DATA ANALYSIS METHODS FOR THE DETECTION OF THE EPOCH OF REIONIZATION

Carina Cheng[1], et al.

[1]Astronomy Dept., U. California, Berkeley, CA

ABSTRACT

The Epoch of Reionization (EoR) is an uncharted era in our Universe's history during which the birth of the first stars and galaxies led to the ionization of the omnipresent neutral hydrogen. This important epoch of our cosmic dawn harbors a wealth of information regarding the environment during this transformative time, including insight into the nature of the first luminous sources and implications about cosmological parameters. There are many experiments investigating the EoR by tracing the 21 cm line of neutral hydrogen, a signal which is very faint and difficult to isolate. As instruments become more sensitive and a statistical power spectrum detection is in our foreseeable future, it has become increasingly important to develop techniques that help ensure that a potential detection isn't washed out or mistaken for something else. In this paper, we focus on two crucial ingredients for a power spectrum detection - a method for the accurate injection and recovery of an EoR signal, and jack-knife tests that can tease out foregrounds and systematics that contaminate our signal. We describe and apply these methods using data taken by the 64-element configuration of the Donald C. Baker Precision Array for Probing the Epoch of Reionization (PAPER).

1. INTRODUCTION

By about one billion years after the Big Bang, the very first stars and galaxies are thought to have ionized all the neutral hydrogen that dominated the Universe's early life. This important transition, during which the first luminous structures formed from gravitational collapse and emit intense radiation, transforming the cold neutral gas into a plasma, is known as the epoch of reionization (EoR). The EoR represents an unexplored era in our cosmic dawn, whose history encodes important information regarding the nature of the first galaxies and the process of structure formation. A direct detection of the EoR would unlock powerful information about the intergalactic medium, revealing connections between the smooth matter distribution exhibited via cosmic microwave background (CMB) studies and the highly structured web of galaxies we observe today.

One promising technique to probe the EoR is to target the 21 cm wavelength emission that is emitted by neutral hydrogen via its spin-flip transition. This technique is powerful because it can be observed as a function of redshift - that is, the wavelength of the signal reaching our telescopes can be directly mapped to a distance from where the emission originated before stretching out as it traveled through expanding space. The 21 cm line therefore offers a window into following the evolution of ionization, temperature, and density fluctuations on cosmic scales.

Although a detection of the EoR has not currently been made to-date, there are several radio telescope experiments that have succeeded in using the 21 cm signal from hydrogen in order to place constraints on the brightness of the EoR. Examples of experiments investigating the mean brightness temperature of the EoR relative to the CMB are EDGES (Bowman & Rogers 2010), the LWA (Ellingson et al. 2009), LEDA (Greenhill & Bernardi 2012), DARE (Burns et al. 2012), SciHi (Voytek et al. 2014), BIGHORNS (Sokolowski et al. 2015), and SARAS (Patra et al. 2015). Major interferometers, which seek to measure statistical power spectra, include the GMRT (Paciga et al. 2013), LOFAR (van Haarlem et al. 2013), the MWA (Tingay et al. 2013), the 21CMA (Peterson 2004, Wu 2009), and PAPER (Parsons et al. 2010). The Hydrogen Epoch of Reionization Array (HERA), which is currently being built, is a next-generation instrument that hopes to combine lessons learned from previous experiments and is forecasted to be able to make a successful [?σ] detection with an eventual [?] elements.

The major challenge that faces all 21 cm experiments is in isolating a small signal that is buried underneath foregrounds and instrumental systematics that are 4-5 orders of magnitude brighter. A clean measurement therefore requires an intimate understanding of the instrument and a very careful, thorough understanding of data analysis choices. With HERA on the horizon and continual progress being made in the field, it is becoming increasingly more important to develop methods that can both reinforce a potential real detection and explain any false ones. In this paper, we present approaches to both of these issues using data from the 64-element configuration of PAPER. We place special emphasis on the accurate injection and recovery of a fake EoR signal in order to assess signal loss associated with different weighting techniques, as well as the im-

portance of using jack-knife tests to investigate sources of bias.

This paper is organized as follows. In section 2 we give a brief overview of the PAPER-64 array and observations. In section 3 we describe our power spectrum estimation pipeline. Section 4 presents two key components of any power spectrum analysis pipeline - accurately quantifying signal loss that may result from data analysis choices, and choosing a weighting scheme to maximum foreground down-weighting without amassing large amounts of signal loss. In Section 5 we highlight useful jack-knife tests that can help discriminate between excess detections from foregrounds. We conclude in Section 6.

2. OBSERVATIONS

The Donald C. Baker Precision Array for Probing the Epoch of Reionization (PAPER) is a dedicated 21 cm experiment located in the Karoo Desert in South Africa. The PAPER-64 configuration consists of 64 dual-polarization drift-scan elements, each 2 m on a side. The antenna layout is formatted in a grid layout (Figure 1), with 8 antennas on a side, 30 m spacing between antennas along the East/West direction, and 4 m spacing between antennas along the North/South direction. For the rest of this paper, we focus on data from only the 30 m pure East/West baselines.

PAPER-64 observed for a total of 135 nights between 2012-2013. The correlator processes a bandwidth of 100-200 MHz, corresponding to a redshift range of 6-12. For more information about the backend system of PAPER-64 and its observations, we refer the reader to [??] and Ali et al. (2015).

Because there is a detailed discussion of the PAPER-64 data reduction pipeline in Ali et al. (2015), here we will only briefly summarize the data processing steps prior to the power spectrum analysis.

Beginning with compressed data products which have been cleaned of radio frequency interference (RFI) at the 6σ level and then down-sampled (to 42.9 s integrations, 203 frequency channels), we then employ the technique of redundant calibration as a means of calibrating for individual antenna gains without needing to use information about the sky. Because PAPER is laid out on a grid, antenna calibration parameters can be found through the use redundancy. Namely, baselines of the same length and orientation should measure the same sky signal, but in practice there are differences due to instrumental effects caused by antennas, cables, and receivers. Using the basis of redundancy, however, allows us to set up a system of equations to solve for a complex gain per antenna that brings all baseline of a particular type into alignment. We do this using the package OMNICAL.

We next solve for the array's overall phase and gain calibration parameters using a standard self calibration method. We used radio sources Pictor A, Fornax A, and the Crab Nebula to fit for an overall phase solution, and set our overall flux scale using Pictor A as a calibrator source. Although PAPER-64 data exists for all 4 polarizations, we only use the xx and yy polarization data to form Stokes I as $V_I = \frac{1}{2}(V_{XX} + V_{YY})$.

Eliminating bright foregrounds remains a challenging yet crucial component of any 21 cm data analysis. There are many techniques to go about foreground removal, including spectral polynomial fitting, principle component analysis, Fourier-mode filtering, non-parametric subtractions, and inverse covariance weighting. PAPER uses a delay-spectrum filtering method which was first explained and applied in Parsons et al. (2014). In short, the delay-filtering technique employs the spectrally smooth nature of foregrounds, which are consequently localized in delay-space, the Fourier dual to frequency. We subtract all components that fall within the horizon limit for a specific baseline type in delay-space, plus a 15 ns buffer, thus gaining about ~ 4 orders of magnitude in sensitivity.

Finally, after removing an additional layer of RFI (values 3σ above the median), we average together all our data into two datasets: even julian dates and odd julian dates. A total of 124 nights of data comprises this average, and we use these two datasets as a starting point for constructing power spectrum estimates.

3. POWER SPECTRUM ANALYSIS

3.1. *Fringe-Rate-Filtering*

As shown in Parsons et al. (2016), a fringe-rate filter allows us to maximize our sensitivity even further by averaging visibilities in time. We apply this filter in the fringe-rate domain, the Fourier-dual to time, and it effectively amounts to increasing our data's integration time. Another way to describe the filter is that it weights different parts of the sky (which rotate at different "fringe-rates") by the antenna beam. This allows us to up-weight parts of the sky high in the primary beam, and down-weight those that are less sensitive.

In Ali et al. (2015), the filter applied was degraded (widened in fringe-rate space) from an optimal one that can be constructed based on the integration time of the data. This was chosen to decrease the resulting integration time, therefore increasing the number of independent modes and reducing signal loss associated with applying optimal quadratic estimators (OQE) on data with too few modes.

New to our updated PAPER-64 analysis is the use of the optimal fringe-rate filter, which maximizes our sensitivity. With the development of a robust method for assessing signal loss (see Section 4), we feel comfortable using a narrow filter (Figure 2), resulting in an effective integration time of 3857 s (see Section 3.3 for calculation) and 8 total independent modes for our 8.5 hours of LST.

In practice, the optimal fringe-rate profile is computed for a fiducial 30 m baseline at 150 MHz, the center frequency in our band. The filter is implemented on a per baseline basis by convolving the time-domain visibili-
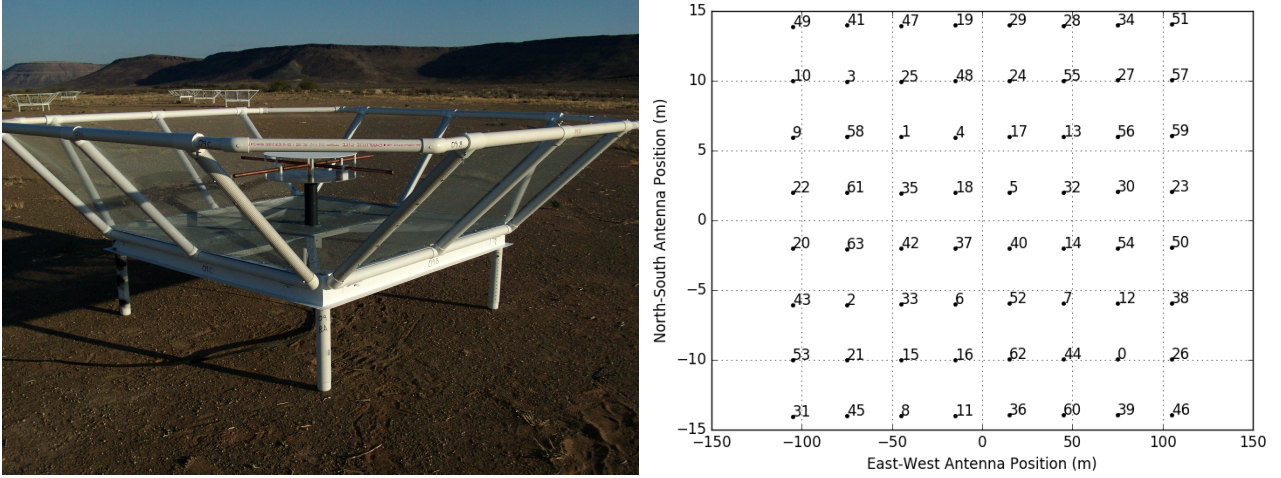
Figure 1. PAPER dipole in South Africa (left) and PAPER-64 antenna layout (right).
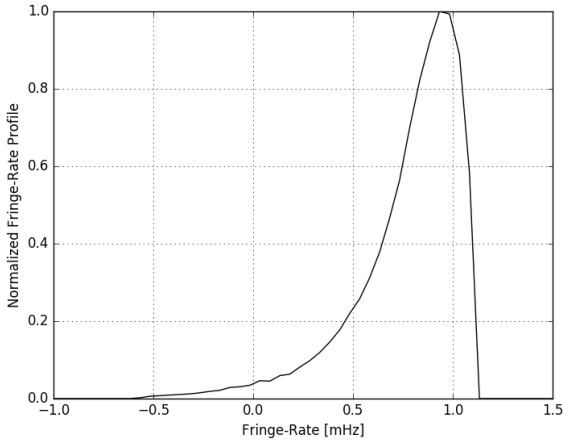

Figure 2. The optimal fringe-rate filter used in the analysis, normalized to 1.

ties with the Fourier transform of the fringe-rate filter, resulting in averaged visibilities yielding ~ 2 orders of magnitude in sensitivity.

3.2. *Power Spectrum Estimation*

To form power spectrum quantities using our two fringe-rate filtered, LST-binned datasets, we use OQE methods as described by [??]. A summary of the method is as follows.

We begin with our data vector $\mathbf{x}$, which contains our visibilities that have the shape (*times*, *frequencies*) for each baseline and data group ('even' or 'odd'). As part of our bootstrapping routine (and to speed things up), we divide up our total number of baselines into 5 random, equal-sized groups, where no baselines are repeated within or between groups. For the Ali et al. (2015) method, each group is then sampled with replacement to create a new group, of the same size, that can have repeated baselines inside it. We discover that in doing so, we are sacrificing some of our sensitivity since this results in there being 3-4 repeated baselines

per group. In order to maximize our sensitivity but still apply random sampling for use in error estimation, we instead form new groups using all independent baselines except the very last one. For example, if we have 10 baselines in a group, we use the first 9 to guarantee at least 9 independent measurements, and then fill the last slot randomly out of the 10. We do this for all 5 groups. This is still a valid means of bootstrapping because there are many more possibilities of baseline groupings than the number of bootstraps we run for this analysis ($nboots = 20$).

Taking $\mathbf{x}$ to be averaged data from all baselines within a group, we now have 10 possible data vectors (for 5 groups and 2 datasets). We next form the quantity $\hat{q}_\alpha$:

$$\hat{q}_\alpha = \frac{1}{2}\mathbf{x}^\dagger \mathbf{C}^{-1}\mathbf{Q}_\alpha \mathbf{C}^{-1}\mathbf{x} - b_\alpha \qquad (1)$$

where $\mathbf{C}$ is the covariance matrix defined to be $\mathbf{C} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$. The matrix $\mathbf{Q}$ is an operator that takes our frequency-domain visibilities and Fourier-transforms them into power spectrum space. A noise bias b_α is subtracted off. The index α represents each $k_\parallel$ bin, where $k_\parallel$ is the Fourier-dual to frequency.

Notice that there are two quantities of $\mathbf{C}^{-1}\mathbf{x}$ in the expression for $\hat{q}$, representing two copies of inverse covariance weighted data. We perform all possible cross-multiplications of this quantity, except between the same datasets ('even' with 'even', for example) and same groups ('baseline group 1' with 'baseline group 1', for example). We avoid these computations to prevent the introduction of biases from shared auto power.

The rest of the OQE formalism is identical to Ali et al. (2015). We normalize our power spectrum estimates using the matrix $\mathbf{M}$:

$$\hat{\mathbf{p}} = \mathbf{M}\hat{\mathbf{q}} \qquad (2)$$

where $\hat{\mathbf{p}}$ is the normalized estimate of the true power spectrum $\mathbf{p}$. We compute $\mathbf{M}$ using the Fisher matrix $\mathbf{F}$, defined as:

$$\mathbf{F}_{\alpha\beta} = \frac{1}{2} tr(\mathbf{C}^{-1} \mathbf{Q}_\alpha \mathbf{C}^{-1} \mathbf{Q}_\beta). \qquad (3)$$

$$(4)$$

We have a choice for $\mathbf{M}$, and we choose to compute it by taking the Cholesky decomposition of $\mathbf{F}$. Namely, we decompose the Fisher matrix such that $\mathbf{F} = \mathbf{LL}^\dagger$, where $\mathbf{L}$ is a lower triangular matrix. Next, we construct $\mathbf{M} = \mathbf{DL}^{-1}$, where $\mathbf{D}$ is a diagonal matrix. Hence, our window function $\mathbf{W} = \mathbf{MF}$ becomes $\mathbf{W} = \mathbf{DL}^\dagger$ and is an upper triangular matrix. This window function was constructed in a way to prevent the leakage of foreground power from low k to high k modes. Specifically, we order the elements in $\mathbf{F}$ in such a way so that power can leak from high k modes to low k modes, but not vice versa. Since most foreground power shows up at low k's, this method ensures a window function that retains clean, noise-dominated measurements.

Finally, our final power spectrum is:

$$\hat{\mathbf{p}} = \mathbf{Wp} \qquad (5)$$

In practice, we end up with a 2-dimensional power spectrum quantity, which is a function of both time and k. We compute 20 bootstraps of these quantities, where each bootstrap creates baseline groups randomly.

In Ali et al. (2015), a second round of bootstrapping occurs over the bootstrap and time axes simultaneously. Random values are sampled with replacement along both axes, drawing as many values as there are number of bootstraps and times. Final power spectrum limits are then computed by taking the mean and standard deviation over this second bootstrap axis.

However, we have found that this method greatly underestimates power spectrum errors, especially for fringe-rate filtered data. This can be explained by the fact that fringe-rate filtered data has a dramatically reduced number of independent modes. Hence, drawing 100 samples out of a length-100 dataset that only has 5 independent modes in it, for example, results in a narrower distribution of values that leads to a false error estimation.

To avoid this issue, we instead take a simple average along the time axis. Our final power spectrum limits are computed by taking the mean and standard deviation over our single bootstrap axis.

In summary, our power spectrum pipeline begins with data with four axes (datasets, baselines, times, frequencies), bootstraps over baselines, uses OQE formalism, and produces 1-dimensional power spectra for both $\mathbf{P(k)}$ and the folded version $\Delta^2(\mathbf{k})$, defined as:

$$\Delta^2(\mathbf{k}) = \frac{k^3}{2\pi^2} \hat{\mathbf{p}}(\mathbf{k}) \qquad (6)$$

One critical new component of our power spectrum pipeline is that we have many different power spectrum channels that simultaneously get processed at the same time and are especially useful for signal loss (Section 4)

and sensitivity (Section **??**) verification. Two important channels, in addition to our PAPER-64 data, are a noise dataset and an EoR dataset (in addition to these, we do different combinations of all three). We will briefly describe each.

We create random noise with a scale determined by the sensitivity of our instrument and parameters of our dataset. We calculate our system temperature for our frequencies of interest as:

$$T_{sys} = 180 \left(\frac{\nu}{0.18}\right)^{-2.55} + T_{rcvr} \qquad (7)$$

where ν are frequencies in GHz. We use a receiver temperature of 200 K, yielding $T_{sys} = 487$ at 150 MHz. This is lower than in Ali et al. (2015) because **[why?]**. We convert this temperature into a variance statistic using:

$$T_{rms} = \frac{T_{sys}}{\sqrt{\Delta\nu\Delta t N_{days} N_{pols}}} \qquad (8)$$

where $\Delta\nu$ is channel spacing, Δt is integration time, N_{days} is the number of data samples for a particular time and frequency that went into our LST binned set, and N_{pols} is the number of polarizations (2 for Stokes I). Our simulated noise has the same shape as our data, and we fringe-rate filter it in the same way to best mimic the real noise floor of our data.

A second additional important channel in our pipeline is a simulated EoR signal. We again simulate random noise (with a default scale of 1) with the same shape as our data, and we fringe-rate filter this signal twice. The first filter transforms the unattached noise into a signal that's attached to the sky (what our instrument observes). The second filter represents the fringe-rate filtering step in our data analysis pipeline. As described in Section 4, we create an EoR signal at various amplitude levels.

Our multi-channel power spectrum pipeline has been essential in helping us understand the nuances in our pipeline. Because there are many different components and countless subtle effects affecting our final limits, it has been imperative to carry all our channels through the analysis to validate each step.

3.3. *Power Spectrum Sensitivity*

[Add info on sensitivity equation, how we calculate effective inttime and adjust beam factor because of FRF, etc.]

4. SIGNAL LOSS

4.1. *Formalism*

Based on our analysis pipeline, potential signal loss is a real and significant issue. More specifically, when applying inverse covariance weighting, $\mathbf{C}^{-1}$ is empirically estimated from the data itself, which has the consequence of over-fitting the noise in the data, producing power spectra values well below the thermal noise

limit that is predicted based on observation parameters. This is especially prevalent when weighting fringe-rate-filtered data, which has so few independent time modes to begin with, leading to a noisier dataset. Being able to accurately quantify this loss is crucial in interpreting and providing credibility to any power spectrum limits.

New to the PAPER-64 analysis is a robust method to estimate signal loss associated with inverse covariance weighting. This method, explained below, is now a standard analysis step for all PAPER analyses and one that will be used for HERA moving forward.

As discussed previously, our power spectrum pipeline runs on a standardized set of channels (pure data, pure noise, pure EoR, and combinations of the three). As part of our signal loss routine, we also compute power spectra with various levels of the created EoR signal, dialing its amplitude from well-below the data level, to well-above. Suppose that $\mathbf{e}$ is the injected EoR (at some amplitude level), and $\mathbf{x}$ is our data vector. We define $\mathbf{r}$ to be:

$$\mathbf{r} = \mathbf{x} + \mathbf{e} \qquad (9)$$

Using our OQE formalism, we are interested in the following two quantities: P_{in} and P_{out}. The input power spectrum, P_{in} represents the unweighted power spectrum of only $\mathbf{e}$, our simulated EoR signal. The output power spectrum, P_{out}, is the weighted power spectrum of $\mathbf{e}$ that would result from our pipeline if the signal was mixed with our data. Comparing the two quantities yields insight into how much of $\mathbf{e}$ is lost due to our choice of weighting. Ignoring normalization, factors:

$$P_{in} \propto \mathbf{e}^{\dagger} \mathbf{I}^{-1} \mathbf{Q} \mathbf{I}^{-1} \mathbf{e} \qquad (10)$$

$$\begin{aligned} P_{out} &\equiv \mathbf{P}_e = \mathbf{P}_r - \mathbf{P}_x \\ &\propto \mathbf{r}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} \mathbf{r} - \mathbf{x}^{\dagger} \mathbf{C}_x^{-1} \mathbf{Q} \mathbf{C}_x^{-1} \mathbf{x} \end{aligned} \qquad (11)$$

It is noted that the output power spectrum is comprised of two terms: the covariance treated power spectrum associated with $\mathbf{r}$, and that of data $\mathbf{x}$ alone.

One may wonder why P_{out} cannot be computed simply as the weighted power spectrum of $\mathbf{e}$ alone, namely $P_{out} \propto \mathbf{e}^{\dagger} \mathbf{C}_e^{-1} \mathbf{Q} \mathbf{C}_e^{-1} \mathbf{e}$. Expanding Equation 11:

$$\begin{aligned} P_{out} &\propto (\mathbf{x} + \mathbf{e})^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} (\mathbf{x} + \mathbf{e}) - \mathbf{x}^{\dagger} \mathbf{C}_x^{-1} \mathbf{Q} \mathbf{C}_x^{-1} \mathbf{x} \\ &\propto \mathbf{x}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} \mathbf{x} + \mathbf{e}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} \mathbf{e} + \mathbf{x}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} \mathbf{e} \\ &+ \mathbf{e}^{\dagger} \mathbf{C}_r^{-1} \mathbf{Q} \mathbf{C}_r^{-1} \mathbf{x} - \mathbf{x}^{\dagger} \mathbf{C}_x^{-1} \mathbf{Q} \mathbf{C}_x^{-1} \mathbf{x} \end{aligned}$$

And taking the case of very large $\mathbf{e}$, so that $\mathbf{C}_r^{-1} \sim \mathbf{C}_e^{-1}$ and any terms involving only $\mathbf{x}$ are small:

$$\begin{aligned} P_{out, \mathbf{e} \gg \mathbf{x}} &\propto \mathbf{e}^{\dagger} \mathbf{C}_e^{-1} \mathbf{Q} \mathbf{C}_e^{-1} \mathbf{e} + \mathbf{x}^{\dagger} \mathbf{C}_e^{-1} \mathbf{Q} \mathbf{C}_e^{-1} \mathbf{e} \\ &+ \mathbf{e}^{\dagger} \mathbf{C}_e^{-1} \mathbf{Q} \mathbf{C}_e^{-1} \mathbf{x} \end{aligned} \qquad (12)$$

We see that our naive expression for P_{out} is the first term, but there also two additional terms. An initial

assumption would be that the cross-terms that involve both $\mathbf{e}$ and $\mathbf{x}$ should be zero, since the two quantities are un-correlated. However, [**need explanation about power in cross-terms here**]. Therefore, in our investigation of signal loss, we use the full quantity for P_{out} as in Equation 11.

For the unweighted case ($\mathbf{C} \equiv \mathbf{I}$), we expect P_{out} and P_{in} to be equal, and hence the ratio of P_{in}/P_{out} to be 1. For the weighted case, this is not true due to signal loss. In order to quantify the loss, we look at the ratio of P_{in} to P_{out} as the amplitude level of the injected signal $\mathbf{e}$ is increased. In the next section, we highlight two methods that yield similar results for the determination of signal loss using P_{in} and P_{out}

4.2. Application

Recall that fringe-rate filtered noise, which mimics the level of noise in our actual PAPER-64 dataset, is a channel in our power spectrum pipeline. We can compute signal loss quantities of interest for the noise $\mathbf{n}$ similar to the expressions we featured previously for data $\mathbf{x}$.

$$P_{in} \propto \mathbf{e}^{\dagger} \mathbf{I}^{-1} \mathbf{Q} \mathbf{I}^{-1} \mathbf{e} \qquad (13)$$

$$\mathbf{s} = \mathbf{n} + \mathbf{e} \qquad (14)$$

$$P_{out} \propto \mathbf{s}^{\dagger} \mathbf{C}_s^{-1} \mathbf{Q} \mathbf{C}_s^{-1} \mathbf{s} - \mathbf{n}^{\dagger} \mathbf{C}_n^{-1} \mathbf{Q} \mathbf{C}_n^{-1} \mathbf{n} \qquad (15)$$

Using the input and output power spectra for range of EoR amplitudes, we use two methods to determine signal loss associated with the over-fitting of noise during inverse covariance weighting. We first look at signal loss for the pure noise case (no data), to show that we can successfully inject EoR signals, determine signal loss, and then recover the signal.

The first method is very straightforward. Post-bootstrapping, our final power spectra are 1-dimensional and only a function of k. For each k, we simply look at P_{out} as a function of P_{in}, as shown in Figure 3. The shape of this function can be explained as follows. At small injection levels (small $\mathbf{e}$), P_{out} and P_{in} are equal, and there is no signal loss. As the amplitude of EoR increases, we then move into a regime where the final output power spectrum is lower than the unweighted input one. This is dangerous, because without correcting for this effect one might be led to underestimate the EoR signal. The peculiar tail at very low injection levels (where $P_{out} > P_{in}$) is an unphysical feature, but rather illustrates that there is some non-negligible cross-term power between $\mathbf{n}$ and $\mathbf{e}$.

For this method, we interpolate the signal loss factor (per k), computed as P_{in}/P_{out}, at a P_{out} value equal to the 2σ power spectrum upper limit of noise alone. In other words, we look at $P_{noise} \propto \mathbf{n}^{\dagger} \mathbf{C}_n^{-1} \mathbf{Q} \mathbf{C}_n^{-1} \mathbf{n}$, compute its 2σ upper limit (mean over bootstraps + $2 \times$ standard deviation over bootstraps), and interpolate the value of P_{in}/P_{out} at this value. We therefore end up with one signal loss correction factor per k.

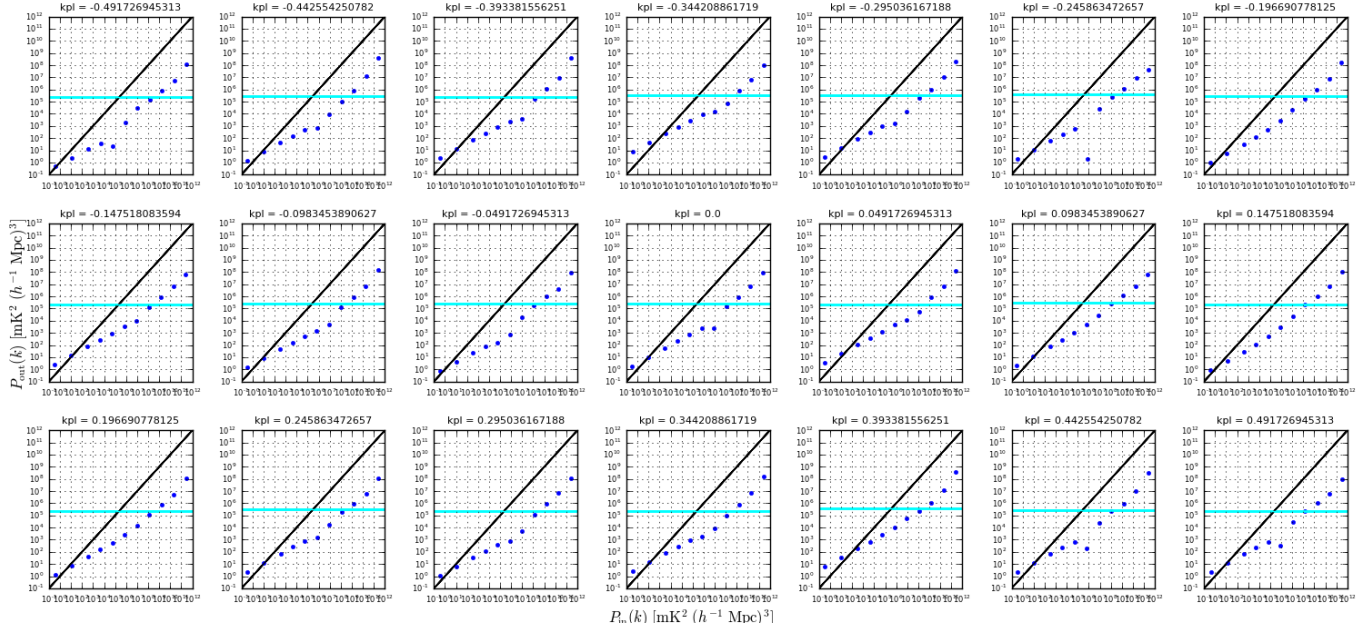Figure 4 shows the power spectrum of our noise simulation, using full inverse covariance weighting, both be-

Figure 3. P_{in} vs. P_{out} (blue points) for 15 injection levels and 21 k's. The solid cyan line is the 2σ upper limit for the weighted power spectrum of noise alone, and it is at this level where the signal loss factor P_{in}/P_{out} is computed by interpolation. [**Maybe re-do this plot with closer-together points**]

fore and after signal loss correction. Prior to signal loss correction, it is obvious that the power spectrum is unfeasible because it is well below the theoretical noise level prediction. Post-correction, the power spectrum values blow up much higher than both the theory and unweighted power spectrum. This is an effect caused by the steep nature of the eigenspectrum of **C**, and is explained more in Section 4.3.

[**Need a plot that shows our signal loss factors are CORRECT. How to do that??**]

Our second method for estimating signal loss is similar to the first, but more comprehensive in a statistical sense. Instead of looking at input and output power spectra after bootstrapping, we now look at their values for every bootstrap in order to get a sense of their distributions. Figure 5 plots P_{in} vs. P_{out} for 20 bootstraps, and as expected, the function now has a spread in the width-direction in comparison to what was plotted in Figure 3, but otherwise shows a familiar trend. Similarly, our weighted noise power spectra also has a defined spread due to bootstrapping.

Using these two distributions (P_{in}/P_{out} and P_{noise}), we can create bins along the P_{noise} axis to yield histograms of signal loss factors for each bin. We similarly sort the values of P_{noise} into the same bins, and multiply the probability of P_{noise} per bin (the number of values falling into that bin, divided by the total) with the signal loss factors in that bin, essentially computing a weighted average across all bins to obtain a final signal loss factor per k. As shown in Figure 6, the results are very similar to the previous method. For future power spectrum results, we choose to use the second method because it computes final signal loss values using our full

distributions of measurements.

One thing to note is that for both methods, we have been careful to validate that the computations yield no signal loss (signal loss factors of 1) for the unweighted power spectrum case, as is expected. This is important in confirming that signal loss is a direct result of the choice for **C**.

4.3. Data Weighting

With our signal loss formalism established, we now have the capability of experimenting with different weighting options for **C**. Our goal here is to choose a weighting method that successfully down-weights foregrounds and systematics in our data without generating large amounts of signal loss. We have found that the balance between the two is a delicate one and requires [**?? finish sentence...**].

We now turn our attention to power spectra using the 30 m East/West baselines of PAPER-64. Our dataset spans 8.5 hours of LST (.1-8.6 hrs), includes a total of 51 baselines, and is fringe-rate filtered using an optimal fringe-rate filter. We have two datasets (even days and odd days), and only cross-multiply data from different days and different baselines. We are interested in 21 frequency channels (channels 95-115), which yields a power spectrum for a redshift of $z = 8.4$.

Using full inverse covariance weighting, our results are not too dissimilar to that of pure noise. Signal loss factors (Figure 7) are of similar order of magnitude, and our power spectrum blows up past the unweighted version after signal loss correction (Figure 8).

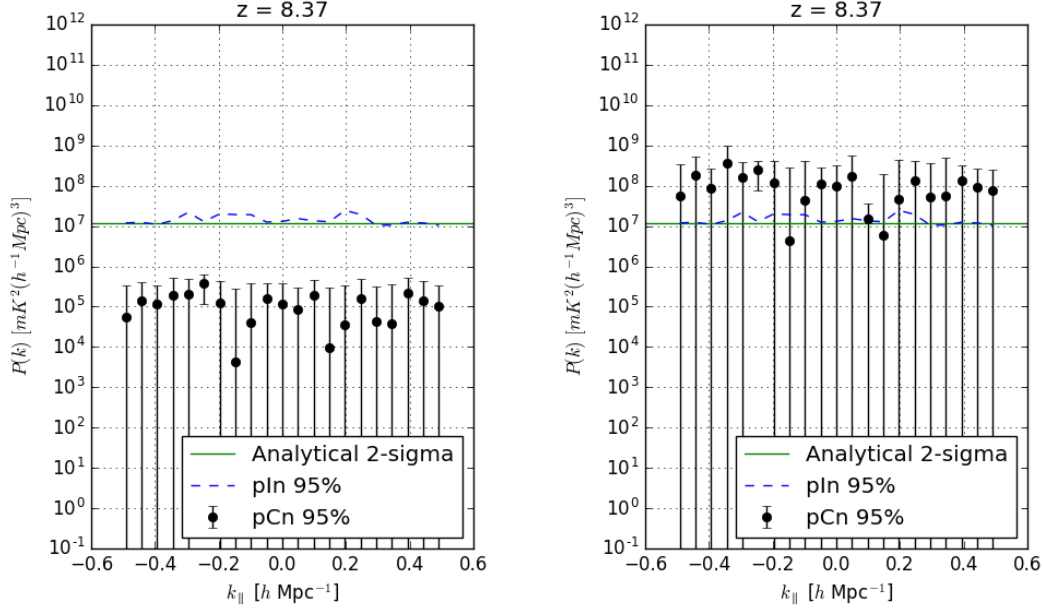Looking into this behavior in more detail, we investigate the shape of the eigenspectrum of **C** for a typical

Figure 4. Full inverse covariance weighted power spectrum of pure noise (black points, with 2σ error bars) before signal loss correction (left) and after (right). The dashed blue line is the unweighted power spectrum (2σ upper limit). The solid green line is the theoretical noise level prediction based on observational parameters. **[Color negative points grey]**
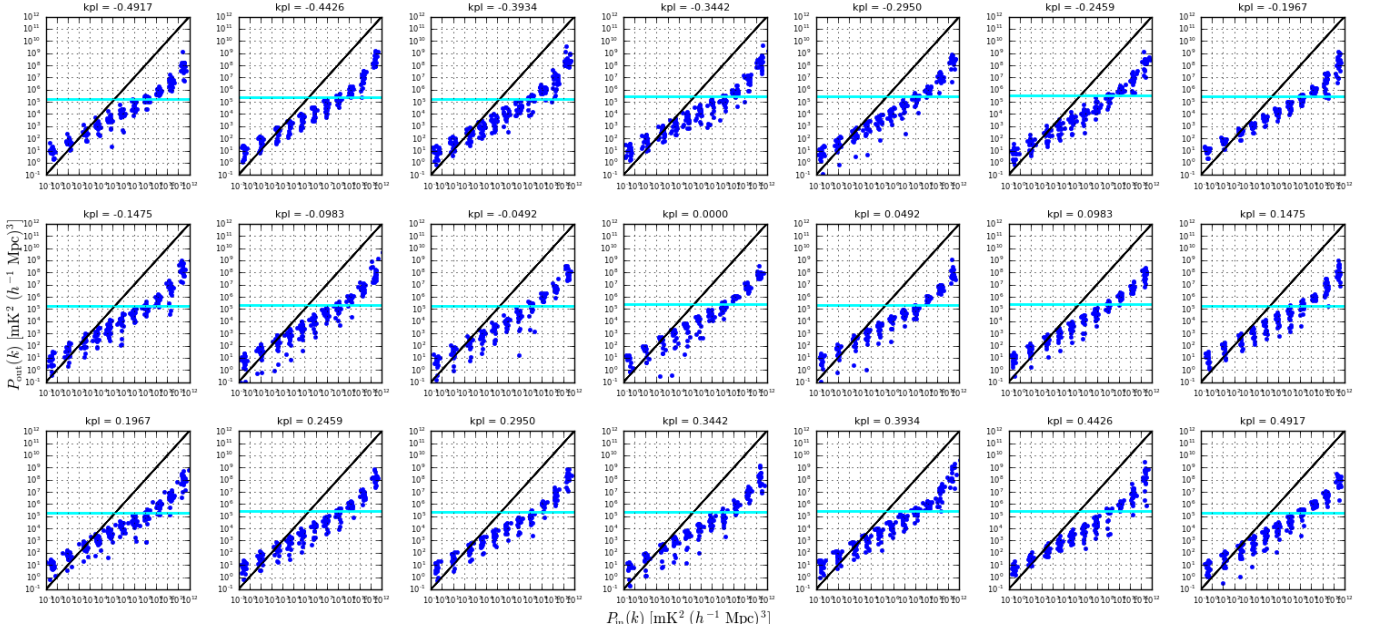

Figure 5. P_{in} vs. P_{out} (blue) for 15 injection levels, 20 bootstraps, and 21 k's. **[Plot a semi-transparent cyan range of pCn values instead of just the max]**

baseline used in the analysis. Figure 9 shows this spectrum for baseline (1,4). Most obviously, the spectrum is steep, spanning 4 orders of magnitude. Not as obvious is the effect of this shape on our results. When the matrix $\mathbf{C}$ is inverted to form $\mathbf{C}^{-1}$, the effect of the steepness of the eigenspectrum is to up-weight very few modes of the sky while the rest are drastically down-weighted. More specifically, our fringe-rate filtered data contains a finite,

small number of independent modes, thereby resulting in a covariance matrix that can be described by just a few modes. Beyond the first few modes, the eigenvalues of each additional mode falls of dramatically. When inverting, we end up not only down-weighting those initial modes but severely up-weighting a few insignificant ones. Because of our weighting choice, signal loss blows up as it thinks we only have a couple modes in our data.
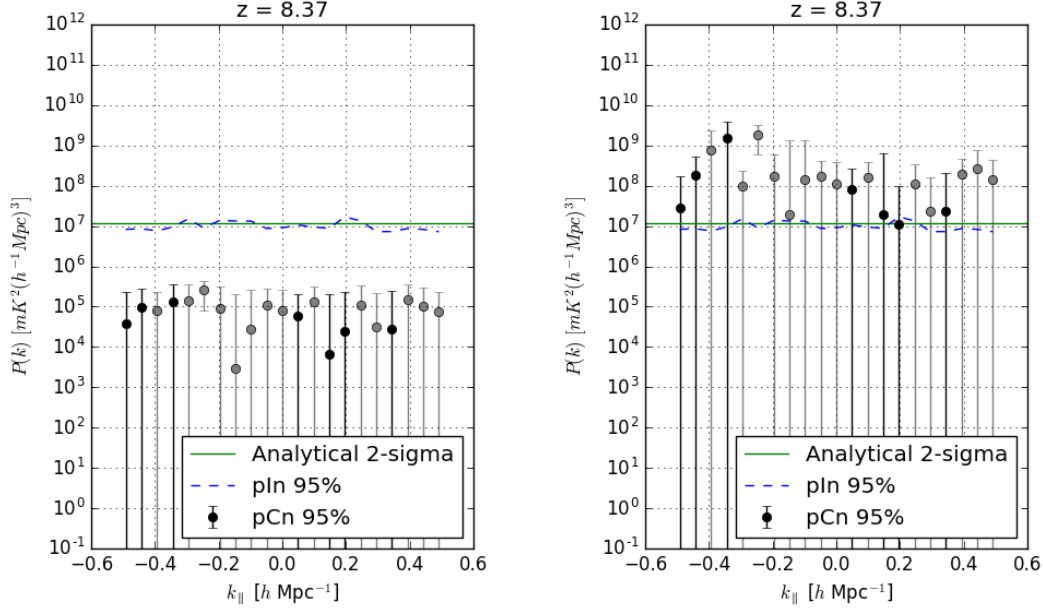
Figure 6. Full inverse covariance weighted power spectrum of pure noise (black and grey points, with 2σ error bars) before signal loss correction (left) and after (right). Black points correspond to positive values, while grey points correspond to originally negative values that have been made positive for plotting. The dashed blue line is the unweighted power spectrum (2σ upper limit). The solid green line is the theoretical noise level prediction based on observational parameters.

Clearly the full inverse covariance treatment of our data is suboptimal to even the unweighted case, but we would like to find a weighting method that does successfully down-weight contaminants in our data and make some improvement over the unweighted power spectrum. There are many choices for determining the covariance matrix $\mathbf{C}$, but here we will illustrate [?] promising ones as applied to PAPER-64.

[**TO DO: decide on/explain/show different weightings**]

5. JACK-KNIVES

6. CONCLUSION

7. ACKNOWLEDGEMENTS

REFERENCES

Ali, Z. S., et al. 2015, ApJ, 809, 61

Bowman, J. D., & Rogers, A. E. E. 2010, Nature, 468, 796

Burns, J. O., et al. 2012, Advances in Space Research, 49, 433

Ellingson, S. W., Clarke, T. E., Cohen, A., Craig, J., Kassim, N. E., Pihlstrom, Y., Rickard, L. J., & Taylor, G. B. 2009, IEEE Proceedings, 97, 1421

Greenhill, L. J., & Bernardi, G. 2012, ArXiv e-prints

Paciga, G., et al. 2013, MNRAS

Parsons, A. R., Liu, A., Ali, Z. S., & Cheng, C. 2016, ApJ, 820, 51

Parsons, A. R., et al. 2010, AJ, 139, 1468

—. 2014, ApJ, 788, 106

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

Peterson, U.-L. P. X.-P. W. J. 2004, ArXiv Astrophysics e-prints

Sokolowski, M., et al. 2015, PASA, 32, e004

Tingay, S. J., et al. 2013, PASA, 30, 7

van Haarlem, M. P., et al. 2013, A&A, 556, A2

Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. 2014, ApJL, 782, L9

Wu, X. 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts #213, 474
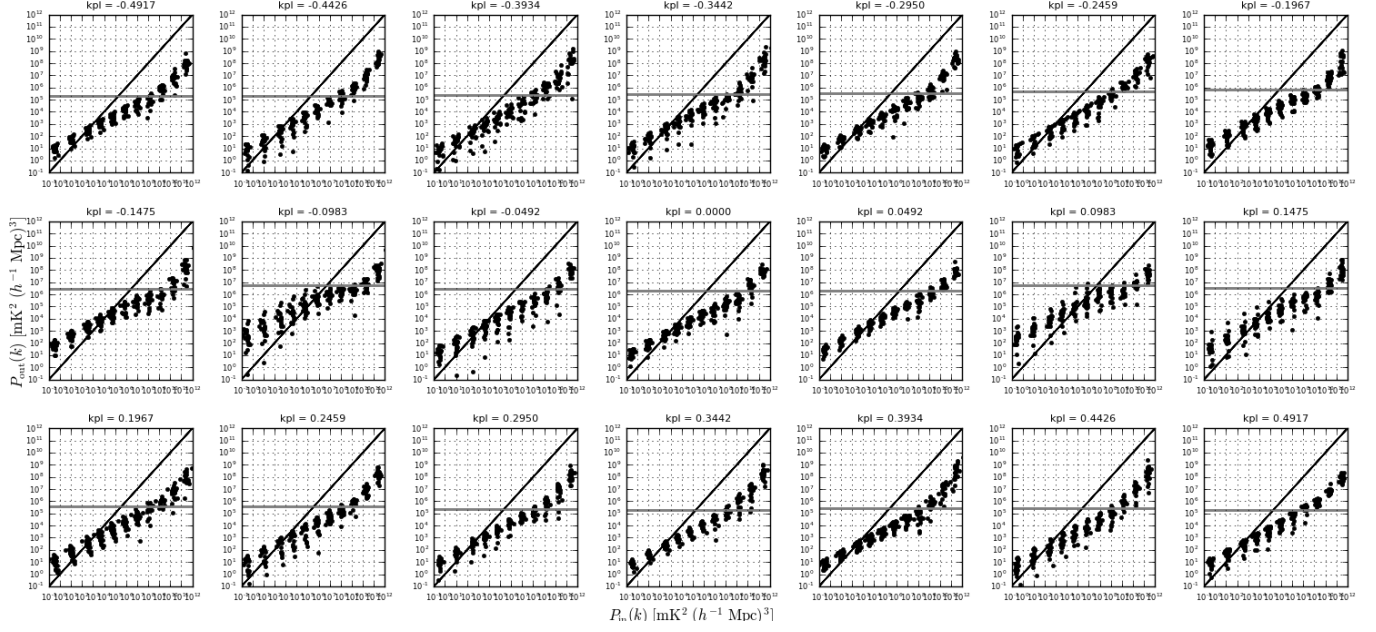
Figure 7. P_{in} vs. P_{out} (black) for 15 injection levels, 20 bootstraps, and 21 k's. [**Plot a semi-transparent grey range of pCv values instead of just the max**]
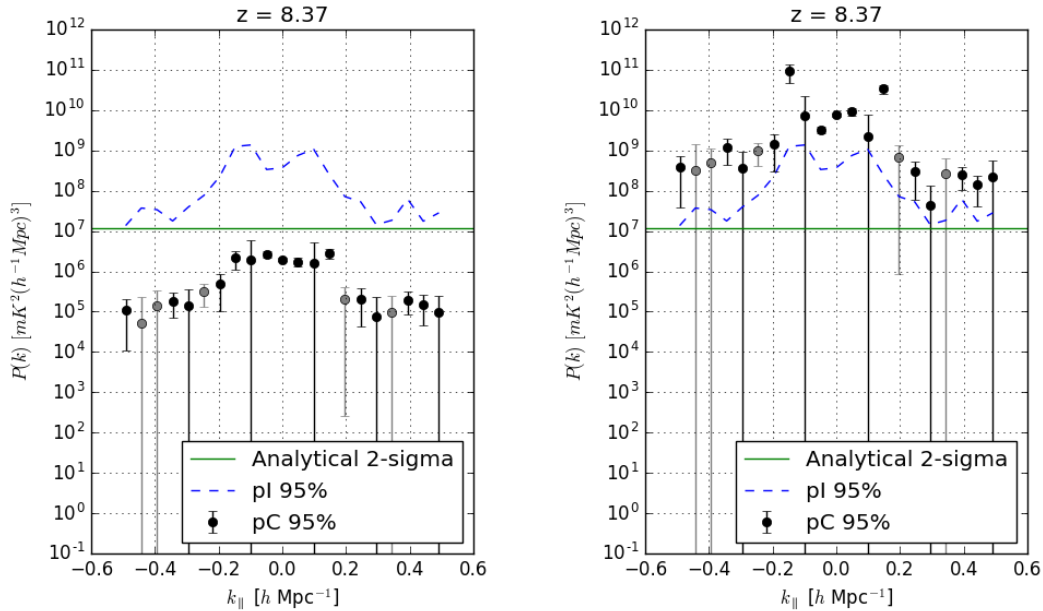

Figure 8. Full inverse covariance weighted power spectrum of PAPER-64 data (black and grey points, with 2σ error bars) before signal loss correction (left) and after (right). Black points correspond to positive values, while grey points correspond to originally negative values that have been made positive for plotting. The dashed blue line is the unweighted power spectrum (2σ upper limit). The solid green line is the theoretical noise level prediction based on observational parameters.
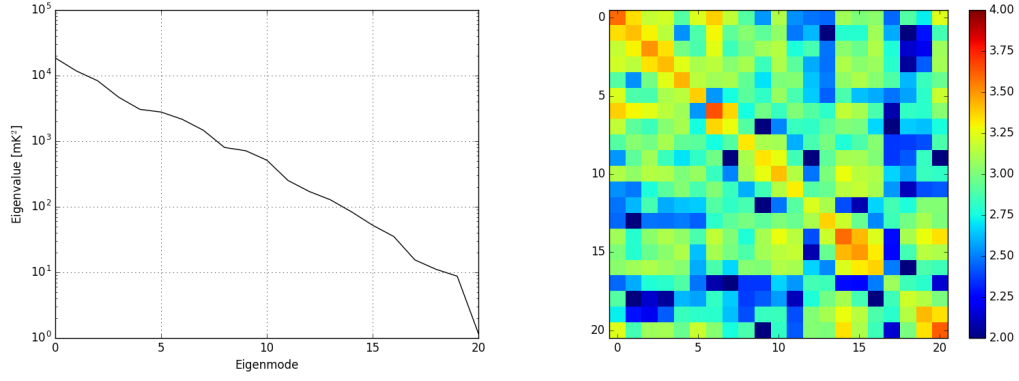
Figure 9. Eigenspectrum for **C** for baseline (1,4) for the 21 channels ofs interest (left) and covariance matrix **C** for the same baseline (right).