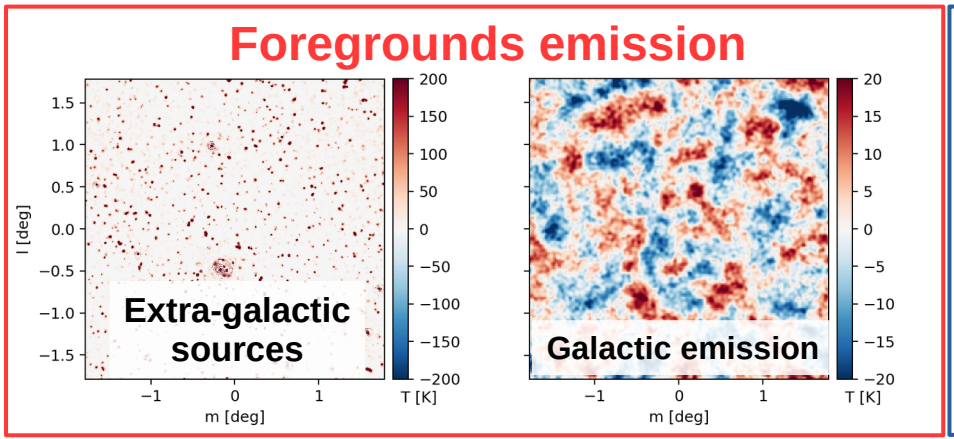
ML-enhanced foregrounds mitigation method for 21-cm experiments

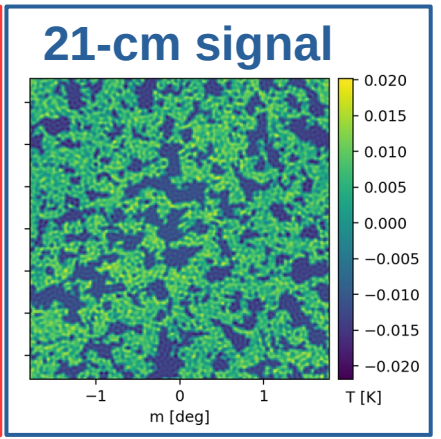
Florent Mertens (LERMA, Paris Observatory)

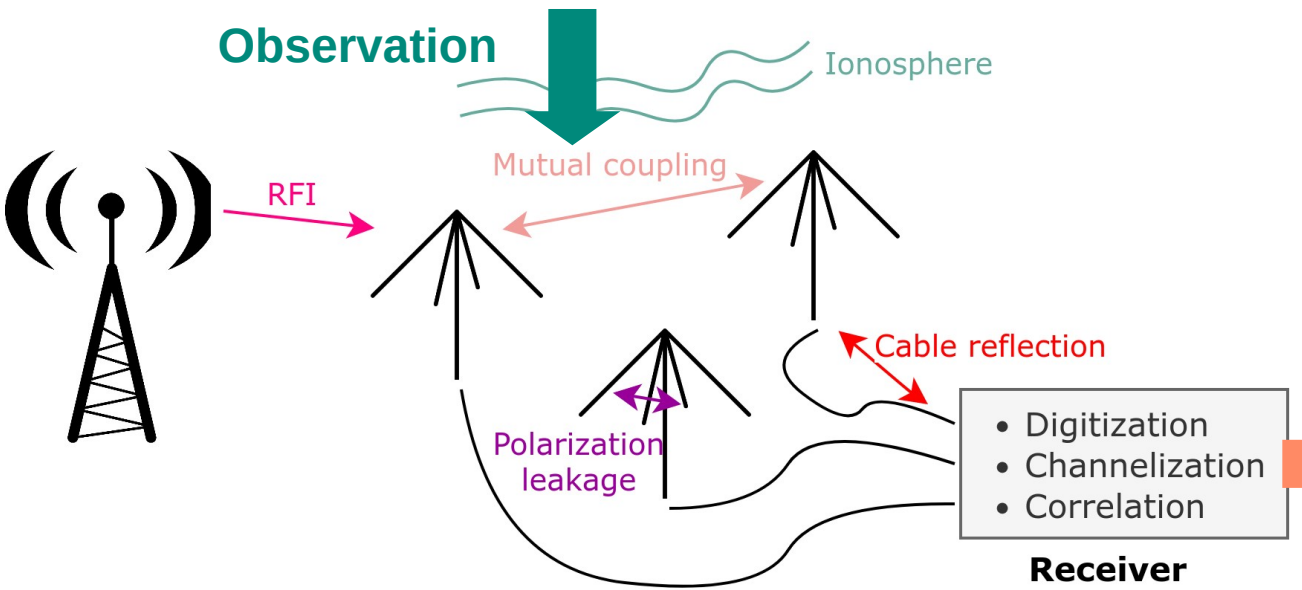
Cosmology in the Alps, Les Diablerets, 21/03/2024

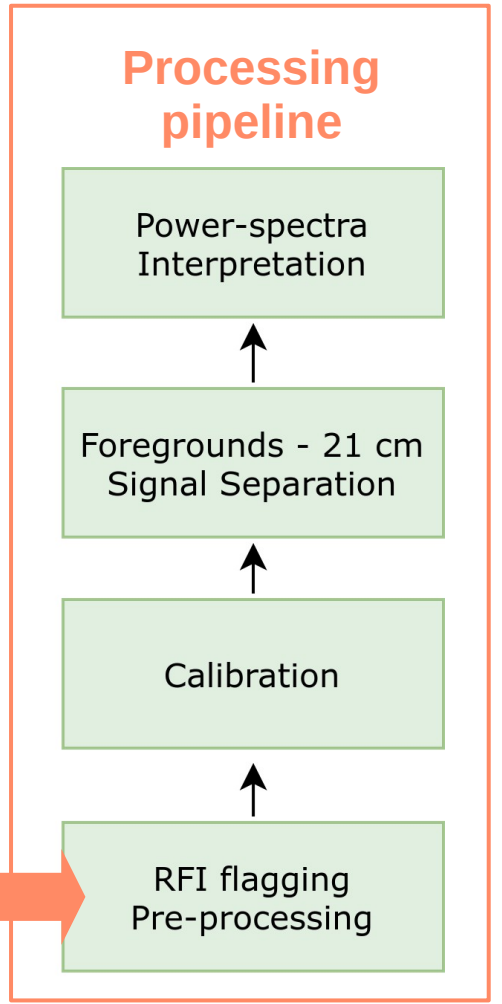
Image credit: Michale Goh/ICRAR-Curtin

A challenging experiment



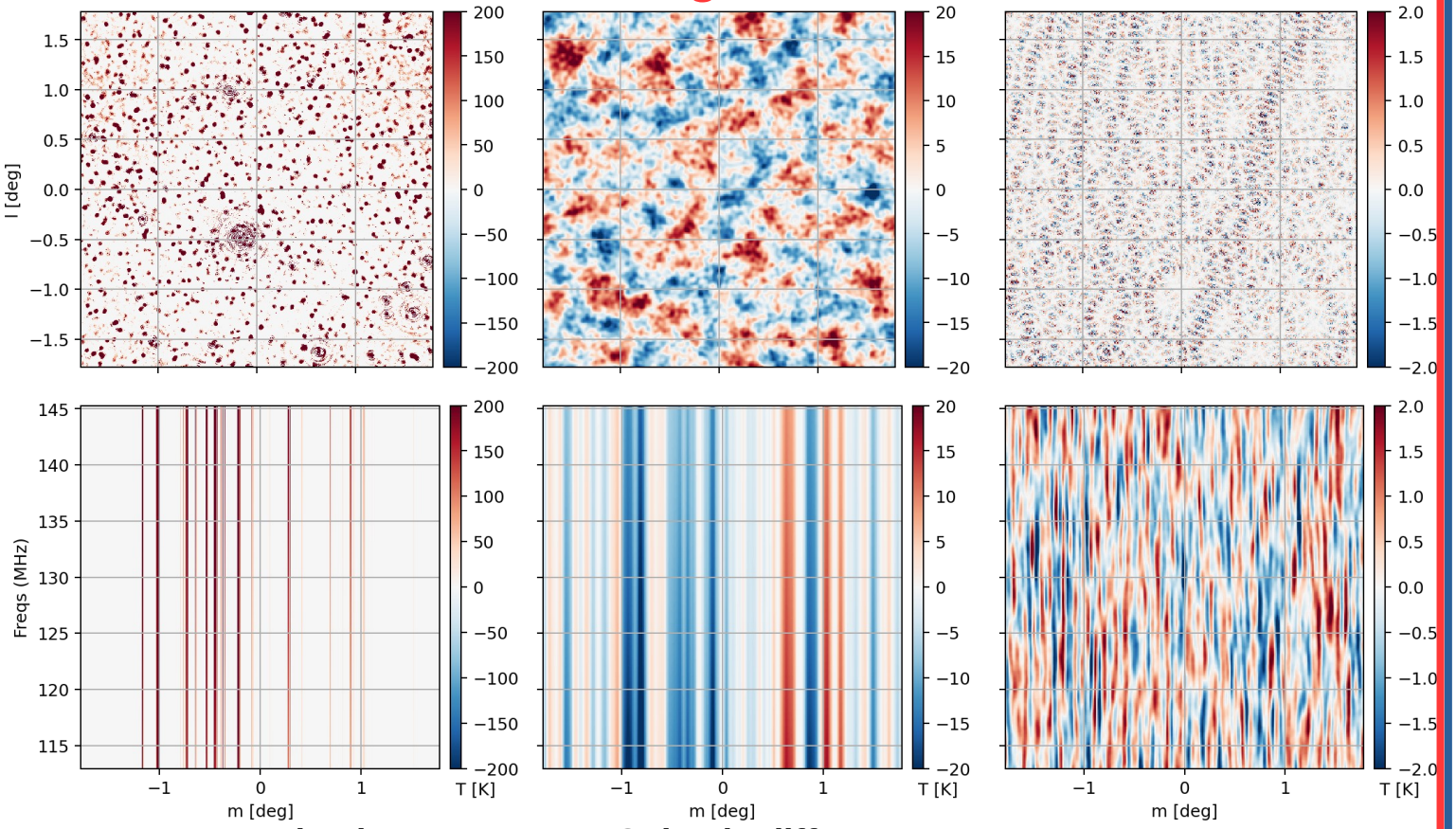






The challenge of the foregrounds

Foregrounds

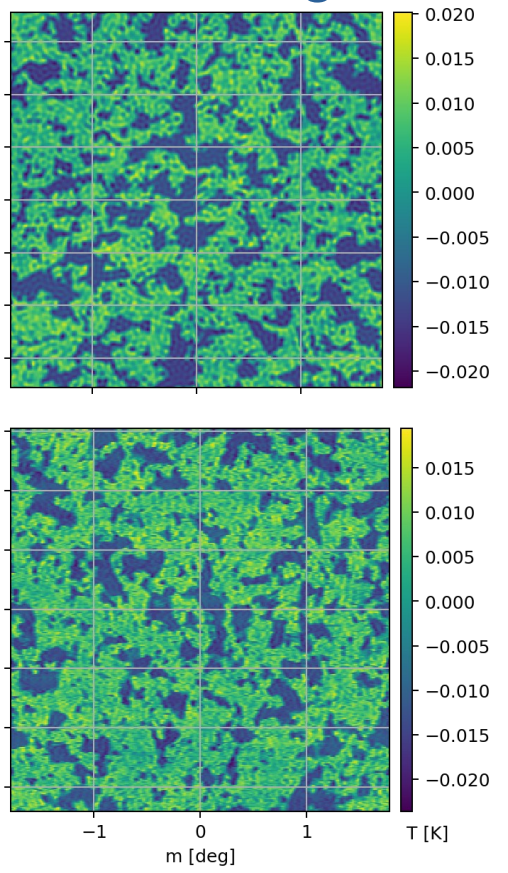


Extragalactic point sources

Galactic diffuse emission

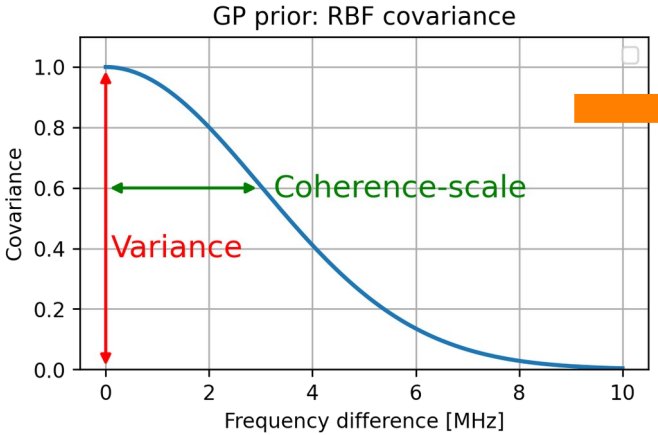
Instrumental effects

21-cm signal

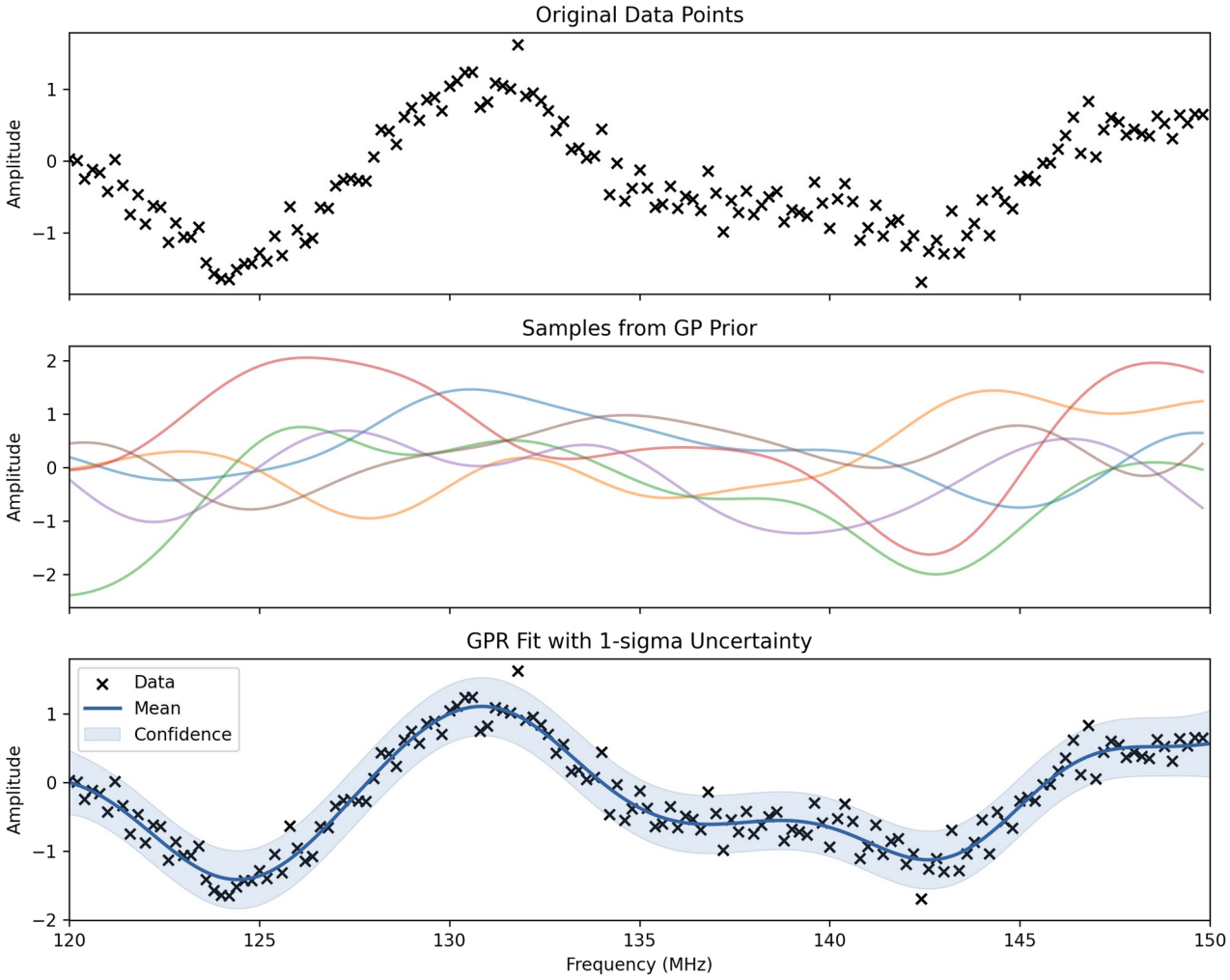


Gaussian Process Regression

- **GPR:** Fits data without assuming a specific functional form.
- **Prior Information:** Encoded through a parametrized covariance function.

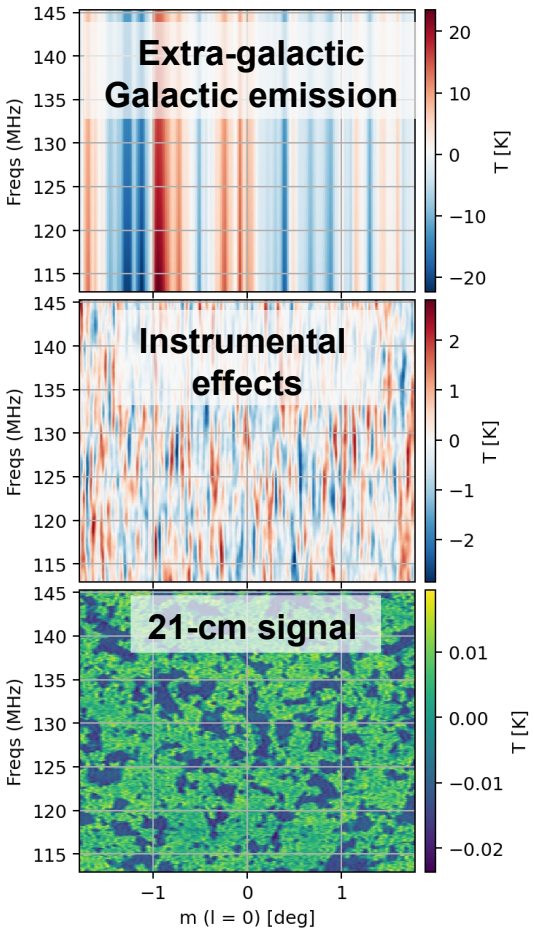


- **Parameters Optimization:** Covariance parameters are determined by maximizing the marginal likelihood.
- **Data fitting:** Conditioning the prior model to the data, we obtain fit + uncertainty.

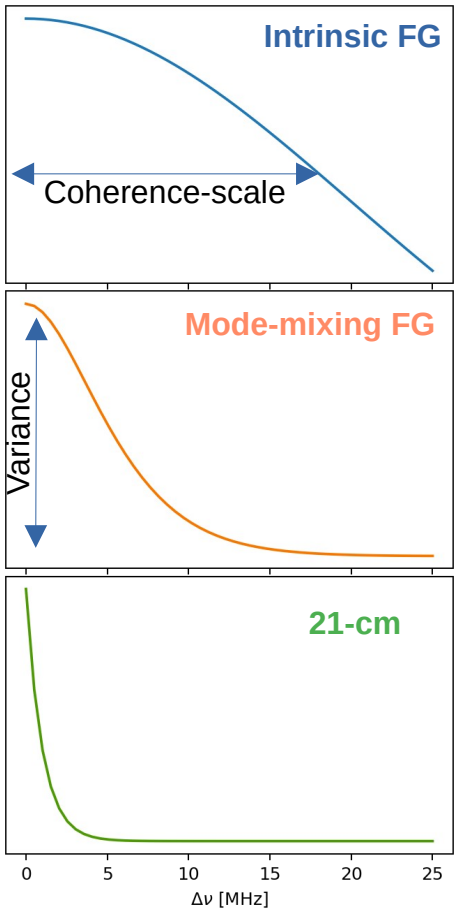


GPR for 21-cm experiments

Frequency slice



Freq-freq covariance



No functional forms but very different spectral characteristic

→ Statistical model prior made of Gaussian Process (GP).

→ Learnt kernel is used for the 21-cm prior covariance.

$$K = K_{\rm fg} + K_{21} + \mathrm{K}_{\rm noise} + \mathrm{K}_{\rm other}$$

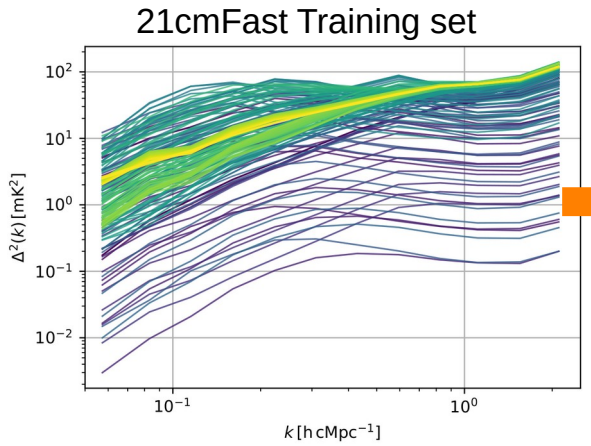
Hyper-parameters of the covariance prior to be optimized with the data

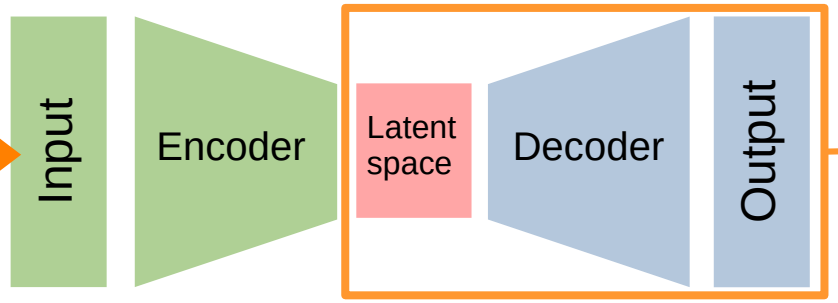
$$E(\mathbf{f}_{\rm fg}) = K_{\rm fg}\left[K_{\rm fg} + K_{21} + \sigma_n^2 I\right]^{-1}\mathbf{d}$$

$$\mathrm{cov}(\mathbf{f}_{\rm fg}) = K_{\rm fg} - K_{\rm fg}\left[K_{\rm fg} + K_{21} + \sigma_n^2 I\right]^{-1}K_{\rm fg}$$

Mertens et al. 2018
Mertens, Bobin, Carucci 2024

Learned covariance function



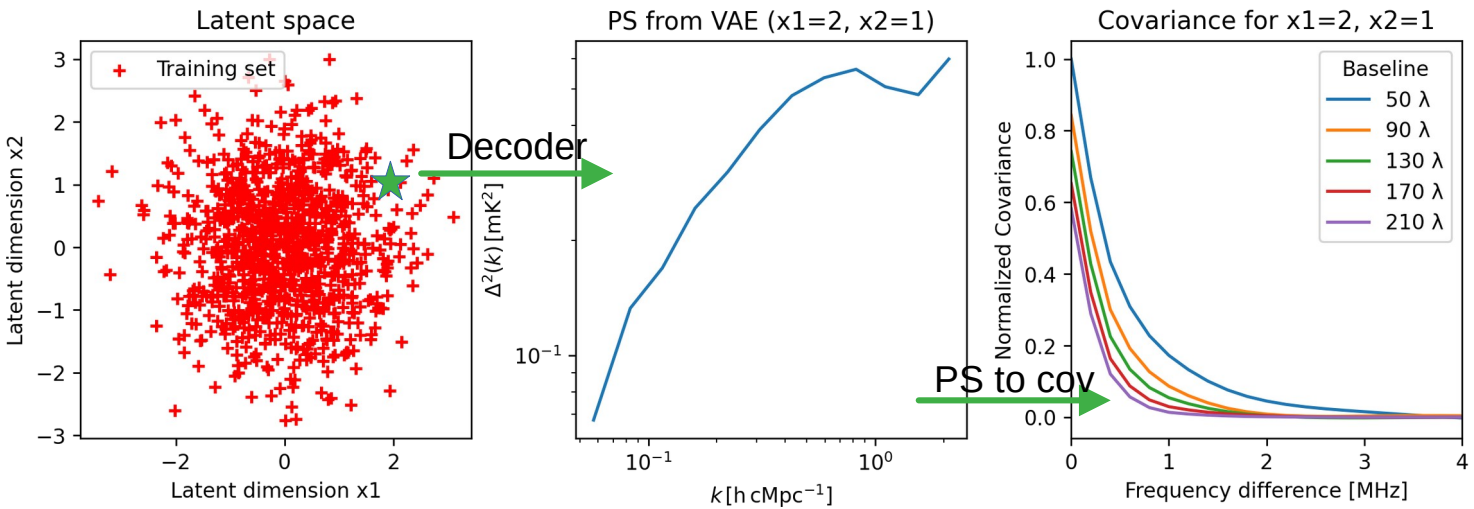


Variational Auto-Encoder (VAE)

21-cm covariance
Latent-space parameters

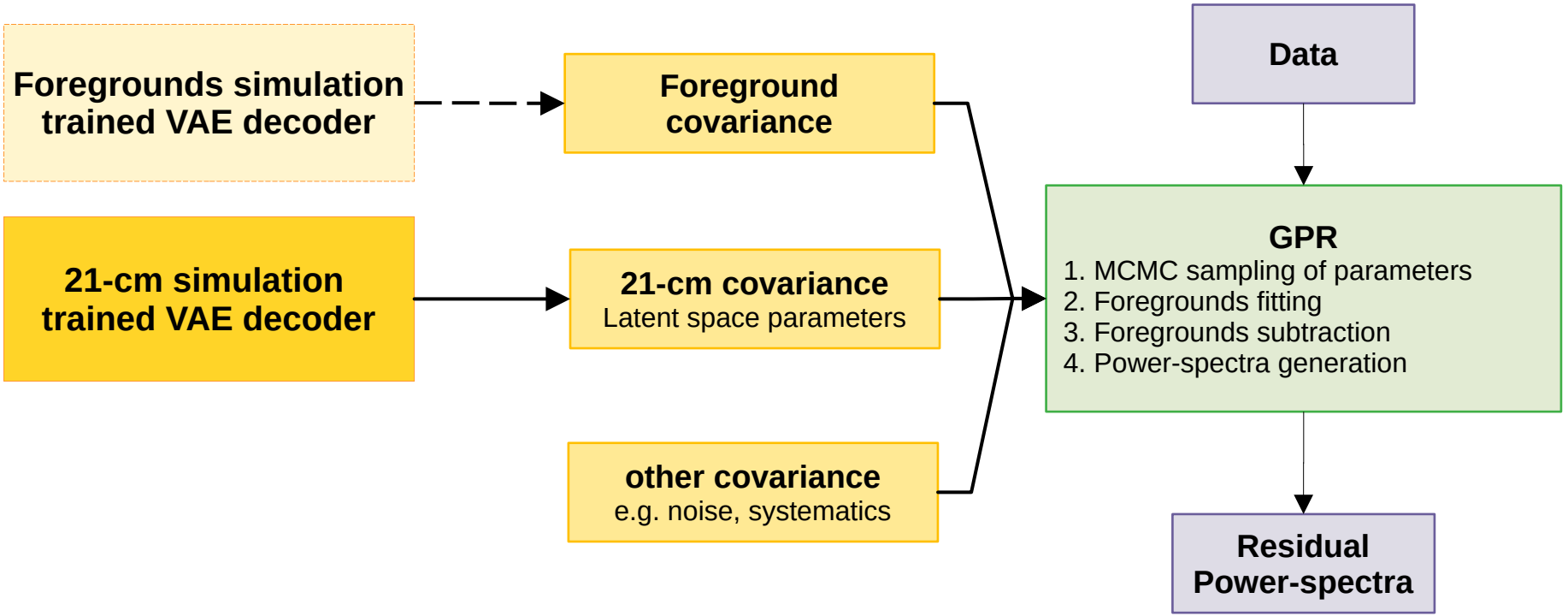
VAE: Trained to minimize:
- ➔ Reconstruction error.
- ➔ KL divergence to standard Gaussian in latent space.

- ✔ Compressed information (lower dimension latent space).
- ✔ Generative.



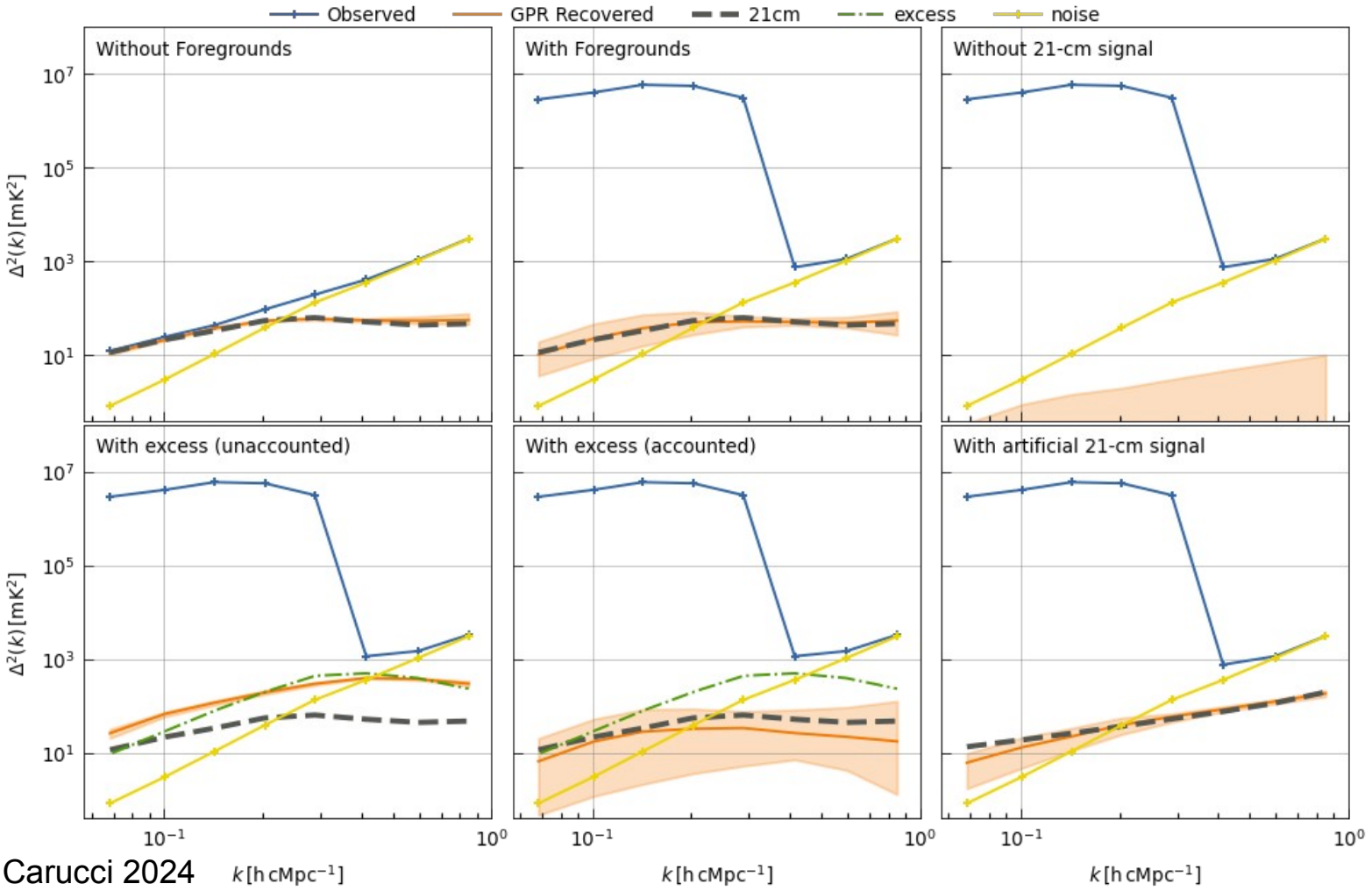
Mertens, Bobin, Carucci 2024

The ML-GPR foreground removal method



Mertens et al. 2018
Mertens, Bobin, Carucci 2024

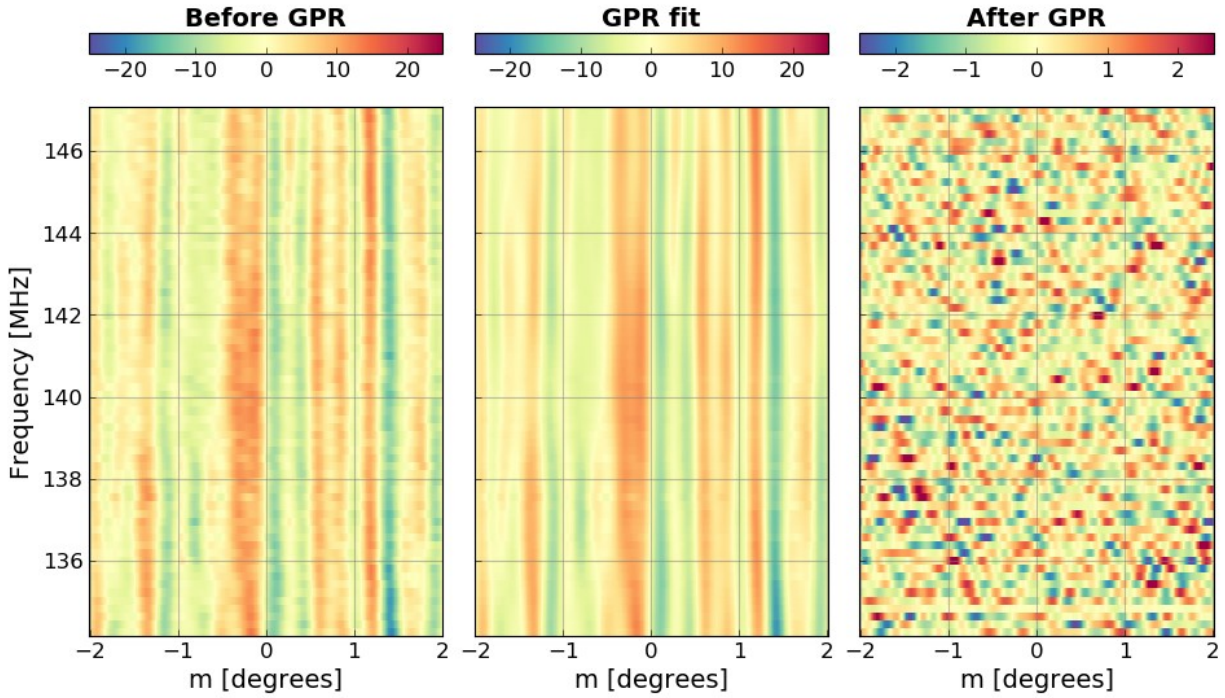
ML-GPR SKA simulation (100 hrs)

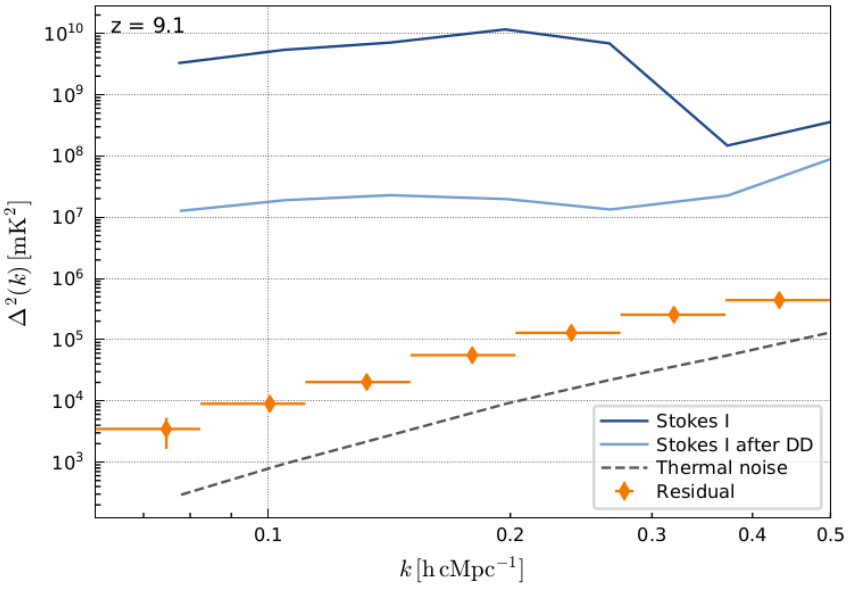


Mertens, Bobin, Carucci 2024

GPR on LOFAR data

NCP field, 140 hours, 134-146 MHz, z ~ 9.1

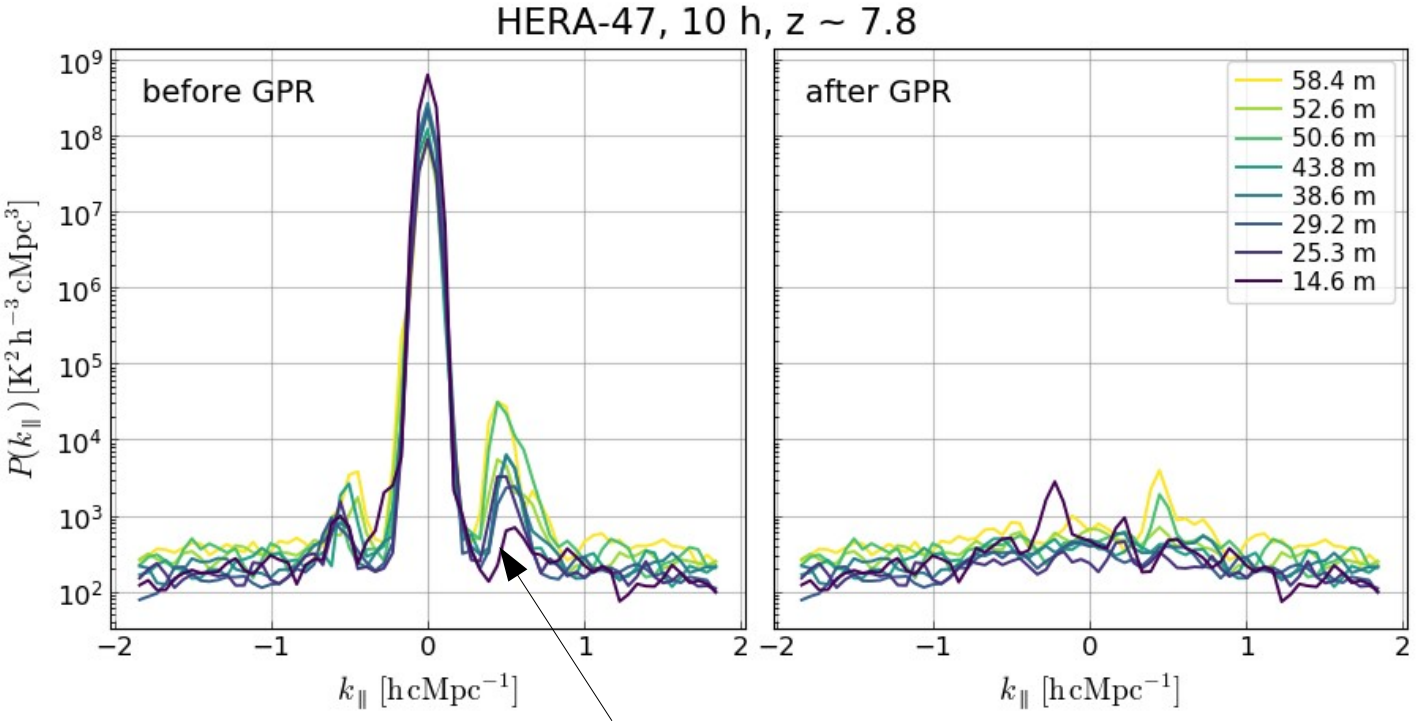


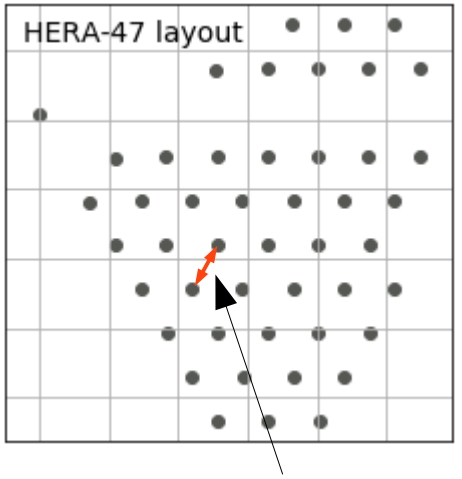


GPR remove the frequency-coherent structure
Residual power level close to thermal noise

Mertens et al. 2020

GPR on HERA data





14.6 m baseline

Periodic component: additional covariance kernel

$$K_{\mathrm{fg}} = K_{\mathrm{sky}} + K_{\mathrm{mix}} + K_{\mathrm{per}}$$

$$\kappa_{\mathrm{per}}(\nu_p, \nu_q) = \sigma^2_{\mathrm{per}} \exp\left(-\frac{r^2}{2l^2_{\mathrm{per}}}\right) \cos\left(\frac{2\pi r}{p_{per}}\right)$$

Ghosh, Mertens et al. 2020

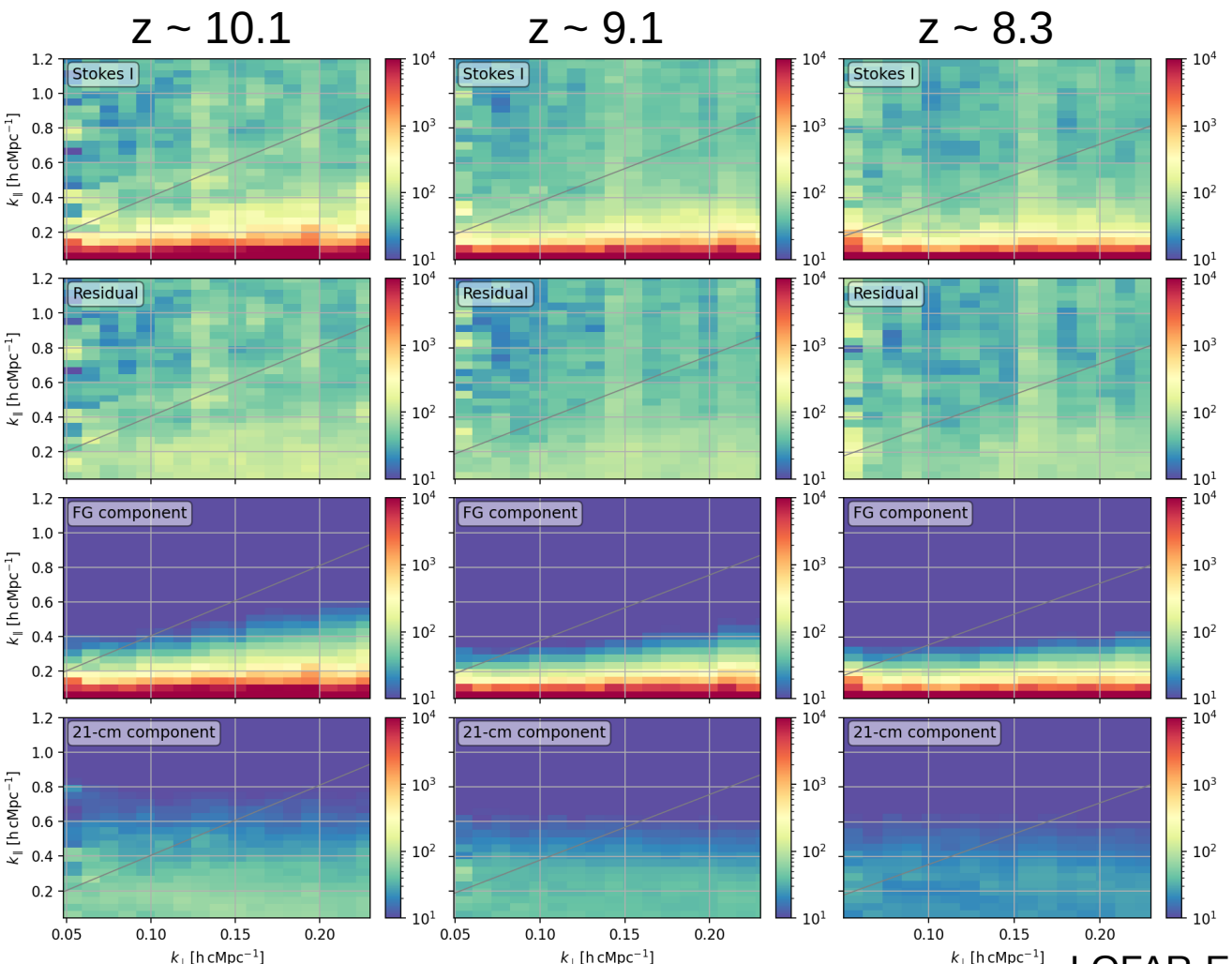
ML-GPR on LOFAR data

Observed data

Residual after Foregrounds removed

Foregrounds component

21-cm (VAE) component



Preliminary results

LOFAR-EoR collab., in prep

SKA Data Challenge 3a

The goal: recover the power-spectra of the 21-cm signal from a simulated SKA data cubes.

The DOTSS-21 team :
24 members (FR : 10, NL : 11)

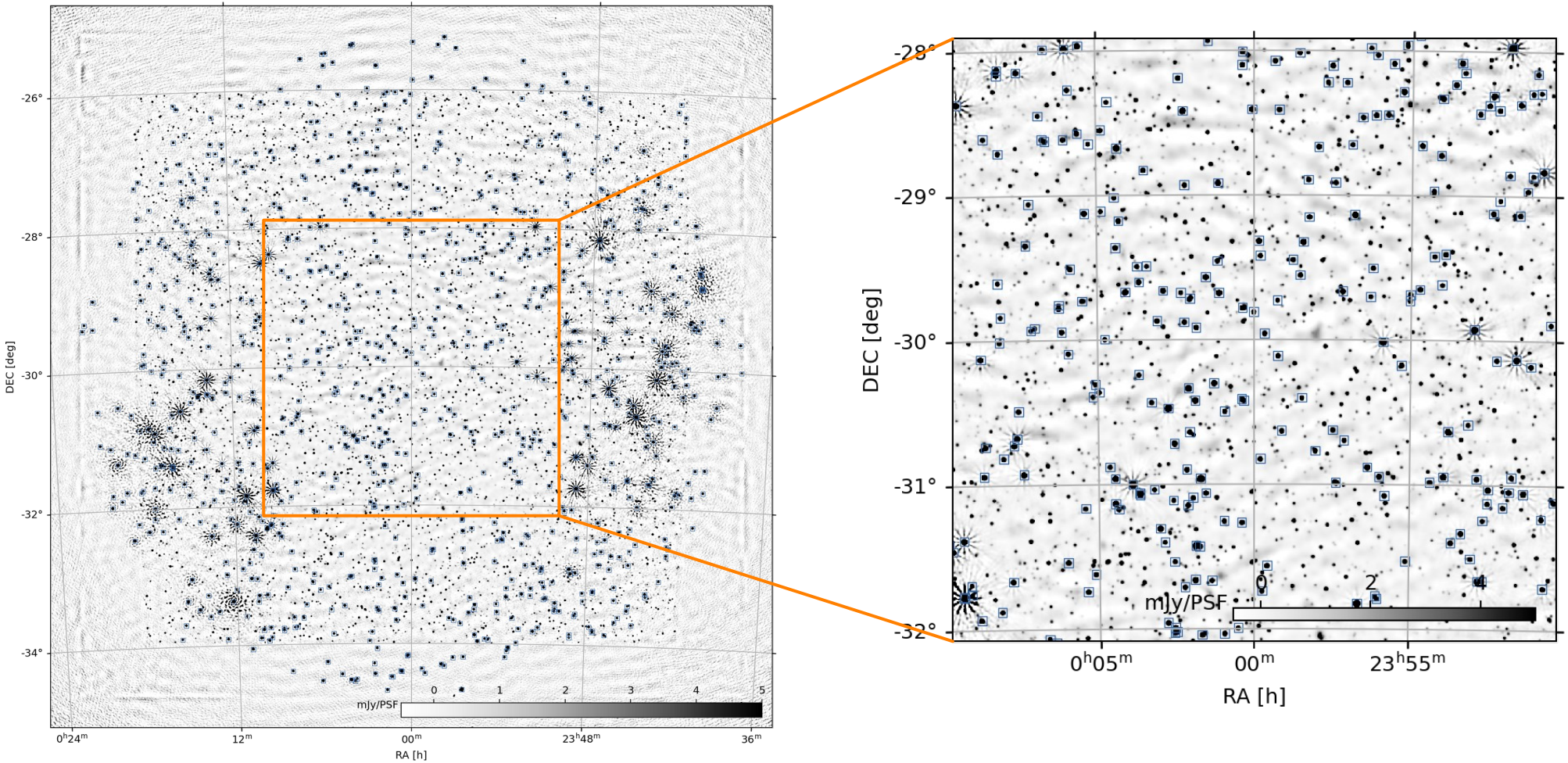
Out team finished second !

Our approach:

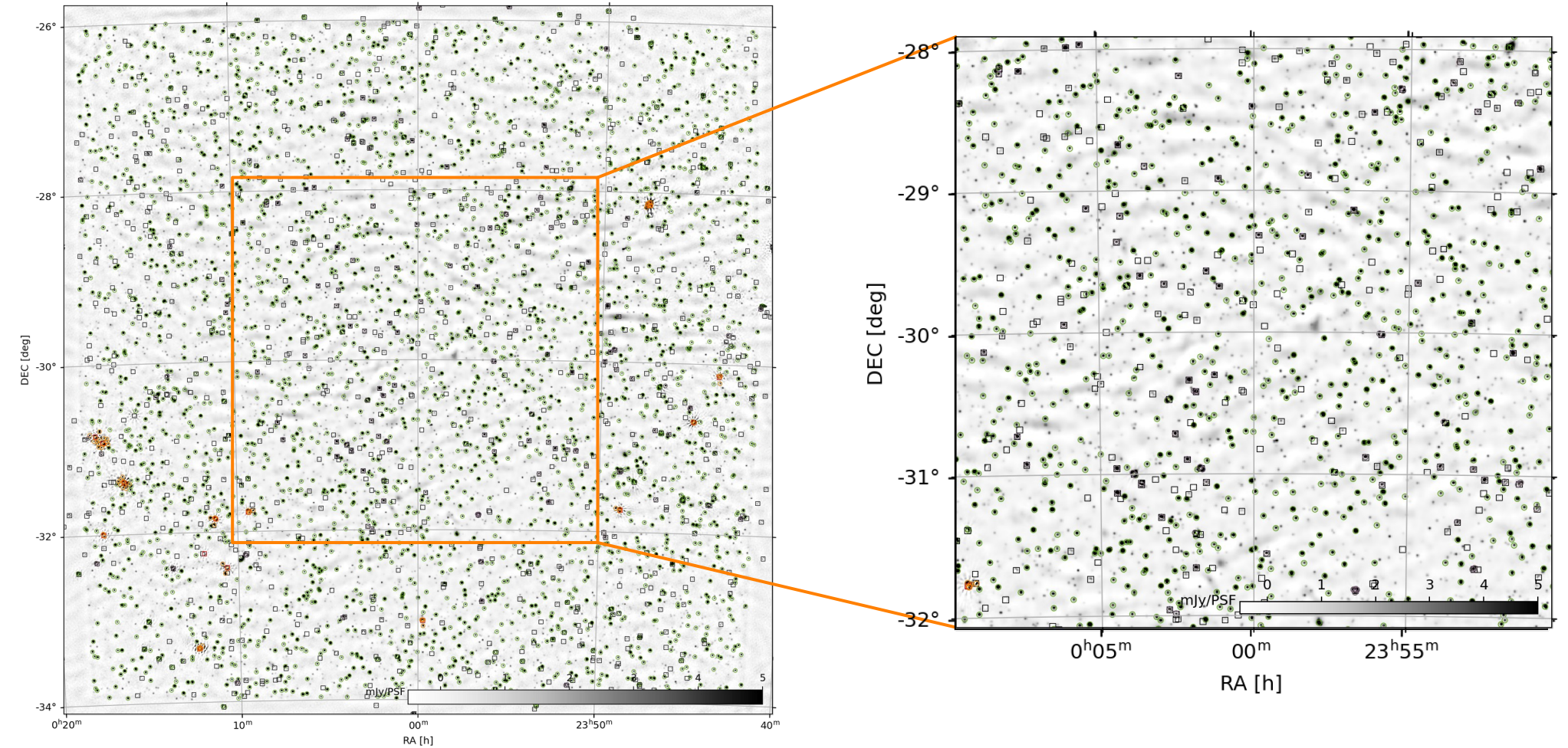
➔ Build on the LOFAR-EoR and NenuFAR Cosmic Dawn experience.

➔ Detect and subtract compact sources
➔ Model and subtract the Galactic diffuse emission
➔ Extract the 21-cm signal with ML-GPR
➔ Produce the power-spectra

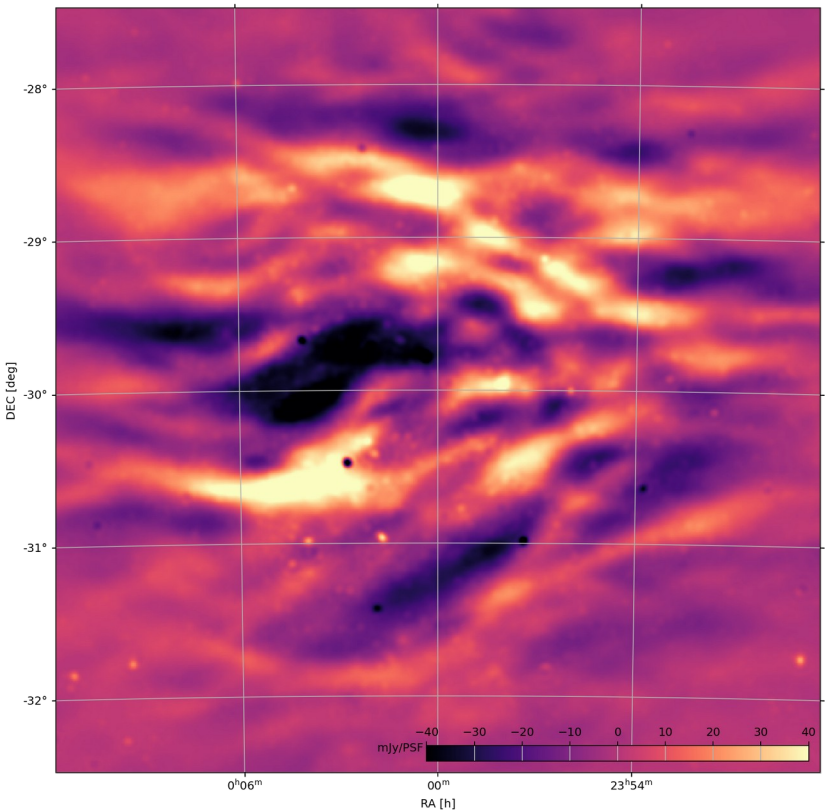
Bright point source removal



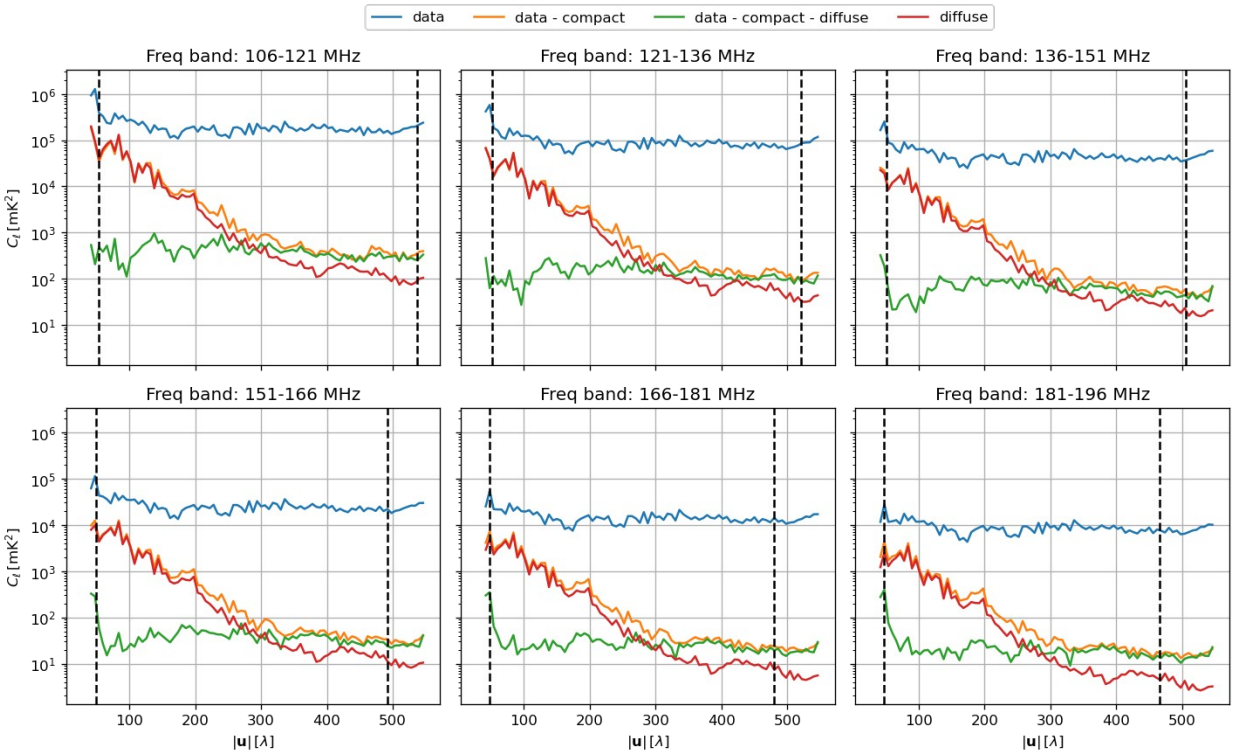
Faint point source removal



Diffuse emission modeling and subtraction

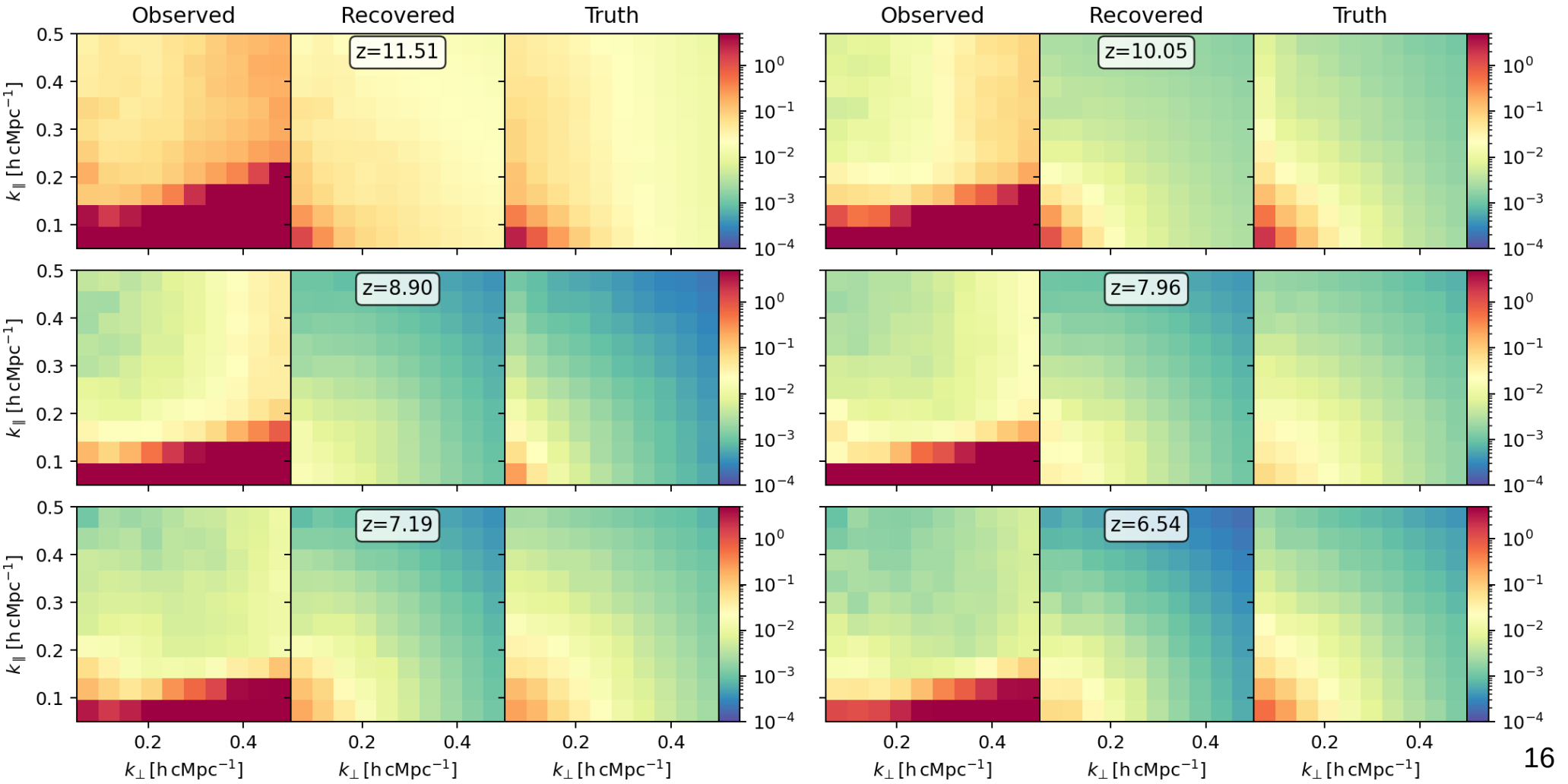


Diffuse emission model, build using WSClean mult-scale cleaning method.



Angular power-spectra in the 6 frequency bands.

SKA Data Challenge 3a - Result



Summary

- The 21-cm signal from the Epoch of Reionization and Cosmic Dawn promises a new and unique probe of the first billion year of the Universe, **but very challenging experiment.**
- Main challenge: **Foregrounds.**
- The **Gaussian Process Regression** method allows to make a statistical separation of the Foregrounds and the 21-cm signal.
- **ML-GPR: learn covariance prior function from simulations** (FG and 21-cm simulations).
- Simulations with ML-learned 21-cm covariance shows **excellent recovery** of the signal even in the presence of excess component, when it can be accounted for.
- Used on several 21-cm experiments (**LOFAR, AARTFAAC, NenuFAR**). Tested on HERA data.
- **Next:** Foregrounds trained covariance, combining multiple redshift bins.