HERA Memo #093
Received 3 Feb 2021

Statistical Methods for Investigating Reflection Systematics

Dixin Chen (Supervisor: Dr. Adrian Liu)

August 22, 2019

1 Introduction

This memo focuses on the systematics originated from cable reflections in the signal chain of an antenna, and investigates how this form of instrumental systematics contaminates the signals occupying high delay modes by performing various statistical tests.

When signals are transmitted forward along the signal chain, part of the incident signal will be reflected back if there is an impedance mismatching between the transmitting and receiving surfaces in the cable. This creates a copy of the foreground signal at a discrete delay dominated by EoR signals and thermal noise. The amplitude of the reflected signal will typically be a fraction of the incident signal amplitude. For example, Figure 1 showed a clear cable reflection feature at around 1000 ns in a power spectrum due to reflections in the 150-meter coaxial cable connecting the antenna and the amplification module.

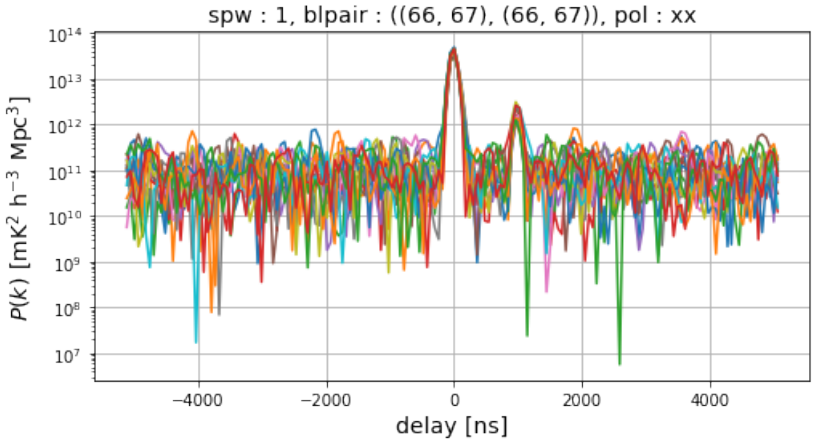


Figure 1: Power spectrum with prominent systematic features of cable reflections in the coaxial cable between antenna and post-amplifier.

2 Distribution models for high delay detections

A probability density function (PDF) describing the distribution of the real part of the power for noise signal was derived and fitted to high-delay signals from power spectra with different prominence of reflection systematics. The derived distribution was compared with a Gaussian distribution by measuring the goodness-of-fit using Kolmogorov-Smirnov test (KS test). The high delay region was roughly defined to be above 1500 ns. We investigated signals there since the systematic tail extends over the high delays where EoR signals and thermal noise occupy.

2.1 Data and model selection

Normality of high delay detections was examined by fitting a Gaussian distribution to signals in an auto-baseline power spectrum with clear cable reflection features and another without clear features. The Gaussian distribution was fitted to the real part of the power at delays between 2000 ns and 4000 ns in each power spectrum. Note that the power we fitted to was before averaging by time or frequency.

Power in auto-baseline delay spectrum is proportional to the square of absolute value of visibilities in delay space. If we assume the visibilities to be random noise with zero mean value, then the power is not expected to be Gaussian distributed since it is now proportional to the product of two complex random variables. Therefore, besides fitting the Gaussian distribution to high-delay power, we also fitted the data with the probability density function (PDF) $\mathrm{pr}_{\mathcal{CNN}}(P)$ of the real part of the product of two complex Gaussian distribution (see derivation in Section 6.1):

$$\mathrm{pr}_{\mathcal{CNN}}(P) = \frac{1}{\sigma}\, e^{-\frac{2|P|}{\sigma}}, \tag{1}$$

where P is the real part of the power, the parameter σ is the standard deviation of P and $\mathcal{CNN}$ represents the complex double normal distribution. During the derivation, we have assumed that the variances of visibilities of the two baselines forming one power spectrum are the same. We then fitted the $\mathcal{CNN}$ distribution to the power at delays between 2000 ns and 4000 ns.

KS tests were performed for each fit with data in each power spectrum in order to measure the goodness of fit. The KS test finds the maximum absolute distance (D) between two cumulative distribution functions (CDFs). For example, if we want to test for the normality of data, the two CDFs will be a given empirical CDF and a normal CDF. If the test statistic is greater than a critical value, the null hypothesis (the data follow a proposed distribution) will be rejected. Another way to interpret the test result is to compare its p-value, which is the probability of observing a test statistic as large as what we found if the null hypothesis is true. If the p-value is smaller than the significance level, the null hypothesis will be rejected.

2.2 Fitting results

Figure 2 showed two power spectra with different prominence of reflection systematic features and the CDFs of models fitted to the power at high delays. For both pairs of baselines, the discrepancy between the data CDF and the Gaussian distribution CDF with fitted parameters plugged in is greater than the discrepancy for the $\mathcal{CNN}$ distribution fit. Figure 3 showed the KS test statistics for fitting with each distribution.

The KS tests reject the hypothesis that the data follow a Gaussian distribution at the 5% level for either pair of baselines, while the KS test statistics accept the hypothesis that the data are $\mathcal{CNN}$ distributed for both pairs. In other words, signals at high delays do not follow a normal distribution but a $\mathcal{CNN}$ distribution regardless of the prominence of cable reflection features in the power spectrum.

We also fitted both distributions to the high-delay signals after subtracting a systematic model and performed KS tests (see Figure 3 and Figure 11 in appendix for plots of CDFs and KS test statistics before and after systematic model subtraction). The KS test result for the Gaussian fit suggests that the null hypothesis is still rejected after subtracting the systematic model. As for the $\mathcal{CNN}$ distribution, the null hypothesis remains to be accepted, and the model fitted slightly better to the data with systematic model subtracted as the KS test statistics decreased by 0.34% after subtracting the systematic model for this particular pair of baselines.

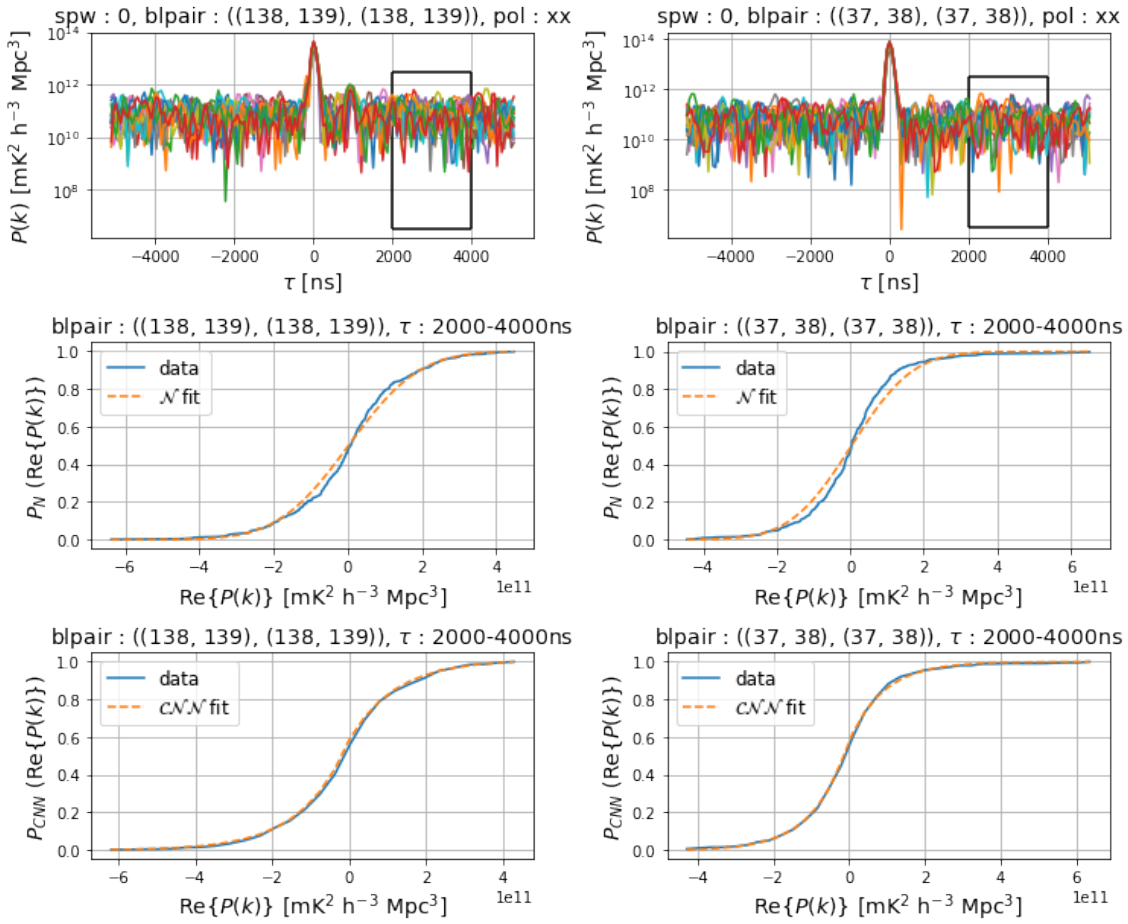


Figure 2: Two auto-baseline power spectra (top row) with different prominence of systematic features. Normalized histogram of real part of the power taken from the boxed region was fitted with the probability density functions of the proposed models. The CDFs of the fitted Gaussian (middle row) and $\mathcal{CNN}$ (bottom row) distributions were plotted along with the data CDF to measure the goodness-of-fit using KS tests.

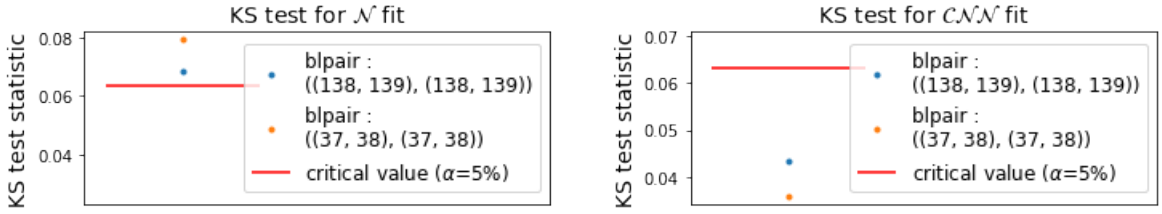


Figure 3: KS tests for fitting the Gaussian distribution (left) and the $\mathcal{CNN}$ distribution (right) to the power at delays 2000 - 4000 ns in Figure 2.

2.3 Baseline-averaged KS test statistics

Multiple pairs of baselines with clear reflection features shown in their power spectra were selected in order to see the change in goodness-of-fit for data with different sizes. Powers at different ranges of delays from each spectrum were fitted with the $\mathcal{CNN}$ distribution. The KS test statistics for the fittings were averaged by baselines and plotted as a function of data size in Figure 4 along with the baseline-averaged KS test statistics for fittings to the data from multiple power spectra without clear reflection features.

Figure 4 showed that the standard error of the average KS test statistics for power spectra with clear reflection features are generally greater than those for power spectra without clear reflection features. This suggests that the consistency of getting a good fit for high-delay signals from a power spectrum with clear reflection features is less than from one without clear features.

Besides, when the sample size is small, i.e., data selected from a narrow range of delays, the goodness-of-fit depends less on the prominence of cable reflection features appeared in power spectrum. As we prolong the range of delays by fixing the lower limit while increasing the upper limit of the delays, the power spectrum with clear cable reflection features generally unexpectedly fits better with the $\mathcal{CNN}$ distribution. This was further discussed in Section 3.2 where we showed the baseline-averaged KS test statistics of fittings for null tests.

2.4 Noise estimation

In order to obtain the standard deviation σ in the PDF of the $\mathcal{CNN}$ distribution, we can directly fit the PDF to the data (Section 2.1) and get an empirical standard deviation σ_e. We can also predict it using noise estimation in frequency space. Let the predicted standard deviation be σ_p.

The prediction of σ_p involves three steps. The first step is to estimate the variance of the noise σ_n^2 for baseline formed by antennas i and j in frequency space using Equation 2:

$$\sigma_n^2 = \frac{V_{ii}V_{jj}}{Bt}, \tag{2}$$

where V_{ii} and V_{jj} are auto-correlation visibilities in frequency space, B represents the

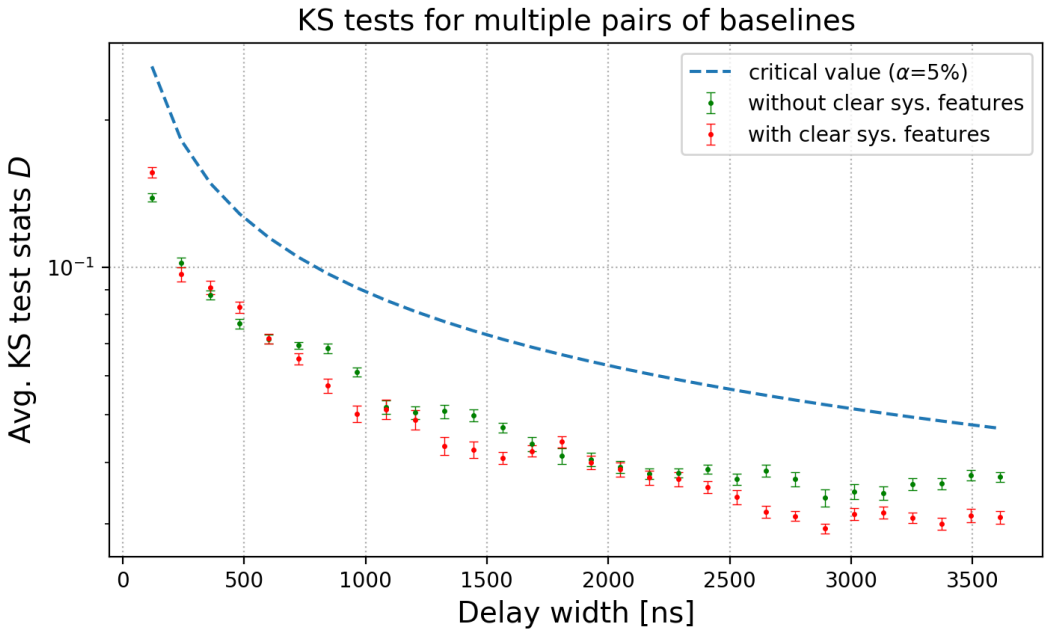


Figure 4: Baseline-averaged KS test statistics for $\mathcal{CNN}$ fittings with high-delay data of different widths of delay from multiple power spectra without (green) and with (red) clear systematic features of cable reflections.

channel width and t is the integration time. The second step is to compute the variance $\sigma_{\tilde{n}}^2$ of the noise after being Fourier transformed into delay space. This was done using the Parseval's theorem, which connects the sum of the square of a function with the sum of the square of its transform. Equation 3 shows the Parseval's theorem for the discrete Fourier transform.

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2, \tag{3}$$

where $X[k]$ is the discrete Fourier transform of function $x[n]$ both of length N. Since the mean value of the independent random noise is zero, the variance is the sum of the square of the noise. We therefore derived Equation 4 to compute the variance of noise in delay space:

$$\sigma_{\tilde{n}}^2 = N\,\sigma_n^2. \tag{4}$$

The last step is to estimate the standard deviation of the power σ_p. We have assumed that the standard deviation of the power in auto-baseline power spectrum of noise is the square of the standard deviation of visibilities in delay space. The standard deviation of

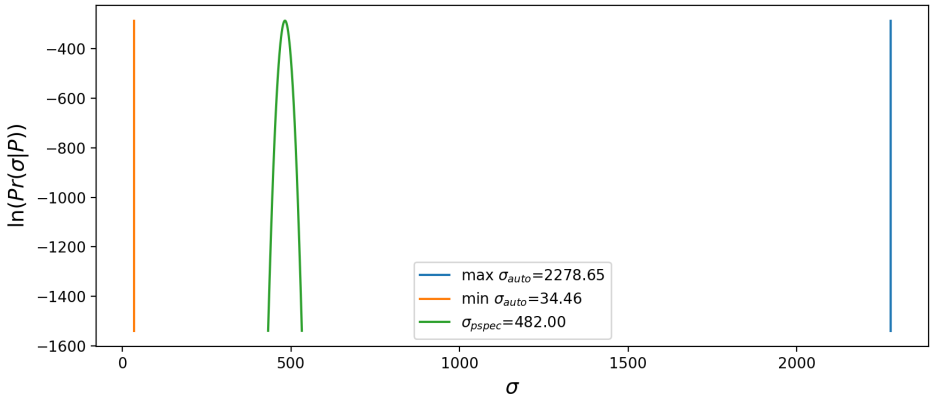


Figure 5: Minimum and maximum values of the standard deviation of time-averaged noise power predicted by auto-correlation visibilities, and the logarithm of the posterior probability function of the empirical standard deviation from fitting the $\mathcal{CNN}$ distribution to high-delay signals.

the power before normalization is

$$\sigma_p = \sqrt{\left(\frac{\sigma^2_{\tilde{n}}}{B_{full}}\right)^2}, \tag{5}$$

where B_{full} is the width of the full band.

The predicted σ_p is an array of values of standard deviation estimated from auto-correlated visibilities at each frequency in the band. We chose the minimum and the maximum values, and compared with the logarithm of a posterior probability function of the empirical standard deviation σ_e with a Gaussian PDF of σ_e as a prior. Figure 5 shows the log of the posterior probability function (green parabola), the maximum and minimum values of predicted σ_p. The probability function of the empirical σ falls inside the predicted region, thus σ_e and σ_p agree with each other to a certain extent.

3 Null tests for consistency in high delay signals

Null tests compute the difference between two data sets and show they are identical when the difference is zero. We constructed null tests between power spectra with different prominence of reflection features to determine the consistency between the data.

3.1 Modelling null test distribution

The null test was constructed between high-delay signals in the power spectra (without averaging by time or delays) formed from the baseline pairs [(83, 84), (83, 84)] (blp1) and [(66, 67), (66, 67)] (blp2). We also took the difference between data from the same delay range of the power spectra formed from blp1 and another pair of baselines [(37, 38), (37, 38)] (blp3). Among the three pairs, only blp2 has prominent features of cable reflections in its power spectrum. The Gaussian distribution was fitted to the difference signals for each group of baseline pairs. Since the difference signal is the difference between the real part of two complex double normal distributions, we also derived a PDF $\mathrm{pr}_{\Delta\mathcal{CNN}}$ for this distribution of differences (see derivation in Section 6.4):

$$\mathrm{pr}_{\Delta\mathcal{CNN}}(\Delta P(k)) = \frac{1}{2\sigma^2}\, e^{-\frac{2|\Delta P(k)|}{\sigma}}(\sigma + 2|\Delta P(k)|), \tag{6}$$

where σ is the standard deviation of the differences between two power spectra, $\Delta P(k)$ is the difference between two power values and $\Delta\mathcal{CNN}$ represents the distribution of the difference between two complex double normal distributions.

The above process was repeated for data taken from three groups of power spectra in order to obtain a more generalized statement of the consistency as well as the goodness of fit in terms of sample size. Group1 contains 11 power spectra with clear cable reflection features and Group2 contains 11 power spectra without clear cable reflection features. Group3 contains the same 11 power spectra in Group 2 but shifted by one index, for example, if group2 = [ps_1, ps_2, ..., ps_{11}], then group3 = [ps_{11}, ps_1, ..., ps_{10}]. Put another way, the aforementioned blp1, blp2 and blp3 correspondingly represent one of the baseline pairs that forms the power spectra in group1, group2 and group3. The null tests were performed between the data of different sizes taken from power spectra in Group1 and Group2, and between data from power spectra in Group2 and Group3. Just like obtaining the average test statistics described in Section 2.3, we calculated the baseline-averaged KS test statistics for null tests and plotted it as a function of the data size.

3.2 Results and Discussion

Figure 6 showed power spectra of the three pairs of baselines mentioned in Section 3.1, and the CDFs of the differences from null tests fitted with a Gaussian distribution and the $\Delta\mathcal{CNN}$ distribution described in Equation 6. From the plots of CDFs, one can observe that the systematic features do not affect the high-delay signals significantly by comparing the CDFs of differences from the two null tests. It is also visually apparent that the $\Delta\mathcal{CNN}$ distribution fits better to the differences than the Gaussian distribution.

The KS test statistics in Figure 7 provided a more detailed comparison between the goodness of fit for both distributions. At a 5% level of significance, the distribution of differences between either pair of power spectra does not differ from the given distributions. Recall that the KS test reports the maximum difference between the two CDFs, the smaller KS test statistics for fitting with the $\Delta\mathcal{CNN}$ distribution show that it is a better model for the differences between two power spectra at high delays than the Gaussian distribution. Since we have assumed that the mean is zero while deriving the $\Delta\mathcal{CNN}$ distribution and the differences fit well to a zero-mean distribution, the two data sets in one null test are statistically the same at a 5% significance level.

Figure 8 showed the baseline-averaged KS test statistics of fitting the $\Delta\mathcal{CNN}$ distribution to the null tests between high-delay data of different sizes taken from power spectra in group1 and group2, and between data from power spectra in group2 and group3. Unlike the baseline-averaged statistics in Figure 4, when we increases the data size by prolonging the delay range, some of the null tests involving data from power spectrum with clear systematic features fit worse with the $\Delta\mathcal{CNN}$ model than the null tests between data from power spectrum without clear reflection features. Nevertheless, overall there is no significant difference found in the fitting of null tests for power spectra with different prominence of cable reflection features.

4 Model comparison with different sample sizes

We sampled a relatively large amount of data from the power spectrum in Section 2 and the KS tests rejected the Gaussian model but accepted the $\mathcal{CNN}$ model. If we

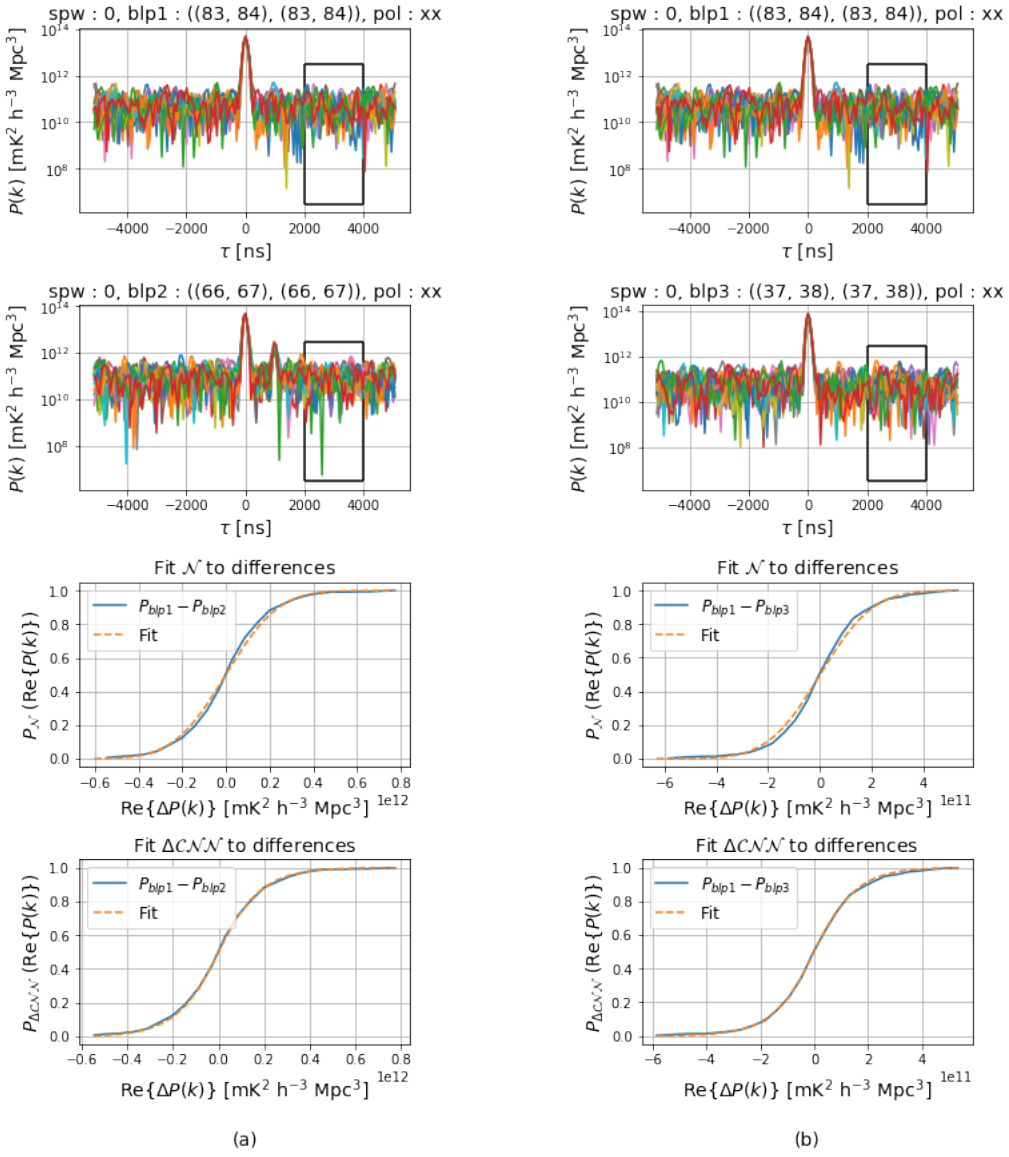


Figure 6: Power spectra in column (a) were formed from blp1 (top row) and blp2 (second row). Power spectra in column (b) were formed from blp1 (top row) and blp3 (second row). Note that only blp2 power spectrum shows clear cable reflection features. The null test were performed between real part of the power in the boxed region of the power spectra in the same column, and the CDFs of the fitted Gaussian distribution (the third row) and the $\Delta\mathcal{CNN}$ distribution (bottom row) were shown along with the data CDF.

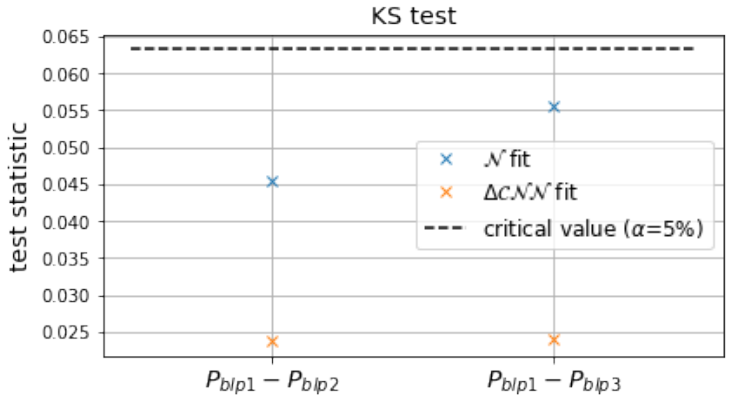


Figure 7: The KS test statistics shows goodness of fit in Figrue 6 at a 5% level of significance. Blue points correspond to the test statistics of comparing the Gaussian distribution with the distribution of differences between blp1 and blp2 (left) or the distribution of differences between blp1 and blp3 (right). Orange points correspond to the test statistics of comparing the $\Delta\mathcal{CNN}$ distribution with the distribution of the differences from null tests.

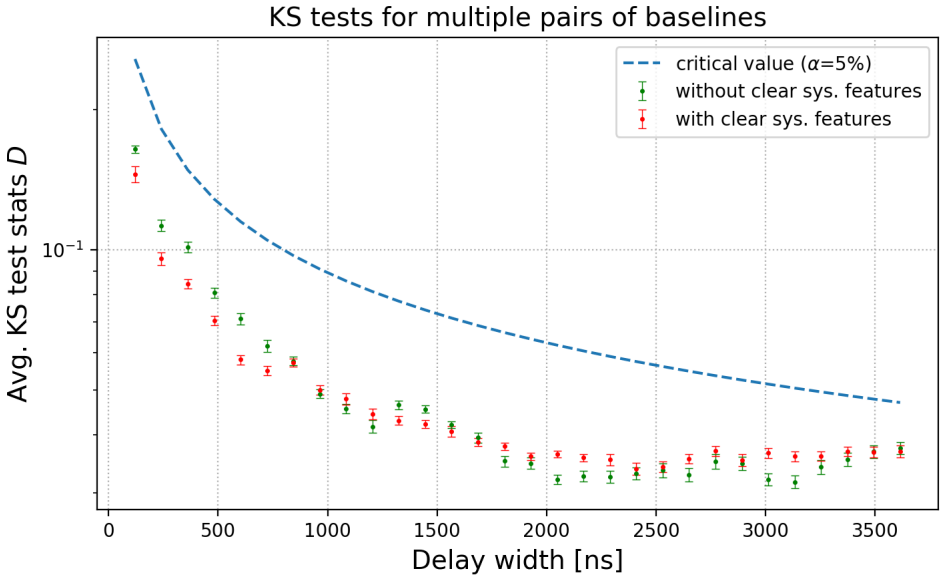


Figure 8: Baseline-averaged KS test stats for $\Delta\mathcal{CNN}$ fittings to null tests between high-delay data of different widths of delay from multiple power spectra without (green) and with (red) clear systematic features.

decrease the size of sampled data, will we still reach the same conclusion? We can ask the same question for the goodness-of-fit test of differences in Section 3 – the KS tests accepted both models, but would it be possible that it could reject one of the models when we increase the sample size? In other words, we wanted to know how robust the KS goodness-of-fit test is for model selection in terms of the sample size.

Due to the lack of information for true distribution of the data, we need another approach to hypothesis testing and the Bayesian model comparison was chosen here. We compared two models M_1 and M_2 with parameters θ_1 and θ_2 correspondingly based on an observed data set $\mathbf{D}$ by computing Bayes factors K given by

$$K = \frac{\Pr(\mathbf{D}|M_1)}{\Pr(\mathbf{D}|M_2)} = \frac{\Pr(M_1|\mathbf{D})}{\Pr(M_2|\mathbf{D})} \frac{\Pr(M_2)}{\Pr(M_1)}, \tag{7}$$

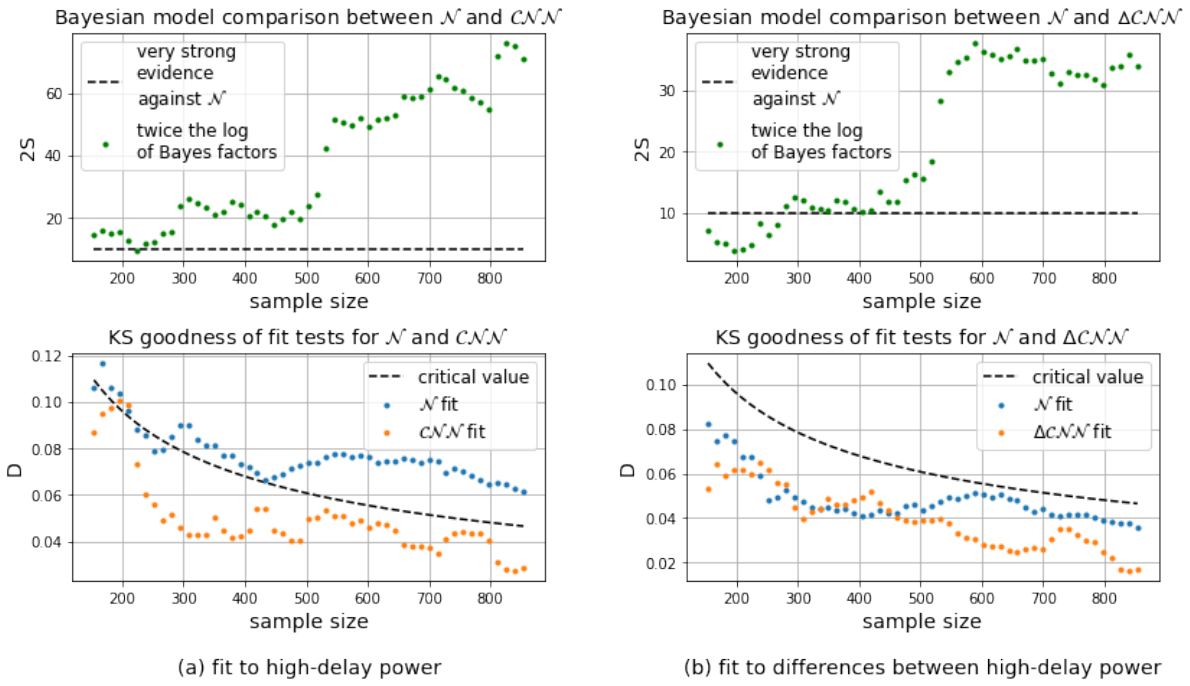
where $\Pr(\mathbf{D}|M_i)$ is a likelihood function, $\Pr(M_i|\mathbf{D})$ is the posterior probability and $\Pr(M_i)$ represents a prior probability of M_i. If we assume the two models are equally probable, the ratio of the priors will be 1 and Bayes factor will then be the ratio of the posterior probabilities of the two models. If K is greater than one, the probability of getting the data given M_1 is greater than that given M_2.

The Schwarz information criterion (SIC) is a reasonably good approximation to the logarithm of Bayes factor and it can avoid the introduction of priors [1]. The Schwarz criterion S is defined as

$$\log B \approx \mathrm{S} = \ln(\hat{L_1}) - \ln(\hat{L_2}) - \frac{1}{2}(d_1 - d_2)\ \ln(n),$$

where $\hat{L}_i$ is the maximized value of the likelihood function of the model M_i, n represents the sample size and d_i corresponds to the number of parameters estimated by the model. Once we have estimated the logarithm of the Bayes factor (S), we compare $2S$ with a scale of interpretation. For instance, a value of $2S$ greater than 10 means that M_1 is more strongly supported by the data than M_2 in Equation 7.

Figure 9(a) showed the model comparison between models fitted to the power values at high delays for blp2 after systematic model subtraction. The Bayes factors and KS tests for the fitting of the real part of power values generally agree with each other. Both



(a) fit to high-delay power

(b) fit to differences between high-delay power

Figure 9: Model comparison between the Gaussian distribution and each of the two derived distributions fitting to the data with different sample sizes. The data are (a) the real part of blp2 power values at high delays (left column), or (b) the differences from the null tests between blp1 and blp2 (right column). Recall only blp2 has clear cable reflection features in the delay spectrum. Interpretation key: the Gaussian model is less preferred when $2S$ is greater than the strong evidence line (dotted lines in top-row plots); the null hypothesis (no differences between the data CDF and the model CDF) is rejected when the KS-test stats are greater the critical values (dotted lines in bottom-row plots).

show that the $\mathcal{CNN}$ model is more strongly supported by the power values than the Gaussian model even when the sample size is relatively small.

Figure 9(b) showed the model comparison between models fitted to the differences between high-delay power values of blp1 and blp2 mentioned in Section 3.1 after systematic model subtraction. Only the Bayes factors of comparing the Gaussian model with the $\Delta\mathcal{CNN}$ model show that the later is more strongly supported by the data when the data was taken at a delay range greater than 1205 ns. However, the KS test only provides a stable preference between the two models for the data from a delay range greater than 2048 ns. Thus, we can see that the Bayes factor is a more robust model comparison method than the KS goodness-of-fit test.

5 Conclusion

Statistical tests were performed in order to investigate to what extent the reflection systematics affect high delay detections. We constructed null tests between power spectra with different reflection feature prominence, and found the null tests failed to provide strong enough evidence to support the statement that cable reflection systematics contaminate the high delay region significantly. We also derived two distributions to describe the real part of the power for noise signals and their differences. KS tests and Bayes factor estimation showed that both derived models fitted better to high delay detections than Gaussian distributions regardless of the prominence of reflection features in the power spectrum, and the Bayes factors give a more stable preference in model selection than the KS goodness-of-fit tests.

References

[1] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[2] M Sanders. Characteristic function of the central chi-squared distribution. Technical report, Retrieved 2009–03-06, 2009.

6 Appendix

6.1 Deriavation of Eq.1

Suppose N_1[1] and N_2 are two complex Gaussian random variables with zero expectation such that

$$n_1 = a_1 + ib_1$$
$$n_2 = a_2 + ib_2,$$

where a_1, a_2, b_1, b_2 are real coefficients and the letter i denotes $\sqrt{-1}$. We wish to compute the distribution of the real part of $N_1 N_2^*$. We have

$$\begin{aligned} n_1 n_2^* &= (a_1 + ib_1)(a_2 - ib_2) \\ &= a_1 a_2 + i a_2 b_1 - i a_1 b_2 + b_1 b_2 \\ &= a_1 a_2 + b_1 b_2 + i(a_2 b_1 - a_1 b_2), \end{aligned}$$

and its real part is

$$\mathrm{Re}\{n_1 n_2^*\} = a_1 a_2 + b_1 b_2.$$

Then, product distribution we want to find becomes the joint distribution of $A_1 A_2$ and $B_1 B_2$. Since A_1, A_2, B_1, B_2 are Gaussian random variables, $A_1 A_2$ and $B_1 B_2$ have the same probability density function (PDF). Here, we compute the distribution of $A_1 A_2$. We have

$$\begin{aligned} a_1 a_2 &= \frac{1}{4}[(a_1 + a_2)^2 - (a_1 - a_2)^2] \\ &\equiv \frac{1}{4}({a'_+}^2 - {a'_-}^2) \\ &\equiv \frac{1}{4}u. \end{aligned}$$

[1]We represent variables by capital letters, X, while lower-case letters, x, denote specific instances of a variable.

Similarly, we have $b_1 b_2 = \frac{1}{4}v$. Let m denote $u + v$, then

$$\mathrm{Re}\{n_1 n_2^*\} = \frac{1}{4}u + \frac{1}{4}v = \frac{1}{4}m. \tag{8}$$

We first compute the characteristic function of U and V, and then use convolution theorem to find probability function of M.

Let probability functions of U and $A_\pm'^2$ be $p_U(u)$ and $p_{A_\pm'^2}(a_\pm'^{\,2} = x)$ respectively, we then have

$$p_U(u) = \int_{-\infty}^{\infty} p_{A_+'^2}(u + x) \cdot p_{A_-'^2}(x) dx \tag{9}$$

We can represent $p_{A_\pm'^2}(x)$ by their corresponding inverse Fourier transformed characteristic function $\tilde{p}_{A_\pm'^2}(k)$:

$$p_{A_\pm'^2}(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{-ikx}\, \tilde{p}_{A_\pm'^2}(k). \tag{10}$$

Substituting Eq.(10) into Eq.(9) with proper change of Fourier space variables gives

$$\begin{aligned}
p_U(u) &= \int_{-\infty}^{\infty} dx\, \frac{dk}{2\pi}\frac{dq}{2\pi}\, e^{ik(u+x)}\, \tilde{p}_{A_+'^2}(k) e^{iqx}\, \tilde{p}_{A_-'^2}(q) \\
&= \int_{-\infty}^{\infty} dx\, e^{i(k+q)x} \int_{-\infty}^{\infty} \frac{dk}{2\pi}\frac{dq}{2\pi}\, e^{iku}\, \tilde{p}_{A_+'^2}(k)\, \tilde{p}_{A_-'^2}(q) \\
&= 2\pi\, \delta(k+q) \int_{-\infty}^{\infty} \frac{dk}{2\pi}\frac{dq}{2\pi}\, e^{iku}\, \tilde{p}_{A_+'^2}(k)\, \tilde{p}_{A_-'^2}(q) \\
&= \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{iku}\, \tilde{p}_{A_+'^2}(k) \int_{-\infty}^{\infty} dq\, \tilde{p}_{A_-'^2}(q)\, \delta[q - (-k)] \\
&= \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{iku}\, \tilde{p}_{A_+'^2}(k)\, \tilde{p}_{A_-'^2}(-k).
\end{aligned}$$

However,

$$p_U(u) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{iku}\, \tilde{p}_U(k),$$

then,

$$p_U(u) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{iku}\, \tilde{p}_U(k) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{iku}\, \tilde{p}_{A_+'^2}(k)\, \tilde{p}_{A_-'^2}(-k),$$

and thus

$$\tilde{p}_U(k) = \tilde{p}_{A'^2_+}(k)\, \tilde{p}_{A'^2_-}(-k). \tag{11}$$

Since A_1 and A_2 are Gaussian distributions, their linear combination $A'_\pm = A_1 \pm A_2$ is also a Gaussian distribution. Then, by definition of chi-square distribution, the distribution of $A'^2_\pm$ is a χ^2 distribution with 1 degree of freedom. The characteristic function of a χ^2 distribution $\tilde{p}_{A'^2_\pm}(\pm k)$ is

$$\tilde{p}_{A'^2_\pm}(\pm k) = \frac{1}{\sqrt{1 \mp i2\sigma^2_\pm k}}, \tag{12}$$

as proved by [2], where $\sigma_\pm$ is the standard deviation of distribution $A'_\pm$. Assuming $\sigma_+ = \sigma_-$ (σ will not have the same unit as $\sigma_\pm$ if they have units) and substituting (12) into (11) gives

$$\tilde{p}_U(k) = \frac{1}{\sqrt{1 - i2\sigma^2_+ k}} \frac{1}{\sqrt{1 + i2\sigma^2_- k}} \overset{\sigma \equiv \sigma_+ \sigma_-}{=} \frac{1}{\sqrt{1 + 4\sigma^2 k^2}}, \tag{13}$$

where σ is the standard deviation of distribution $A'^2_\pm$. Now, we can compute the probability function of M by first computing its characteristic function and then inverse Fourier transforming it to the real space. To get the characteristic function, we square Eq.(13), since 1) we find the joint distribution between U and V by convolving their PDFs, 2) the Fourier transform of a convolution of two PDFs is the pointwise product of their Fourier transforms, which is their corresponding characteristic functions, and 3) Eq.(13) is the characteristic function for both U and V. Thus, squaring Eq.(13) gives

$$\tilde{p}_M(k) = \left(\frac{1}{\sqrt{1 + 4\sigma^2 k^2}} \right)^2 = \frac{1}{1 + 4\sigma^2 k^2}. \tag{14}$$

We then inverse Fourier transform the above characteristic function to get its probability function $p_M(m)$.

$$p_M(m) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, e^{ikm} \frac{1}{1 + 4\sigma^2 k^2} = \frac{1}{4\sigma} e^{-\frac{|m|}{2\sigma}} \text{ (see proof in Section 6.4).} \tag{15}$$

Recall Eq.8, there is a coefficient in front of m. We thus modify Eq.15 to get the proba-

bility function of $N_1 N_2^*$:

$$p_{N_1 N_2^*}(x) = \frac{1}{\frac{1}{4}} \times p_M\left(\frac{m}{\frac{1}{4}}\right) = \frac{1}{\sigma}\, e^{-\frac{2|x|}{\sigma}}. \tag{16}$$

Eq.16 is normalized since

$$\int_{-\infty}^{\infty} p_{N_1 N_2^*}(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sigma}\, e^{-\frac{2|x|}{\sigma}}\, dx = 1.$$

Therefore, the PDF $f(x)$ of $\mathrm{Re}\{N_1 N_2^*\}$ distribution is

$$f(x) = \frac{1}{\sigma}\, e^{-\frac{2|x|}{\sigma}}.$$

6.2 Proof for Eq.15

We prove Eq.15 by Fourier transforming the resulting function. We define $f(x)$ as

$$f(x) = p_M(m = x) = \frac{1}{4\sigma} e^{-\frac{|x|}{2\sigma}}.$$

Fourier transforming $f(x)$ gives

$$\begin{aligned}
\tilde{f}(k) &= \int_{-\infty}^{\infty} f(x)\, e^{-ikx}\, dk \\
&= \int_{-\infty}^{\infty} \frac{1}{4\sigma} e^{-\frac{|x|}{2\sigma}}\, e^{-ikx}\, dk \\
&= \frac{1}{4\sigma} \left(\int_{-\infty}^{0} e^{\frac{x}{2\sigma}}\, e^{-ikx}\, dk + \int_{0}^{\infty} e^{-\frac{x}{2\sigma}}\, e^{-ikx}\, dk \right) \\
&= \frac{1}{4\sigma} \left(\left[\frac{e^{\frac{x}{2\sigma} - ikx}}{\frac{1}{2\sigma} - ik} \right]_{-\infty}^{0} + \left[\frac{e^{-\frac{x}{2\sigma} - ikx}}{-\frac{1}{2\sigma} - ik} \right]_{0}^{\infty} \right) \\
&= \frac{1}{4\sigma} \left(\frac{1}{\frac{1}{2\sigma} - ik} + \frac{1}{\frac{1}{2\sigma} + ik} \right) \\
&= \frac{1}{4\sigma} \left[\frac{\frac{2}{2\sigma}}{\left(\frac{1}{2\sigma}\right)^2 + k^2} \right] \\
&= \frac{1}{4\sigma} \times \frac{\frac{2}{2\sigma}}{\left(\frac{1}{2\sigma}\right)^2} \times \frac{1}{1 + 4\sigma^2 k^2} \\
&= \frac{1}{1 + 4\sigma^2 k^2} = \tilde{p}_M(k).
\end{aligned}$$

6.3 Plots of CDFs and KS test statistics of high-delay signals before and after systematic subtraction

Fig.10 and Fig.11 show fittings for the power spectra of an auto-baseline pair with cable reflection features before and after subtracting a systematic model and KS test statistics for the goodness of fit.

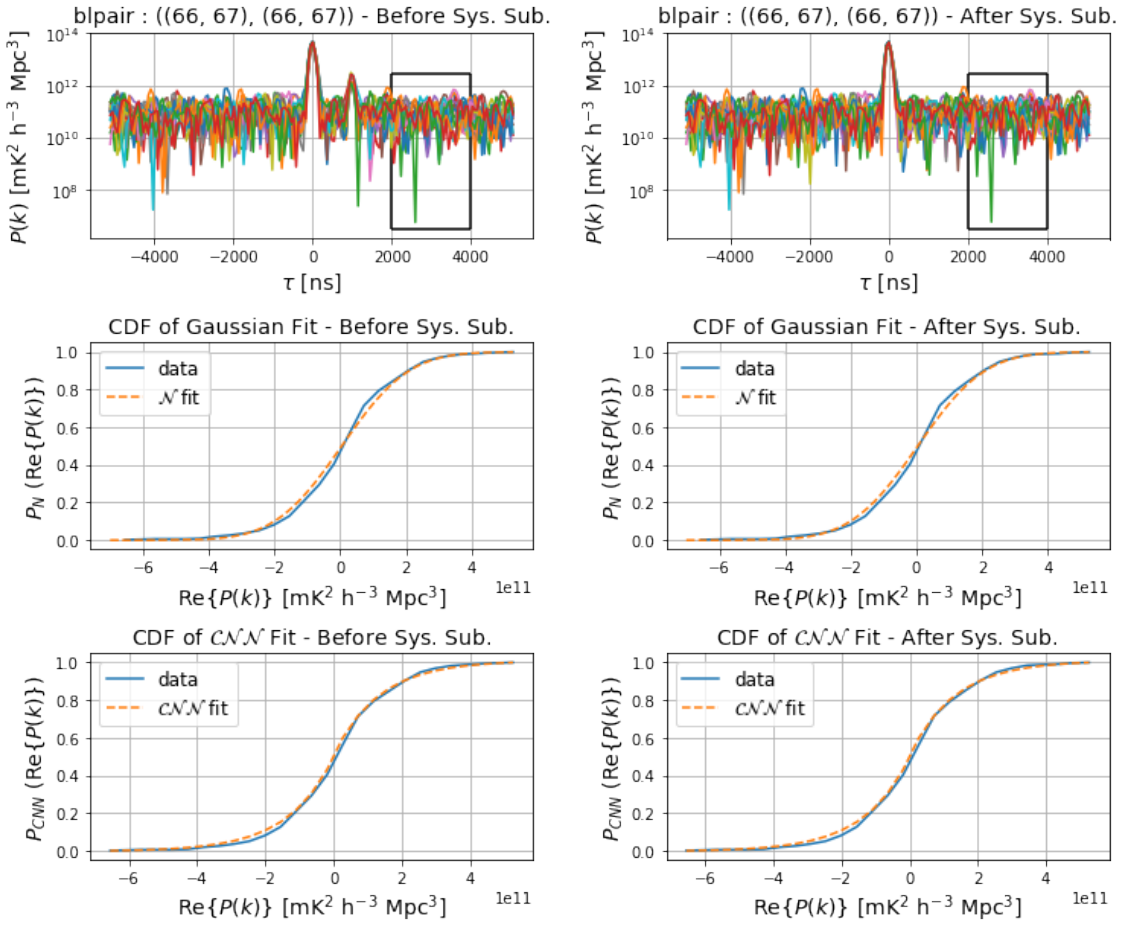


Figure 10: Power spectra (top row) formed from a baseline pair with cable reflections. High-delay signals in the boxed region were fitted with a Gaussian distribution (middle row) and a $\mathcal{CNN}$ distribution (bottom row) for power spectrum before (left column) and after (right column) subtracting the systematic model.

6.4 Derivation of Eq.6

Let D' be the difference between $4\mathrm{Re}\{N_1 N_2^*\}$ and $4\mathrm{Re}\{N_3 N_4^*\}$. Similarly, the characteristic function $\tilde{p}_{D'}(k)$ of distribution D' is the square of characteristic function of $4\mathrm{Re}\{N_i N_j^*\}$

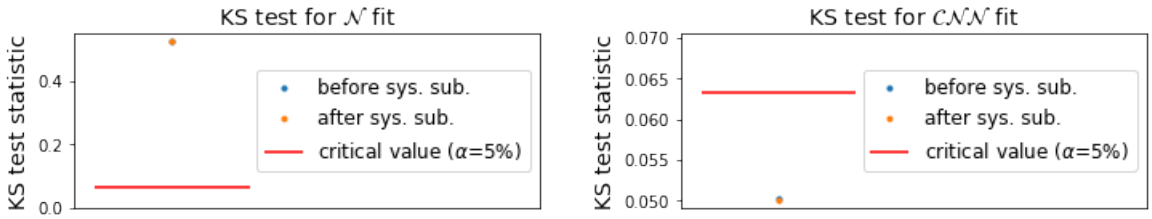


Figure 11: KS tests for normality (left) and the $\mathcal{CNN}$ distribution fitting results (right) of the power at delays 2000 - 4000 ns.

(Eq.14).

$$\tilde{p}_{D'}(k) = \left(\frac{1}{1 + 4\sigma_M^2 k^2}\right)^2, \tag{17}$$

where σ_M is the standard deviation of distribution $4\mathrm{Re}\{N_i N_j^*\}$ assuming they have the same variance. Let σ be the standard deviation of the difference distribution, inverse Fourier transforming Eq.17 gives the probability function $p'_D(x)$:

$$p'_D(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi}\, \tilde{p}_D(k)\, e^{ikx} = \frac{e^{-\frac{|x|}{2\sigma}}(2\sigma + |x|)}{16\sigma^2}. \tag{18}$$

The probability function $p_D(x)$ of the difference distribution between $\mathrm{Re}\{N_1 N_2^*\}$ and $\mathrm{Re}\{N_3 N_4^*\}$ is

$$p_D(x) = \frac{1}{\frac{1}{4}} * p'_D(\frac{x}{\frac{1}{4}}) = \frac{1}{2\sigma^2}\, e^{-\frac{2|x|}{\sigma}}(\sigma + 2|x|). \tag{19}$$

Eq.19 is normalized since

$$\int_{-\infty}^{\infty} p_D(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{2\sigma^2}\, e^{-\frac{2|x|}{\sigma}}(\sigma + 2|x|)dx = 1.$$

Thus, the PDF $f_D(x)$ of D distribution is

$$f_D(x) = \frac{1}{2\sigma^2}\, e^{-\frac{2|x|}{\sigma}}(\sigma + 2|x|).$$