HERA-12 Autocorrelations

Aaron Ewall-Wice

March 2019

Abstract

We investigate the spectral performance of HERA-12 auto-correlations. While charge pump spikes appear to have been mitigated, there is a low level systematics floor at one part in $\sim 10^{-4}$ that is above the level of thermal noise floor. In light of simulation results, this systematics floor is a real artifact in the measurements. In addition, large-delay wings remain at the lowest frequencies though the grounding of suspension springs on Antenna 39 seems to improve these significantly. The -45 dB systematics floor in filtering/clean residuals appears to be dominated by RFI-like spikes that appear across antennas, are higher in the YY polarization, and are relatively stable in time in the autocorrelations. These spikes dissapear when an noise source is hooked up. It follows that the -45 dB floor is likely arising from external RFI whose phase stability in cross-correlations will have important implications on HERA's ultimate success.

1 Note on Julian Date Shorthand

All of the Julian Dates in this memo started with '245'. For example, 2458560. I often neglect to include these first three digits and write, for example, 8560 instead.

2 Introduction

By integrating power over the entire sky, autocorrelations are a fast way to obtain high signal-to-noise measurements of the spectral structure in an interferometer's signal chain (e.g. **(author?)** 3, 4). In this memo, we investigate the systematics in the HERA signal chain by inspecting delay-filtered auto-correlations. In the absence of unflagged RFI, Auto-correlations that are below $\approx 10^{-5}$ the level of the DC mode are a necessary but not sufficient condition for the antenna and RF signal chain to be able to detect the 21 cm signal using proven analysis techniques. This is because an additional smooth noise bias is present in an auto-correlation, that has not necessarily traveled through all stages of the signal chain that sky radiation is subjected too. Thus, fine-scale spectral features have the potential to be suppressed relative to the $\tau = 0$ mode.

3 Auto-correlations as a Diagnostic of spectral structure.

Why are we using auto-correlations as a diagonistic of spectral structure? To build some intuition, lets relate the spectral structure present in an auto-correlation to the spectral structure we observe in a cross-correlation. Suppose the electrical field from the sky, at antenna α, is given by s^α. Each stage in a signal chain adds some noise n_i^α and then a multiplicative gain g_i^α where we let larger i denote gain stages that are further down-stream in the signal chain from the antenna. The voltage measured by the correlator, from antenna α is a gaussian variate with mean $v^\alpha(\nu)$ and standard deviation $\sigma_\alpha^2(\nu)$. To simplify notation, we will assume that the statistics of $v_m^\alpha(\nu, t)$ are stationary (the distribution is time invariant). The voltage measured by the correlator at time t and frequency ν, $v_m^\alpha(\nu, t)$ is

$$v_m^\alpha(\nu, t) \sim \mathcal{N}(v^\alpha(\nu), \sigma_\alpha^2(\nu)). \tag{1}$$

We can write down the expectation value of, $\langle v^\alpha(\nu, t) \rangle = v^\alpha(\nu)$, in terms of a linear combination of the means of the noise sources multiplied by gain terms.

$$v^\alpha(\nu) = \prod_{i=1}^{N} g_i^\alpha(\nu) s^\alpha(\nu) + \sum_{i=1}^{N} \prod_{j=i}^{N} g_j^\alpha(\nu) n_i^\alpha(\nu) \equiv \Gamma_1^\alpha(\nu) s^\alpha(\nu) + \sum_{i=1}^{N} \Gamma_i^\alpha(\nu) n_i^\alpha(\nu). \tag{2}$$

where $\Gamma_i^\alpha(\nu) = \prod_{j=i}^N g_j^\alpha(\nu)$ is the effective gain experience by power injected by noisy components at the input of the ith stage of the signal chain.

We will assume, for now, that the injected noise terms in the signal chain n_i are uncorrelated from antenna to antenna and are zero-mean complex Gaussian variates. If this is the case, then the cross correlation is

$$V^{\alpha\beta}(\nu) \equiv \langle v^\alpha(\nu,t)v^{\beta*}(\nu,t)\rangle = \Gamma_1^\alpha(\nu)\Gamma_1^{\beta\star}(\nu)\langle s^\alpha(\nu,t)s^{\beta*}(\nu,t)\rangle = \Gamma_1^\alpha(\nu)\Gamma_1^{\beta\star}(\nu)S^{\alpha\beta}(\nu) \tag{3}$$

While $S^{\alpha\beta}(\nu)$ is the intrinsic foreground correlation function which occupies the "wedge". Non-smoothness of $\Gamma_1^\alpha(\nu)$ and $\Gamma_1^{\beta*}(\nu)$ leak foreground power out of the wedge via a convolution in delay-space.

On the other hand, the variance of the visibility is

$$\sigma_{\alpha\beta}^2(\nu) = \langle|v^\alpha(\nu,t)|^2\rangle\langle|v_\beta(\nu,t)|^2\rangle = V^{\alpha\alpha}(\nu)V^{\beta\beta}(\nu) \tag{4}$$

For comissioning purposes, we would like to make a high SNR measurement of Γ_1^α and Γ_1^β. Unfortunately, $S^{\alpha\beta}$ is typically much smaller than $V^{\alpha\alpha}$ and $V^{\beta\beta}$ so we must build sensitivity by integrating over many time-steps. A signal chain is "EoR ready" if, at high delays ($\gtrsim 300$ns), it is devoid of spectral structure above ~ -50dB. Because $\langle v^\alpha v^{\beta*}\rangle \ll \sigma_{\alpha\beta}^2$, establishing the absence of such spectral structure in cross-correlations requires a measurement with sensitivity similar to the full 21 cm measurement – an impractical task with a small number of antennas.

For an autocorrelation, the variance is

$$\sigma_{\alpha\alpha}^2(\nu) = \langle|v^\alpha(\nu,t)|^4\rangle - (\langle|v^\alpha(\nu,t)|^2\rangle)^2 = (\langle|v^\alpha(\nu,t)|^2\rangle)^2 = (V^{\alpha\alpha}(\nu))^2 \tag{5}$$

and the mean is given by

$$V^{\alpha\alpha}(\nu) = \langle|v^\alpha(\nu,t)|^2\rangle = |\Gamma_1^\alpha(\nu)|^2\langle s^\alpha(\nu,t)s^{\alpha*}(\nu,t)\rangle + \sum_{i=1}^N |\Gamma_i^\alpha(\nu)|^2\langle|n_i(\nu,t)|^2\rangle \equiv |\Gamma_1^\alpha(\nu)|^2 S^\alpha(\nu) + \sum_{i=1}^N |\Gamma_i^\alpha(\nu)|^2 N_i^\alpha(\nu) \tag{6}$$

Thus, for a single time-step the variance is equal to the mean of the signal we measure (rather than much less as in the cross-correlation case). This allows us to rapidly build sensitivity and we an obtain a high SNR measurement of the signal chain with modest collecting areas and integration times.

As we can see from equation 6, a clear downside to using the auto-correlations to learn about Γ_1^α is that they are biased by system noise terms.

What do we actually learn about an autocorrelation by taking its Fourier transform?

$$\widetilde{V}^{\alpha\alpha}(\tau) = (\widetilde{\Gamma}_1^\alpha(\tau) \star \widetilde{\Gamma}_1^{\alpha*}(-\tau)) \star \widetilde{S}^{\alpha\alpha}(\tau) + \sum_i (\widetilde{\Gamma}_i^\alpha(\tau) \star \widetilde{\Gamma}_i^{\alpha*}(-\tau)) \star \widetilde{N}_i^\alpha(\tau) \tag{7}$$

We can approximate the Fourier transforms of spectrally smooth foreground means $\widetilde{S}^{\alpha\alpha}(\tau)$ and noise biases $\widetilde{N}^{\alpha\alpha}(\tau)$ as proportional to δ_D functions in the Fourier domain so that the convolutions become multiplications by a constant.

$$\widetilde{V}^{\alpha\alpha}(\tau) \approx \widetilde{S}_0^{\alpha\alpha}(\widetilde{\Gamma}_1^\alpha(\tau) \star \widetilde{\Gamma}_1^{\alpha*}(-\tau)) + \sum_i \widetilde{N}_{i,0}^\alpha(\widetilde{\Gamma}_i^\alpha(\tau) \star \widetilde{\Gamma}_i^{\alpha*}(-\tau)) \tag{8}$$

$\widetilde{\Gamma}_1^\alpha(\tau) \star \widetilde{\Gamma}_1^{\alpha*}(-\tau)$ is essentially the delay-kernel of the signal chain convolved with its complex conjugate in Fourier space. If the kernel occupies positive delays, then the Fourier transform of its complex conjugate occupies negative delays. What results is a kernel which to first order is equal to Γ_1^α reflected about itself at $\tau = 0$ and is equal to what the sky in a cross-correlation would be convolved by if $\Gamma_1^\alpha(\nu) = \Gamma_1^\beta(\nu)$.

We now consider a simple scenario to help us understand how an auto-correlation might be biased. First, lets imagine that the signal chain involves three steps: an antenna, g_{ant}, a low-noise amplifier, g_{LNA}, and an RF chain, g_{RF} that leads from the LNA to the receiver. We assume that all of the system noise is injected at the low-noise amplifier, $n_{\text{ant}}(\nu,t) \sim n_{\text{RF}}(\nu,t) \sim \delta_D(0)$. This simplified setup is a standard case for VHF systems where noise terms downstream of the LNA are heavily suppressed.

$$\widetilde{V}^{\alpha\alpha}(\tau) = \widetilde{S}_0^{\alpha\alpha}(\widetilde{\Gamma}_1^\alpha(\tau) \star \widetilde{\Gamma}_1^{\alpha*}(-\tau)) + N_{\text{LNA},0}^\alpha(\widetilde{\Gamma}_2^\alpha(\tau) \star \widetilde{\Gamma}_2^{\alpha*}(-\tau)) \tag{9}$$

Where here, $\Gamma_2^\alpha = \Gamma_1^\alpha/g_{\text{ant}}^\alpha$. If the bias in equation 9 is spectrally smooth, (or at least smoother than Γ_1^α), than it will have the effect of making high delay modes appear smaller, relative to zero delay than is actually the case in Γ_1^α. Since dividing by $g_{\text{]ant}}^\alpha$ eliminates the spectral structures arising from reflections in the antenna (or for a phased array the beamformer-dipole cables), one might expect for this to actually be the case for low and moderate

delay structures. On the other hand, high delay structures, such as RFoF and cable reflections are not affected by dividing out the antenna gain.

Compare the delay spectra of a short baseline normalized to 1 at $\tau = 0$ with an autocorrelation normalized to 1 at $\tau = 0$ to see the actual impact of bias!

In summary, our expectation for a signal chain where most noise power is injected downstream of the antenna, at the LNA, but upstream of the cables leading to the receivers and correlator is that the low-moderate delays that are impacted by intra-dish reflections will appear suppressed (relative to $\tau = 0$) while high delay features that are downstream of the LNA will be unaffected. We therefor consider the observation of spectral structure in the autocorrelations that is suppressed at moderate-low delays below the level necessary for an EoR detection to be a necessary but not sufficient test of whether the system is EoR ready. This wold not be the case if $\langle |n_{\mathrm{LNA}}(\nu, t)|^2 \rangle$ or any of the down-stream noise sources injected spectrally complex noise (as is the case for the RFoF spikes).

3.1 Statistics

Let's discuss the statistics of autocorrelations in more detail.

We have been assuming that all voltage signals are given by complex circularly-symmetric Gaussian variables with equal variances for their real and imaginary components. If this is the case, than each time step is given by an exponential distribution.

$$V_m^{\alpha\alpha}(\nu, t) \sim \frac{1}{\sigma_{\alpha\alpha}} \exp\left[-\frac{V^{\alpha\alpha}(\nu, t)}{\sigma_{\alpha\alpha}} \right] \tag{10}$$

whose mean and variance are given by $\sigma_{\alpha\alpha}(\nu)$ and $\sigma_{\alpha\alpha}^2(\nu)$ respectively.

If we time-average this quantity over $n = \Delta\nu\tau_{\mathrm{acc}}$ times, the resulting distribution of $V_m^{\alpha\alpha}(\nu)$ is a Γ distribution.

$$V_m^{\alpha\alpha}(\nu) \sim \left(n\sigma_{\alpha\alpha}^{-1} \right)^n \frac{[V_m^{\alpha\alpha}(\nu)]^{n-1}}{(n-1)!} \exp\left[-n\sigma_{\alpha\alpha}^{-1}\bar{V}_m^{\alpha\alpha}(\nu) \right] \tag{11}$$

whose mean and variance are given by $\sigma_{\alpha\alpha}$ and $\sigma_{\alpha\alpha}^2/n$ respectively. An important point – this is different from an exponential since, as we might expect, its RMS is now much larger than its variance.

In order to examine spectral structure, we will be investigating the application of delay-transforms and linear filters, (i.e. linear transforms) in frequency of $^{\alpha\alpha}(\nu)$. Suppose we apply a linear operation across frequency,

$$\widetilde{V}_m^{\alpha\alpha}(\tau) \equiv \sum_\nu R_{\tau\nu} V_m^{\alpha\alpha}(\nu) \tag{12}$$

The expectation value of $\widetilde{V}_m^{\alpha\alpha}(\tau)$ is simply

$$\langle \widetilde{V}_m^{\alpha\alpha}(\tau) \equiv \widetilde{V}^{\alpha\alpha}(\tau) = \rangle = \sum_\nu R_{\tau\nu} \langle V_m^{\alpha\alpha}(\nu) \rangle = \sum_\nu R_{\tau\nu} |\Gamma_1^\alpha(\nu)|^2 S^{\alpha\alpha}(\nu) + \sum_\nu R_{\tau\nu} \sum_i |\Gamma_i^\alpha(\nu)|^2 N_i^\alpha(\nu) \tag{13}$$

This is all a very carbon intensive way of saying that, despite the slightly different statistics of the autos (vs the crosses) the Fourier transform of a time-averaged auto-correlation is still just the Fourier transform of a time-averaged auto-correlation. **There is no bias in the Fourier transform of the auto-correlations aside from the second term in equation 6, which is as smooth as or smoother than the signal chain if the receiver noise is spectrally smooth!**

The covariance can be arrived by applying the linear transform and its conjugate to the left and right-hand sides of the frequency-domain covariance matrix.

$$\widetilde{\mathsf{C}} = \mathsf{R}\mathsf{C}\mathsf{R}^H \tag{14}$$

For the typical assumption of statistically independent channels (this should not be confused by the variance and mean varying smoothly from channel to channel), the covariance of our linear transform is

$$\widetilde{\mathsf{C}}^{\alpha\alpha}(\tau, \tau') = \sum_\nu R_{\tau\nu} R_{\tau'\nu}^* \sigma_{\alpha\alpha}^2(\nu) \tag{15}$$

It is instructive to consider the tapered Fourier-Transform case $R_{\tau\nu} R_{\tau'\nu} = |T(\nu)|^2 e^{2\pi i \nu(\tau - \tau')}$. Equation 15 tells us that the covariance between two different τ delays is given by the 1-d Fourier transform in frequency of the data variance tapered by the square of the taper function. It follows that the correlations between delay-bins is set by the fourier transform of the frequency dependence of the variance matrix. This covariance only depends on the

separation between τ channels due to the fact (a) the channels of the autocorrelation are statistically independent and (b) τ only appears in the Fourier transform exponent and its conjugate which is almost always the case for $R_{\tau\nu}$. We note that it is possible for filters to introduce correlations between different frequency channels prior to Fourier transforming and could therefor violate this condition.

By the central limit theorem, the distribution of the Fourier-transformed auto-correlation is close to a complex Gaussian with a mean and covariance matrix given by equations 13 and 14.

$$\widetilde{V}_m^{\alpha\alpha}(\tau) \sim \mathcal{N}\left[\widetilde{V}^{\alpha\alpha}(\tau), \widetilde{\mathsf{C}}^{\alpha\alpha}(\tau,\tau'), \widetilde{\mathcal{C}}^{\alpha\alpha}(\tau,\tau')\right] \tag{16}$$

where $\widetilde{\mathcal{C}}$ is the relation matrix $\langle \widetilde{V}_m^{\alpha\alpha}(\tau)\widetilde{V}_m^{\alpha\alpha}(\tau')\rangle - \langle \widetilde{V}_m^{\alpha\alpha}(\tau)d\rangle\langle\widetilde{V}_m^{\alpha\alpha}(\tau')\rangle$. The relation matrix goes to zero for τ,τ' greater than the width of the Fourier taper function so we will assume that it is zero here (we restrict our analysis to only positive delays since the negative delays of the Fourier transform of a real valued autocorrelation are the complex conjugate of the positive delays). If we were to simply take the absolute value of $\widetilde{V}_\tau^\alpha$ then we would get a positive real number which can have disadvantages if you are trying to discern between data that is dominated by random noise or dominated by a bias. Fortunately,one can estimate the absolute value of the Fourier transform of a signal and by taking the conjugate product of two independent realizations [2]. We do this in the following way.

The HERA correlator outputs independent "even" and "odd" data sets lumped into a Δ file formed from the difference of even and odd, and a Σ file, formed from the sum of even and odd data.

$$\Sigma_\alpha = V_+^\alpha + V_-^\alpha \tag{17}$$

$$\Delta_\alpha = V_+^\alpha - V_-^\alpha \tag{18}$$

The expectation value of Δ^α is 0 while the expectation value of Σ^α is $2V^\alpha$ We can form even and oddd data sets by taking

$$V^{\alpha+} = \frac{\Sigma_\alpha + \Delta_\alpha}{2} \tag{19}$$

$$V^{\alpha-} = \frac{\Sigma\alpha - \Delta_\alpha}{2} \tag{20}$$

$\widetilde{V^{\alpha+}}$ and $\widetilde{V^{\alpha-}}$ are independent complex random normal variates with means and covariances (in frequency) given by equations 13 and 14. Their product $\widetilde{V^{\alpha+}}\widetilde{V^{\alpha-*}}$ is a random variate whose distribution is given by the "complex double Gaussian distribution" [5]. For people who appreciate closed form solutions, the "complex double Gaussian" is a real downer – it involves infinite sums of modified Bessel functions. The mean and second moments of this product are simple enough.

$$\langle \widetilde{V^{\alpha+}}\widetilde{V^{\alpha-*}}\rangle = \langle\widetilde{V^{\alpha+}}\rangle\langle\widetilde{V^{\alpha-}}^*\rangle = \langle|\widetilde{V}^\alpha|^2\rangle \tag{21}$$

An unbiased estimator of the autocorrelation power spectrum. Great! The variance is

$$\widetilde{\mathsf{C}}^{+-}(\tau,\tau') = |\widetilde{\mathsf{C}}(\tau,\tau')|^2 + 2\mathrm{Re}\left[\widetilde{V}_\alpha(\tau)\widetilde{V}_\alpha^*(\tau')\widetilde{C}(\tau,\tau')\right] \tag{22}$$

The first term, associated with noise, averages down with time, the second term, averages down slower.

4 The Data

We analyze approximately five hours of auto-correlation data from JD2458560. Before attempting any filtering or delay-domain analysis, we inspect the low and high band auto-correlations between $45 - 85\,\mathrm{MHz}$ and $110 - 170\,\mathrm{MHz}$ respectively. Antennas 2X, 12X, and 23X,Y display anomalous behavior.

We run `xrfi` on the autocorrelations to remove RFI. Fig. 2. We find that XRFI does not always catch all RFI and in response, we impelment more conservative flagging that throws away channels that have greater than 2.5% of times flagged by xrfi and times where greater than 5% of channels are flagged. After this flagging, we generally find that our auto-correlations are RFI free (to the level that we can determine from visual inspection).

Fig 1 shows plots of the total autocorrelation amplitudes over the high and low bands that we analyze.

In addition, there are some interesting gain jumps at LST below 9 hours so we only end up looking at data from rough fdoes noty 9 to 12 hours LST.
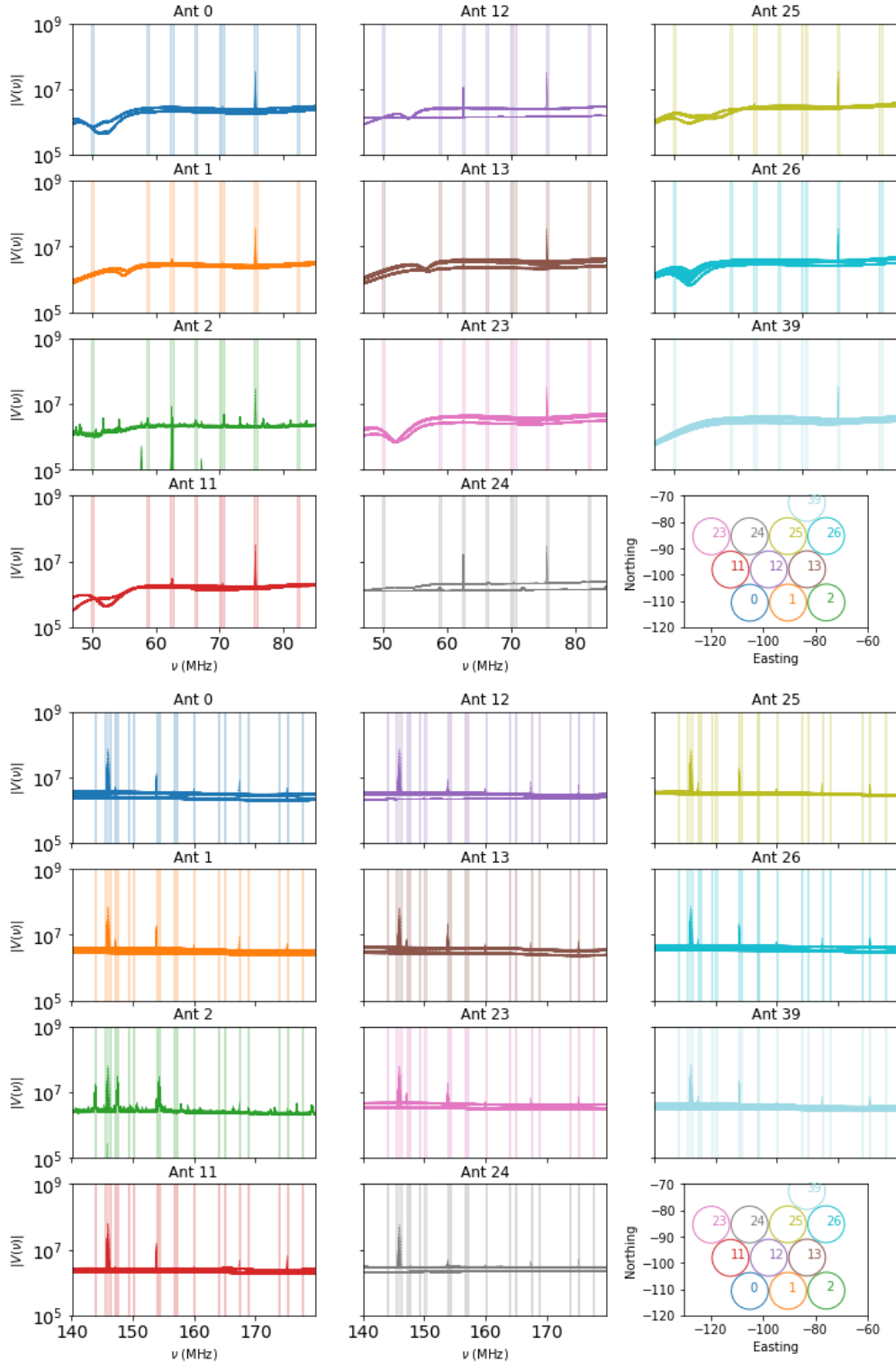
Figure 1: Top: Low-band autocorrelations between 45 and 85 MHz. Bottom, high band autocorrelations between 130 and 170 MHz. Times over all examined LST are present in each plot. Solid lines denote XX and dashed lines denote YY. In the low band, the divot appears reduced in antenna 39 though other spectral structures are now present. Major issues are present in antennas 12X, 24X and Y, and antenna 2X. Orbcom is clearly visible in the high band at ~ 137 MHz along with two other persistent RFI spikes. Over the low-band, ~ 75 MHz is heavily contaminated.

5
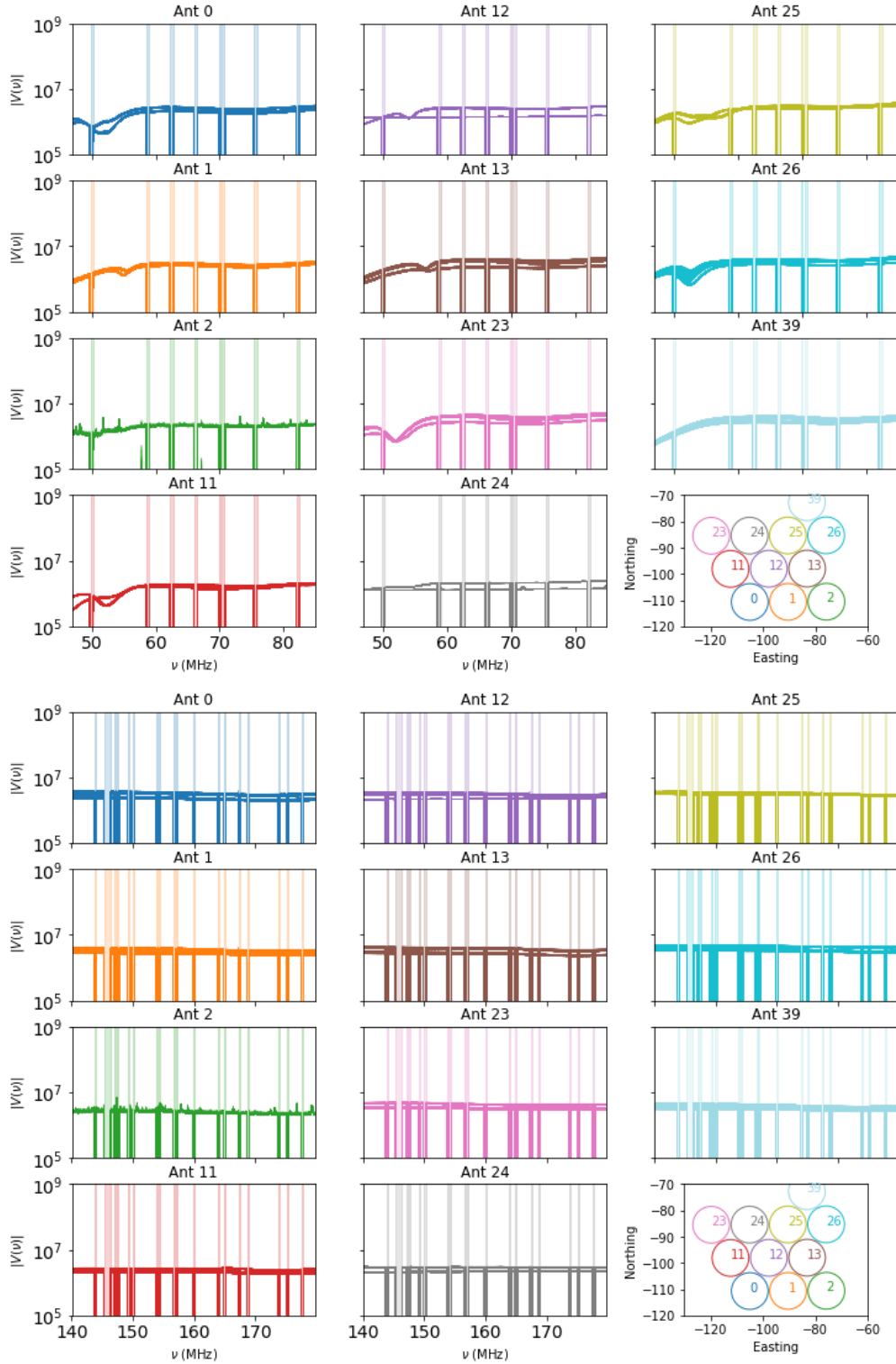
Figure 2: The same as Fig. 1 with XRFI flags applied along with a threshhold that flags all times at a channel with RFI in greater than 2.5% of data and all channels at each time where RFI is present in greater than 5% of the data. By visual inspection, the RFI spikes are absent from the data. Antennas 2, 12, 24 do not survive flagging.

6

We apply a linear delay-filter to our data and inspect the residual waterfalls in delay space (Fig. 3, Fig. 4). Over the low-band, there are significant low-delay wings that are clearly visible after a single integration. Such structures are not present in the high-band.

Antenna-0, which is the most external antenna on the array appears to have the least amount of high-delay contamination, hinting at a potential cross-coupling issue. Antennas 1X, 25X, 26X, and 39X along with 0Y, 11Y, 26Y, and 39Y have RFoF reflections visible above the noise after a single time-step at roughly 3000 ns (1 km) round trip. 26X appears to have two different reflections (and potentially 39Y?) While the divot does not clearly appear in antenna 39, the high-delay wings remain. This raises the possibility that other structures, in addition to the divot and their associated causes (which may be related to but do not exclusively involve the springs), are responsible for the high delay wings in the auto-correlations.

Many of the high band antennas do not high-delay structures at the level observed in the low band data. However, there may be some exceptions. Antennas 25 and 39 both show high-delay wings which may be related. Antenna 25X's high delay structure is transient in time and strangely enough gets worse at 11 hours LST. The time at which the wings appear in the high band of antenna 25X corresopnds to a time where the RFoF reflection drops. This suggests that, at least where it appears in the high band, the low delay wings and RFoF reflection might be related. Something similar happens in the low-band on antenna 25X. Antenna 26X's reflection seems to weaken at a similar time. I will examine specific strange behaviors such as this one in § **??**.

5 Antenna Performance from Averaged Autocorrelations.

To gain a high signal-to-noise look at spectral structure, we coherently average delay-transformed and filtered auto-correlations over three hours of LST. We plot the averaged autocorrelations normalized by their mean value over all times and frequencies in Figs. 5 and 6.

More specifically, the correlator outputs a summed and difference data set given by

$$
\begin{aligned}
\Sigma^\alpha &= v^{\alpha o} + v^{\alpha e} \\
\Delta^\alpha &= v^{\alpha e} - v^{\alpha o}
\end{aligned}
\tag{23}
$$

We therefor derive even and odd data sets by taking

$$
\begin{aligned}
\widetilde{v}^{\alpha e} &= \frac{\widetilde{\Sigma}^\alpha + \widetilde{\Delta}^\alpha}{2} \\
\widetilde{v}^{\alpha o} &= \frac{\widetilde{\Sigma}^\alpha - \widetilde{\Delta}^\alpha}{2}
\end{aligned}
$$

$$\tag{24}$$

We compare these autocorrelations with noise only diffed files. Present in all of the antennas and polarizations, at all frequencies, is a systematics floor of roughly $-45\,\mathrm{dB}$. This floor appears at roughly the same level in both the high and low bands, and as we will see, using validation simulations in § **??**, does not originate from low delay spectral structure or the filtering/cleaning algorithms themselves. We plot our simulated data in such a way that negative values

Significant wings exist over the low band

6 Sources of Spectral structure

Broadly, we observe three types of spectral structure in the delay plots of our data. These structures are

1. Low-delay wings that are prominent in the low-band and significantly reduced in the high band.

2. A flat systematic floor at $\approx -45\,\mathrm{dB}$ that exists at similar levels in both the high and low band (although it is slightly lower over the low-band). This floor is probably caused by low-level external RFI (see Table 3).

3. Discrete reflection like features at ≈ 3000ns. These features disappear when we turn on noise sources up-stream of the RFoF system so they are actually probably not reflections. Their amplitudes and delays also change dramatically with frequency (see Table 2) which are unusually large for typical reflections caused by impedance mismatches.
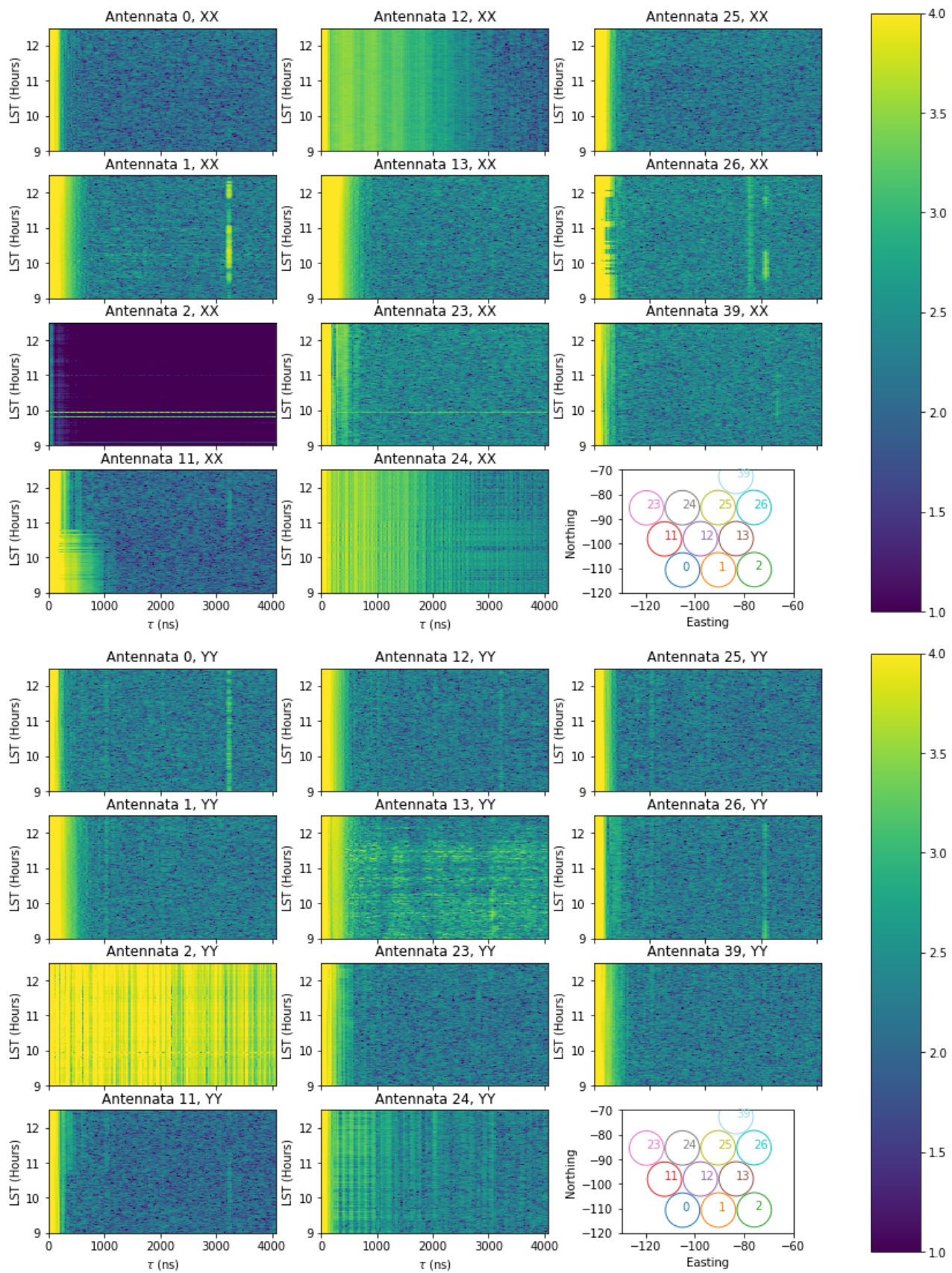
Figure 3: Waterfalls of delay-cleaned and restored data. Top: X Pols. Bottom: Y Pols. Antenna-0, which is the most external antenna on the array appears to have the least amount of high-delay contamination, hinting at a potential cross-coupling issue. Antennas 1X, 25X, 26X, and 39X along with 0Y, 11Y, 26Y, and 39Y have RFoF reflections visible above the noise after a single time-step at roughly 3000 ns (1 km) round trip. 26X appears to have two different reflections (and potentially 39Y?)

8
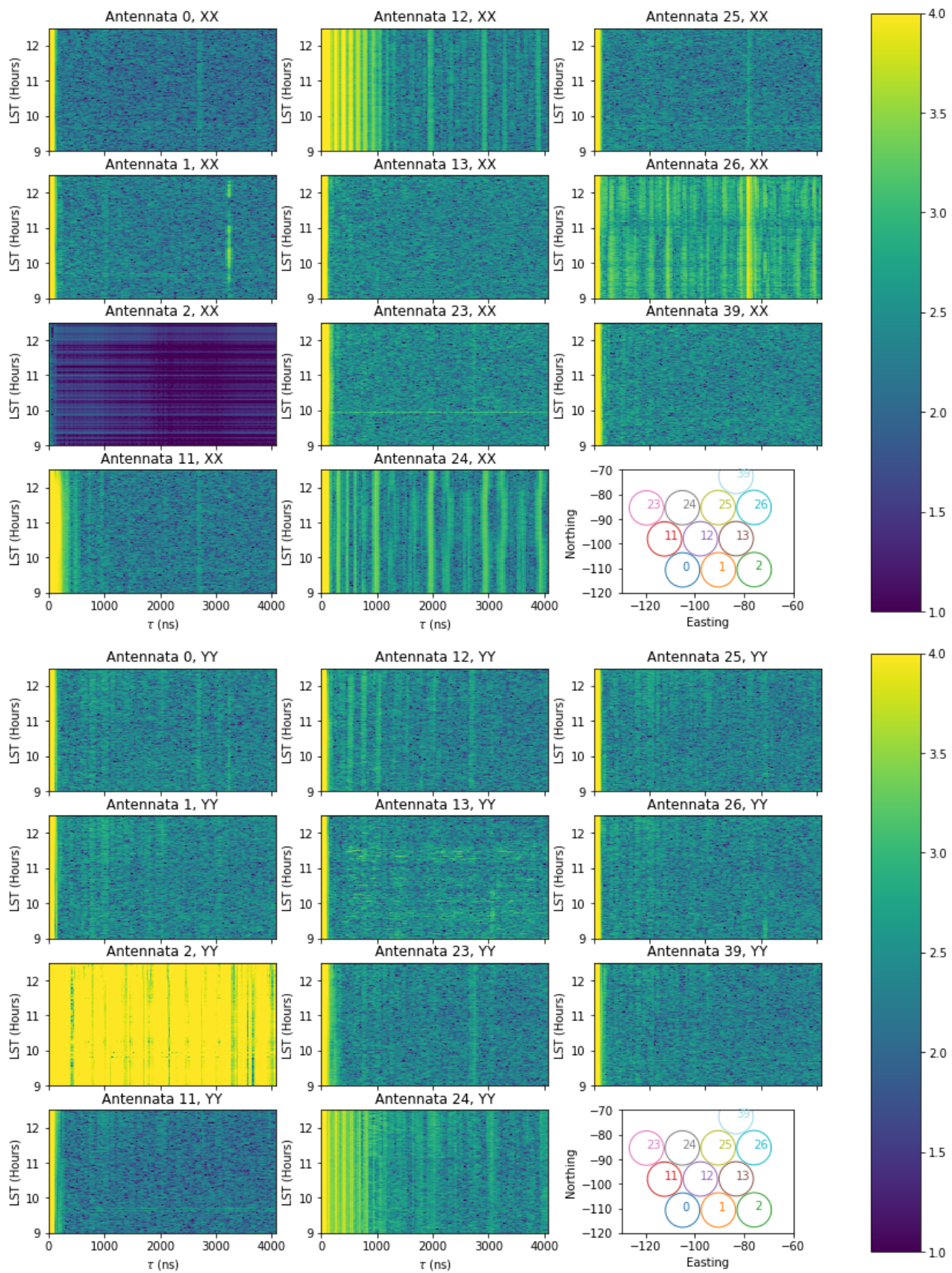
Figure 4: The same as Fig. 3 but now for the high band (130-170 MHz). We no longer observe the low delay wings that are present in the low-band data. The double reflettion is still present in 26 X. Several of teh antennas have repeated spectral stripes that may be low-level, unflagged RFI (26X, 39X, 26Y, 39 Y).

9
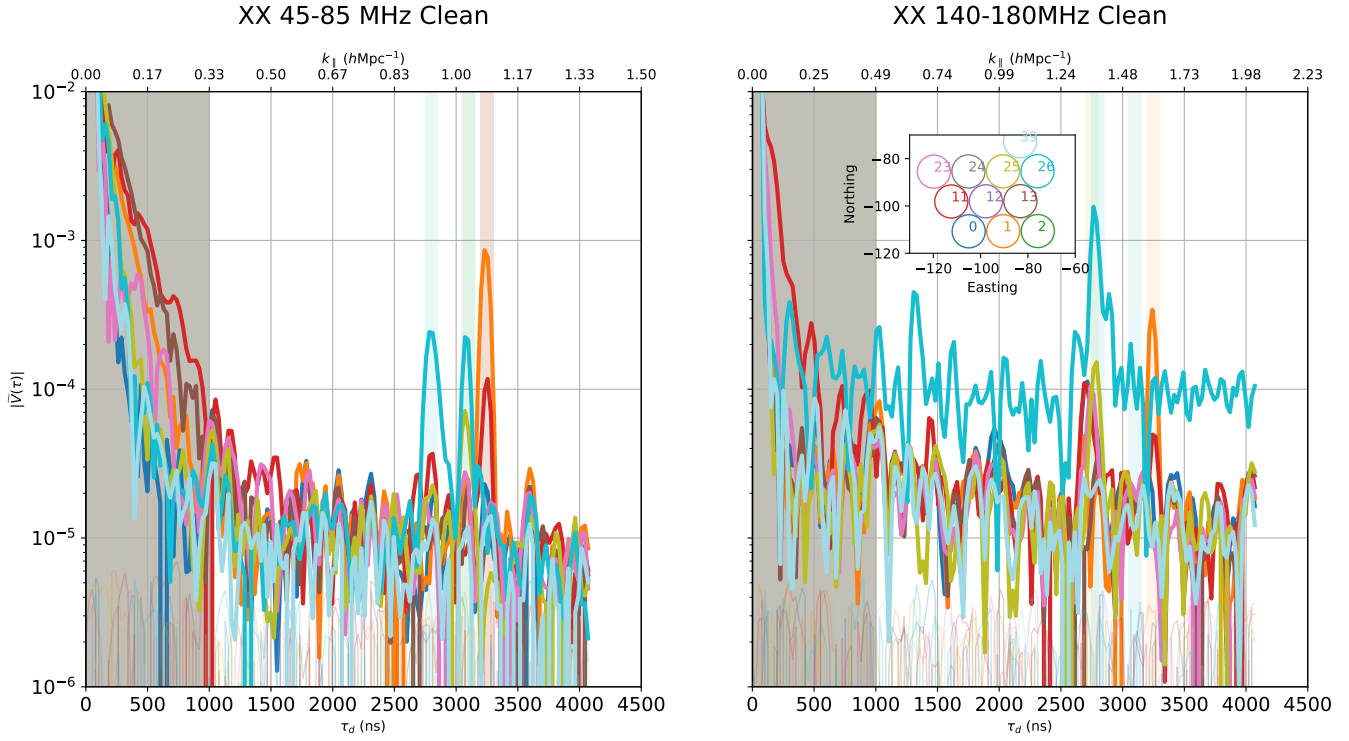
Figure 5: Autocorrelations over the low band, cleaned and restored out to 1000 ns and over narrow windows correponding to the delays of potential RFoF reflections for XX. On the left hand plot we show 45-85 MHz. A flat systematics floor exists at roughly $10^{-4} - 10^{-5}$ the level of $\tau = 0$ foregrounds. We show the high band (140-180 MHz) (right). Shaded regions indicate the area in which cleaning was allowed to detect peaks.
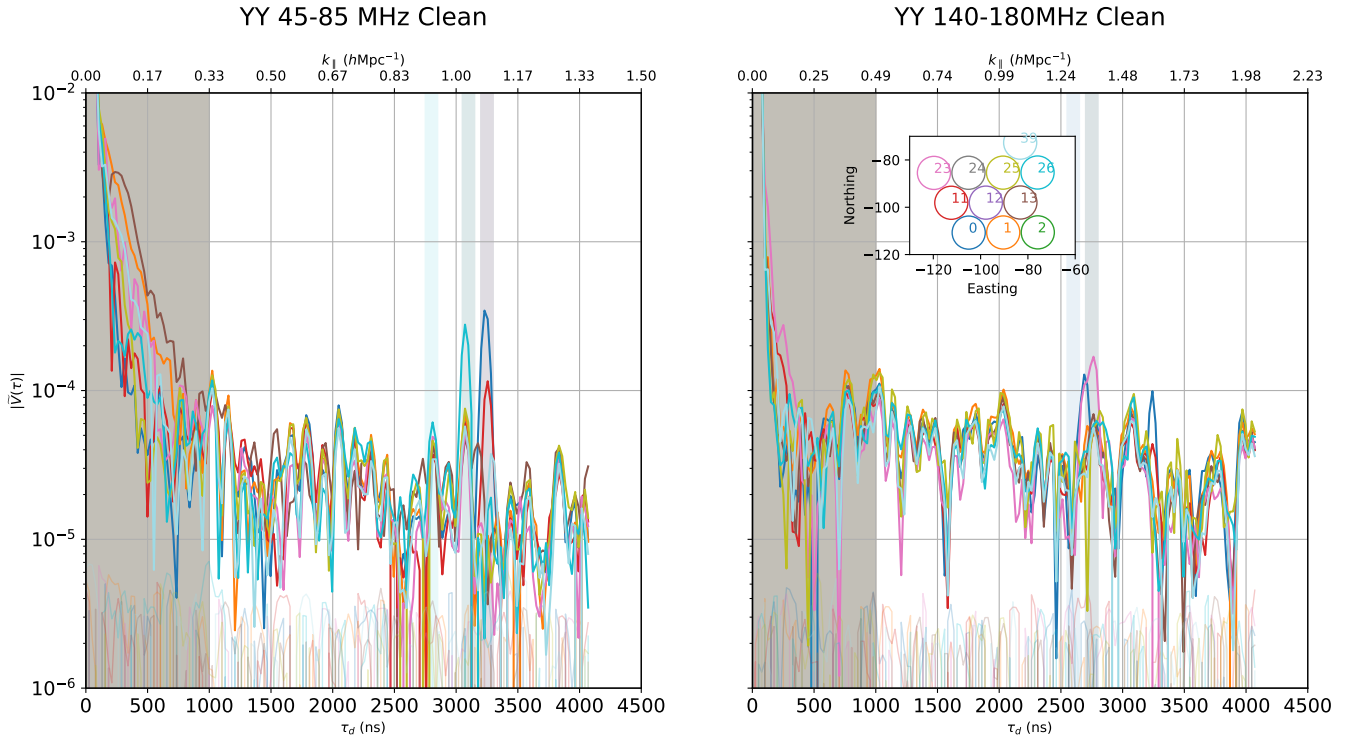
Figure 6: The same as Fig. 5 but now for the YY polarization. Unlike the XX polarization, the residuals in YY are highly correlated in the delay domain at hhigh delays, suggesting the same underyling frequency structure that is the same from antenna to antenna which might be expected for environmental RFI.
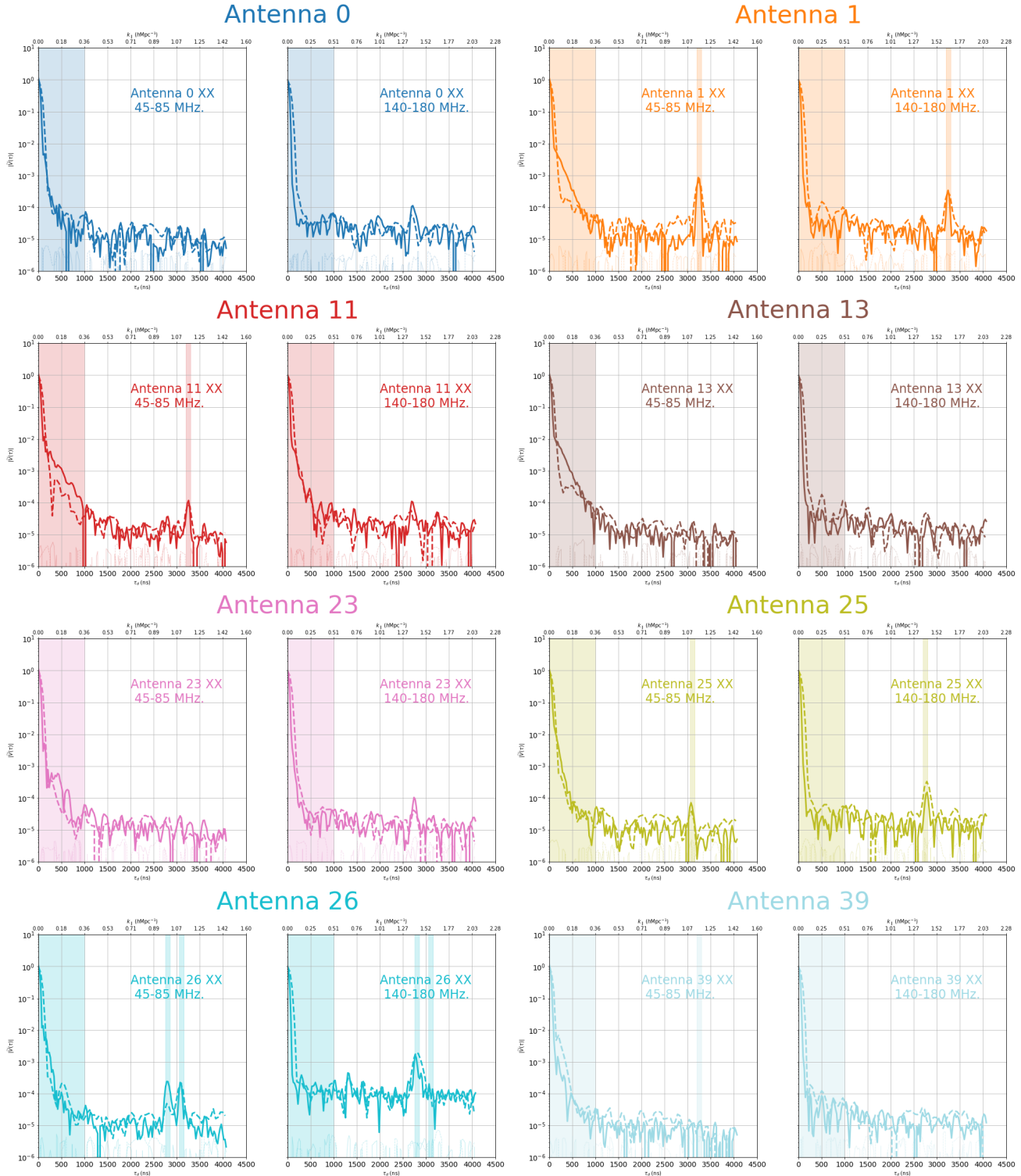
Figure 7: Left: Comparisons between full low-band (45-85 MHz) (solid lines) and restricted low band (65-86 MHz) (dashed lines). While the autocorrelations wings are significantly improved in some of the antennas when the 45-65 MHz band is excluded, roughly half of the antennas (11, 23, 25, 26) do not show much improvement. Antennas that do not improve in one polarization also do not improve in the other while antennas that do improve in one polarization also improve in the other. Note that while the wings are reduced by excluding the lowest frequencies, there are in many cases still wings present (e.g. antenna-13 XX, antenna-26 XX, Antenna-1 XX etc...). Right: The same but for the high band (130-170 MHz). A restricted band that excludes orbcomm (150-170 MHz) is shown with dashed lines. Several antennas have reflections that move in amplitude and delay between the low and high bands!

11
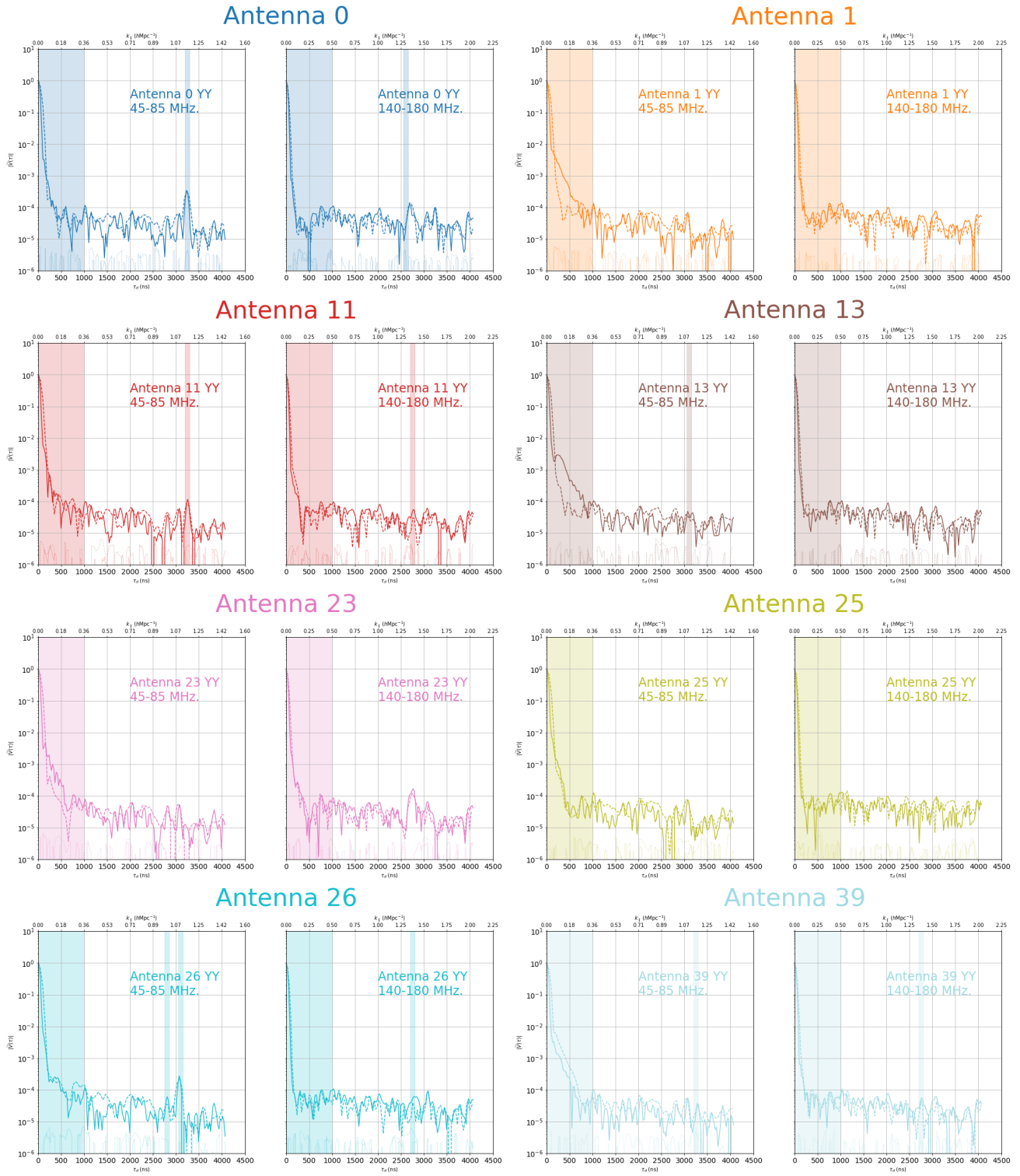
Figure 8: The same as Fig. ?? except for YY.

Hypothesis	Test	Result	Fig
Systematic introduced by analysis code bug.	Perform clean on simulated autocorrelations and perform linear filtering as an alternative check	simulated autocorrelations do not have residuals after clean	36
Sidelobes of low delay structure	Extend filtering region to 1000 ns to encapsulate all low-delay power in high and low band. Compare high/low bands since low band has more low-delay power.	Systematics floor is the same for low and high bands despite different levels of low-delay contamination. High band filter subtracts all low-delay power but does not reduce the systematics floor.	6; 5
Sidelobes of high-delay RFOF reflections	Simulated reflections. Run filtering and cleaning on simulations and also on data.	side-lobes exist at similar levels to data. A -50 dB sidelobe floor shows up in simulations that is effectively removed by filtering out reflection delays.	37
Internal RFI or reflections	see if residuals are present when noise source is on	residuals significantly lower and different – source is not downstream of noise sources	30, 21

Table 1: A table of hypotheses for the causes of the -45 dB systematics floor in auto-correlations and test results. The leading hypothesis consistent with these tests is low-level external RFI.

We demonstrate the standard contribution level of each systematic on Antenna 1 XX in Figs. 9 and 10 where we show residual levels after filtering out various systematics, inspecting residuals in the frequency and delay domains. Several prominent RFI spikes, including ORBCOMM dominate super-horizon emission at roughly $\sim 10 - 100\%$ the sky level and are clearly visible in the top panel of Figs. 9 and 10 when RFI has not been flagged. Flagging this RFI and out to 300 ns reveals the cable reflections and the systematic floor at higher delays. In the low band (Figs. 9 second panel), the wings are still prominent below 1000 ns and we see that their residuals are broad structures that tend to peak at ≈ 55 MHz. The scalloping in the frequency residuals are likely an artifact of the rapid cutoff imposed by the filter at its low cutoff of 300 ns. Extending the delay filter out to ≈ 1000 ns to remove all of the wings is shown in the third from top panels in Figs. 3 and Figs 4. After filtering the wings, the frequency domain residuals are dominated by reflections which correspond roughly over the RFoF path delay which we filter out in the bottom panels of the Figures. A puzzling aspect of these reflections is that some of them appear to shift significantly in delay over each sub-band. For example, the reflection in Antenna 11 XX appears to shift from 3250 ns over the 45-85 MHz band to 2750 ns over the 130-170 MHz band. A potential explanation of this apparent shift is that there is actually a system of reflections spread out over several hundred ns as is clearly the case on Antenna 25 XX. In addition, the reflections seem to die when noise sources are turned on (Fig. 21). It follows that the "reflections" are not reflections at all. What are they?!

After cleaing the not-reflections, residuals are dominated by spiky structures that resemble a forest of low-level RFI that could be environmental or generated within the array (our tests suggests that they originate upstream of the noise diodes). We will discuss these spikes further below.

6.0.1 The "floor"

What is the source of the -45 dB systematics floor? The bottom right panels of Figs. 9 and 10 show an averaged frequency domain view of the floor in antenna 1 XX. While the frequency strucutre is quite complicated, it does appear to be dominated by spike-like prominances that are on average several MHz apart which is suggestive of RFI. We perform several jack-knives to determine whether this RFI is generated inside of the signal chain, similar ot the RFoF charge pump spikes that were observed in earlier iterations of the system or whether it is caused by environmental RFI which could have enhanced contamination over the previous system due to changes in the directivity of the beam towards the horizon.

6.0.2 Polarization Jackknife.

We first take a jackknife in polarization with the rational that environmental RFI originating from a relatively small number of sources that are not isotropically distributed should appear differently between the X and Y pols. We show
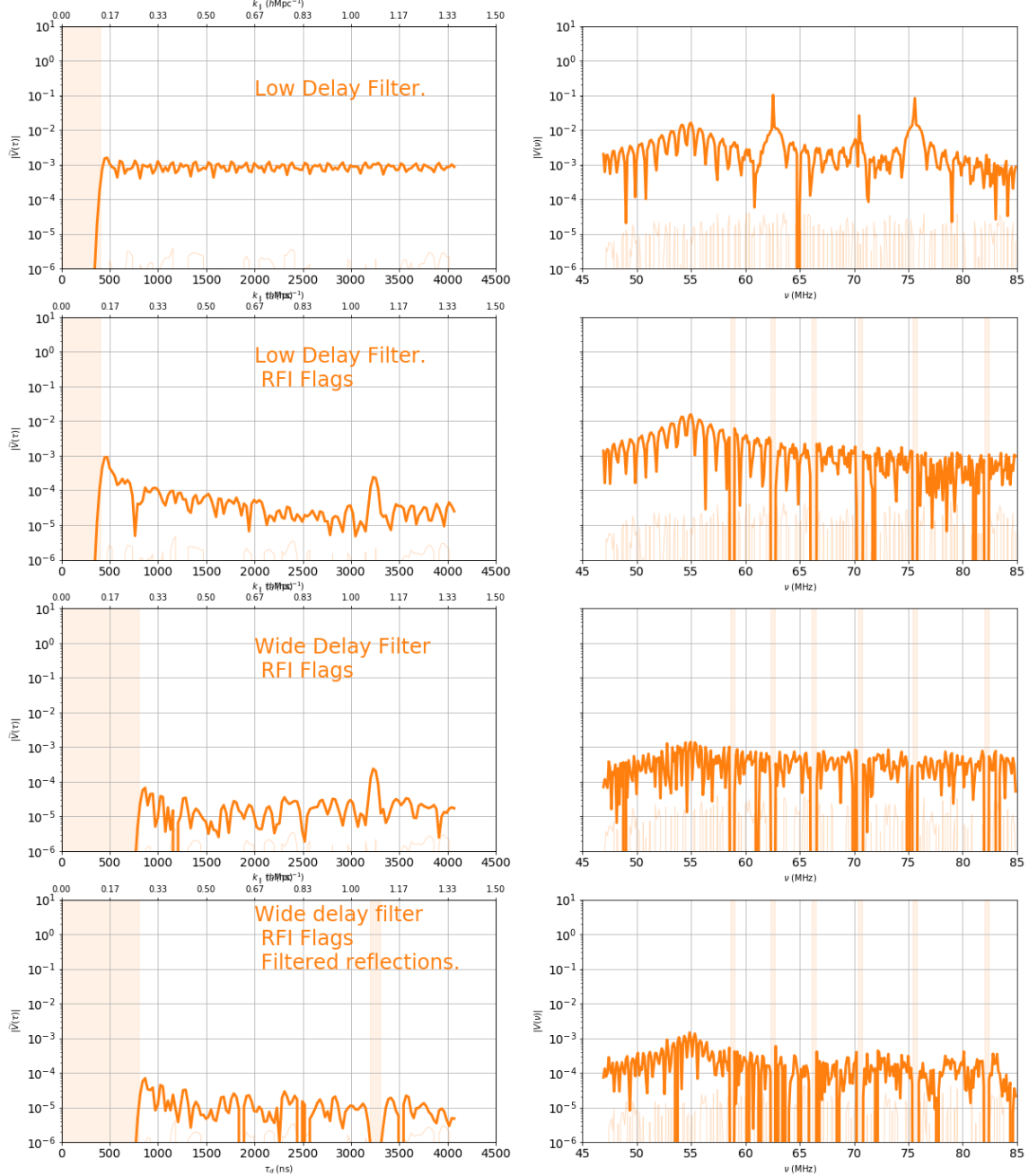
Figure 9: A sequence of delay (left) and frequency (right) domain residuals integrated coherently on an autocorrelation over ≈ 3 hours. Top: Residuals after filtering between ±300 ns but without applying any flagging of RFI. The residuals are dominated by RFI two prominant RFI spikes at ≈ 62.5 MHz and 75 MHz, leading to a characteristic flat delay-domain residual. The residual that dominates after the RFI flagging is the "wing" which is characterized as a delay-domain slope which only reaches down to -50 dB after ≈ 1000 ns. In the frequency domain, the wings form a shallow, but large amplitude peak at ≈ 55 MHz. It is possible that the scalloping that is present in the wings is simply a consequence of the filter-imposed cutoff at 300 ns which causes a power peak at roughly 500 ns, yielding the 2 MHz periodicity. The dominant residual after removing the majority of the wings (second from bottom) is a reflection-like peak at ≈ 3000 ns which appears as a (relatively) monochromatic ripple in the frequency domain. Removing this peak (bottom panel) yields an underlying systematic floor at ≈ −50 dB which consists of the high delay components of the wings which we can still see peaking at ≈ 55 MHz along with a myriad of RFI-like spike features which may be RFI. We will find that the spikes are correlated between antennas, supporting a narrow-band RFI hypothesis.
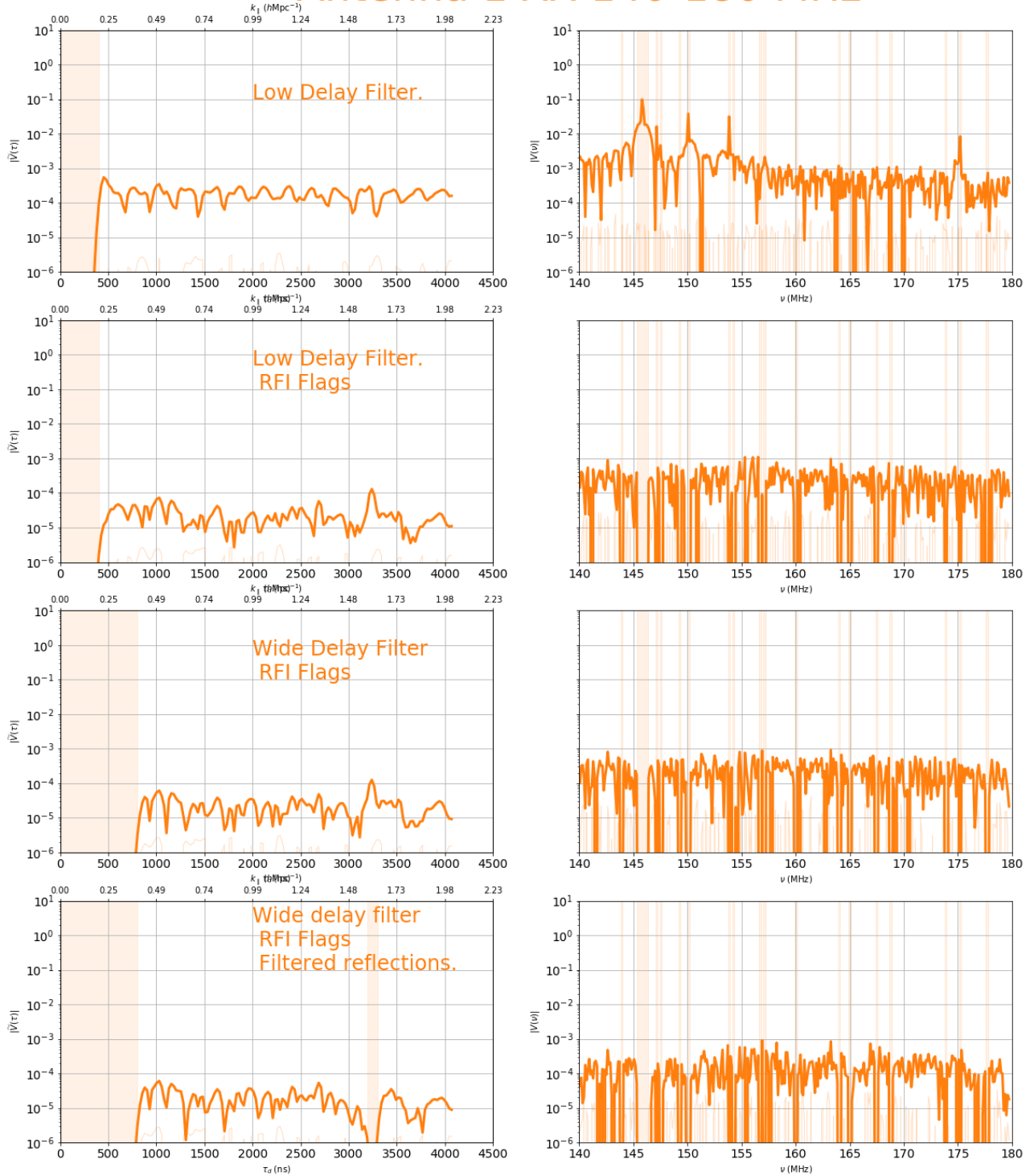
Figure 10: The same as Fig. 9 but now focusing on the "High" band between 140-170 MHz. The 3250 MHz reflection is somewhat less prominent over this band, suggesting a non-trivial degree of frequency evolution in the reflection properties that might complicate the simplest reflection removal strategies in calibration. The major RFI spikes in the high band occur over the ORBCOMM band ($\approx 136 - 138$ MHz) as well as ≈ 147 MHz, 150 MHz, and 154 MHz. The systematics floor in the high band is slightly higher than the low band, at -45 dB. This could be because there are more nummerous and higher amplitude spikes (at least 5 above -40 dB.
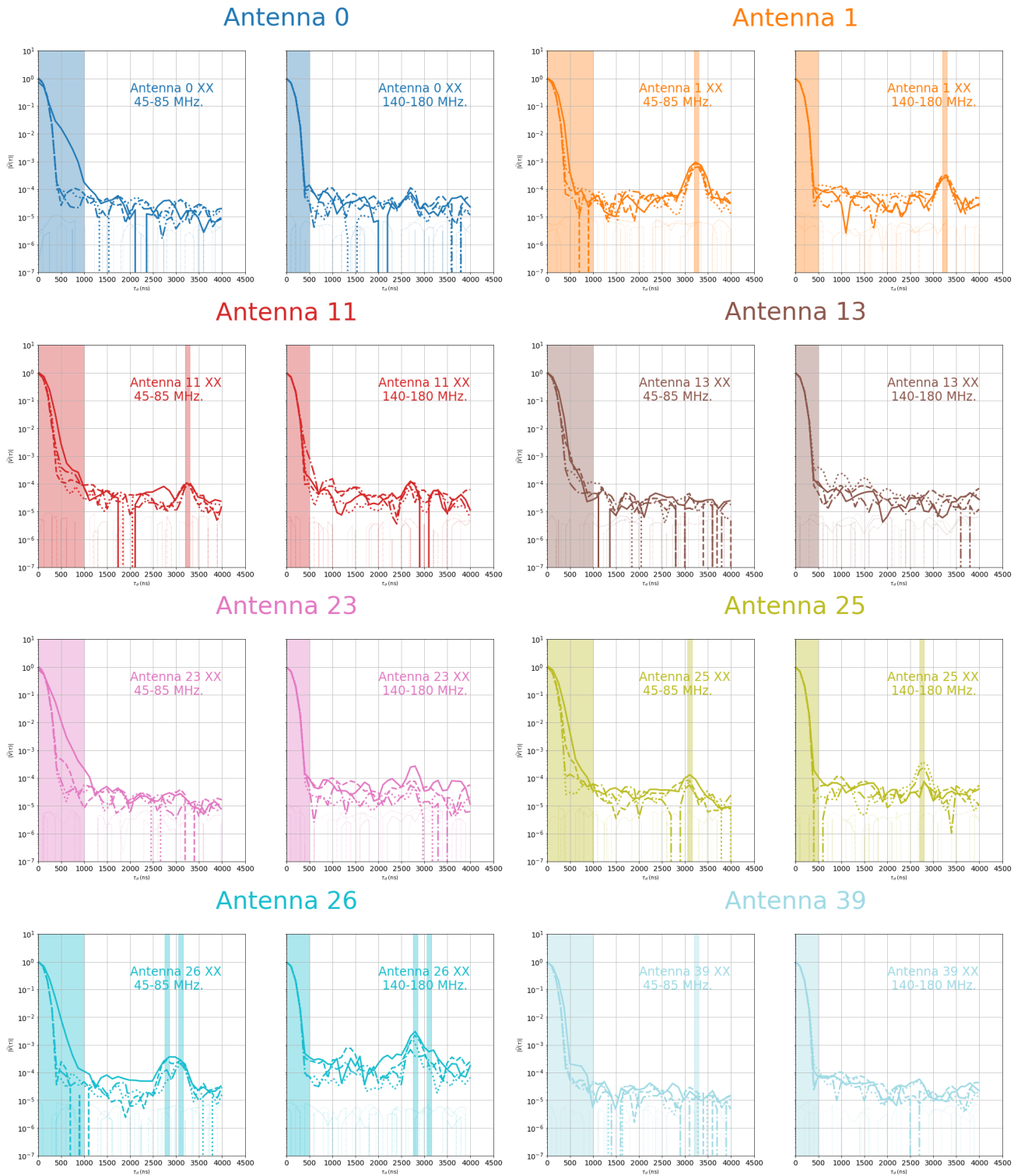
Figure 11: Comparisons between 10 MHz subbands in order of rising frequency (solid, dashed, dashed-dotted and dotted).
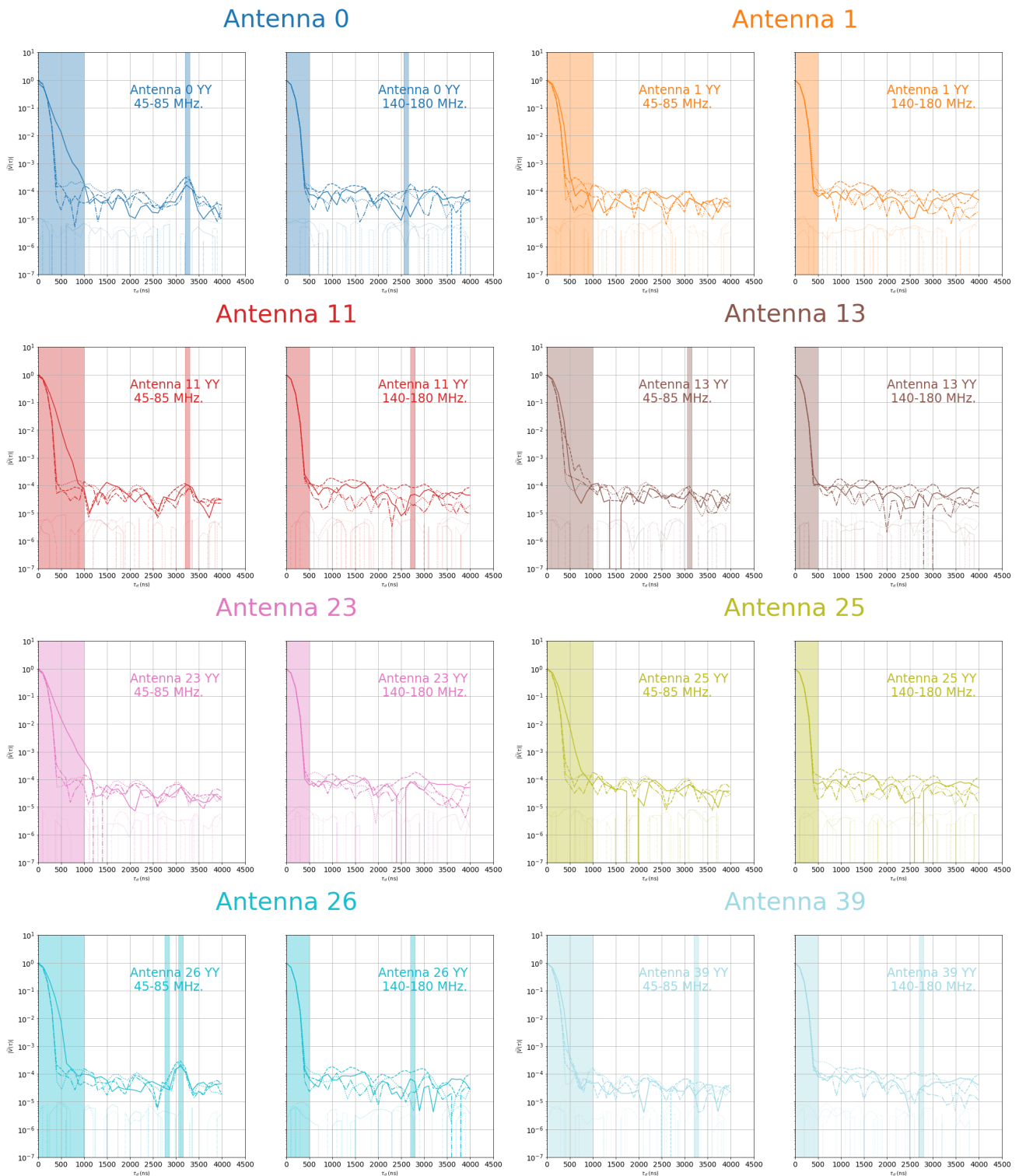
Figure 12: Comparisons between 10 MHz subbands in order of rising frequency (solid, dashed, dashed-dotted and dotted).

Figure 13:

Antenna	XX 45-85 MHz	XX 130-170 MHz	YY 45-85 MHz	YY 130-170 MHz
0 NA	NA	3250 ns	2600 ns	
1	3250 ns	3250 ns	NA	NA
11	3250 ns	2750	3250 ns	2750 ns
13	NA	NA	3100 ns	NA
23	NA	NA	NA	NA
25	3100 ns	2750 ns	NA	NA
26	3100& 2800 ns	3100& 2800 ns	3100& 2800 ns	2750
39	3250 ns	NA	3250 ns	3250 ns
2750 ns				

Table 2: Delays of observed RFoF Reflections (derived from visual inspection). Note that several antennas have RFoF reflections that are only observed over one of the bands (39XX, 13XX) while several have reflections that move down in delay. For example, Antenna 11 XX has reflections that appear at ≈ 2750 ns over 130-170 MHz but 3250 ns over 45-85 MHz. (See Figs. 7,8). One possibility for the apparent shifts in delay is that there actually exist a system of reflections associated with the RFoF (as suggested for example by Antenna 25 XX and 26 XX) and their amplitudes simply evolve rapidly over frequency.

the frequency domain jack-knives in Figs 14 and 15. With the exception of antenna 26 (which had an outlier event in the XX polarization), the floor on YY is systematically higher than XX by several dB. Over many antennas, the spikes are higher by several dB.

6.0.3 Antenna Jackknife.

We next look into how the residuals compare from antenna to antenna. Environmental RFI should appear correlated from antenna to antenna while RFI and or coupling structures arising in the individual signal chains of each antenna should appear somewhat different.

We plot comparisons of the residuals between different antennas in frequency in Fig. 16 and Fig. 17.

When comparing the different antennas, we find that many of the frequency peaks line up very well between antenna to antenna which indicates reduced likelihood that the systematics spikes are caused by per-antenna signal chain artifacts and are instead being generated by some common RFI sources.

6.0.4 Time Jackknife

How do the residuals evolve with time? We explore this by dividing our data sets into two, three hour intervals ranging from 9-12 hours of LST and 13-16 hours of LST. Many of the spiked features are common between the two time intervals, indicating that the source of the systematics floor has large coherence times, as we might expect for low-level persistent RFI. (Fig. 18).

6.0.5 Waterfalls and comparison between days.

We take a closer look at the time evolution of the systematics floor and compare its evolution across days in Fig. ?? 20. In these plots, we focus on a 20 MHz chunk and see residual RFI that has escaped `xrfi` and exists at the SNR~ 3 level when integrating 10 time-steps (86 seconds).

Here is what we know about the streaks we see in the data

1. They are generally bigger in YY (EW).

2. They repeat over multiple nights.

3. They appear stable over $\gtrsim$ hour time-scales.

4. They are at the ~ -40 to ~ -20 dB level in the autocorrelation frequency domain and (mostly) evade `xrfi`.

5. They appear to be spaced at $\lesssim 1$ MHz apart.

6. their frequencies appear stable over hour time-scales (not charge pump spikes)?

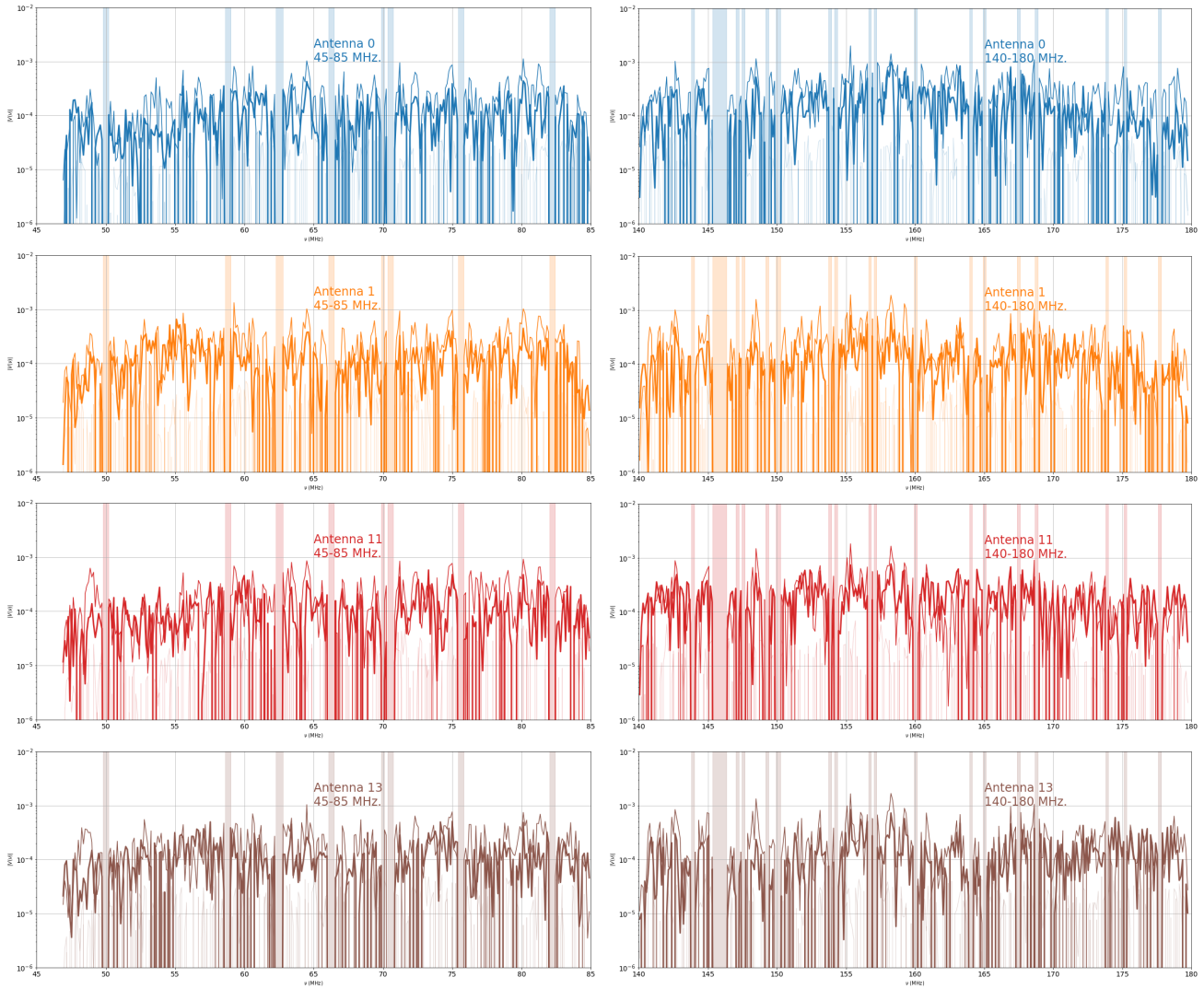Now here are some better formatted jackknife plots

Figure 14: Left column: Comparisons between XX and YY polarizations over the low band (45-85 MHz). Right column: THe same for the high band (140-180 MHz).
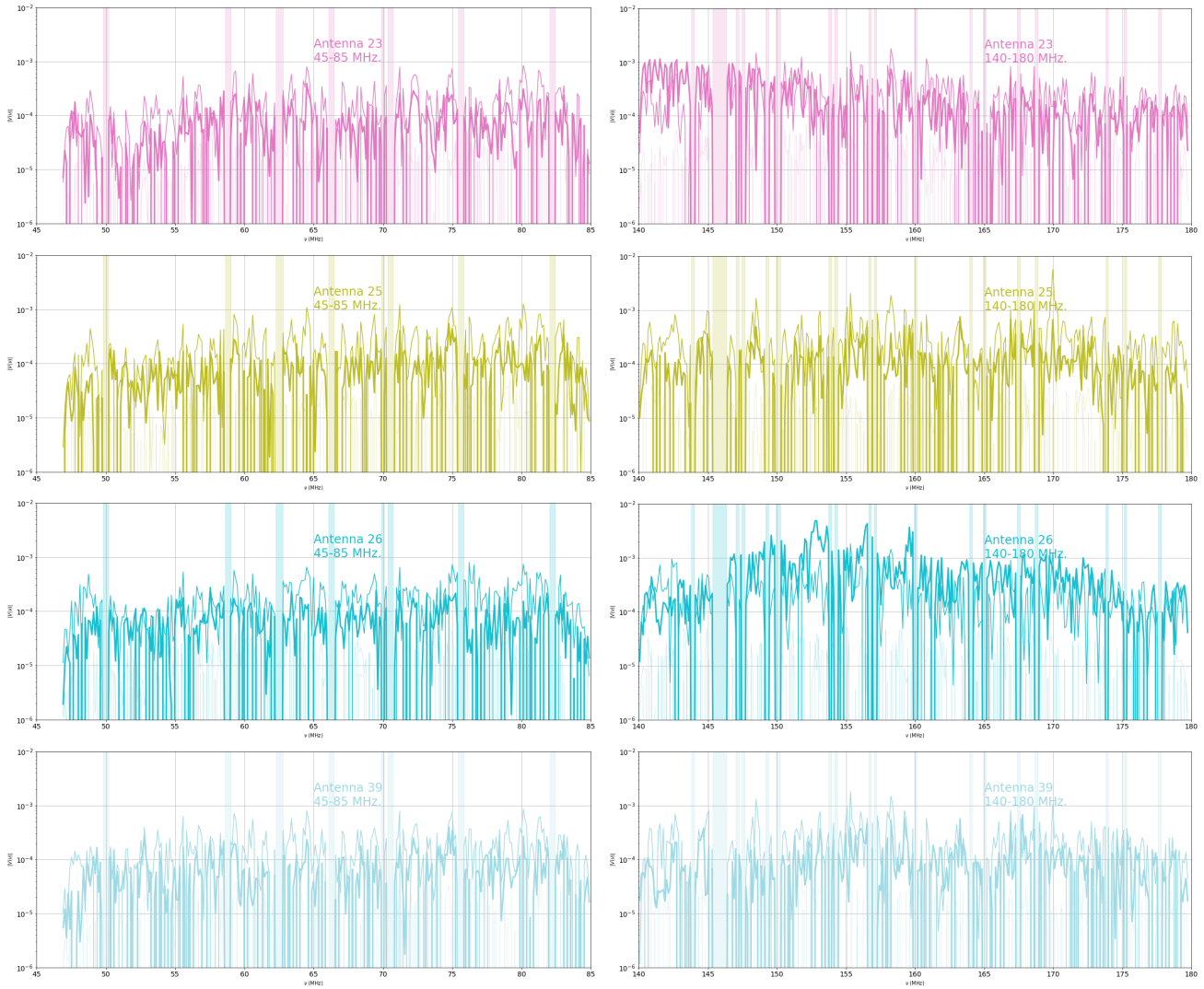
Figure 15: Left column: Comparisons between XX and YY polarizations over the low band (45-85 MHz). Right column: THe same for the high band (140-180 MHz).
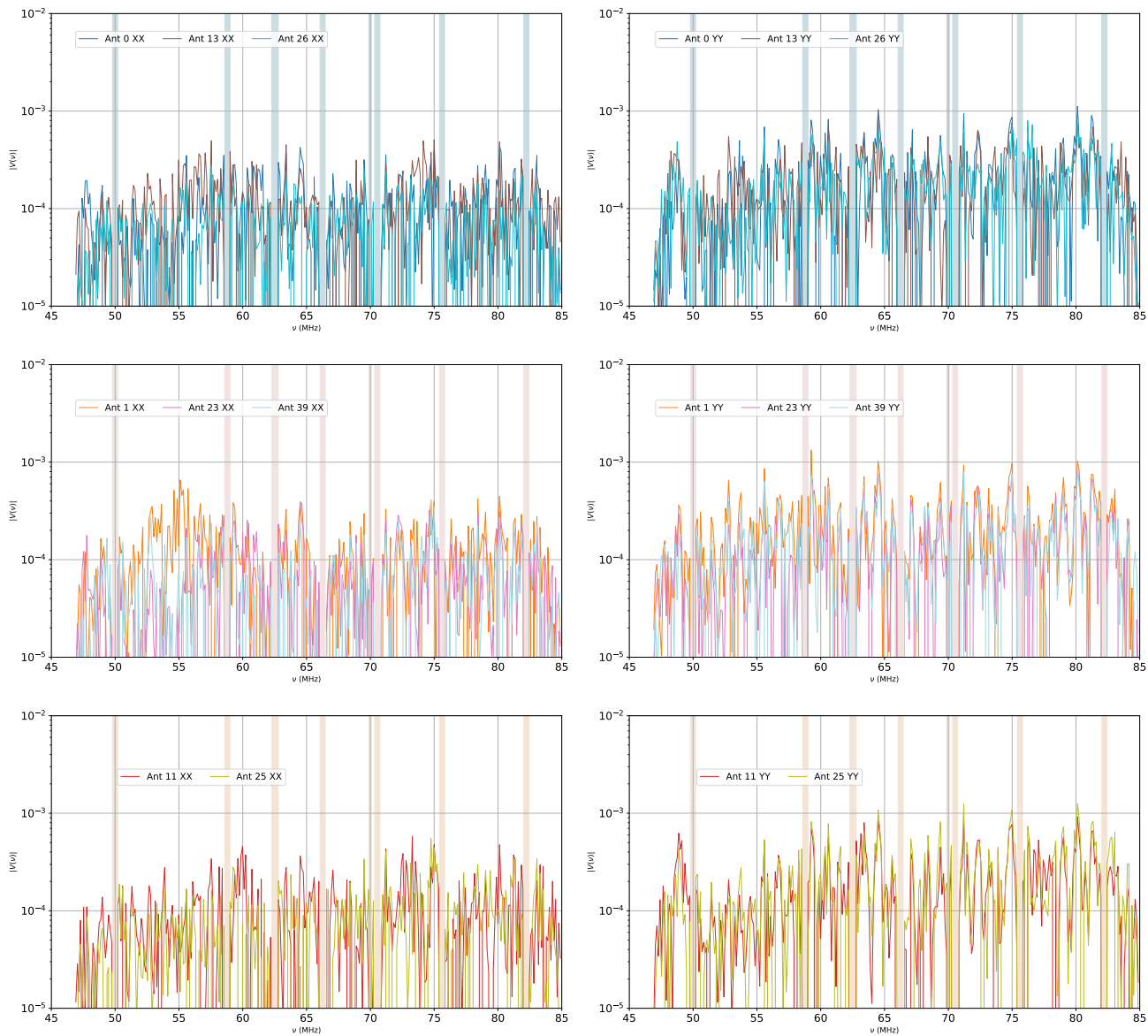
Figure 16: Left, Comparisons between time averaged frequency domain residuals on all antennas over the low band (45-65 MHz). Right: The same for YY.
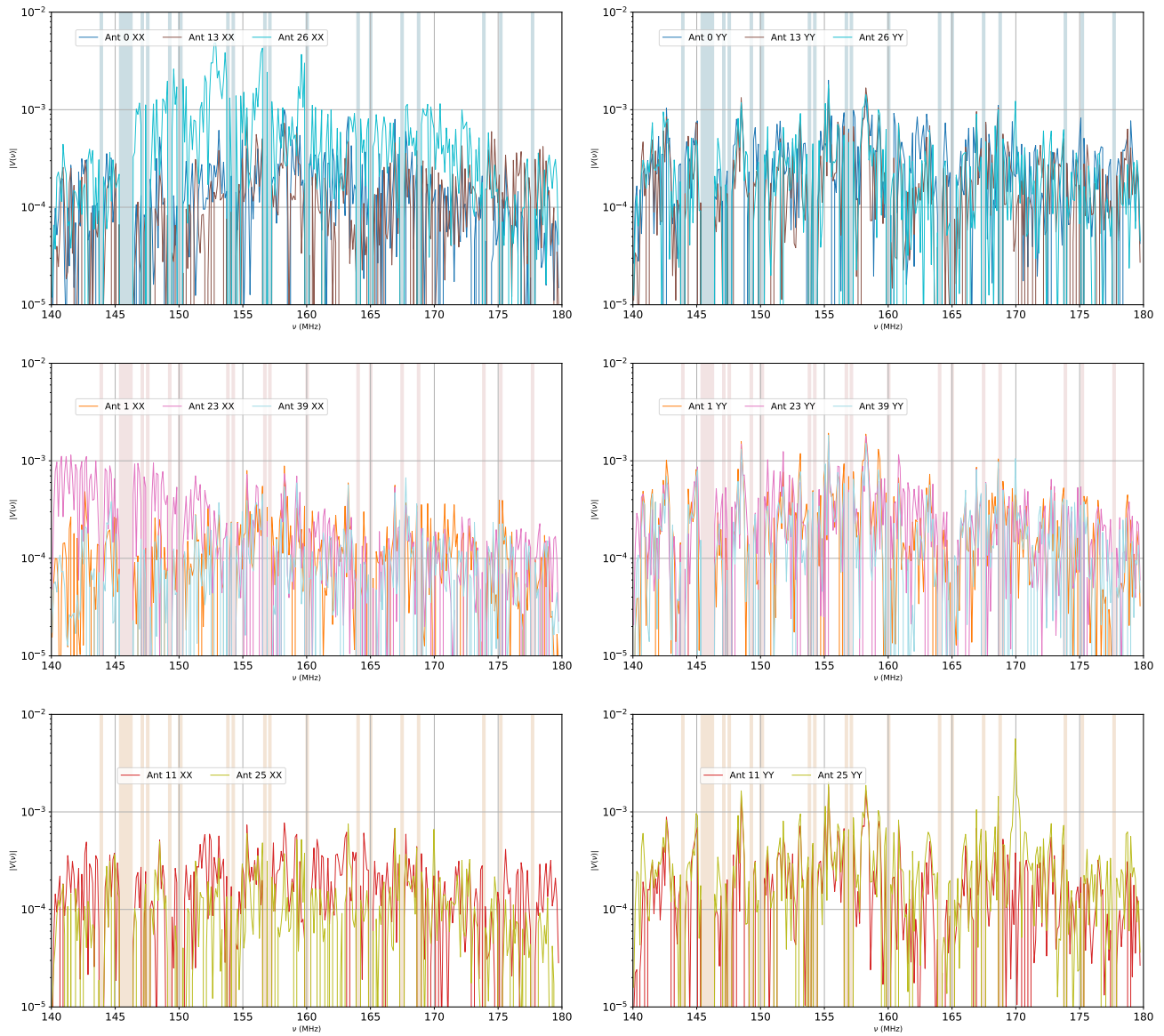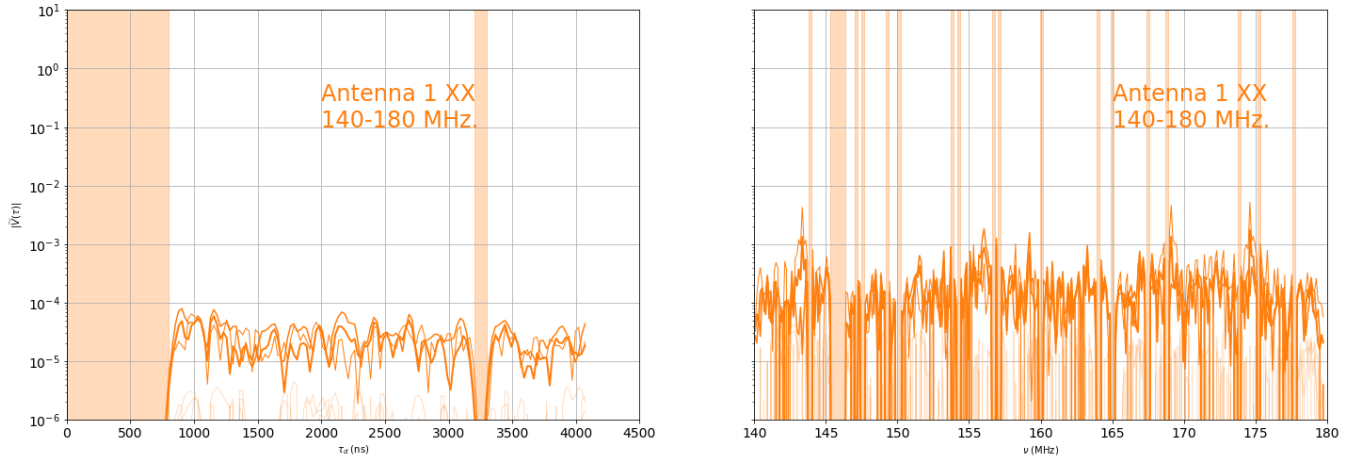
Figure 17: Same as Fig. 16 but now for the high band (140-180 MHz)

Antenna 1

Antenna 1
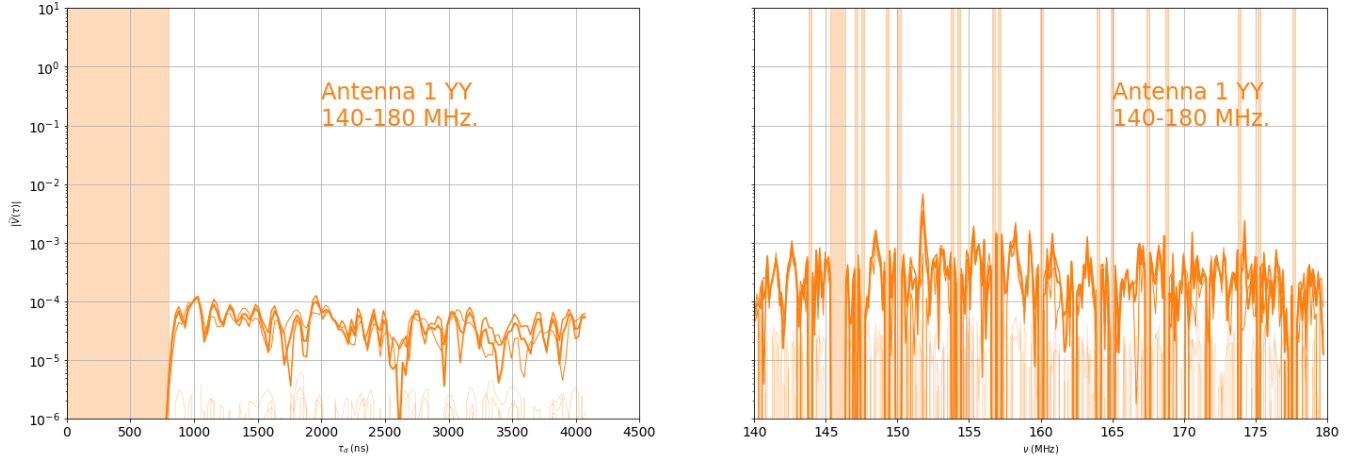
Figure 18: Thick lines – averaged residuals from 9-12 hours LST. Thin lines – averaged residuals from 13-16 hours of LST. Top: XX polarization. Bottom: YY. Many of the detected residual structures are common in both LSTs, suggesting that the systematics floor is formed by persistent sources (rather than rare transient events with large amplitudes).
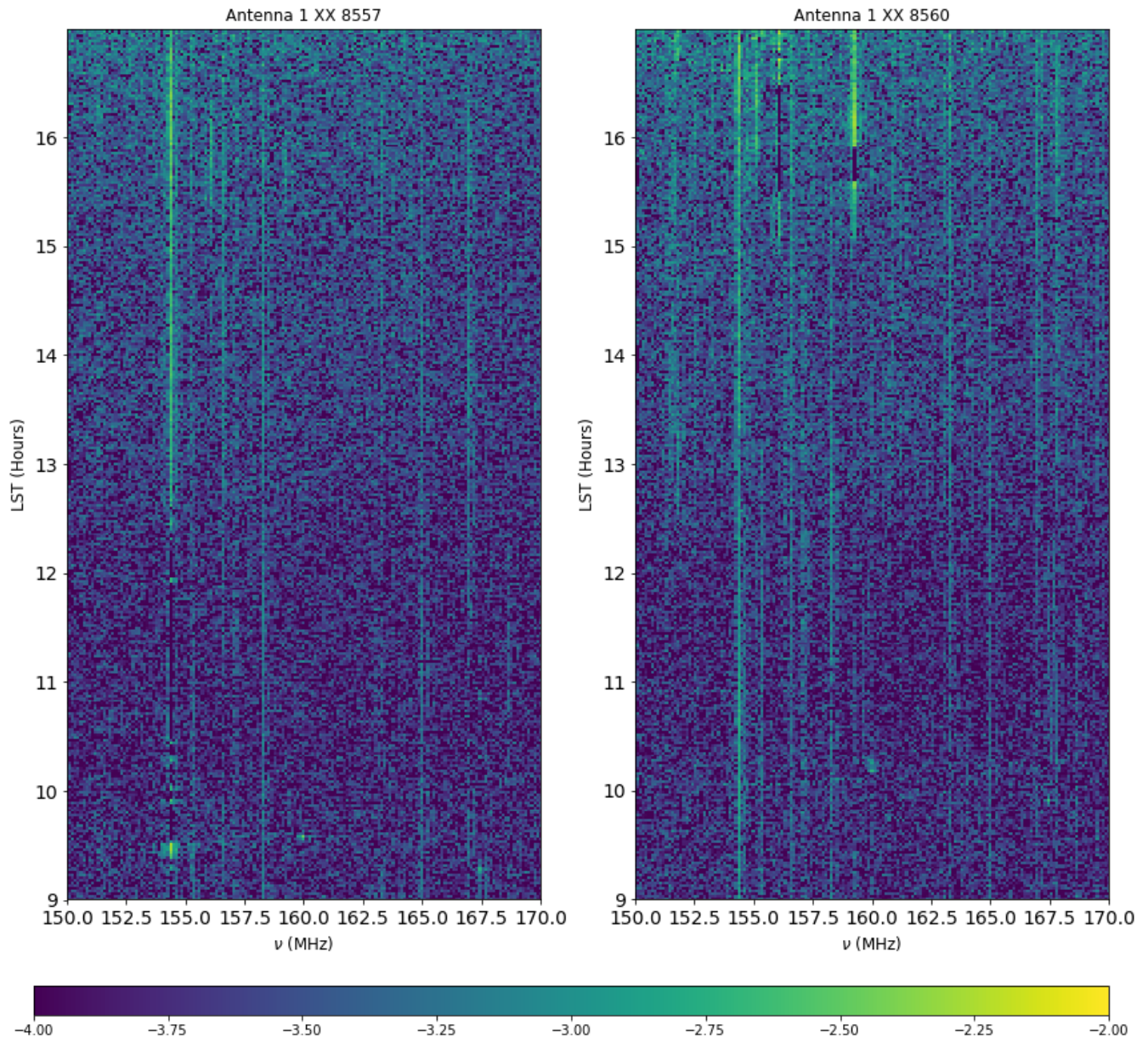
Figure 19: Waterfalls in which all power has been filtered below 1000 ns on XX. Each time cell represents 86 seconds of averaging. Streaks of low-level narrow band power are present in both days of observations and do not appear to drift in frequency which is indicative of persistent low-level RFI.
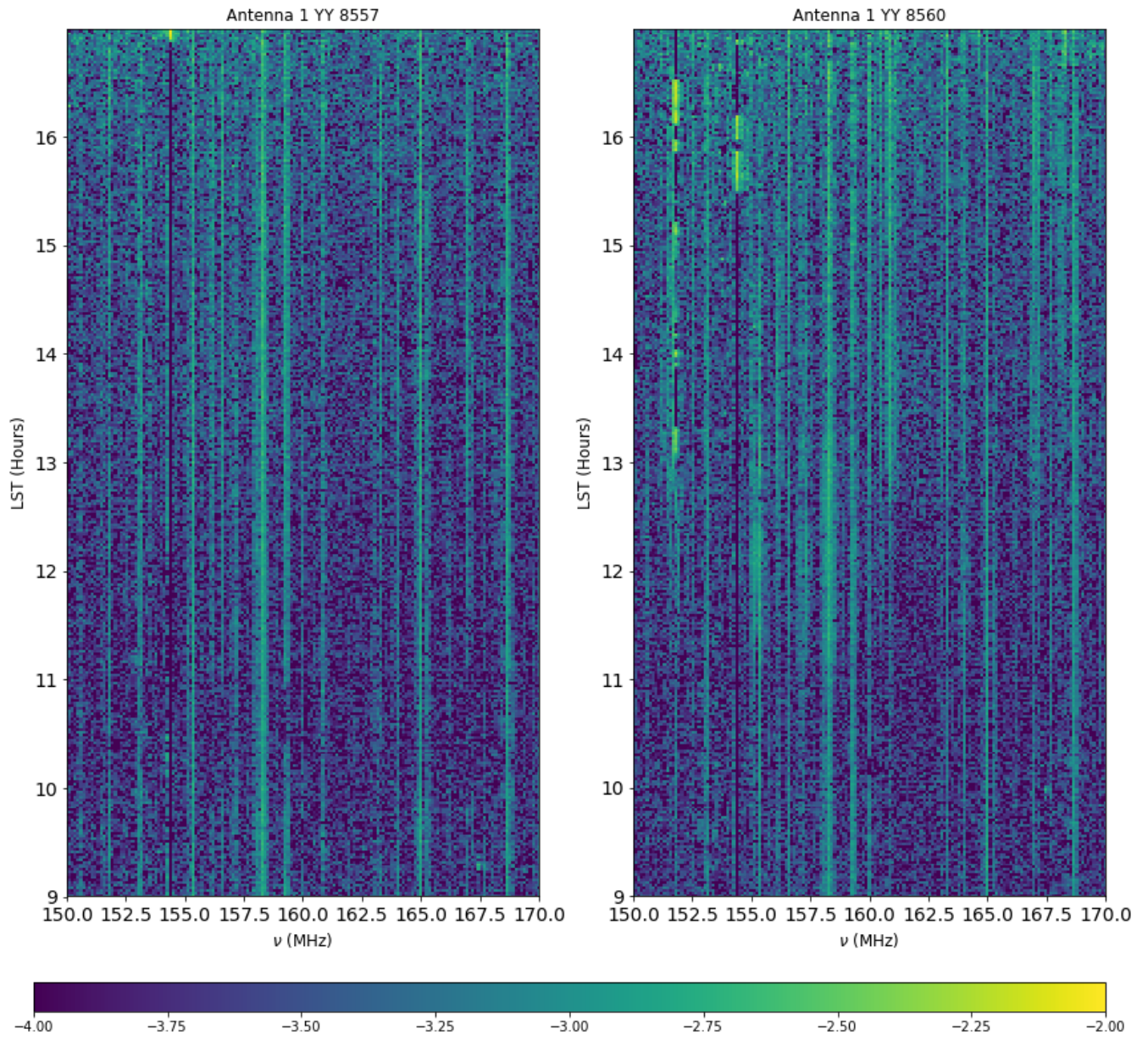
Figure 20: Same as Fig. **??** but the YY polarization.
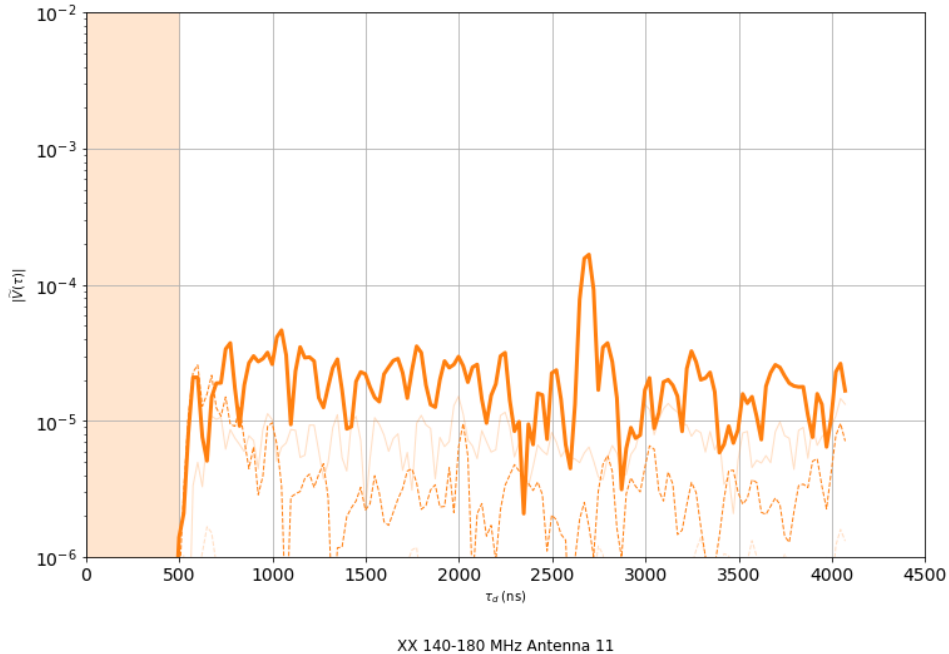
Figure 21: 1 hour averaged delay-domain linear filter residuals with the noise source on (dashed) and the noise source off (solid). Delay-residuals off the sky are 10dB lower, similar to the frequency domain results in Fig. 30. Interestingly enough, the reflection-like spike at 3250 ns is not present when the noise source is on.

6.0.6 Tests with Noise Diodes.

On JD 2458590, noise sources are activated in antennas 0, 1, 2, 13, 23. The rest of the antennas appear to see the sky. I compare waterfalls between (a) two antennas that see the sky from 2458560 and 2458590 along with (b) an antenna on 24258590 with a noise source versus the same antenna on 2458560 when it looking at the sky. In Fig. 28, we inspect residuals form antenna 11 on days 2458560 and 2458590. On both of these days, antenna 11 saw the sky. The same streaked pattern in frequency repeats on 2458560 and 2458590. We Then compare an antenna with a noise source on day 8590 to 8560 when it did not have a noise source (antenna 1 YY) in Fig. 29. On day 2458560, the streaked pattern is nearly identical (sans noise) to the pattern seen by antenna 11 YY on days 2458590 and 2458560. On day 2458590, when antenna 1 YY is hooked up to a noise source, we do not observe the same pattern but rather a different set of residuals that drift slightly in frequency (with time) and are spaced by $\approx 2.5\,\mathrm{MHz}$. These lines are likely the charge pump spikes.

An a later date, JD2458605, observations were taken where the noise source is chopped over 30 minute intervals. We inspect the on/off linear clean residuals (normalized to the on/off average amplitudes of unfiltered data) in fig. 30. Residuals with the noise sources on are roughly 5-10below residuals with the noise sources off and are dominated by a different set of features that drift in frequency, similar to the charge pump spikes observed in earlier iterations of the system. **Since the spikes setting the -45 dB floor that dominate residuals when the noise source is off are not present when the noise source is on, we conclude that these spikes arise upstream of the noise sources and may be coming from the external environment.** The same trend is mirrored in the delay domain (Fig. 21) where the delay-residuals are now down to -50 dB. In addition, the reflection-like feature that was present in the on-sky data is not present in the noise diode data, indicating that this feature is not a typical reflection caused by an impedance mismatch.

6.0.7 What RFI power would cause low-level spikes?

Given a model of HERA's system temperature, it is possible for us to obtain a rough determination of the power within a transmitter that would cause RFI at the level that we observe in our data. Suppose we detect an RFI streak at the level of SNR times the noise level. If an isotropic transmitter emits power with an amplitude of P_i over a
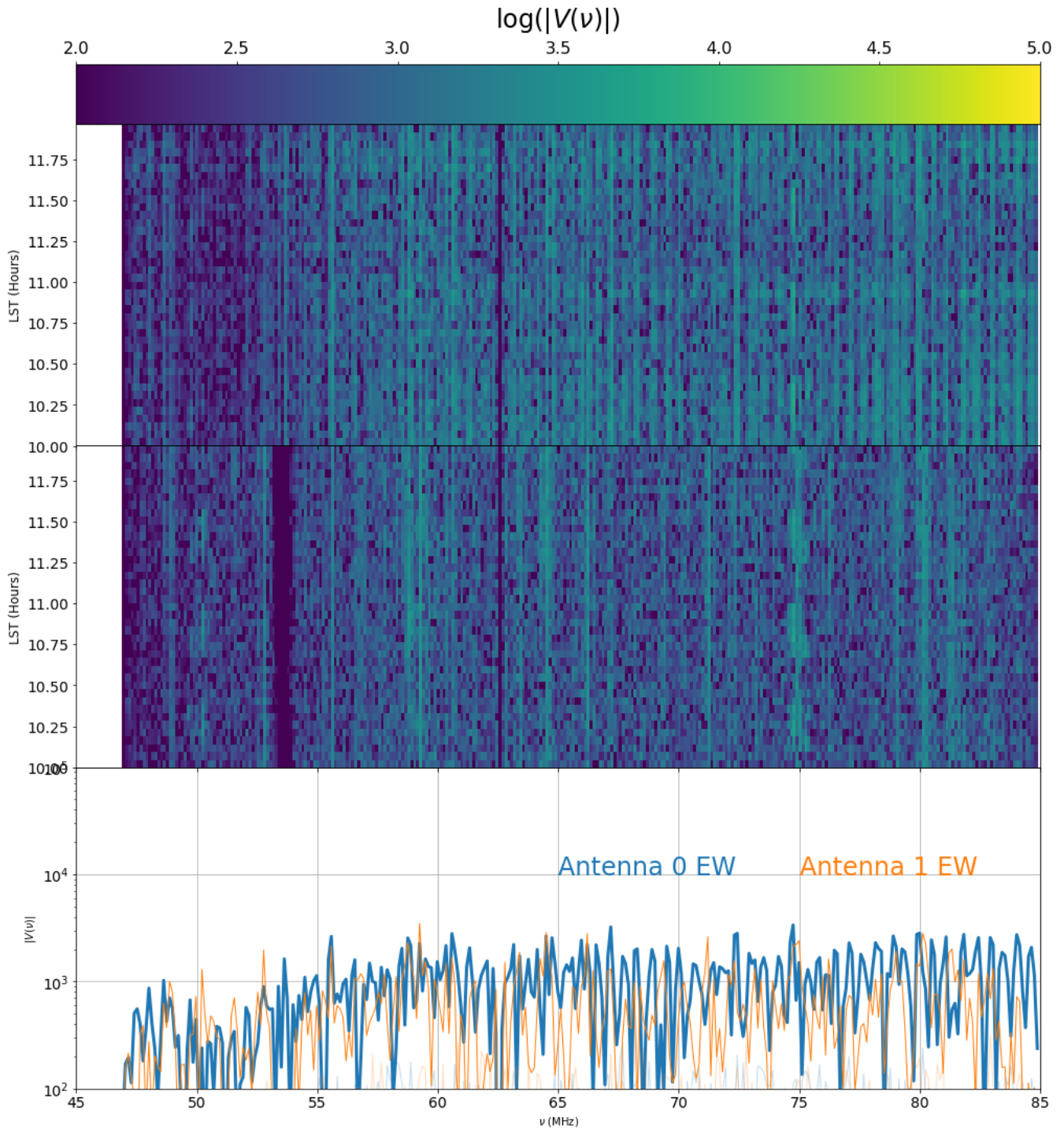
26

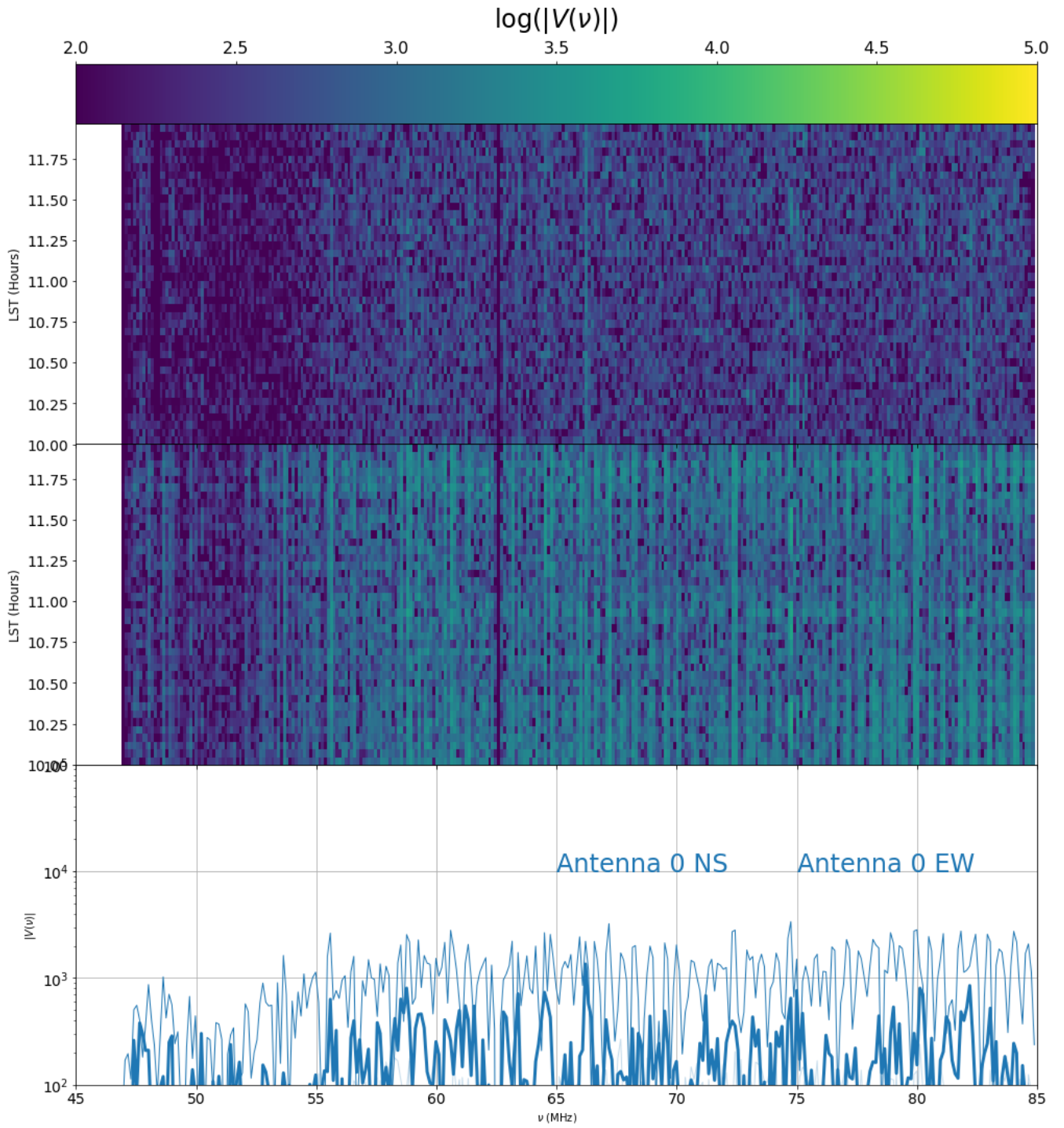Figure 22: a jackknife of filtered residuals across multiple antennas low band.

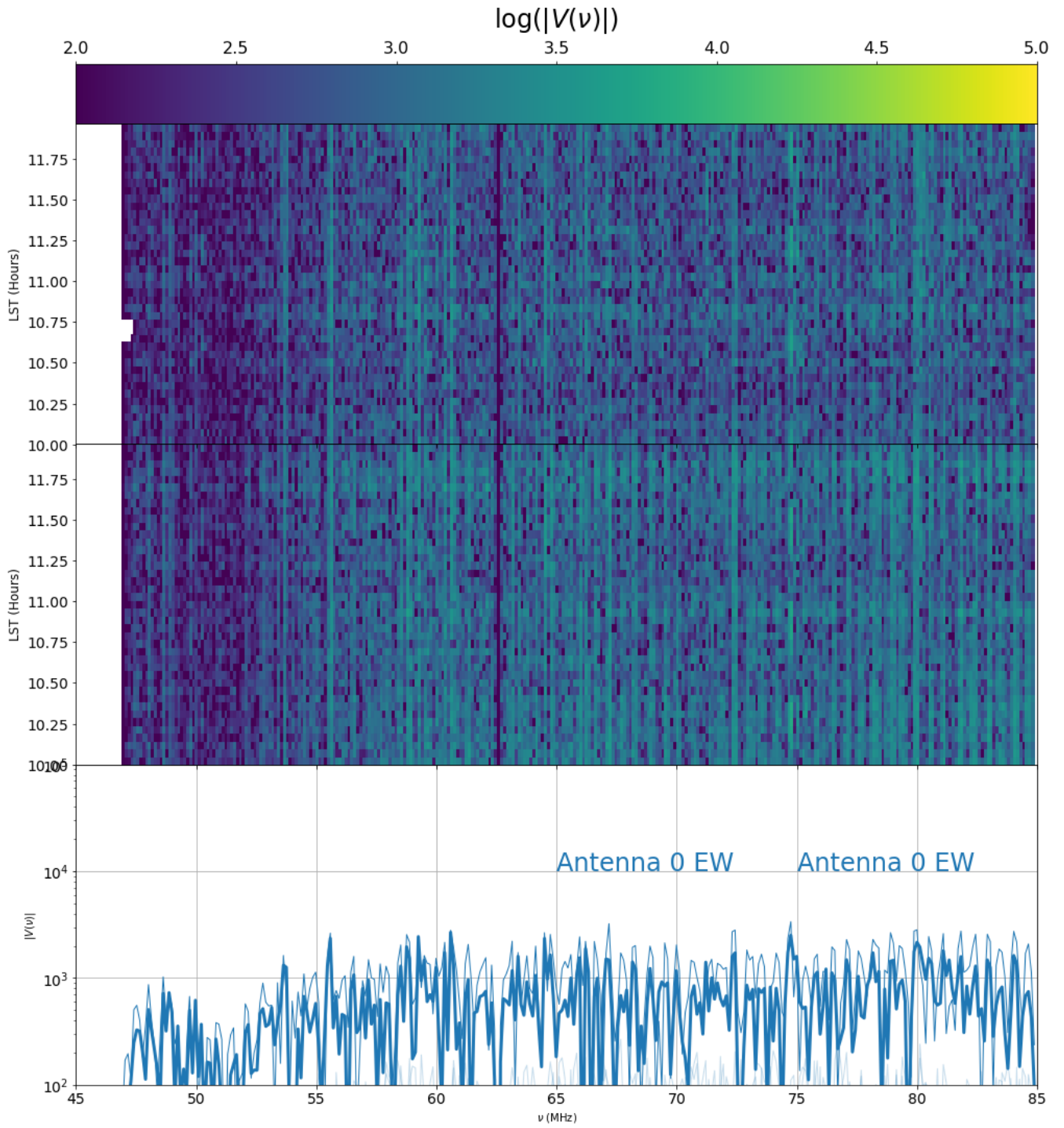Figure 23: a jackknife of filtered residuals across multiple polarization low band

Figure 24: A jackknife of filtered residuals across multiple days low band
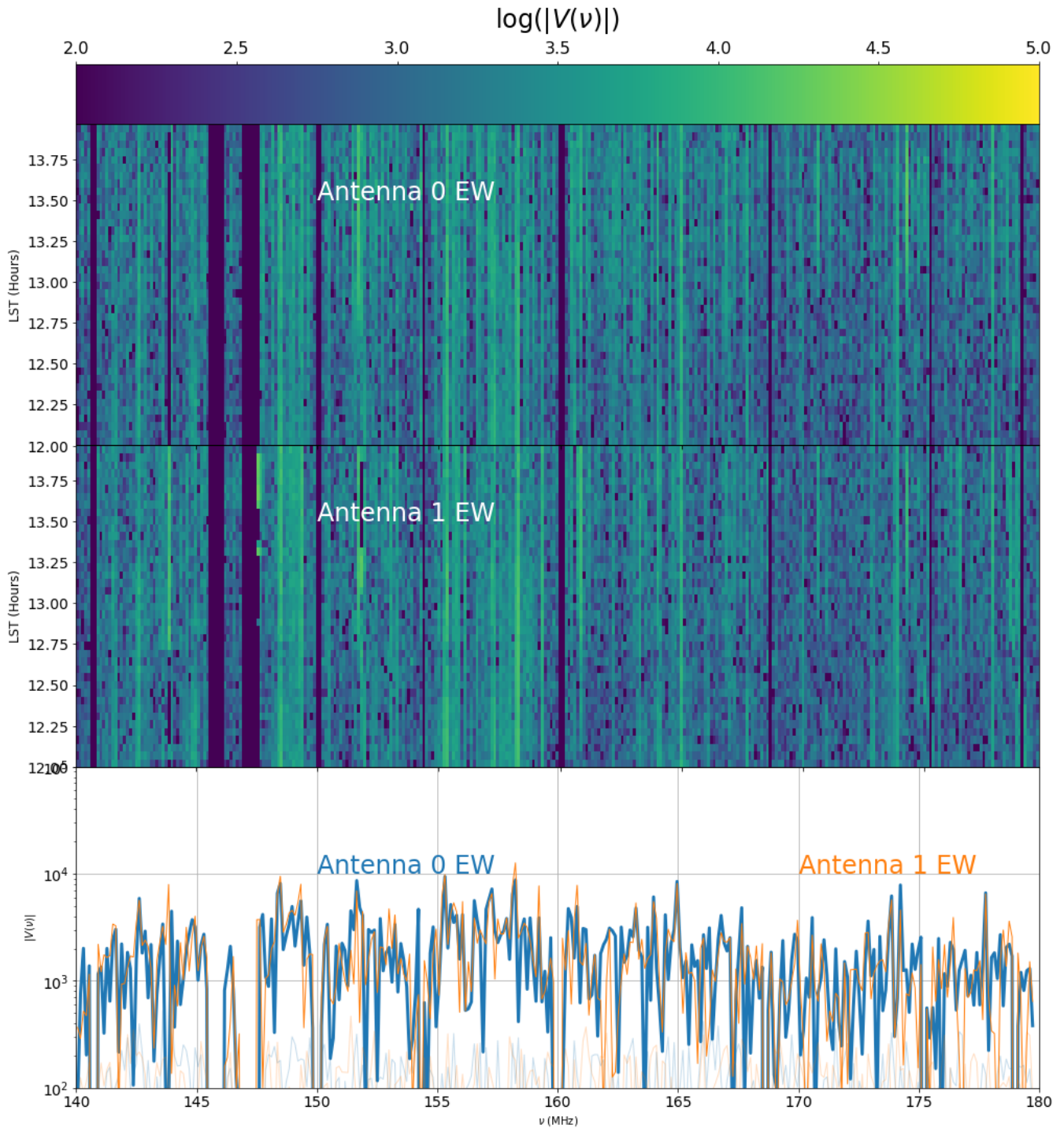
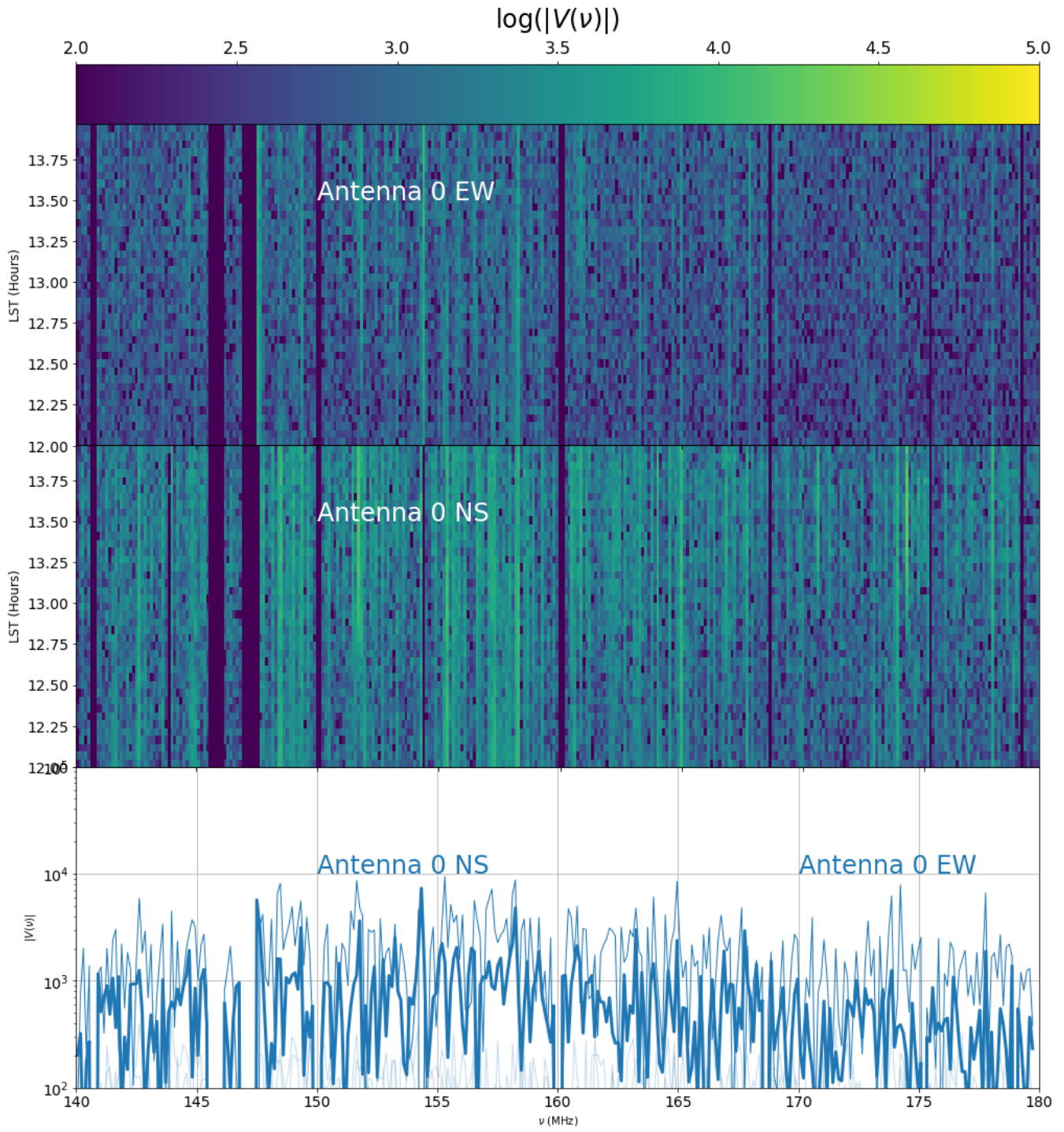Figure 25: A jackknife across antennas of filter residuals high band.

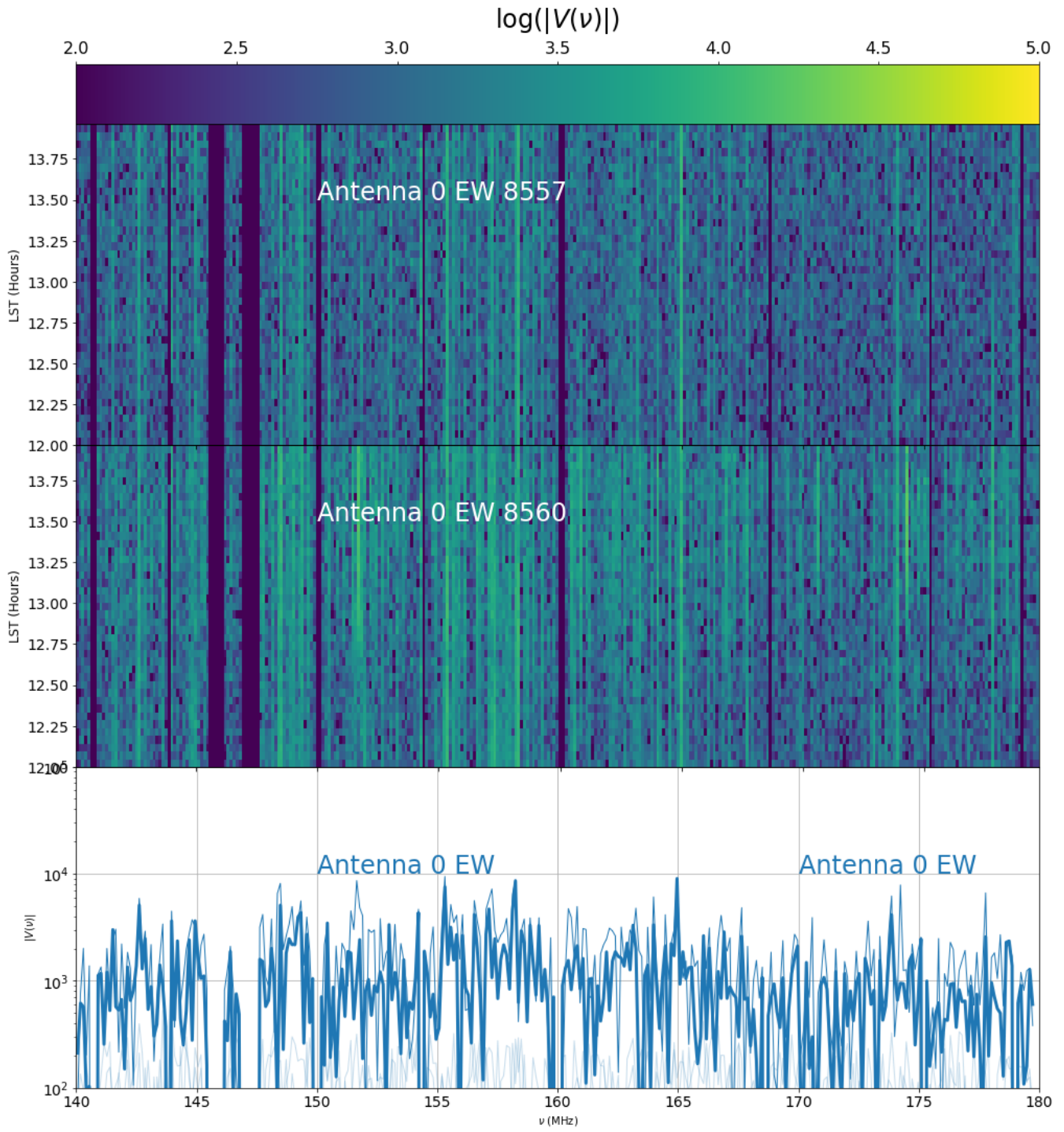Figure 26: A jackknife across polarizations of filter residuals high band.

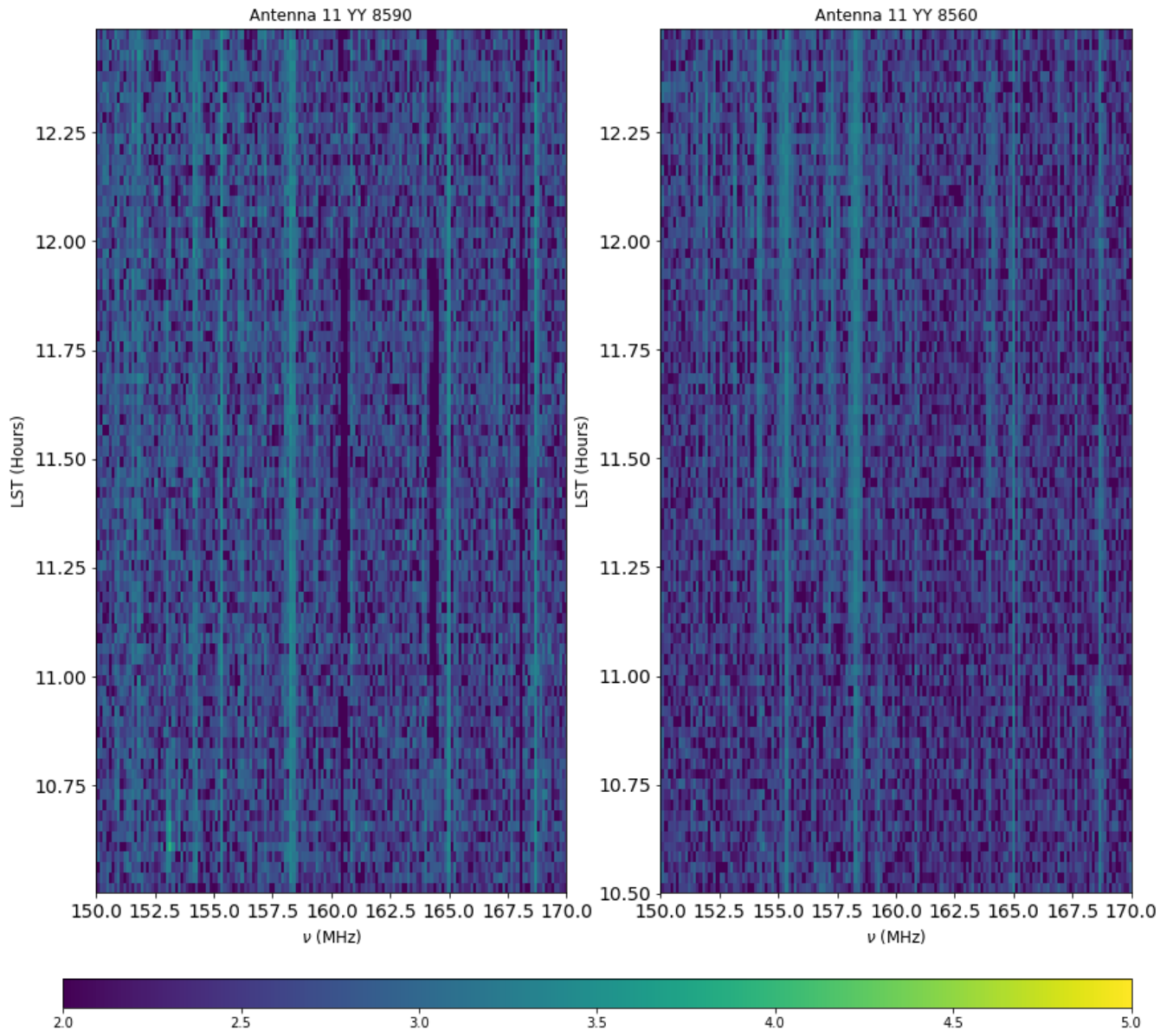Figure 27: Jackknife of filter residuals across multiple days high band.

Figure 28: Comparisons between day 8590 and 8560 on antenna 11, where the antenna sees the sky on both days. Vertical streaks repeat over both days.
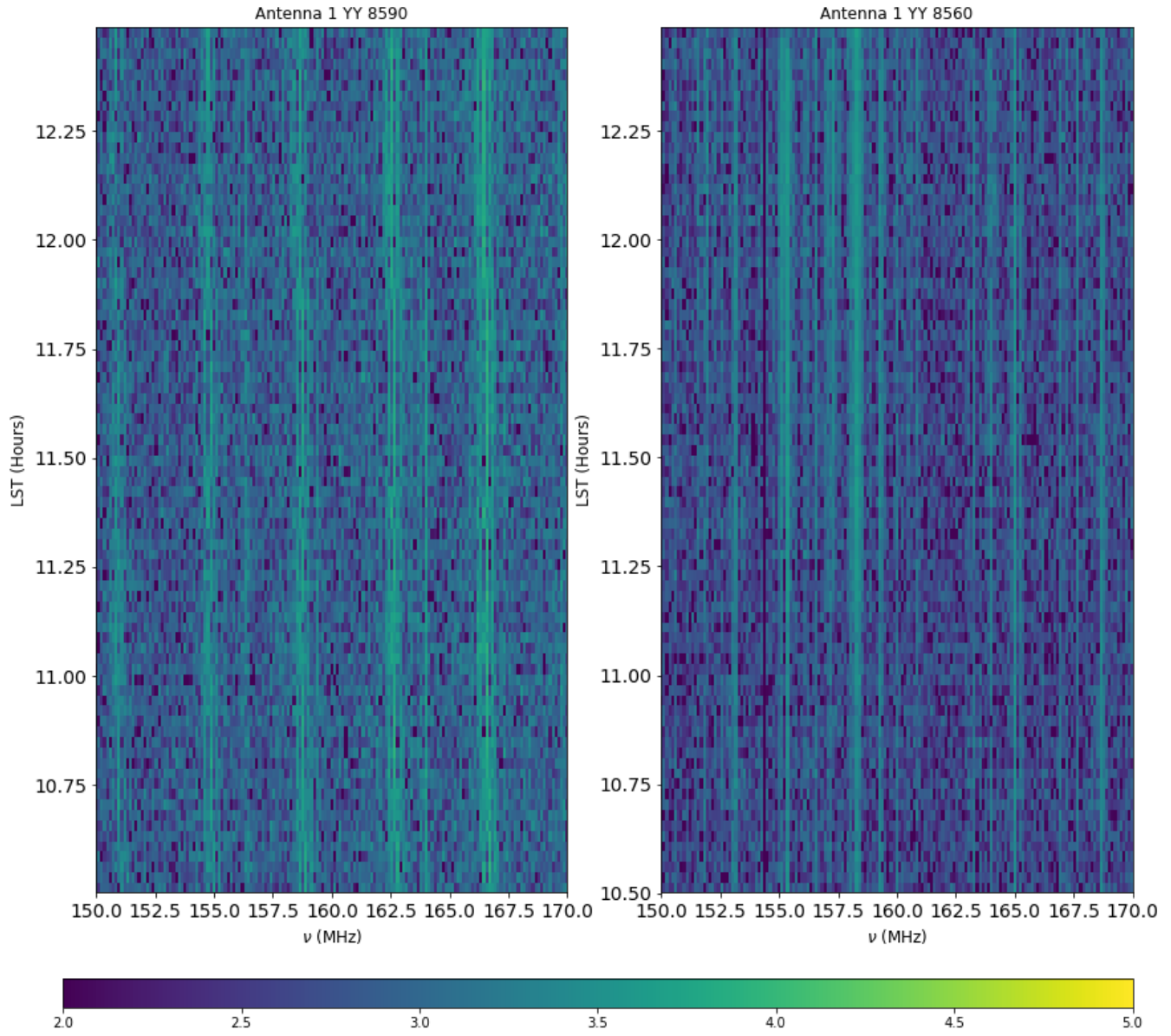
Figure 29: The same as Fig. 28 but now focusing on antenna 1 which has a noise source connected on 8590. We see the same streaking pattern is different on 8590 and drifts, as has been observed in the laser charge pump spikes.
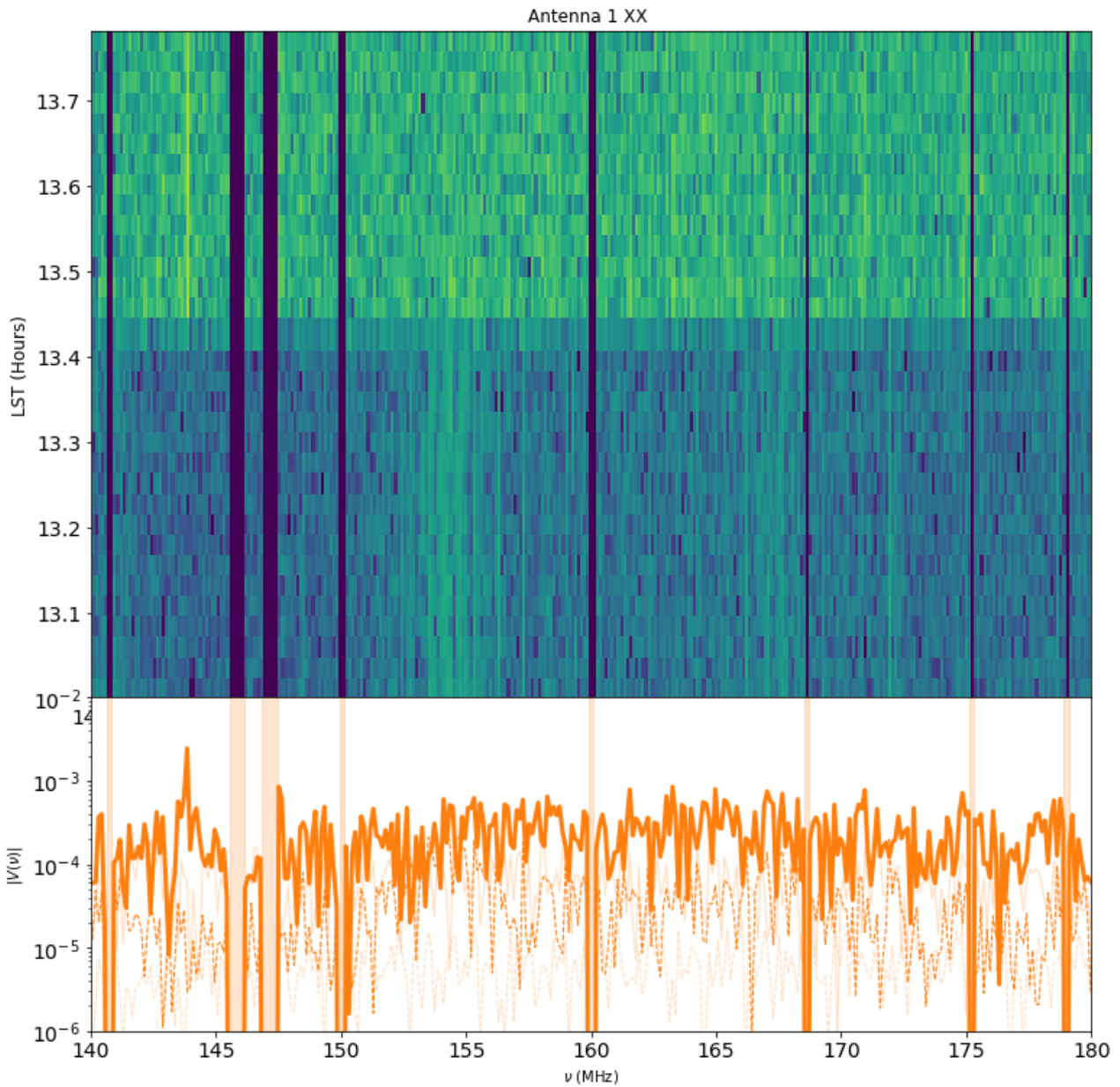
Figure 30: Top, a waterfall of frequency domain linear clean residuals from antenna 1 XX taken over an hour on JD 2458605 when the antenna was chopped between noise source and sky over 30 minute intervals. Residuals before and after noise source is turned on are normalized to the overall average value of the data over the on/off interval. We see that the amplitude of residuals when the noise source is turned off is on the order of 5-10 dB lower and dominated by a different set of systematic then the sky (namely drifting noise spikes).
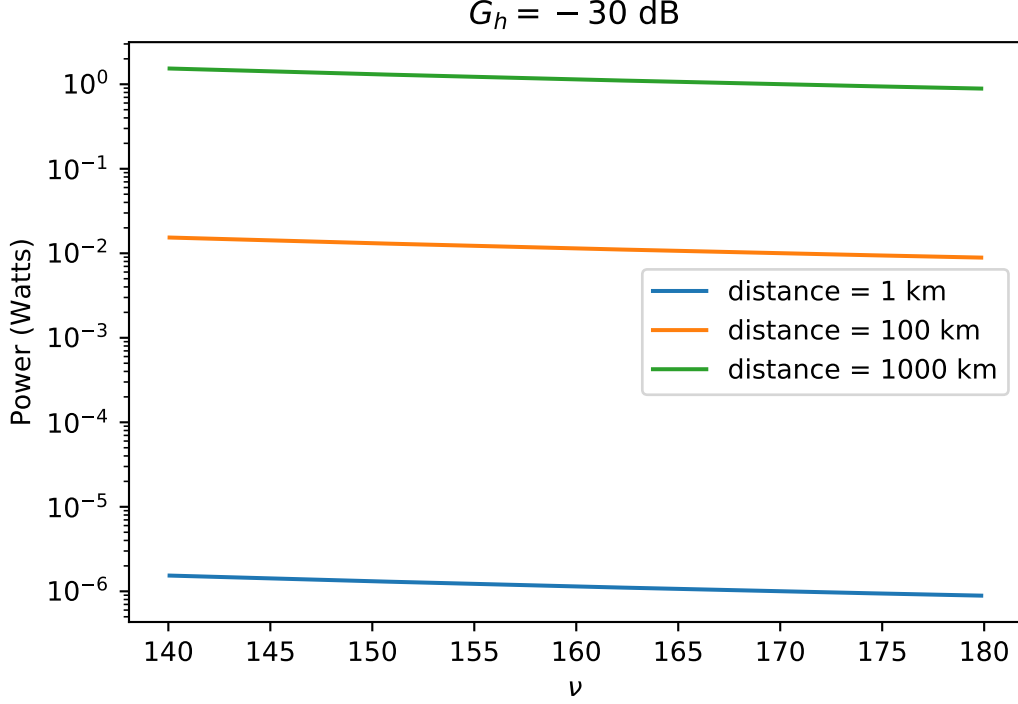
Figure 31: The rough power level necessary for an RFI transmitter to appear at SNR=3 in our autocorrelation observations with $\Delta\nu = 122\,\mathrm{kHz}$, and 86 seconds of integration.

bandwidth of $\Delta\nu_i$ and is located R_i away from the array and RFI enters the antenna through the dish side-lobes with an amplitude of G_h relative to zenith, then

$$G_\mathrm{h}\frac{P_i}{4\pi R_i^2 \Delta\nu_i} = \mathrm{SNR}\frac{k_B T_\mathrm{sys}}{\sqrt{\Delta\nu\tau}A_\mathrm{eff}} \tag{25}$$

so

$$P_i = \mathrm{SNR}\frac{k_B T_\mathrm{sys}}{\sqrt{\Delta\nu\tau}}\frac{4\pi R_i^2}{A_\mathrm{eff}}\Delta\nu_i \tag{26}$$

where T_sys is the system temperature, A_eff is the antenna effective area, $\Delta\nu$ is the observing bandwidth, τ is the integration time. I obtain an order of magnitude estimate of P_i by setting $\tau = 86$ seconds, $A_\mathrm{eff} = \pi(14.6/2\mathrm{m})^2$, $\Delta\nu = \Delta\nu_i = 122\,\mathrm{kHz}$, and $G_\mathrm{h} = -30\,\mathrm{dB}$. We see that a transmitter that is emitting the equivalent of one Watt/Sr in the direction of HERA's side-lobes from a distance of $1000\,\mathrm{km}$ could conceivably produce RFI spikes similar to what we observe. Transmitters for radio stations can often be significantly more powerful (tends of kW) and might be observable at far greater distances.

6.0.8   Is the RFI being generated by Antenna 2?

Short answer – probably not. See Figure 32

6.1   Exploring the impact of RFI flagging strategies

The preferential polarization and spiked morphology suggest that the systematic floor is contributed to by some form of external RFI. The flagging procedure above used manual flagging by which flags were inserted by visual inspection. This sort of flagging can miss intermittant events in time and potentially be the cause of the RFI systematics floor so in this section we experiment with more agressive flagging strategies. First, we perform flagging by running the `H1C` flagging algorithm only autocorrelations only ("autos"). Next, we use the latest iteration of flagging operating on all baselines in the 12 antenna dataset but do not apply any thresholding on the flags (in thresholding, channels and time steps with a certain fraction of flags are entirely flagged). We dub this method "all". Finally, we employ a
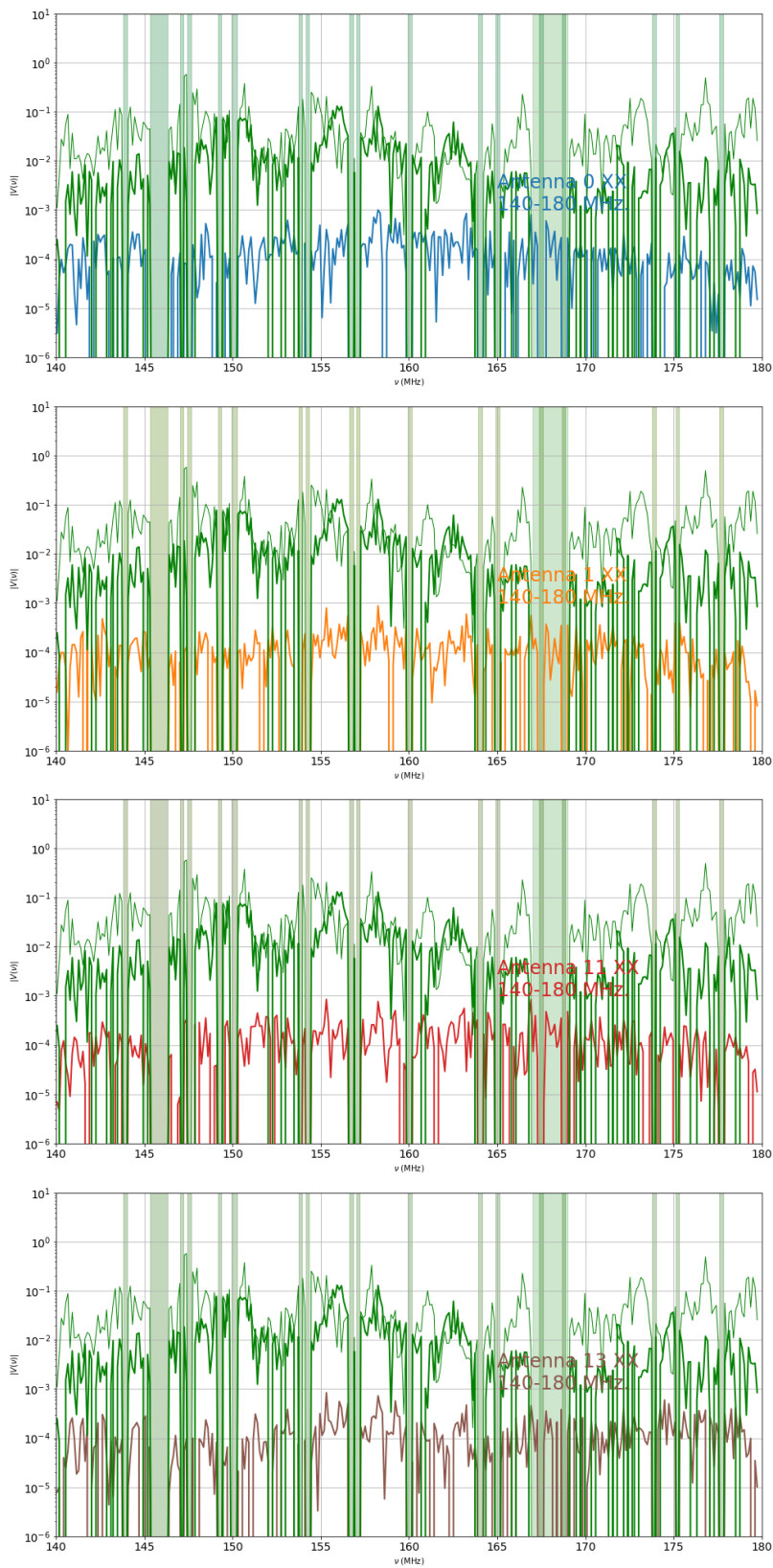
Figure 32: Comparison between the residuals of several of the closest antennas to Antenna 2 and antenna 2 residuals (green) which was thrown out due to the presence of lots of RFI structure but could also potentially be broadcasting. XX and YY are denoted by thick and thin lines for antenna 2. Since the peaks in neither of the pols of antenna 2 line up with peaks in antenna 1, it does not appear that antenna 2 is a major contributor to the residual floor present in the autocorrelations.

Hypothesis	Test	Result	Fig
Narrow-band feature associated with "divot"	Truncate band and mitigate divot in antenna 39 by grounding springs.	Wings are present in antenna 39 even though the divot is out of band, suggesting a broad-band feature that the divot may be associated with. Truncating the band only helps in a subset of antennas and does not correlate with the presence of a divot. Antenna 11 has a clear divot at 55 MHz but truncating the band does not help with the wings so the divot is not a clear exclusive cause of the wings although frequency domain residuals tend to be dominated by it.	5
Cross-coupling	Obtain autocorrelations from an antenna that is isolated from the others.	**This test has not been run**	

Table 3: A table of hypotheses for the causes of the low-delay reflections systematics that are especially prominant in the low band. Update: This is probably cause by springs which have been removed from above the dish (credit Nicholas Fagnoni).

very agressive flagging strategy where we throw out all times and frequencies with greater than 20% of data flagged. In Fig. 34 we demonstrate the agressiveness of our various flagging methods by showing there percentage of channels and times flagged on a representative antenna (1 XX).

7 Validation Simulations

How do we know that the biases we are seeing at high delay are not an artifact of our cleaning or filtering procedure? To validate that our analysis does not leave spurious residuals in regions of delay space where the don't actually exist, we repeat our analysis on a set of idealized simulations of auto-correlations generated using the `hera-sim` software package. We first generate a simulated set of auto-correlations assuming unity gains along with the same integration times and LSTs used in our observations. We also apply the same RFI flags that are present in our data (though at first we don't include any simulated RFI). Waterfalls are shown in Fig. **??**. At the cadence of one time-step, there is no evidence of the low level rumble effects, wings, or reflections that we observe in Fig. 4. This supports the idea that these features arise from real properties of the data (such as reflections, digital artifacts, low-level unflagged RFI, etc...) and not from inauspicious choices in our analysis.

7.1 Complicating our Simulations

We experiment with adding non-idealities to our simulations including adding simulated RFI and reflections at both high and low delays. We observe spurious structures appearing data when we (a) don't filter agressively enough at low delays,leaving low-delay power in, (Fig. 37 top panels) and when only low-delay power is entirely removed (Fig. 35 left panels). In either case, not all intrinsic power is removed and because of significant flagging side-lobes, power is bled out nearly $\sim 10\%$ the level of the intrinsic power itself. In particular, the sidelobes of a 10^{-3} level reflection end up at a similar level to the $-45\,\mathrm{dB}$ systematics we are seeing in our data.

Inspired by this observation, we attempt to lower the systematics floor in our data by performing flagging of the RFoF reflections.

8 Validation with data from the EDGEs Experiment

If ever an experiment existed with a smooth bandpass, its EDGEs. Here we validate our filtering method using the spectrum measured by [1]. In this section, we verify whether its actually possible to build an antenna and signal
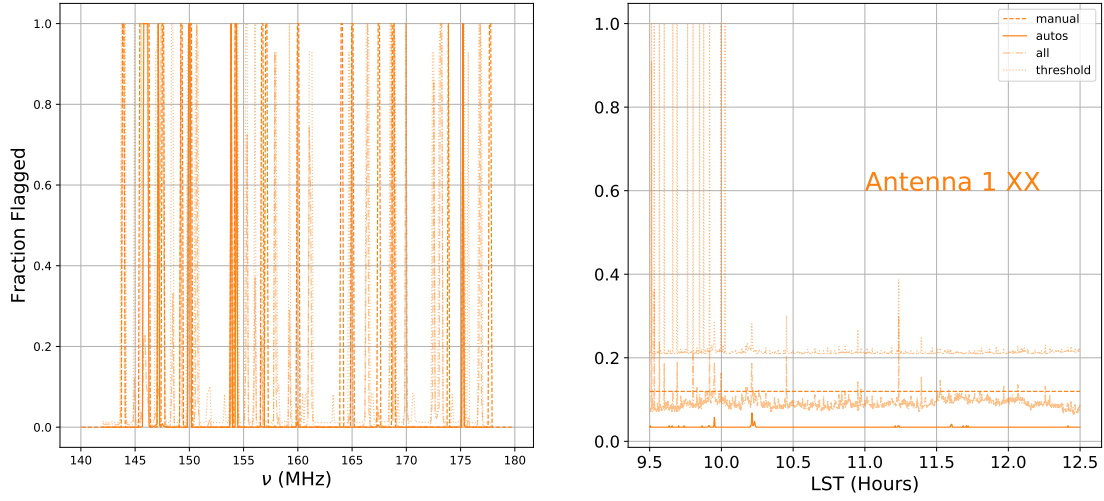
Figure 33: Flagging percentages for a representative antenna (1 XX).
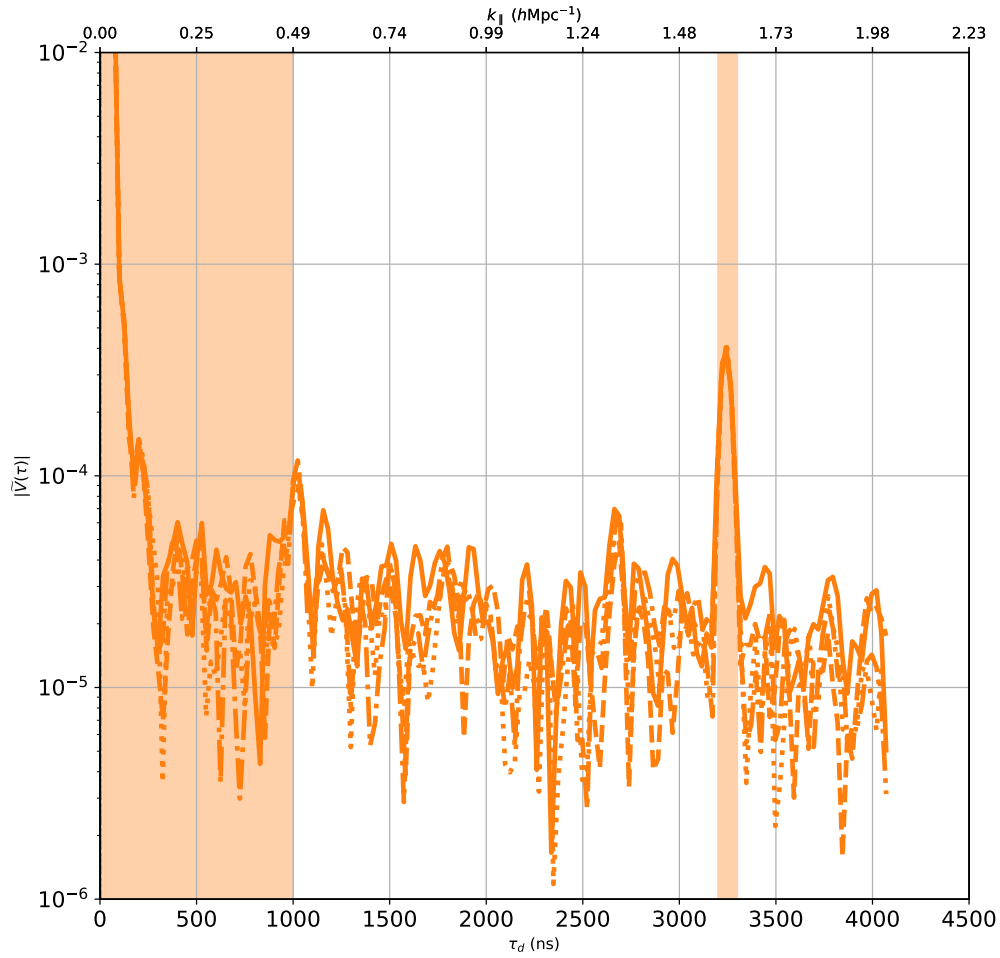

Figure 34: Comparison of clean residuals for various flagging strategies. Manual (dashed), autos (solid), all (dot-dashed), and thresholding (dotted). Various flagging strategies in do not produce significantly different results.
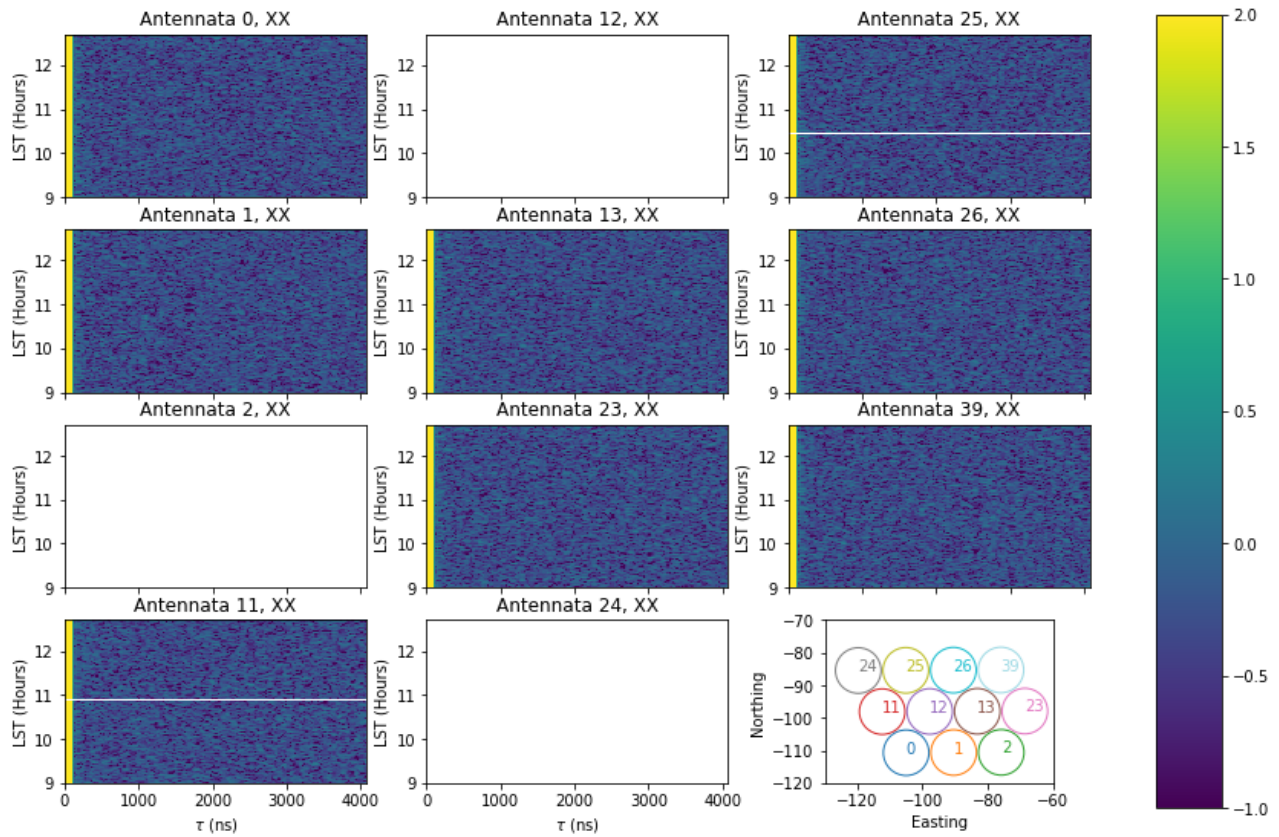
Figure 35: Waterfalls of **simulated** cleaned and restored auto-correlations between 130 and 170 MHz using the same flags and analysis as in Fig. 4 (though no actual RFI is present in the simulation). The simulated cleaned and restored auto-correlations do not show any of the artifacts (low level rumble, reflections, wings etc..) that are present in the actual data.
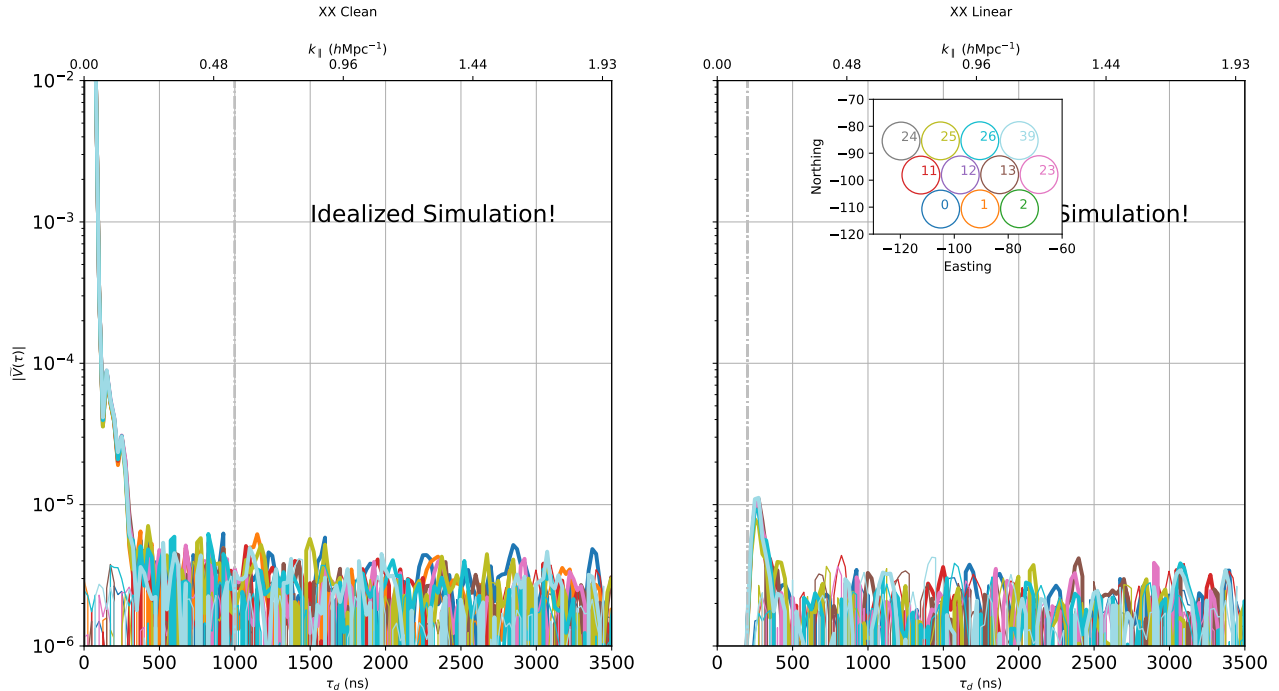
Figure 36: Deep integrations of simulated autocorrelations between 130 and 170 MHz. In the absence of any spectral structure (aside from the variation of the beam in `hera-sim`, the averaged autocorrelations are free of power down to the noise level and are noise like at high delays. We compare the performance of `1d` clean (left) to the linear filtering technique (right). Both techniques filter the autocorrelations down to below the level of the thermal noise beyond several hundred nanoseconds and we do not observe any significant biases.

chain with the delay isolation we require at any point on he surface of the Earth that will pass our auto-correlation test. In Fig. 39 we show the delay and frequency domains of the EDGES autocorrelation before and after filtering within 25 ns. In order to verify that our technique still works when RFI flagging occurs, we generate 20 random RFI gaps. In Fig. 39, we see that outside of 100 ns, the EDGES autocorrelation is consistent with thermal noise. Hence there exists a location on Earth (MRO) along with a smooth and RFI free antenna/signal chain (EDGES) that meet our criterion for 21 cm power spectrum measurements.

9   Comparison with The Old HERA signal chain (H1C)

In this section, we address the question: "Have we done no harm?". In other words, have we made things better or worse when upgrading the HERA system from H1C to H2C which involved replacing the receivers, feeds, and analog signal chains. To answer this question, I perform clean and delay filtering analyses on data from H1C, JD 2458101, employing all hours in this data set. I combine `xrfi` flags with manual flags, and in order to avoid `orbcomm` i compare data between 140 and 180 MHz. I note that the integration times and LSTs in the H1C and H2C data sets are not the same although the noise floors end up being very close to eachother so the comparison is still somewhat objective.

Below 1500 ns, we see that the answer to our hippocratic question is "most definitely yes!". Rippling structures only ≈ 30 dB down are improved dramatically in the new system, probably due to the move from coaxial cable to optical fiber.

On the other hand, beyond 1500 ns, the answer is "no". While the H1C data dips into the noise at ≈ −55 − −60 dB shortly after the coaxial reflections, the H2C data remains at the level of the −45 dB over all delays.

We compare the residuals of H1C and H2C in the frequency domain the better understand the nature of the systematics floor in Figure 41. While the H1C residuals tend to pass through zero frequently, as we might expect for data comprising primarily of noise, the H2C data remains above zero over all non-flagged delays. In addition, the H2C residuals have several prominent spike-like features in frequency while the H1C data does not (though a one or two channels do look RFI like). The morphological differences between H1C and H2C support the theory that there is some additional RFI present in the H2C data that is either generated internally or is environmental that is (a) new
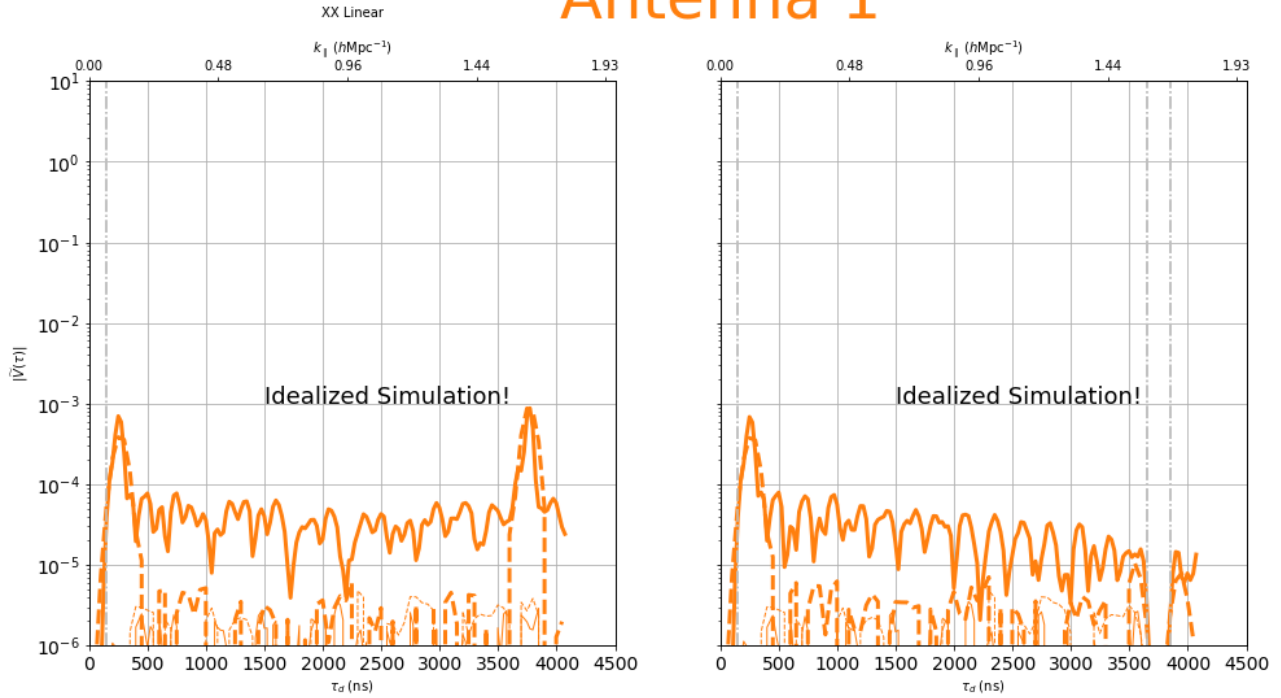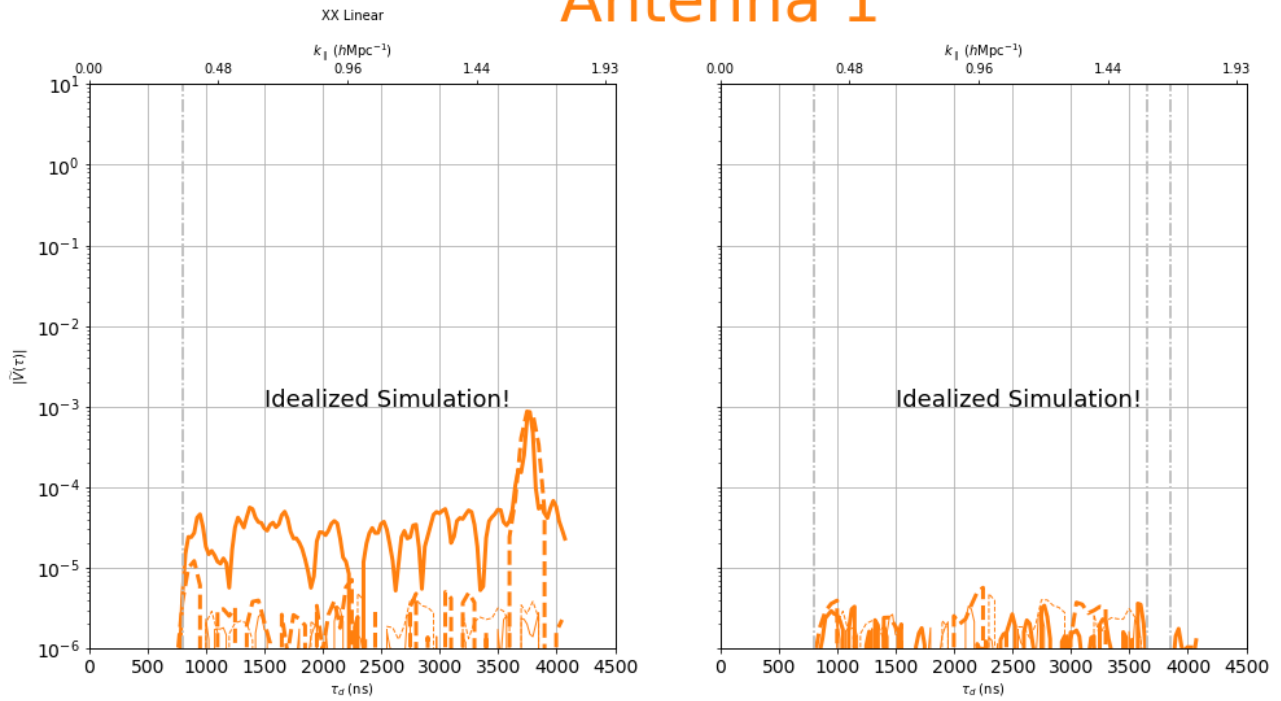
Figure 37: Top Left: A simulated autocorrelation with RFI and low and high delay reflections. Top Left: Filtering is performed out to 150 ns which does not remove all of the power associated with low delay reflections. RFI side-lobes leak power out to high delays across the entire band. A truncated band (dashed line 150-170 MHz) performs better since it does not include Orbcomm. Top Right: The high delay reflection is filtered out but side-lobes remain from the low-delay reflections. Bottom Left: The low-delay filter is extended to 800 ns, allowing it to remove all of the low-delay reflections. Side-lobes from the high delay RFOF reflection still poses a problem. Not flagging or filtering the reflection leads to ≈ 10% sidelobes given our levels of flagging which might explain the $10^{-4.5}$ systematic floor we observe in our measurements. Bottom Right: The same simulated data when a filter that subtracts power within 200 ns of the 3900 ns reflection along with all low delay power. Filtering out the reflection removes this systematic floor. The low-delay foregrounds in the autocorrelation simulation have already been filtered below 1000 ns.
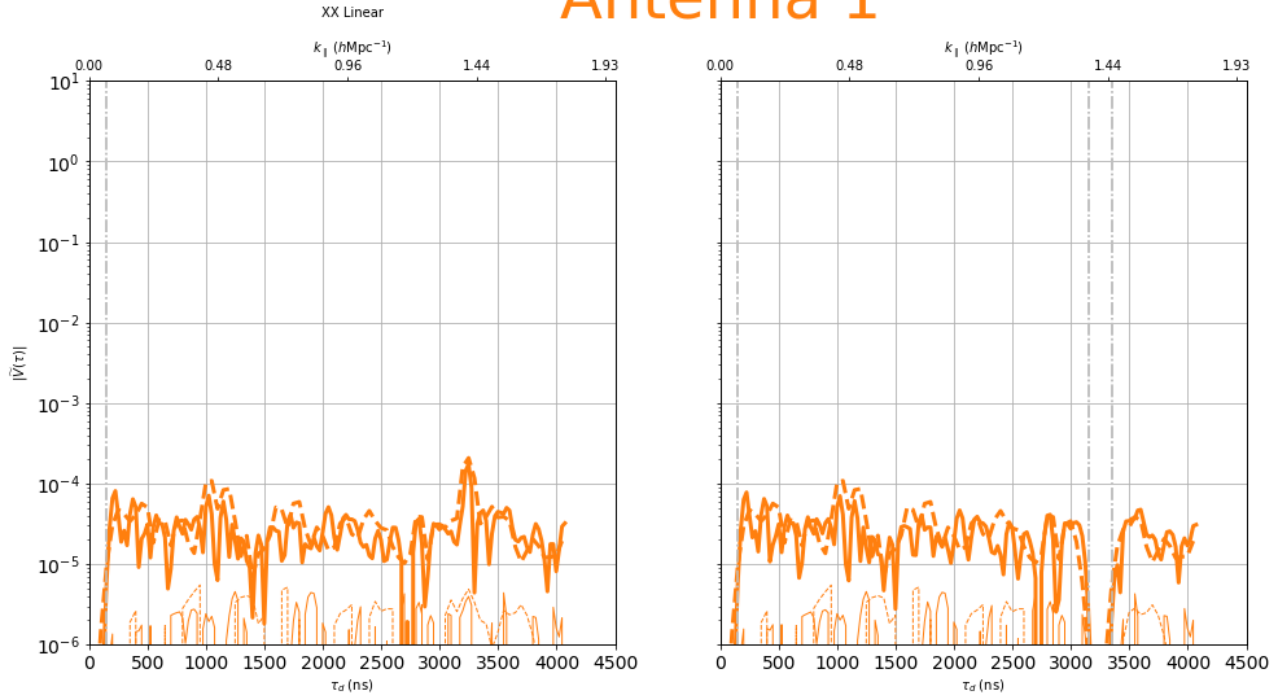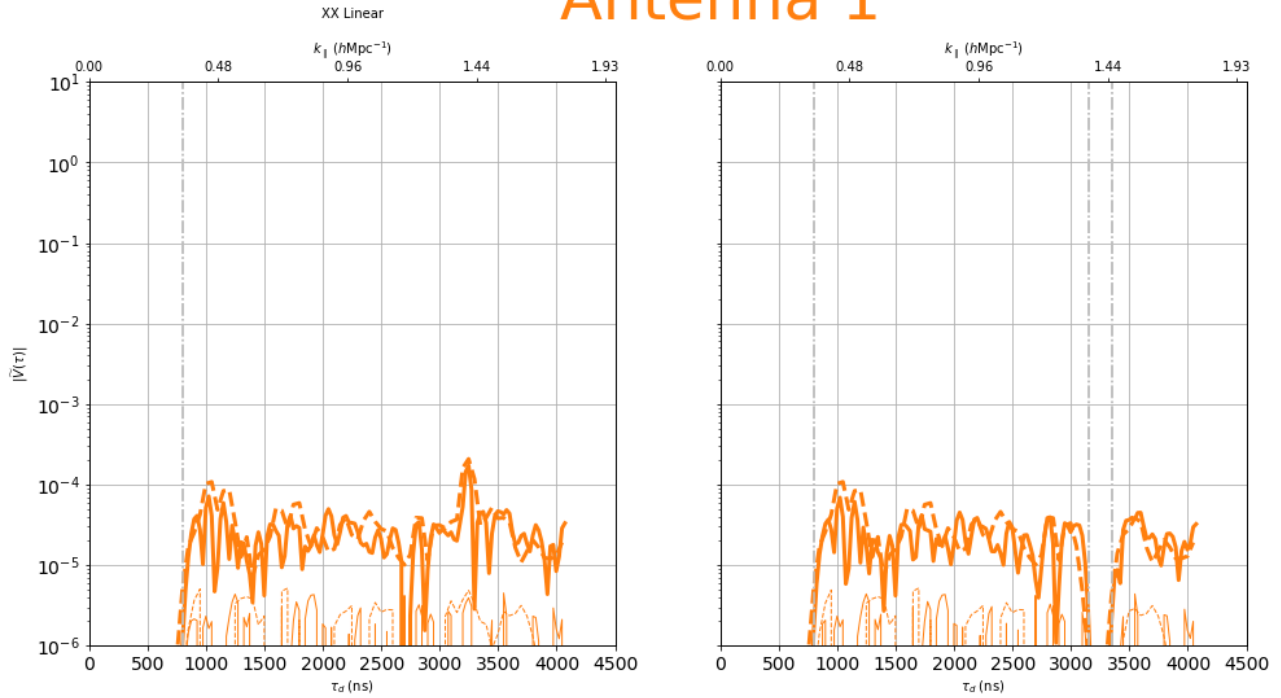
Figure 38: Same as Fig. 37 but now showing actual data from Antenna 1. We see in the lower right that applying filters to all of the reflection regions does not remove the systematics floor. Thus, structure arising from side-lobes from a single high delay reflection and the low delay antenna structures does not explain the origin of the systematics floor.
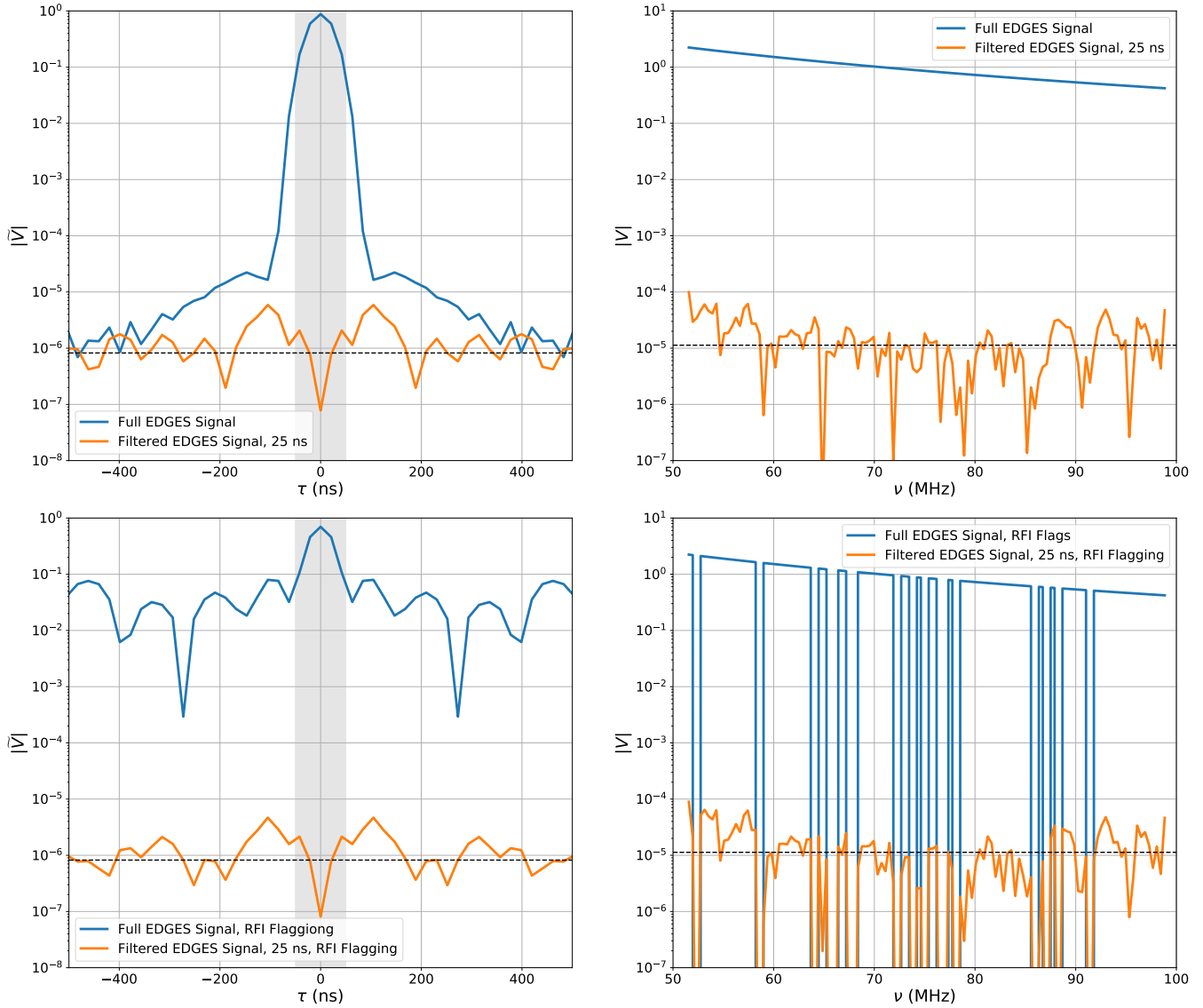
Figure 39: Filtered and unfiltered EDGEs data in the frequency (top-right) and delay (top-left) domain. Black horizontal line denotes the level of residuals after fitting and subtracting the EDGEs model for foregrounds and the 21 cm signal. These plots demonstrate that there do exist locations on the earth that are RF free enough and signal chains that are smooth enough to in principal pass our requirements to measure the 21 cm fluctuations.

(did not exist during H1C) or (b) still existed during H1C but was better shielded by the H1C system (potentially through feeds with lower horizon directivity).

10   Conclusion

References

[1] Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen, and Nivedita Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. , 555(7694):67–70, Mar 2018.

[2] Joshua S. Dillon, Adrian Liu, Christopher L. Williams, Jacqueline N. Hewitt, Max Tegmark, Edward H. Morgan, Alan M. Levine, Miguel F. Morales, Steven J. Tingay, Gianni Bernardi, Judd D. Bowman, Frank H. Briggs, Roger C. Cappallo, David Emrich, Daniel A. Mitchell, Divya Oberoi, Thiagaraj Prabu, Randall Wayth, and Rachel L. Webster. Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data. , 89:023002, Jan 2014.

[3] A. Ewall-Wice, Joshua S. Dillon, J. N. Hewitt, A. Loeb, A. Mesinger, A. R. Neben, A. R. Offringa, M. Tegmark, N. Barry, A. P. Beardsley, G. Bernardi, Judd D. Bowman, F. Briggs, R. J. Cappallo, P. Carroll, B. E. Corey, A. de Oliveira-Costa, D. Emrich, L. Feng, B. M. Gaensler, R. Goeke, L. J. Greenhill, B. J. Hazelton, N. Hurley-Walker, M. Johnston-Hollitt, Daniel C. Jacobs, D. L. Kaplan, J. C. Kasper, HS Kim, E. Kratzenberg, E. Lenc, J. Line, C. J. Lonsdale, M. J. Lynch, B. McKinley, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, Nithyanandan Thyagarajan, D. Oberoi, S. M. Ord, S. Paul, B. Pindor, J. C. Pober, T. Prabu, P. Procopio, J. Riding, A. E. E. Rogers, A. Roshi, N. Udaya Shankar, Shiv K. Sethi, K. S. Srivani, R. Subrahmanyan, I. S. Sullivan, S. J. Tingay, C. M. Trott, M. Waterson, R. B. Wayth, R. L. Webster, A. R. Whitney, A. Williams, C. L. Williams, C. Wu, and J. S. B. Wyithe. First limits on the 21 cm power spectrum during the Epoch of X-ray heating. , 460(4):4320–4347, Aug 2016.

[4] Nicholas S. Kern, Aaron R. Parsons, Joshua S. Dillon, Adam E. Lanman, Adrian Liu, Philip Bull, Aaron Ewall-Wice, Zara Abdurashidova, James E. Aguirre, Paul Alexander, Zaki S. Ali, Yanga Balfour, Adam P. Beardsley, Gianni Bernardi, Judd D. Bowman, Richard F. Bradley, Jacob Burba, Chris L. Carilli, Carina Cheng, David R. DeBoer, Matt Dexter, Eloy de Lera Acedo, Nicolas Fagnoni, Randall Fritz, Steve R. Furlanetto, Brian Glendenning, Deepthi Gorthi, Bradley Greig, Jasper Grobbelaar, Ziyaad Halday, Bryna J. Hazelton, Jacqueline N. Hewitt, Jack Hickish, Daniel C. Jacobs, Austin Julius, Joshua Kerrigan, Piyanat Kittiwisit, Saul A. Kohn, Matthew Kolopanis, Paul La Plante, Telalo Lekalake, David MacMahon, Lourence Malan, Cresshim Malgas, Matthys Maree, Zachary E. Martinot, Eunice Matsetela, Andrei Mesinger, Mathakane Molewa, Miguel F. Morales, Tshegofalang Mosiane, Steven G. Murray, Abraham R. Neben, Aaron R. Parsons, Nipanjana Patra, Samantha Pieterse, Jonathan C. Pober, Nima Razavi-Ghods, Jon Ringuette, James Robnett, Kathryn Rosie, Peter Sims, Craig Smith, Angelo Syce, Nithyanandan Thyagarajan, Peter K. G. Williams, and Haoxuan Zheng. Mitigating Internal Instrument Coupling II: A Method Demonstration with the Hydrogen Epoch of Reionization Array. *arXiv e-prints*, page arXiv:1909.11733, Sep 2019.

[5] N. O'Donoughue and J. M. F. Moura. On the product of independent complex gaussians. *IEEE Transactions on Signal Processing*, 60(3):1050–1063, March 2012.
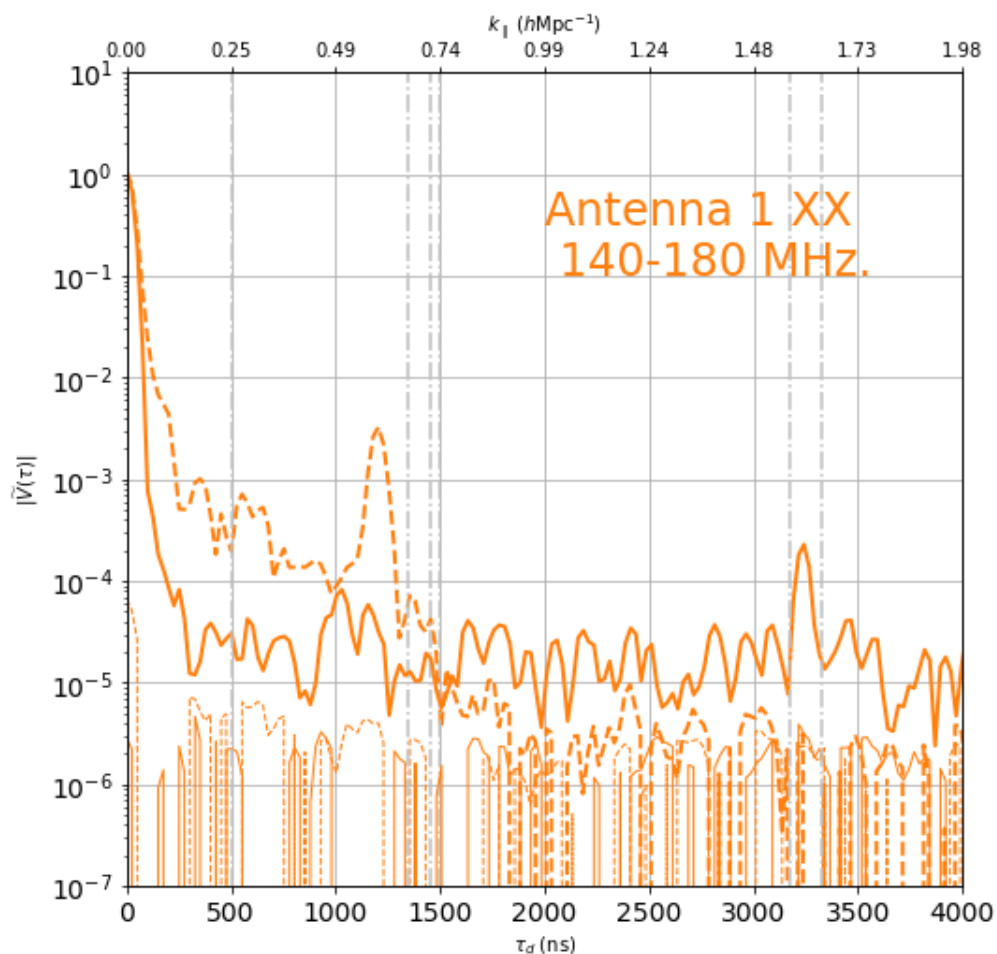
Figure 40: A comparison between restored and cleaned data from antenna over H1C (dashed lines) and H2C (solid lines) in the delay domain. We see that below 1500 ns, H1C does considerably worse than H2C, with residuals on the order of only -30 dB which are likely caused by sub-reflections in the coaxial cables. We note however that some intrinsic power does extend below the coaxial reflection and is above the noise up to ≈ 2000 ns – potentially a continuation of sub reflections due involving the second round trip on coaxial cable. Beyond 2000 ns, H1C data is consistent with noise while H2C data is not.
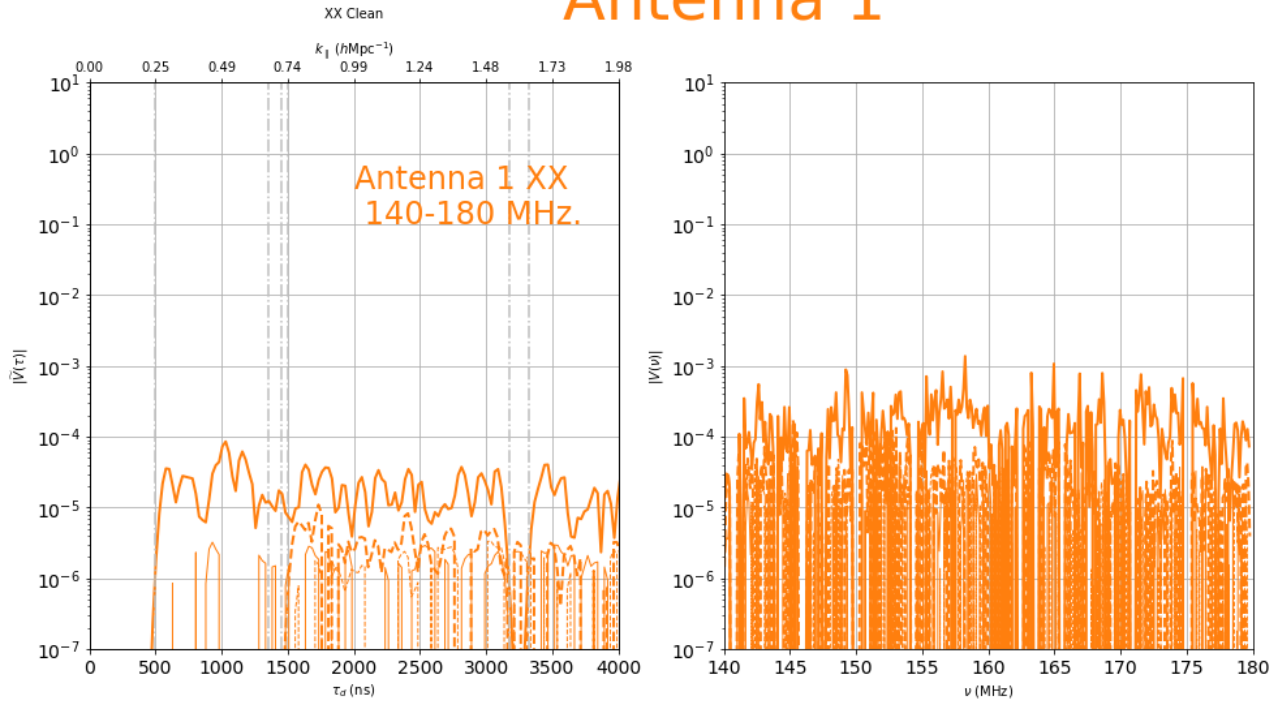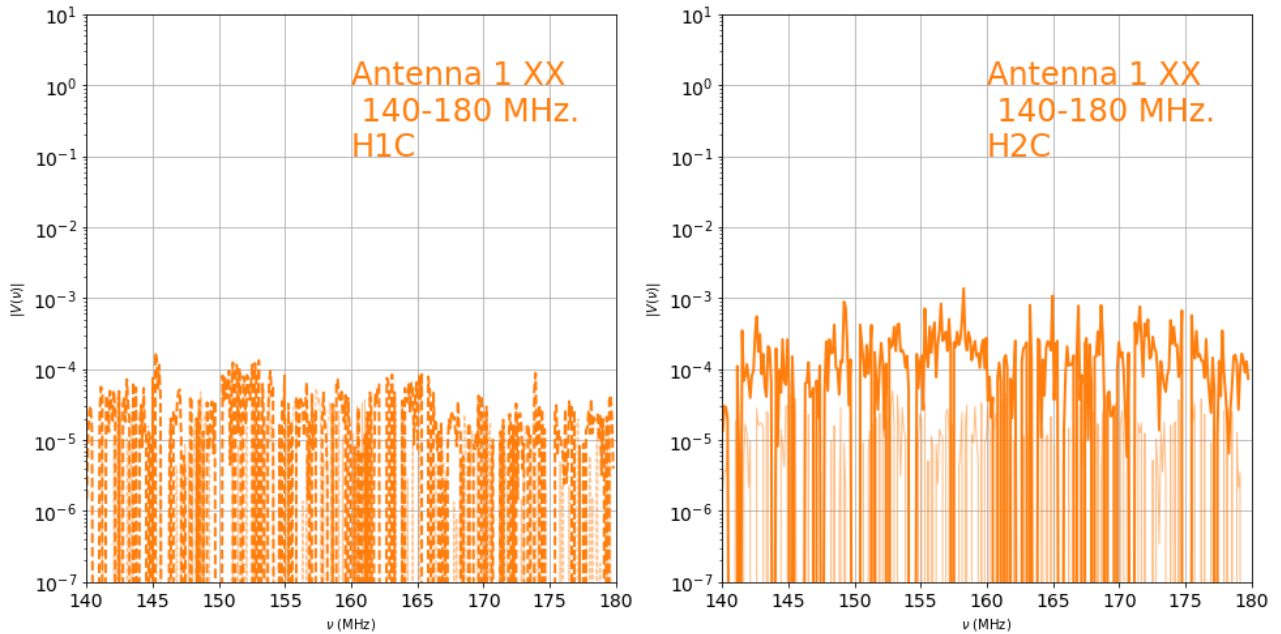
Figure 41: Left: A comparison of H1C and H2C data after being passed through a linear filter that removes all power below 500 ns for H2C and 1500 ns from H1C (we use a larger filter width for H1C to remove delay side-lobes of the rumble in the coaxial cables. We see that the $-45\,\mathrm{dB}$ systematics floor exists in the H2C data but if it exists, it is below the noise in H1C. Bottom: Frequency residuals from H1C and H2C on two separate plots for better direct comparison.