for both Bands 1 and 2. While this is an intensive exercise, it provides a good feel for the data. Visual inspection of large amounts of data is laborious, meaning that anomalous modes can be missed. Furthermore, logarithmic scales can be deceiving; a more rigorous approach is needed.

We develop detection process whereby the modified-Z score is computed for each scaleogram, where the location and scales in the computation of Z^{mod} are the median and the median absolute deviation (MAD) across times and baselines, both of which are shown in Fig. 5 for Band 2. CWT coefficients that have support outside of the CoI are also masked to minimize false detections; practically, we mask coefficients with $\sigma^{\mathrm{mad}}_{\mathrm{B1}} > 2 \times 10^{1}$ and $\sigma^{\mathrm{mad}}_{\mathrm{B2}} > 7 \times 10^{-2}$ for Bands 1 and 2 (determined by essay), respectively, as the edge-effect artefacts leak further into the data than the estimated $\sqrt{2}s$ CoI.

The effect of flagging channels 570 – 573 (within Band 2) can be clearly seen in Fig. 5: before the preprocessing pipelines, certain channel ranges are flagged for exhibiting narrowband features in the inspection of residual visibilities after wideband delay CLEAN. These flagged channels are then inpainted with a CLEAN model, which will be smooth. It is, therefore, expected to see lower power at high delays (low scales) for those channels. Interestingly, there is slightly elevated power for those channels from scales ~ 7, which corresponds to the 2 µs delay threshold used for the CLEAN based inpainting. By re-running the preprocessing pipeline, we find that the final spherically averaged PS

[1] https://github.com/PyWavelets/pywt

[2] Defined as the linear combination of the instrumental polarization visibilities: $V_{\mathrm{pI}} = (V_{\mathrm{EE}} + V_{\mathrm{NN}})/2$ (Hamaker et al., 1996). Such pI visibilities are used as we expect the 21 cm signal to be unpolarized (Moore et al., 2013). Here, E and N denote the east and north alignment of the feeds, such that EE instrumental polarization corresponds to an east-facing feed correlated with another east-facing feed (likewise for NN)

[3] https://napari.org/

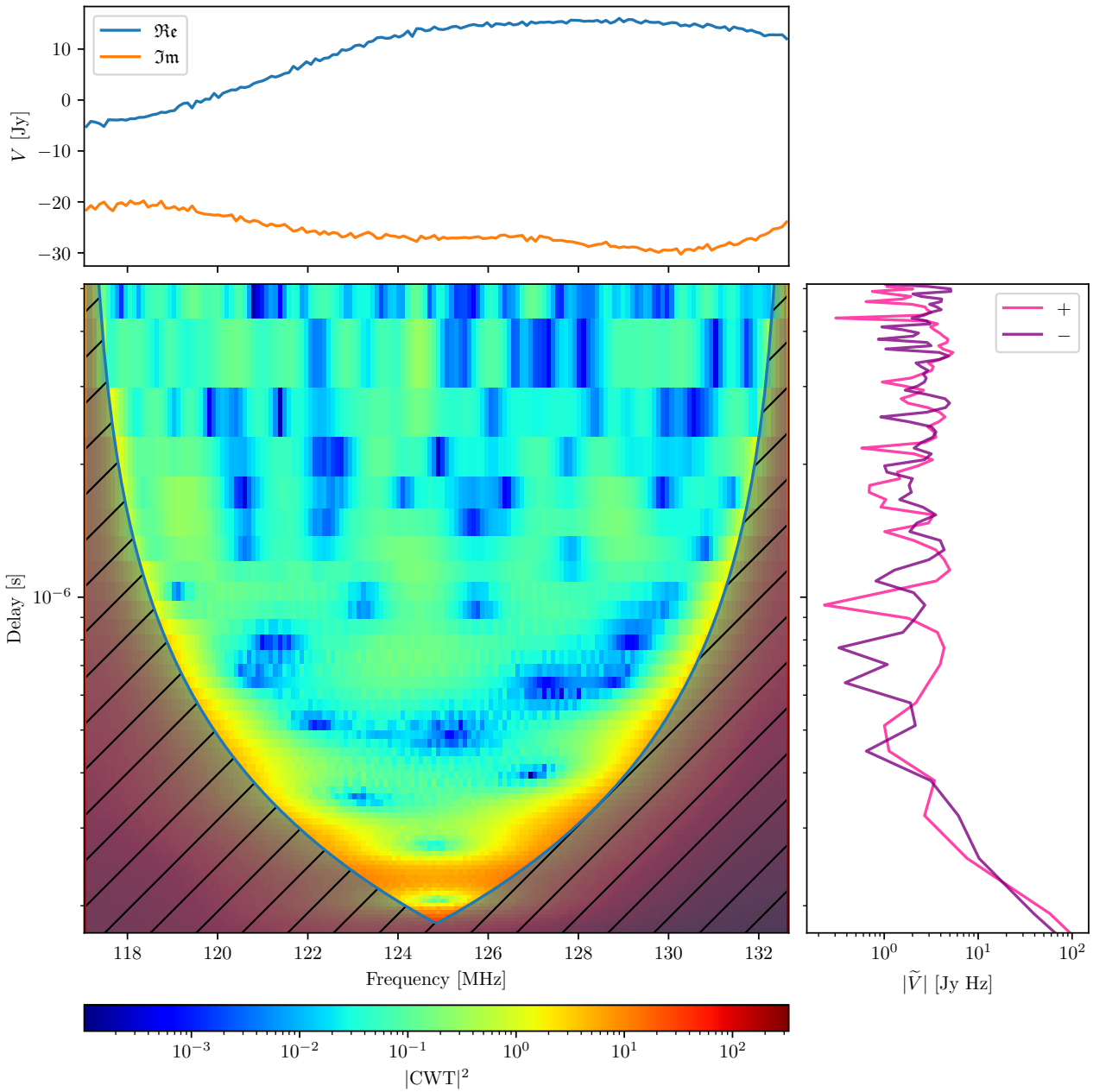


Figure 4: Continuous wavelet transform scaleogram, with complex visibility signal shown in the upper plot and its FT, using a Blackman-Harris window, shown on the right plot. Wavelet scales are from 2 – 59, corresponding to delays of 5.120 – 0.174 µs; the starting wavelet scale of 2 is chosen to match the Nyquist frequency. The cone of influence (CoI) is clearly delineated in the scaleogram, showing the regions of frequency-time space that are corrupted by edge-effect artefacts. In the FT plot, we show $|\widetilde{V}|$ for both positive and negative delays, for completeness. This particular example is for the final visibility of the 14 m (12, 13, pI) EW baseline for a 214 s time bin at Local Sidereal Time (LST) 2.12 h (in Field 1) and for Band 1, after all the reduction steps from the analysis and preprocessing pipelines. This particular data slice is well-behaved, with no anomalies appearing across the three plots.

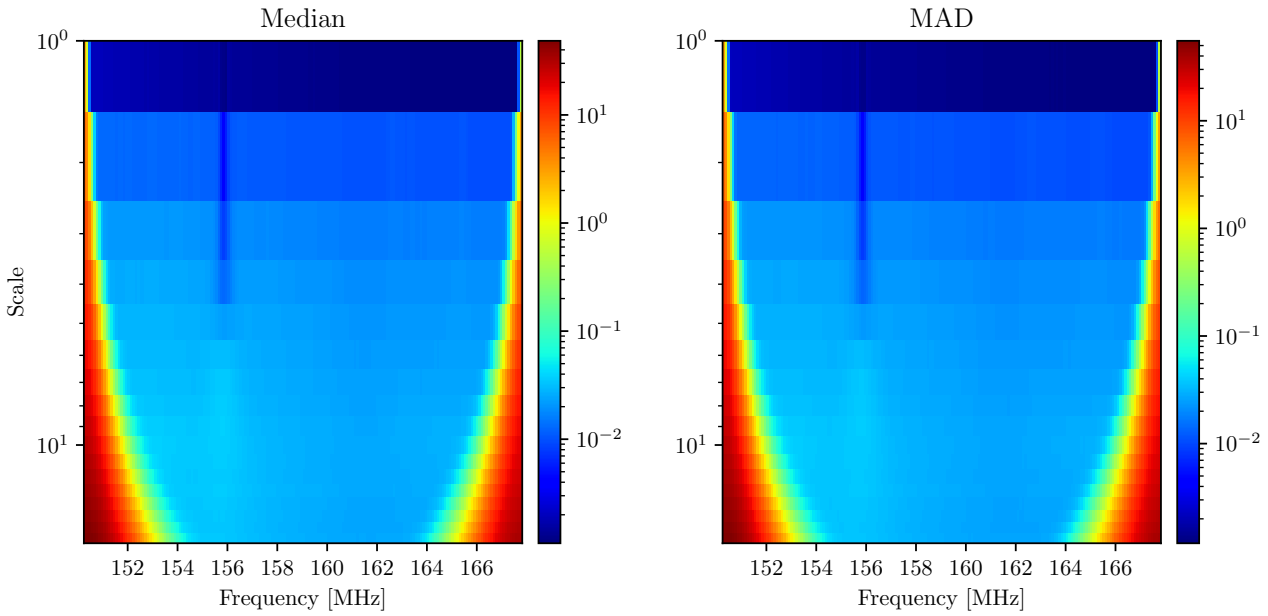


Figure 5: Median and MAD of WPS scaleograms for Band 2, both having units [Jy^2 Hz^2]. The additional flagged channels in preprocessing clearly leave an imprint around frequency 155.8 MHz.

results are largely unchanged by the flagging of these channels.

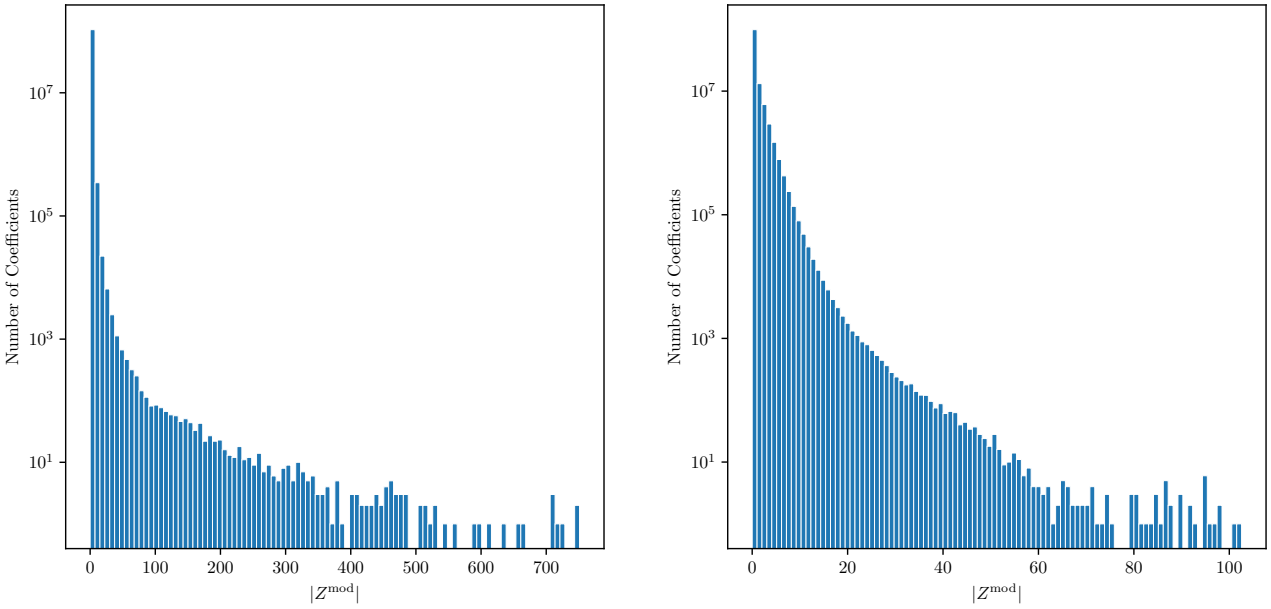
Histograms of the number of power coefficients in the scaleograms as a function of $|Z^{\text{mod}}|$ following the outlier detection procedure are shown in Fig. 6 for both bands. A fair number of outliers are detected from this method, with a non-negligible amount attaining extremely high $Z^{\text{mod}}|$ scores; these anomalies, if not removed, can corrupt final PS estimates when the data is further averaged.

As an example of a particularly bad data slice, we look at baseline (66, 85, pI) at LST 8.52 h that has an isolated region of higher power around frequency 154 MHz and delay ≈ 1 µs. We show its scaleogram with signal and FT appended to the relevant axes in Fig. 7, and further examine it in comparison to its redundant group in Fig. 8. The region of high power is a significant outlier with peak $|Z^{\text{mod}}| = 103$. It seems to indicate a transient event that is not present in adjacent time integrations or other baselines. The anomaly is not as appreciable when solely looking at the signal and Fourier domains, but clearly stands out by looking at the scaleograms and $|Z^{\text{mod}}|$ heatmaps.

We further reduce the outlier detection data to make it more manageable and to potentially identify patterns. The maximum $|Z^{\text{mod}}|$ is taken for each scaleogram and is plotted as a function of time and baseline in Fig. 9 for all bands and fields. From this visualization, it is clear that certain time bins and baseline groups suffer more from these transient effects in their WPS: e.g. the bottom time bin of Field 1 (more prominent in Band 1) and top bin of Field 3 have noticeably more anomalies than the other bins. The short baseline groups, particularly those with length ≈ 25 m (corresponding to baseline indices 81 – 158), also seem to have more outliers, as seen from the brighter rectangular regions at the lower end of the baseline indices.

We repeat this analysis but look at the maximum $|Z^{\text{mod}}|$ per antenna, with results plotted in Fig. 10. The findings are similar to Fig. 9, with the last time bin of Band 1 Field 1 and the first time bin of Band 2 Field 3 showing the worst outliers. There is no antenna that is evidently faulty, although it seems that antennas closer to the centre of the array configuration seem to have a higher prevalence of strong outlier events.

As a further illustration of the affected baselines, we plot baselines that have outliers in their scaleograms $|Z^{\text{mod}}| > 5$ for Band 1 in Fig. 11, with the methodology for drawing the baselines explained in



(a) Band 1 **(b)** Band 2

Figure 6: Number of points in scaleograms binned by $|Z^{\mathrm{mod}}|$ score. 1.39% of Band 1 and 1.56% of Band 2 power coefficients exceed $|Z^{\mathrm{mod}}| = 5$. Band 1 also has more extreme outliers than Band 2. Note that each count corresponds to a point on a scale-frequency scaleogram; therefore, if there is a region of higher power (consisting of many points) for a given scaleogram, that scaleogram will be counted multiple times.

the caption. From such plots, baselines consisting of antennas closer to the array's centre can be seen to have a higher proportion of severe outliers.

From all these considerations, we infer that Band 1 as a whole is of worse quality than Band 2, with many more extreme outliers seen in the baseline and antenna heatmaps in Figs. 9 and 10. These could be the source of the elevated power at lower k modes seen in The HERA Collaboration et al. (2022) that does not agree with the theoretical noise floor. Moreover, the transient anomalous effects should be considered as per-baseline issues, and any additional flagging should be as such; there is no clear evidence to discard any antennas. Shorter baselines also seem to be more affected, and some time bins at the edges of the windows of observations for the fields also appear more problematic.

5 21 cm wavelet power spectrum

To demonstrate the potential of wavelets and their WPS as combined spatial and spectral estimator of the 21 cm signal, we form cross wavelet power spectrum (CWPS) in a similar fashion to the HERA PS analysis:

$$\mathrm{CWPS}_{m,a} = \gamma^{i}_{m,a}{}^{*}\gamma^{j}_{m,a} \tag{8}$$

where CWT coefficient are cross-multiplied, with i, j the indices of either baselines or time bins, with $i \neq j$.

CWT scaleograms are formed for all baselines and times. Each scaleogram slice, containing the CWT coefficients for different scales and frequencies, is cross-multiplied with the one adjacent in time (separated by 214 s), with the cross-multiplication being done with all baselines in its redundant set (that share the same length and orientation). Considering all baseline permutations in the Fourier domain rather than redundant averaging in the visibility domain reduces the impact of baseline-dependent systematics. Cross-multiplying adjacent time bins further reduces any noise bias. Next, all such cross-power scaleograms are incoherently averaged across redundant baseline sets and across the time axis.

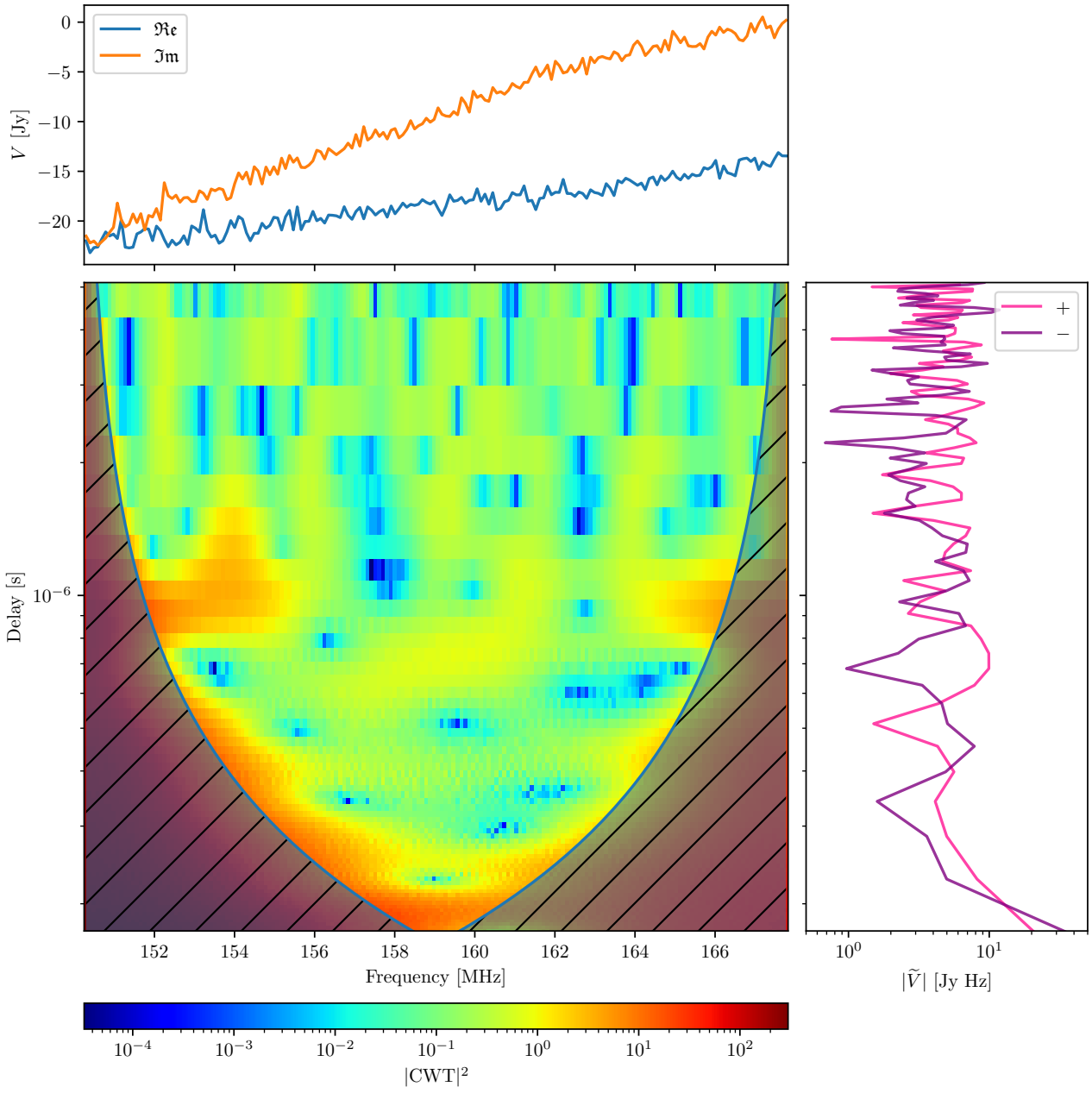


Figure 7: Visibility, FT and scaleogram for visibility with baseline (66, 85, pI) for a 214 s time bin at LST 8.52 h (in Field 3) and for Band 2. Wavelet scales are from 2 – 59, corresponding to delays of 5.120 – 0.174 µs, same as Fig. 4. We see a clear triangular shape region of higher power at frequencies ≈ 154 MHz and delays ≈ 1 µs, which is not evident in the visibility and FT plot, despite corresponding to a peak $|Z^{\mathrm{mod}}| = 103$ (as shown in Fig. 8).

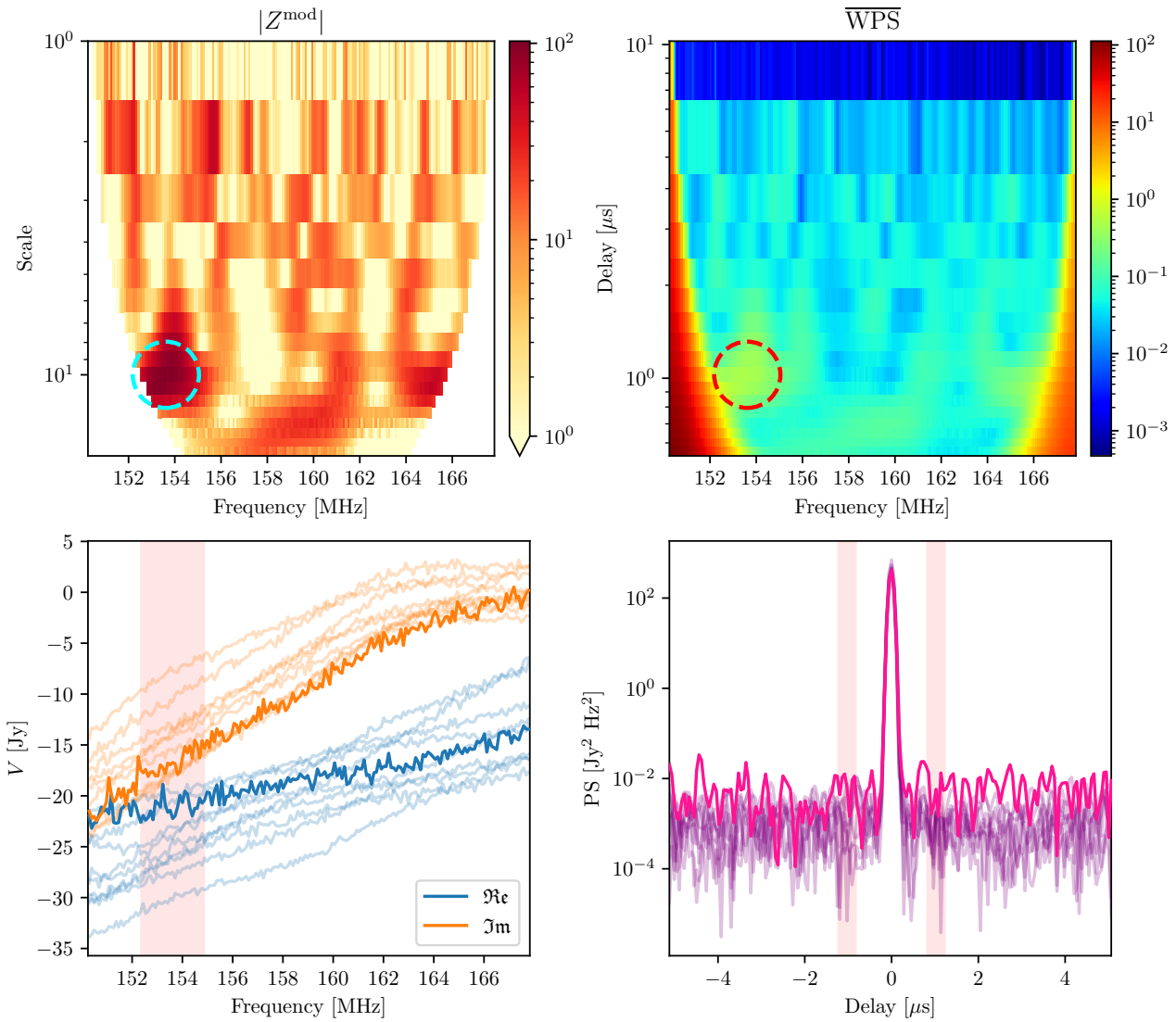
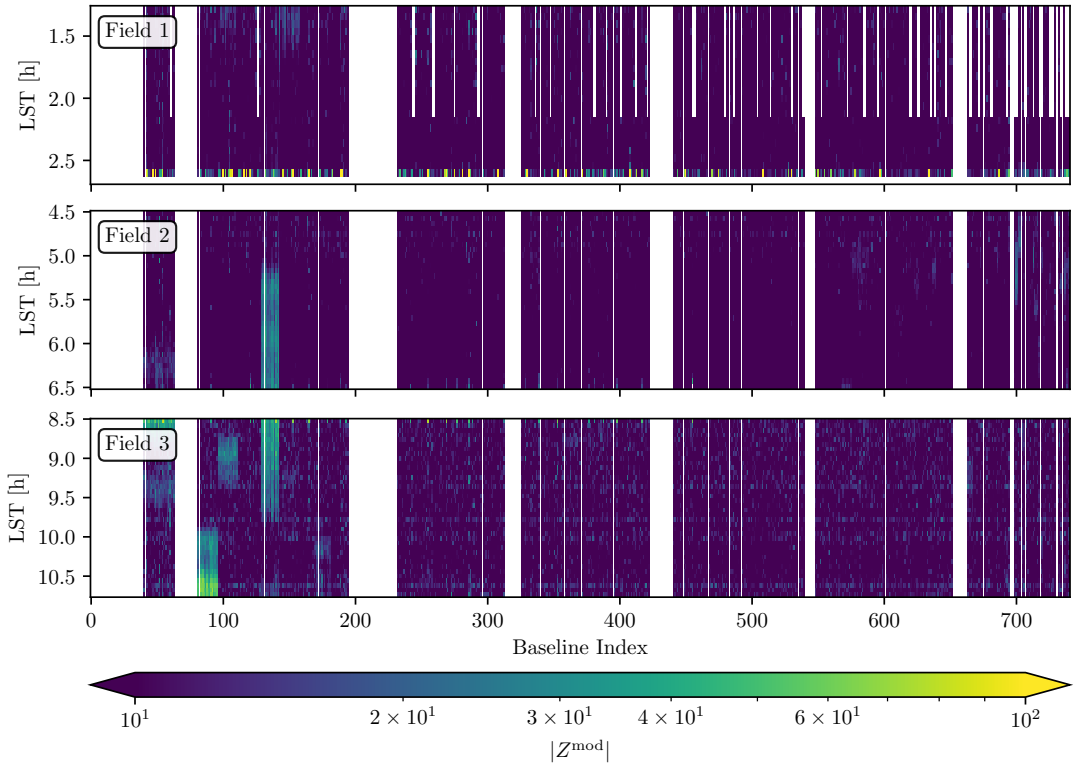


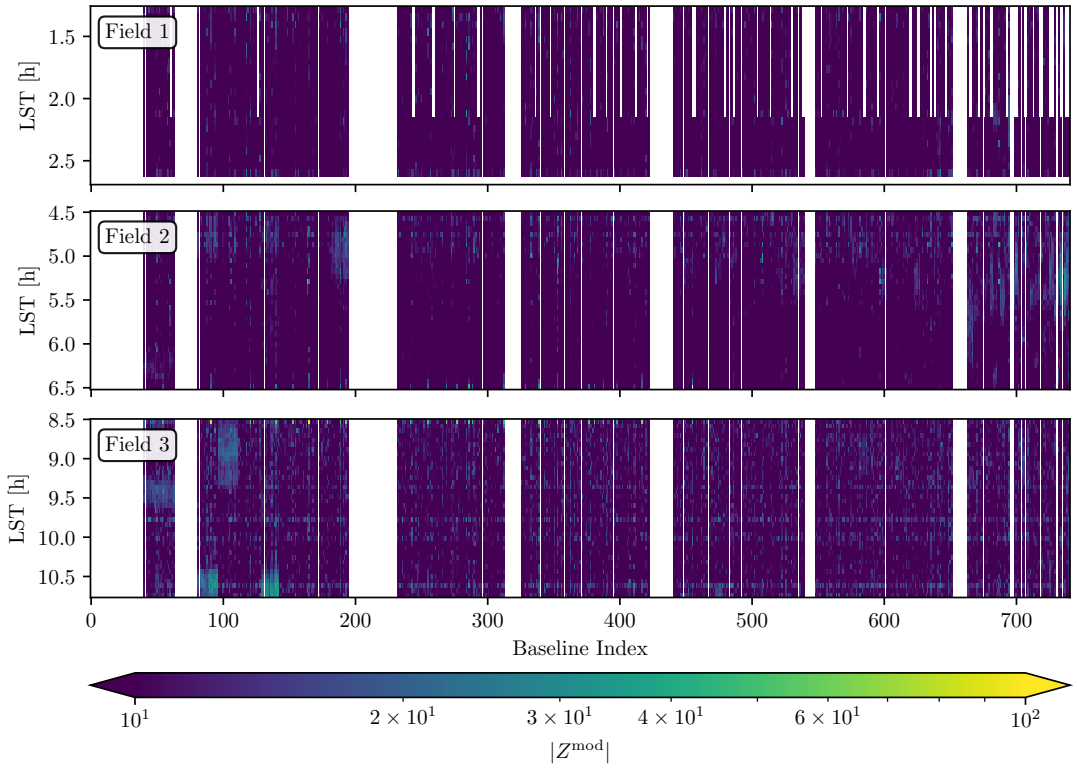
Figure 8: Top row: $|Z^{\mathrm{mod}}|$ for the scaleogram shown in Fig. 7 (left) and the mean WPS for the redundant group (right). Bottom row: plot of the complex visibilities (left) and their PS (right) for the baselines in the redundant group. The visibility and PS for the faulty baseline are highlighted. Areas of interest in frequency and delay space are also shaded in red; it is not entirely evident that these are areas of concern if purely looking at the visibility and PS plots.

We show the final CWPS scaleograms for 14 m and 29 m EW baselines in Figs. 12 and 13, respectively, for all bands and fields. These final scaleograms all show a horizontal feature at delays $\approx 0.4\,\mu$s (or, in $k_\parallel$ units, $0.20\,h\,\mathrm{Mpc}^{-1}$ at $z = 10.4$, and $0.22\,h\,\mathrm{Mpc}^{-1}$ at $z = 7.9$), which is close to the HERA H1C_IDR2.2 upper limits set in The HERA Collaboration et al. (2022) and could be cause for concern. The computation time for the calculations to obtain the shown CWPS for all bands and fields from the fully reduced and calibrated pI (pseudo-I) visibilities is $\sim O(10\,\mathrm{min})$.

A standardized way of presenting WPS results in 21 cm cosmology is needed. Converting the scaleograms in Figs. 12 and 13 to cosmological units is a two-step process that involves first converting visibilities from Jy to mK and then multiplying the WPS by a cosmological scalar quantity to finally get units $\mathrm{mK}^2\,\mathrm{h}^{-3}\,\mathrm{Mpc}^3$. Such a process, however, leads to added complexity, as $k_\parallel = k_\parallel(\eta, z)$ (where η is delay, so some regridding would be required since each vertical slice in the scaleograms of Figs. 12 and 13 would be slightly shifted. In addition, WPS from different baselines must be combined appropriately. Moreover, the scaleograms here are created using the CWT, which takes the product of the signal and wavelet at each frequency and for all specified wavelet scales. These scaled functions that are continuously shifted over the signal do not form an orthogonal basis; therefore, the resulting

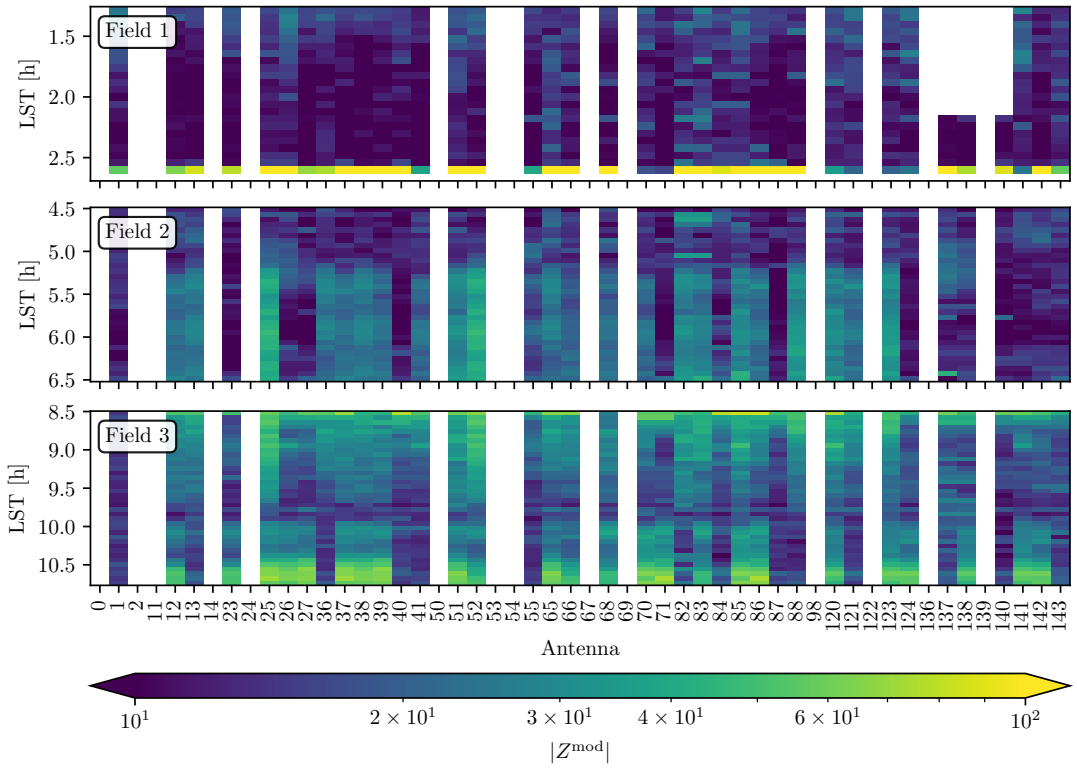


(a) Band 1

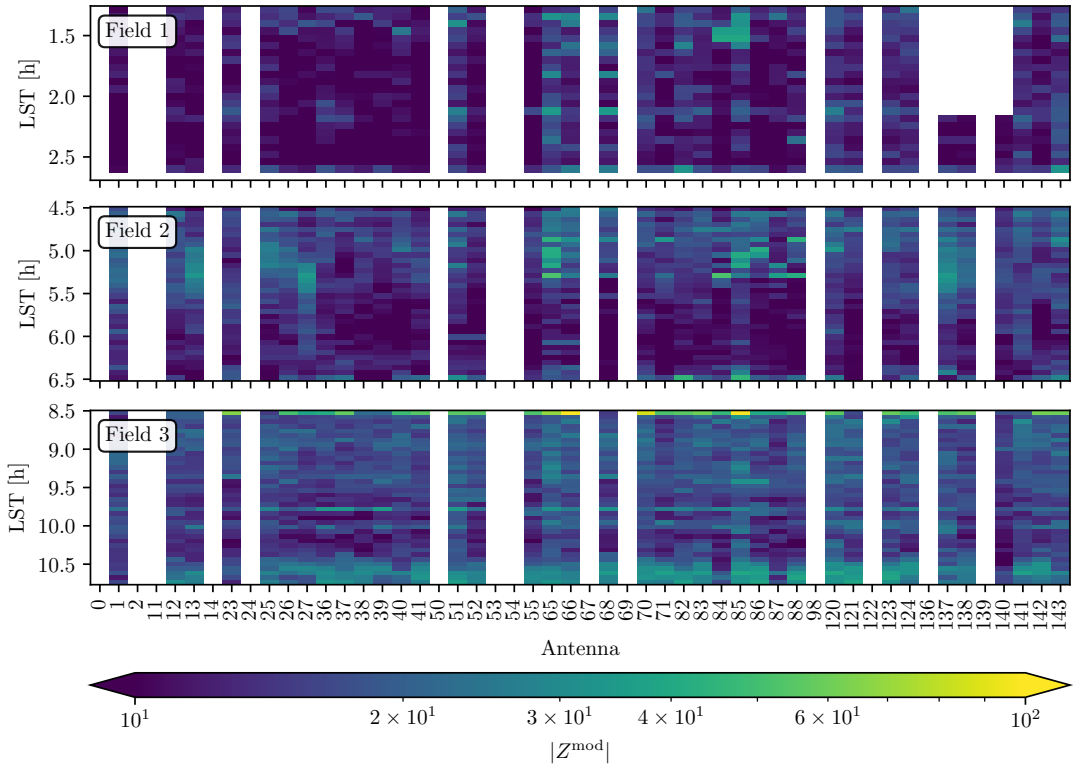


(b) Band 2

Figure 9: Maximum $|Z^{\mathrm{mod}}|$ for WPS scaleograms by baseline. Baseline length increases with baseline index. Only baselines taken forward for PS computation in the standard HERA pipeline are shown. The bottom time bin of Band 1 Field 1 and the top time bin of Field 3 (both bands) contain many extreme outliers. Some of the shorter baselines are also afflicted by strong outliers, indicated by the brighter regions in both bands.



(a) Band 1



(b) Band 2

Figure 10: Maximum $|Z^{\mathrm{mod}}|$ for WPS scaleograms by antenna. As with Fig. 9, the biggest outliers are contained in the bottom time bin of Band 1 Field 1 and the top time bin of Field 3 (both bands). No antenna is clearly faulty from this metric.

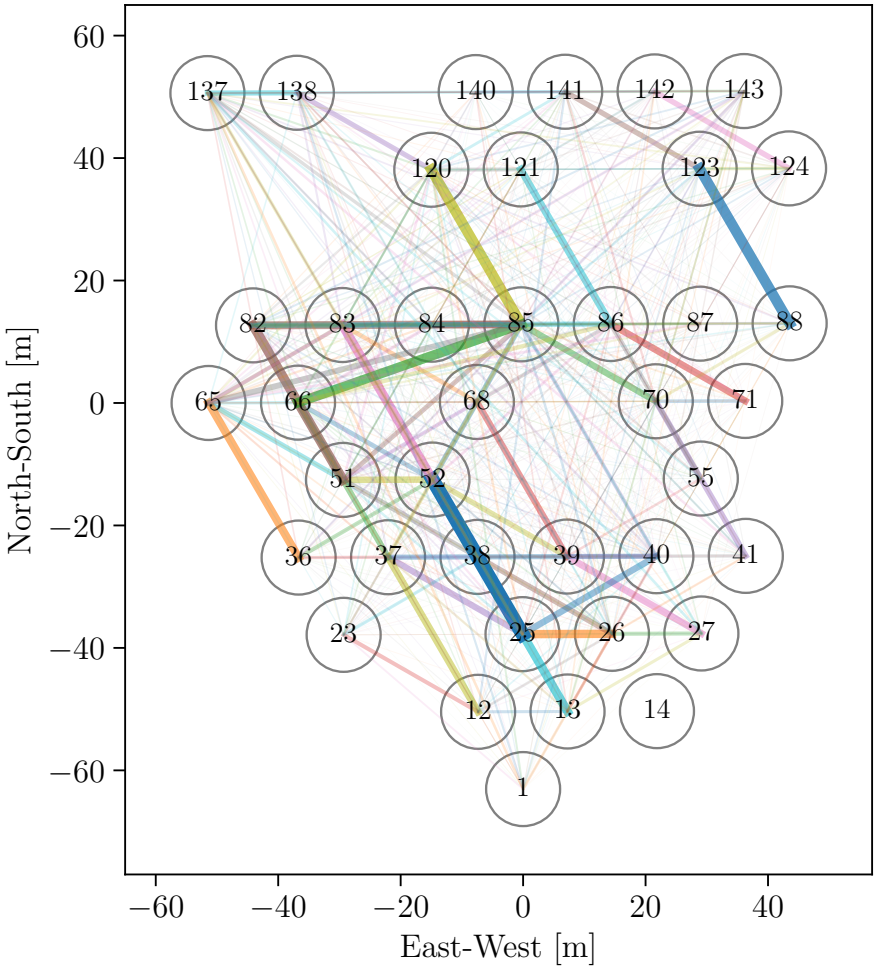


Figure 11: Array layout of unflagged antennas with baselines drawn according to their $|Z^{\mathrm{mod}}|$ score for Band 1. First, all baselines/time slices with (at least) a $|Z^{\mathrm{mod}}| > 5$ data point in its scaleogram are found (there are 39,966), and the maximum $|Z^{\mathrm{mod}}|$ is taken for each scaleogram. We then average the $|Z^{\mathrm{mod}}|$ score across the time dimension, subtract the minimum $|Z^{\mathrm{mod}}|$ score from all points and then divide by the maximum $|Z^{\mathrm{mod}}|$ score to get a weight for each baseline. This weight is then used for the thickness and transparency parameters in drawing these baselines. The thicker and more opaque baselines are those with, on average, more serious outliers in their scaleograms. Baselines with central antennas appear to be more affected. This plot is merely illustrative, as there are many ways to fold the data and adjust the plotting parameters. Even so, from looking at various metrics, antennas 85, 86, 52, 40, 25 and 66 recur as being involved in baselines that suffer from the worst outliers, although no statistic is significant enough to justify discarding them.

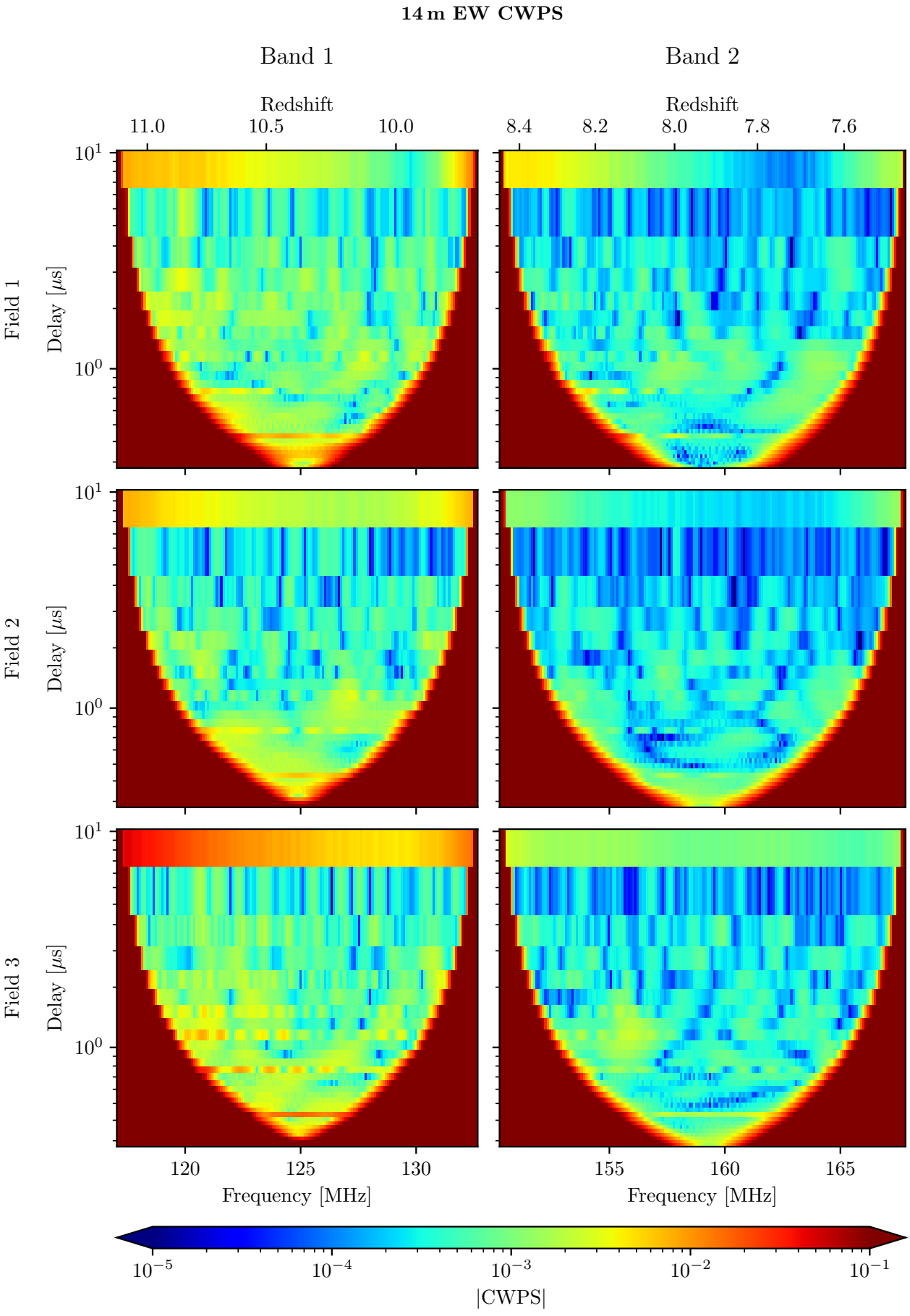


Figure 12: CWPS for 14 m EW baselines for all bands and fields of the H1C_IDR2.2 analysis.

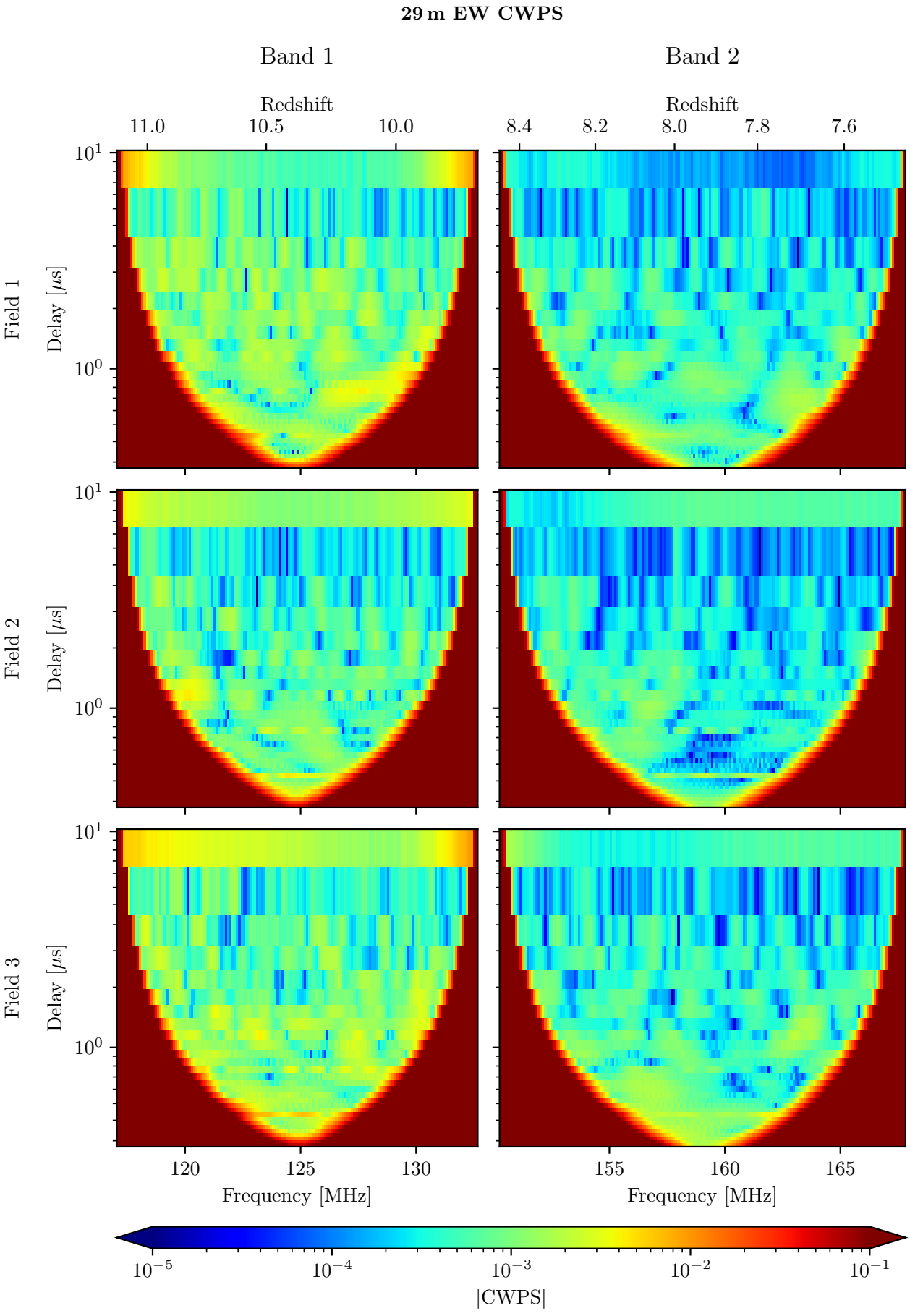


Figure 13: CWPS for 29 m EW baselines for all bands and fields of the H1C_IDR2.2 analysis.

coefficients form a highly redundant representation of the signal. Wavelets can be discretized such that they are scaled and translated in discrete steps, with the modified transform operation called the DWT, which provides an optimal representation of the signal. The CWT remains useful, though, as this oversampling provides a better way of picturing the signal. See §E for a brief rundown of the DWT.

6 Discussion

Wavelets and their transform provide a method of spectral decomposition of a signal while also retaining spatial locality. Such multiresolution analysis is useful when dealing with a dynamic signal and has been used extensively in other branches of physics. The 21 cm signal is expected to evolve over redshift (light-cone effect), so any PS estimate over a considerable bandwidth will be biased, which in turn reduces the detectability of the signal. Wavelets may be capable of providing an unbiased estimator of spectral power that can fully characterize the 3D nature of the EoR. The efficiency of wavelet transforms also is beneficial as data volumes grow (increased number of antennas and evenings of observation) and frequency resolution increases.

This memorandum outlines the WT theory, and shows proof of concept CWPS results for the 14 m and 29 m EW HERA baselines. A standardized way of presenting WPS scaleograms is needed for consistency between different experiments, which may be challenging due to the increased dimensionality of the result. Simulations with e.g. 21cmFAST (Mesinger et al., 2011; Murray et al., 2020) are also needed to compute fiducial scaleograms for comparison.

The use of WPS can also be extended to improve error detection of fully calibrated and reduced visibilities. By inspecting frequency-delay scaleograms, anomalous regions of higher than average power can be detected. RFI or other transient events would then manifest themselves as pyramids (see e.g. Fig. 7), as the effect would first be noticed at high delays (low scales) and would then be subsequently detected by bigger wavelets (higher scales) at lower delays. The analysis of the data in this dual space also reveals structure that can be missed if looking at one of the dimensions alone.

From visual and Z^{mod} considerations, scaleograms with regions of high power were identified and problematic baseline/time slices catalogued for each band separately. It appears that some time bins at the edges of the fields of observation should have been discarded for PS analysis, as they contain a considerable number of extreme outliers that would corrupt any estimates after time averaging. The shorter 14 – 28 m baselines experience outlier events over particular time windows, with baselines within a redundant group appearing to be affected together. Further looking at outliers on a per-antenna basis does not reveal that any one antenna is prominently faulty, although there are signs that antennas closer to the centre of the array (and that are, thus, surrounded by more neighbouring antennas) are involved with more severe outliers.

Anomaly detection with WPS scaleograms presents itself as a great flagging tool that can spot erroneous visibility slices. With further automation, this technique can be used in production and can provide per-baseline flags for PS estimation.

Appendices

Appendix A The Fourier transform

We employ the ordinary unitary form of the FT. The FT of a function $\xi(t)$ is given by $\mathcal{F}(\xi(t)) = \hat{\xi}(f)$, with the forward and inverse transforms defined as

$$\hat{\xi}(f) = \int_{-\infty}^{\infty} \xi(t) e^{-2\pi i f t} \, \mathrm{d}t \tag{9}$$

$$\xi(t) = \int_{-\infty}^{\infty} \hat{\xi}(f) e^{2\pi i f t} \, \mathrm{d}f \tag{10}$$

thus forming the FT pair

$$\xi(t) \overset{\mathcal{F}}{\longleftrightarrow} \hat{\xi}(f) \tag{11}$$

Equivalently, for finite equally-spaced sequences x_n, we compute the discrete Fourier transform (DFT) and its inverse, which are given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i \frac{kn}{N}} \tag{12}$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2\pi i \frac{kn}{N}} \tag{13}$$

We use the fast Fourier transform (FFT) algorithm for all DFT computations for its reduced complexity from $\mathcal{O}(N^2) \to \mathcal{O}(N \log N)$, which was popularized by Cooley & Tukey (1965) but first discovered by Gauss in 1805 and published posthumously in Gauss (1866).

Appendix B The power spectrum

The distribution of power of a time signal by frequency component is given by its PS. The energy of a signal is equal in both the time and frequency domain by the Plancherel identity

$$\|\xi\|^2 = \int_{-\infty}^{\infty} |\xi(t)|^2 \, \mathrm{d}t = \int_{-\infty}^{\infty} |\hat{\xi}(f)|^2 \, \mathrm{d}f \tag{14}$$

We, thus, define the PS as

$$S(f) = |\hat{\xi}(f)|^2 = \hat{\xi}^*(f)\hat{\xi}(f) \tag{15}$$

By the Wiener–Khinchin theorem, the PS is also equal to the FT of the two-point autocorrelation function, such that:

$$S(f) = \int_{-\infty}^{\infty} R_{\xi\xi}(t) e^{-2\pi i f t} \, \mathrm{d}t = \hat{R}_{\xi\xi}(f) \tag{16}$$

where

$$R_{\xi\xi}(t) = \int_{-\infty}^{\infty} \xi(t')\xi^*(t'-t) \, \mathrm{d}t' = \int_{-\infty}^{\infty} \xi^*(t')\xi(t'+t) \, \mathrm{d}t' \tag{17}$$

which is also commonly used to define the PS.

In the discrete case, these definitions become

$$S_k = X_k^* X_k \tag{18}$$

$$= \sum_{n=0}^{N-1} R_{xx,n} e^{-2\pi i \frac{kn}{N}} \tag{19}$$

with

$$R_{xx,n} = \sum_{n'=0}^{N-1} x_{n'} x^*_{n'-n} \tag{20}$$

The approach in Eq. (18) is the most direct and can be efficiently implemented owing to the speed-up from the FFT algorithm.

B.1 The cross power spectrum

For two given signals $\xi(t), \zeta(t)$, we can define the cross power spectrum as

$$C_{\xi\zeta}(f) = \int_{-\infty}^{\infty} R_{\xi\zeta}(t) e^{-2\pi i f t}\, \mathrm{d}t = \hat{R}_{\xi\zeta}(f) \tag{21}$$

where $R_{\xi\zeta}(t)$ is the cross-correlation of $\xi(t)$ with $\zeta(t)$, as defined by

$$R_{\xi\zeta}(t) = \int_{-\infty}^{\infty} \xi^*(t')\zeta(t'+t)\, \mathrm{d}t' \tag{22}$$

$C_{\zeta\xi}(f)$ can be found in a similar fashion by swapping the order of the signals. From these definitions, we find that $C^*_{\xi\zeta}(f) = C_{\zeta\xi}(f)$, and the PS is a special case when $\xi(t) = \zeta(t)$. In Eqs. (15) and (16), the (auto) PS $S(f)$ is implied to mean $S_{\xi\xi}(f)$.

Discrete analogies can also be made for the cross power spectrum (CPS). In practice, we compute the product of the complex conjugate of the FT of one signal with the FT of the other:

$$C_k = X^*_k Y_k \tag{23}$$

In CPS computation, we obtain a complex spectrum; for two coherent signals that aim to measure/detect the same phenomenon (e.g. redundant baselines in radio interferometry seeking to detect the 21 cm signal), we are usually only interested in the $\mathfrak{Re}$ part of the CPS, since that is where the phenomenon of interest lives, with the $\mathfrak{Im}$ part being a realization of the noise.

We refer to the PS and CPS somewhat interchangeably, unless specified otherwise.

Appendix C The uncertainty principle

If we adjust the temporal spread of a function $\xi(t)$ we can contract or dilate the function by a factor $s > 0$ such that

$$\xi_s(t) = \frac{1}{\sqrt{s}} \xi\left(\frac{t}{s}\right) \tag{24}$$

where energy normalization is ensured, such that $\|\xi_s\|_2 = \|\xi\|_2$. The spread of $\xi(t)$ is increased for $s > 1$ and decreased for $s < 1$. Taking the FT of $\xi_s(t)$, we obtain

$$\hat{\xi}_s(f) = \sqrt{s}\hat{\xi}(sf) \tag{25}$$

We see that the scaling has opposite effects in Eqs. (24) and (25). For a temporal dilation with $s > 1$, we observe a decrease in spread in the frequency domain. Thus, there is a trade-off in how localized the function is in time and frequency.

Taking the limiting case of a Dirac $\delta(t)$, the function is precisely localized in time. The FT gives $\hat{\delta}(f) = 1$; thus, the FT of the function is uniformly spread across all frequencies. Conversely, a function $\varepsilon(t) = e^{iat}$ has FT $\hat{\varepsilon}(f) = \delta(f - \frac{a}{2\pi})$, so is completely localized in frequency but not in time.

This negotiation is at the foundation of time-frequency analysis and is quantified by the uncertainty principle.

Theorem 1 (The uncertainty principle)**.** *The temporal and frequency standard deviations of a function* $\xi \in L^2(\mathbb{R})$ *must satisfy*

$$\sigma_t \sigma_f \geq \frac{1}{4\pi} \tag{26}$$

Proof. Following Mallat (2009), we prove the uncertainty principle (with the original proof done by Weyl (1928)) for Schwartz class functions, $\xi \in \mathcal{S}(\mathbb{R})$, which are infinitely differentiable functions $\xi : \mathbb{R} \to \mathbb{C}$ such that $\xi^{(n)}(t)$ rapidly decreases $\forall n \geq 0$; formally, we write this as

$$\sup_{t \in \mathbb{R}} |t|^m |\xi^{(n)}(t)| < \infty, \quad \forall m, n \geq 0 \tag{27}$$

The FT is well defined for $\xi \in \mathcal{S}(\mathbb{R})$, with $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$. For this proof, we only need Eq. (27) to hold for $m = 1/2$, $n = 1$. This proof can be extended to any square-integrable function $\xi(t) \in \mathbf{L}^2(\mathbb{R})$, such that $\int_{-\infty}^{\infty} |\xi(t)|^2 \, \mathrm{d}t < \infty$; this is, however, a much more involved exercise, which we do not show in full here.

The temporal variance of $\xi(t)$ is given by

$$\sigma_t^2 = \frac{\int_{-\infty}^{\infty} (t - \bar{t})^2 \xi(t) \xi^*(t) \, \mathrm{d}t}{\int_{-\infty}^{\infty} \xi(t) \xi^*(t) \, \mathrm{d}t} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} (t - \bar{t})^2 |\xi(t)|^2 \, \mathrm{d}t \tag{28}$$

where $\bar{t}$ is the mean value of $\xi(t)$, which is given by

$$\bar{t} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} t |\xi(t)|^2 \, \mathrm{d}t \tag{29}$$

In the frequency domain, the variance is similarly given by

$$\sigma_f^2 = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} (f - \bar{f})^2 |\hat{\xi}(f)|^2 \, \mathrm{d}f \tag{30}$$

with

$$\bar{f} = \frac{1}{\|\xi\|^2} \int_{-\infty}^{\infty} f |\hat{\xi}(f)|^2 \, \mathrm{d}f \tag{31}$$

where we have used the Plancherel identity to get $1/\|\hat{\xi}\|^2 = 1/\|\xi\|^2$ in front of the integrals in Eqs. (30) and (31).

For mean localizations in time $\bar{t}$ and frequency $\bar{f}$ of $\xi(t)$, the mean average time and frequency location of $e^{-2\pi i \bar{f} f} \xi(t + \bar{t})$ are null, where we have simply applied a translation in time ξ by $\bar{t}$ and a modulation in frequency by $\bar{f}$. Hence, we find it sufficient to prove the theorem for $\bar{t} = 0$ and $\bar{f} = 0$.

Let us consider $\sigma_t^2 \sigma_f^2$

$$\sigma_t^2 \sigma_f^2 = \frac{1}{\|\xi\|^4} \int_{-\infty}^{\infty} |t \xi(t)|^2 \, \mathrm{d}t \int_{-\infty}^{\infty} |f \hat{\xi}(f)|^2 \, \mathrm{d}f \tag{32}$$

Now, $\mathcal{F}(\xi'(t)) = 2\pi i f \hat{\xi}(f)$, therefore, by the Plancherel identity we get

$$\sigma_t^2 \sigma_f^2 = \frac{1}{4\pi^2 \|\xi\|^4} \int_{-\infty}^{\infty} |t \xi(t)|^2 \, \mathrm{d}t \int_{-\infty}^{\infty} |\xi'(t)|^2 \, \mathrm{d}t \tag{33}$$

We now then use the Cauchy-Schwarz inequality

$$\int |\xi(t)\zeta^*(t)|^2 \,\mathrm{d}t \leq \int |\xi(t)|^2 \,\mathrm{d}t \int |\zeta(t)|^2 \,\mathrm{d}t \tag{34}$$

valid for $\xi, \zeta \in \mathbf{L}^2(\mathbb{R})$, to obtain

$$\sigma_t^2 \sigma_f^2 \geq \frac{1}{4\pi^2 \|\xi\|^4} \left[\int_{-\infty}^{\infty} |t\xi'(t)\xi^*(t)| \,\mathrm{d}t \right]^2 \tag{35}$$

$$\geq \frac{1}{4\pi^2 \|\xi\|^4} \left[\int_{-\infty}^{\infty} \frac{t}{2} \left(\xi'(t)\xi^*(t) + \xi'^*(t)\xi(t) \right) \mathrm{d}t \right]^2 \tag{36}$$

$$\geq \frac{1}{16\pi^2 \|\xi\|^4} \left[\int_{-\infty}^{\infty} t \left(|\xi(t)|^2 \right)' \mathrm{d}t \right]^2 \tag{37}$$

Integrating by parts:

$$\sigma_t^2 \sigma_f^2 \geq \frac{1}{16\pi^2 \|\xi\|^4} \left[t|\xi(t)|^2 \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} |\xi(t)|^2 \,\mathrm{d}t \right]^2 \tag{38}$$

The first term in brackets disappears for $\xi \in \mathcal{S}(\mathbb{R})$ as $\lim_{|t|\to\infty} t^{1/2}\xi^{(1)}(t) = 0$, and the remaining terms in $\|\xi\|$ cancel. Taking the square root on both sides, we obtain Eq. (26).

□

Appendix D Morlet wavelets

Morlet wavelets originate from the motivation to find a wavelet function $\psi(t)$ that simultaneously minimizes its standard deviation in time and frequency.

From the uncertainty principle in Theorem 1, we know that there is a fundamental lower bound on the spreads in time and frequency, regardless of the original wavelet function $\psi(t)$. To force equality on this lower bound, we must force equality in the Cauchy-Schwartz inequality, which is satisfied iff $\xi(t) \propto \zeta(t)$ in Eq. (34). The Cauchy-Schwartz inequality applied at Eq. (33), therefore, requires $t\psi(t) \propto \psi'(t)$. We solve the differential equation

$$\frac{\mathrm{d}\psi(t)}{\mathrm{d}t} = -2pt\psi(t) \tag{39}$$

for some $p \in \mathbb{C}$.

Rearranging and separating variables, we obtain

$$\int \frac{\mathrm{d}\psi(t)}{\psi(t)} = -2p \int t \,\mathrm{d}t \tag{40}$$

which gives us

$$\ln \psi(t) = -pt^2 + c \tag{41}$$

We, therefore, obtain Gaussian solutions of the form

$$\psi(t) = qe^{-p(t-a)^2} e^{ibx} \tag{42}$$

Therefore, we have shown that the unique family of signals that achieve the lower bound $\sigma_t \sigma_f = \frac{1}{4\pi}$ are complex exponentials multiplied by a Gaussian, thus, giving the Morlet wavelet functions in Eq. (5).

Appendix E Discrete wavelet transform

The CWT is a redundant measure as it uses a non-orthogonal set of wavelets, which means that the resulting coefficients are highly correlated. In addition, the translation τ and scale a factors for the mother wavelets in Eq. (4) are continuous, so we effectively have an infinite number of wavelets. For these reasons, we are motivated to reduce the parameterization of the WT in order to spectrally describe the signal with a minimal number of components.

In the DWT, we discretize the mother wavelet of Eq. (4):

$$\psi_{j,k}(t) = \frac{1}{\sqrt{a_0^j}}\psi\left(\frac{t - k\tau_0 a_0^j}{a_0^j}\right) \tag{43}$$

where $j, k \in \mathbb{Z}$ scale and translate the mother wavelet, such that j sets the wavelet's width and k sets its position. $a_0 > 1$ is a fixed dilation step and τ_0 is the translation factor. The wavelets in Eq. (43) are usually continuous functions; it is the sampling in the time-scale space that is now discrete. The convention $a_0 = 2, \tau_0 = 1$ is usually followed, which corresponds to dyadic grid sampling; this is a natural choice and is also computationally efficient.

Any signal $\xi(t)$ can, hence, be reconstructed by and expressed as the superposition

$$\xi(t) = \sum_j \sum_k \varsigma_{j,k}\psi_{j,k}(t) \tag{44}$$

for wavelet coefficients $\varsigma_{j,k}$

$$\varsigma_{j,k} = \int_{\infty}^{\infty} \xi(t)\psi_{j,k}(t)\,\mathrm{d}t \tag{45}$$

as long as the energy of the wavelet coefficients are bound by constants A, B (Daubechies, 1990), such that

$$A\|\xi(t)\| \leq \sum_j \sum_k \left|\int_{\infty}^{\infty} \varsigma_{j,k}\psi_{j,k}(t)\,\mathrm{d}t\right|^2 \leq B\|\xi(t)\| \tag{46}$$

with $0 < A < B < \infty$. When $A = B$, the discrete wavelets form an orthonormal basis:

$$\int_{\infty}^{\infty} \psi_{j,k}(t)\psi_{j',k'}(t)\,\mathrm{d}t = \delta_{j,j'}\delta_{k,k'} \tag{47}$$

For a complete signal representation, the spectrum needs to be covered all the way to zero; stretching the wavelet in time by a factor of 2 contracts the frequency response by a factor of 2; thus, as per Zeno's dichotomy paradox, an infinite number of wavelets should still be needed. Practically, we choose a suitable cutoff (i.e. set a maximum j) and then fill this gap with a low-pass spectrum that we call the *scaling function* (or father wavelet). Intuitively, the scaling function extracts the global features that are missed by the wavelet coefficients. See Mallat (2009); Debnath & Shah (2015) for further information on the scaling function and wavelet decomposition/reconstruction.

Acronyms

CoI	cone of influence
CPS	cross power spectrum
CWPS	cross wavelet power spectrum
CWT	continuous wavelet transform
DFT	discrete Fourier transform
DWT	discrete wavelet transform
EE	east-east
EoR	Epoch of Reionization
EW	east-west
FFT	fast Fourier transform
FT	Fourier transform
HERA	Hydrogen Epoch of Reionization Array
LST	Local Sidereal Time
MAD	median absolute deviation
NN	north-north
PS	power spectrum
RFI	radio-frequency interference
STFT	short-time Fourier transform
WPS	wavelet power spectrum
WT	wavelet transform

References

Barkana R., Loeb A., 2006, Monthly Notices of the Royal Astronomical Society: Letters, 372, L43

Cooley J. W., Tukey J. W., 1965, Mathematics of Computation, 19, 297

Daubechies I., 1990, IEEE Transactions on Information Theory, 36, 961

Debnath L., Shah F. A., 2015, Wavelet Transforms and Their Applications, second edn. Birkhäuser, Boston, MA, doi:10.1007/978-0-8176-8418-1

Gabor D., 1946, Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering, 93, 429

Gauss C. F., 1866, Theoria Interpolationis Methodo Nova Tractata. Carl Friedrich Gauss Werke Vol. 3, Königlichen Gesellschaft der Wissenschaften, Göttingen

Hamaker J. P., Bregman J. D., Sault R. J., 1996, Astronomy and Astrophysics Supplement Series, 117, 137

Lee G. R., Gommers R., Waselewski F., Wohlfahrt K., O'Leary A., 2019, Journal of Open Source Software, 4, 1237

Mallat S., 2009, A Wavelet Tour of Signal Processing, third edn. Academic Press, Boston, doi:10.1016/B978-0-12-374370-1.X0001-8

Mesinger A., Furlanetto S., Cen R., 2011, Monthly Notices of the Royal Astronomical Society, 411, 955

Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, The Astrophysical Journal, 769, 154

Murray S., Greig B., Mesinger A., Muñoz J., Qin Y., Park J., Watkinson C., 2020, The Journal of Open Source Software, 5, 2582

Sofroniew N., et al., 2022, napari: A Multi-dimensional Image Viewer for Python, doi:10.5281/zenodo.6598542

The HERA Collaboration et al., 2022, The Astrophysical Journal, 925, 221

Trott C. M., 2016, Monthly Notices of the Royal Astronomical Society, 461, 126

Weyl H., 1928, Gruppentheorie und Quantenmechanik. S. Hirzel, Leipzig