

Deconvolving the wedge: Robust Foregrounds removal for 21-cm experiments

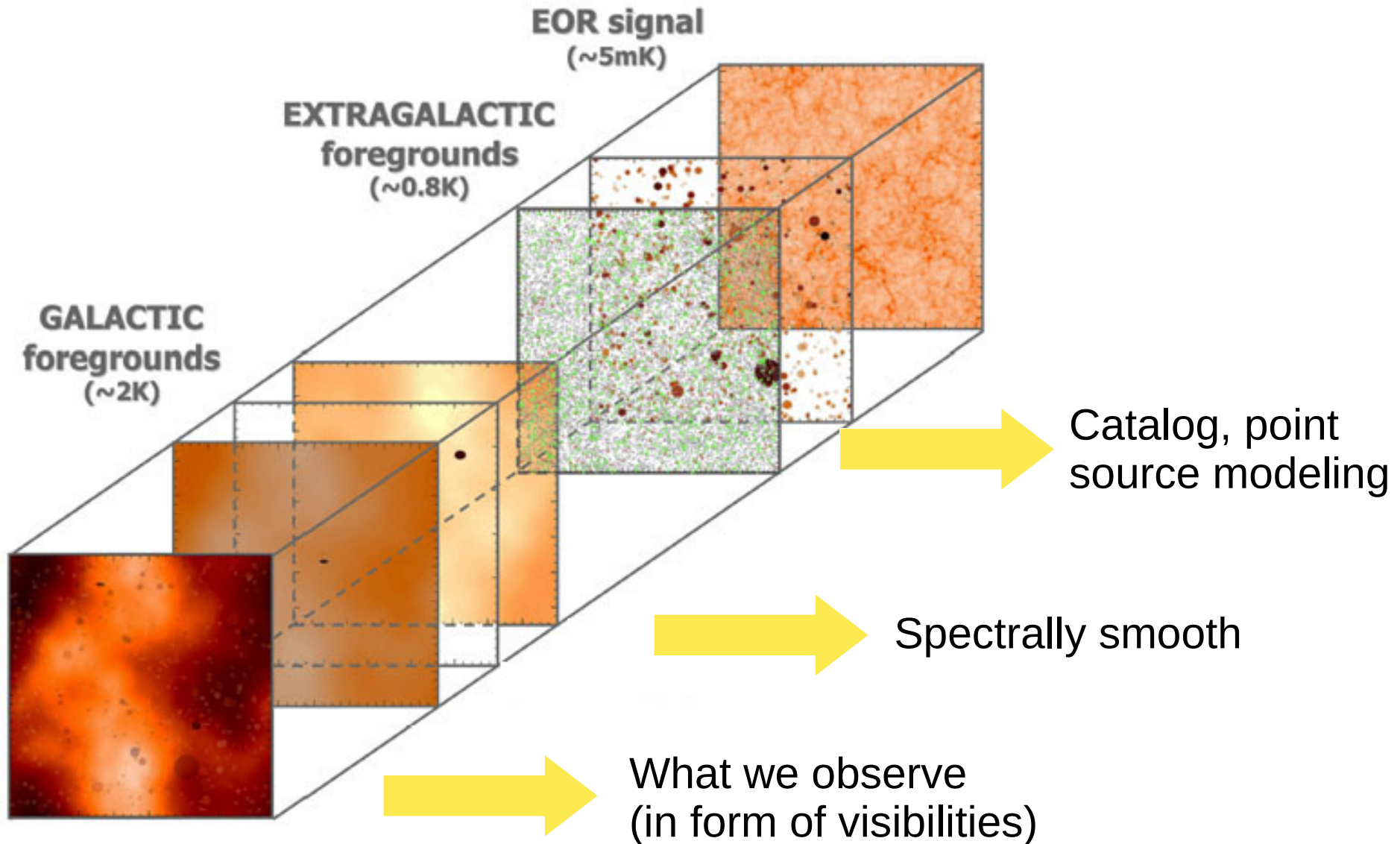


LOFAR

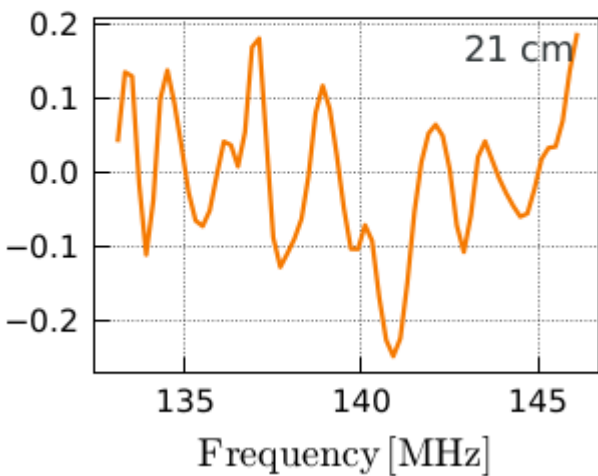
Florent Mertens, Abhik Ghosh, Leon Koopmans
and the LOFAR-EoR team



Detecting the EoR

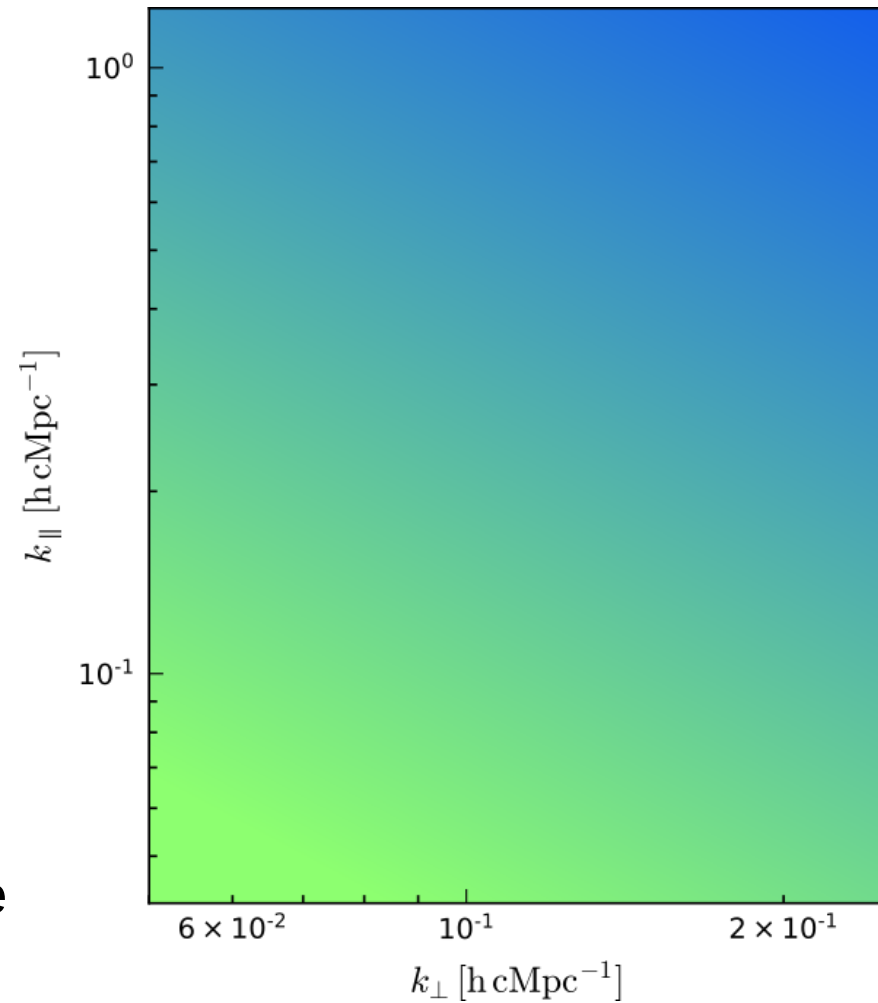


Detecting the EoR

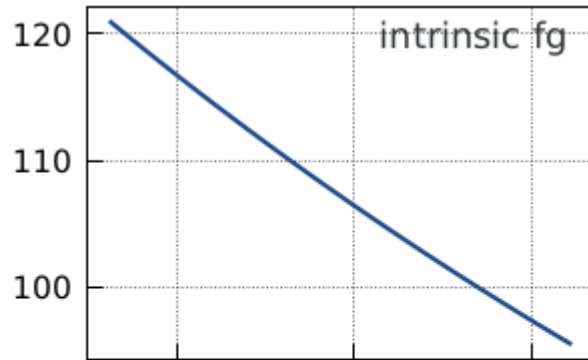


21-cm signal:

- Uncorrelated \sim MHz
- Isotropic in 3D volume

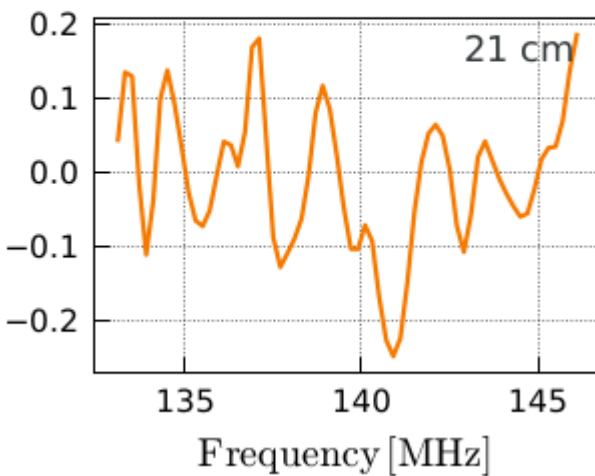


Detecting the EoR



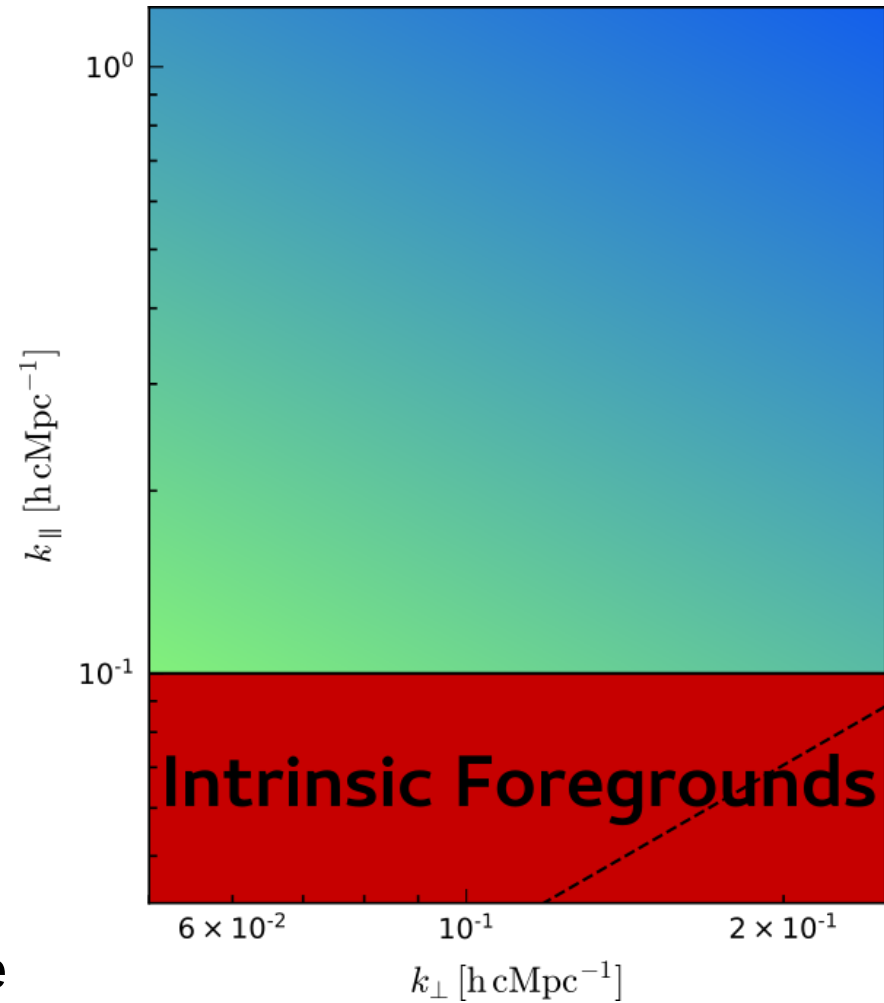
Diffuse emission and residual extragalactic sources:

- Smooth in frequency

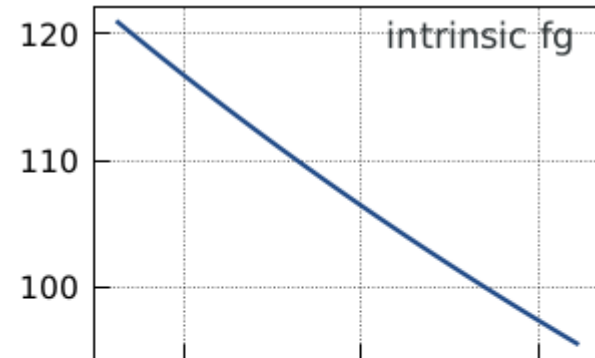


21-cm signal:

- Uncorrelated \sim MHz
- Isotropic in 3D volume

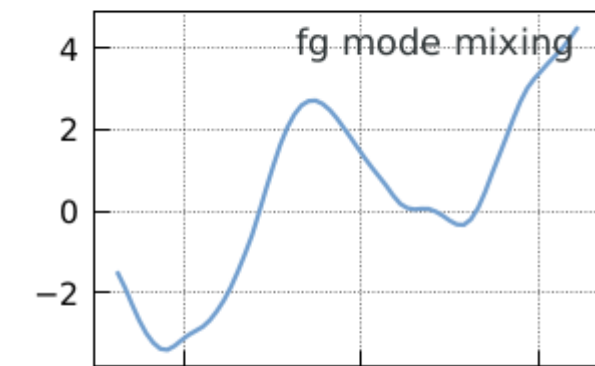


Detecting the EoR



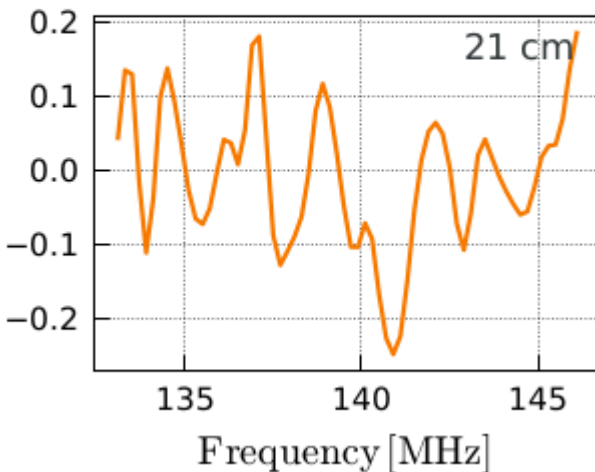
Diffuse emission and residual extragalactic sources:

- Smooth in frequency



Mode mixing:

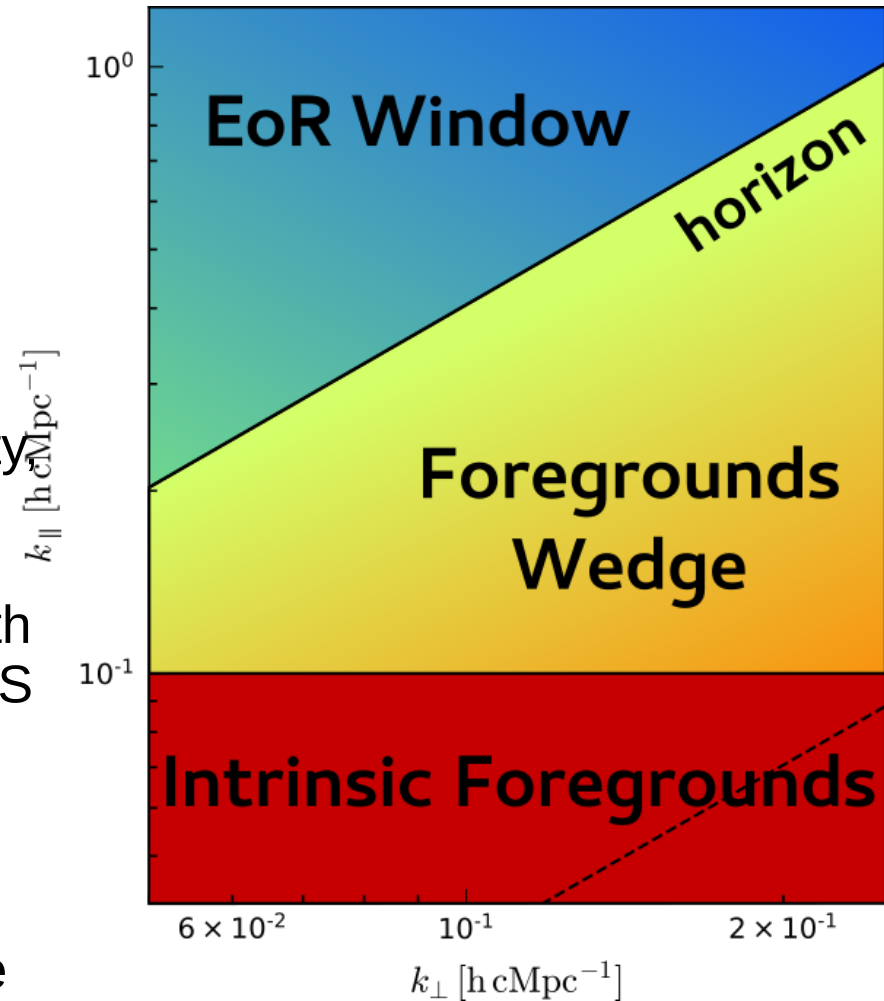
- instrument chromaticity, imperfect calibration, ionosphere...
- Less frequency smooth
- Wedge shape in 2D PS



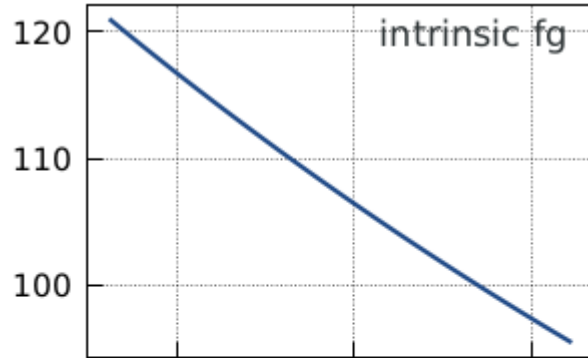
21-cm signal:

- Uncorrelated \sim MHz
- Isotropic in 3D volume

+ Noise

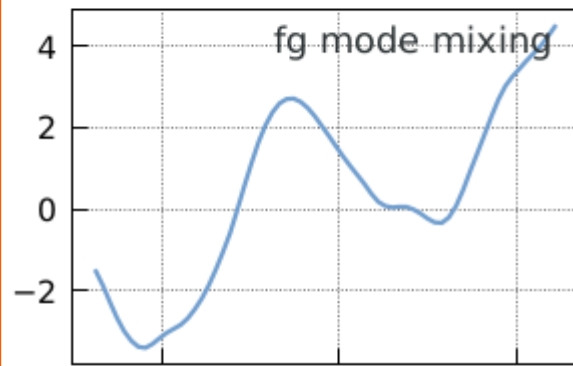


Detecting the EoR



Diffuse emission and residual extragalactic sources:

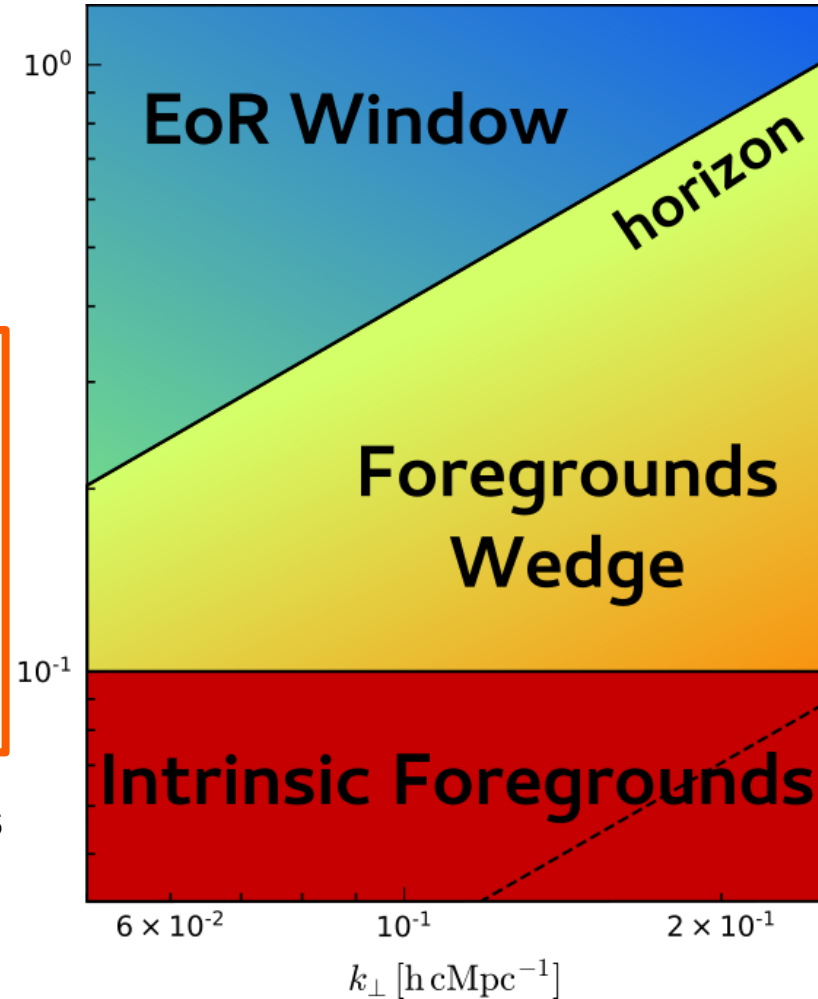
- Smooth in frequency



Mode mixing:

- instrument chromaticity, imperfect calibration, ionosphere...
- Less frequency smooth
- Wedge shape in 2D PS

k_{\parallel} [h cMpc⁻¹]



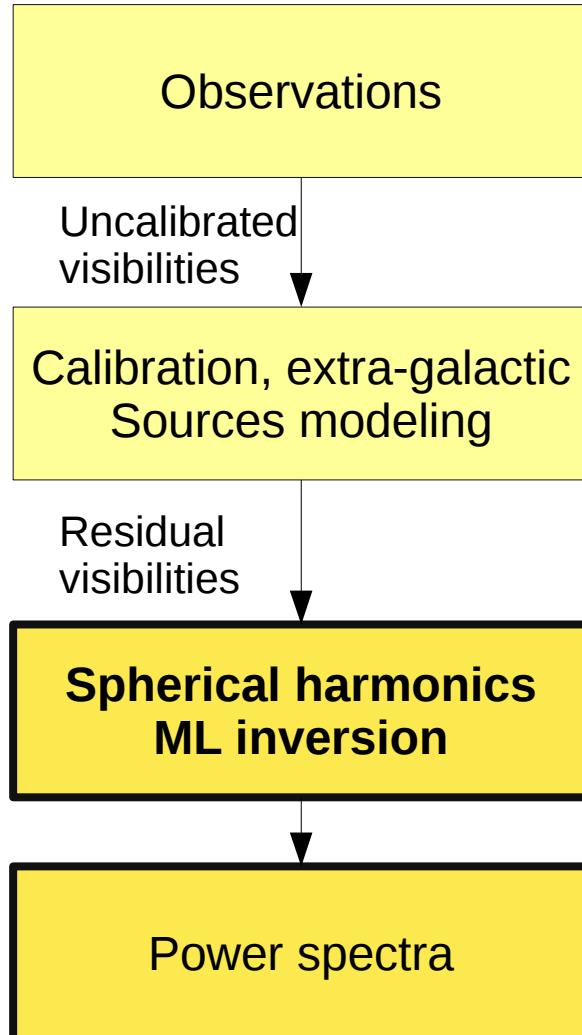
Major obstacle of residual foregrounds modeling and removal.



What we need:

1. Reduce the effect of instrument chromaticity:
ML Spherical Wave visibility modeling (Ghosh, FM, LK, sub.)
2. Model the Foregrounds including mode mixing:
Gaussian Process Regression (Mertens, AG, LK, in prep.)

Spherical Harmonics ML inversion



- Orthogonal basis on the sphere:
Full sky, no flat sky approximation
- Relation with visibilities (McEwen 2008, Carozzi 2015):

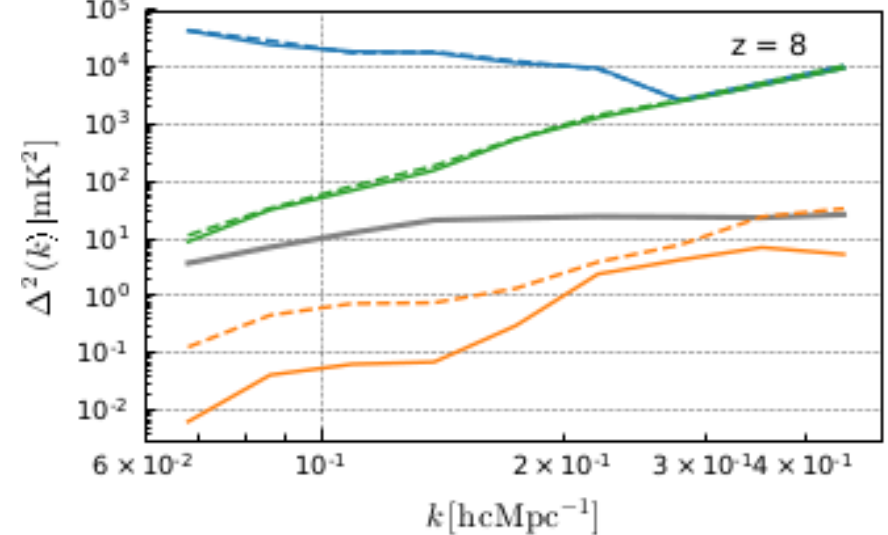
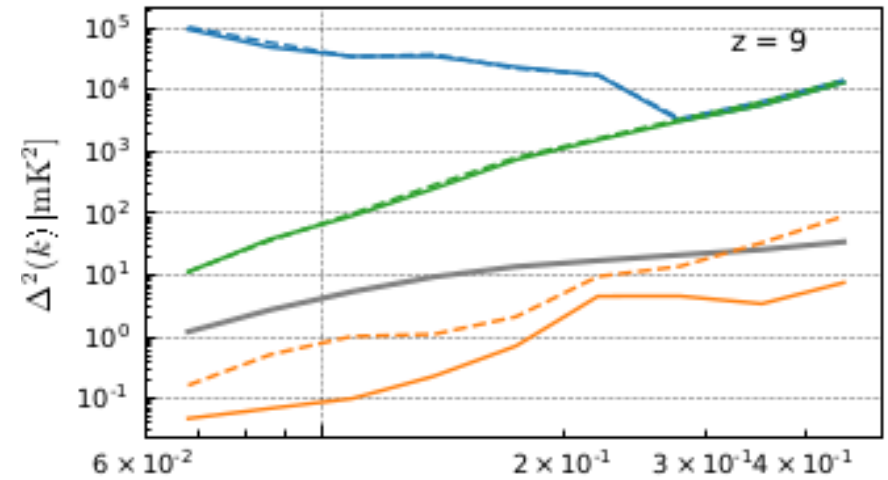
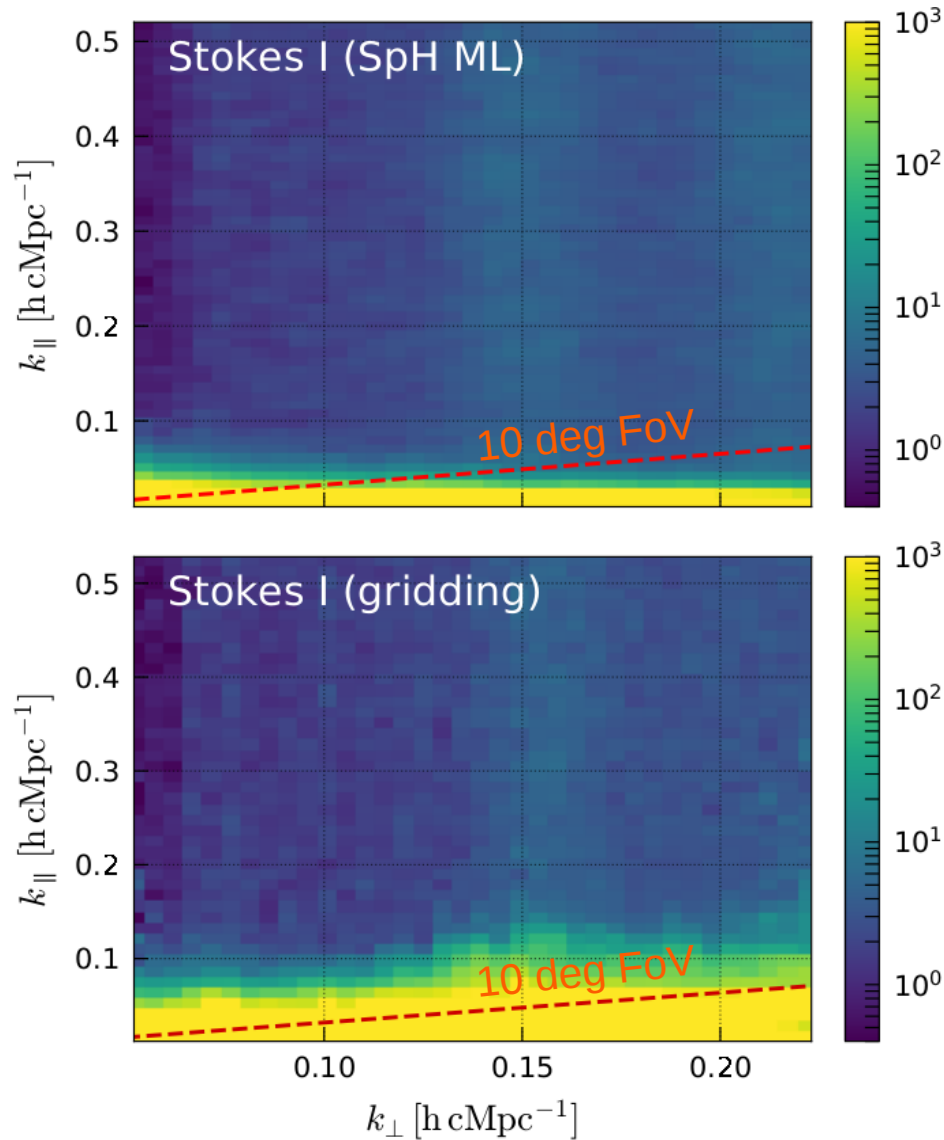
$$\mathcal{V} = \sum_{lm} 4\pi(-i)^\ell b_{\ell m} j_\ell(kr) Y_{\ell m}(\Omega_k)$$

→ ML inversion

- Direct estimate of the power spectra:

$$C_\ell = \frac{4\pi}{\Omega_{\text{PB}}} \frac{1}{\ell + 1} \sum_{m=0}^{\ell} |b_{\ell m}|^2$$

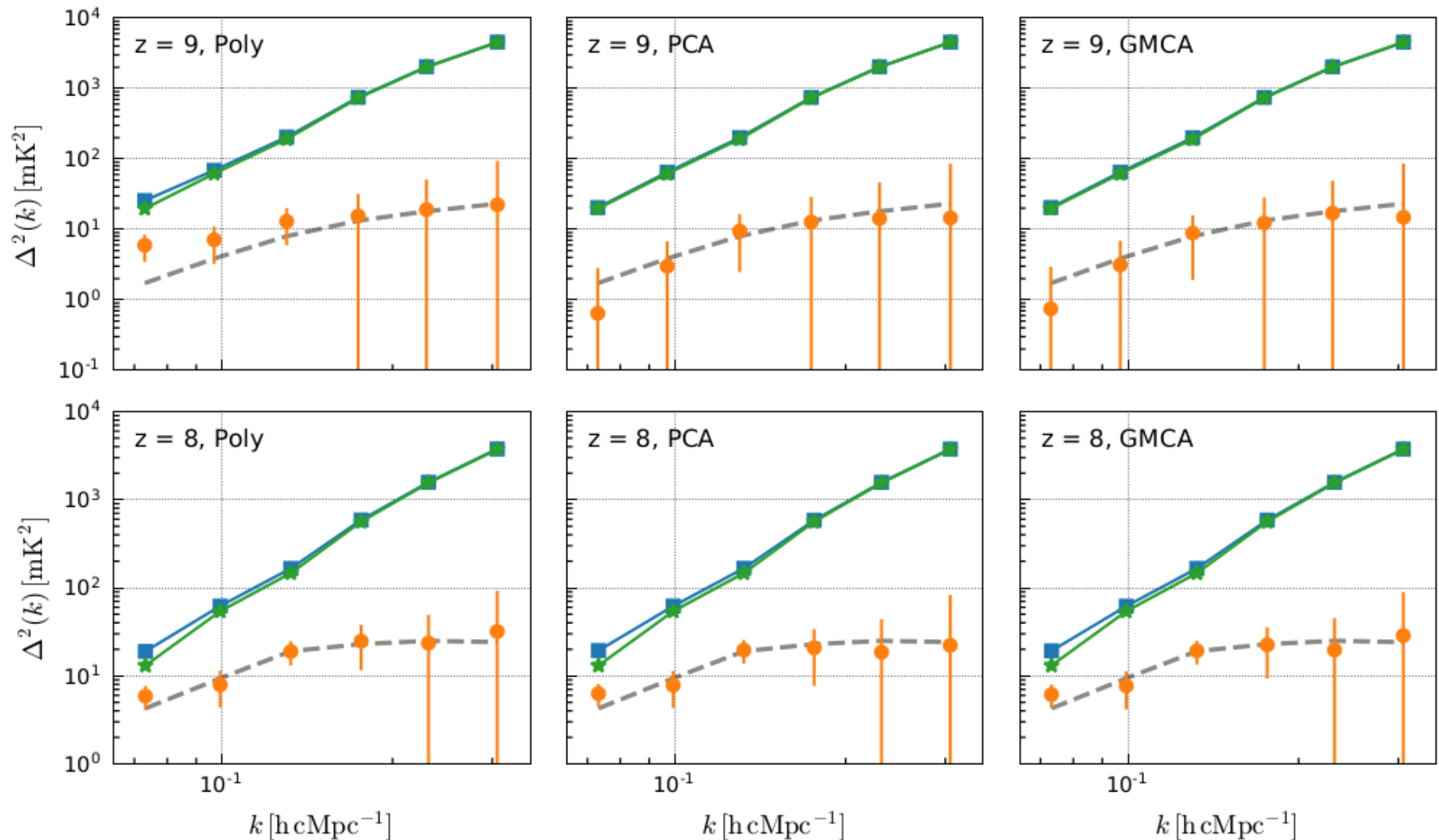
SpH ML inversion: simulation



- Mode mixing contaminants considerably reduced.
- The wedge is effectively deconvolved

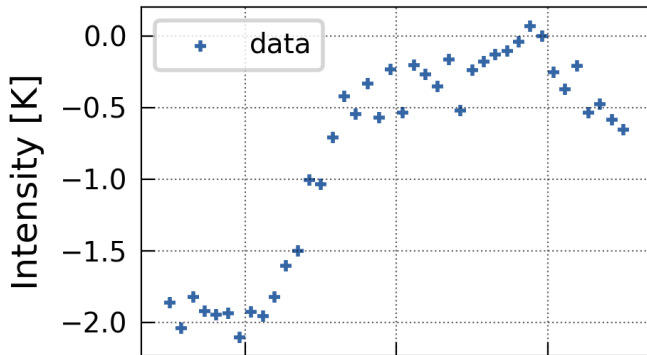
Estimator error (orange line) below the 21-cm input signal (gray).

SpH ML inversion: EoR detection



- Recovered sky frequency-smooth: foregrounds easy to remove with low order polynomial fit or blind source separation methods.
- Subtracting the noise bias (green) we can recover the PS of the 21-cm line signal (orange) which is compared to the input 21-cm signal (gray dashed line).

Gaussian Process Regression



A Gaussian Process (GP) is a collection of random variables with a joint Gaussian distribution:

$$\mathbf{f} \sim \mathcal{N}(\mathbf{0}, K)$$

Joint distribution of observed data y and prior model f :

$$P(y, \mathbf{f}) \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(x, x) + \sigma_n^2 I & K(x, x') \\ K(x', x) & K(x', x') \end{bmatrix} \right)$$

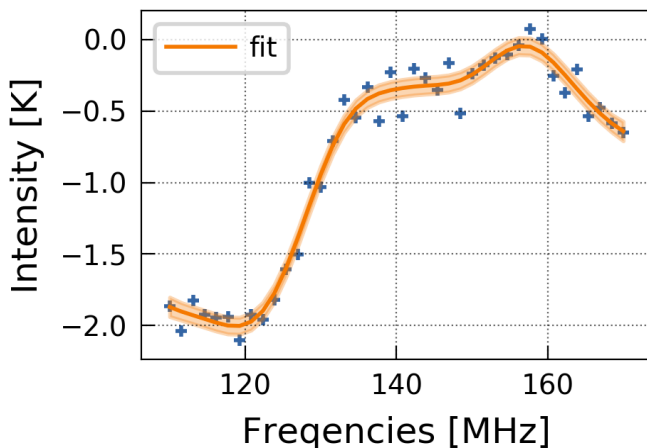
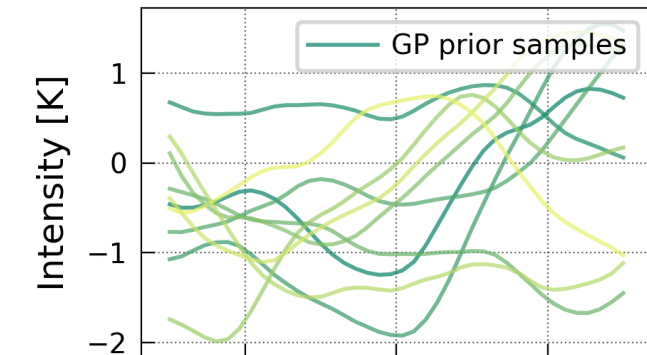
Conditioning the prior model to the data, we obtain another GP which is analytically computable:

$$P(\mathbf{f}|y) = \mathcal{N}(\bar{\mathbf{f}}, \text{cov}(\mathbf{f}))$$

with:

$$\bar{\mathbf{f}} = K(x', x) \left[K(x, x) + \sigma_n^2 I \right]^{-1} y$$

$$\text{cov}(\mathbf{f}) = K(x', x') - K(x', x) \left[K(x, x) + \sigma_n^2 I \right]^{-1} K(x, x')$$



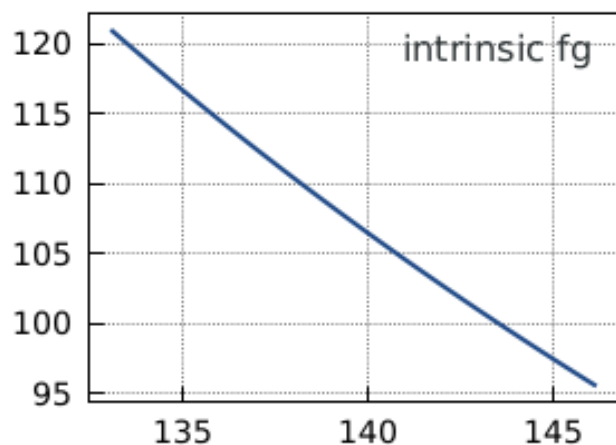
GPR modeling for EoR experiments

Observed data can be decomposed in three main components.
Joint distribution of observed data y and prior model f :

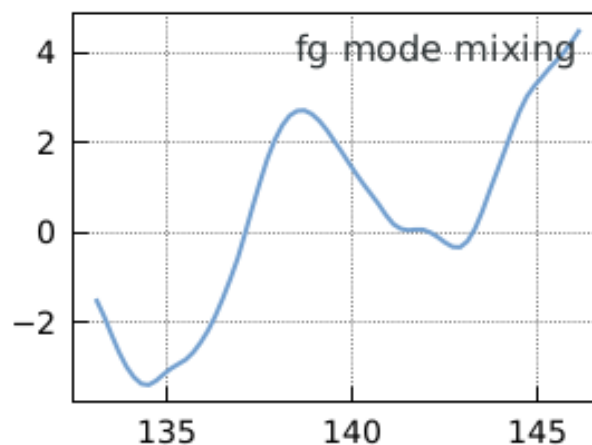
$$P(y, \mathbf{f}) \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_{fg}(x, x) + K_{eor}(x, x) + \sigma_n^2 I & K_{fg}(x, x') \\ K_{fg}(x', x) & K_{fg}(x', x') \end{bmatrix} \right)$$

$$K_{fg} = K_{fg_int} + K_{fg_mix}$$

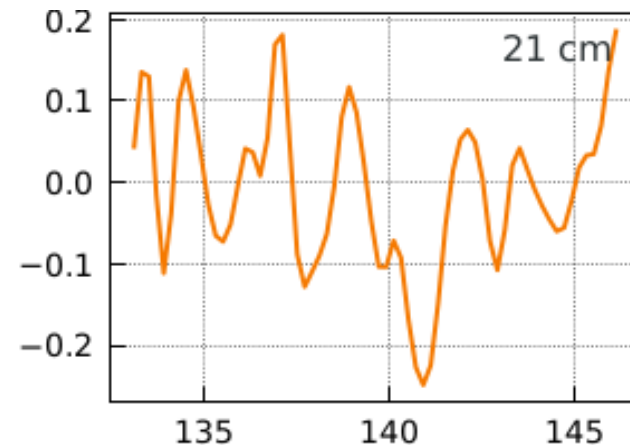
Two hyper-parameters per covariance functions: variance and length-scale
→ 6 in total, optimized by maximizing the marginal likelihood.



Diffuse emission and residual extragalactic sources:
Smooth in frequency



Mode mixing:
Less frequency smooth

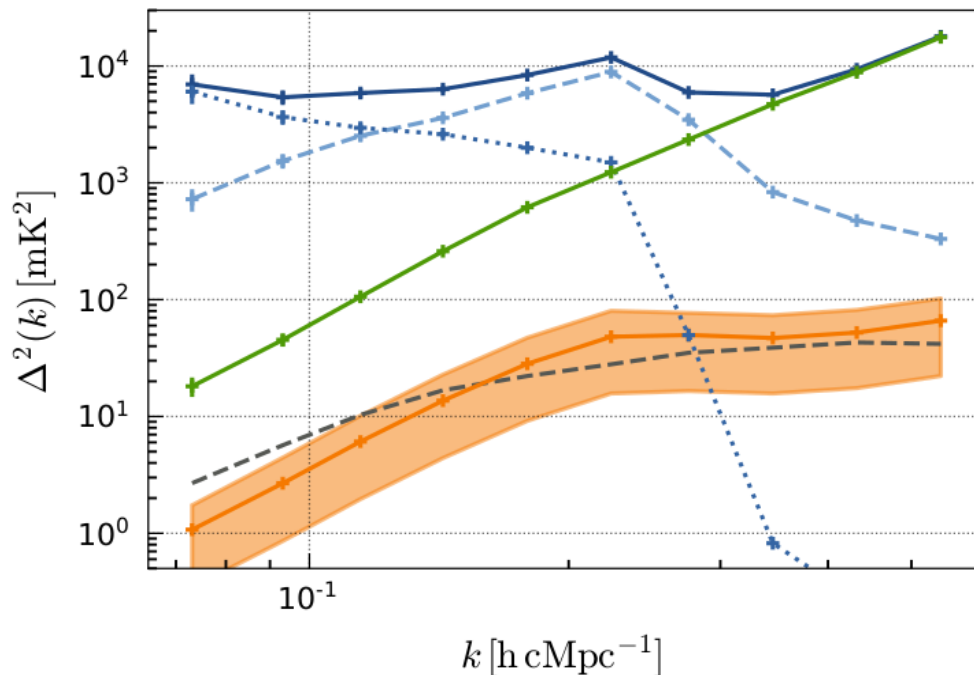


21-cm signal:
Uncorrelated ~ MHz

GPR simulation

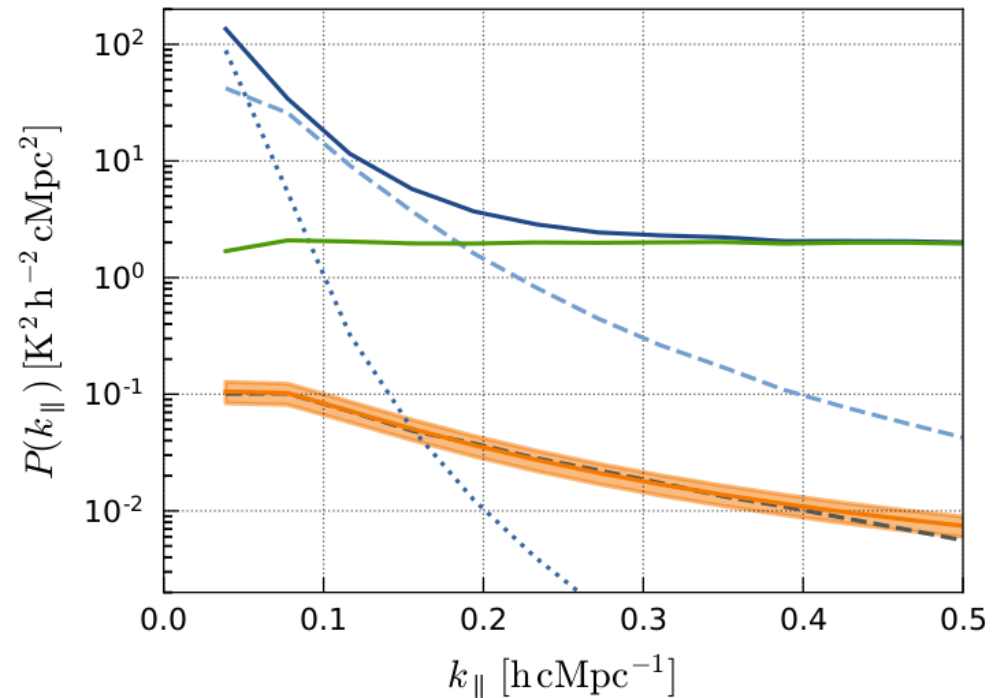
Simulation:

- **Intrinsic foregrounds:** galactic diffuse emission
- **Mode mixing foregrounds:** GP model with length-scale of 2 MHz
- EoR input signal: simulated from 21 cm fast
- **noise:** equivalent to 1200 hours of LOFAR



Spherically averaged power spectra.

Orange line and filled region are the mean and rms of the recovered EoR signal using GPR



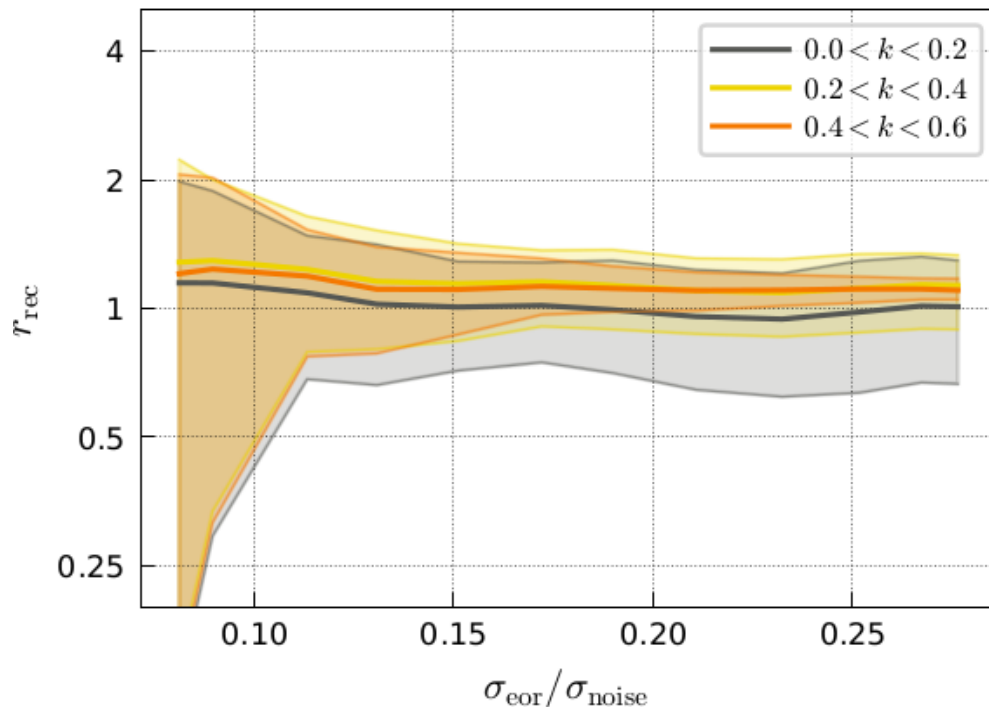
Cylindrically-averaged power spectra as function of k_{par}

GPR simulation

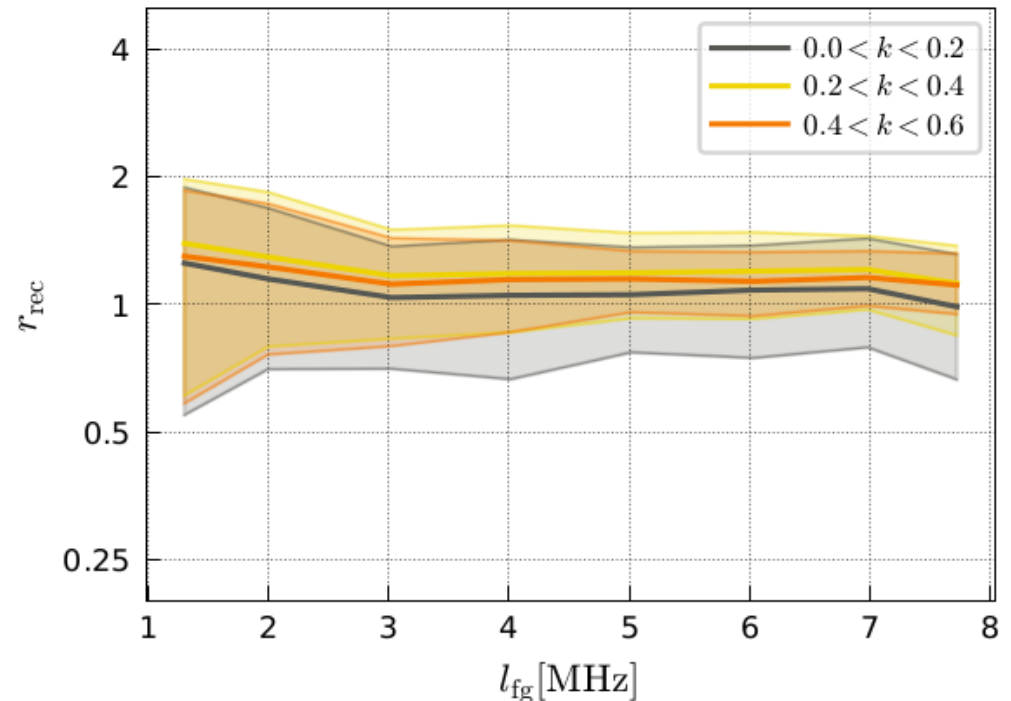
Estimating the biases over the parameter space:

Ratio of recovered over input simulated 21-cm signal with varying

21-cm signal intensity



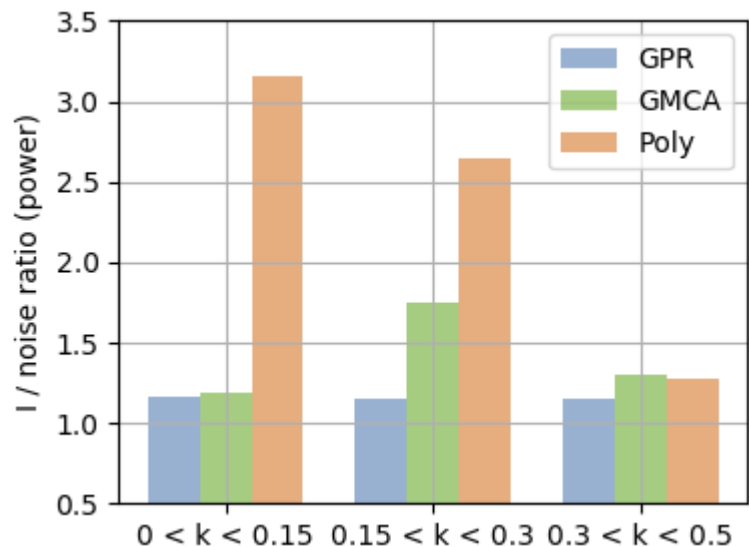
foregrounds mode-mixing length-scale



- Limited biases. Could be folded in MCMC determination of physical EoR parameters.
- Limiting the wedge in intensity and extend also reduce the biases.

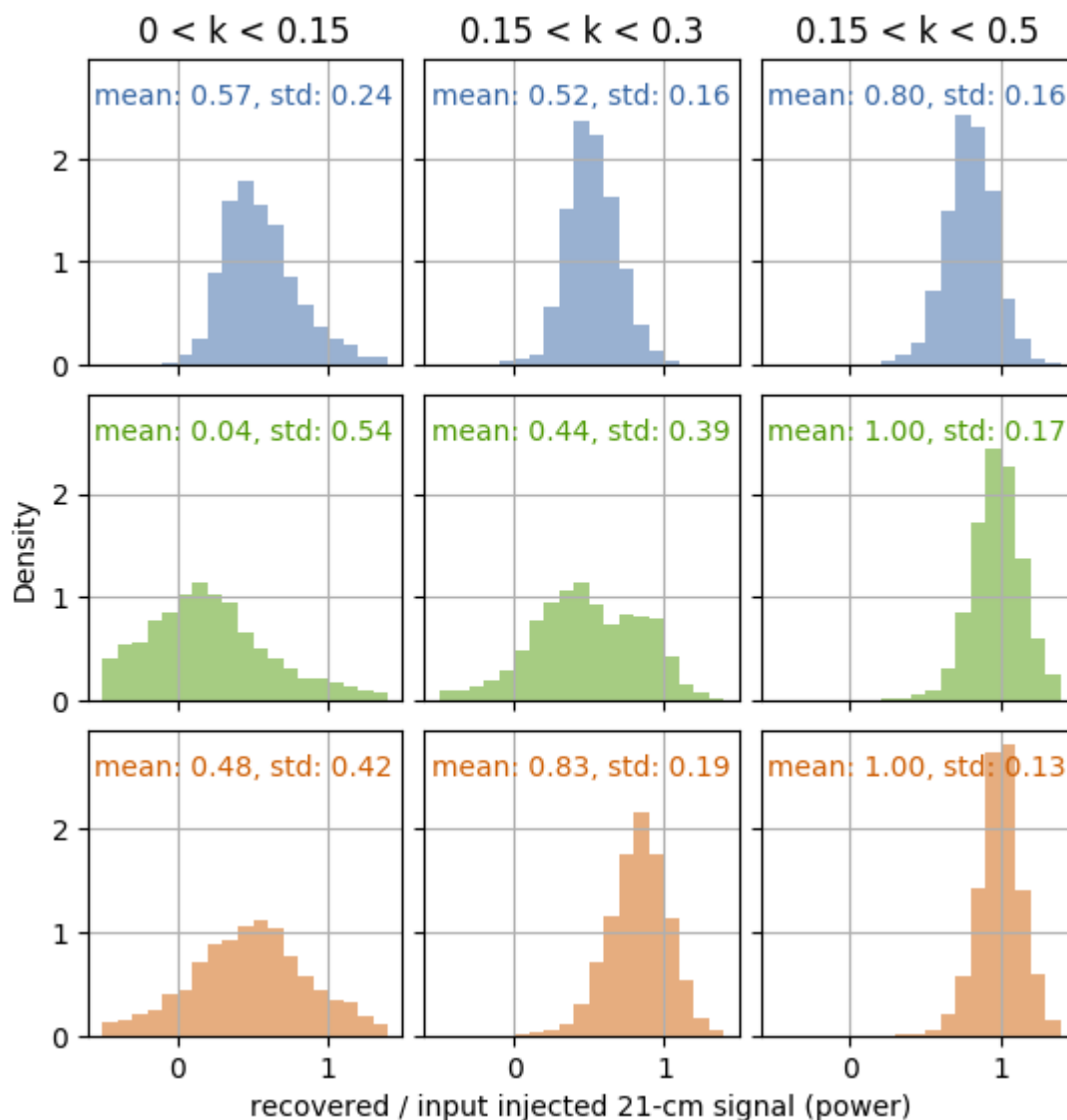
Injection test with LOFAR data

Ratio of residual I over noise:
How much foregrounds is removed ?



- At large scale GPR perform better than any other methods while keeping about 60 % of the signal.
- Bias is also more predictable.

21-cm signal injection test:
How much signal is suppressed ?



Outlook

- Spherical Harmonics ML inversion promising alternative to classical imaging: reduce PSF mode-mixing and side lobe noise.
- Gaussian Process Regression (GPR), a non-parametric but not blind foregrounds removal technique.
- GPR outperforms currently available methods: better model the data and reduce over-fitting.
- GPR performance tested using full simulation and with 21-cm signal injection test.