Optimal Techniques for analyzing Line Intensity Mapping data

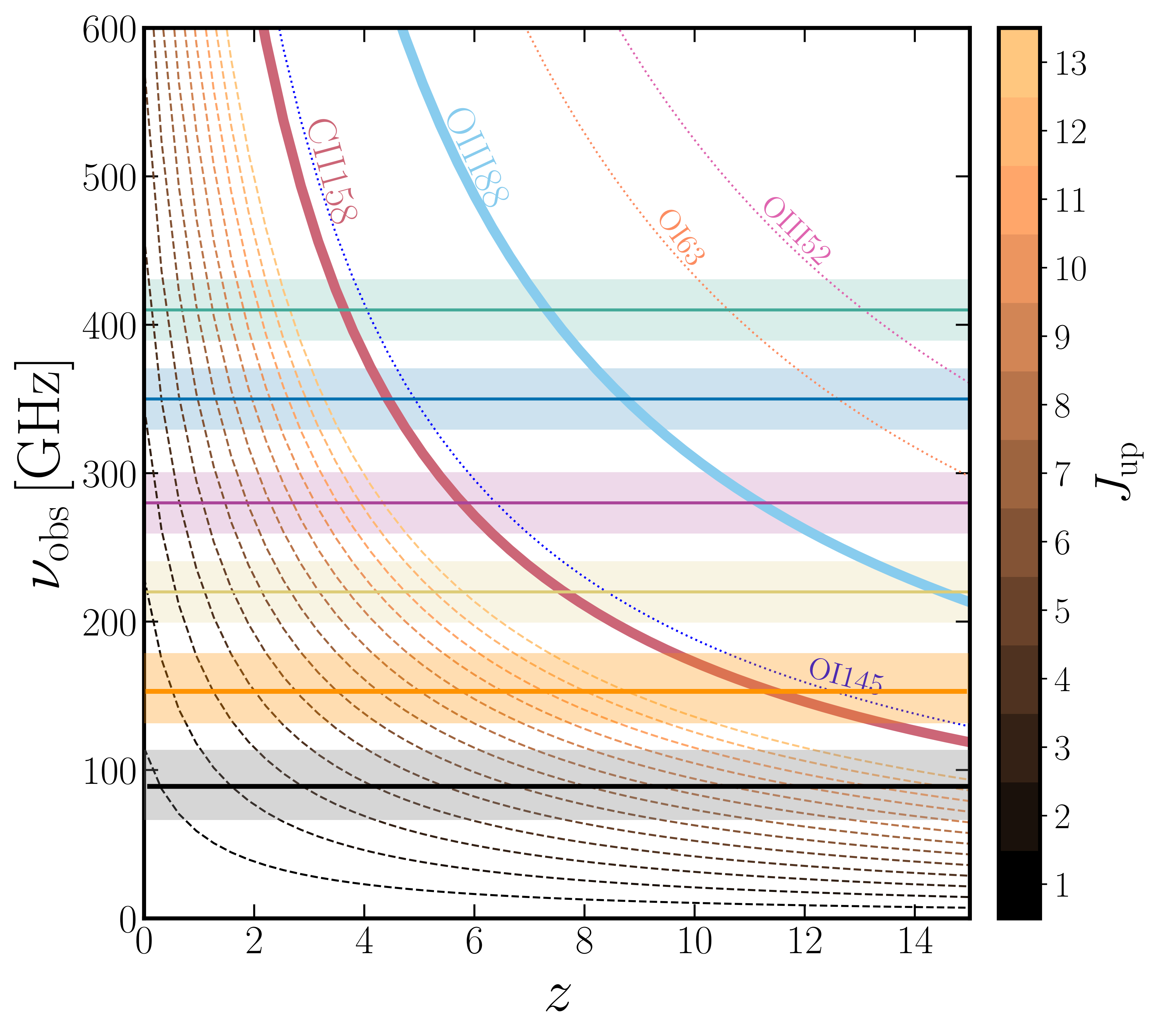
Anirban Roy

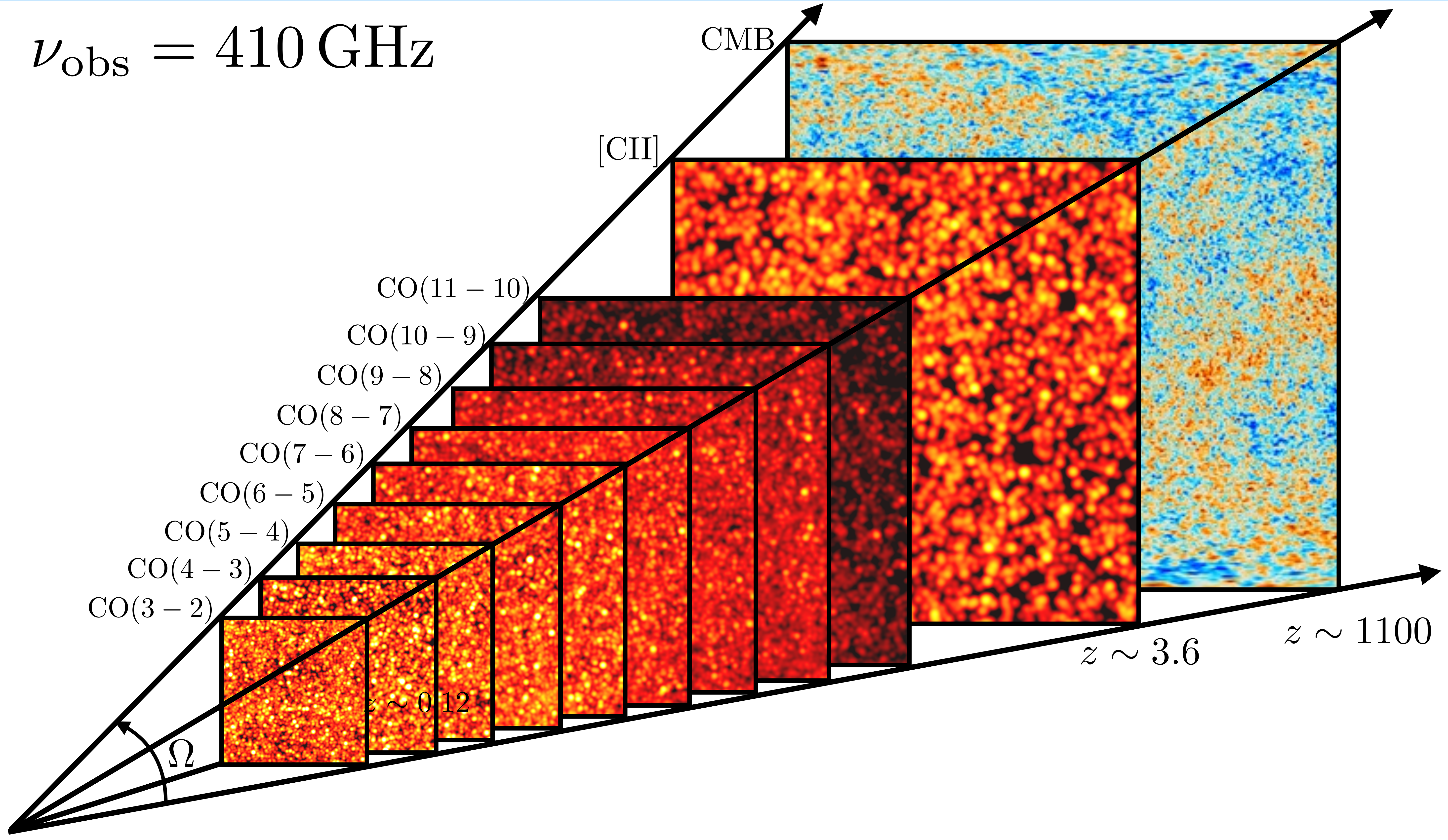
New York University

In collaboration with

Anthony Pullen (NYU), Rachel Somerville (Flatiron), Patrick Breysse (SMU), Nick Battaglia (Cornell)

Tomography with LIM

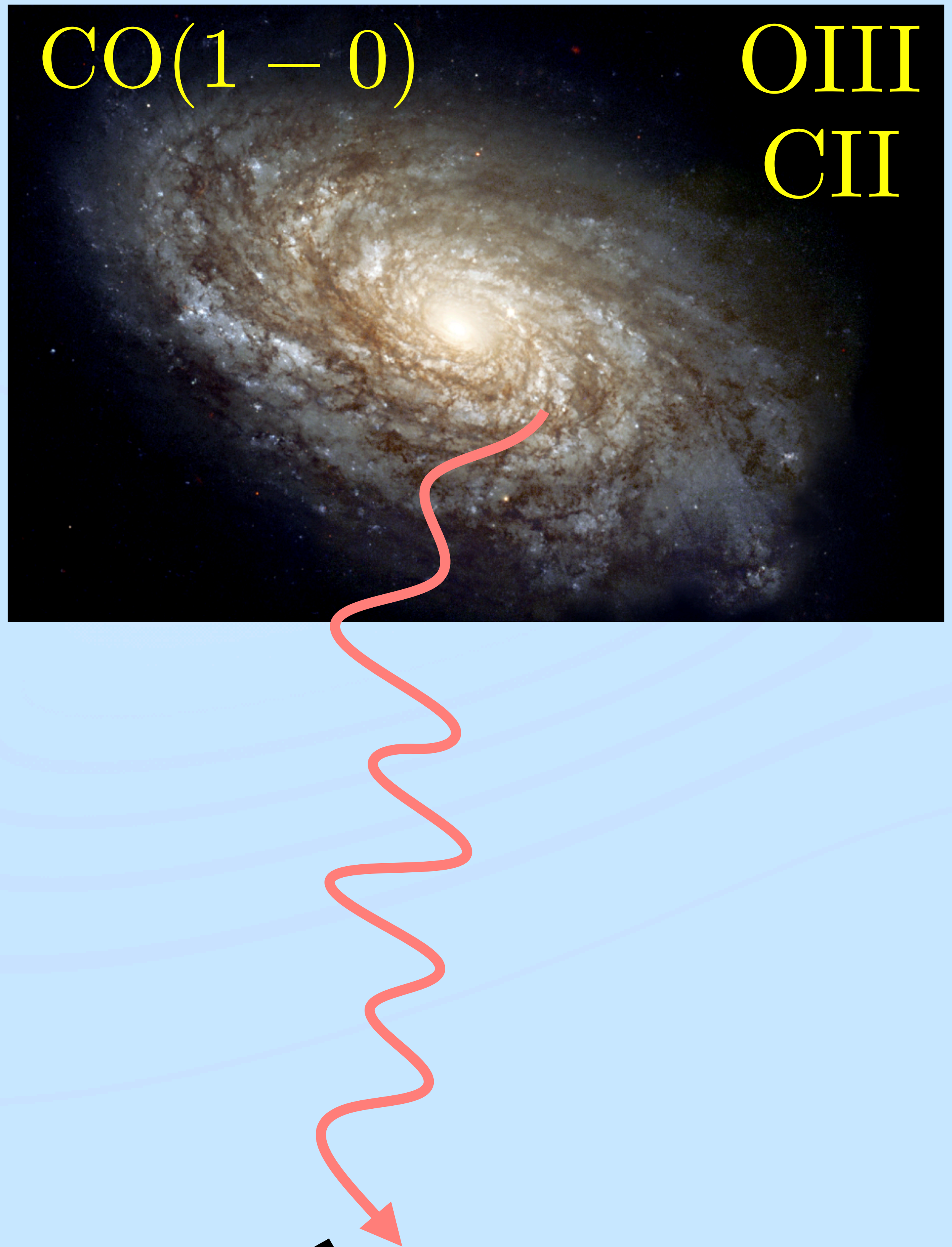




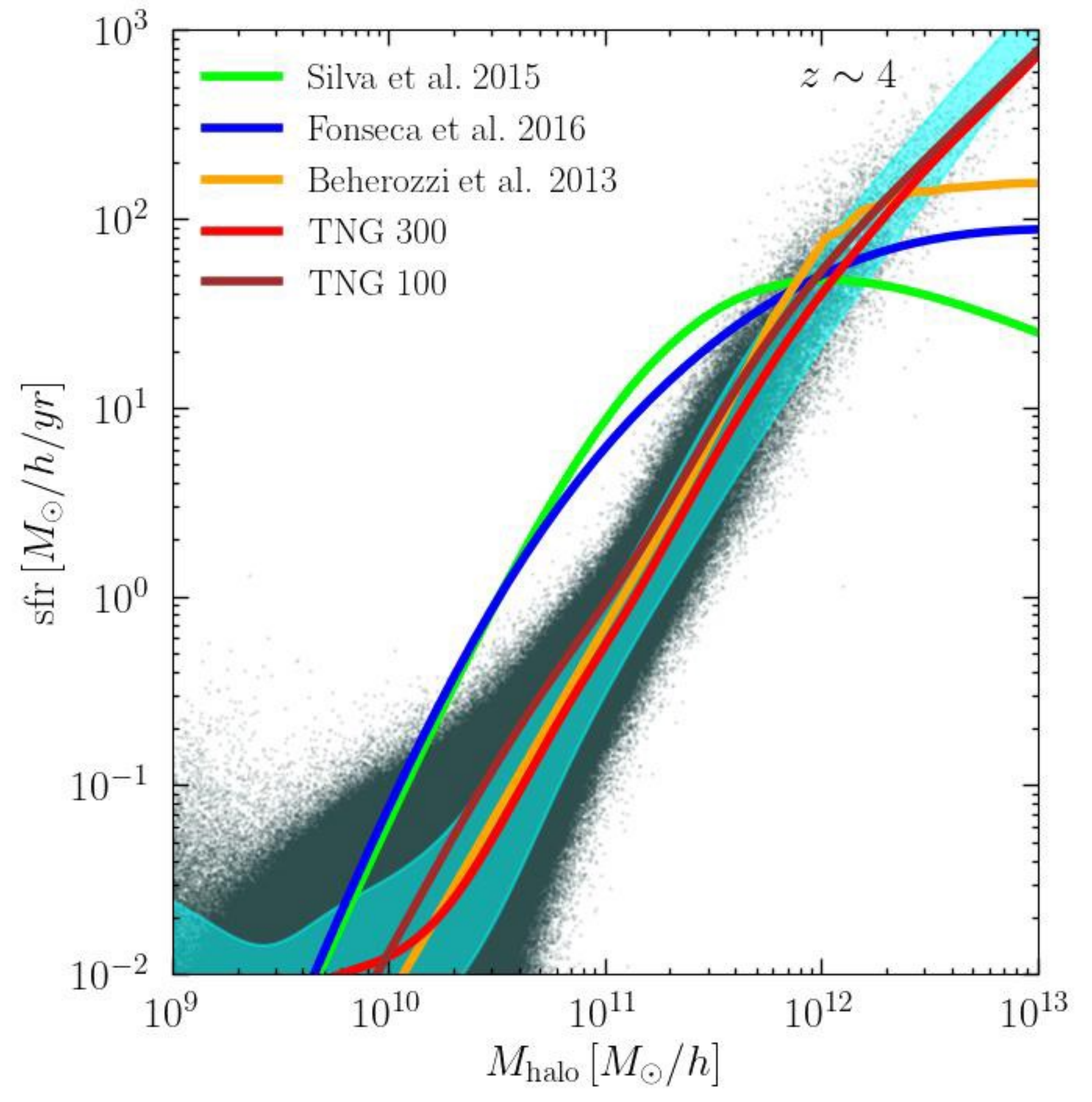
Multiple frequency channels ⟶ **Observation of multiple lines** ⟶ **Tomography of the Universe**

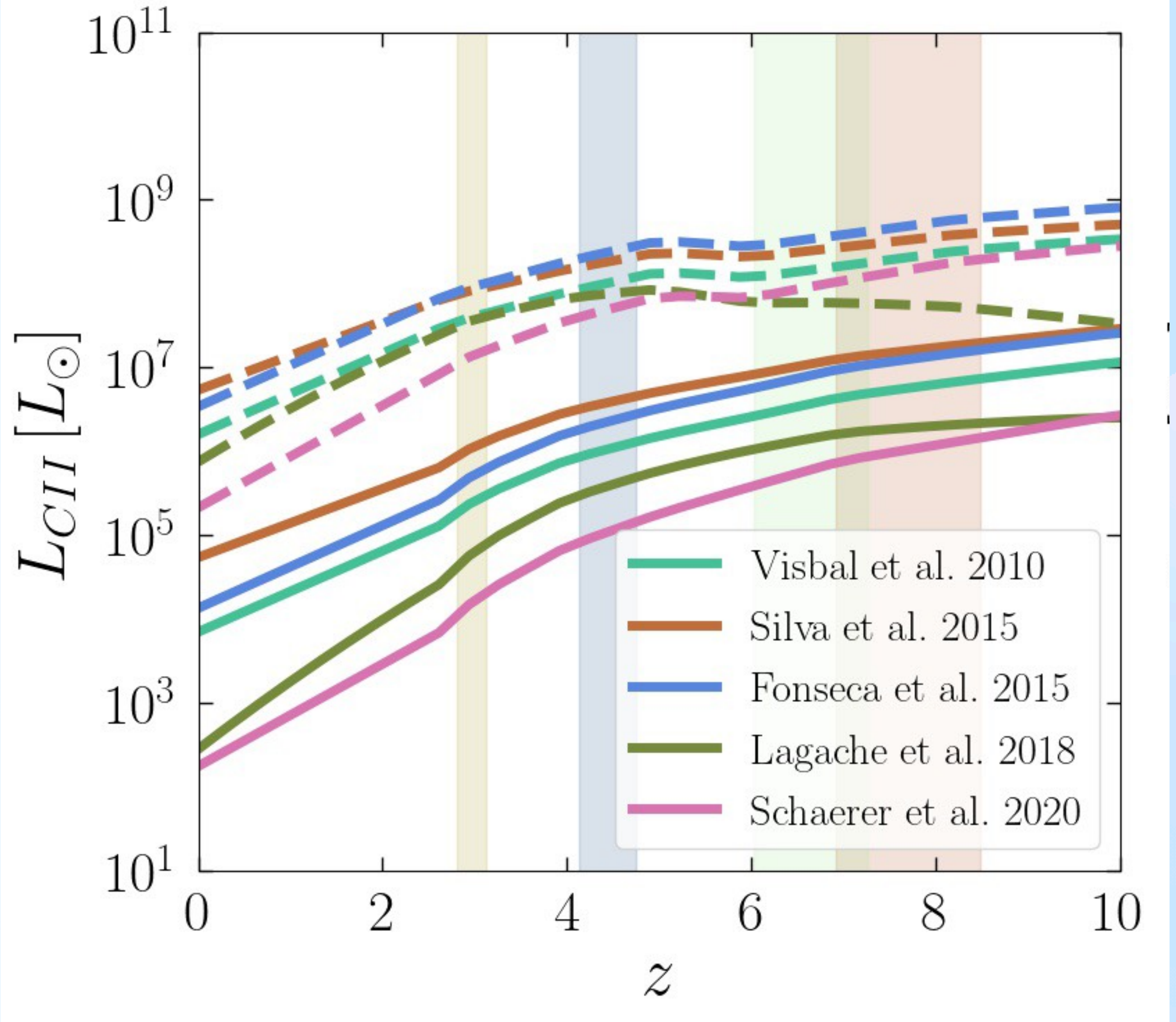
Roy+ (2024, 2023)

General steps for LIM modeling



$$\nu_{\rm obs} = \frac{\nu_{\rm em}}{1+z}$$



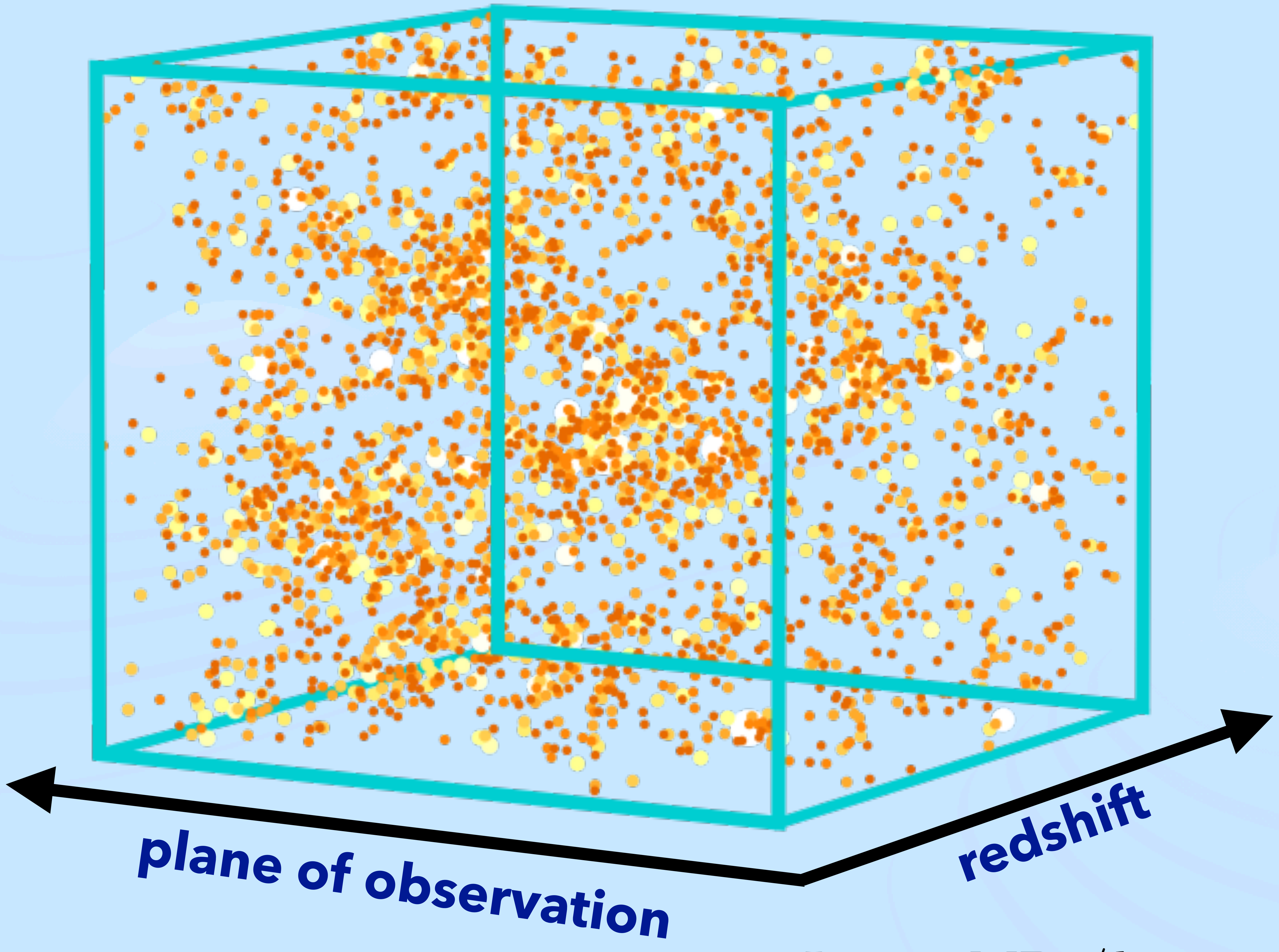


$$\log L_{\rm line} = a_{\rm off} + b_{\rm off} \log\left(\rm SFR\right)$$

Roy+ (2023)

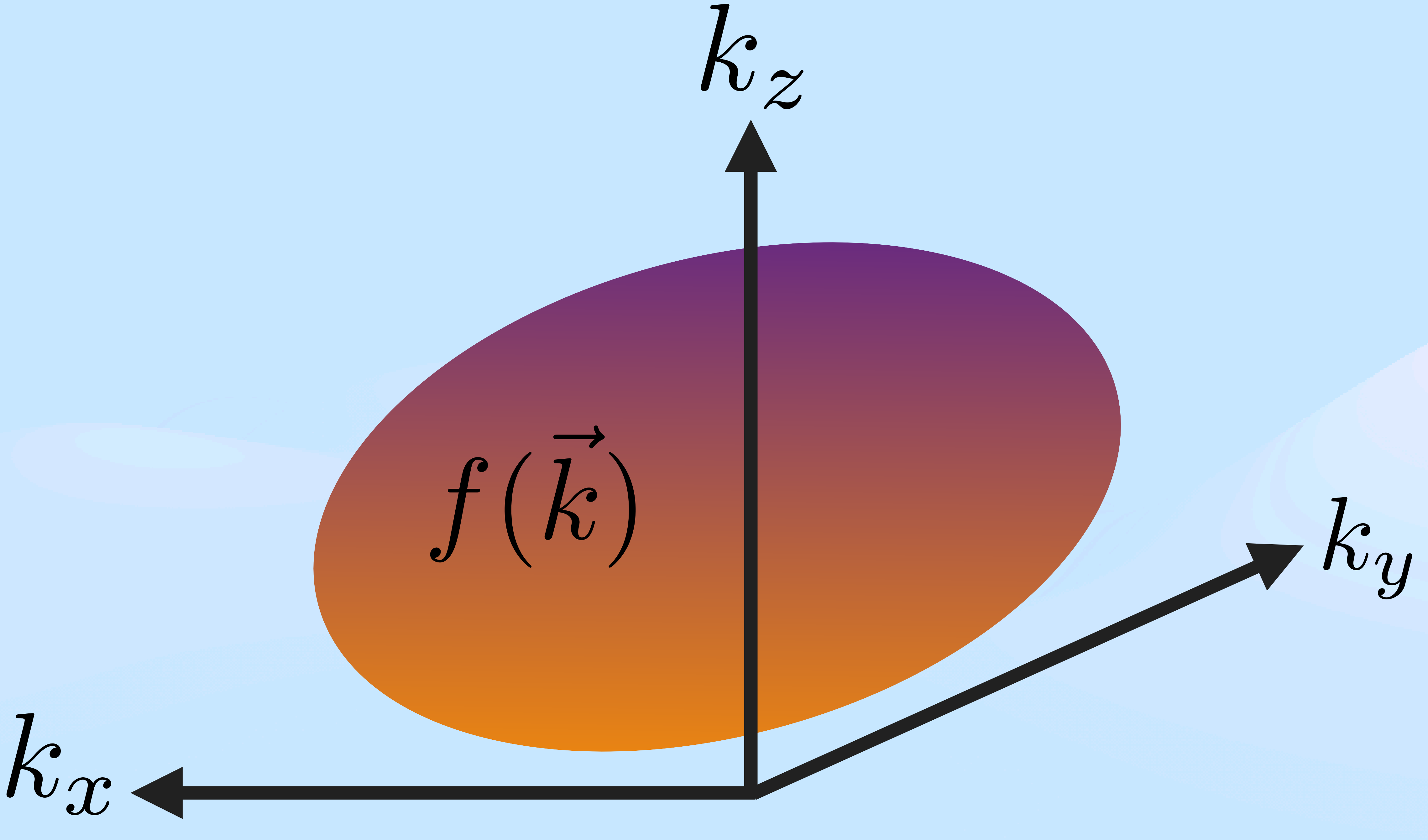
LIM data analysis

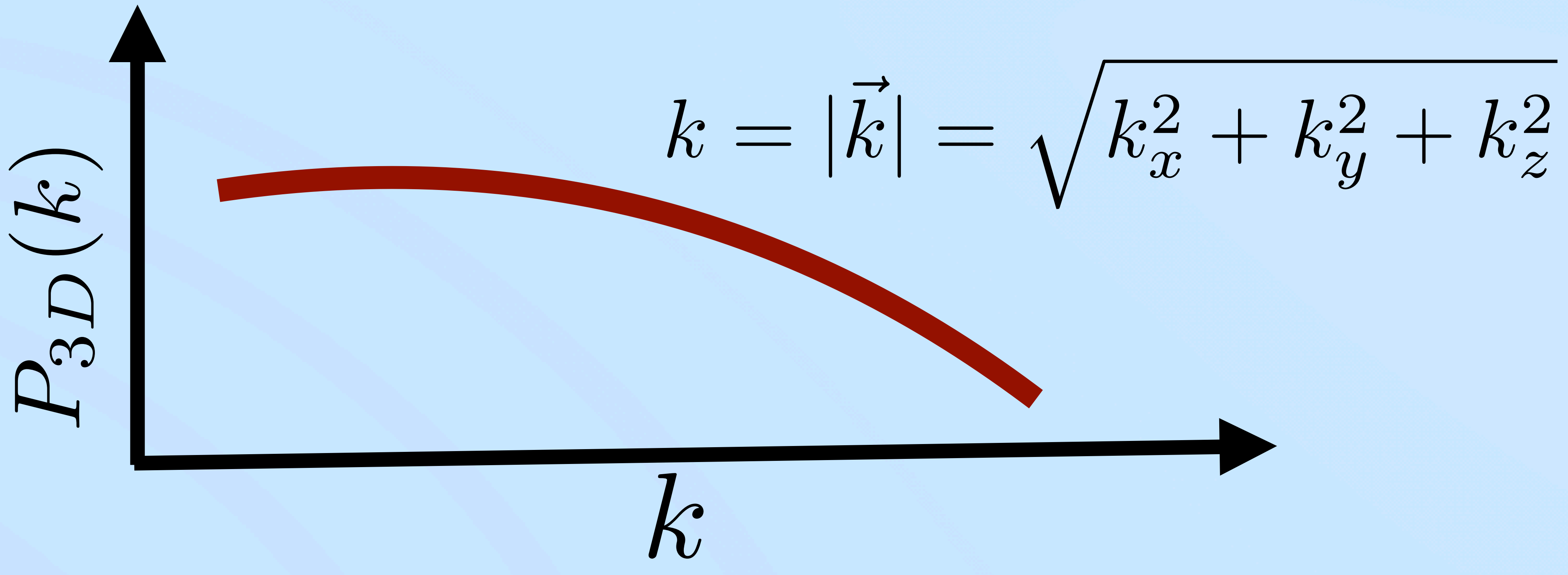
Simulations



$$L = 80\,\mathrm{MPc}/h$$
$$\nu_{\rm obs} = 280\,\mathrm{GHz}$$
$$z_{\rm obs} = 5.8$$
$$M_{\rm min} = 10^{11}\,M_\odot/h$$

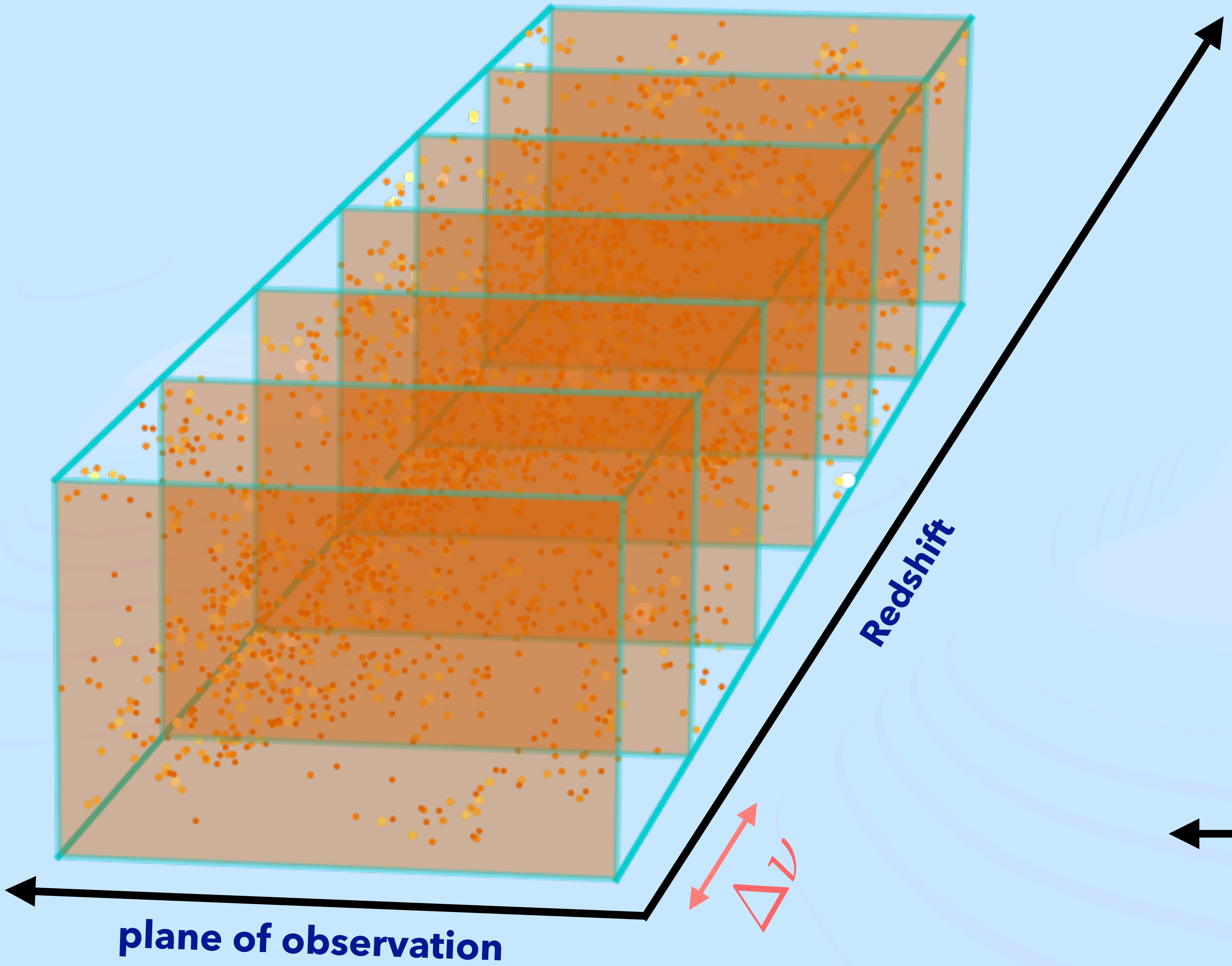
Summary Statistics

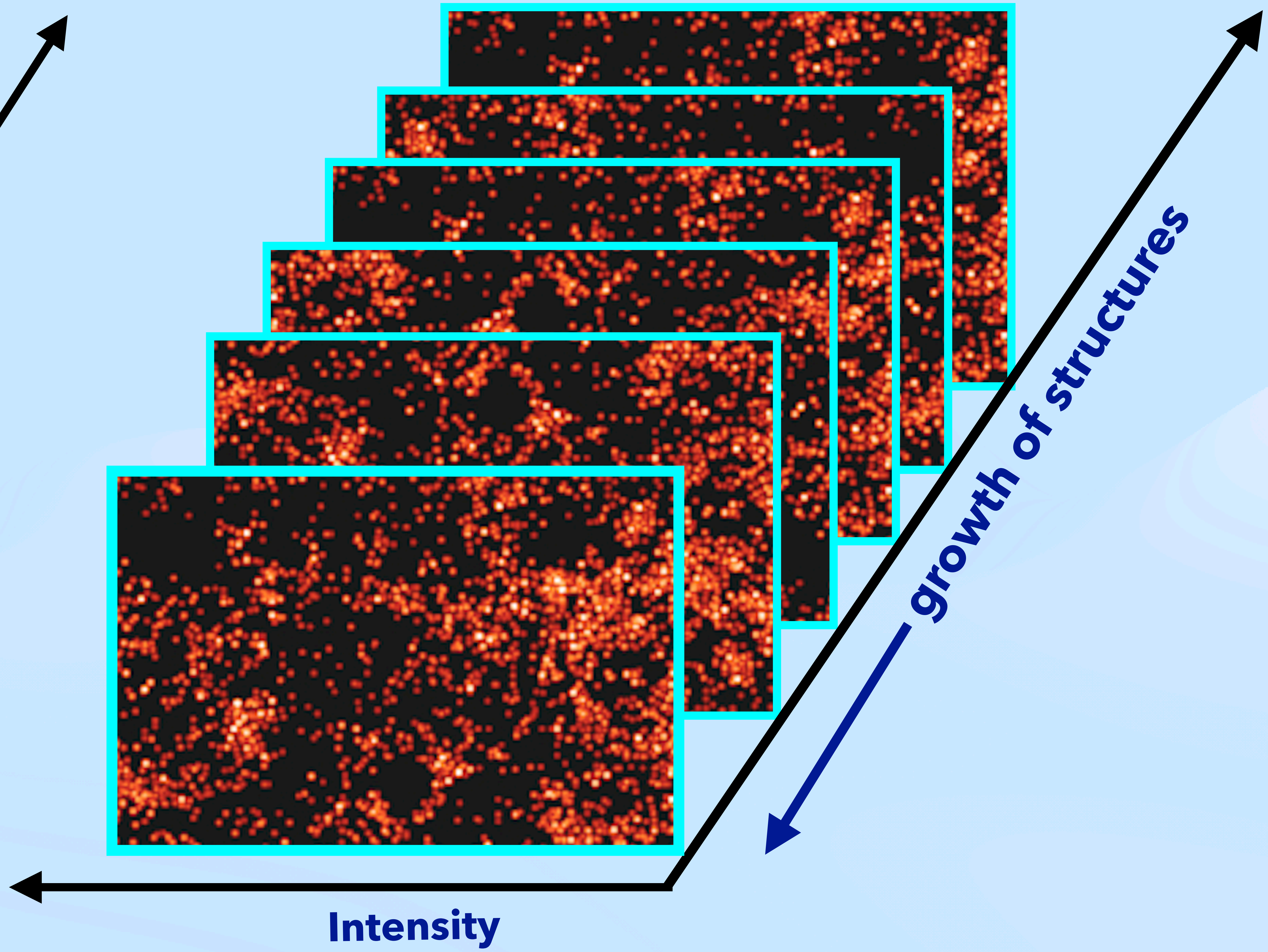




5

LIM data analysis

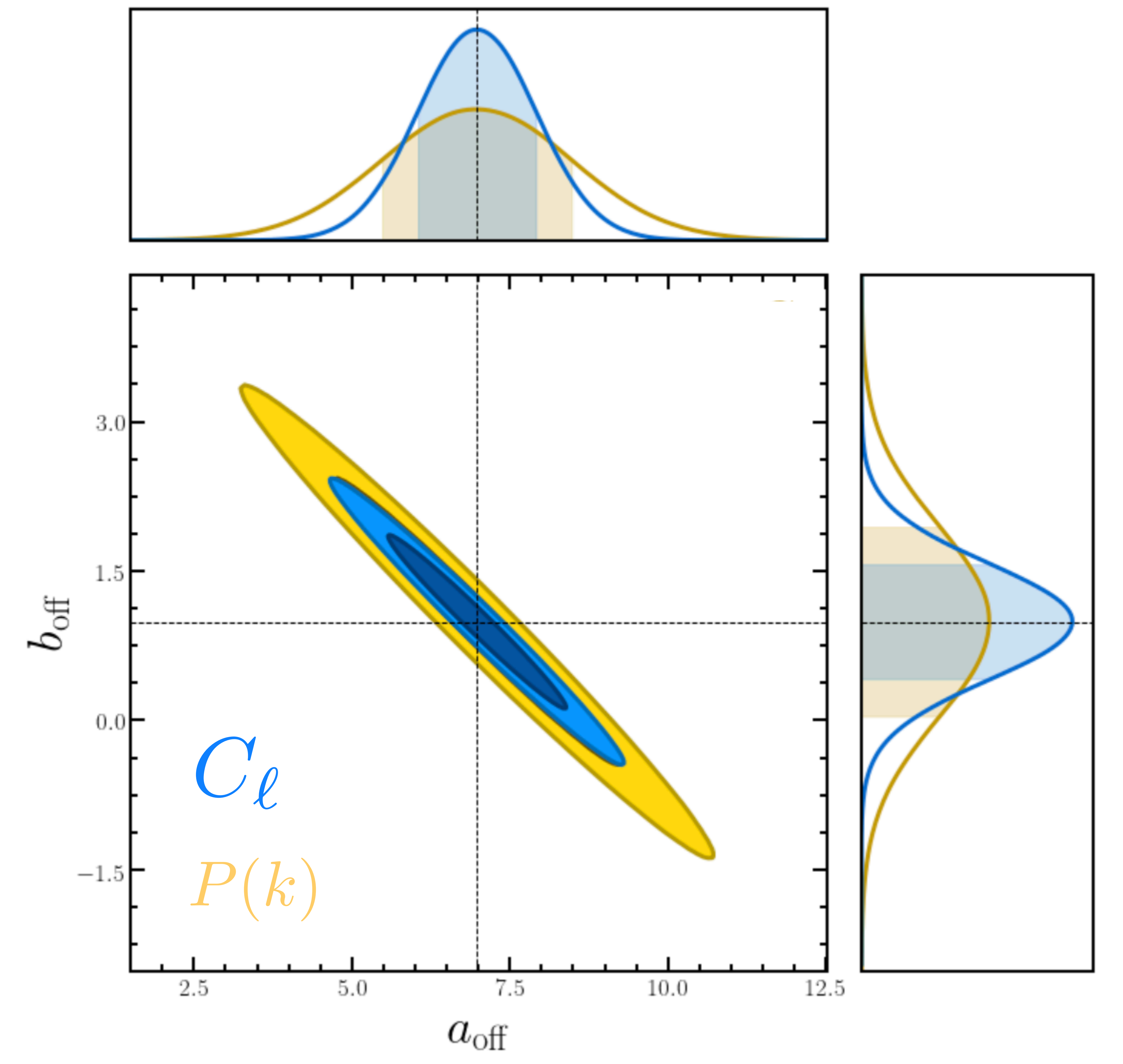




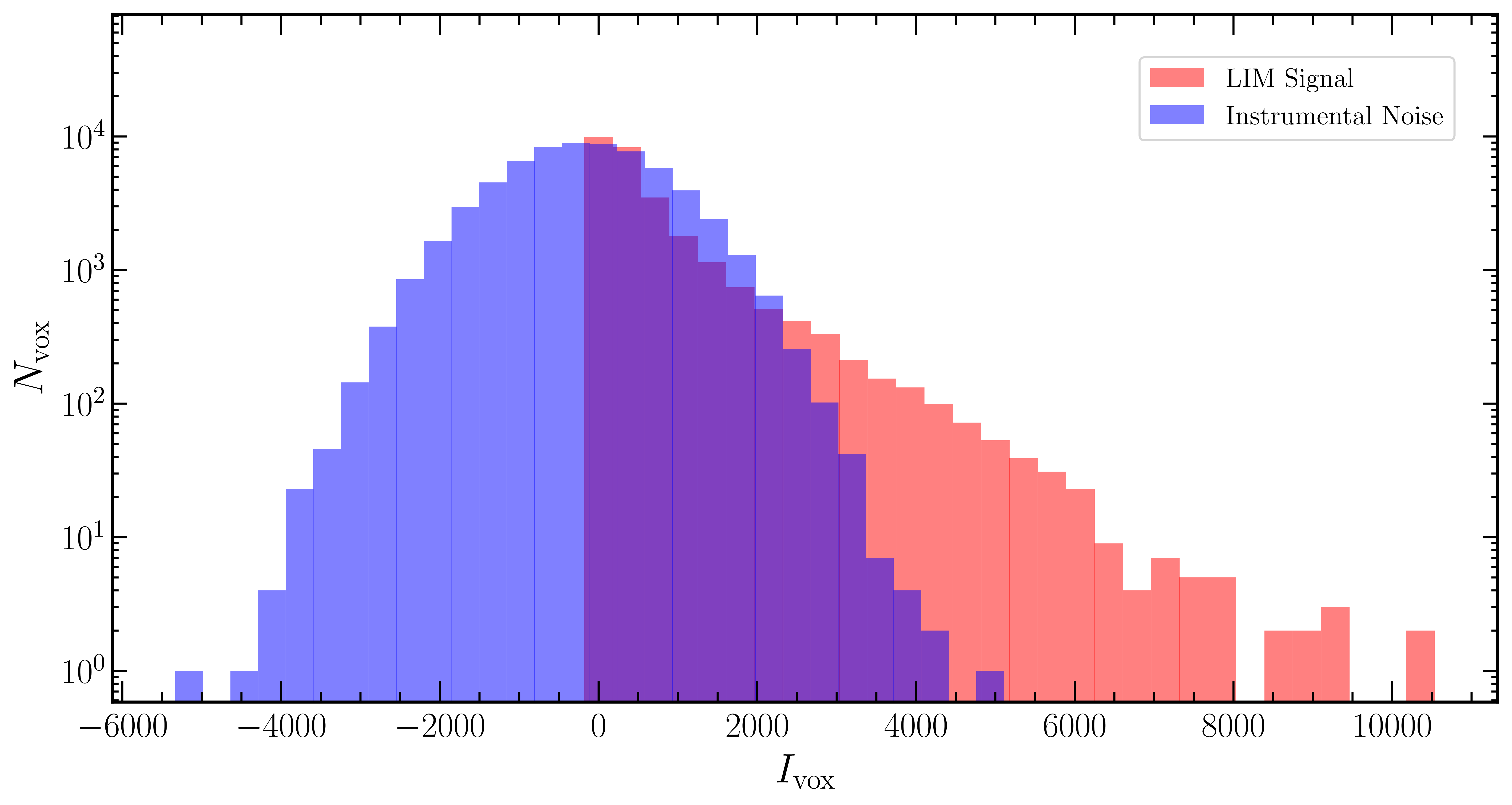
$$L_x = L_y = 80\,\mathrm{MPc}/h$$
$$L_z = 320\,\mathrm{MPc}/h$$

How to analyze the LIM dataset to extract maximum information encoded in it?

LIM estimators



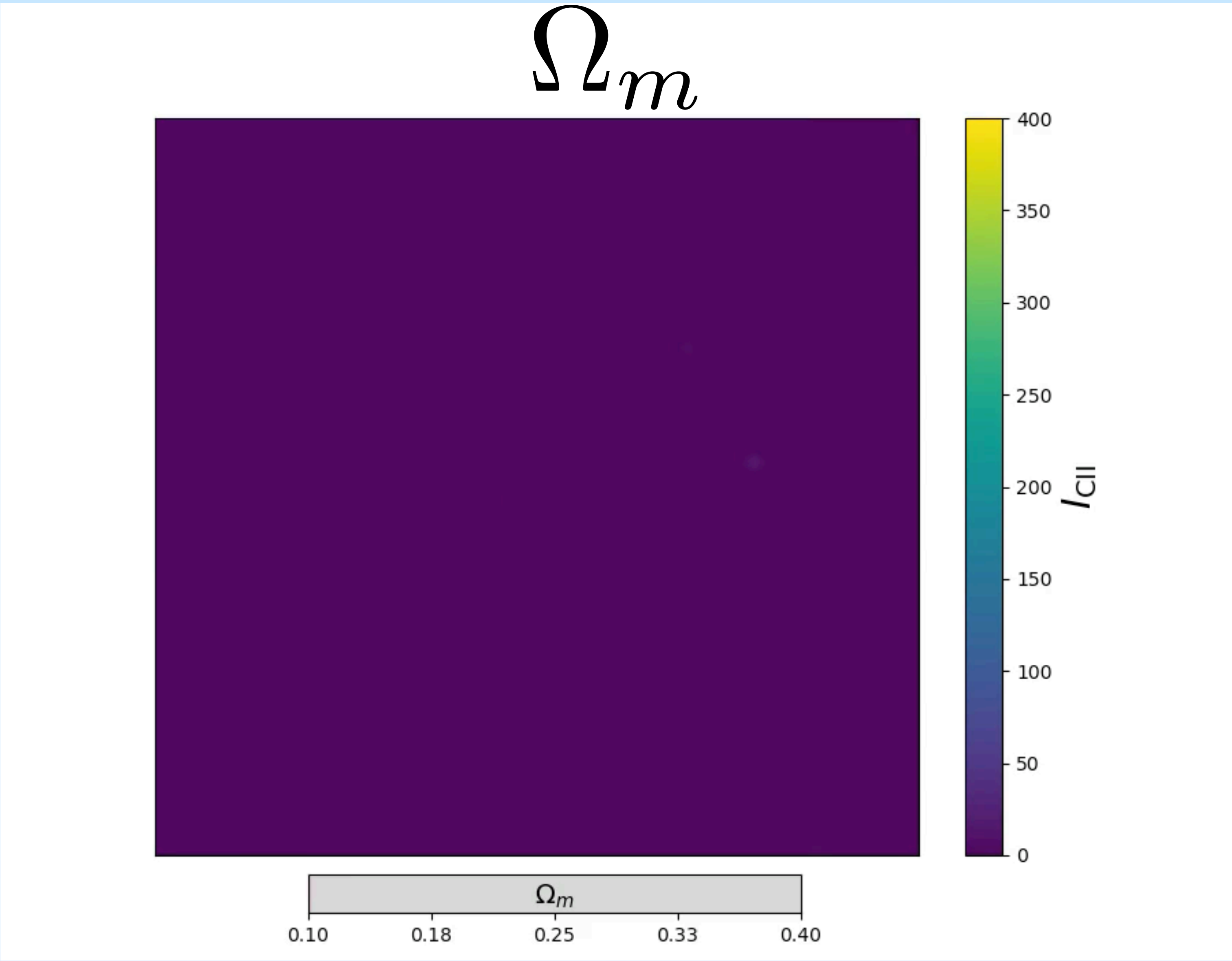
Intensity Distribution Function

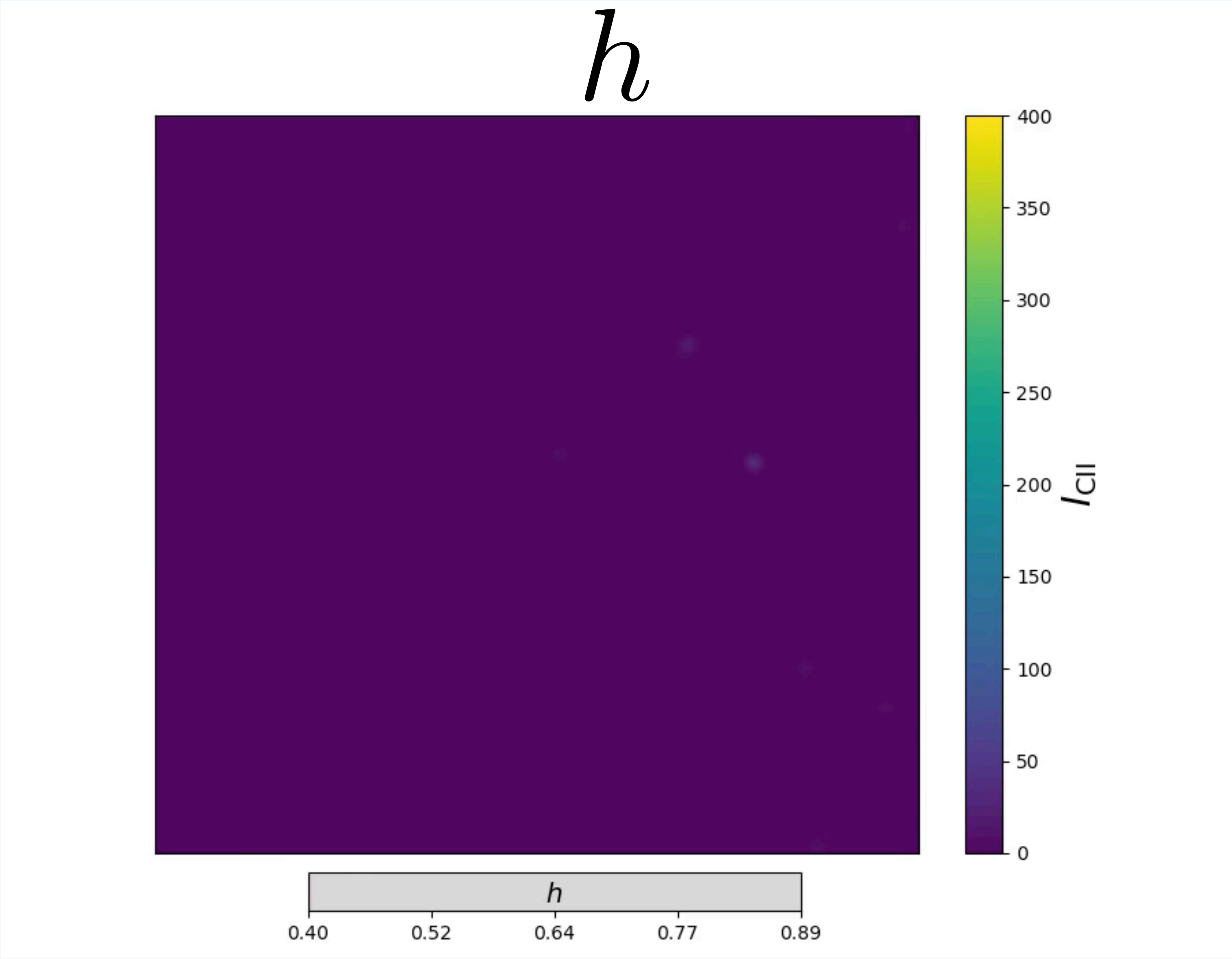


- The long tail is due to the signal, where the SNR becomes larger.
- Several tests are ongoing to determine if we can improve the parameter constraints by using more maps through data augmentation.

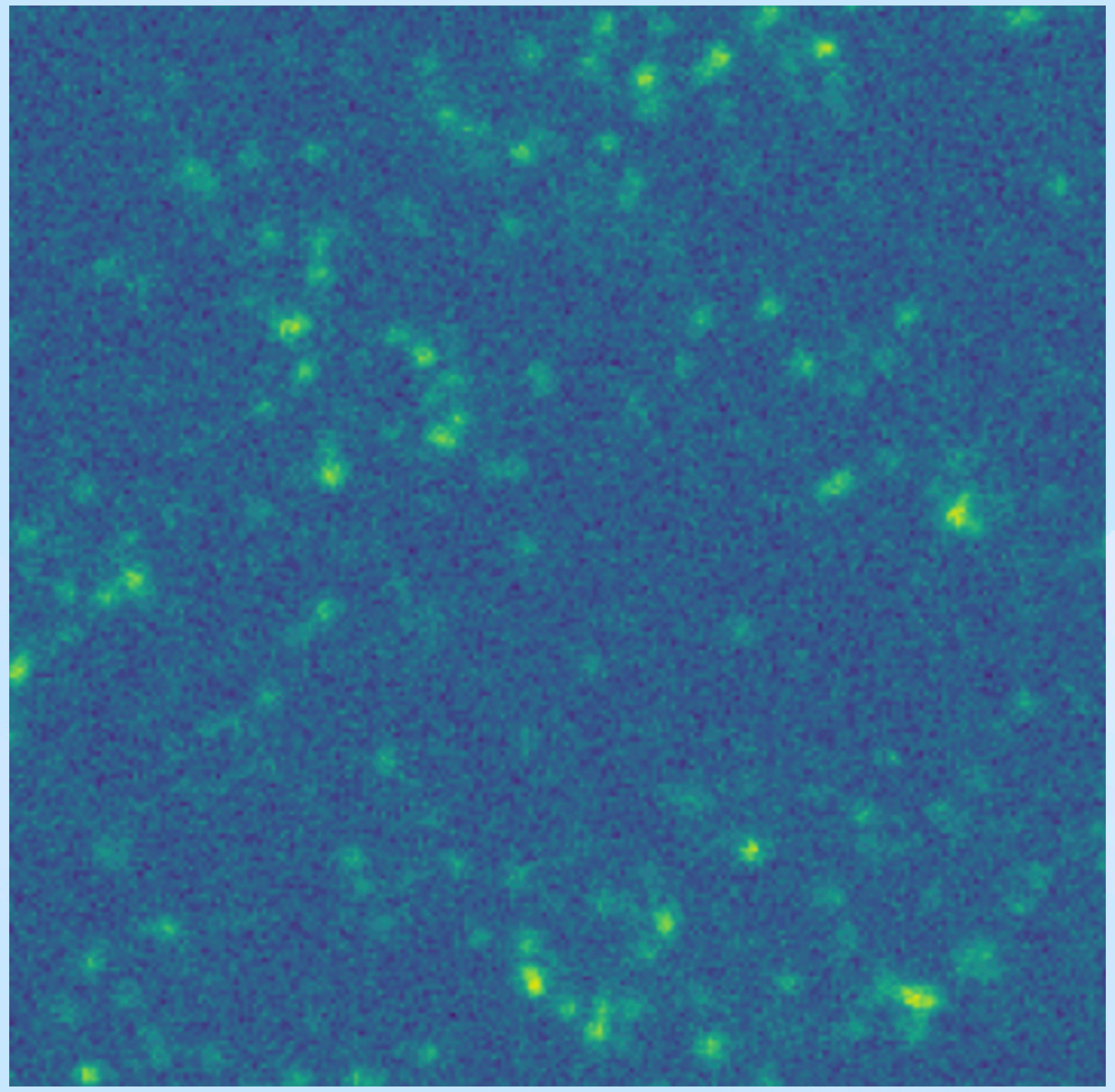
Astro vs Cosmo parameters

Mock signal maps





Observation



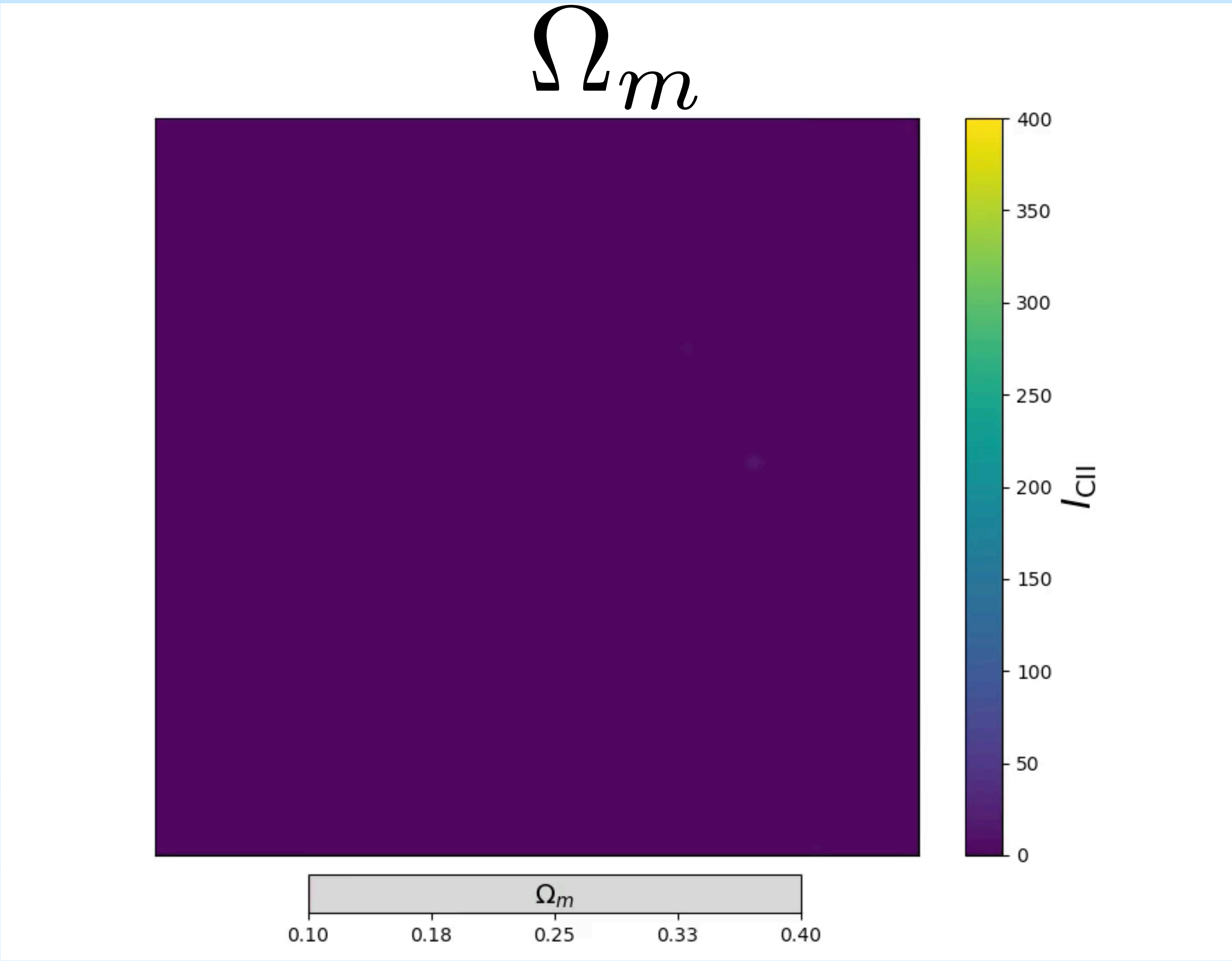
Training over summary statistics or field level analysis

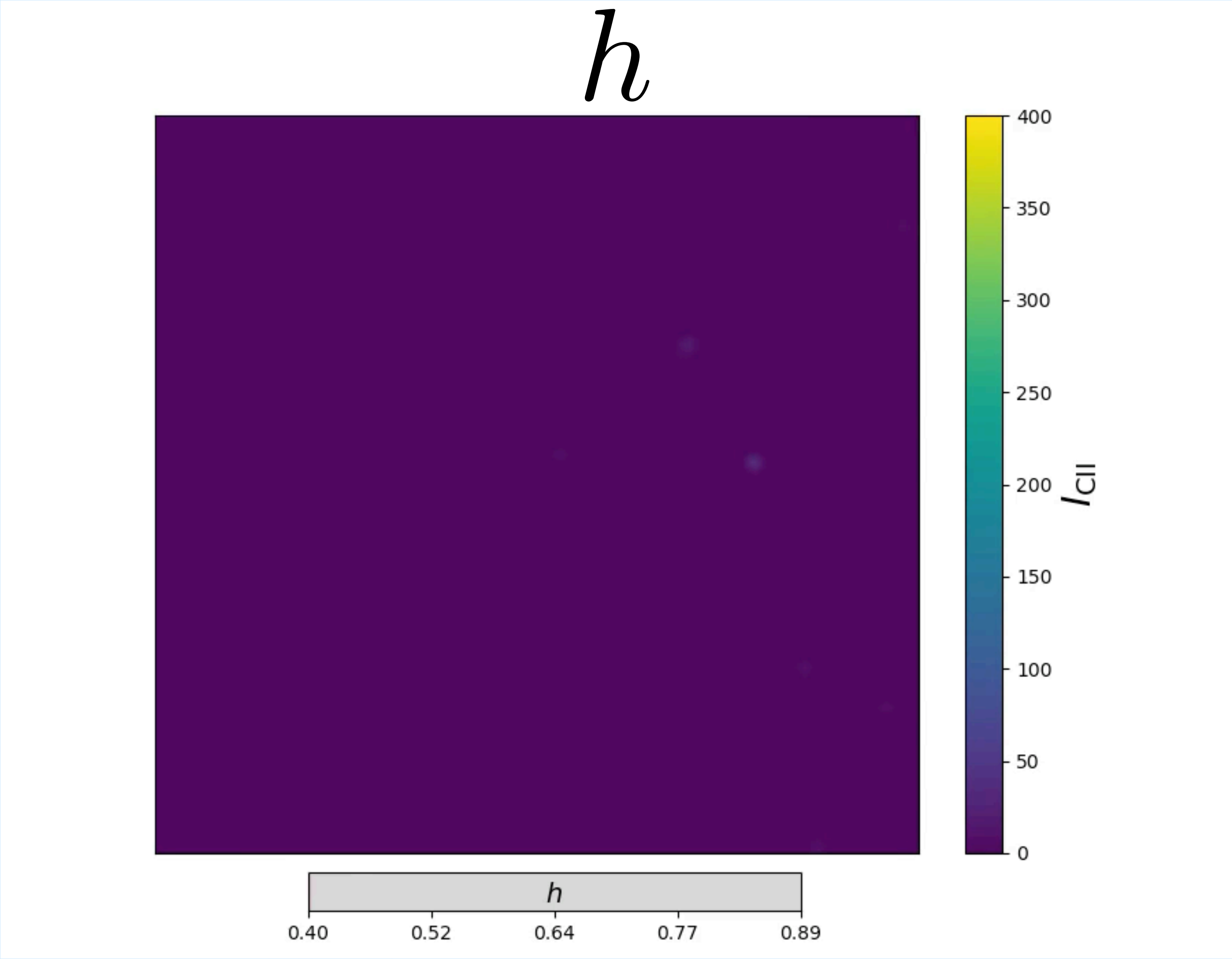
Parameters (?)

We produced ~ 5000 simulations varying h, Ω_{m}, a_{off} and b_{off}.

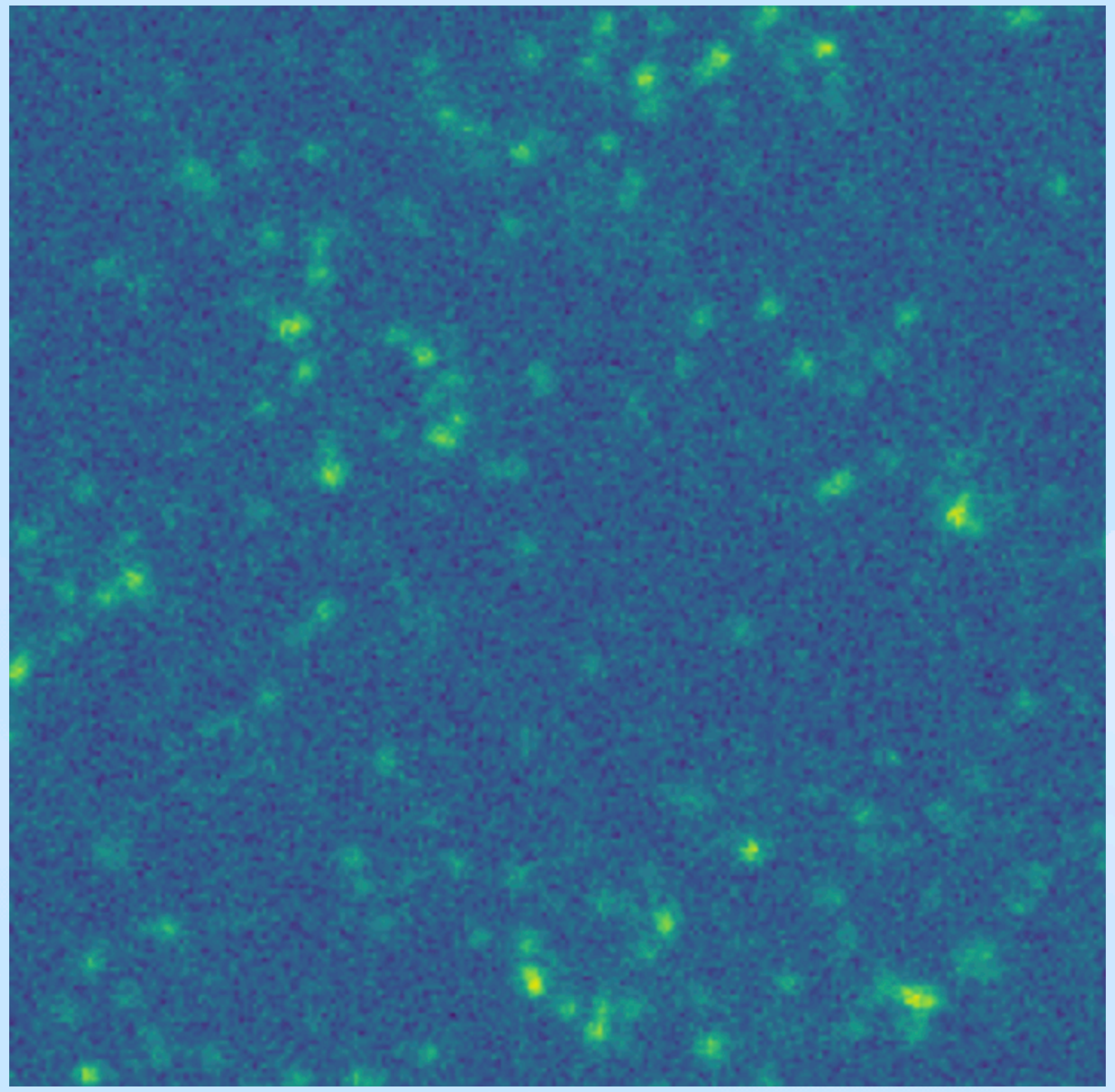
Astro vs Cosmo parameters

Mock signal maps





Observation



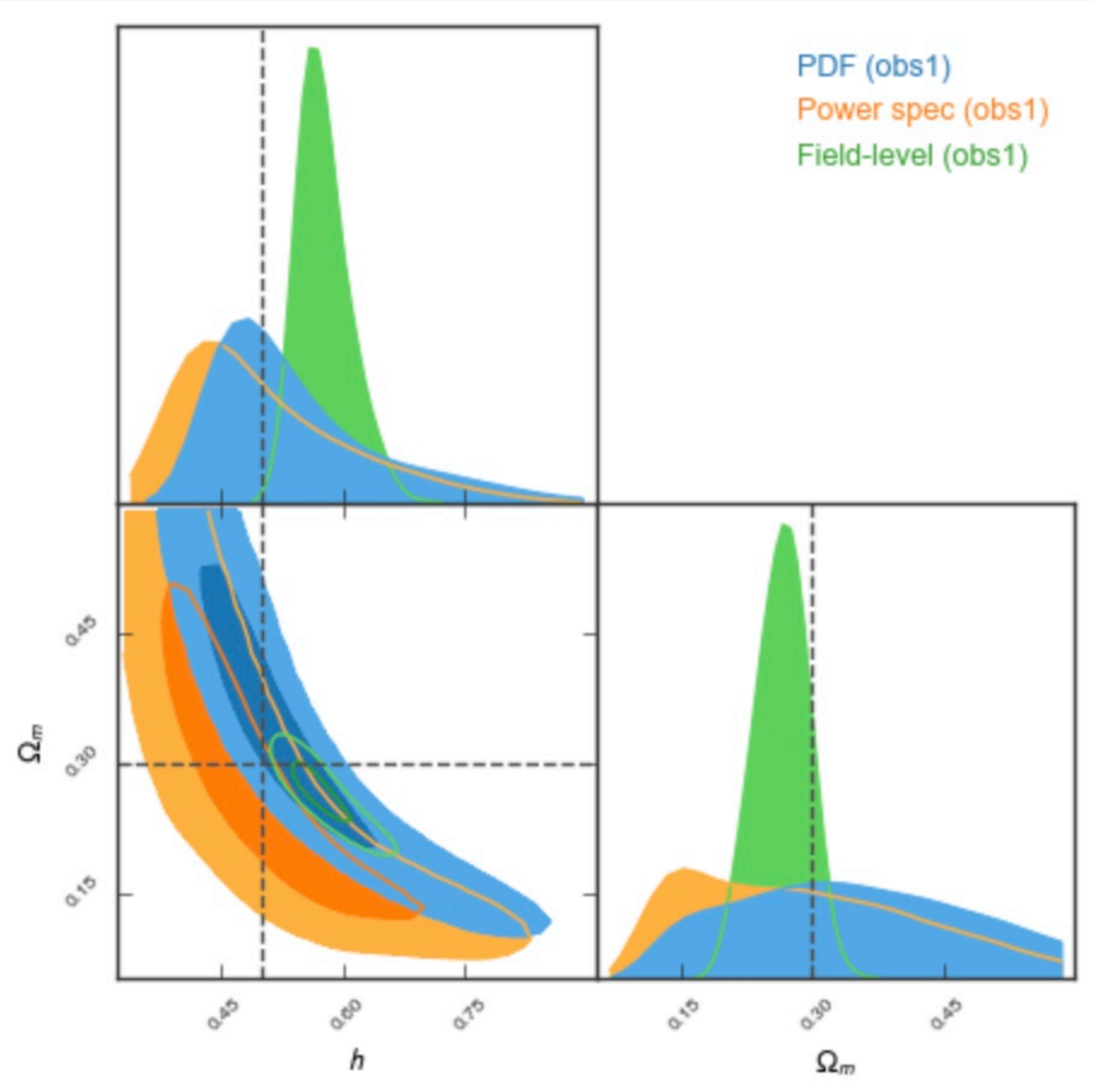
Training over summary statistics or field level analysis

Parameters (?)

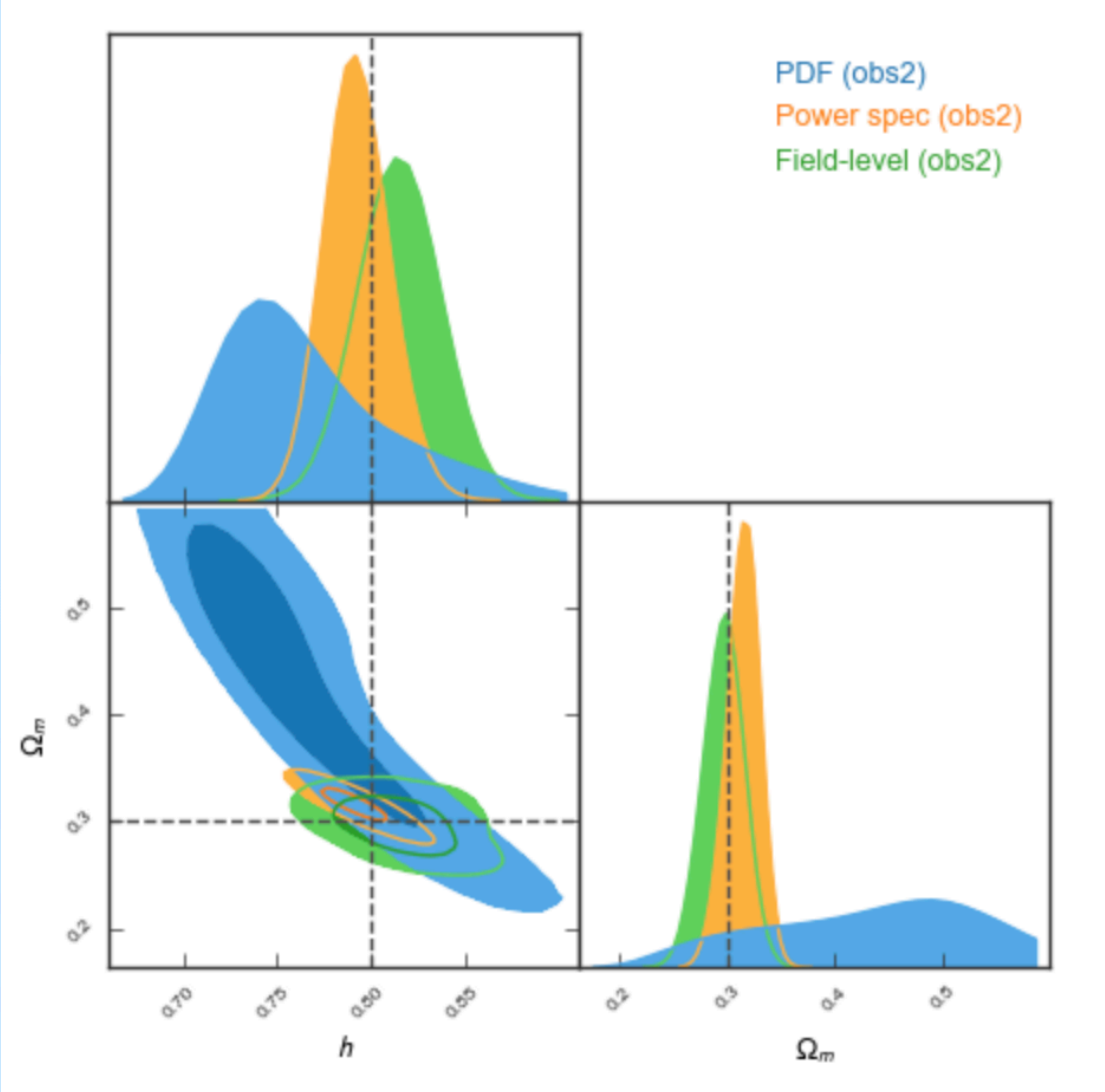
We produced ~ 5000 simulations varying h, $\Omega_{\rm m}$, $a_{\rm off}$ and $b_{\rm off}$.

Field-level analysis

Preliminary

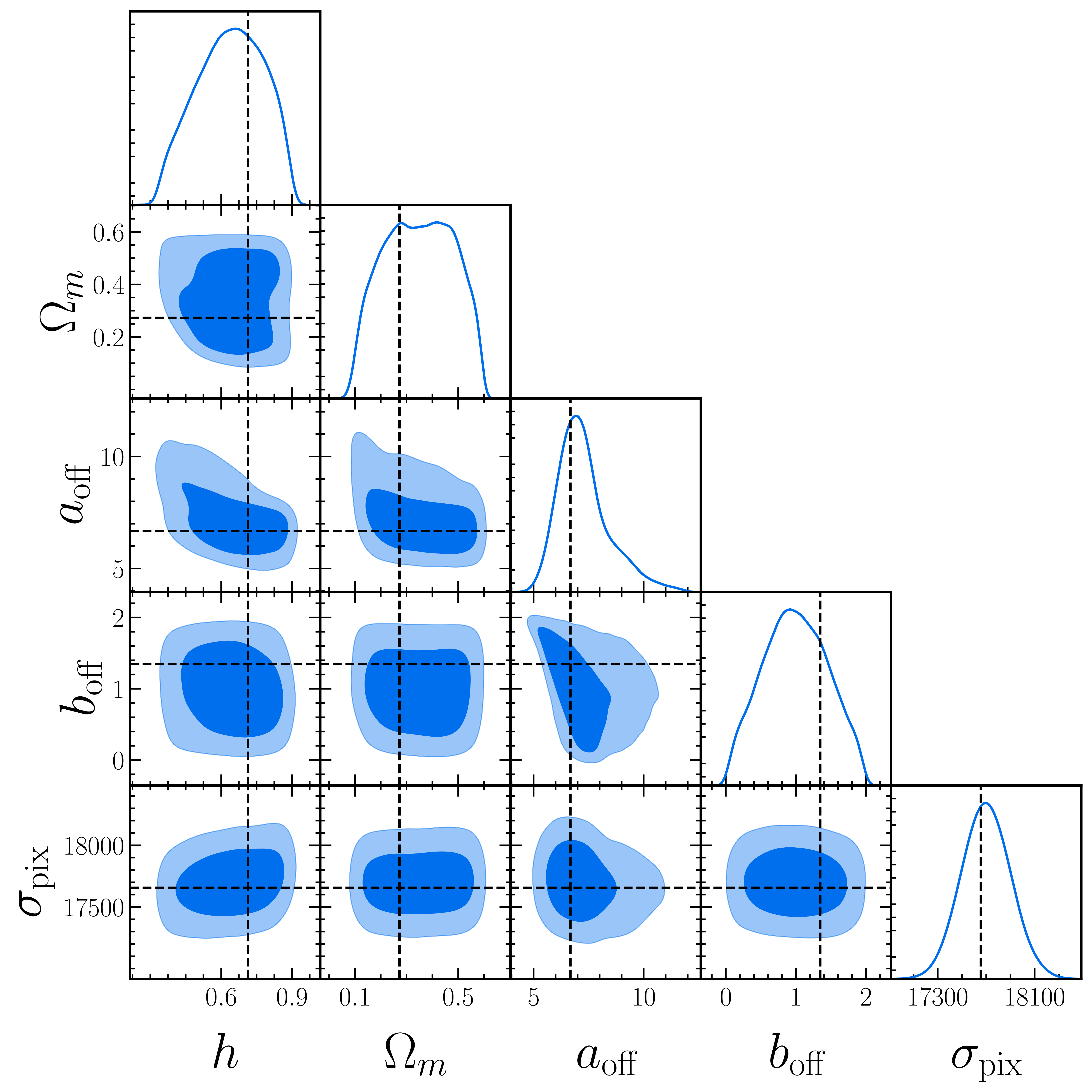


Low SNR case

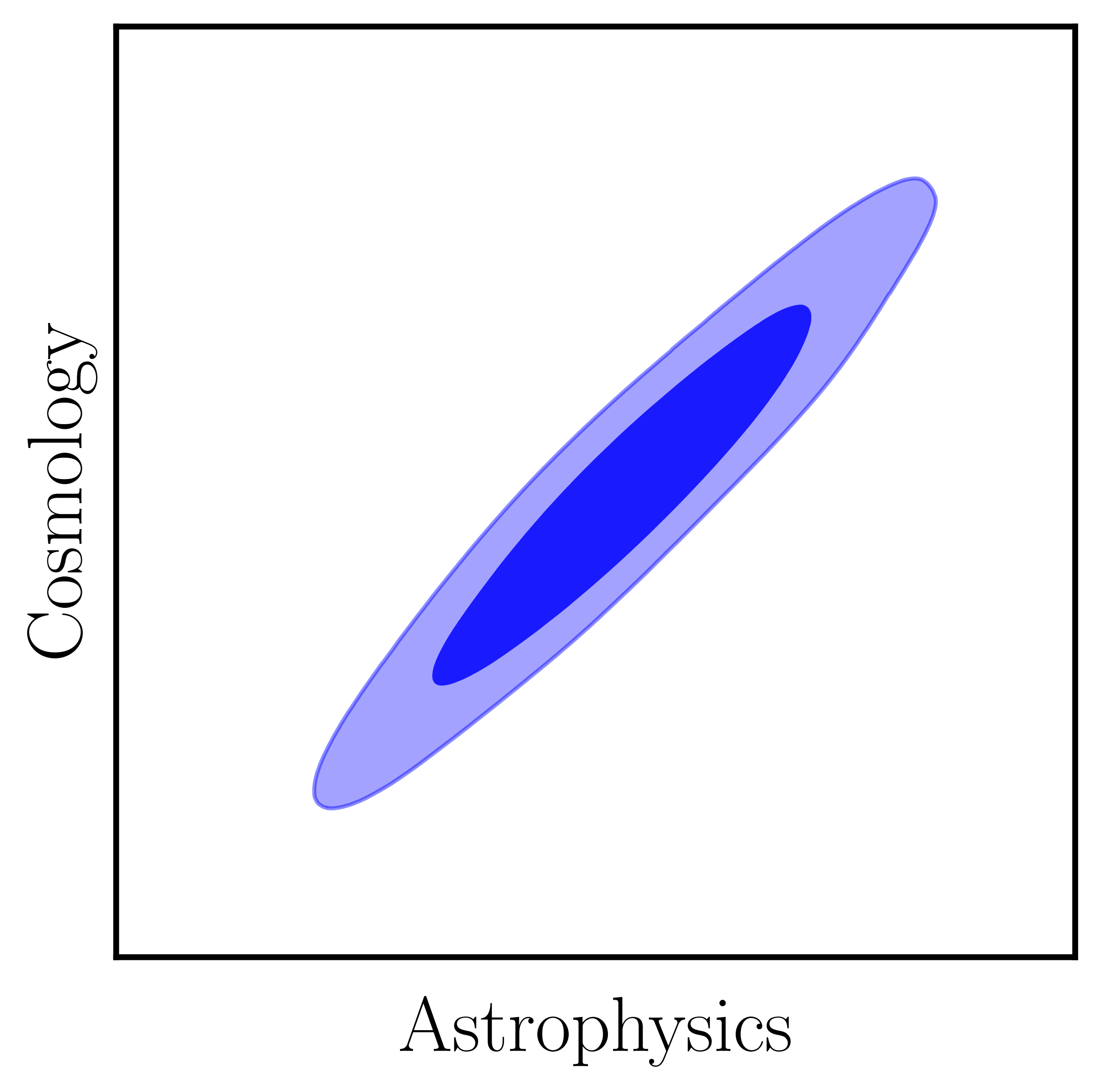


High SNR case

Parameter inference

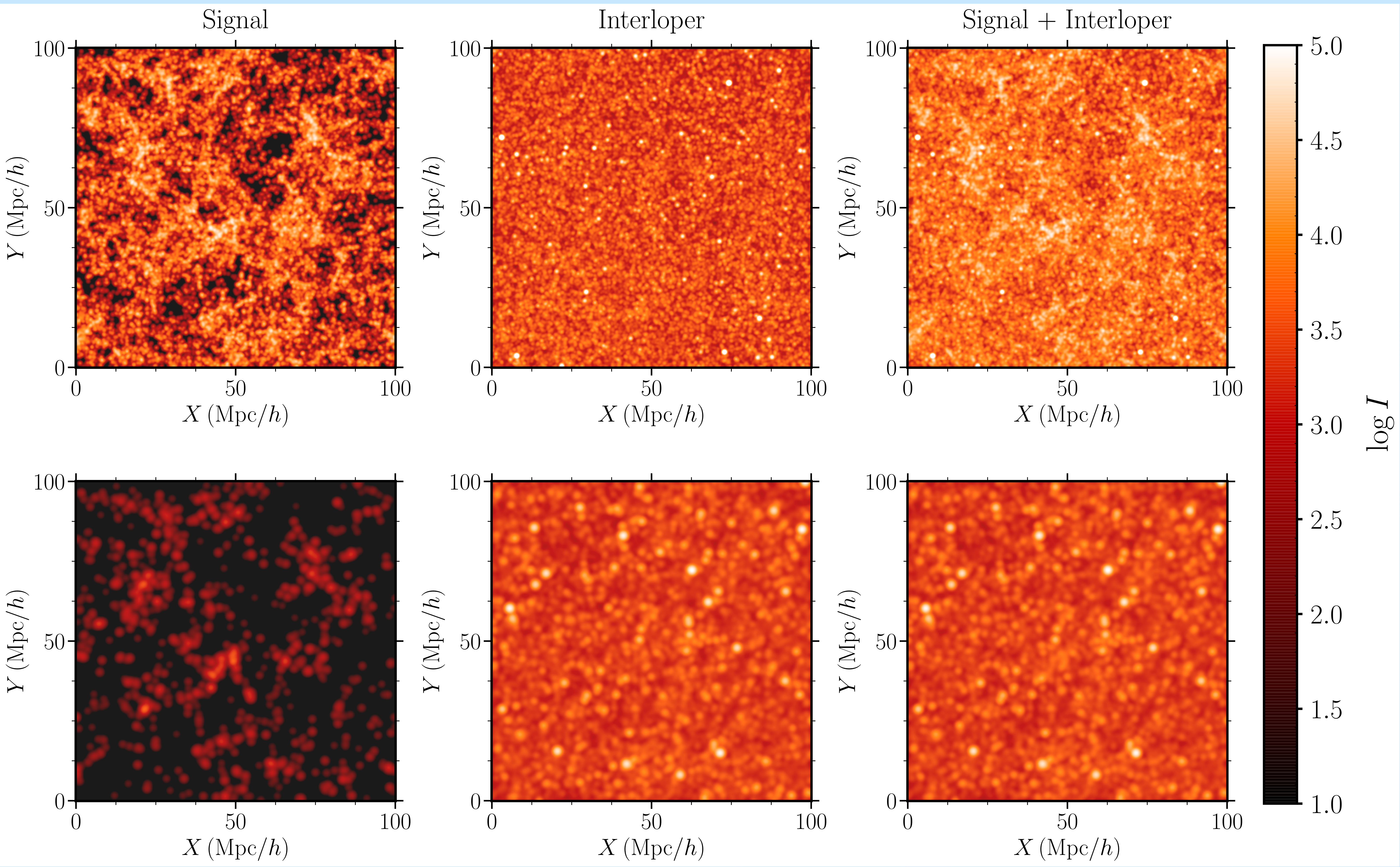


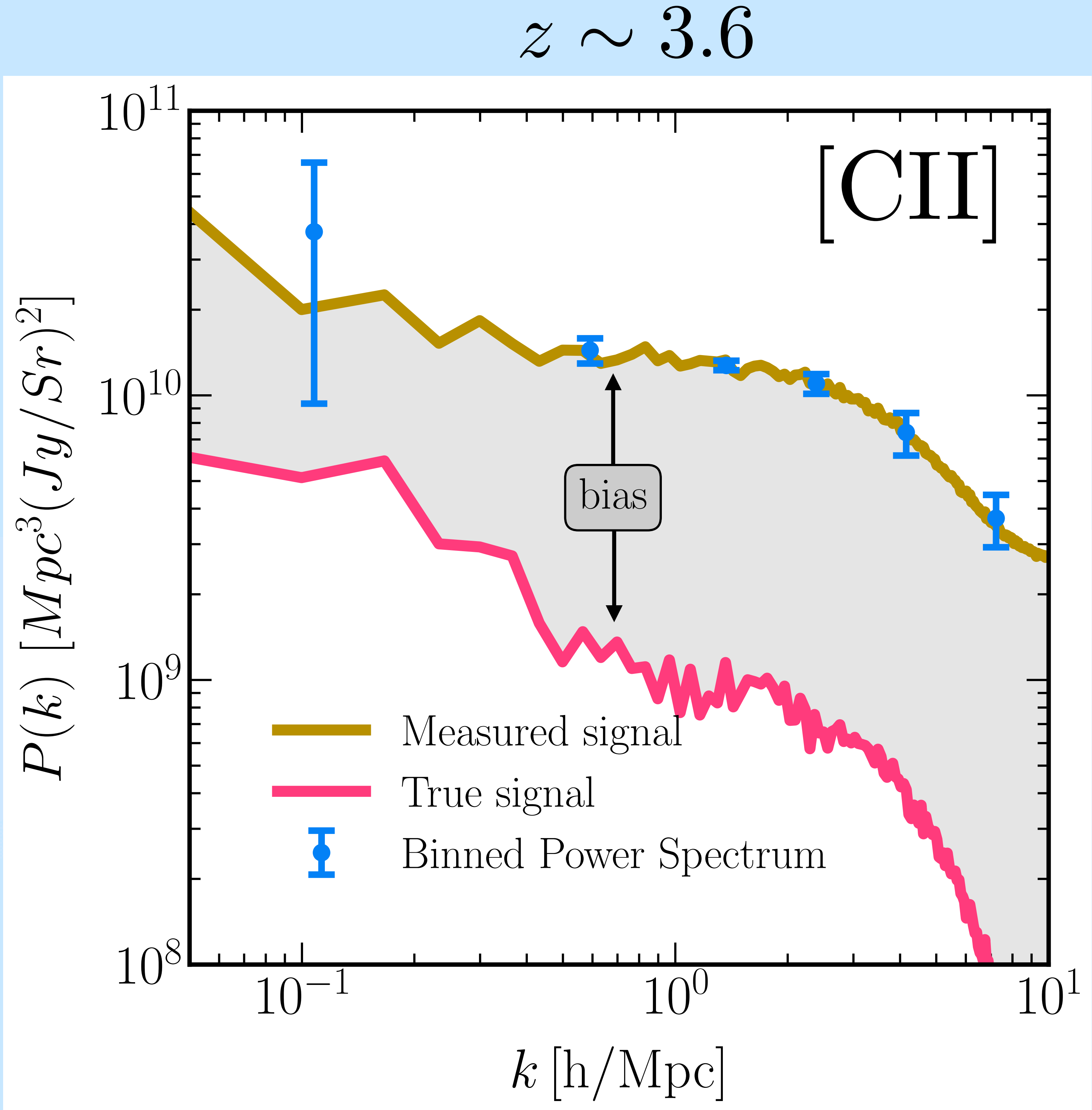
- Inference pipeline is able to estimate the signal and noise contribution in the mock map.
- Can we do cosmology with MLIM?
  YES!
- Can we do **precision cosmology** with MLIM?
  Probably not!



Roy, Pullen, Somerville+ (in prep.)

Target signal and interlopers



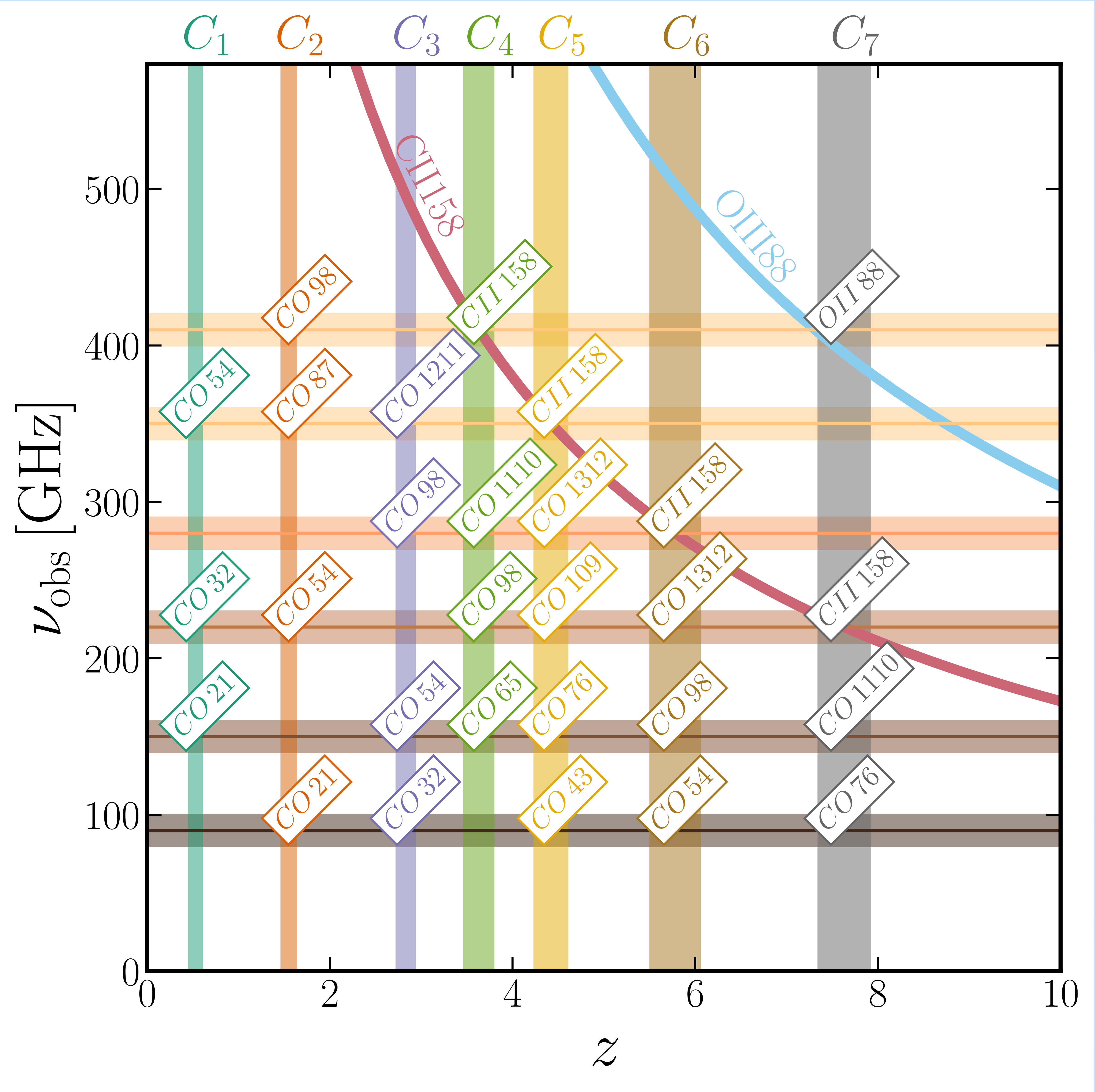


- The signal is significantly biased by interlopers and noise.
- Noise bias can be reduced through data splitting.

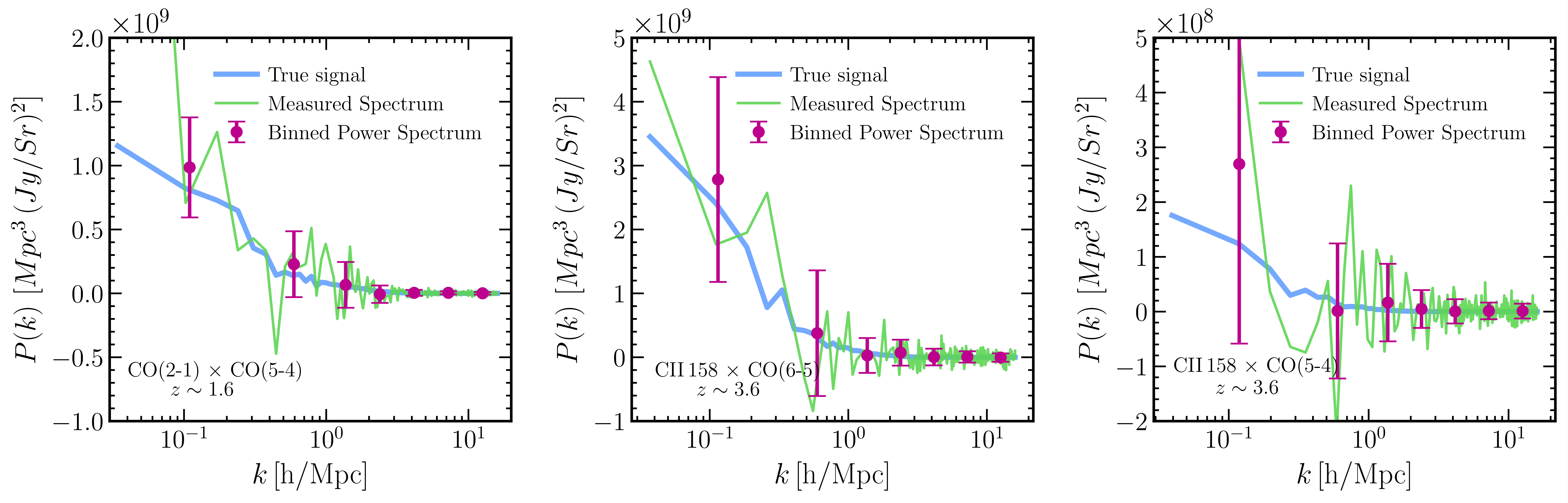
How can interlopers be removed?

Roy+ (2024)

Line-line cross-correlations



Line-line cross-correlations

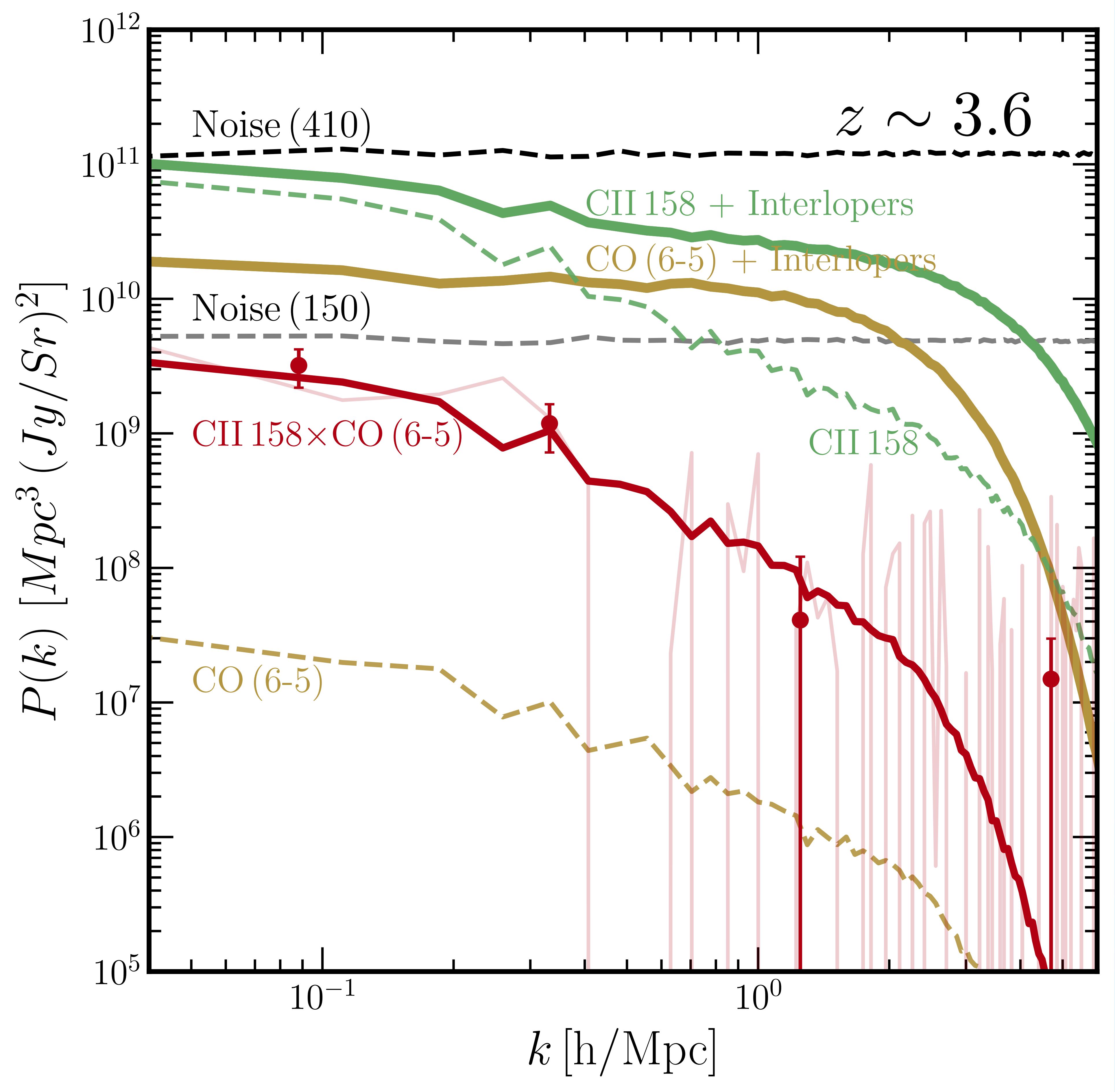


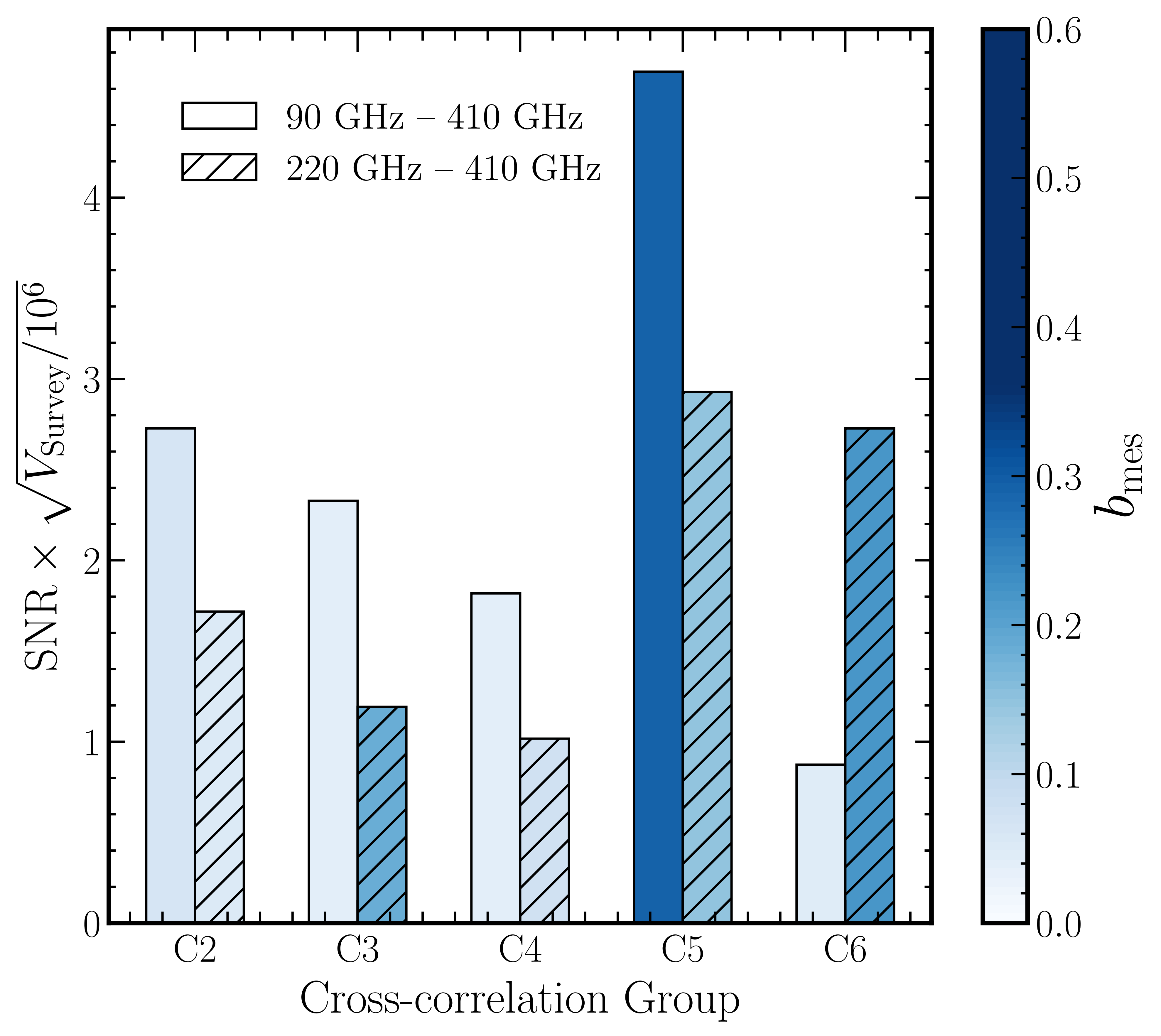
Two distinct frequency channels ⟶ Two or more lines from the same redshift

- The detectability of the cross-correlated signals is 23, 10, and 5, respectively, for a FYST-like experiment.

Roy & Battaglia (2024)

LIM cross-correlations





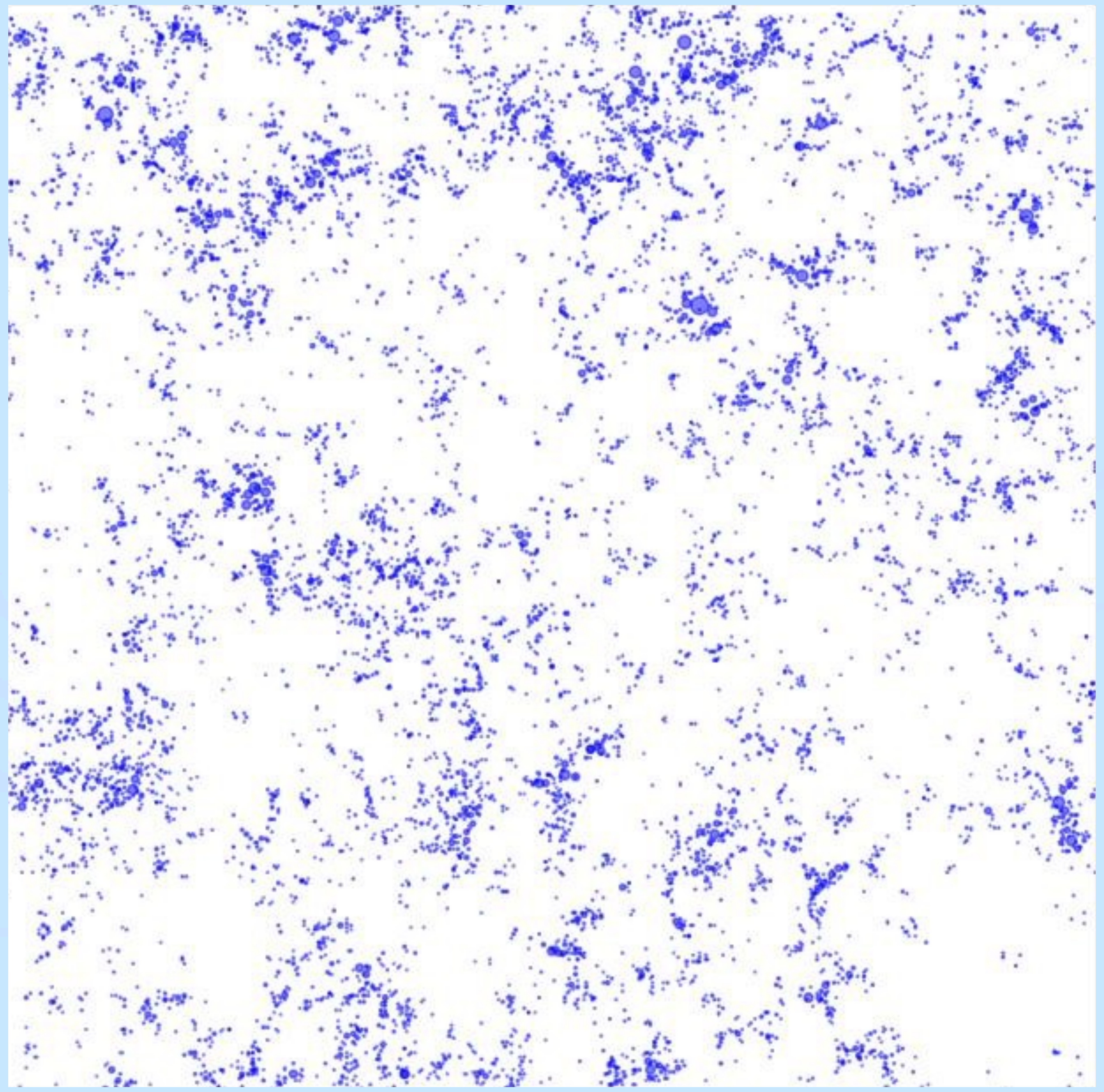
■ Having low-frequency channels, such as 90 GHz and 150 GHz, enables line-line cross-correlations from z = 0.58 to 7.6, with reduced measurement bias.

Roy+ (2024)

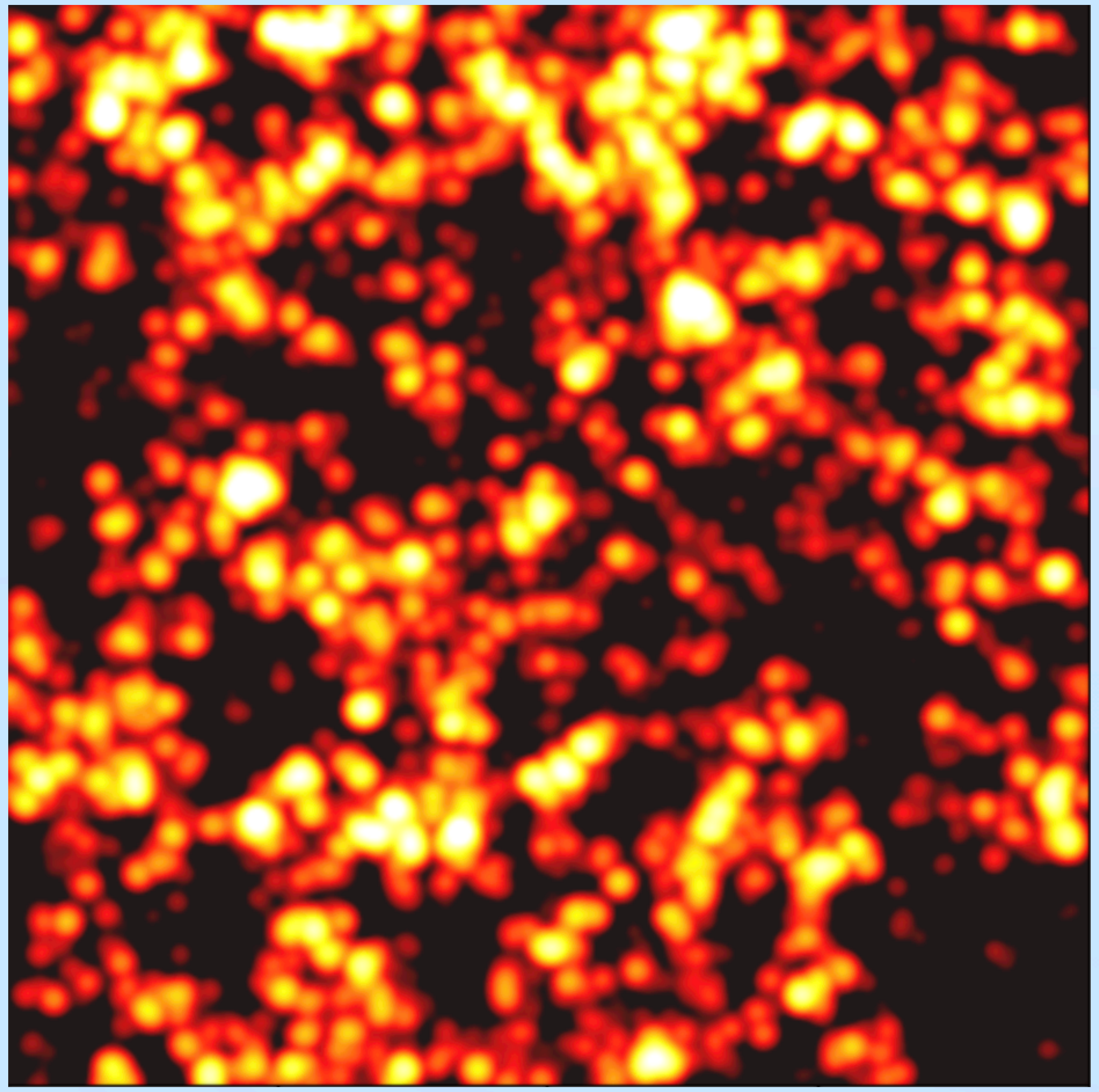
LIM x 21cm

Will it be possible to constrain parameters related to the 21 cm signal and LIM for SKA-like and FYST-like experiments?

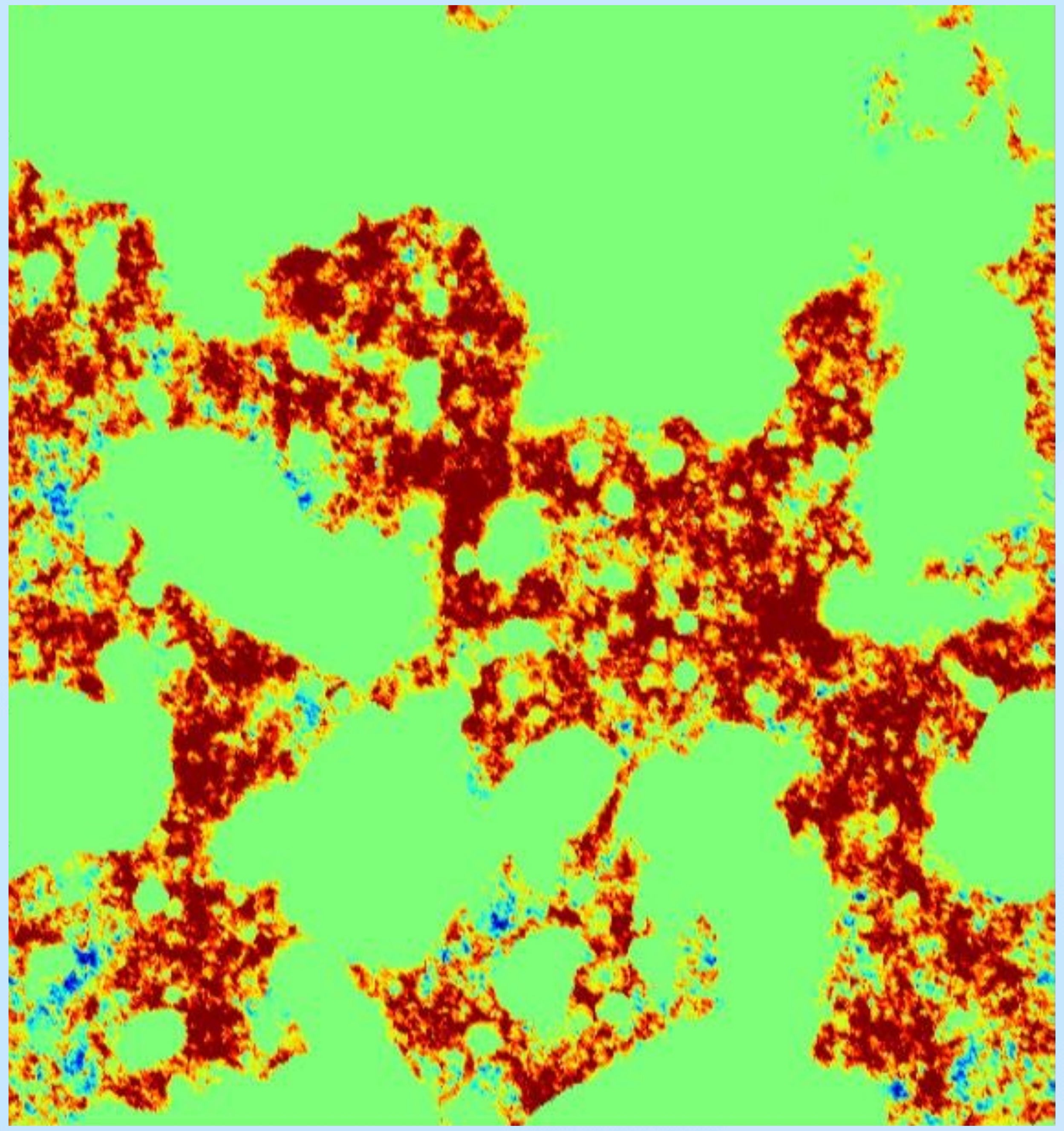
Galaxies



[CII]



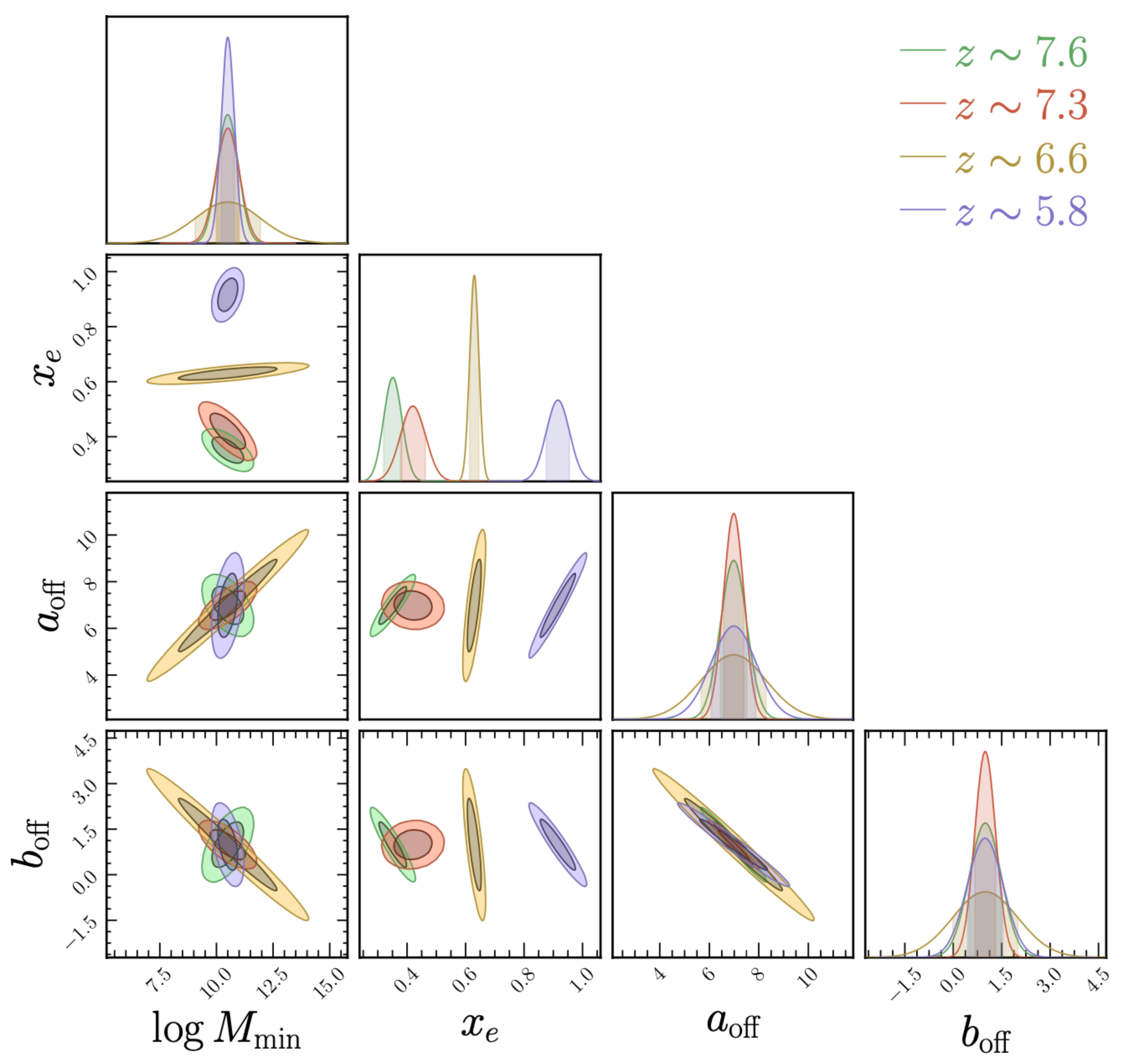
21cm

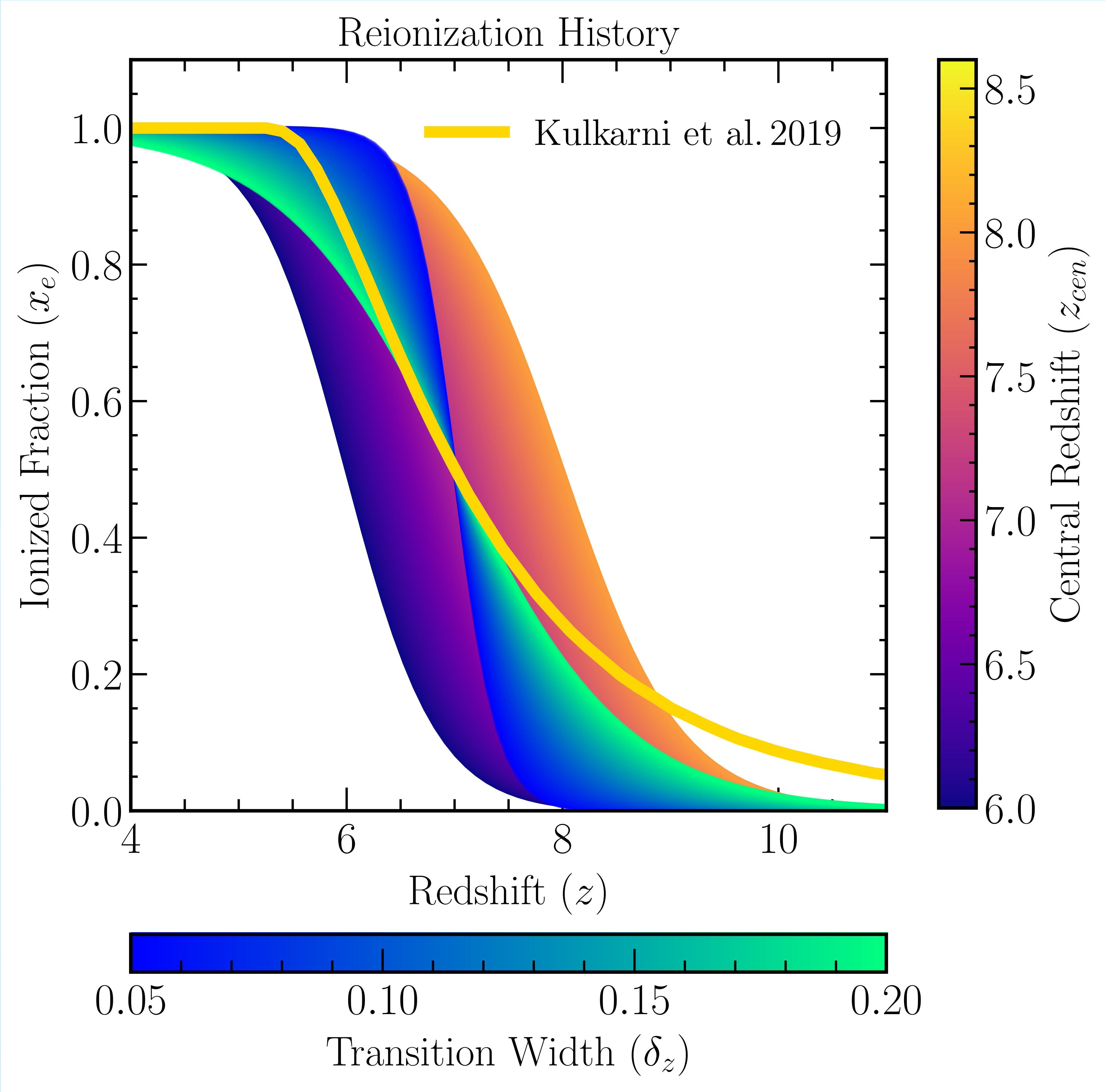


- The 21cm line probes neutral hydrogen, while [CII] probes ionized regions.
- On large scales, these two signals should be anti-correlated.

Fronenberg+ (2024), Drumitru+ (2018)
Roy, Pullen, Somerville+ (in prep.)

Parameters from LIM x 21cm





- LIM × 21cm could be a direct probe to measure the ionization fraction at several redshifts.

Roy, Pullen, Somerville+ (in preparation)

Summary and Outlook

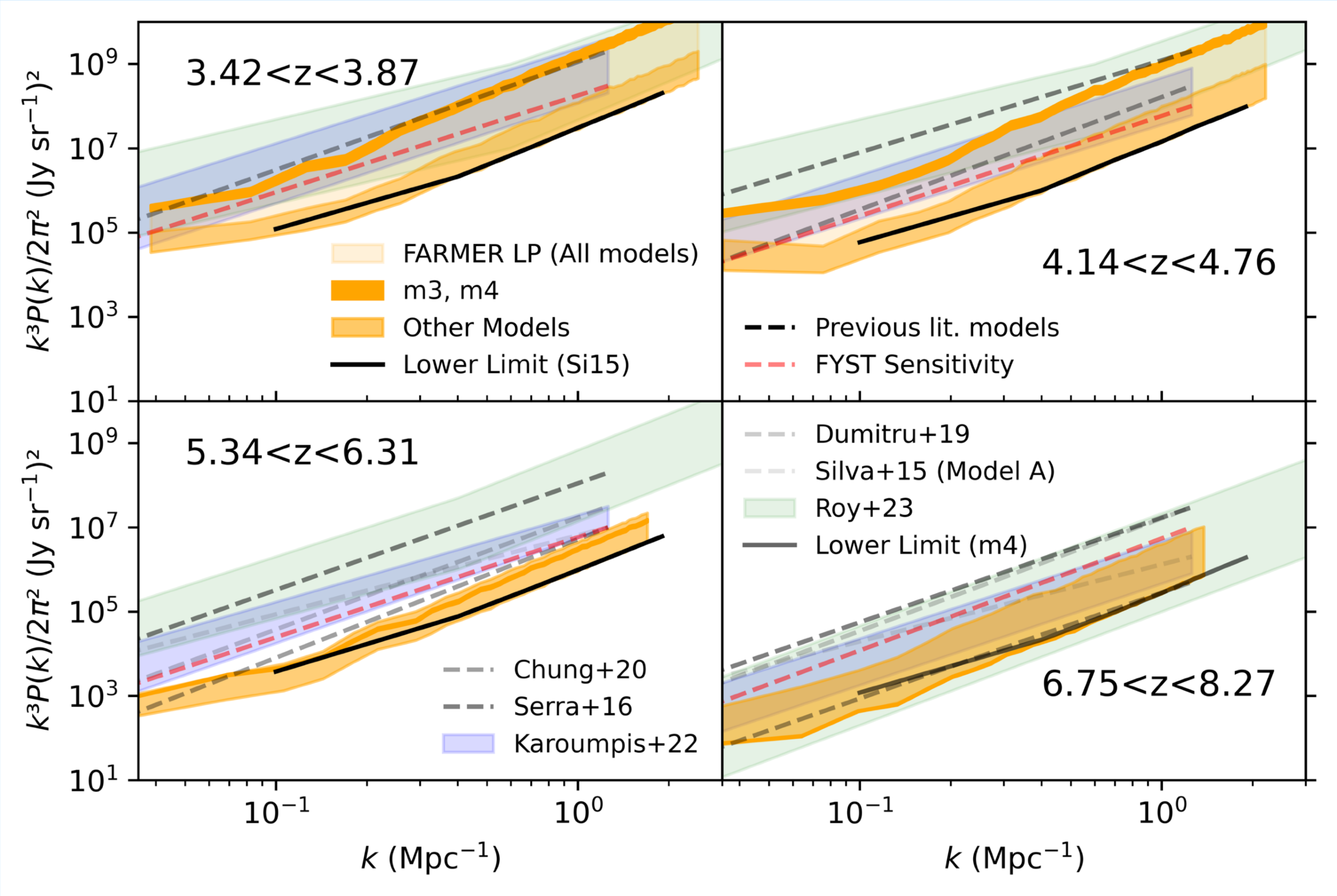
- LIM modeling is very complex, and the astrophysical uncertainty is huge.
  Current and future LIM experiments will be able to rule out some models.

- Using Semi-Analytic Models of galaxy formation will be useful for parameter inference.
  They enable precise constraints on physical parameters using multi-line observations.

- Combining analysis methods such as the power spectrum, VID, and field-level analysis could improve the parameter constraints.
  This is an unexplored territory. The data analysis pipeline needs to be ready soon and thoroughly checked.

- It is feasible to do cosmology at low redshift, similar to the 21cm signal. Of course, there will always be degeneracies between astrophysical and cosmological parameters.
  Assimilating datasets together, using priors from Planck, will help break degeneracies

Comparison of the LIM models



$$P^{\mathrm{line}}(k,z) = A(z) + B(z) \times P_{\mathrm{m}}(k,z)$$

- Clarke+ simulated [CII] maps for COSMOS 2020 galaxy catalogues
- Most of the models are valid for a particular redshift range, and extrapolation is viable for other redshifts.
- FYST forecasts a few binned [CII] power spectrum detections with S/N >1 for most models at z~3.6 and 4.4.
- High-redshift [CII] detection depends on the models and remains challenging.

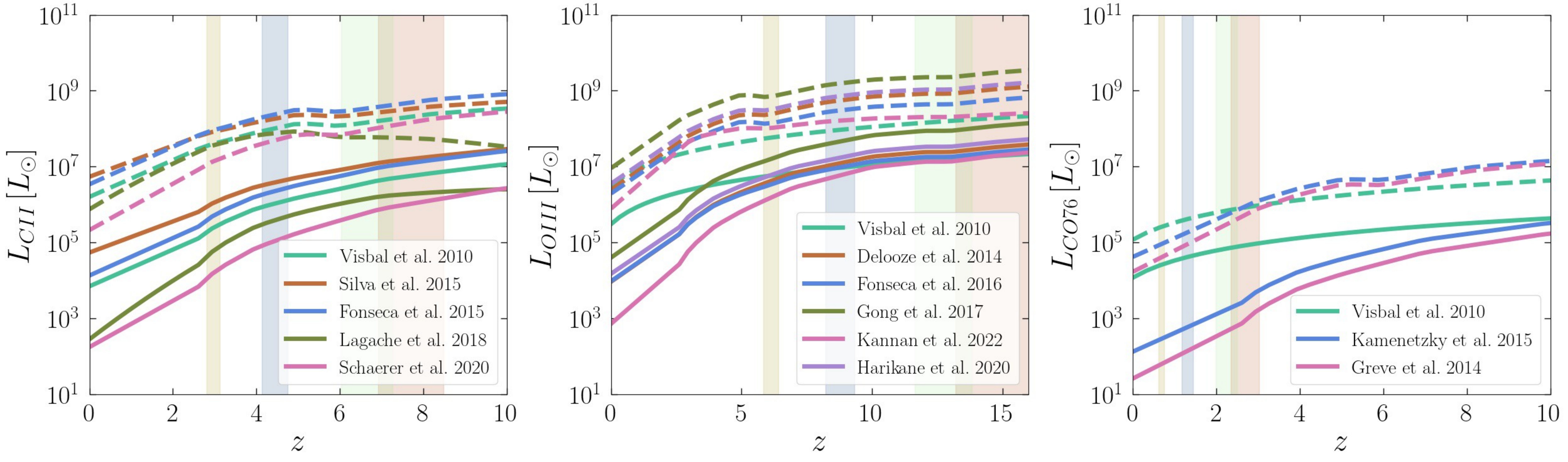
Clarke+ (2024), Roy+ (2023), Karoumpis+ (2022), ….

General steps for LIM modeling

Vertical shaded regions: FYST's redshift coverage

Solid: $M_{\rm halo} = 10^{10}\, M_\odot/h$

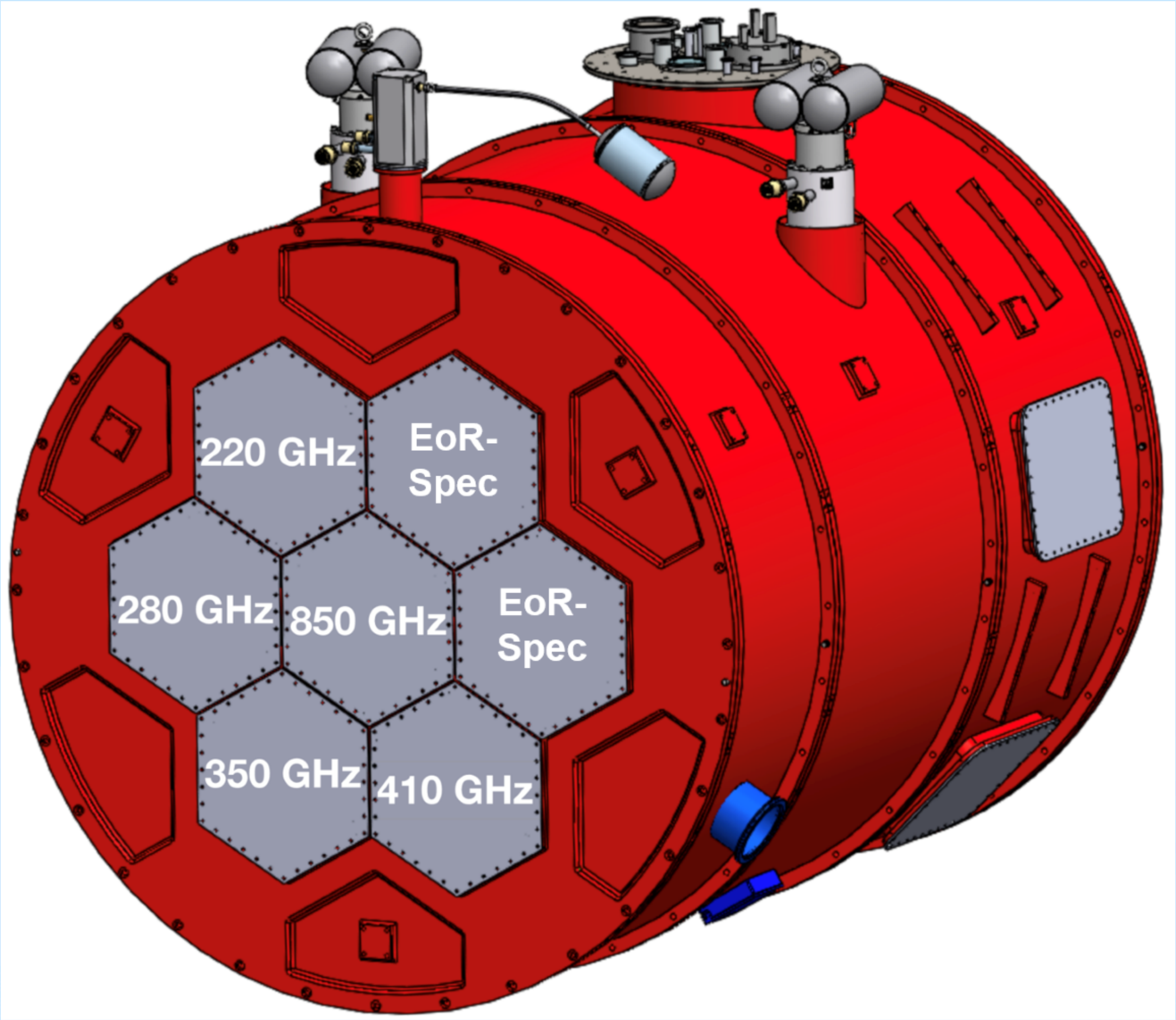
Dashed: $M_{\rm halo} = 10^{11}\, M_\odot/h$



$$I_{\rm line}(z) = \frac{c}{4\pi}\frac{1}{\nu_{\rm rest}H(z_{\rm em})}\int_{M_{\rm min}}^{M_{\rm max}} L_{\rm line}(M,z)\frac{dn}{dM}dM$$

Roy+ (2023), Murmu+ (2023), Leung+ (2020)

FYST



CMB Observations: Characterize foregrounds & Rayleigh scattering

SZ Observations: SZ spectrum from millimeter through to submillimeter

Line Intensity Mapping: LSS back to Reionization with CII and CO Lines

CIB Observations: Galaxy formation from the first billion years to Cosmic Noon

Galactic Polarization: Characterizing magnetic fields and galactic polarization science

Galactic Ecology: Characterizing cloud and star formation in the MW and nearby galaxies

Aravena+ (2022)

Mitigating the interlopers

- **Bayesian analysis** [e.g., Cheng+ (2024)]
  Utilizes the correlated information to perform joint inference for the relevant lines.

- **Line-line cross-correlations** [e.g., Roy & Battaglia 2024]
  Cross-correlate two lines originating from the same redshift. The signals are correlated, whereas noise and interlopers are not.

- **Masking of contaminated pixels** [e.g., Karoumpis+ (2024)]
  Masks the bright pixels contaminated by interlopers based on external catalogs, and then attempts to extract the target signal from high redshifts.

- **Using external tracers** [e.g., Bernal+ (2024)]
  Removes the interlopers by probing the statistical correlations with external tracers on large scales.

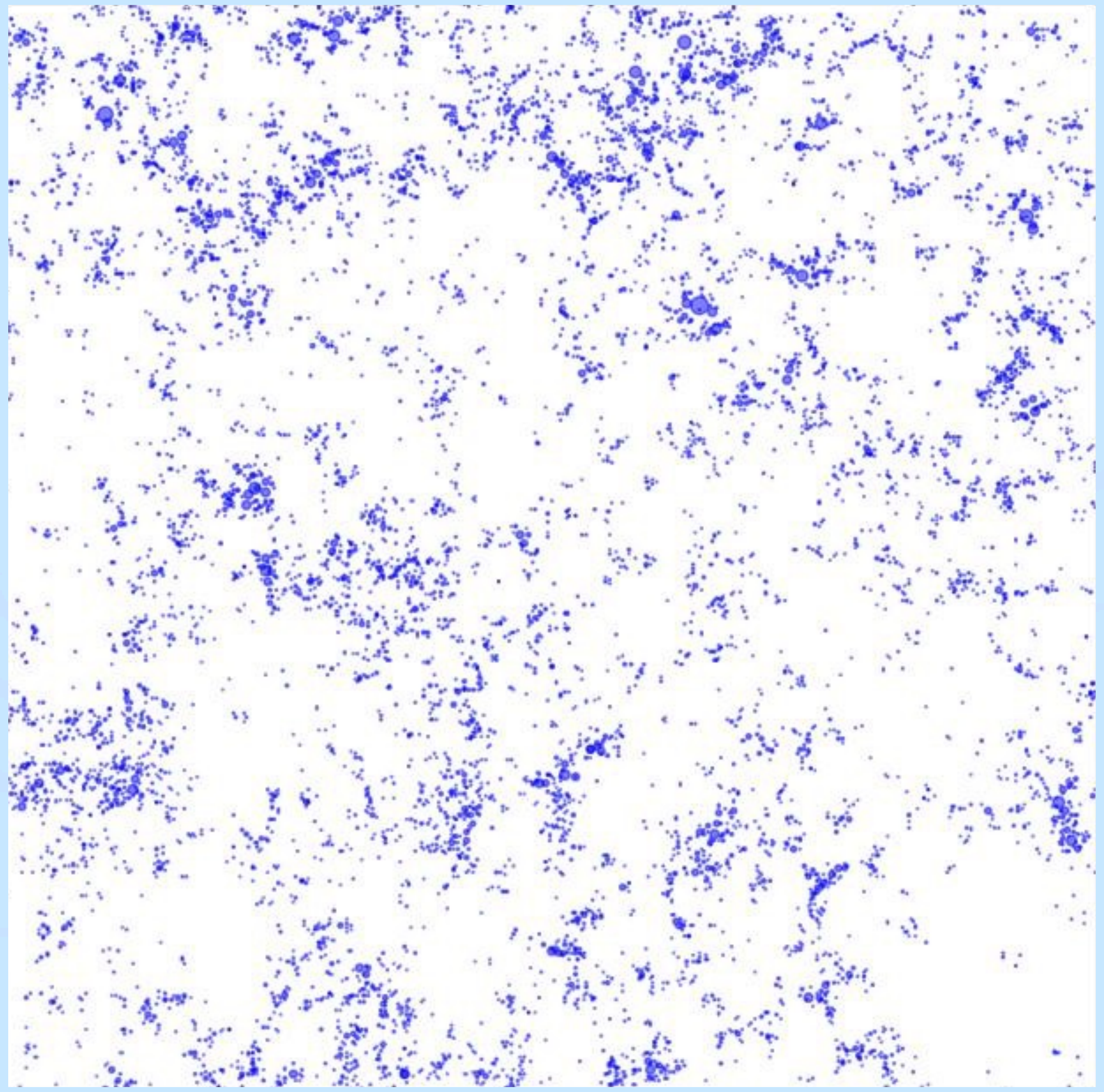
Outline

- **Background**
- **Modeling of line intensities**
- **Analysis techniques**
- **Detection of CO(3-2) using Planck data**
- **Cross-correlations with other probes**
- **Conclusion**

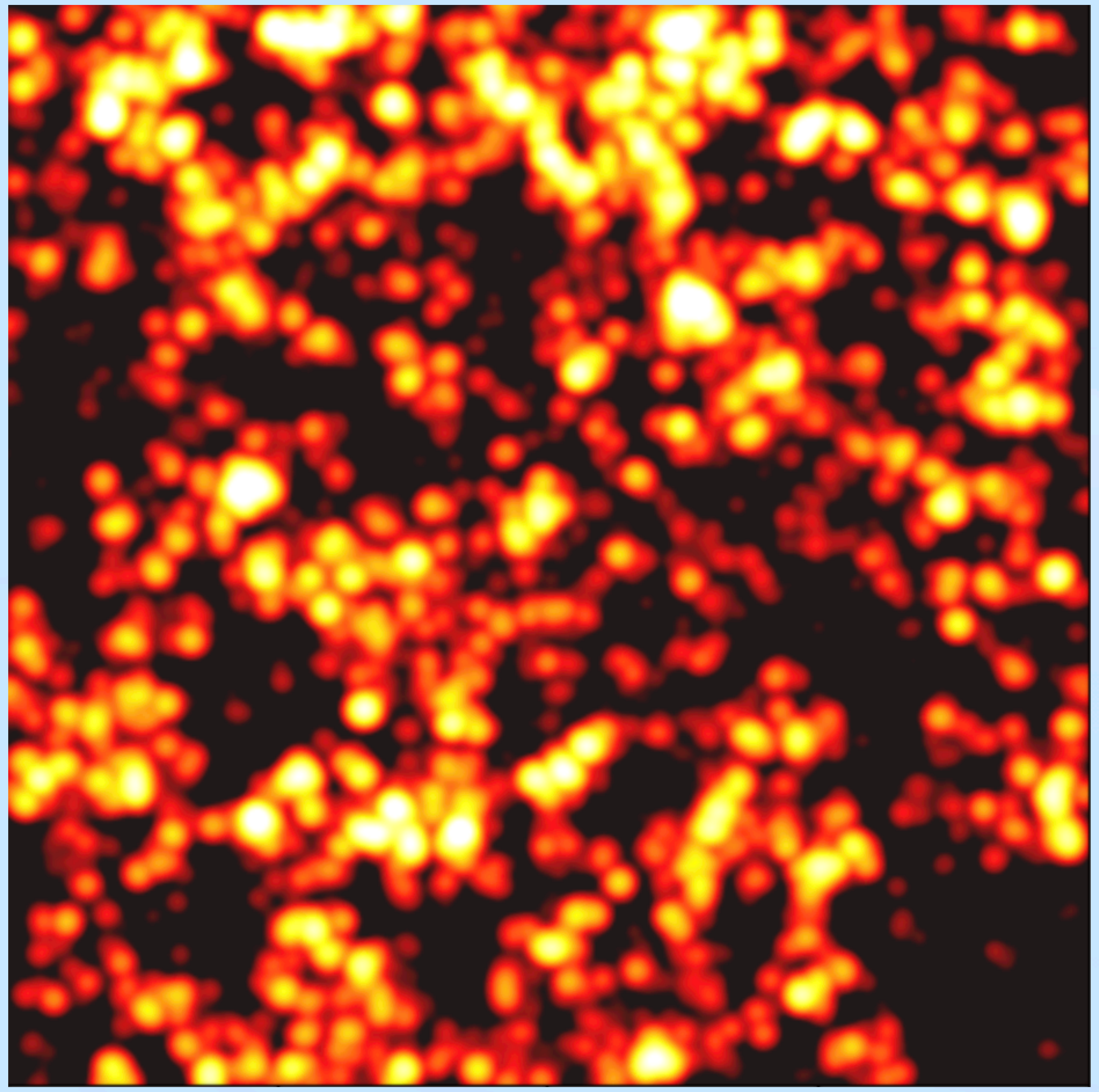
LIM x 21cm

Will it be possible to constrain parameters related to the 21 cm signal and LIM for SKA-like and FYST-like experiments?

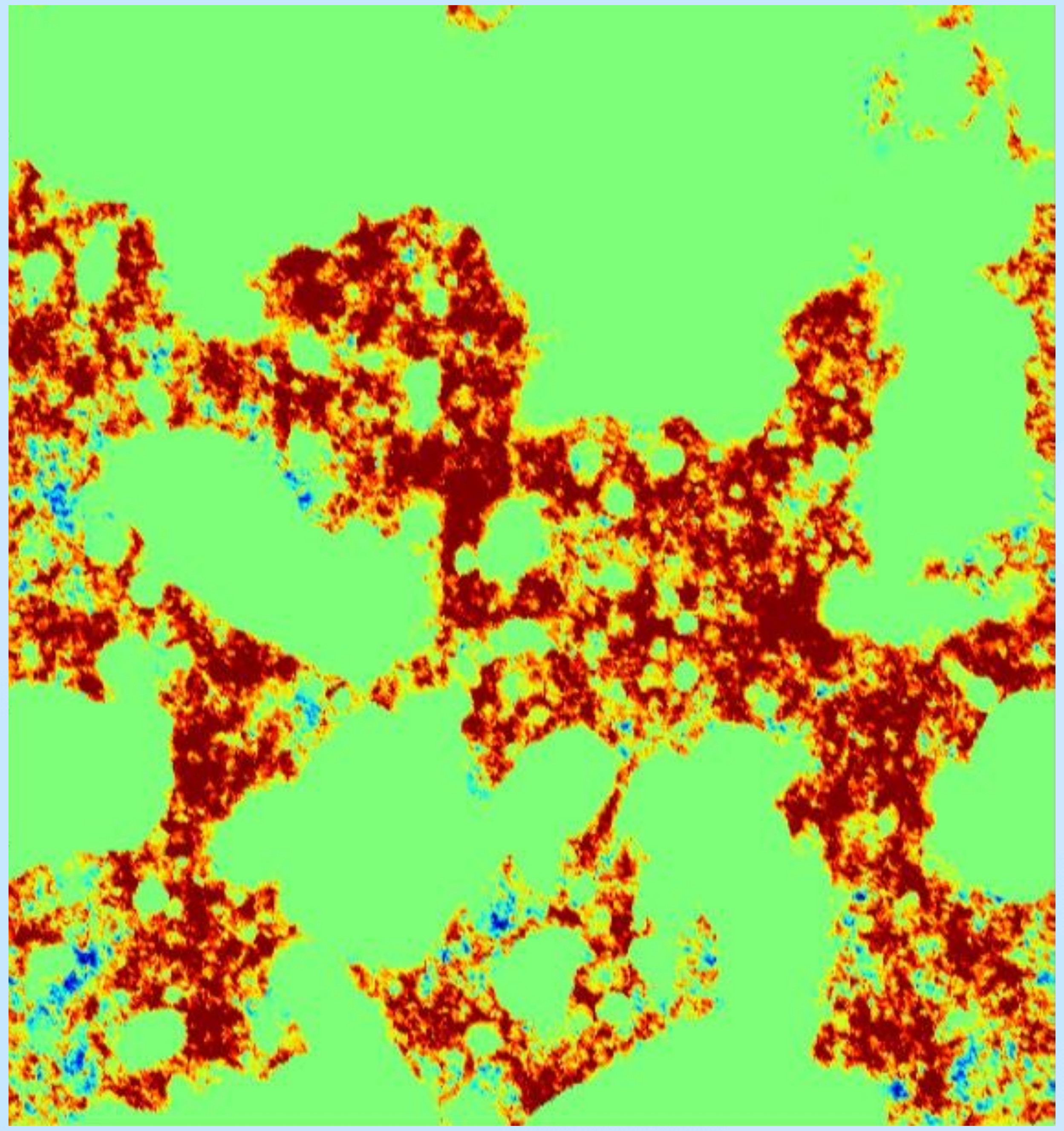
Galaxies



[CII]



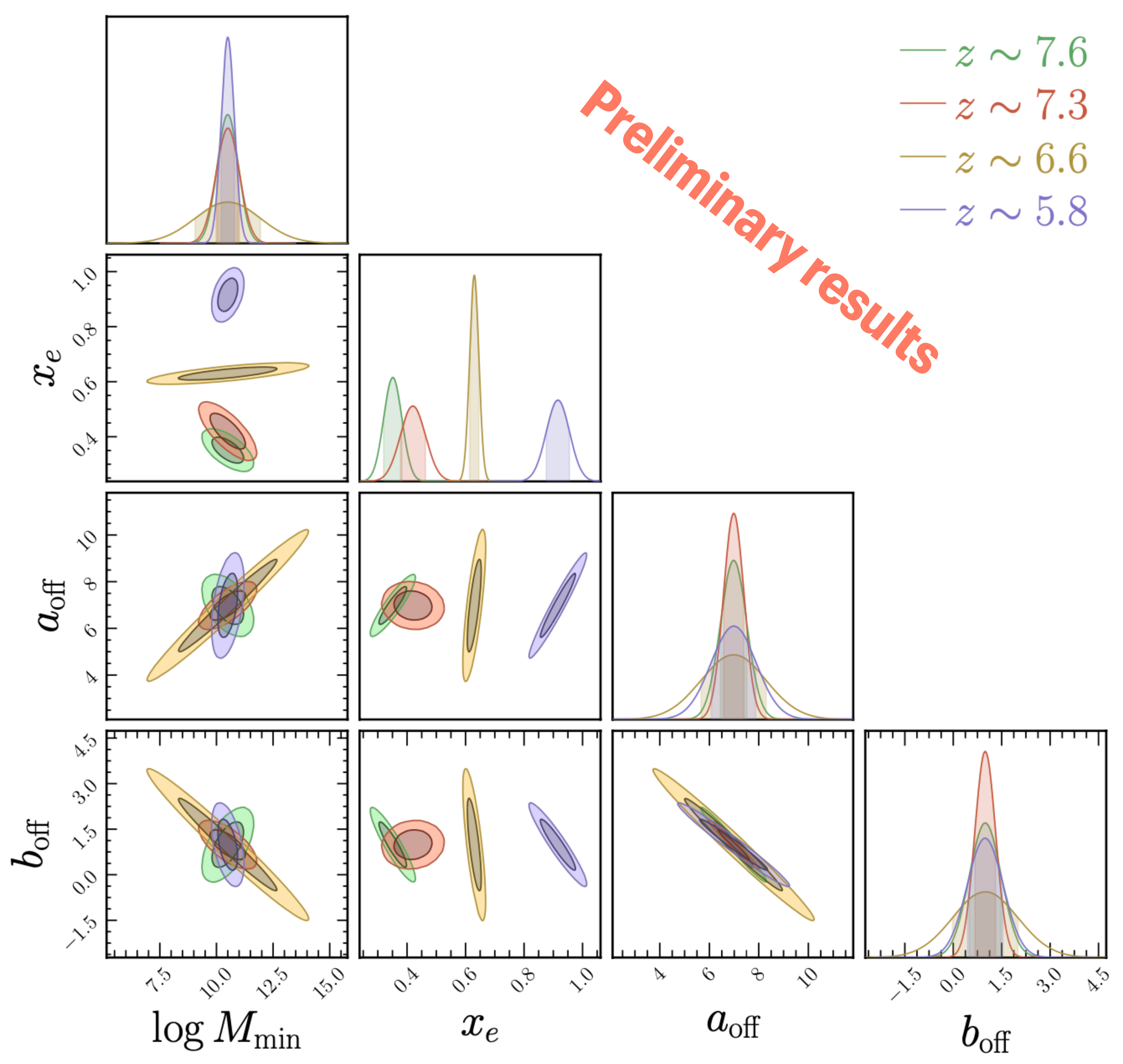
21cm

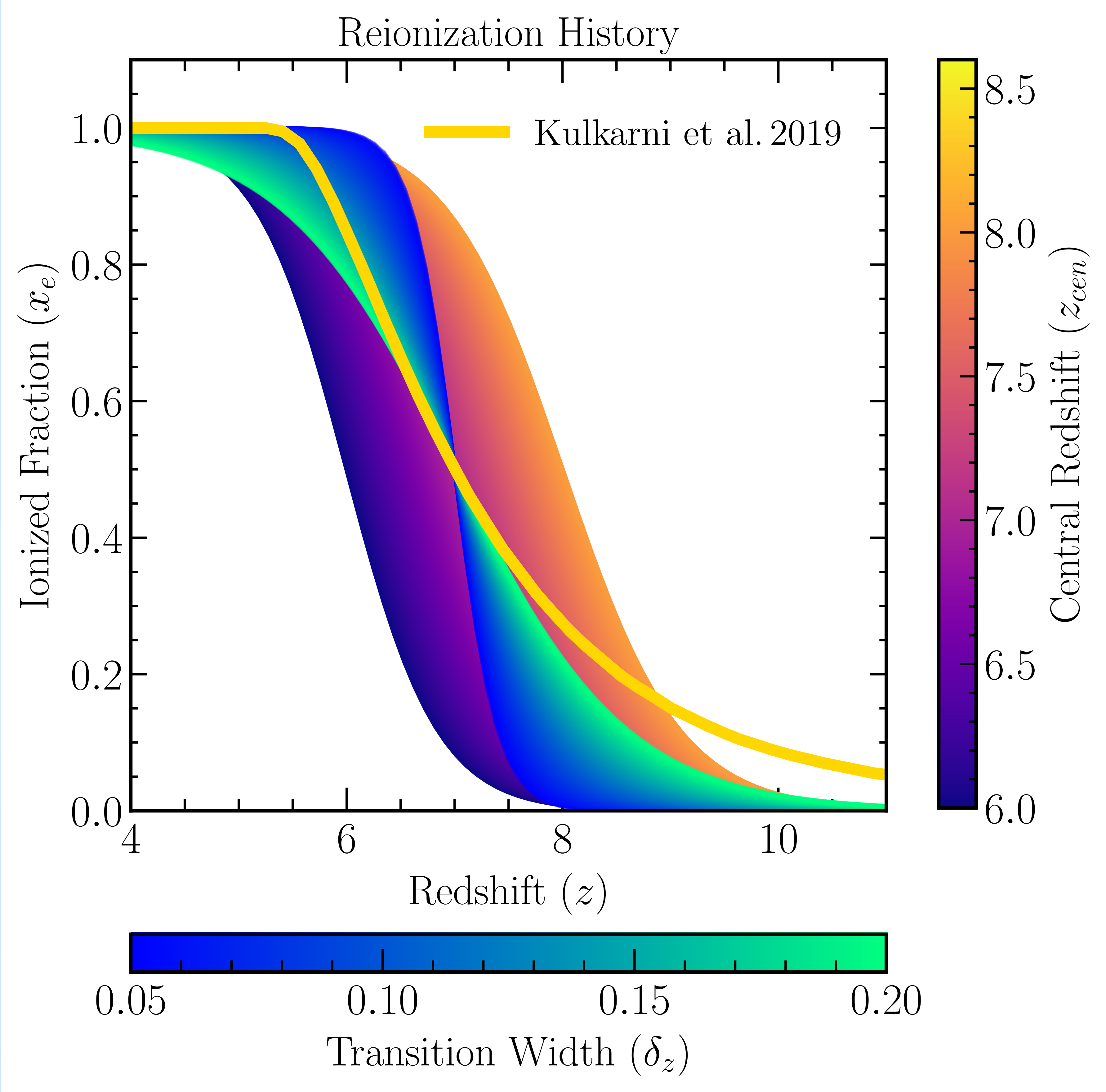


- The 21cm line probes neutral hydrogen, while [CII] probes ionized regions.
- On large scales, these two signals should be anti-correlated.

Fronenberg+ (2024), Drumitru+ (2018)
Roy, Pullen, Somerville+ (in prep.)

Parameters from LIM x 21cm





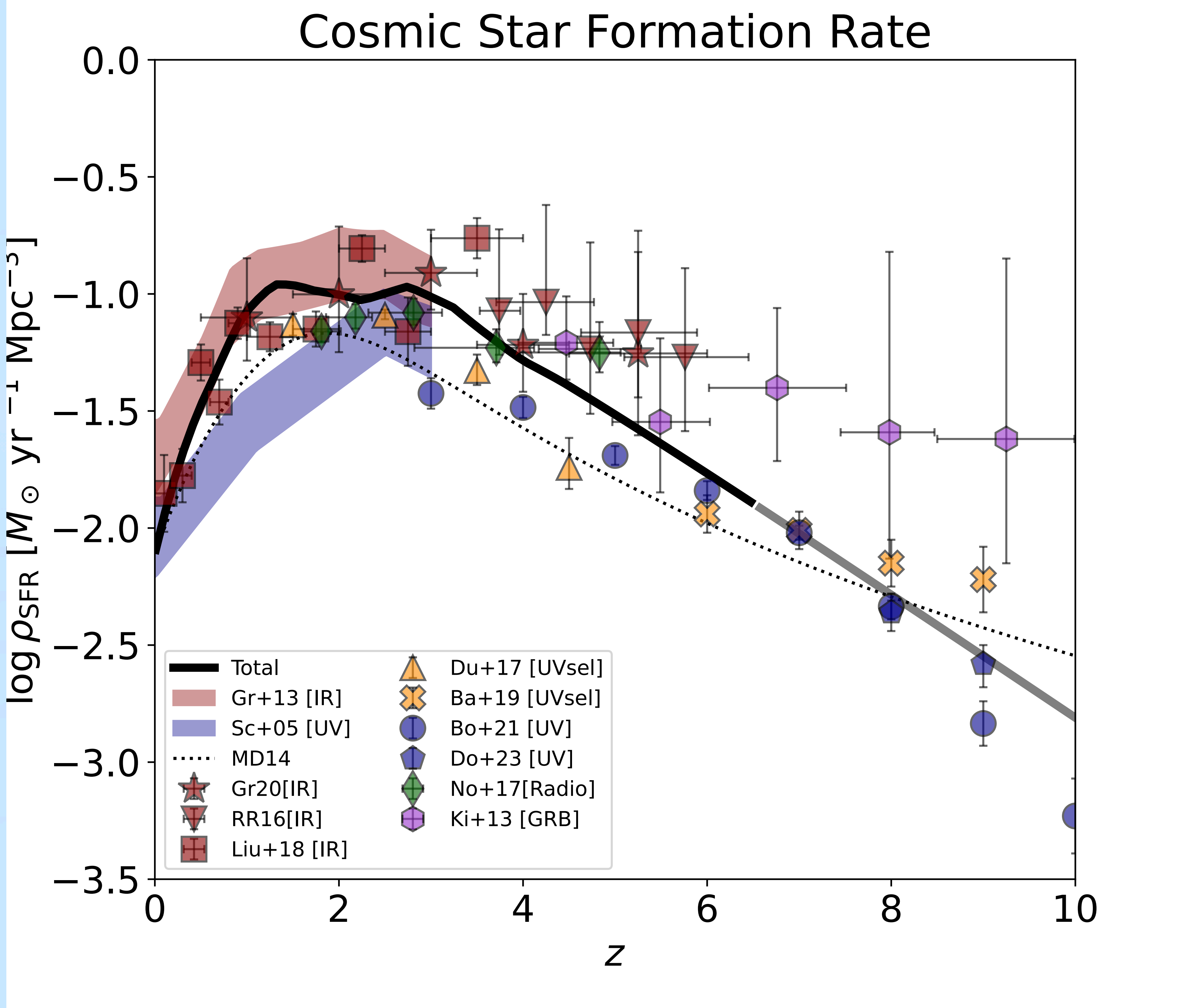
- LIM × 21cm could be a direct probe to measure the ionization fraction at several redshifts.

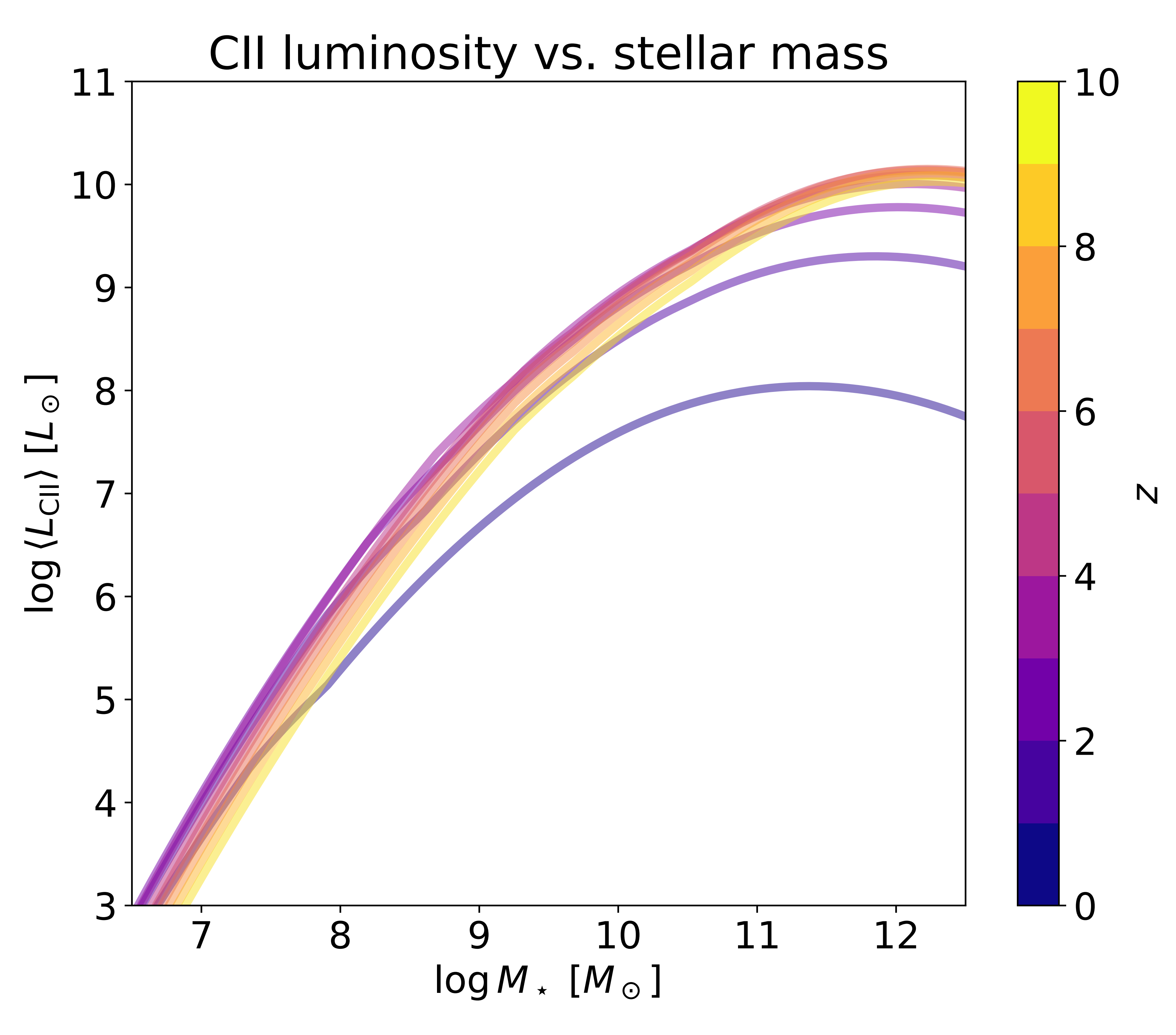
Roy, Pullen, Somerville+ (in preparation)

Back up

Learning from Galaxy observations

- We utilized the relationships among the observed stellar mass function, star formation rate, and fundamental metallicity relation





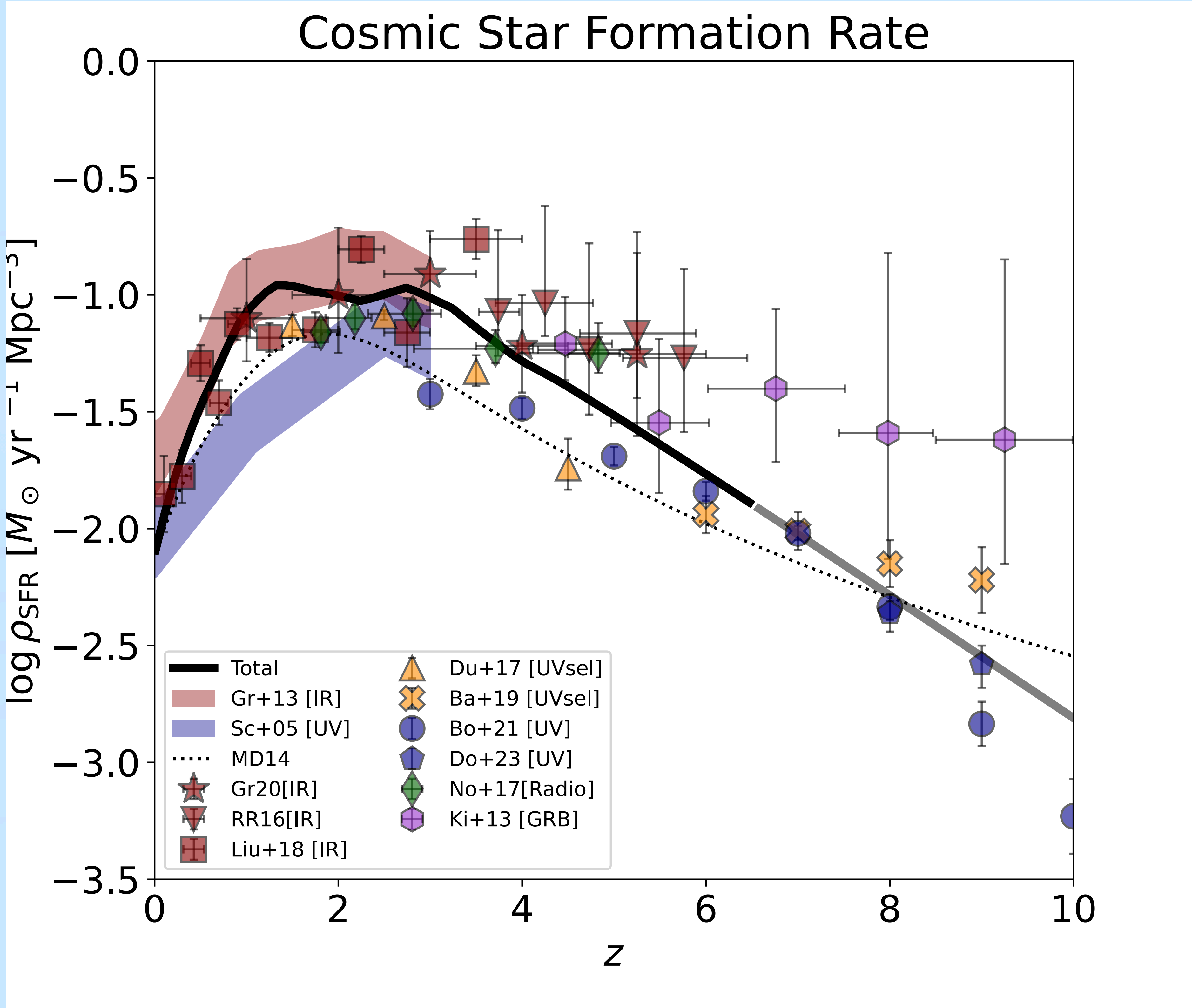
- Next, we combine stellar evolution and radiative transfer presicriptions to infer [CII] intensities

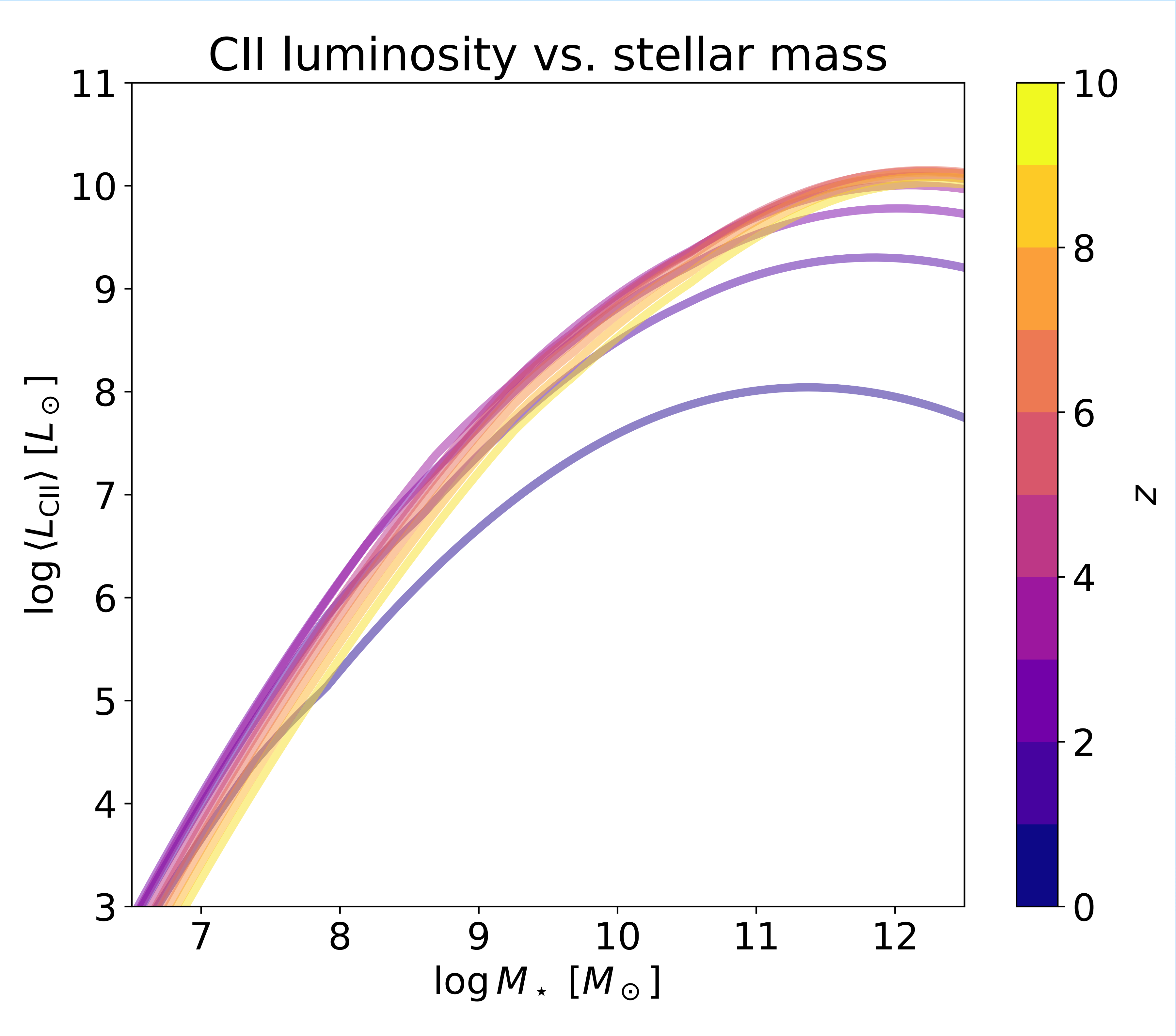
$$\langle L_{[\rm CII]}\rangle(M_\star, z) = \int \mathrm{d}\log\psi \, \frac{\mathrm{d}p}{\mathrm{d}\log\psi}(\psi|M_\star, z) \int \mathrm{d}\log Z \, \frac{\mathrm{d}p}{\mathrm{d}\log Z}(Z|M_\star, \psi, z)$$

Roy & Lapi (2024)

Learning from Galaxy observations

- We utilized the relationships among the observed stellar mass function, star formation rate, and fundamental metallicity relation



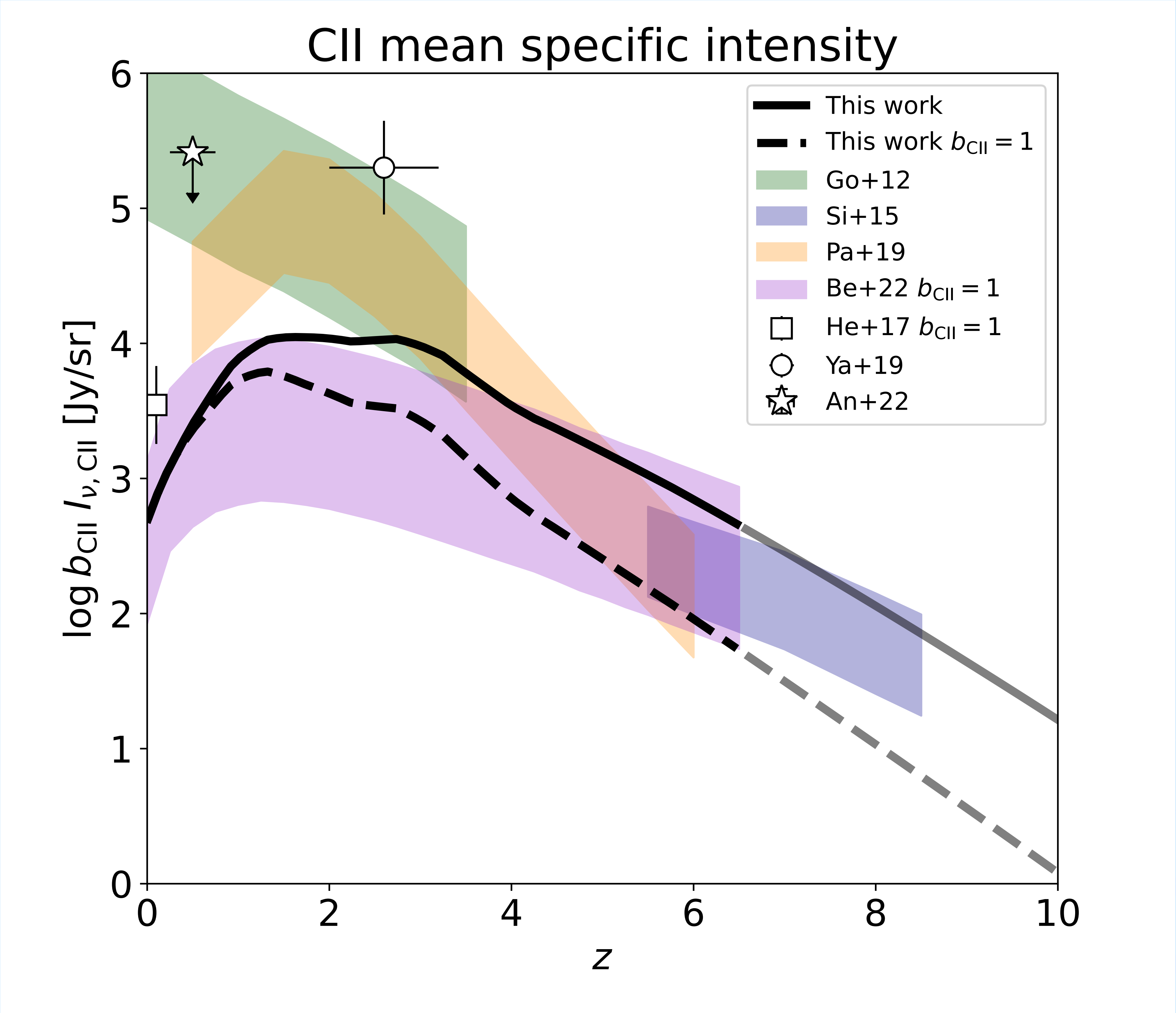


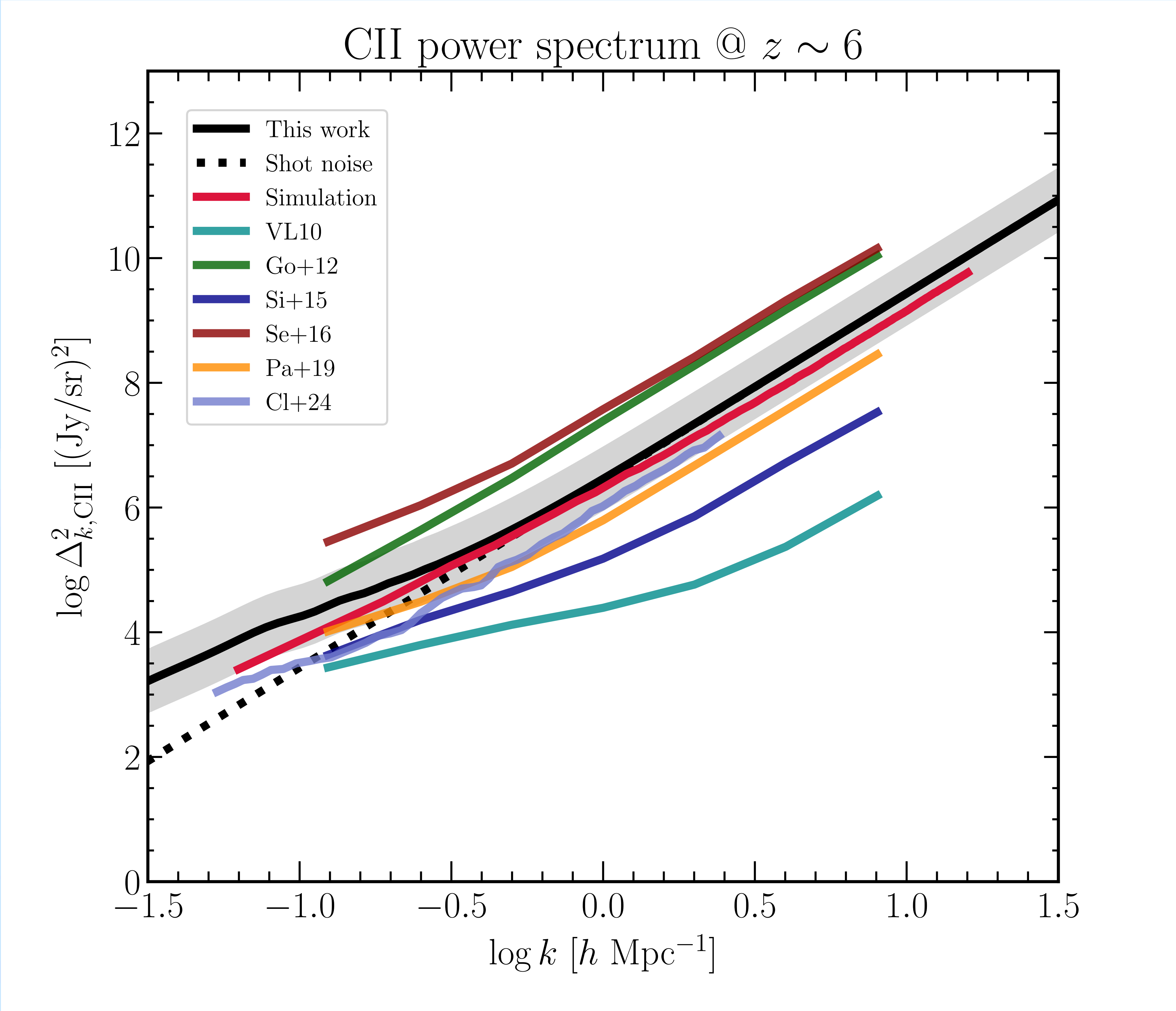
- Next, we combine stellar evolution and radiative transfer presicriptions to infer [CII] intensities

$$\langle L_{[\mathrm{CII}]}\rangle(M_\star, z) = \int \mathrm{d}\log\psi\, \underbrace{\frac{\mathrm{d}p}{\mathrm{d}\log\psi}(\psi|M_\star, z)}_{\text{Observed}} \int \mathrm{d}\log Z\, \underbrace{\frac{\mathrm{d}p}{\mathrm{d}\log Z}(Z|M_\star, \psi, z)}_{\text{Observed}}$$

Roy & Lapi (2024)

Learning from Galaxy observations





- This approach can be extended to the other lines such as CO and [OIII]
- The best-fit model, along with its uncertainty, falls within the range of other models.

Roy & Lapi (2024)