Gaussian Process Regression: An Application in Radio Cosmology

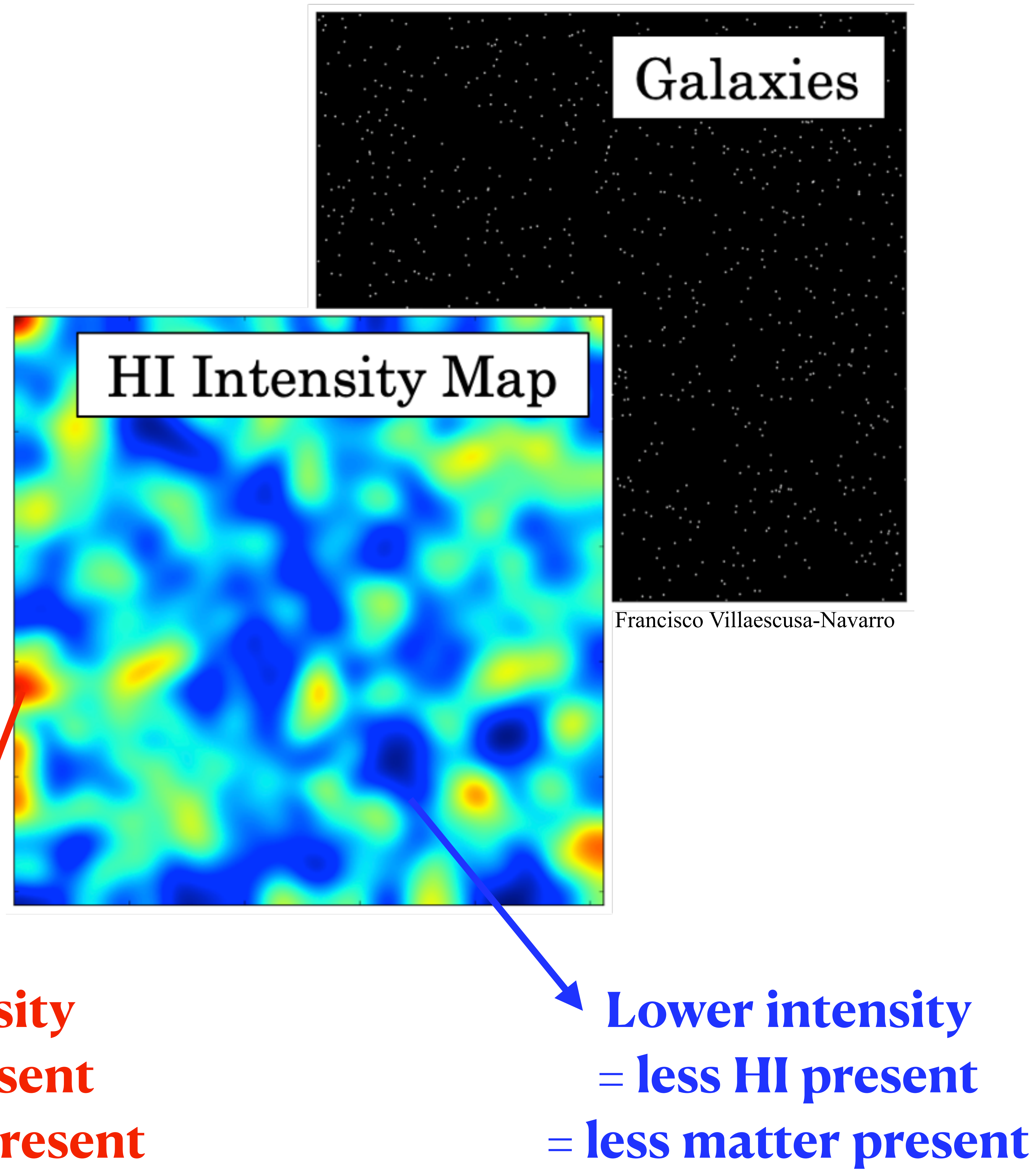
arXiv:2105.12665

Paula Soares

In collaboration with Catherine Watkinson, Steve Cunnington and Alkistis Pourtsidou

HI intensity mapping

- Large-scale structure: how matter is clustered and structured on a large scale in our Universe
- After reionisation, most of the neutral hydrogen (HI) can be found in galaxies
  - HI is a good tracer of the large-scale structure
- Can quickly map **large** areas of the sky
- **But need to remove foregrounds!**



Motivation

- GPR has already been applied as a foreground removal technique successfully in the context of the Epoch of Reionisation (see e.g. Mertens et al. 2018 [arXiv:1711.10834] and public code `ps_eor`[1])
  - * How does GPR perform in the case of low redshift, single-dish Intensity Mapping?
  - * How does it compare to other methods e.g. PCA?
  - * Could we use it for future surveys such as the SKA?

[1]gitlab.com/flomertens/ps_eor

SKA-like simulations

Assume our data, and each of its components (foreground, HI, noise) is a Gaussian process

Our data's covariance function:

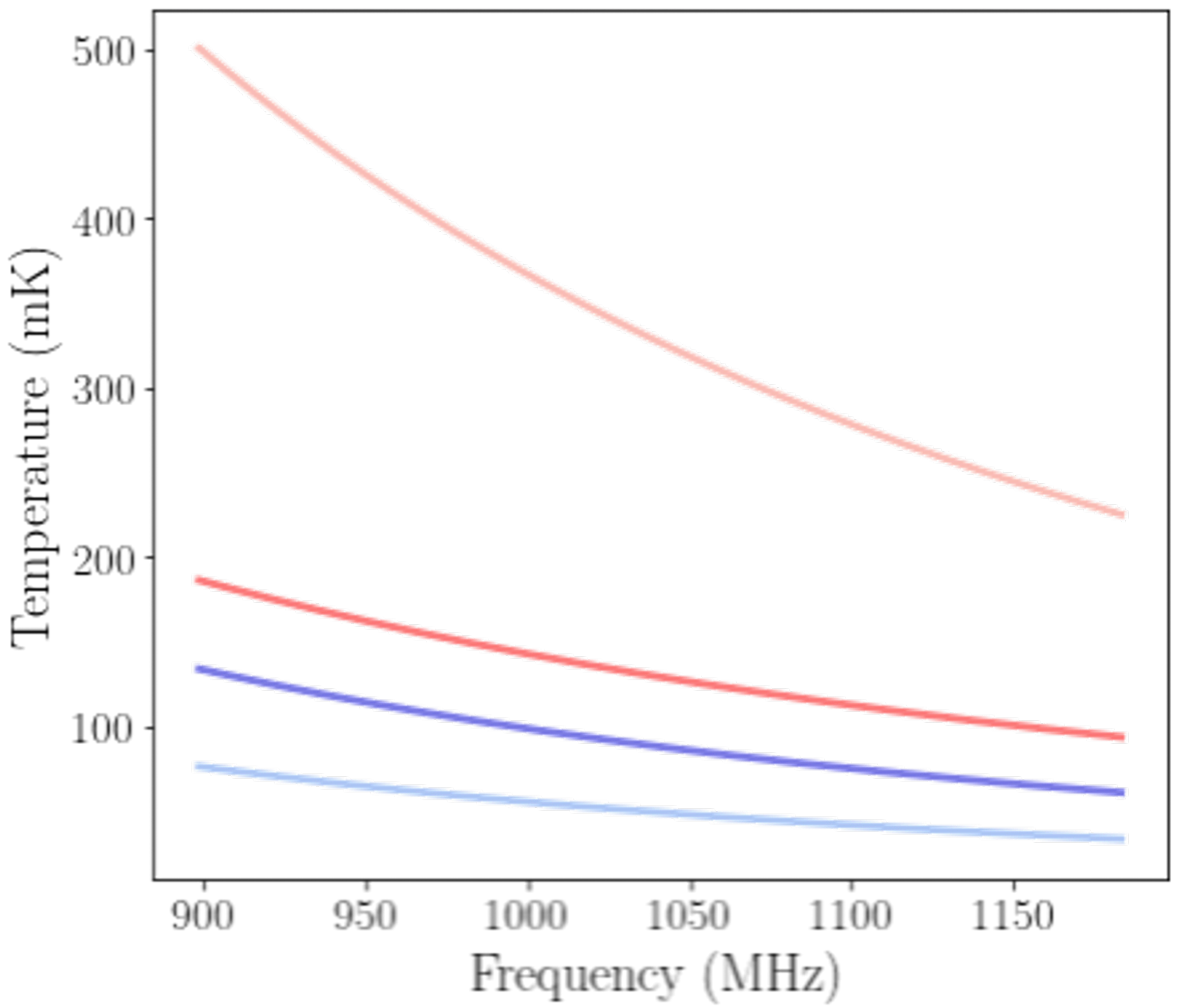
$$K = K_{\text{fg}} + K_{21} + K_{\text{noise}}$$

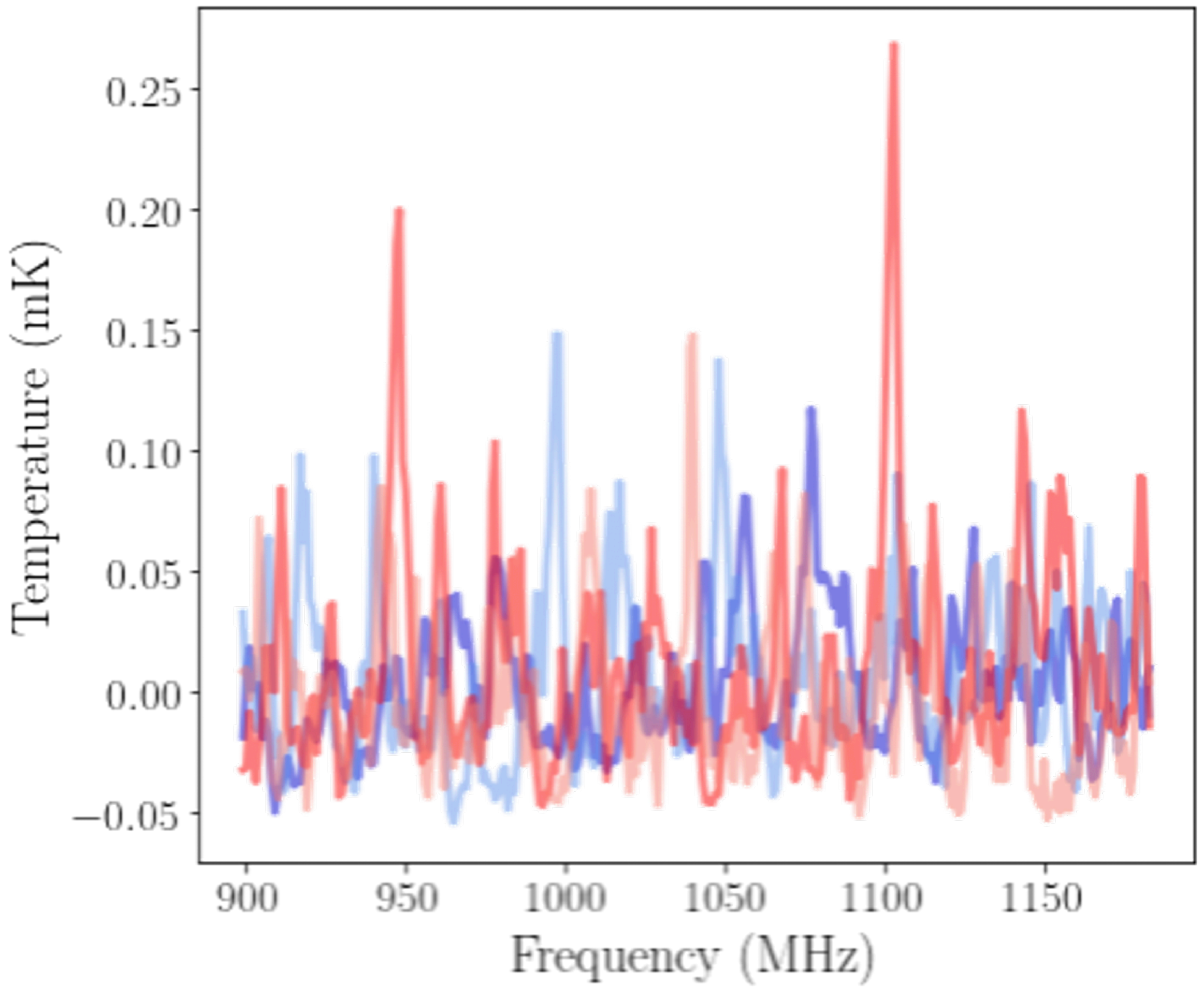
Smooth foregrounds K_{fg}

- Correlated in frequency
- High amplitude
- Overall smooth in frequency

21cm signal K_{21}

- Not correlated in frequency
- Small amplitude
- Not smooth in frequency

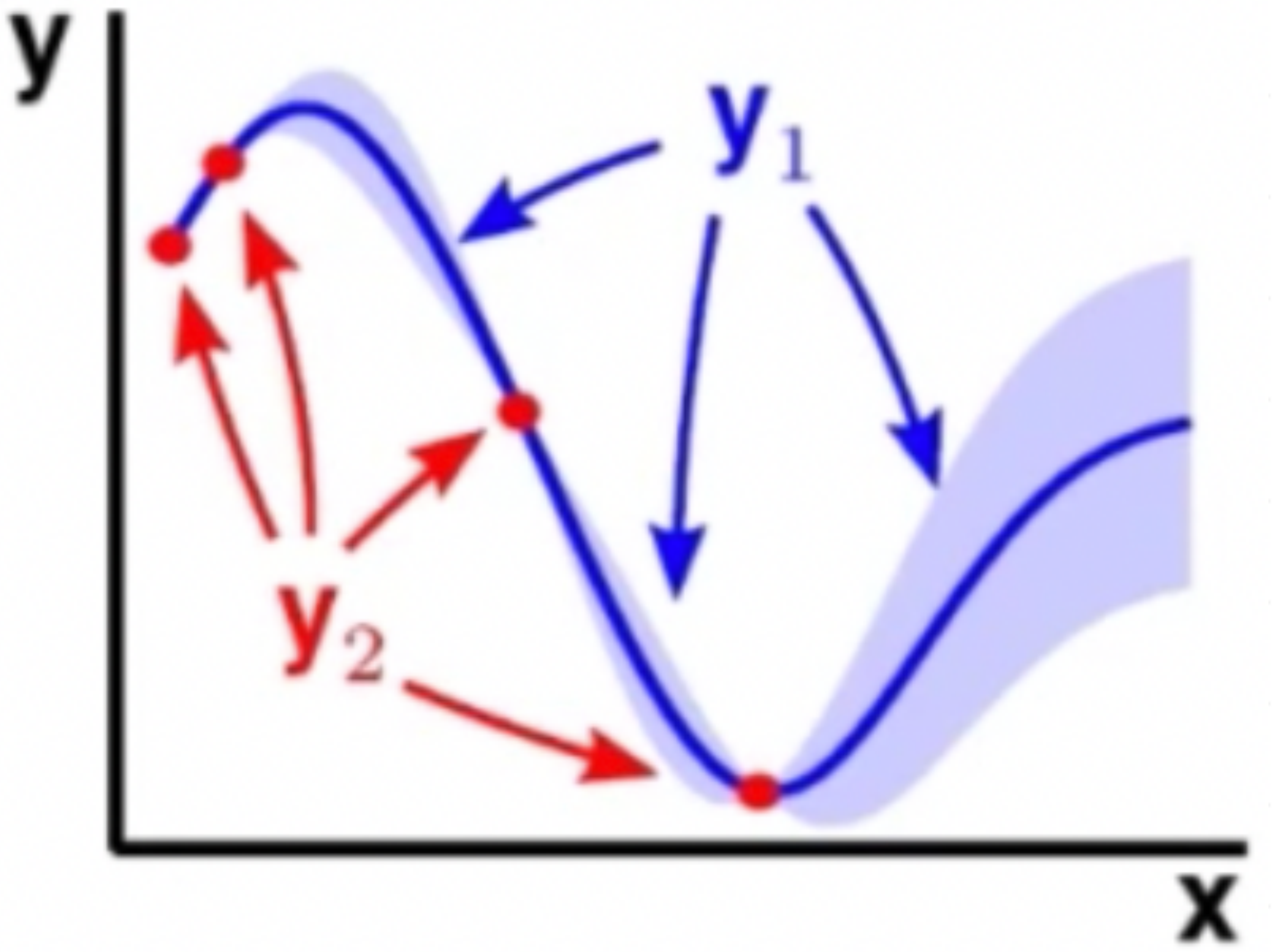




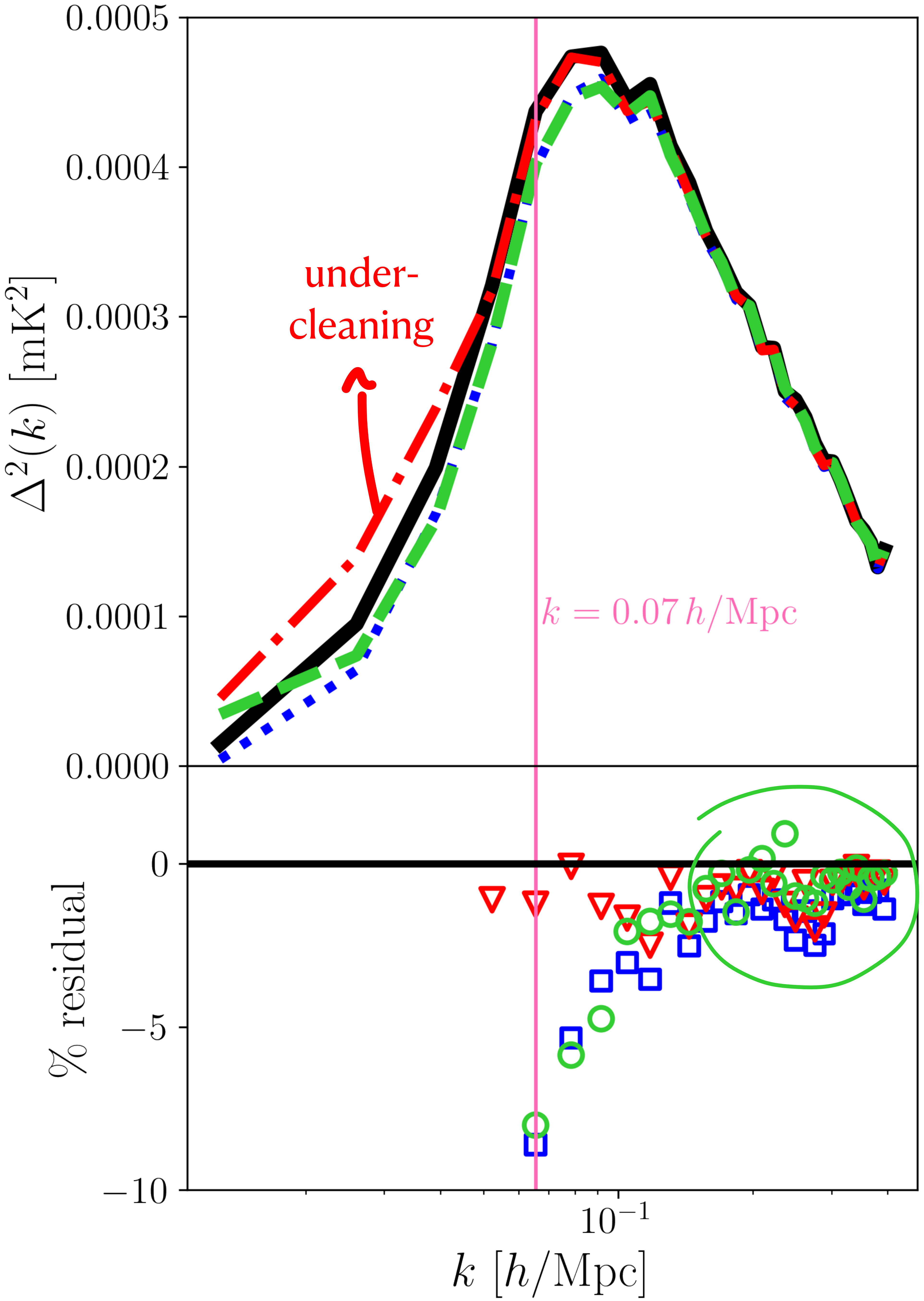
Foreground removal

How does GPR remove foregrounds? By predicting them!

Now we have: our data (d), and its best fitting covariance function $\left(K = K_{\rm fg} + K_{21} + K_{\rm noise}\right)$. We can use this to *predict what the foregrounds look like in our frequency range*:

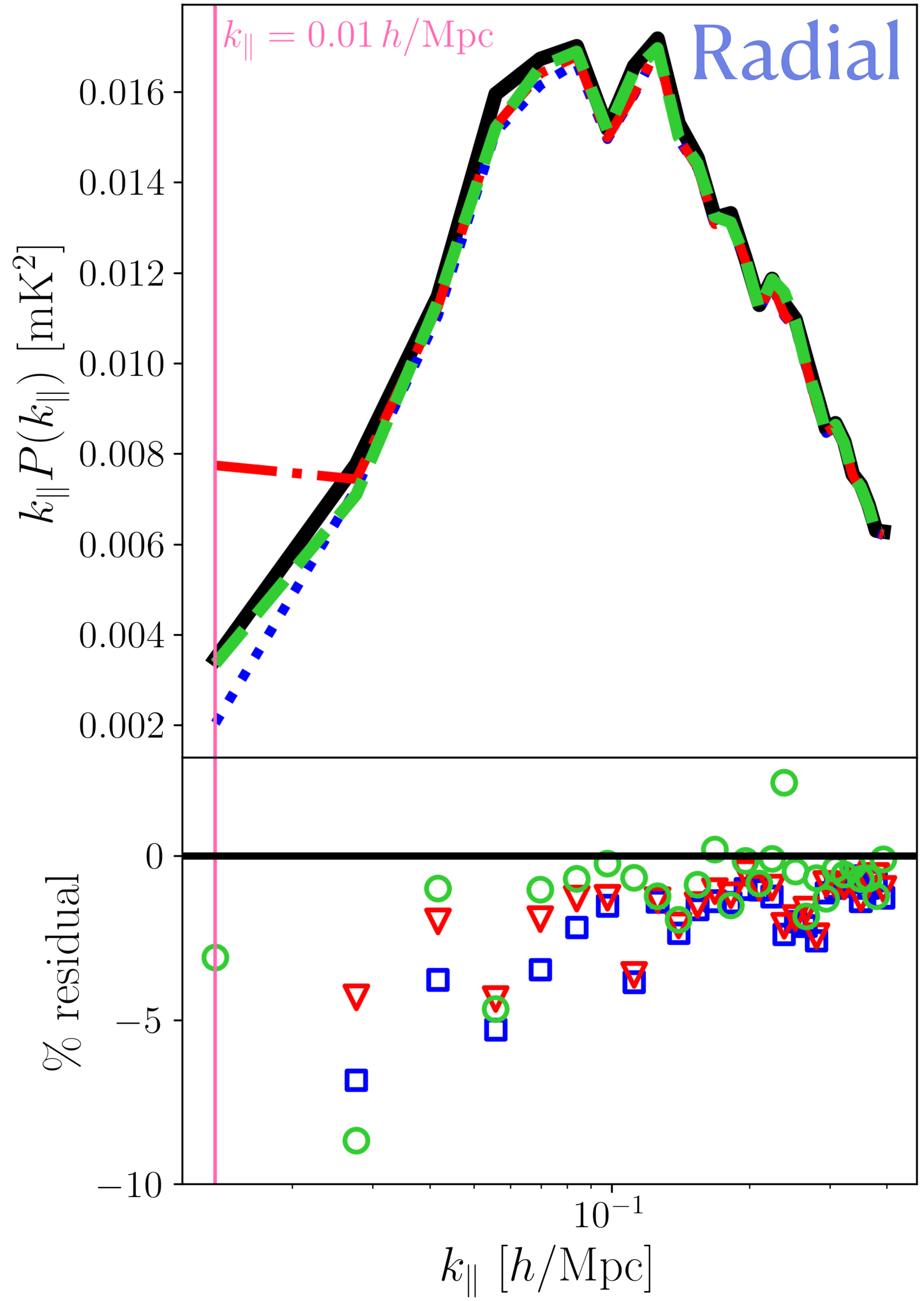


Results



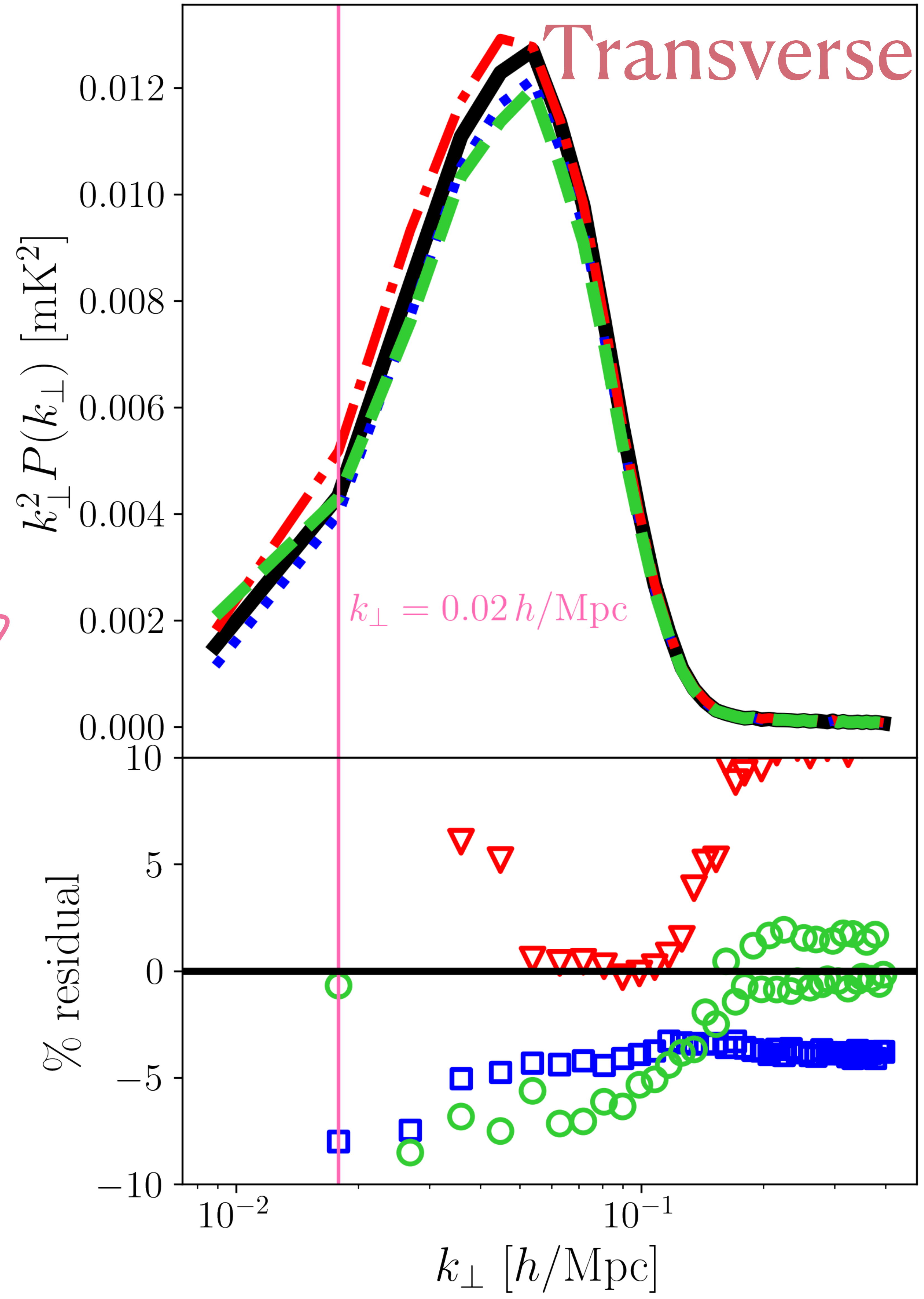
- True HI power spectrum is the black solid line, *what we want to recover*
  - **GPR** results are in **green**
  - PCA results are in **red** ($N_{\rm fg} = 2$) and **blue** ($N_{\rm fg} = 3$)
- Bottom panel shows percentage residual difference from truth

Results



- Very good
- GPR is better than PCA on all scales
- GPR recovers the full range of the radial power spectrum within 10% residual

- Less good
- GPR better on small scales where beam dominates
- GPR cannot recover full range of transverse power spectrum within 10% residual



GPR is better in the radial direction

Key takeaways

- *It is* possible to run **GPR** for foreground removal technique in the case of single-dish, low redshift HI intensity mapping
- **GPR** performs better than PCA on small scales
- **GPR** performs better in the radial direction than in the transverse direction
- For PCA, we constantly needed to change $N_{\rm fg}$ depending on bandwidth size, missing channels, including polarisation, etc.
    - **GPR does not require this fine tuning, it finds the best fitting covariance model given the data**
- Our code is available at github.com/paulassoares/gpr4im