HERA Memo #092
Received 3 Feb 2021

Noise in Delay Spectrum After Systematics Subtraction

Taylor Dibblee-Barkman

Supervised by Adrian Liu and Saurabh Singh

August 25, 2020

This memo presents the results of comparing delay spectra with systematics due to cable reflections before and after the systematics are subtracted to an analytic model of noise in the delay spectrum. The delay spectra generated from data in Sec. 1 can be compared visually in Sec. 2. Sec. 3 shows the cumulative distribution functions (CDFs) of powers for a sample of delays. Sec. 4 shows a statistical comparison of each set of delay spectra to the theory model with a Kolmogorov-Smirnov (KS) test for all delays. The data path is `/lustre/aoc/projects/hera/H1C_IDR2/IDR2_2_pspec/v2/one_group/data`.

Contents

1 Description of Datasets

The three datasets used in this analysis are described as follows: real HERA data with features in the delay spectra due to foregrounds and systematics (A, filename = `"zen.grp1.of1.LST.1.31552.HH.OCRSLP2.uvh5"`), the same dataset after the first-order subtraction of systematics due to cross-coupling (B, filename = `"zen.grp1.of1.LST.1.31631.HH.OCRSLP2X.uvh5"`), and data that simulates noise using Gaussian variables (E, link = `https://github.com/HERA-Team/H1C_IDR2/blob/master/notebooks/null_tests/making_noise_sims.ipynb`), which is referred to as "forward-modelled noise" in this memo. We also obtained a systematics-only dataset (C) by subtracting the visibilities B from A. The following waterfall plots show the visibilities for spectral windows (spw) 175-335 and 515-695, and baseline pairs (blp) ((36, 37), (36, 36)) and ((51, 52), 51, 52)). Note that the delay spectra in Sec. 2 are generated using auto-baseline and time-interleaving.

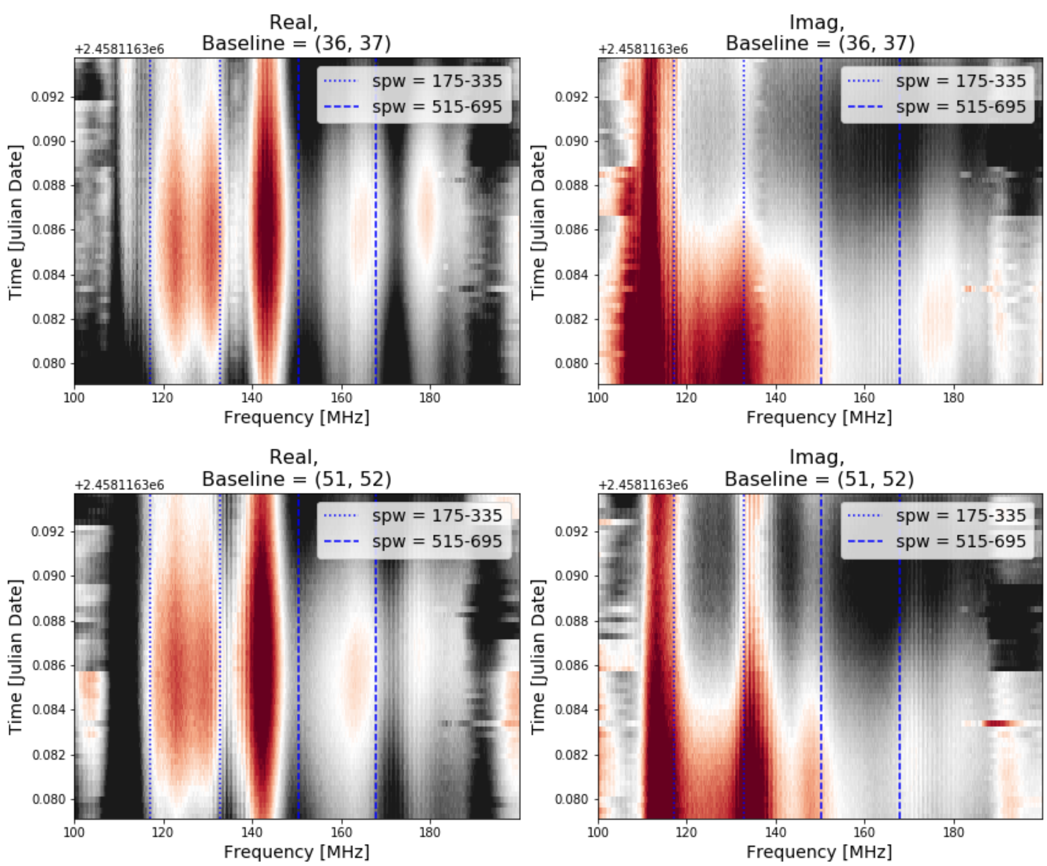


Figure 1: Real and imaginary components of visibilities from the systematics-preserved dataset (A).

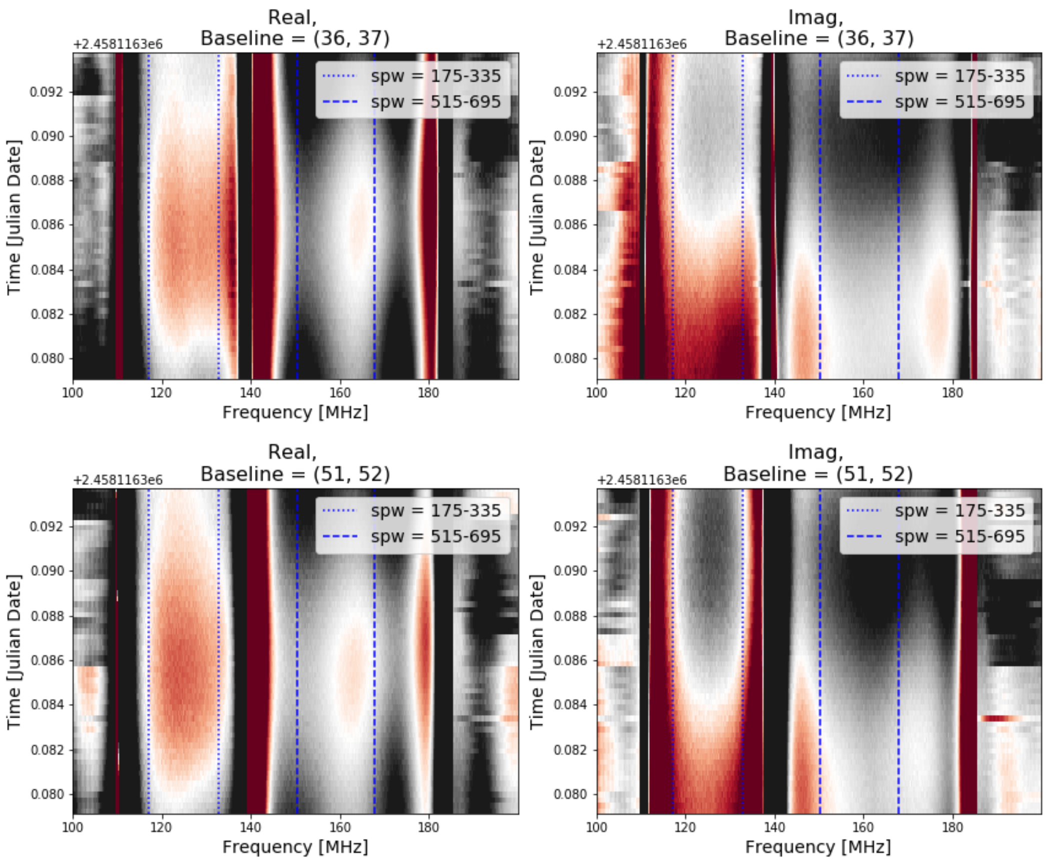


Figure 2: Real and imaginary components of visibilities from the systematics-subtracted dataset (B).

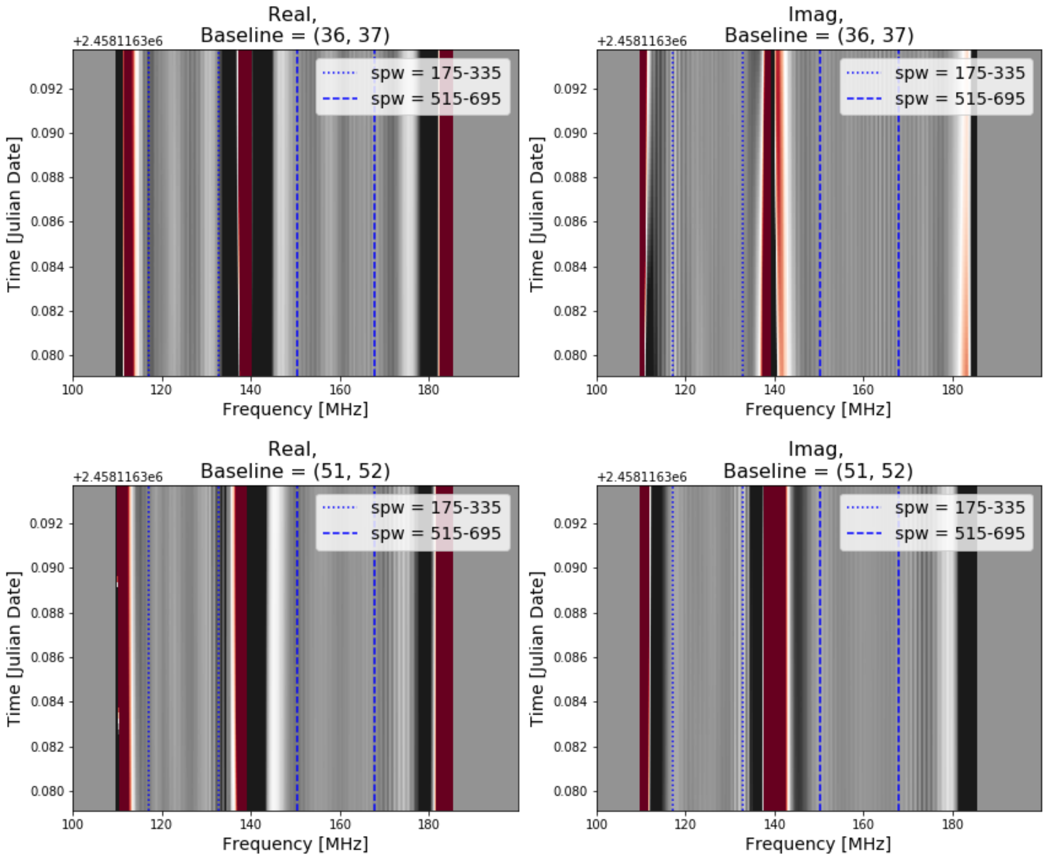


Figure 3: Real and imaginary components of visibilities from the systematics-only dataset (C) obtained by subtracting B from A.

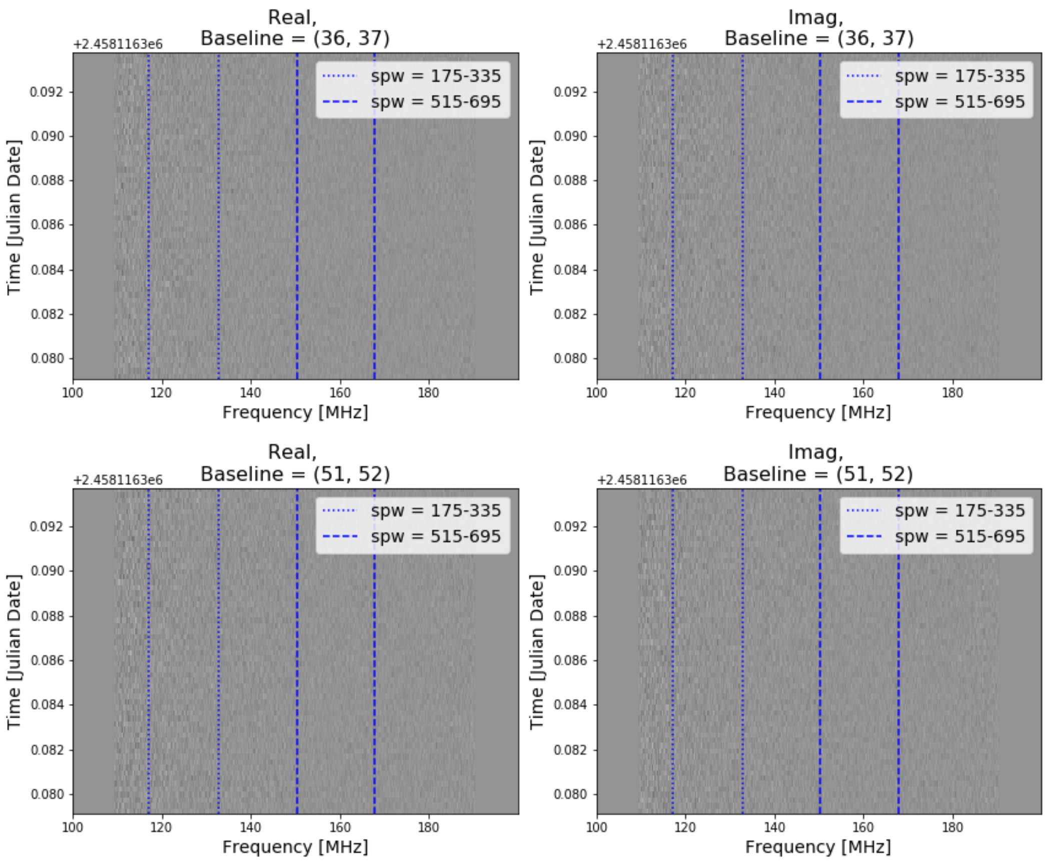


Figure 4: Real and imaginary components of visibilities from the "forward-modelled" noise dataset (E).

2 Delay Spectra

Delay spectra were generated from the datasets described above (A, B, C, E). Another systematics-only set of delay spectra (D) was plotted by subtracting the powers generated from datasets A and B (P(A)-P(B)). Both sets of delay spectra showing systematics-only show that the subtraction is locally peaked with non-local tails. In addition, we generated a set of delay spectra simulating "ideal noise" (F) by randomly drawing powers from a double decaying exponential with time-dependent decay coefficients $\lambda(t)$, which are computed from the "forward-modelled" noise delay spectra (E) at each time sample. The probability distribution for each time sample is given by:

$$p_t(x) = \frac{\lambda(t)}{2} e^{\lambda(t)|x|}, \tag{1}$$

where x is power, $\lambda(t)=\sqrt{2}/\sigma(t)$, and $\sigma(t)$ is the standard deviation in power at each time sample from the set of non-ideal noise delay spectra.

spw = 175-335, blp = ((36, 37), (36, 37)):

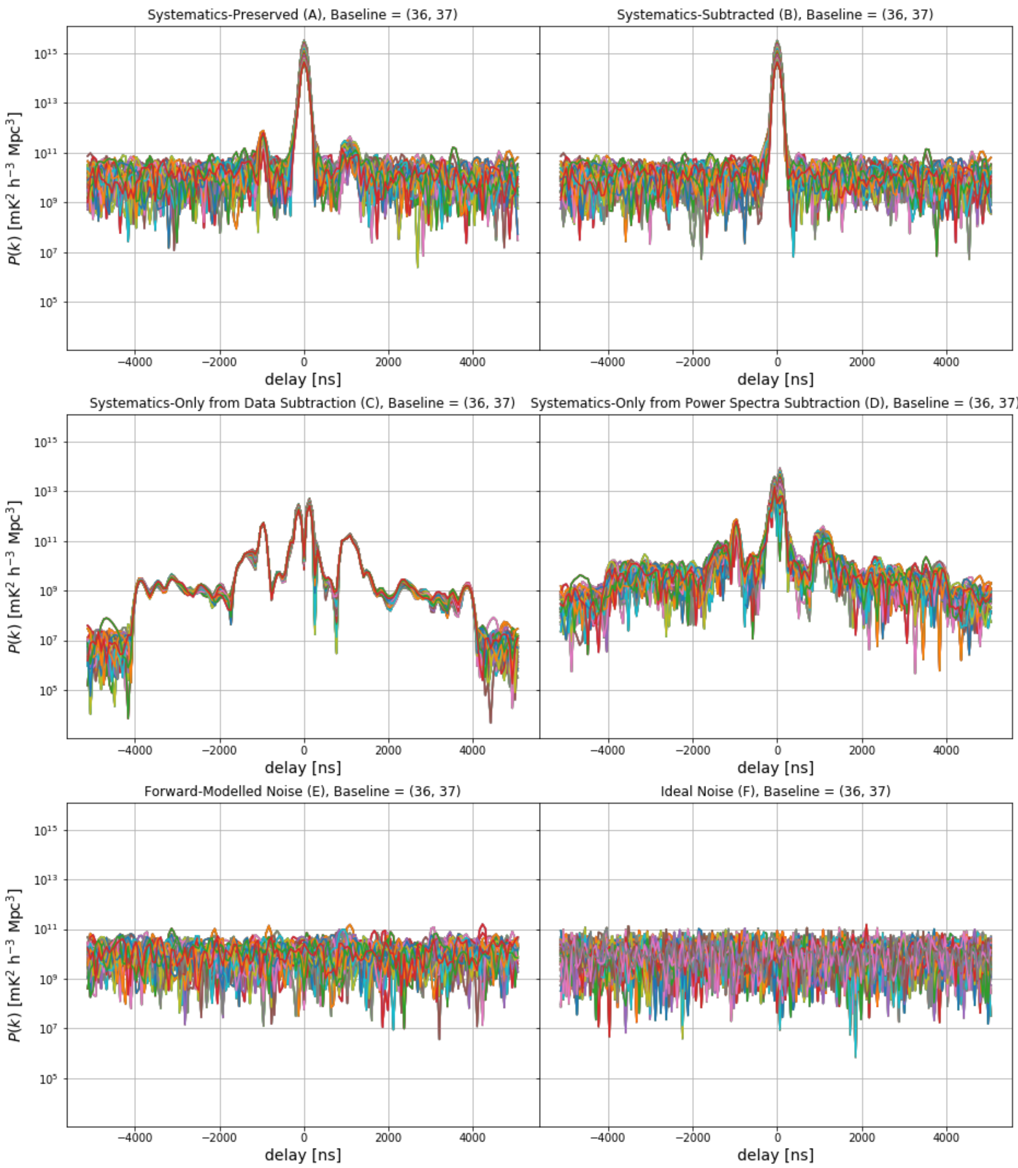


Figure 5: Delay spectra generated from datasets A, B, C, E. Delay spectra D obtained by subtracting spectra generated from B from spectra generated from A. Delay spectra F obtained by randomly drawing powers from double decaying exponential with time-dependent decay coefficients.

spw = 175-335, blp = ((51, 52), (51, 52)):

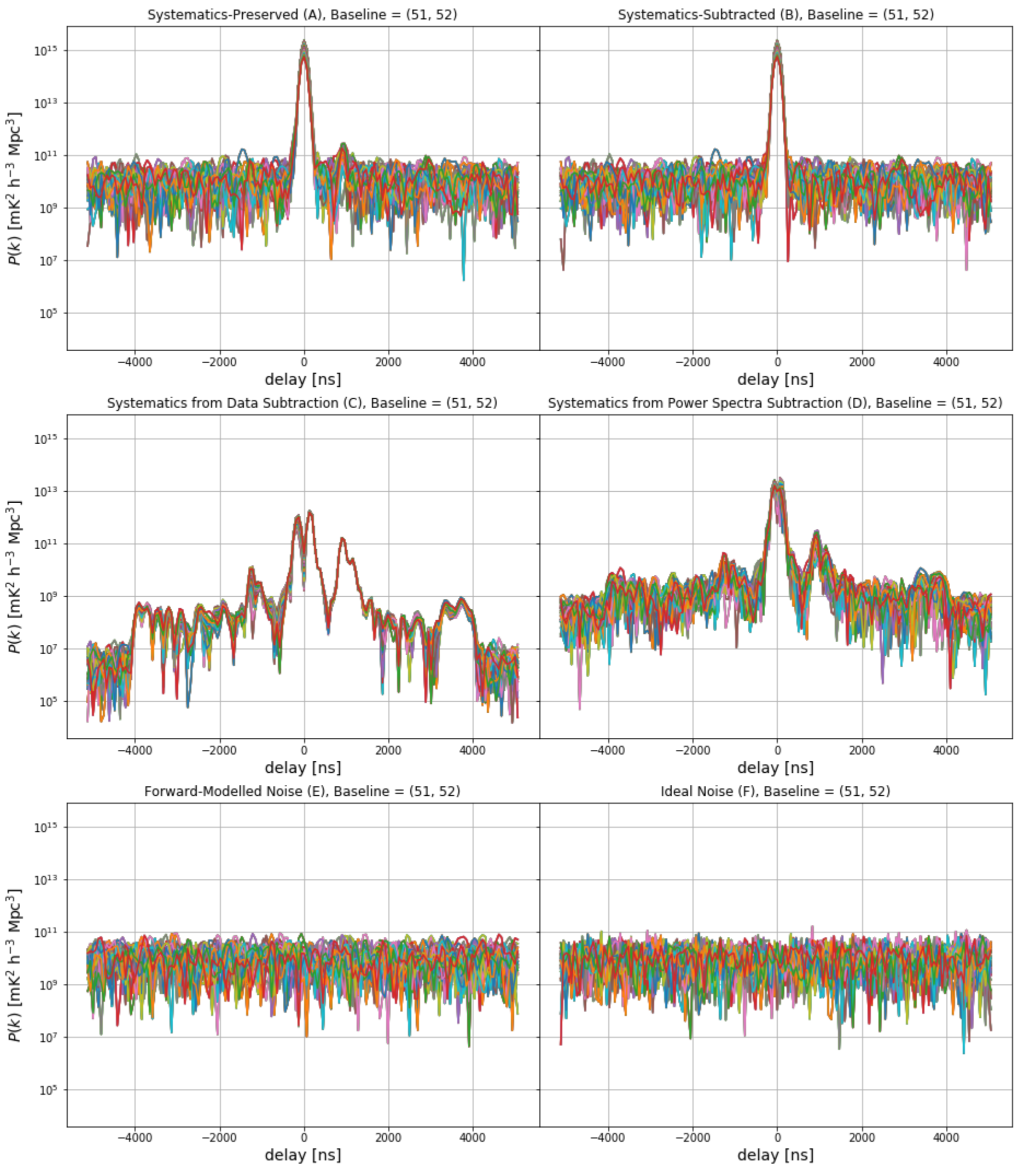


Figure 6: Delay spectra generated from datasets A, B, C, E. Delay spectra D obtained by subtracting spectra generated from B from spectra generated from A. Delay spectra F obtained by randomly drawing powers from double decaying exponential with time-dependent decay coefficients.

spw = 515-695, blp = ((36, 37), (36, 37)):

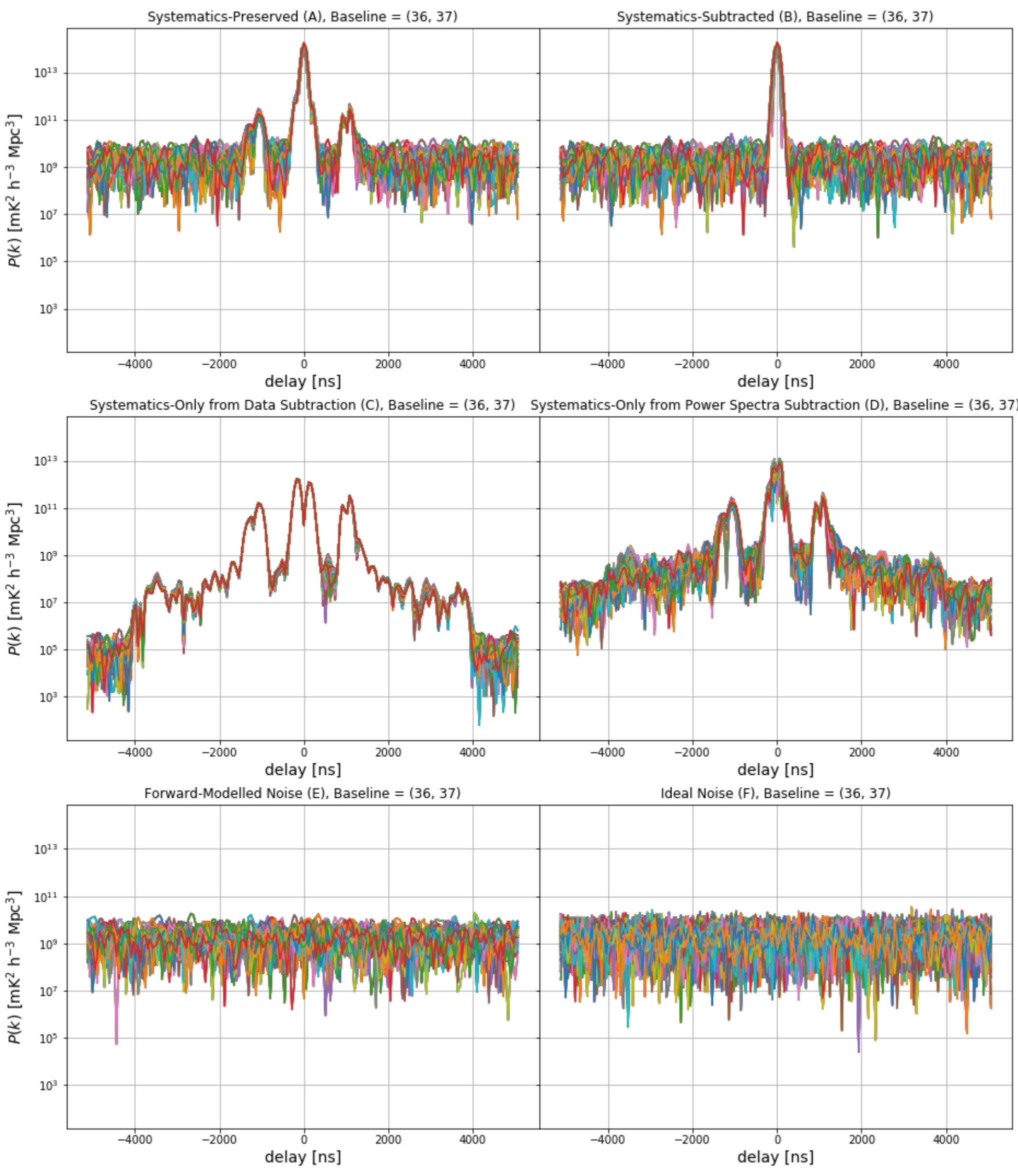


Figure 7: Delay spectra generated from datasets A, B, C, E. Delay spectra D obtained by subtracting spectra generated from B from spectra generated from A. Delay spectra F obtained by randomly drawing powers from double decaying exponential with time-dependent decay coefficients.

spw = 515-695, blp = ((51, 52), (51, 52)):

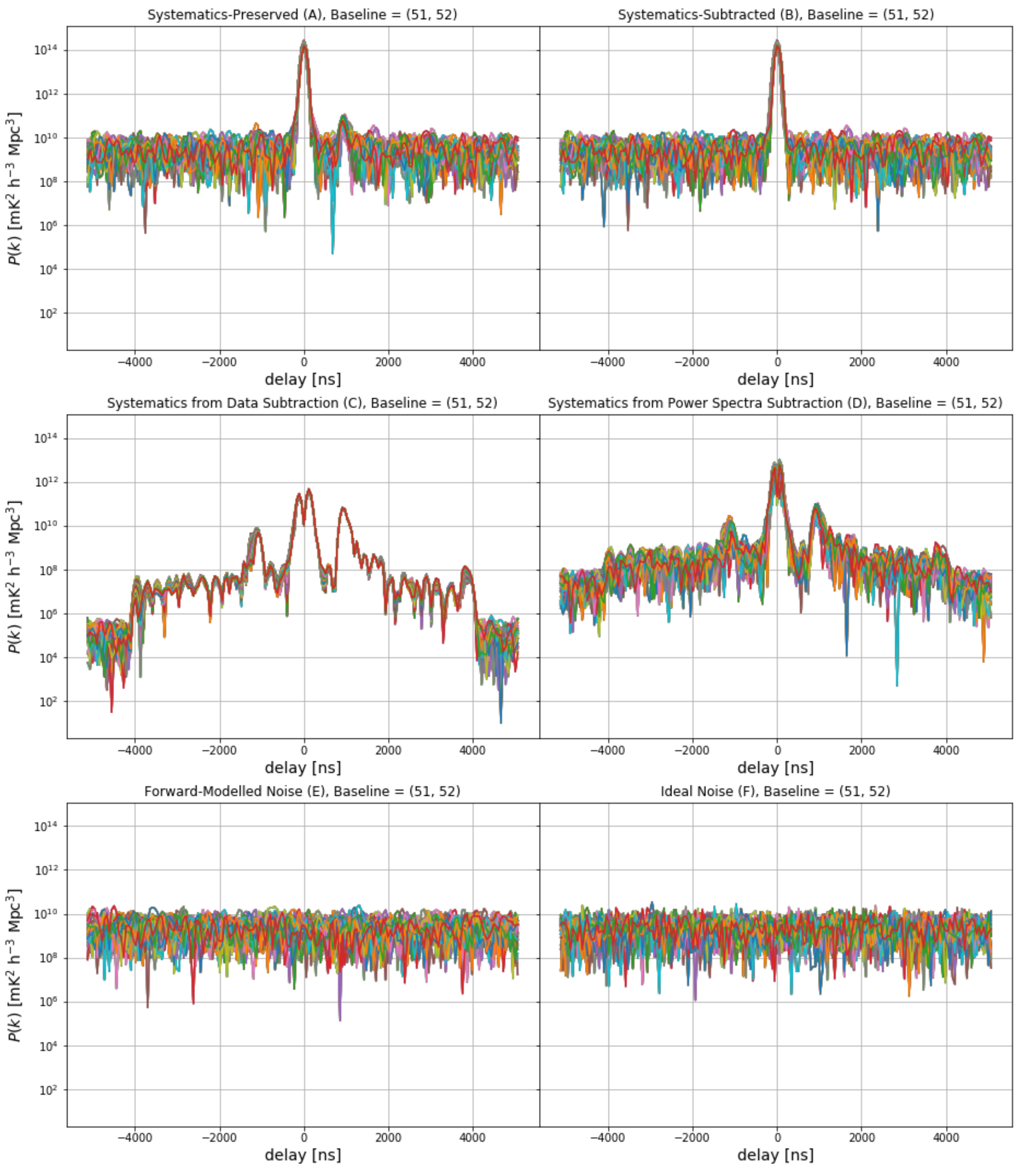


Figure 8: Delay spectra generated from datasets A, B, C, E. Delay spectra D obtained by subtracting spectra generated from B from spectra generated from A. Delay spectra F obtained by randomly drawing powers from double decaying exponential with time-dependent decay coefficients.

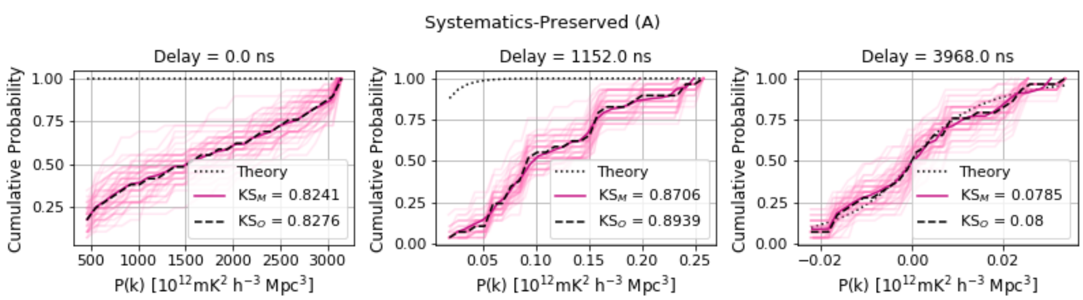
3 Bootstrapped CDFs

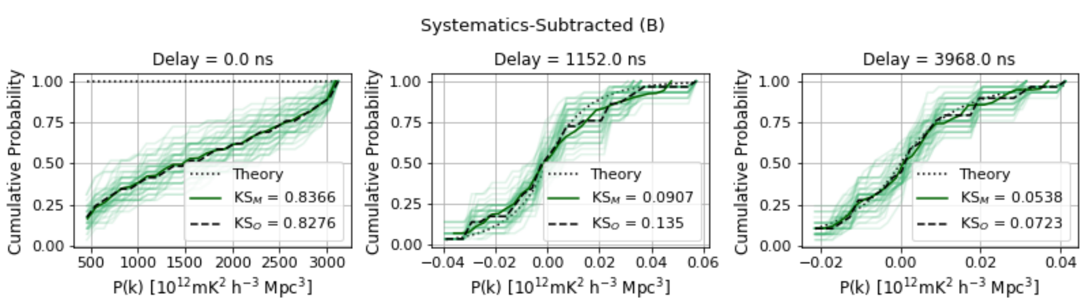
CDFs generated from each of the above sets of delay spectra are shown for a sample of three delays. These are chosen based on where the features from foregrounds, systematics, and noise dominate the delay spectra generated from dataset A. Bootstrapped CDFs are shown with semi-transparent curves, which are generated by randomly drawing from the original sample of powers with repeats allowed. The legend shows the KS results from comparing the theory CDF to the mean (KS_M) over the bootstrapped CDFs or the CDF generated from the original (KS_O) sample of powers. The theory CDF is given by:

$$\mathcal{P}(x) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} \frac{e^{\lambda(t_i)x}}{2} & \text{if } x < 0, \\ \frac{1}{N}\sum_{i=1}^{N} \left(1 - \frac{e^{-\lambda(t_i)x}}{2}\right) & \text{if } x > 0, \end{cases} \tag{2}$$

where N is the number of time samples (in this case, we have N=29 samples of delay spectra), x is power, and t_i is the i$^{\text{th}}$ time sample.

spw = 175-335, blp = ((36, 37), (36, 37)):





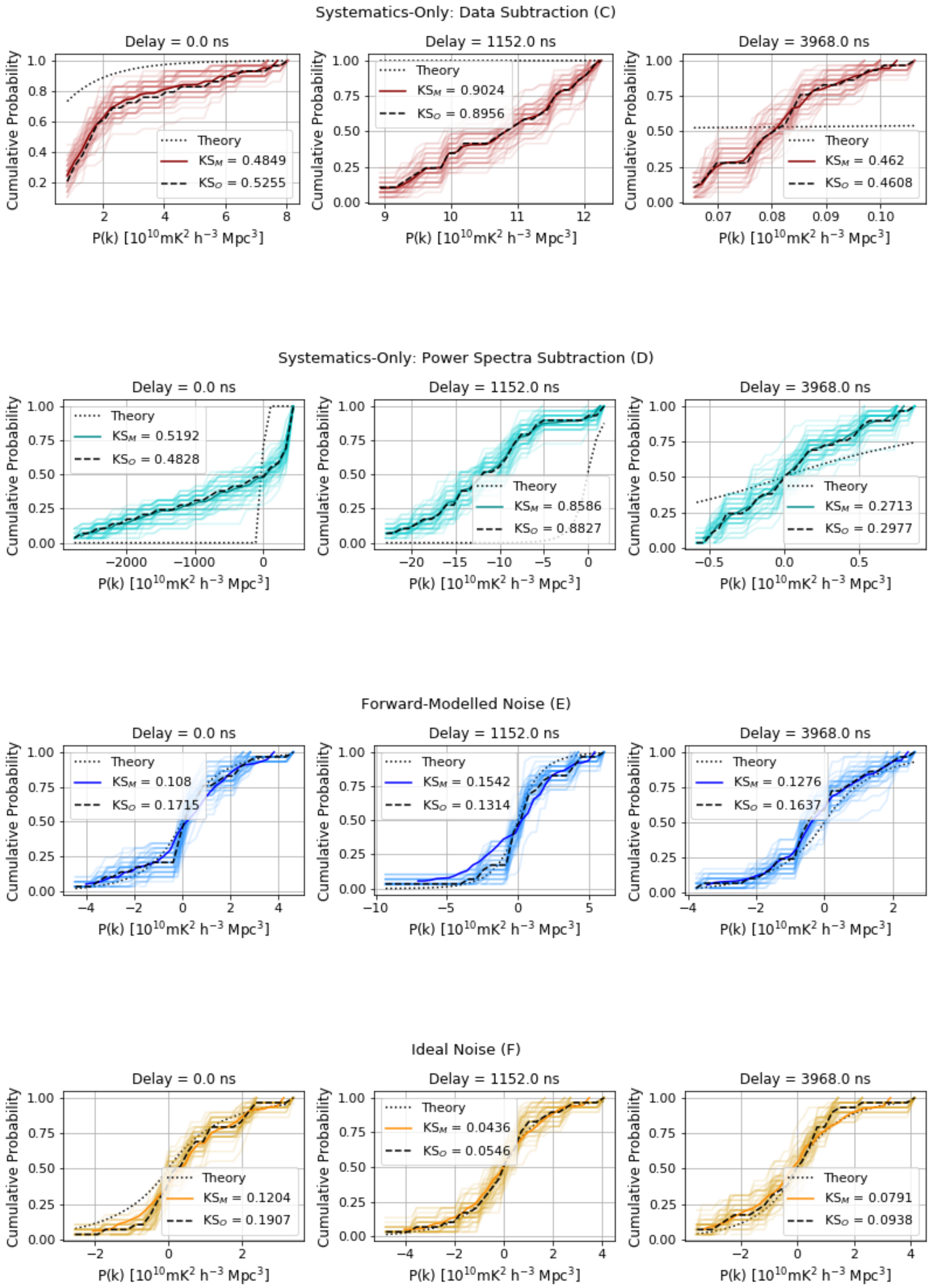
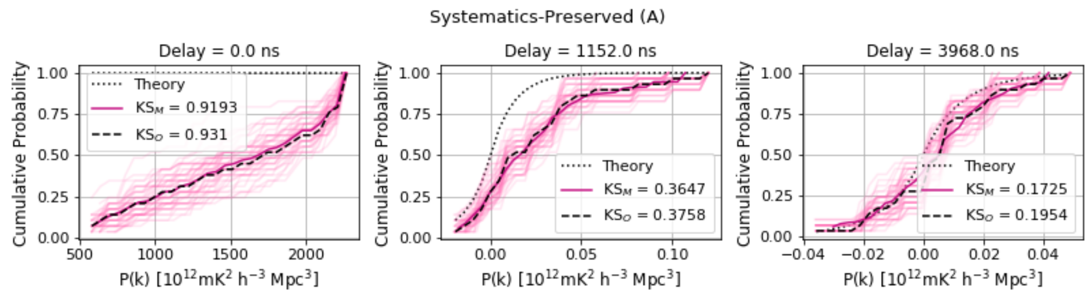
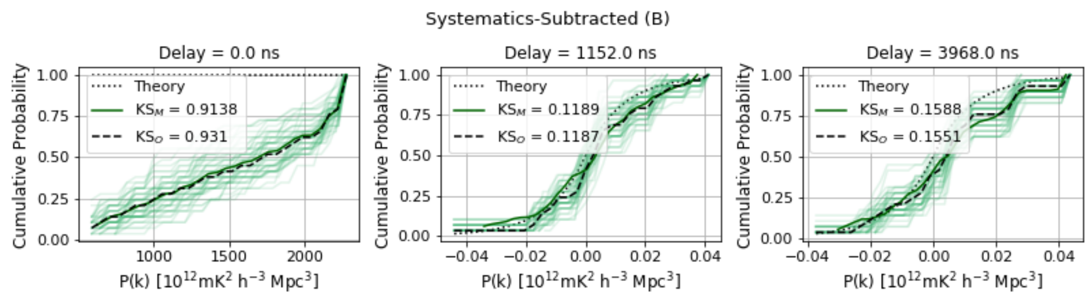
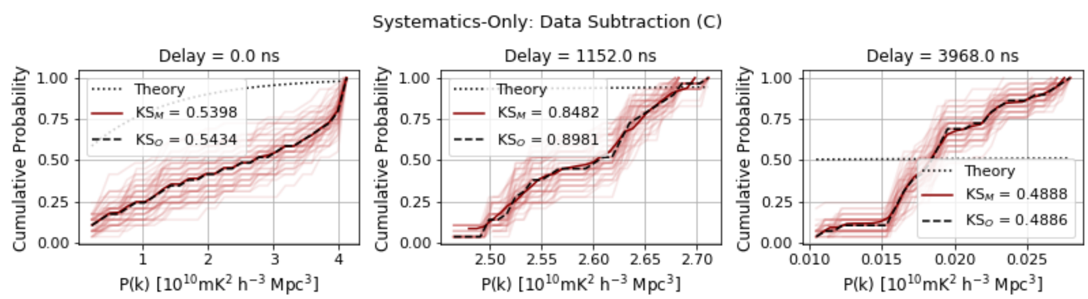


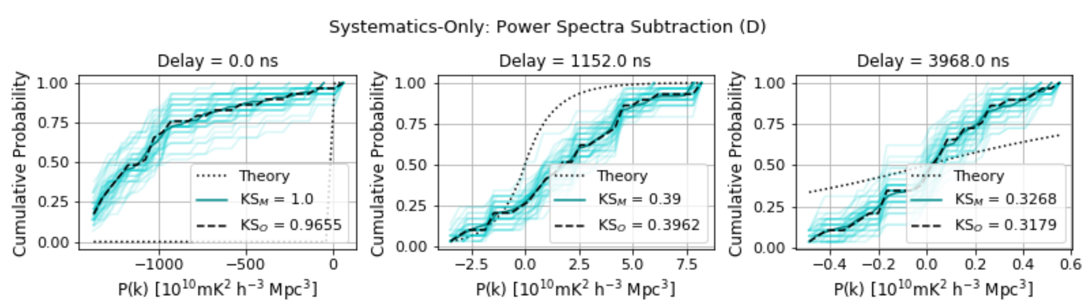
Figure 9: Sample of bootstrapped CDFs (transcluscent), mean CDF taken over bootstraps (solid colour), CDF generated from the original sample of powers (dashed black), theory CDF (dotted black).

spw = 175-335, blp = ((51, 52), (51, 52)):









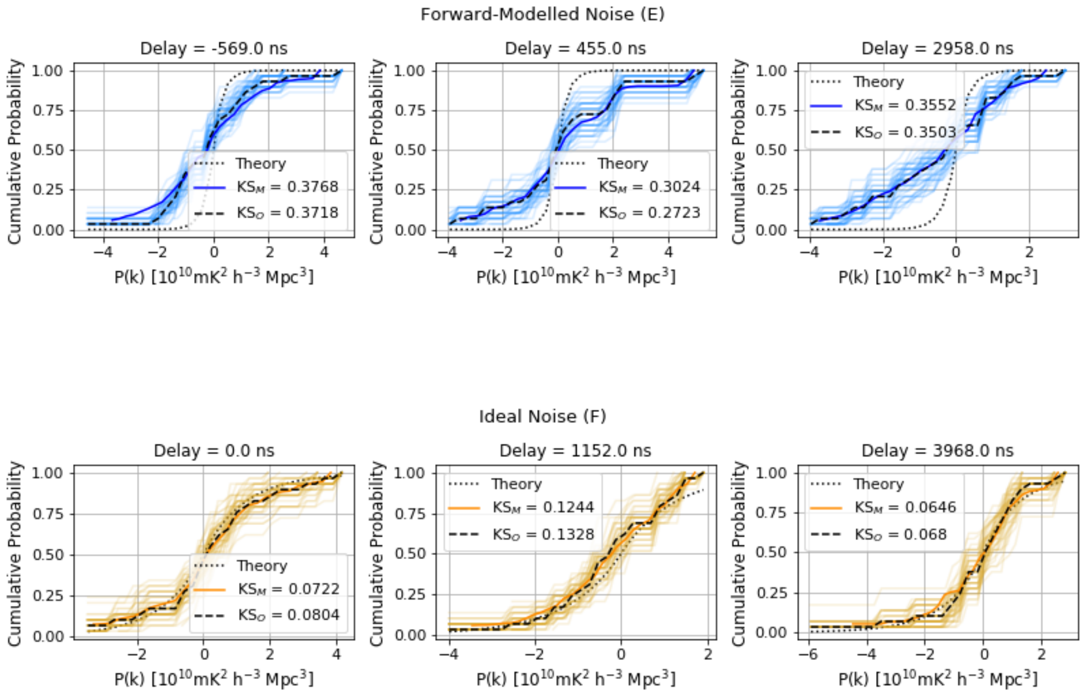
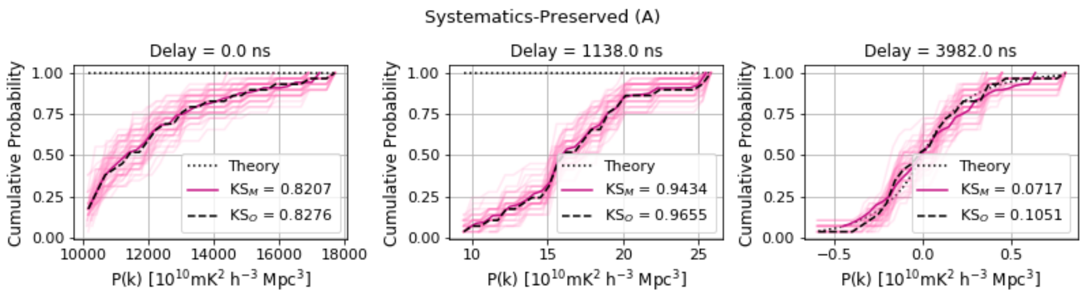
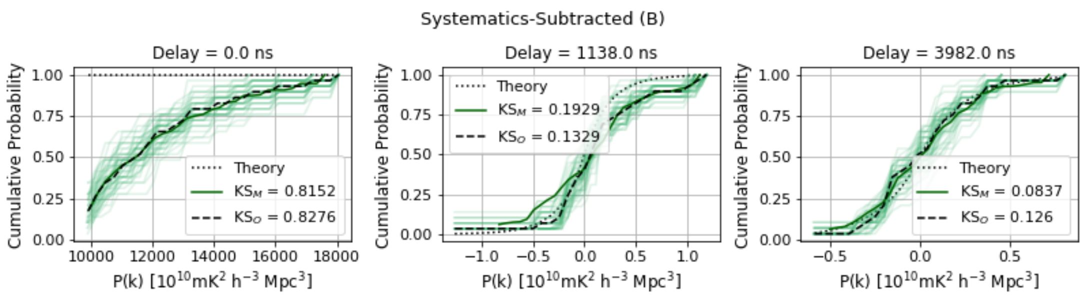
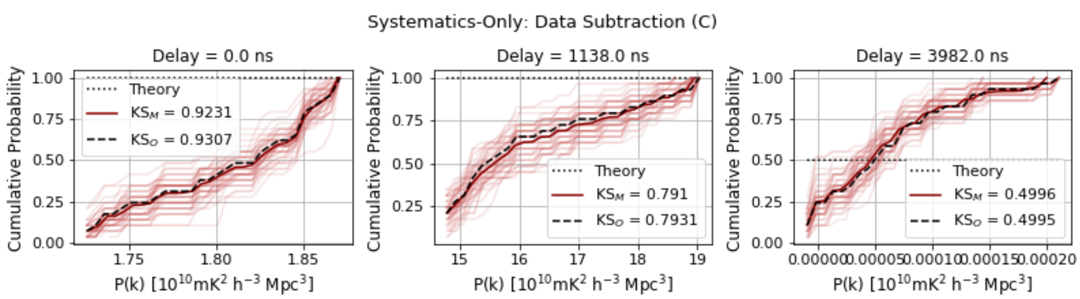


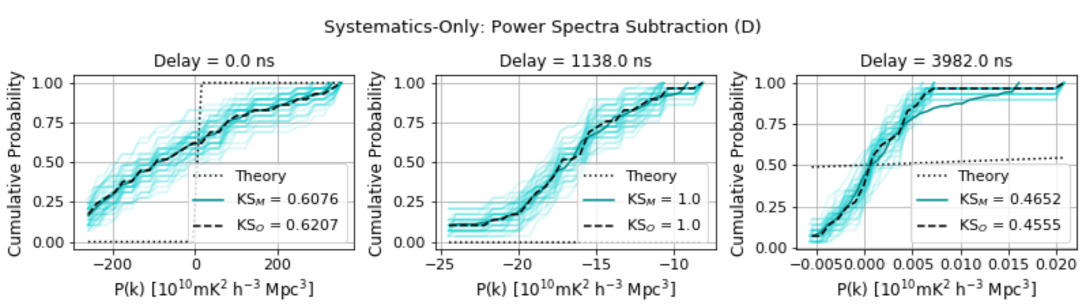
Figure 10: Sample of bootstrapped CDFs (transcluscent), mean CDF taken over bootstraps (solid colour), CDF generated from the original sample of powers (dashed black), theory CDF (dotted black).

spw = 515-695, blp = ((36, 37), (36, 37)):









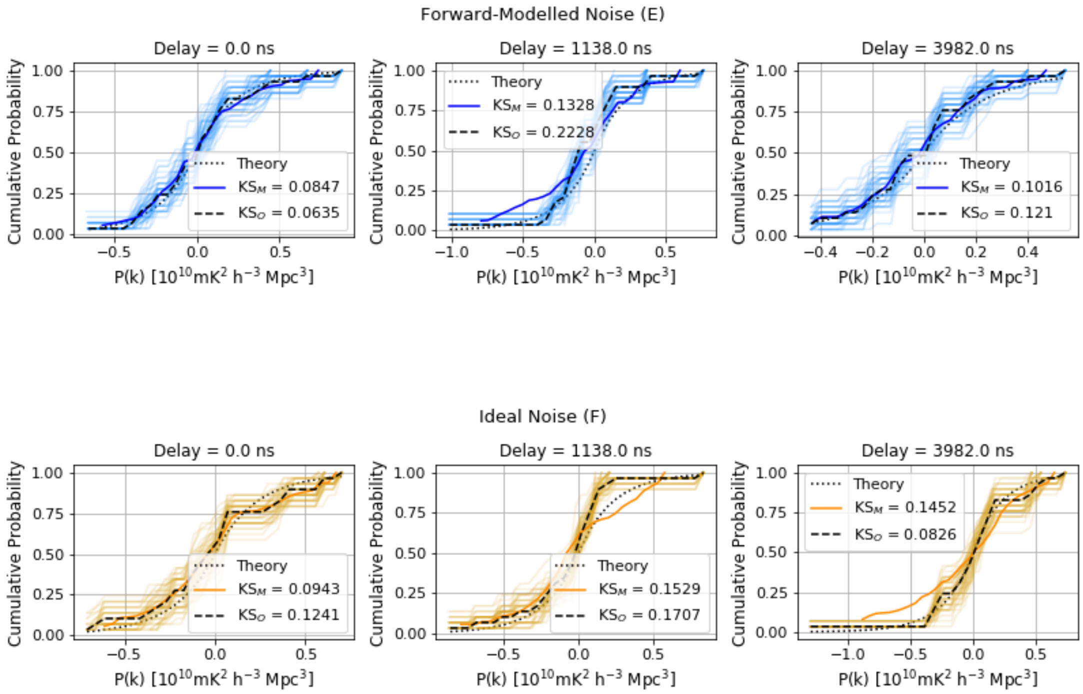
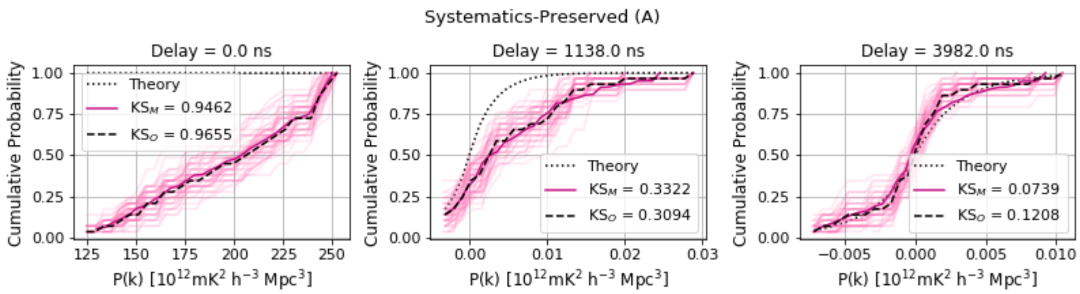
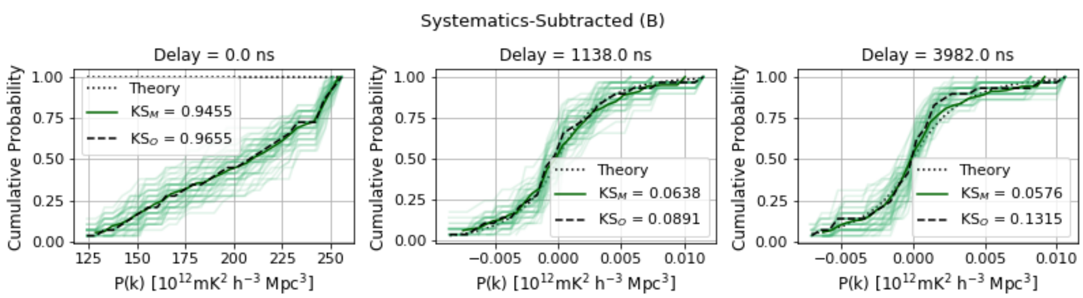
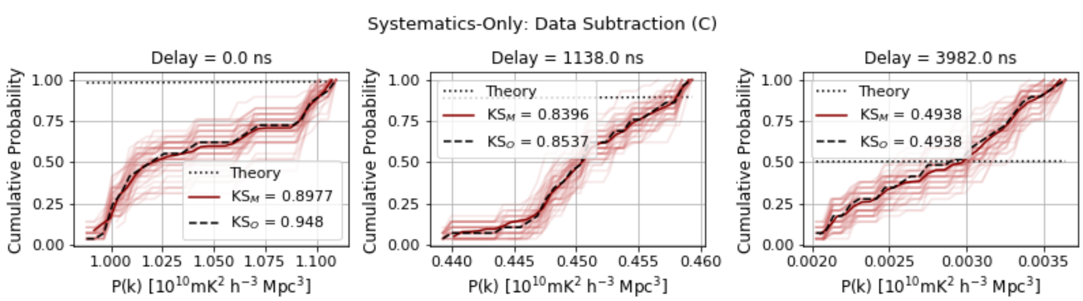


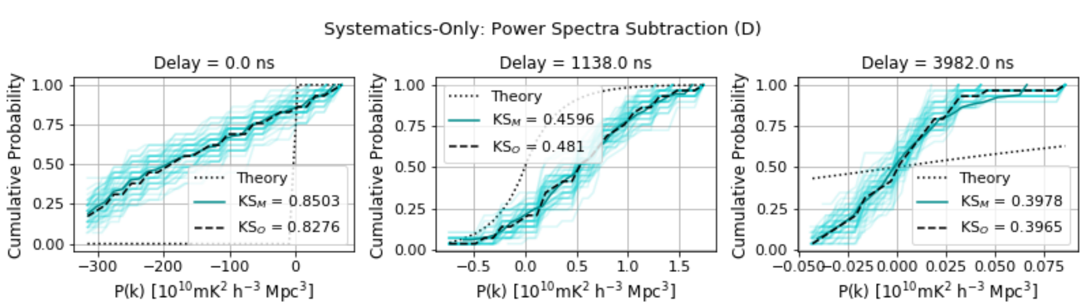
Figure 11: Sample of bootstrapped CDFs (transcluscent), mean CDF taken over bootstraps (solid colour), CDF generated from the original sample of powers (dashed black), theory CDF (dotted black).

spw = 515-695, blp = ((51, 52), (51, 52)):









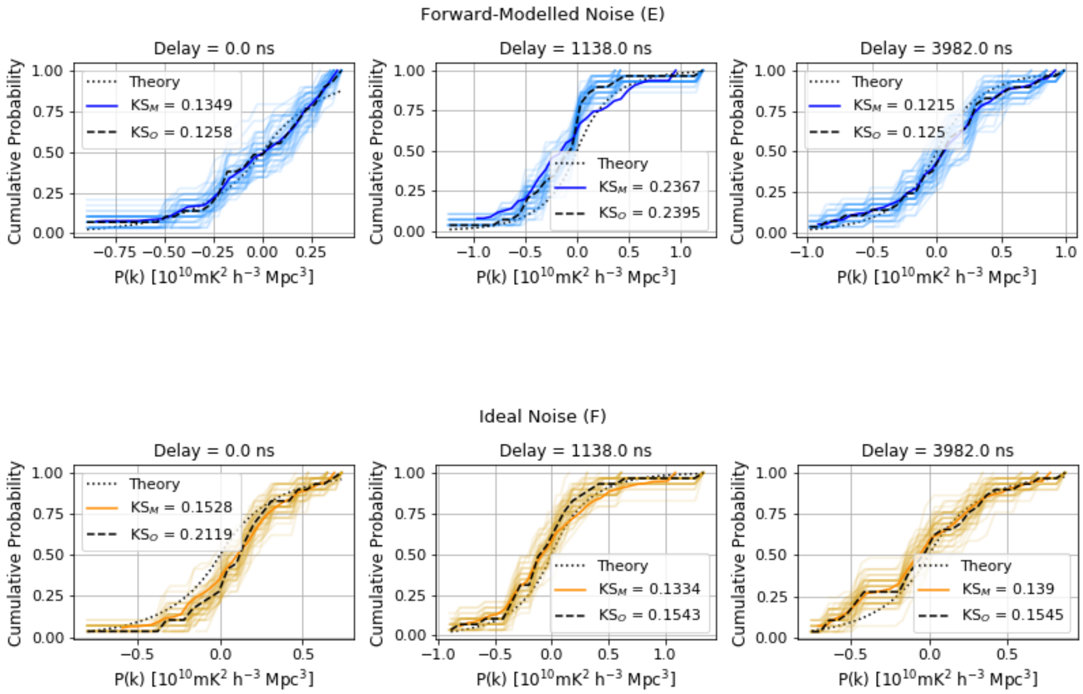


Figure 12: Sample of bootstrapped CDFs (transcluscent), mean CDF taken over bootstraps (solid colour), CDF generated from the original sample of powers (dashed black), theory CDF (dotted black).

4 KS Results

The KS test results are plotted against delay for each dataset. The test determines whether the two samples being compared are drawn from the same distribution. KS results below the critical value confirm the null hypothesis with a confidence level of 95%. Both datasets without features from foregrounds or systematics (E, F) have most KS results below the critical value. The systematics-removed dataset (B) gives KS results below the critical value at delays that are not dominated by foregrounds and the KS results appear to be consistent with those from dataset A at high delays. This suggests that the systematics removal has not affected the delay spectra at noise-dominated delays. The peak of KS results close to zero delay appear to be more narrow for B than for A, suggesting that the systematics removal has had an effect on the foreground peak in the delay spectra. As expected, the systematics-only datasets give KS results that are generally above the critical value, thus rejecting the null hypothesis.

spw = 175-335, blp = ((36, 37), (36, 37)):

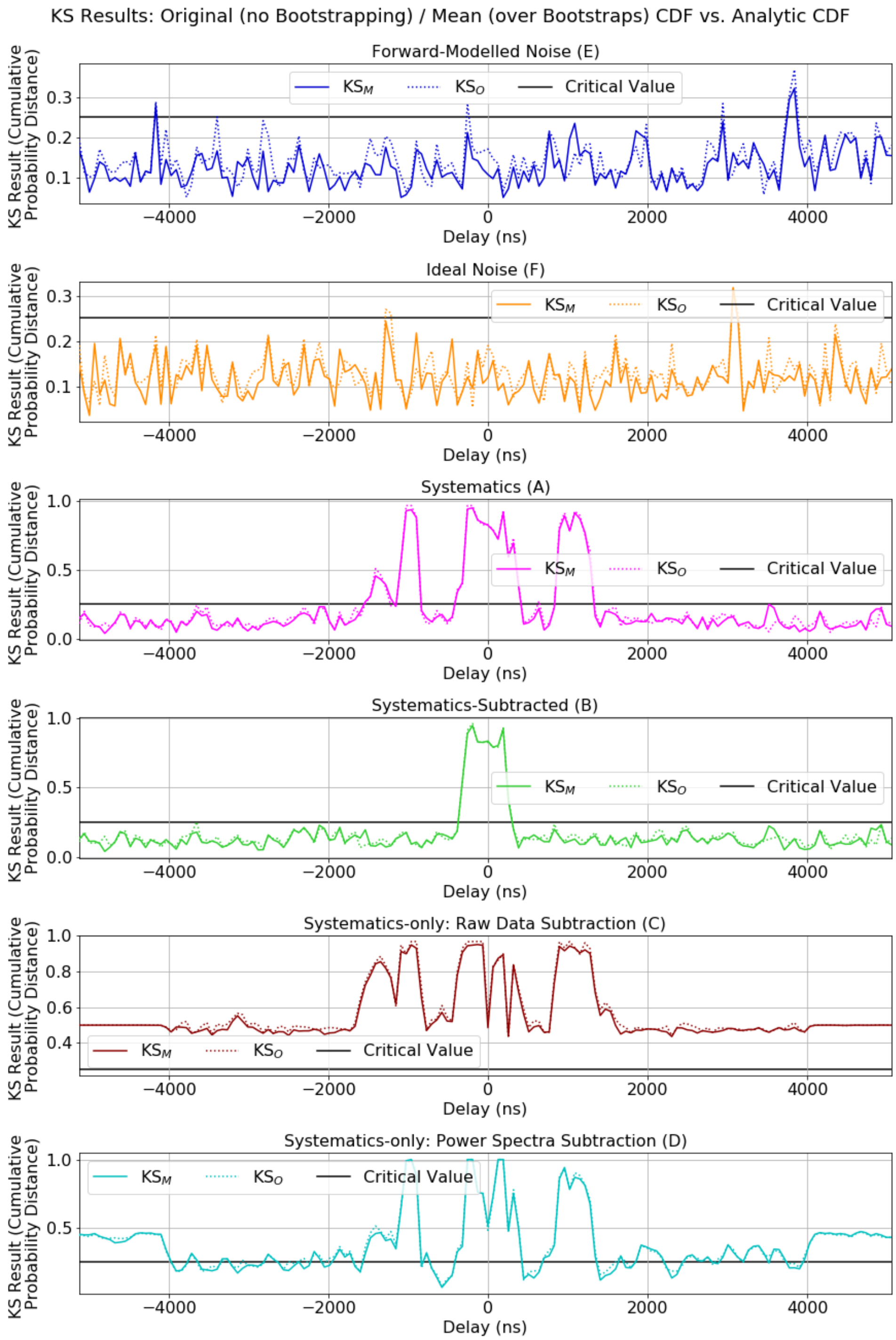


Figure 13: KS Results from comparing the theory CDF to the mean CDF (KS_M) and the original CDF (KS_O) against delay. Critical value is shown to indicate at which delays the samples are drawn from the same distribution with a confidence level of 95%.

spw = 175-335, blp = ((51, 52), (51, 52)):

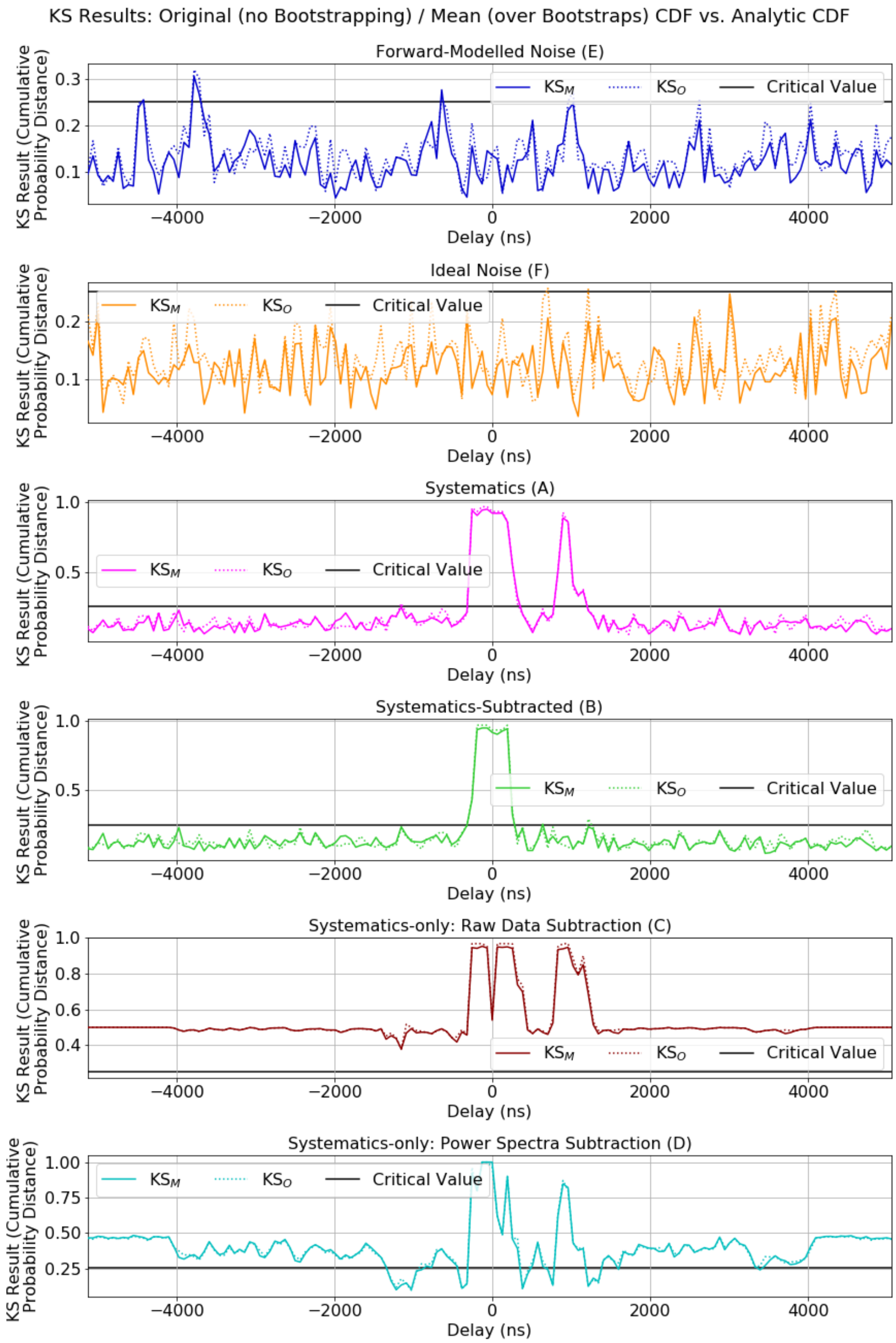


Figure 14: KS Results from comparing the theory CDF to the mean CDF (KS_M) and the original CDF (KS_O) against delay. Critical value is shown to indicate at which delays the samples are drawn from the same distribution with a confidence level of 95%.

spw = 515-695, blp = ((36, 37), (36, 37)):

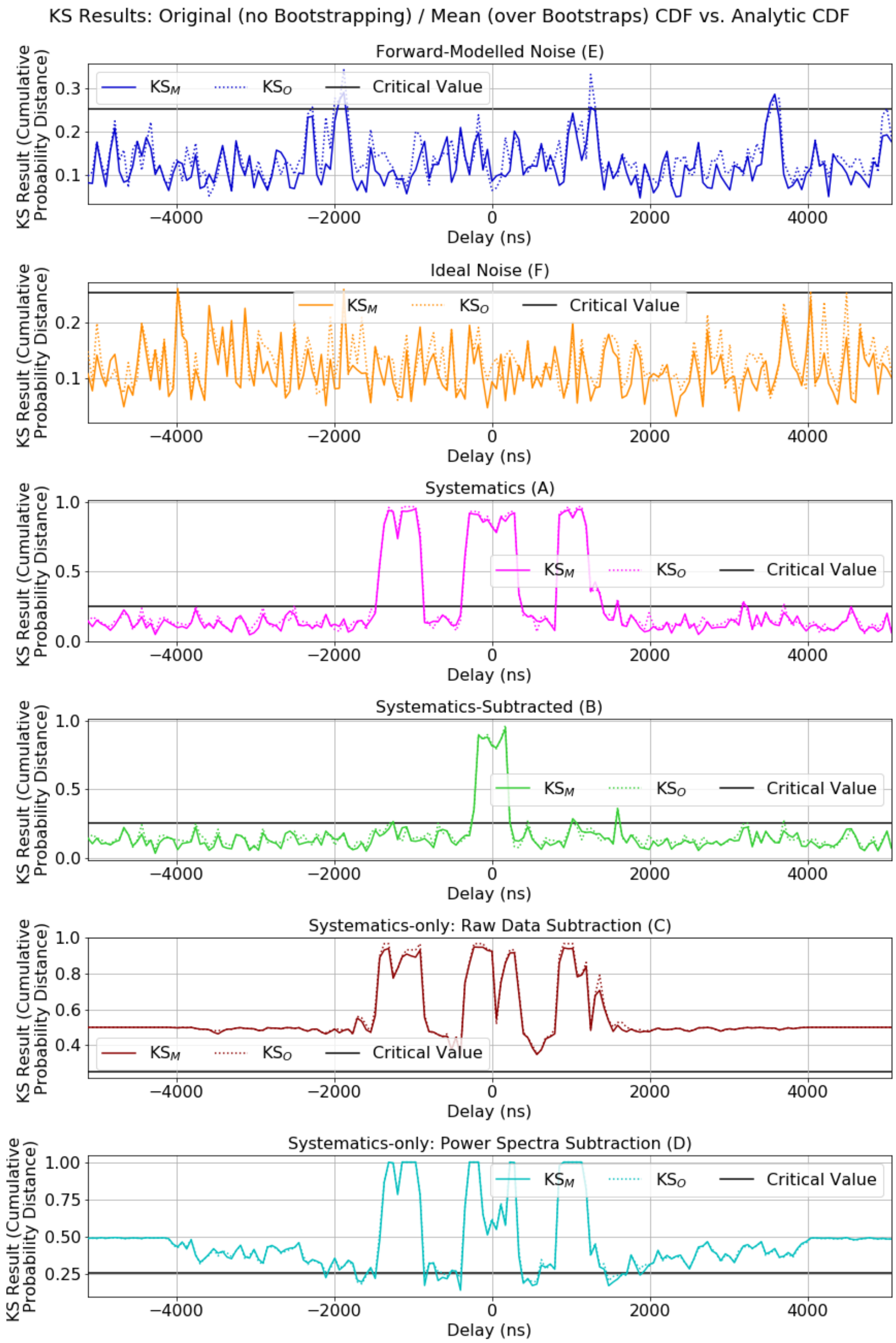


Figure 15: KS Results from comparing the theory CDF to the mean CDF (KS$_M$) and the original CDF (KS$_O$) against delay. Critical value is shown to indicate at which delays the samples are drawn from the same distribution with a confidence level of 95%.

spw = 515-695, blp = ((51, 52), (51, 52)):

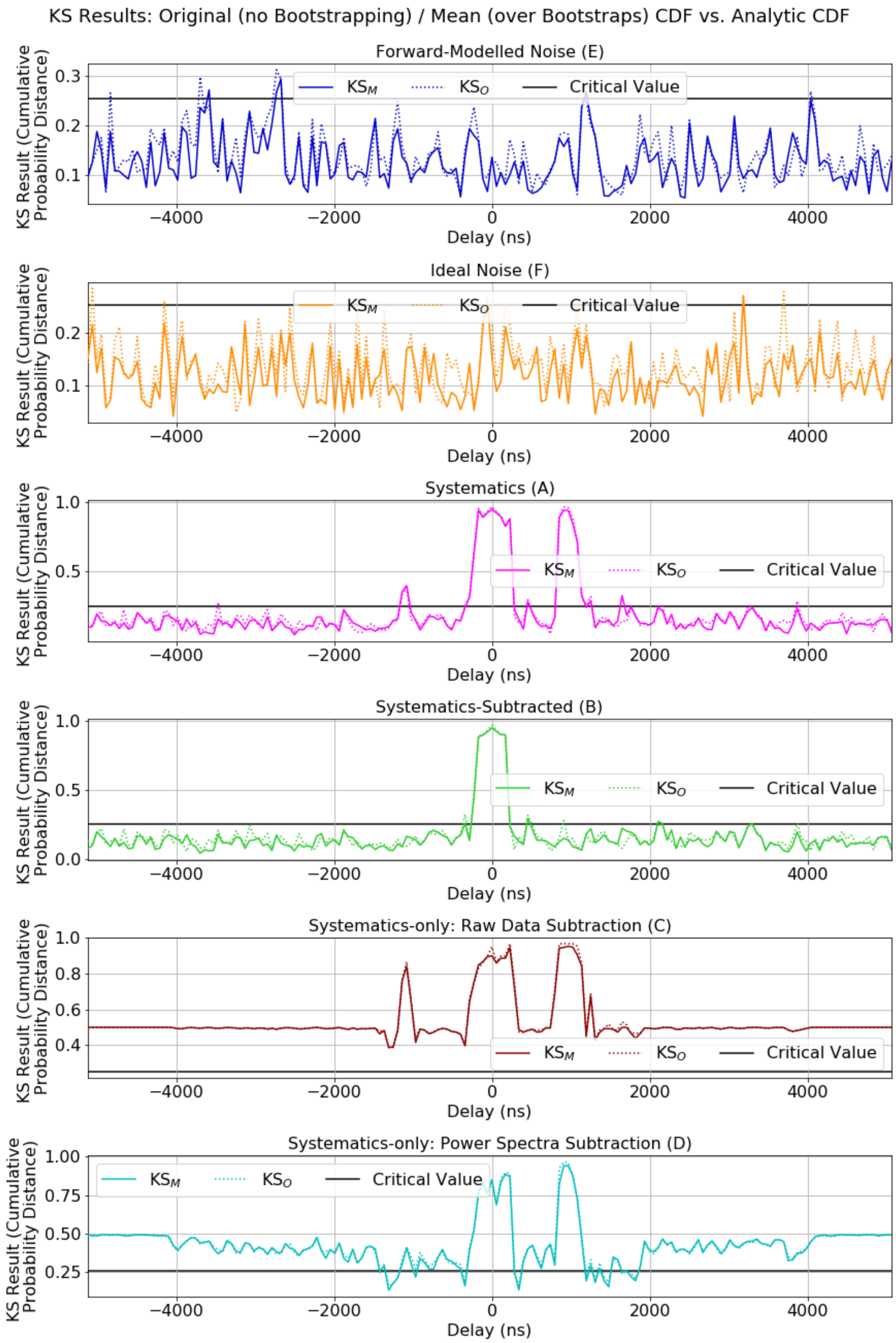


Figure 16: KS Results from comparing the theory CDF to the mean CDF (KS_M) and the original CDF (KS_O) against delay. Critical value is shown to indicate at which delays the samples are drawn from the same distribution with a confidence level of 95%.

5 Summary

The delay spectra generated from the systematics-subtracted dataset appear to be consistent with the theoretical model for noise at delays that are not dominated by foregrounds, as shown from the statistical comparison of CDFs using a KS test.