Atmospheric Noise Removal for FYST: Current Methods and ML Prospects

Ankur Dev
Universität Bonn

A novel sub-mm telescope: 2025

The Fred Young Submillimeter Telescope (FYST): wide-field, 6-m aperture sub-mm telescope.

Site location: at 5600 meters on Cerro Chajnantor in northern Chile.

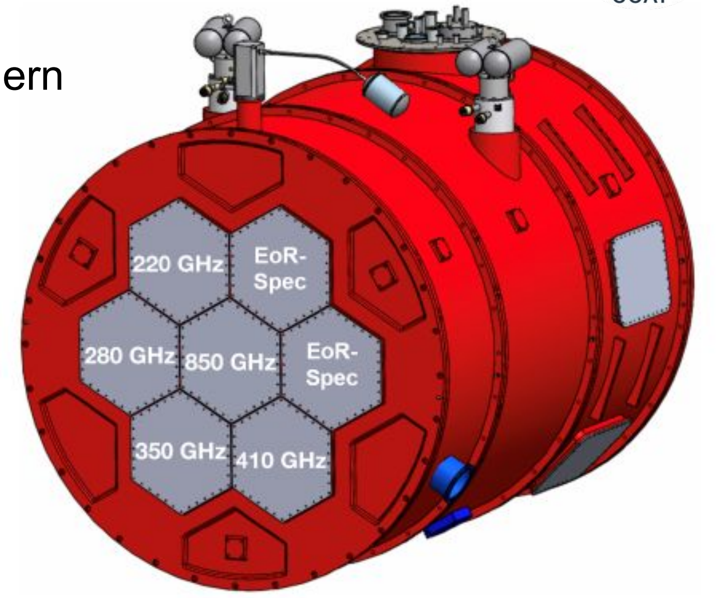
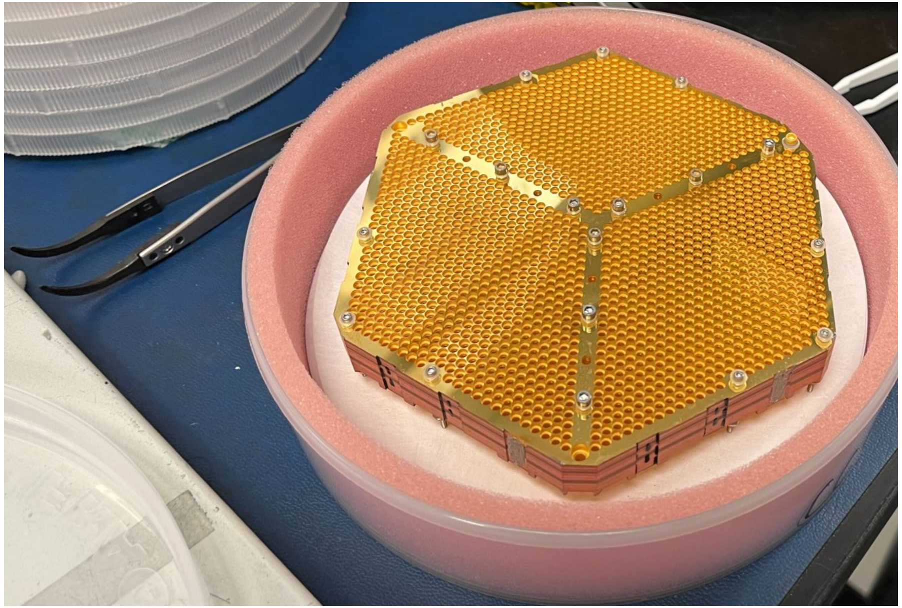
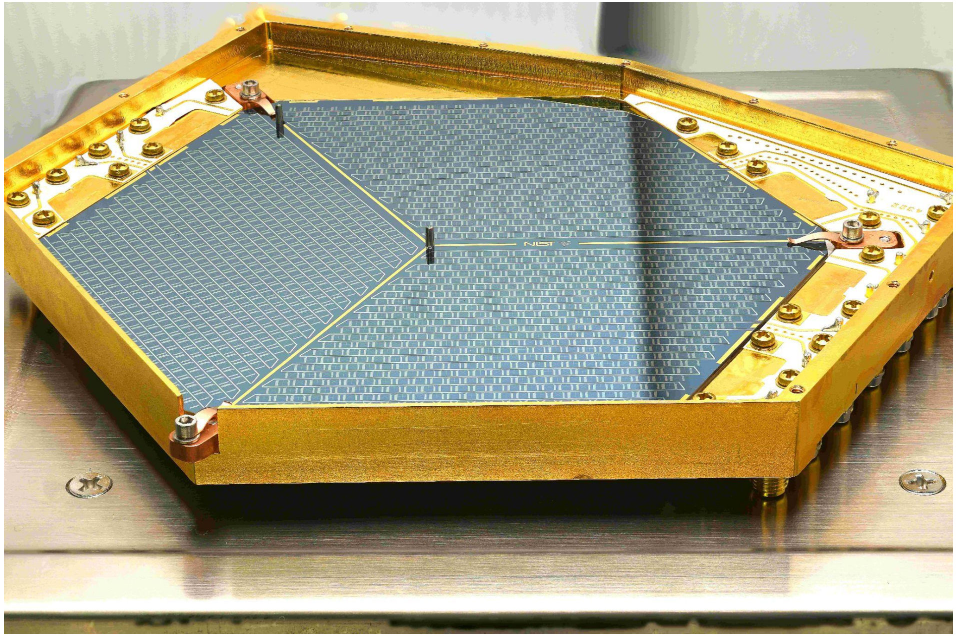


Fig above: FYST model at Cerro Chajnantor (www.ccatobservatory.org)
Fig below: Prime-Cam instrument design with the seven instrument modules (Vavagiakis+2018)

Under development : FYST Detector Arrays

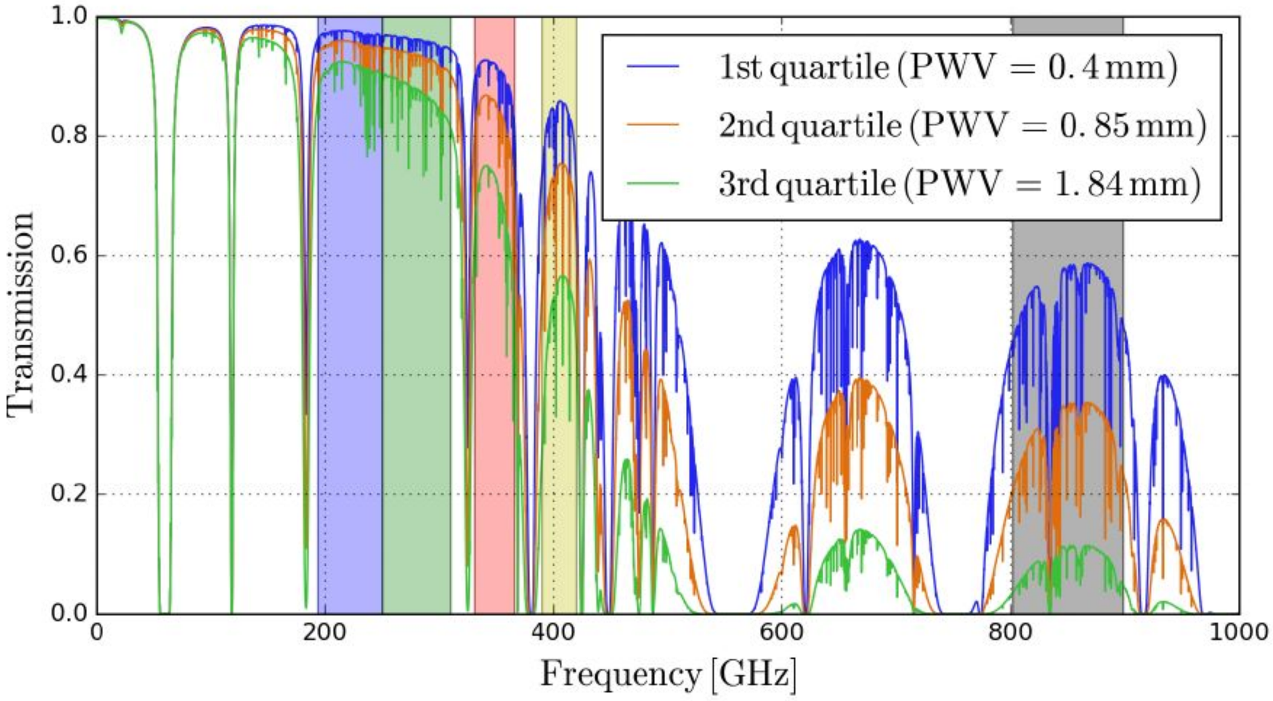




~ 3500 detectors per array times x3 ;
400Hz sampling rate
⇒ Big data volume

Fig Credit: Silicon feedhorn package, and 280GHz aluminum array (Cody Duell, Jordan Wheeler)

Correlated Atmospheric Noise is a Challenge



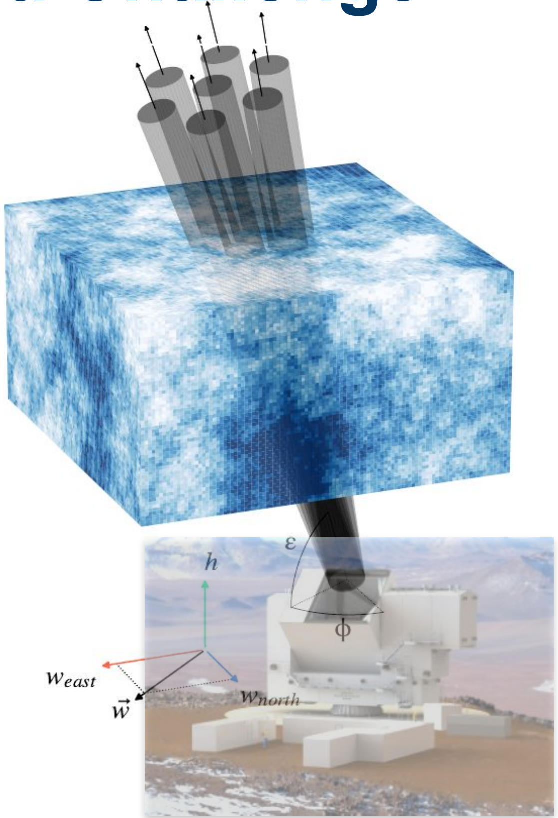


Fig left: Atmospheric transmission spectra for FYST Site (Choi et al 2020)

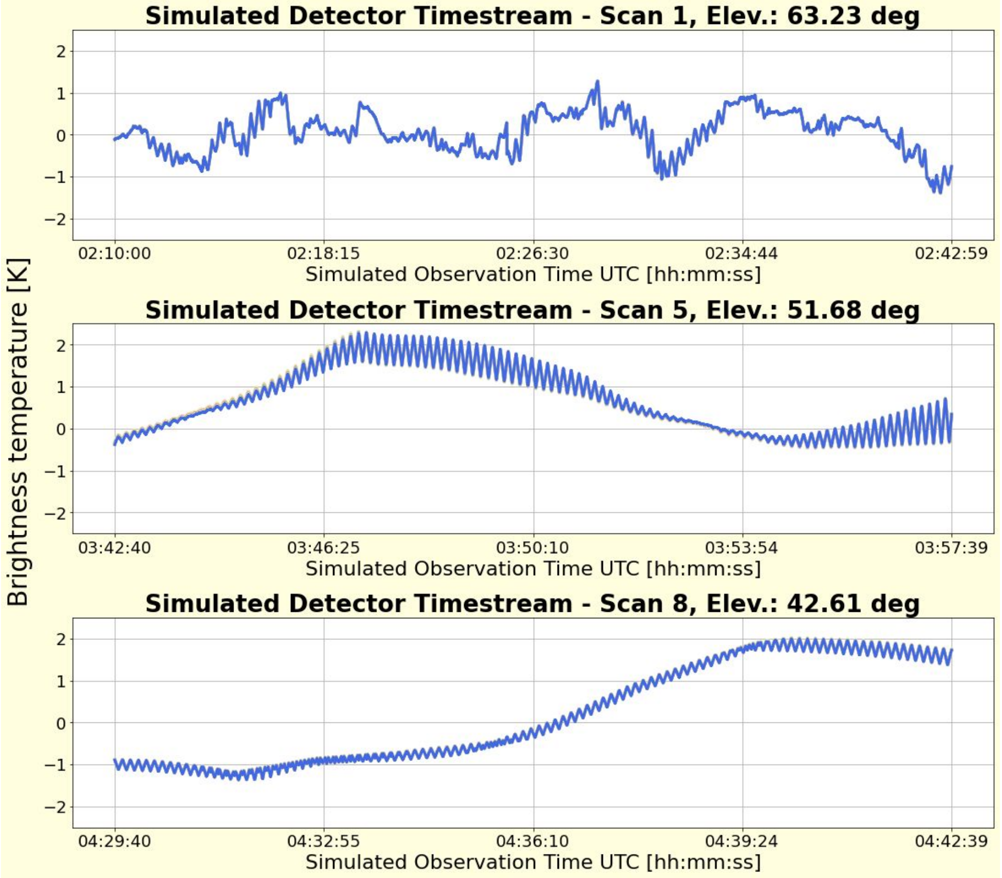
Fig right: Array of detectors observing through inhomogeneous atmosphere (Morris+2022) ; Adapted with FYST model

Goal: Removing low-frequency Correlated Noise

We want to remove the correlated noise component, while retaining the underlying cosmological signal

$$d_v(t) = \langle d_v \rangle P_{cel}\, \Delta s^v_{science} + n_{white}(t) + n_{corr}(t)$$

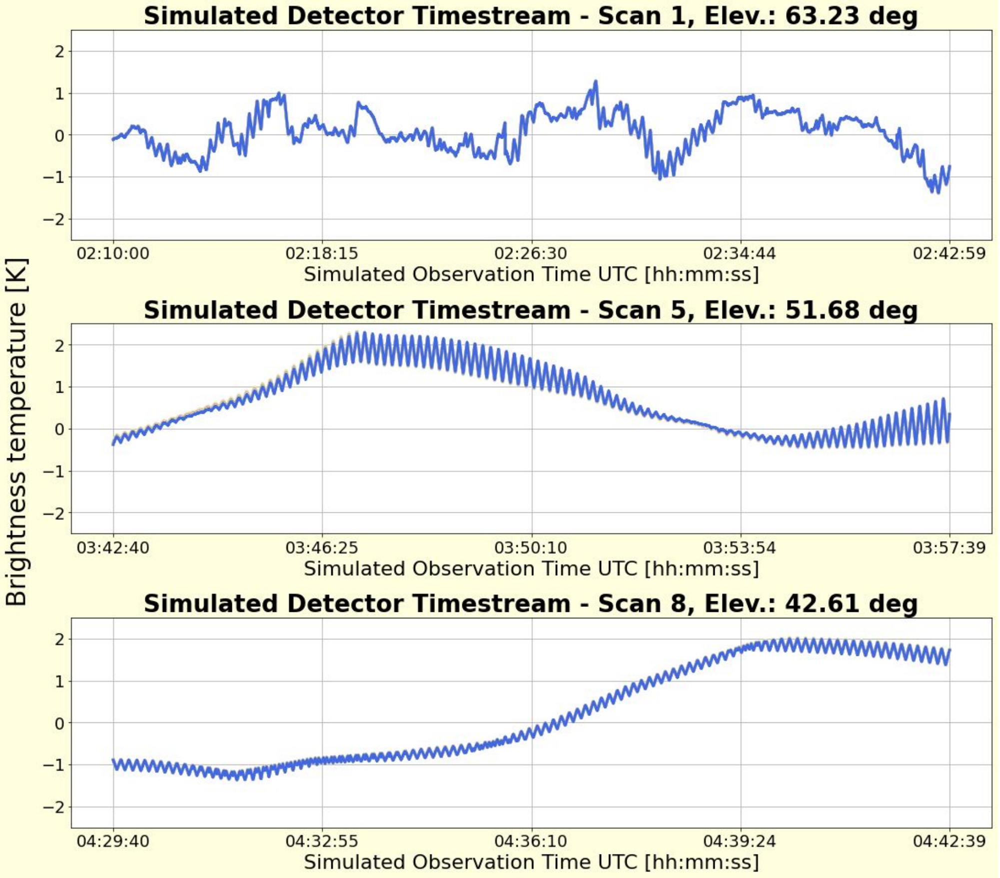
Detector Timestreams / Time-ordered Data (TOD)



Goal: Removing low-frequency Correlated Noise

We want to remove the correlated noise component, while retaining the underlying cosmological signal

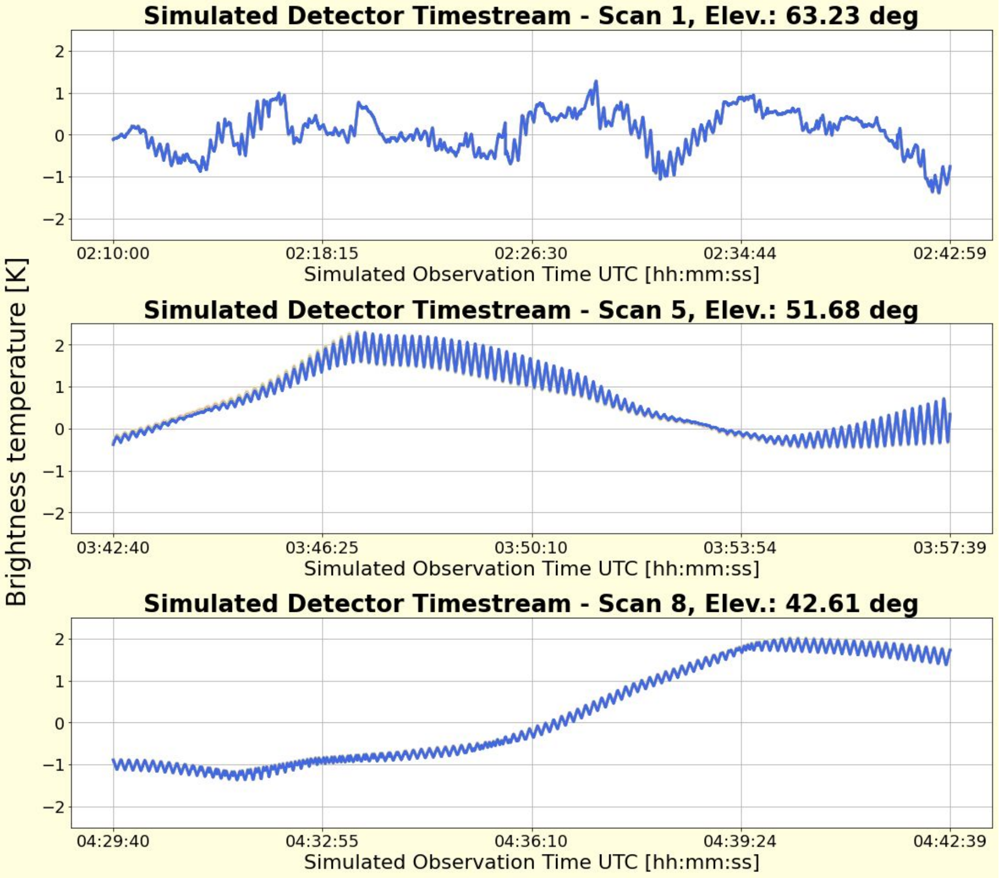
$$d_v(t) = < d_v > P_{cel}\, \Delta s^v_{science} + n_{white}(t) + n_{corr}(t)$$



Goal: Removing low-frequency Correlated Noise

We want to remove the correlated noise component, while retaining the underlying cosmological signal

$$d_v(t) = < d_v > P_{cel}\, \Delta s^{v}_{science} + n_{white}(t) + n_{corr}(t)$$

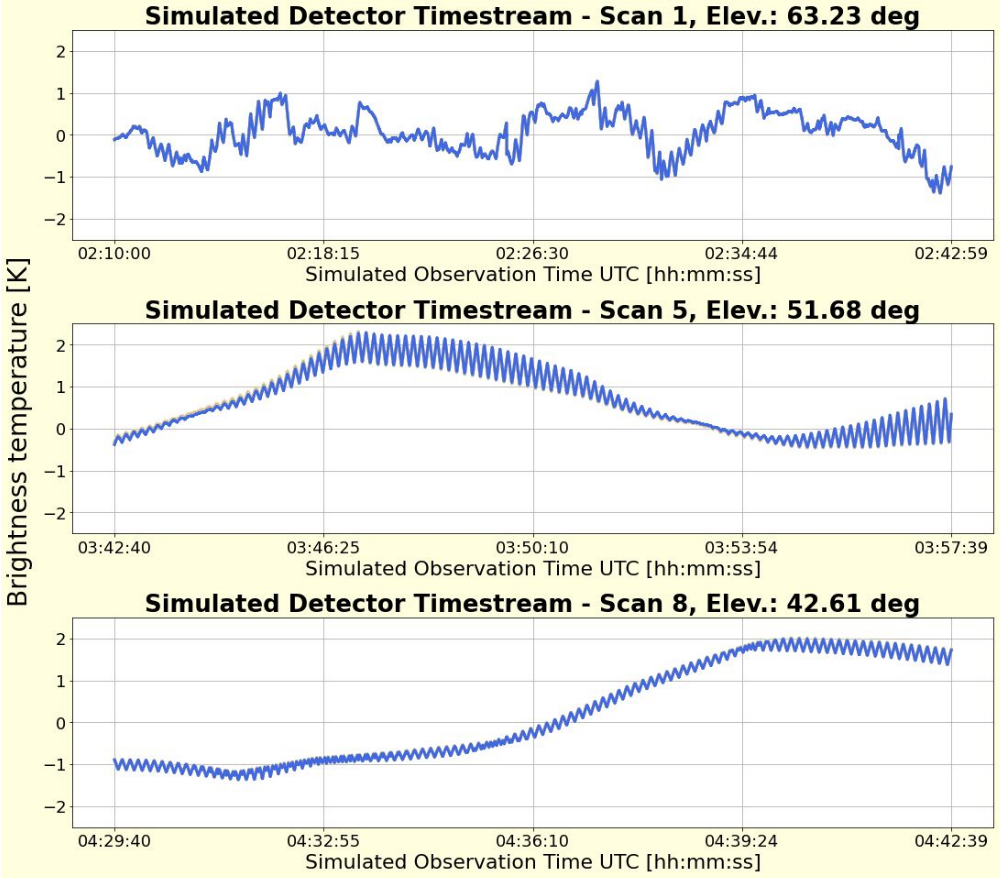


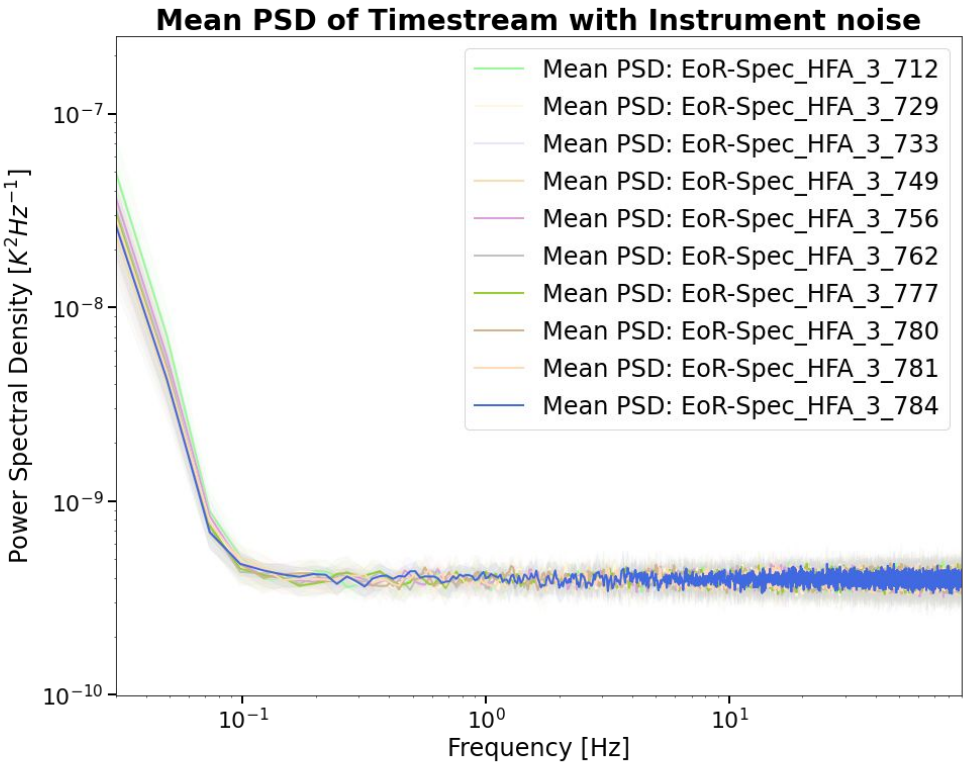
Goal: Removing low-frequency Correlated Noise

We want to remove the correlated noise component, while retaining the underlying cosmological signal

Data reduction includes:

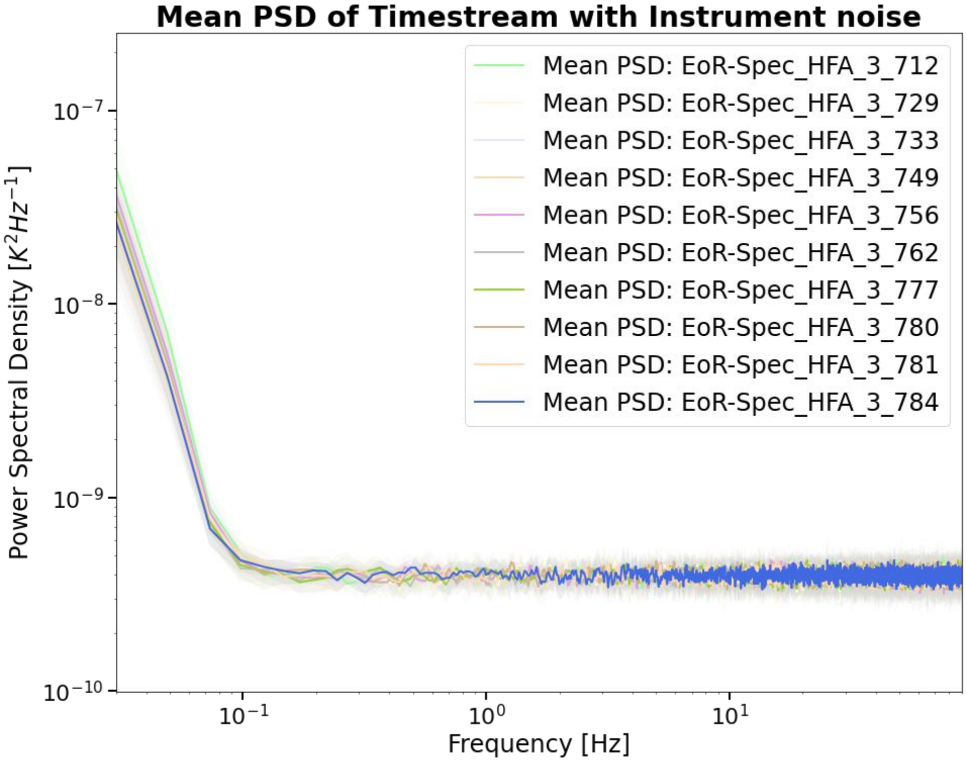
- Data selection
- Detrending
- Filtering : Polynomial and Fourier space
- Principal Component Analysis (PCA)
- Flagging operations
- Map-making from Timestream

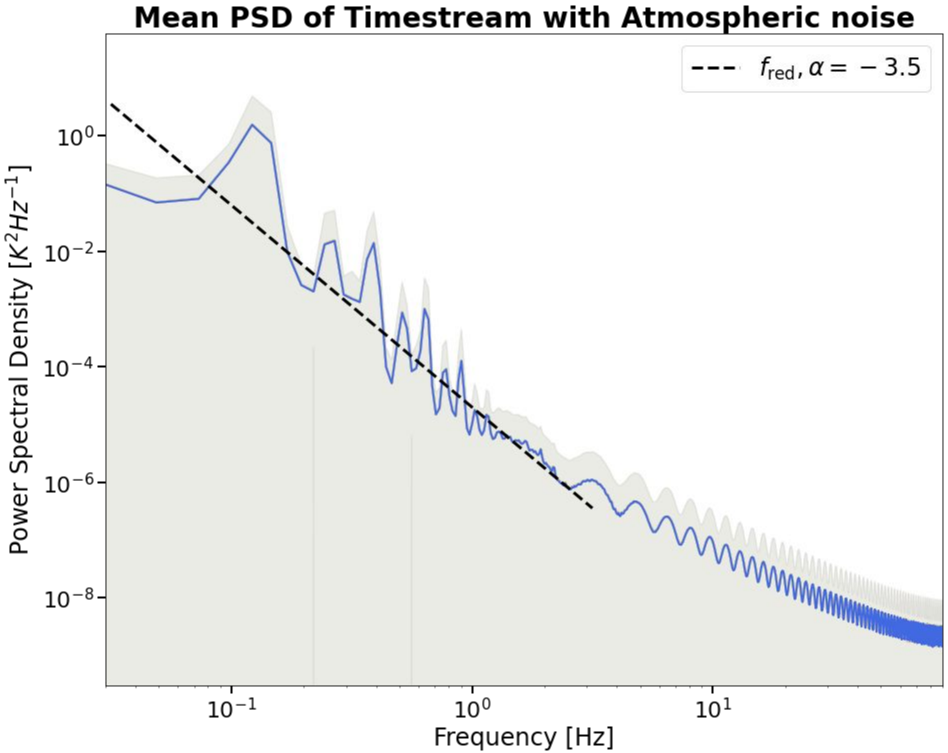



Mean PSD of Timestream with Instrument noise
Mean PSD: EoR-Spec_HFA_3_712
Mean PSD: EoR-Spec_HFA_3_729
Mean PSD: EoR-Spec_HFA_3_733
Mean PSD: EoR-Spec_HFA_3_749
Mean PSD: EoR-Spec_HFA_3_756
Mean PSD: EoR-Spec_HFA_3_762
Mean PSD: EoR-Spec_HFA_3_777
Mean PSD: EoR-Spec_HFA_3_780
Mean PSD: EoR-Spec_HFA_3_781
Mean PSD: EoR-Spec_HFA_3_784
Power Spectral Density [$K^2 Hz^{-1}$]
10^{-7}
10^{-8}
10^{-9}
10^{-10}
10^{-1}
10^{0}
10^{1}
Frequency [Hz]

Power Spectral Density: Atmosphere introduces 1/f Noise

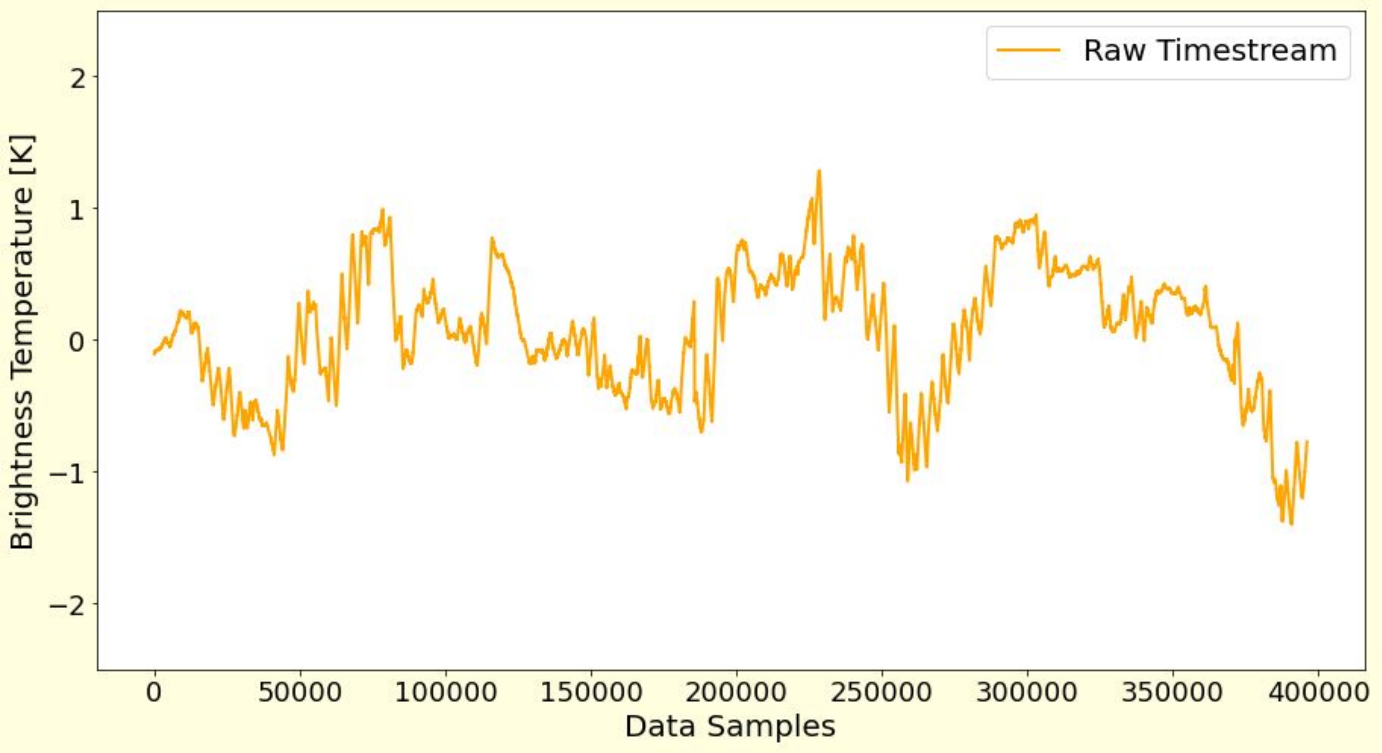
$$N(f) = \sigma_0^2 \left(1 + \left(\frac{f}{f_{\text{knee}}}\right)^{\alpha}\right)$$

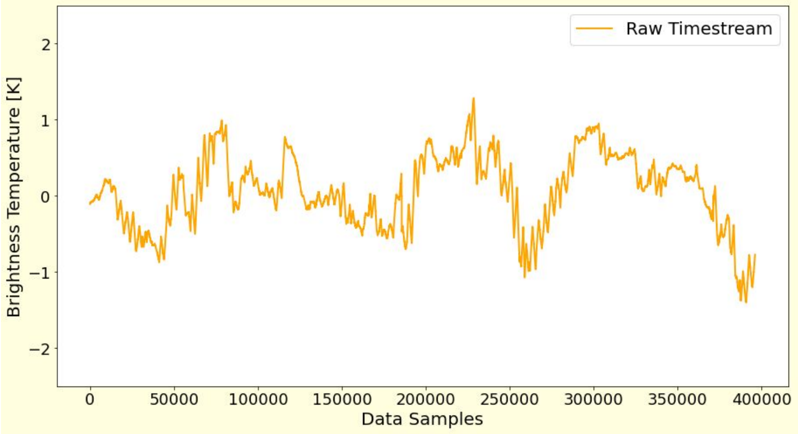




Data-processing for removing Atmospheric Noise

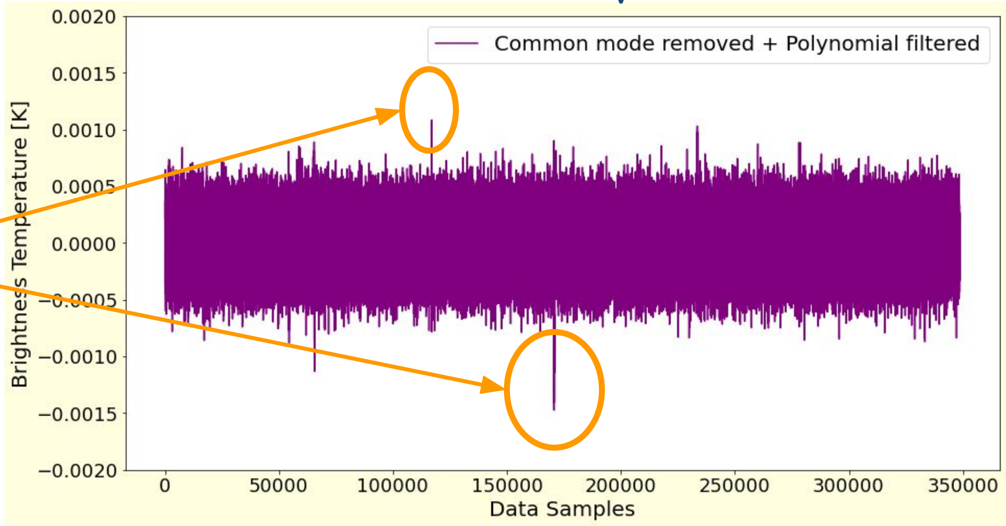
Data:
Detector timestream to be processed for removing trends





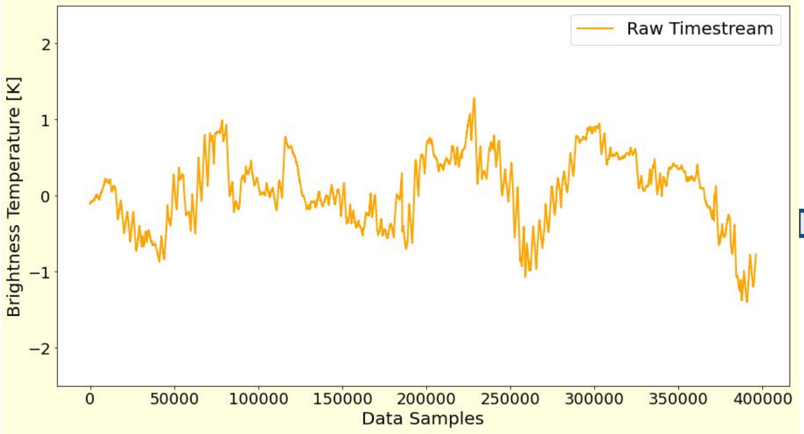
Step 1: Common mode removed from all detectors

Step 2: 0th Order Polynomial subtracted



Few data-samples are poorly behaved

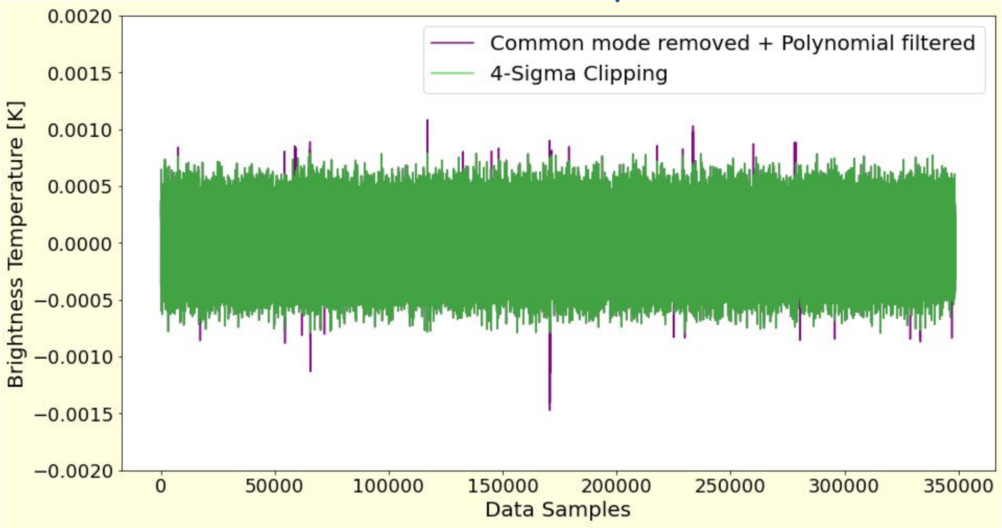
Time-domain analysis



Step 1: Common mode removed from all detectors

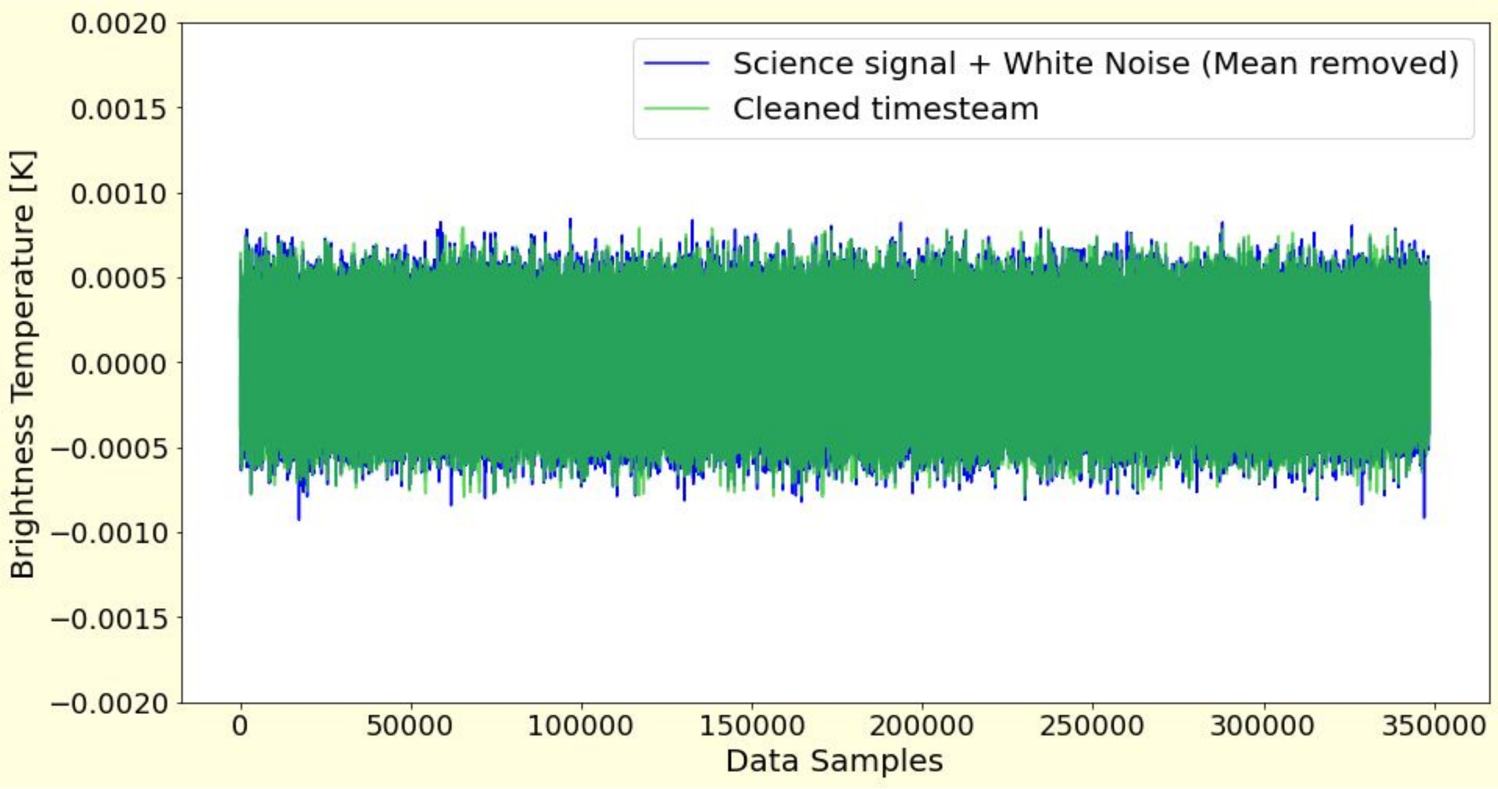
Step 2: 0^{th} Order Polynomial subtracted

Step 3: Excluding extreme statistical outliers (4-sigma clip)

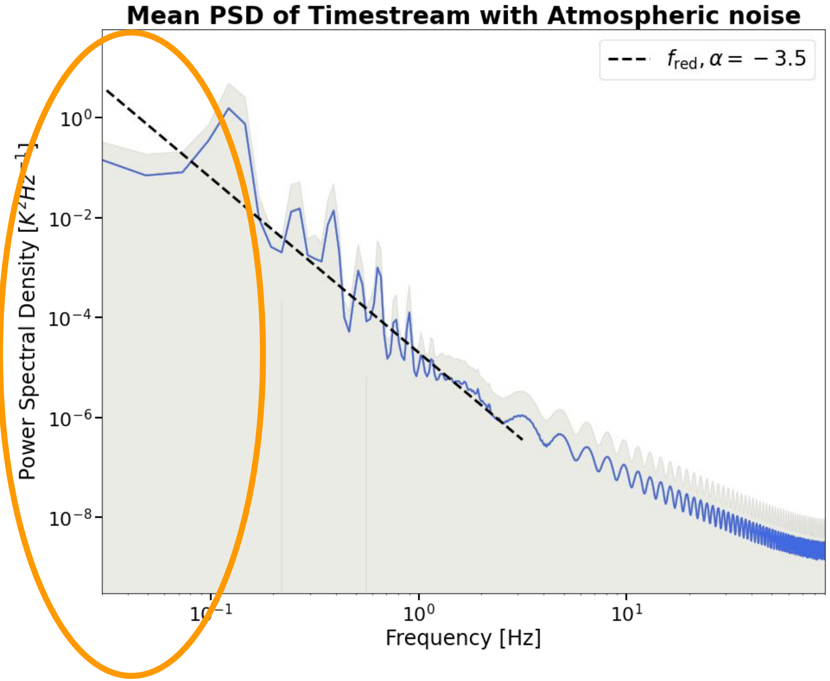


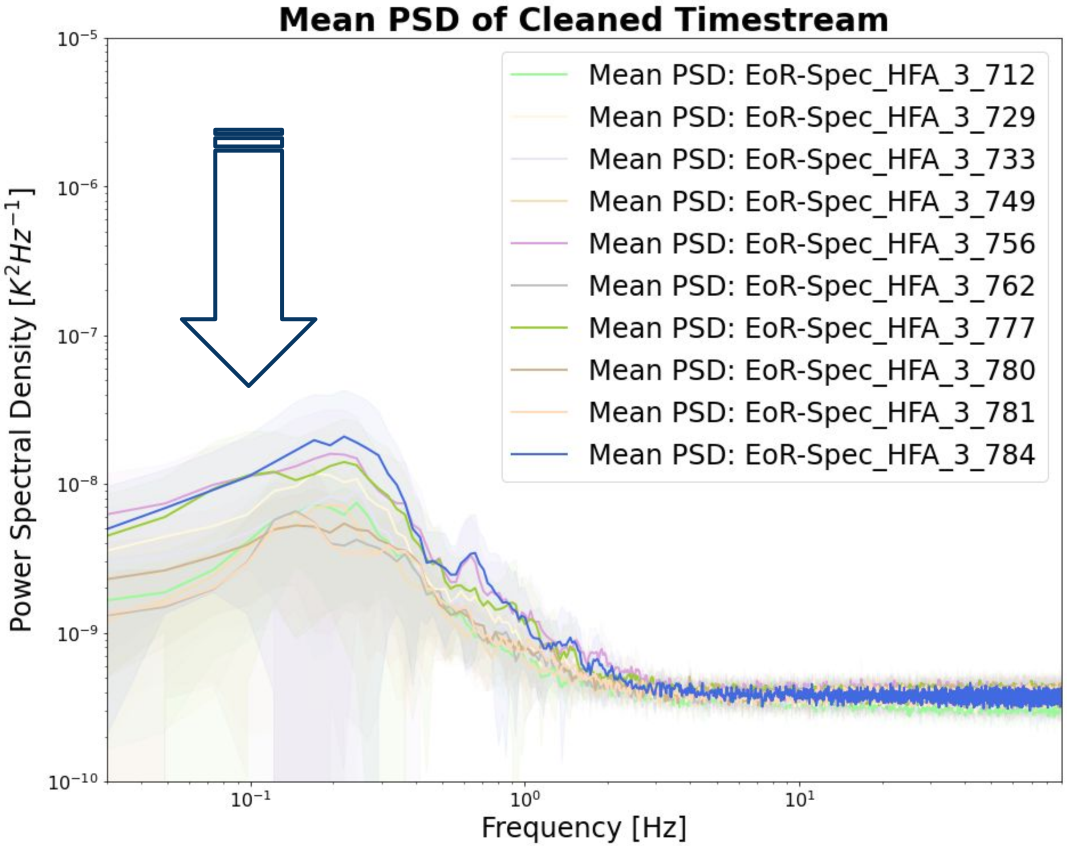
Time-domain analysis

Timestream Cleaned to White Noise Level

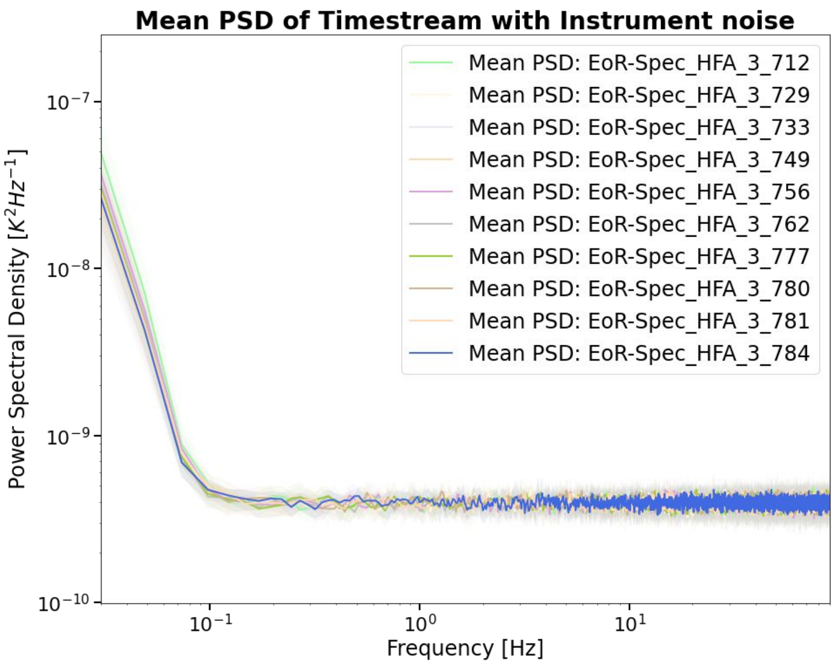


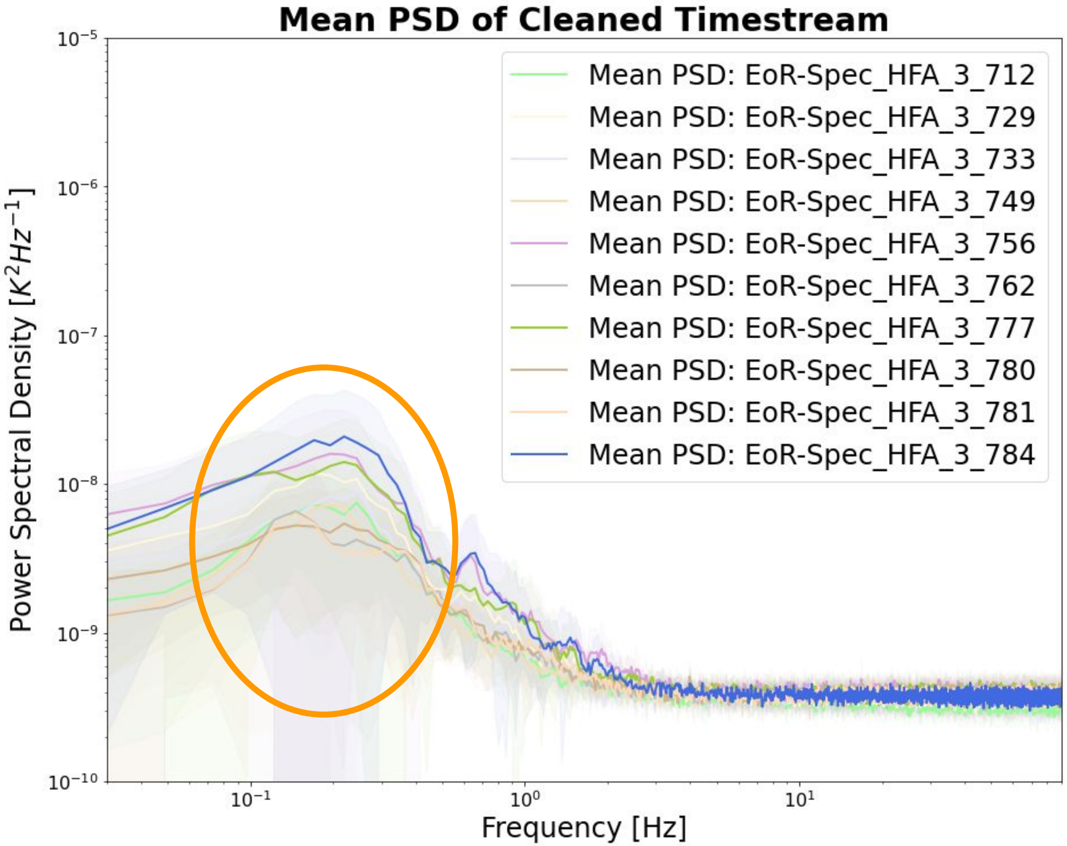
Power Spectral Density Comparison of Cleaned Timestream





Power Spectral Density Comparison of Cleaned Timestream





Can ML methods contribute to current data reduction techniques?

- Convolutional Neural Networks (CNNs)
  - Layered, hierarchical architectures for spatial/temporal pattern recognition.
  - Adaptability in identifying and reducing diverse noise components through encoding-decoding structures.
- Gaussian Process Regression (GPR)
  - non-parametric, Bayesian approach to model and predict time-series data.
  - modeling complex noise structures
- Monitoring Data Quality
  - Outlier Detection: Leveraging ML for identification of anomalies in detector timestreams.
  - Data Selection and Cuts

Can ML methods contribute to current data reduction techniques?