Hydrogen Intensity Mapping:
the ultimate signal is the weakest of all

Isabella Paola Carucci

INAF - OATs

Cosmology 2023 in Miramare

Funded by
the European Union
NextGenerationEU

INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

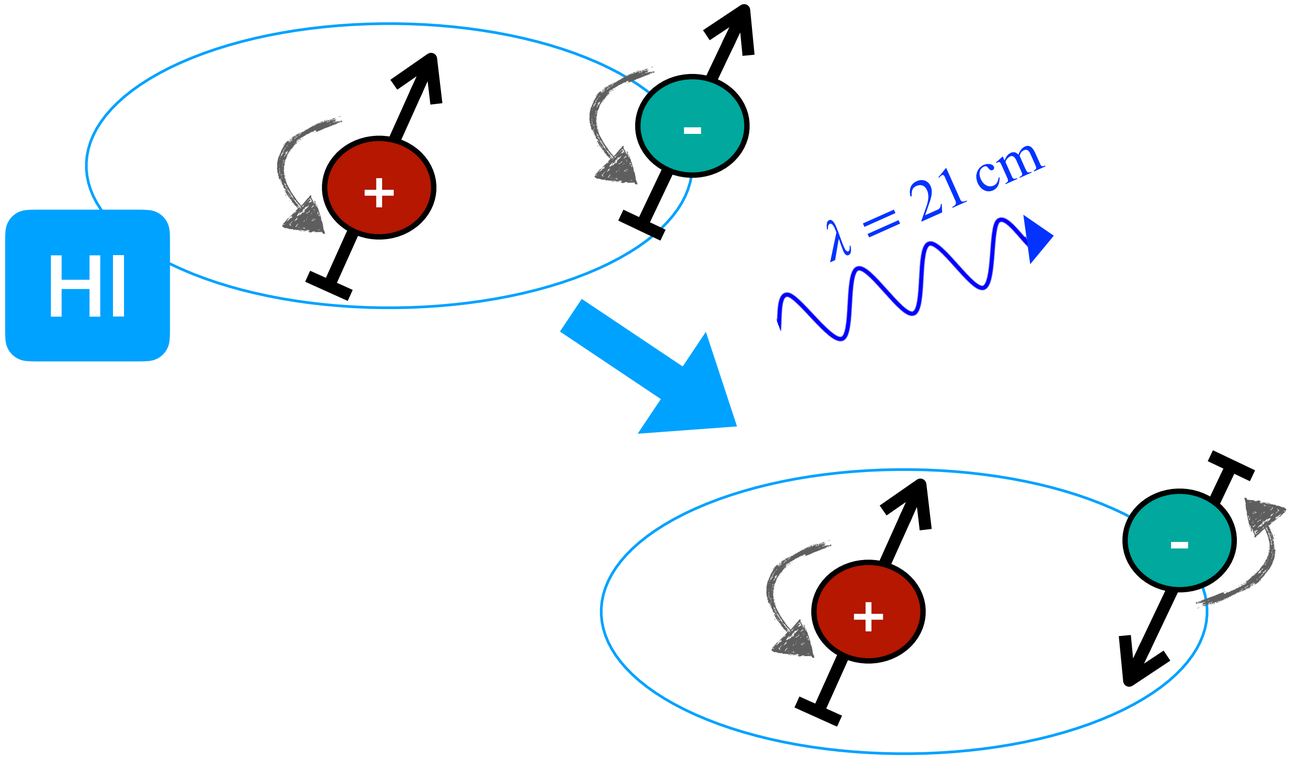
1. **Hydrogen Intensity Mapping (IM):** what is it and why to do it

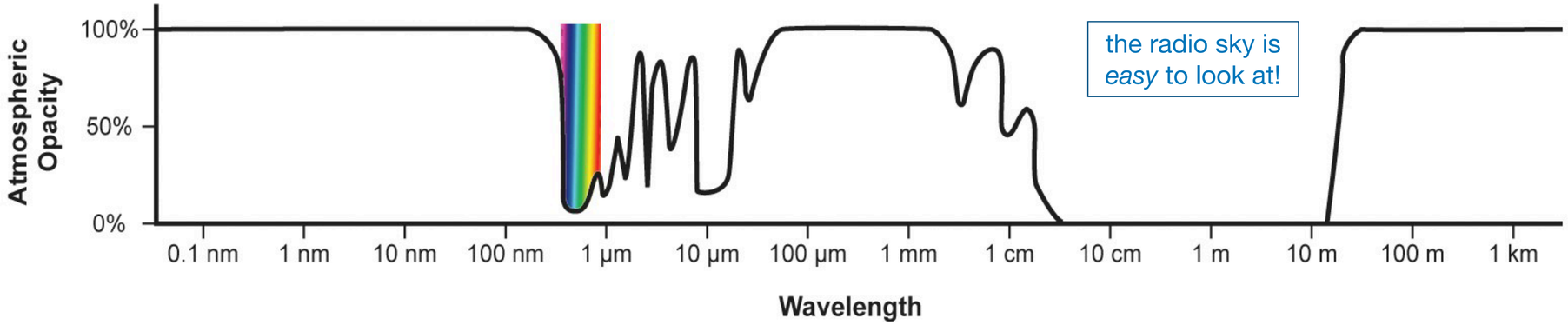
2. **IM is hard!** Biggest challenge: weakness of the IM signal compared to contaminants

3. We are getting there. **MeerKLASS**

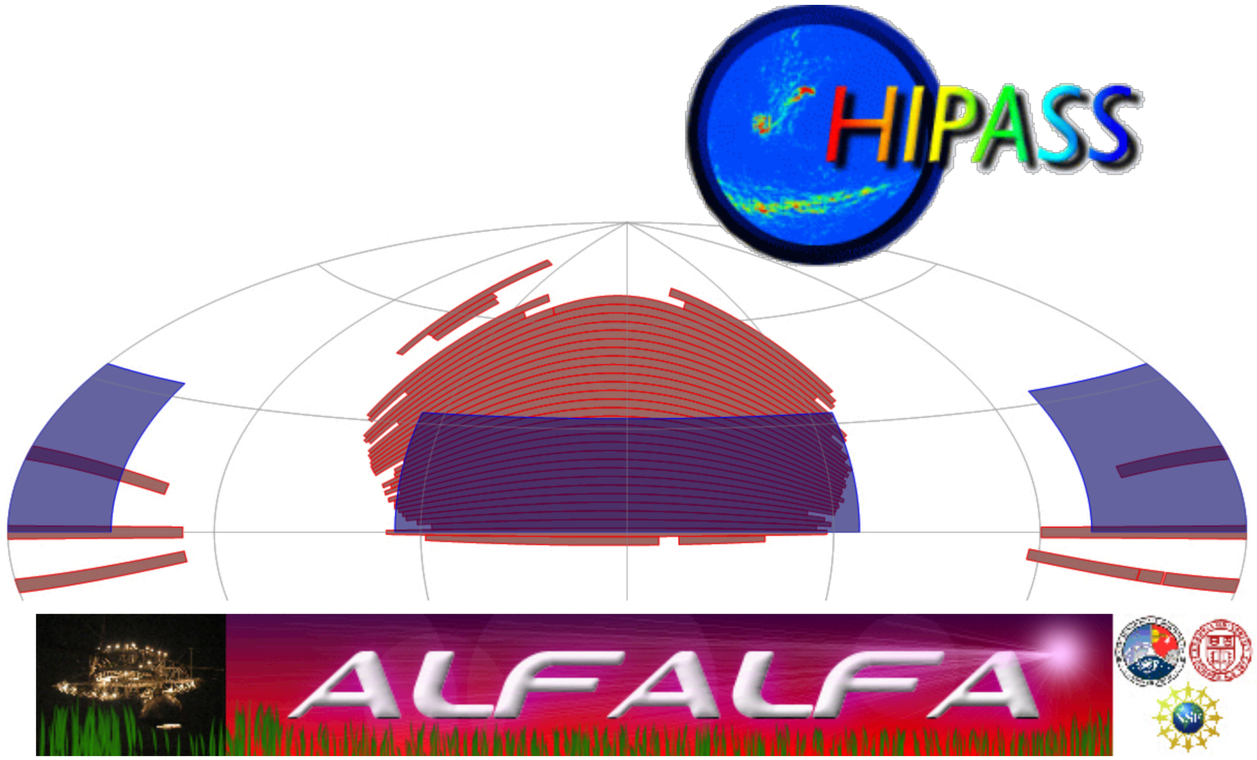
21-cm radiation

- strongly forbidden: $t_{1/2} \sim 10^7$ years
- VERY abundant
- Spectrally isolated
- Small obscuration

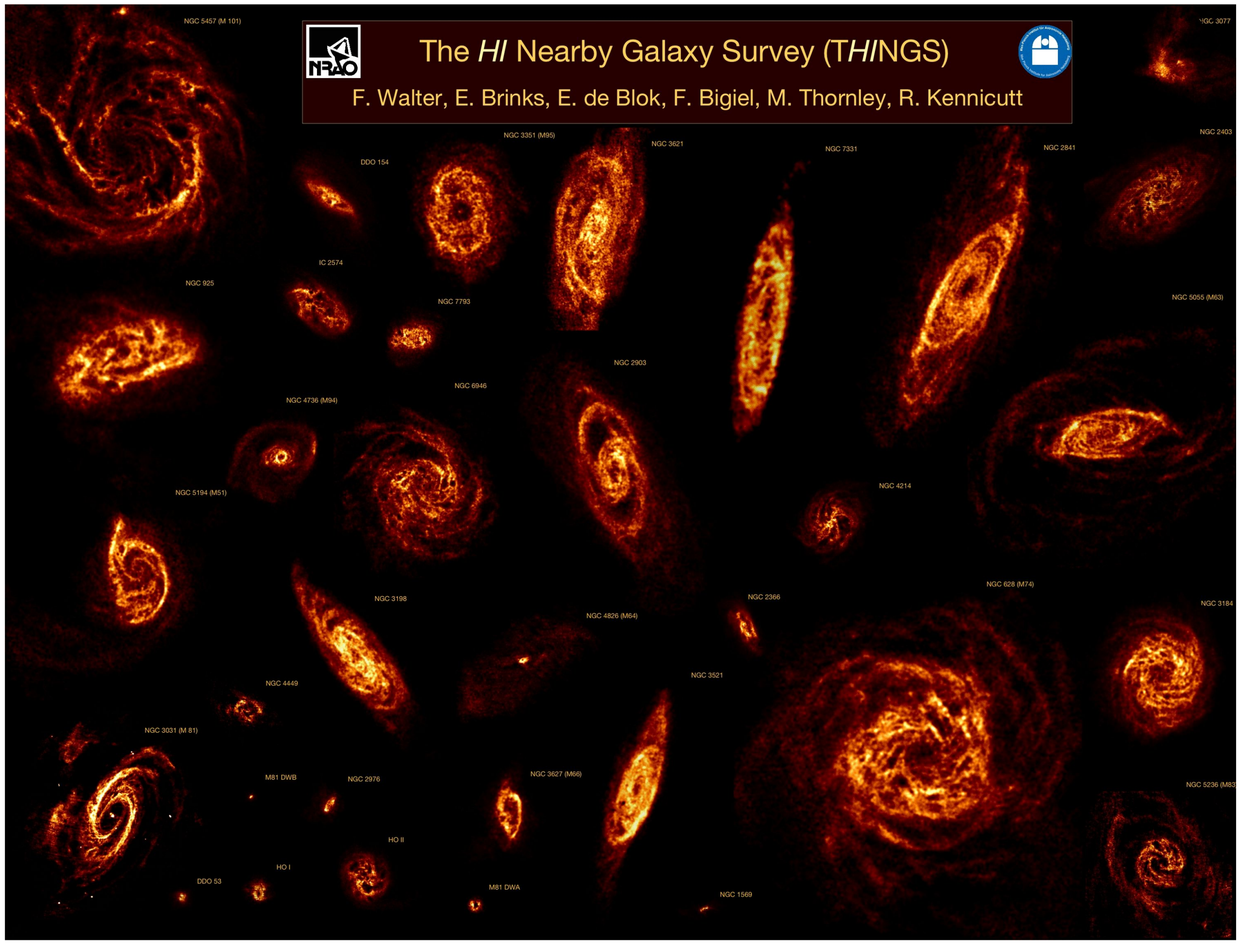




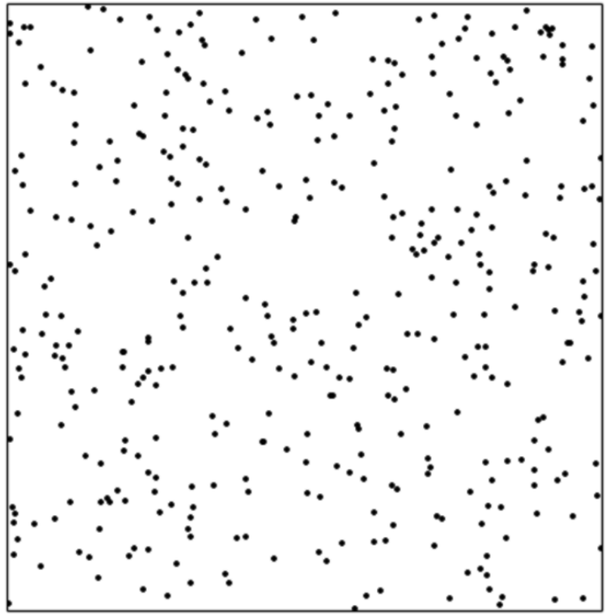
the radio sky is *easy* to look at!



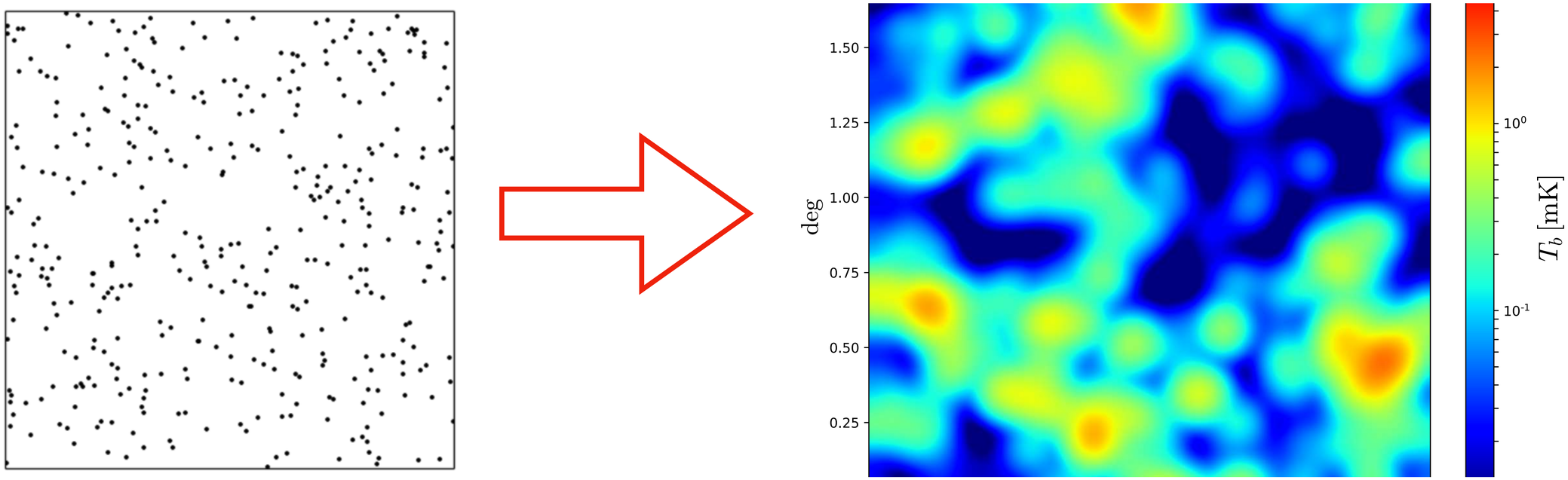
The Arecibo Legacy Fast ALFA Survey



Record z = 0.376 detection of 21cm emitting galaxy with 178 hours from VLA [Fernández et al, 2016]

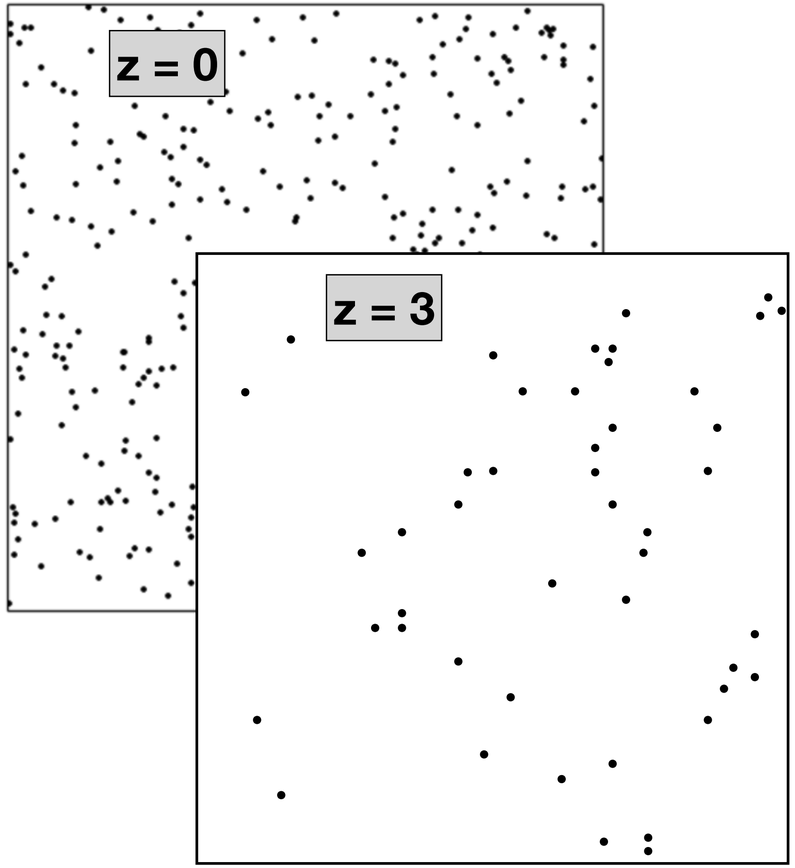


21-cm intensity mapping

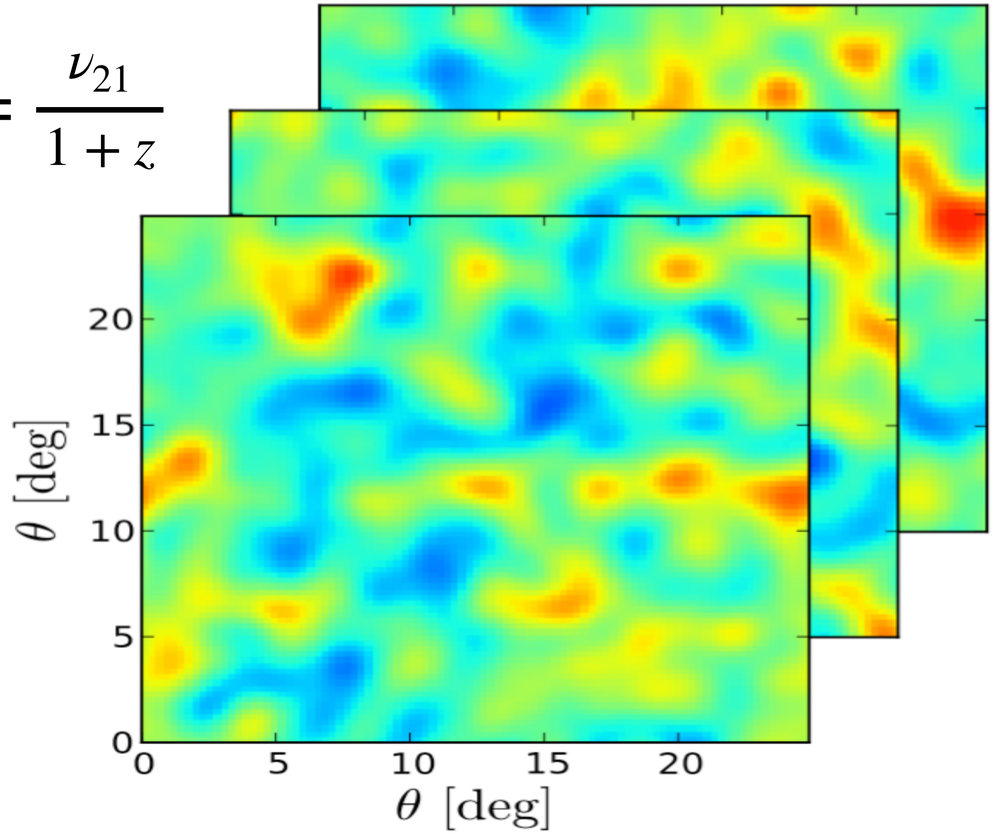


Put signal-to-noise where you really need it: **linear large scale modes**

21-cm intensity mapping



$$\nu = \frac{\nu_{21}}{1+z}$$



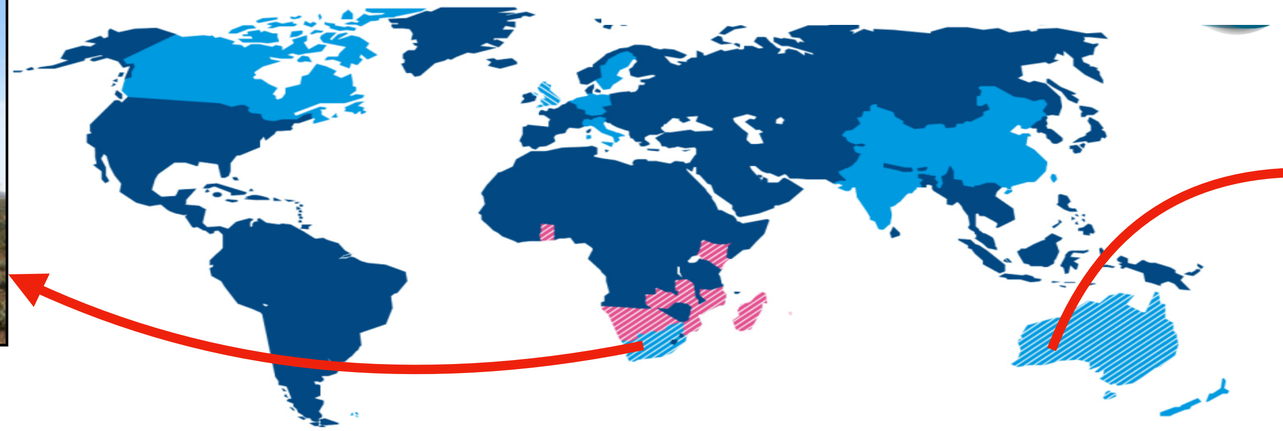
Big volumes (for cheap) and high redshift resolution

A large-scale structure scientist's wish list:

HI Intensity mapping

1. Large areas
2. Deep and accurate redshifts (distances)
3. Better coverage of the Universe epochs

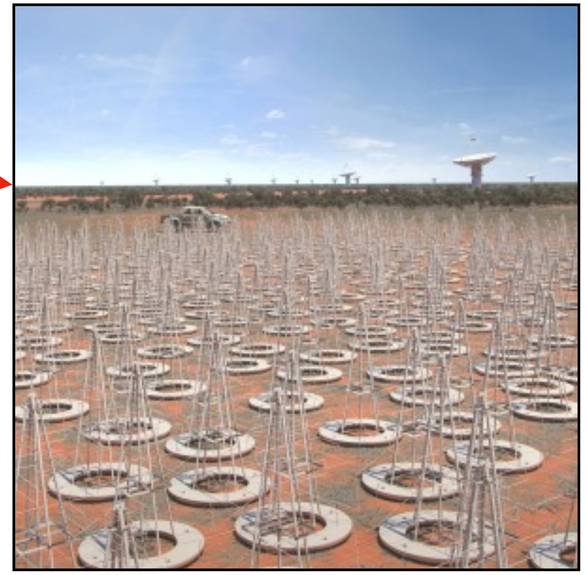
SKAO



Full members
SKA Headquarters host country
SKA Phase 1 and Phase 2 host countries

African partner countries
(non-member SKA Phase 2 host countries)

This map is intended for reference only and is not meant to represent legal borders



0 < z < 3

SKA1-mid
the SKA's mid-frequency instrument

Location:
South Africa

Frequency range:
350 MHz
to
15.3 GHz
with a goal of 24 GHz

197 dishes
(including 64 MeerKAT dishes)

Maximum baseline:
150km

3 < z < 27

SKA1-low
the SKA's low-frequency instrument

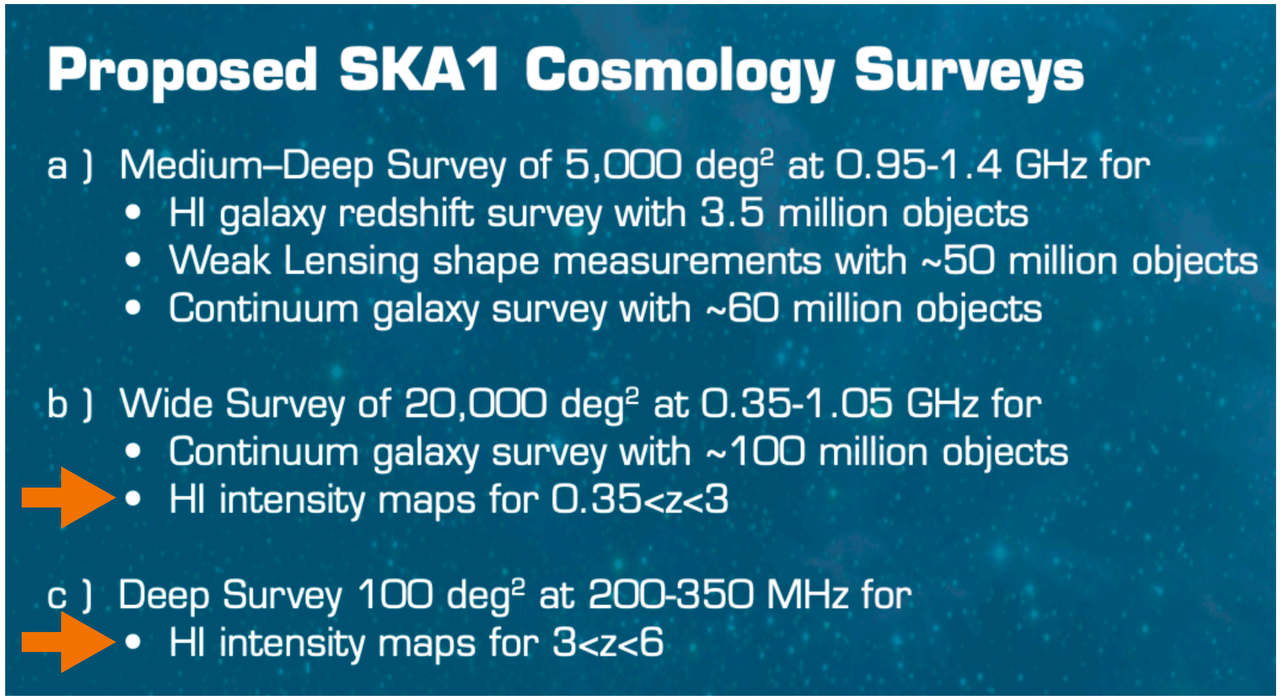
Location: Australia

Frequency range:
50 MHz
to
350 MHz

~131,000
antennas spread between
512 stations

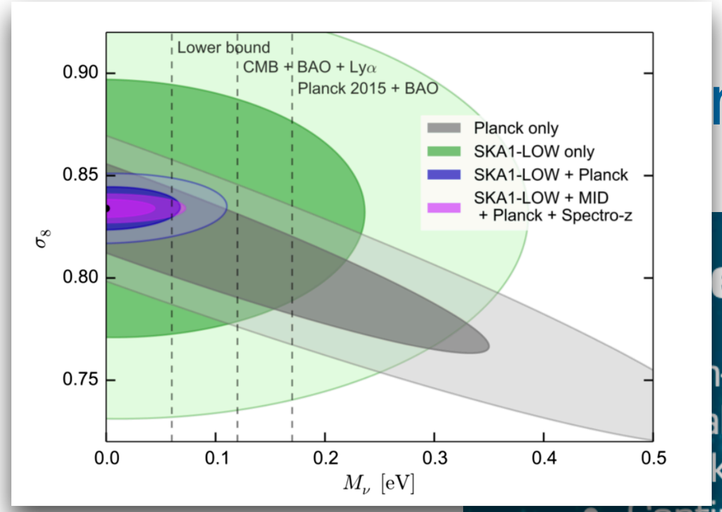
Maximum baseline:
~65km

HI intensity mapping with the SKAO



Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018:
Technical specifications and performance forecasts

nsity mapping with the SKAO

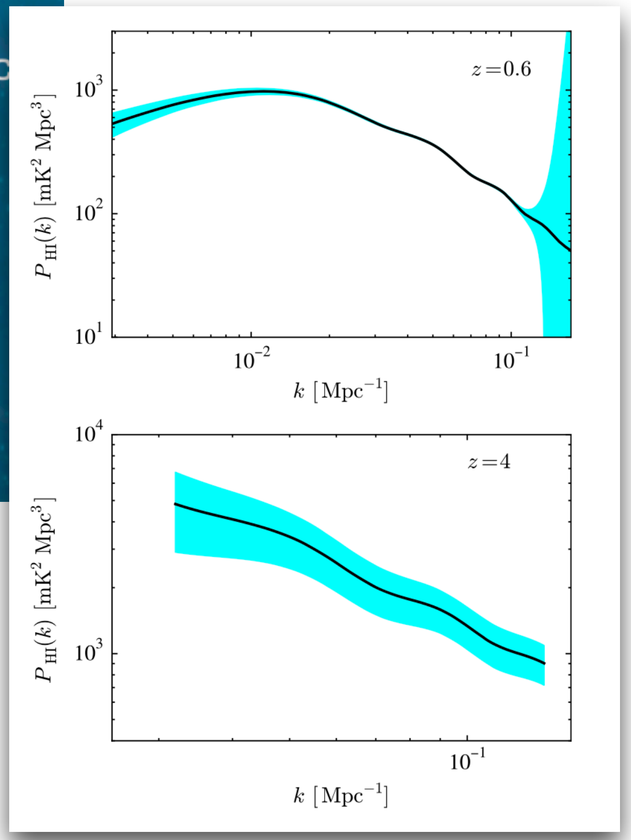


ed SKA1 Cosmology Surveys

–Deep Survey of 5,000 deg² at 0.95-1.4 GHz for
laxy redshift survey with 3.5 million objects
k Lensing shape measurements with ~50 million objec
- Continuum galaxy survey with ~60 million objects

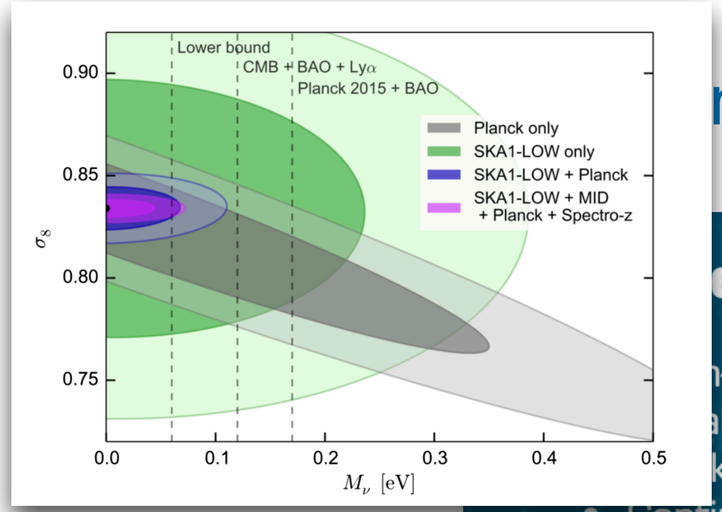
b) Wide Survey of 20,000 deg² at 0.35-1.05 GHz for
- Continuum galaxy survey with ~100 million objects
- HI intensity maps for 0.35<z<3

c) Deep Survey 100 deg² at 200-350 MHz for
- HI intensity maps for 3<z<6



Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018:
Technical specifications and performance forecasts

nsity mapping with

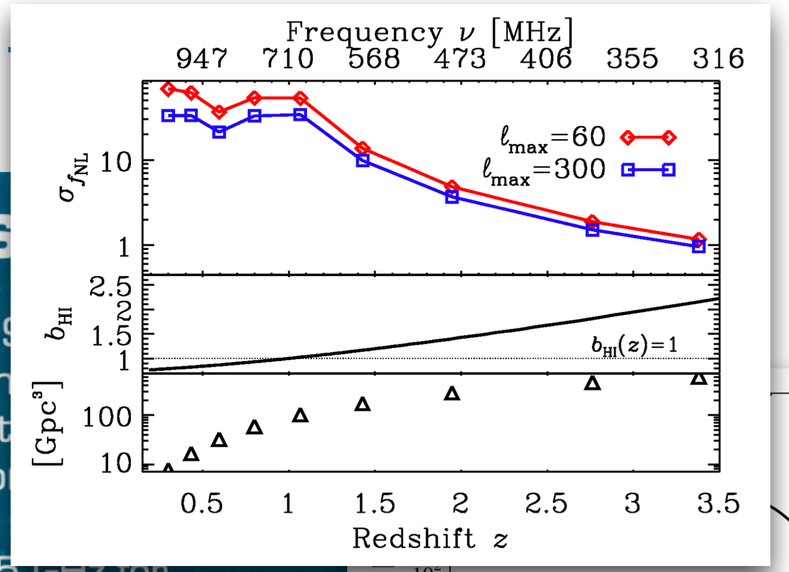


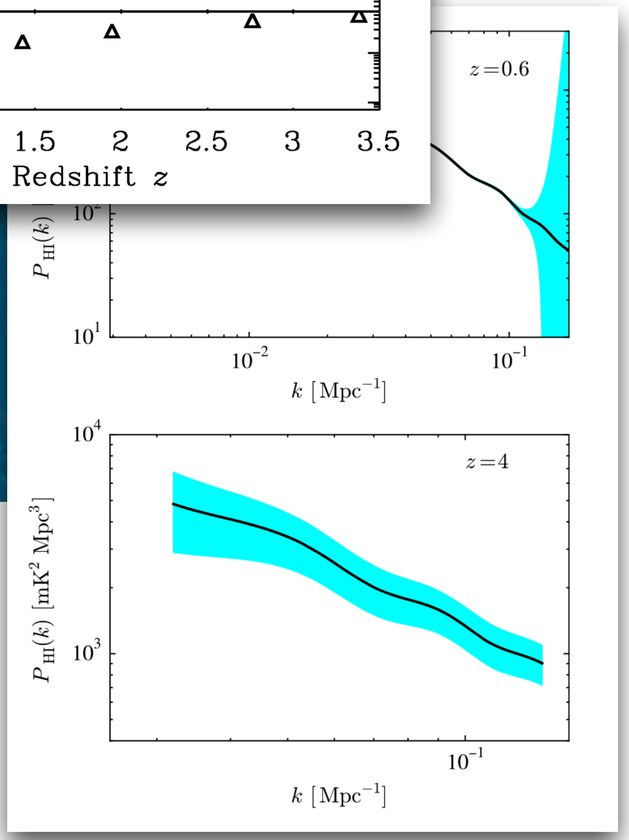
ed SKA1 Cosmology S

–Deep Survey of 5,000 deg² at 0.
laxy redshift survey with 3.5 million
k Lensing shape measurements wit
- Continuum galaxy survey with ~60 millio

b) Wide Survey of 20,000 deg² at 0.35-1.05 GHz for
- Continuum galaxy survey with ~100 million objects
- HI intensity maps for 0.35<z<3

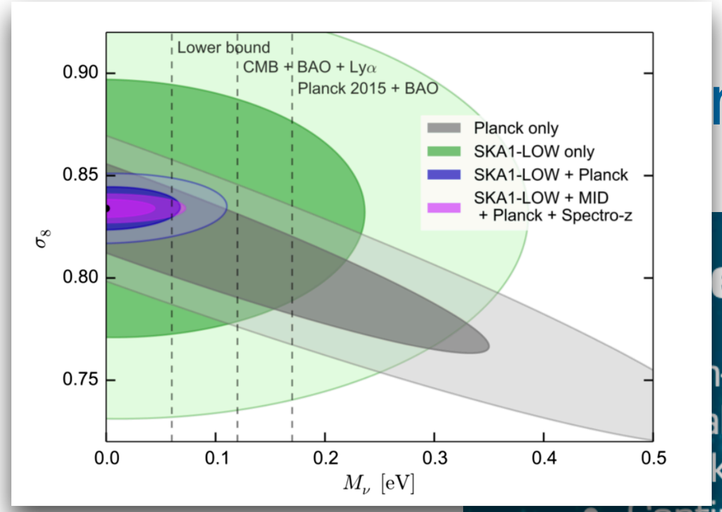
c) Deep Survey 100 deg² at 200-350 MHz for
- HI intensity maps for 3<z<6





Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018:
Technical specifications and performance forecasts

nsity mapping with

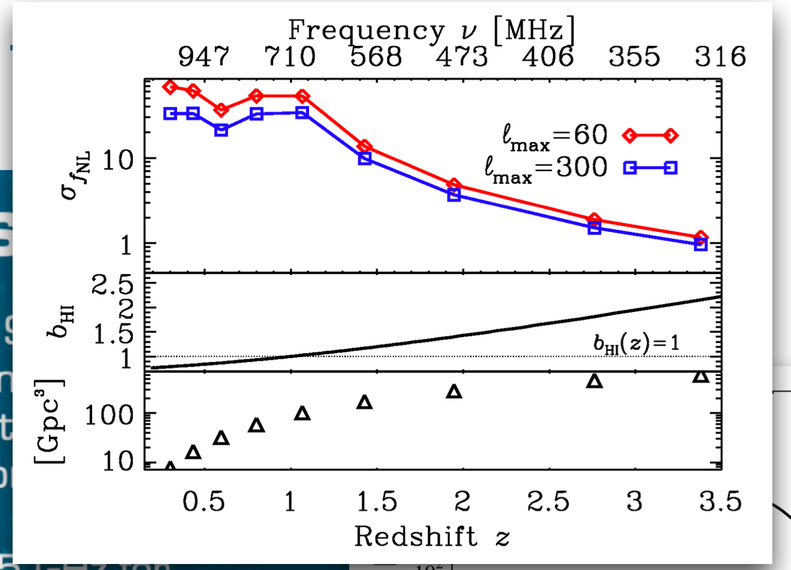


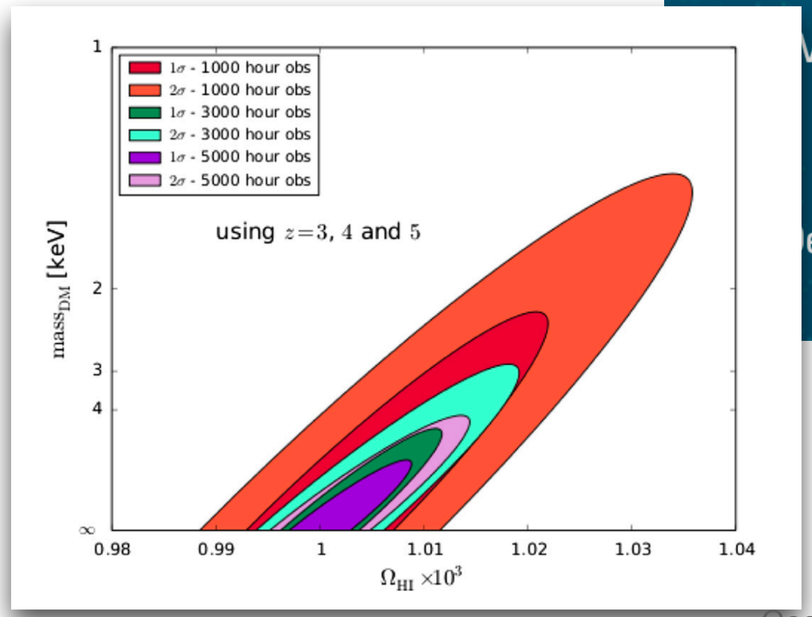
ed SKA1 Cosmology S

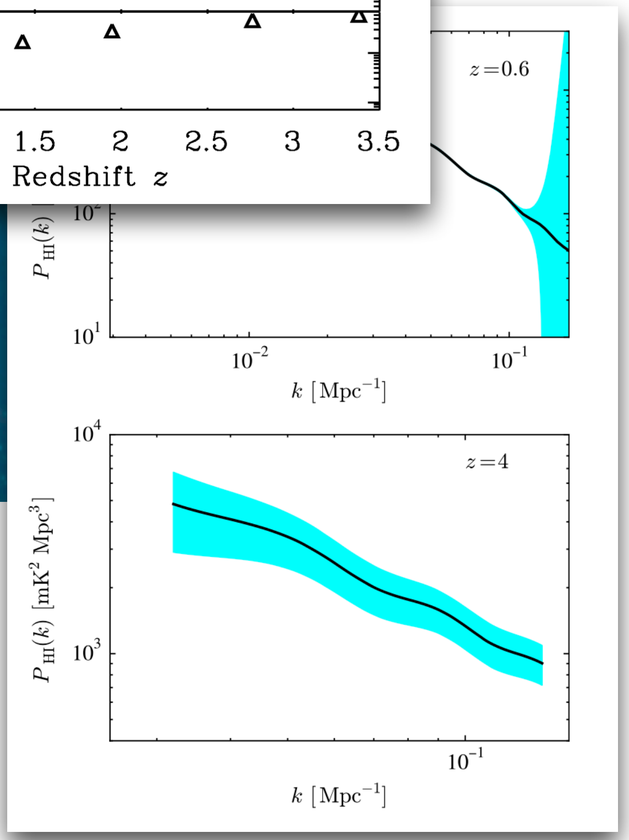
- –Deep Survey of 5,000 deg² at 0.
- laxy redshift survey with 3.5 million
- k Lensing shape measurements wit
- Continuum galaxy survey with ~60 millio

Vide Survey of 20,000 deg² at 0.35-1.05 GHz for
- Continuum galaxy survey with ~100 million objects
- HI intensity maps for 0.35<z<3

eep Survey 100 deg² at 200-350 MHz for
- HI intensity maps for 3<z<6



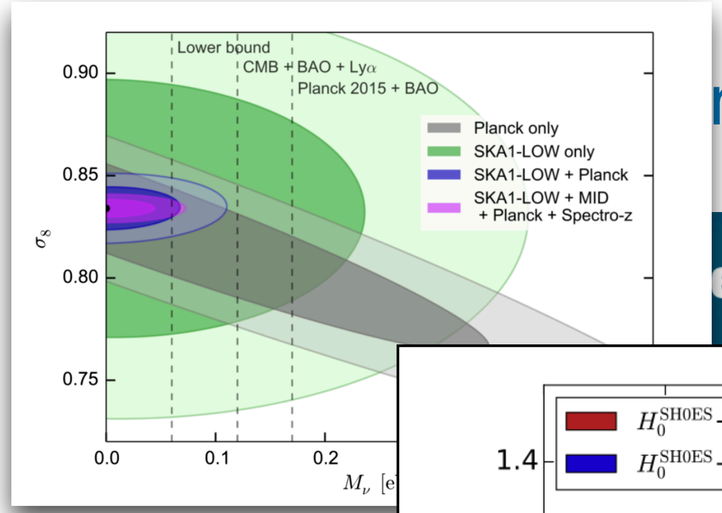


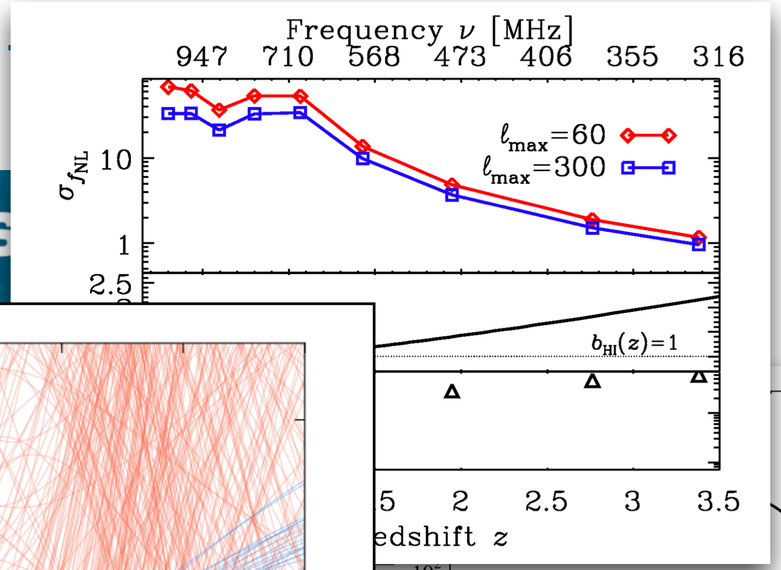


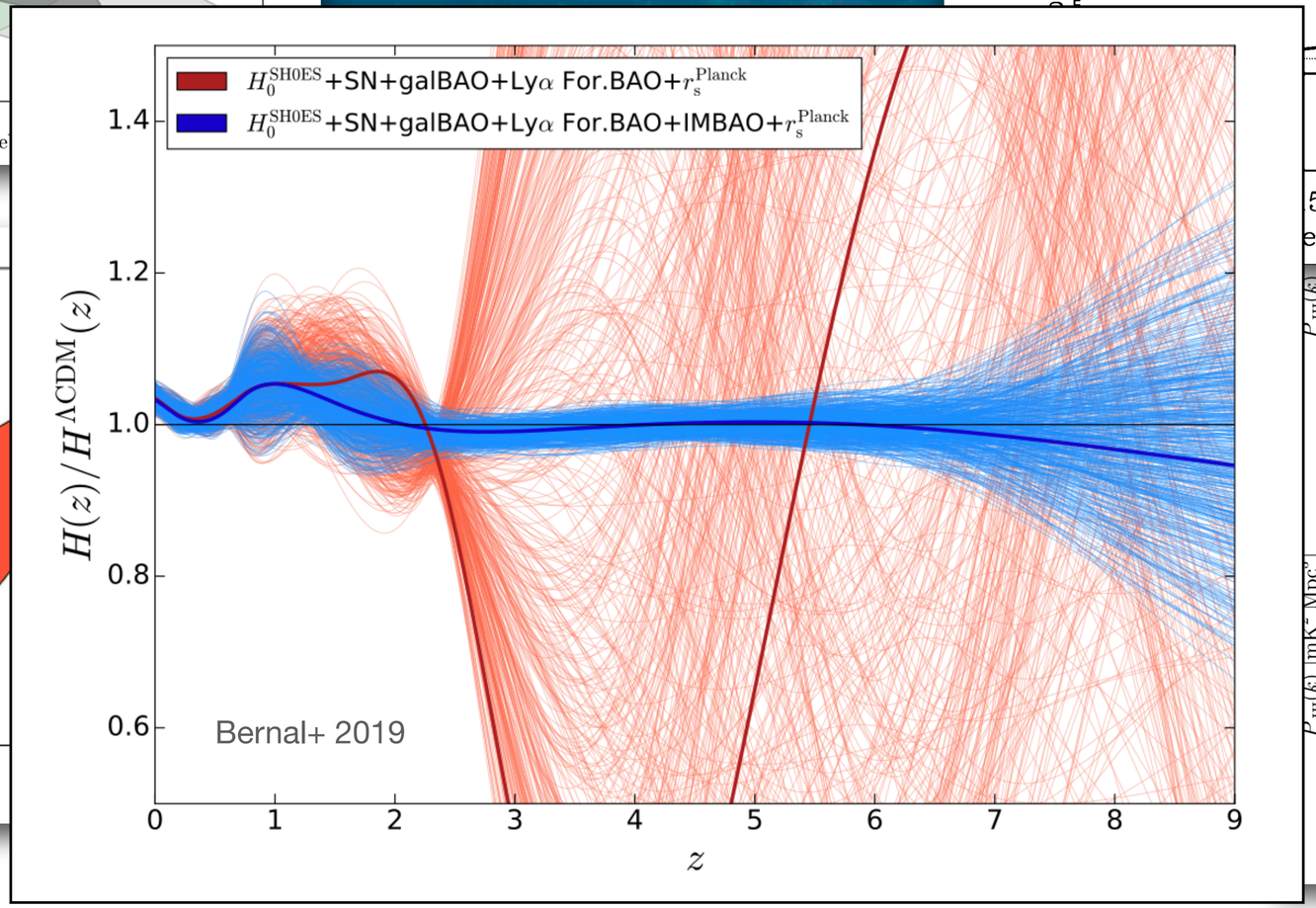
Cosmology with Phase 1 of the Square Kilometre Array **Red Book** 2018: Technical specifications and performance forecasts

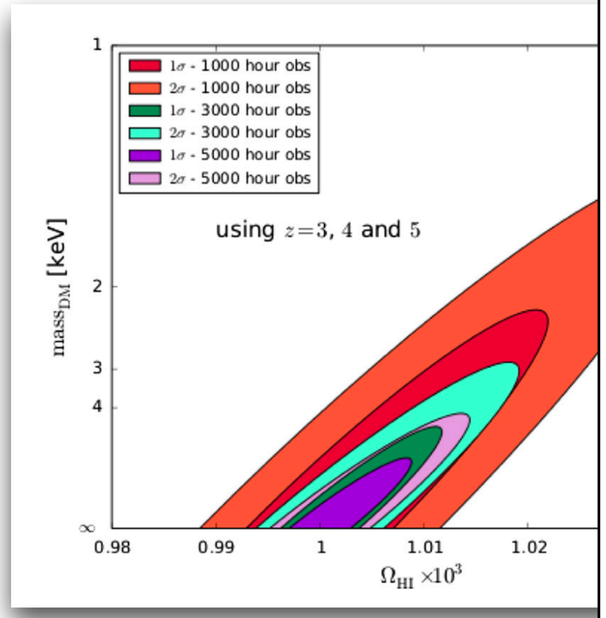
nsity mapping with

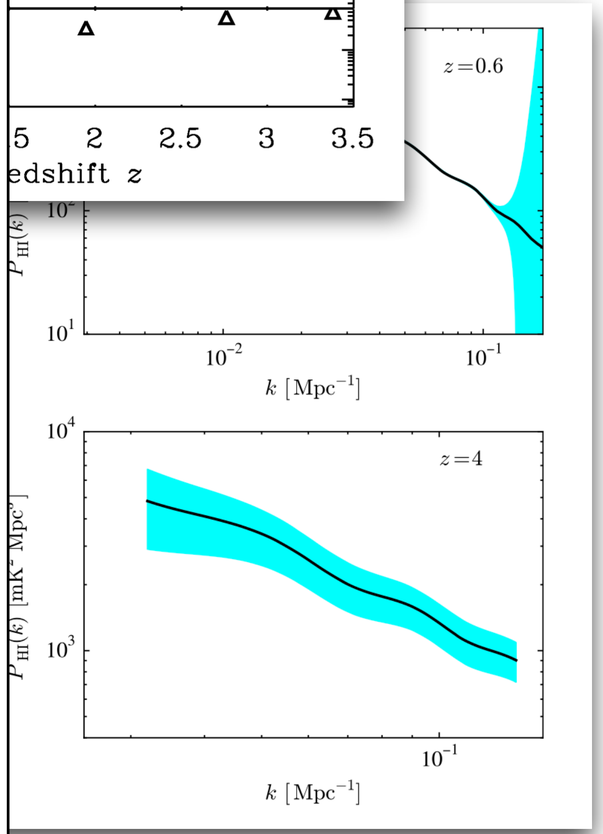
ed SKA1 Cosmology S





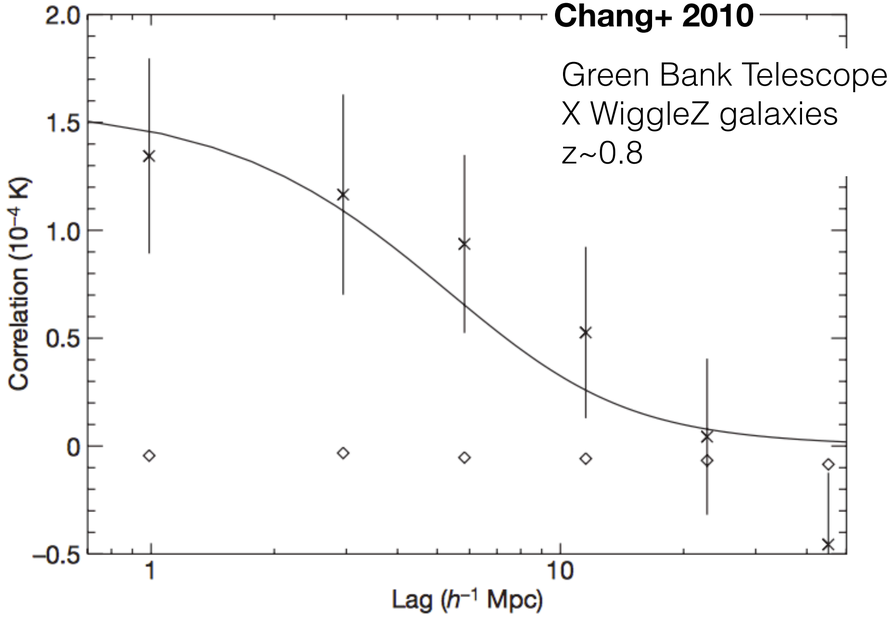




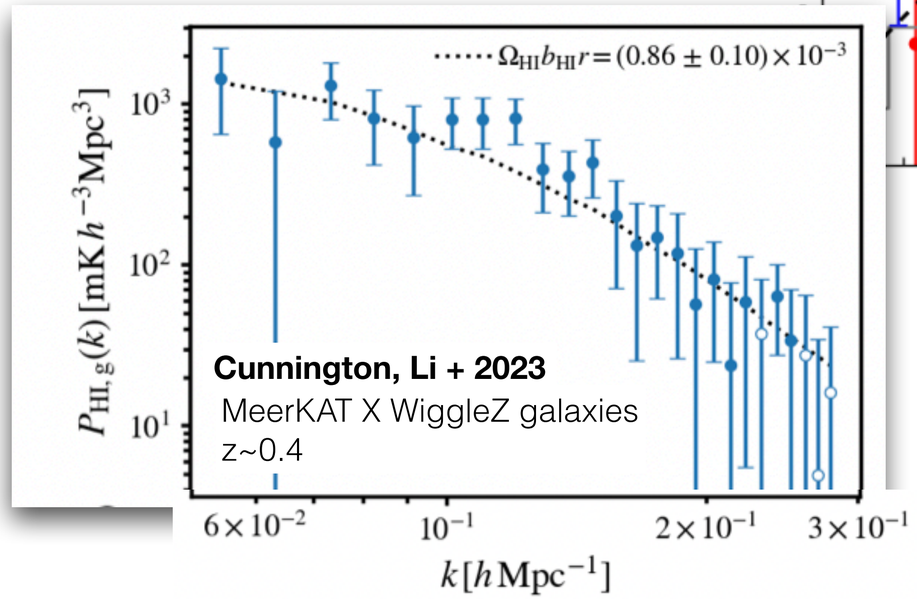


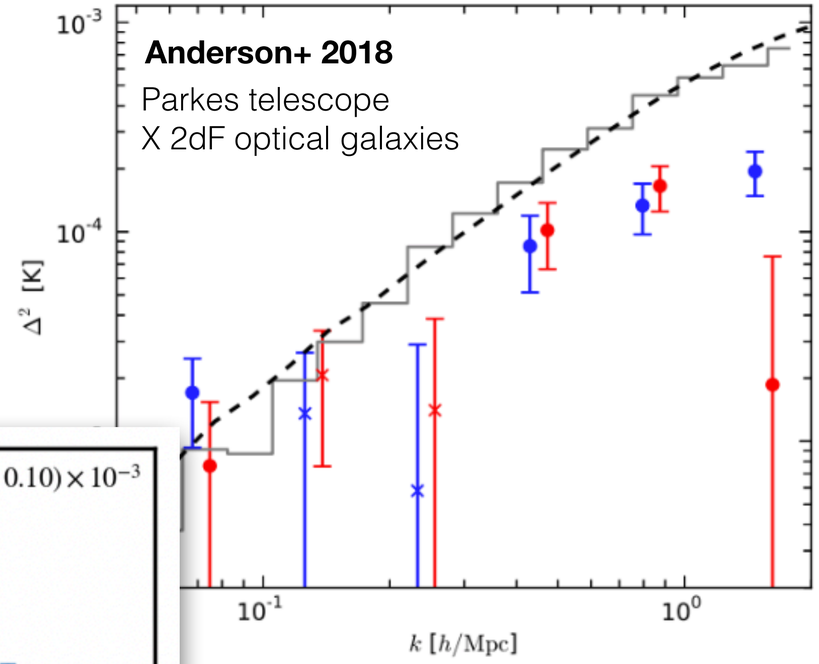
HI intensity mapping

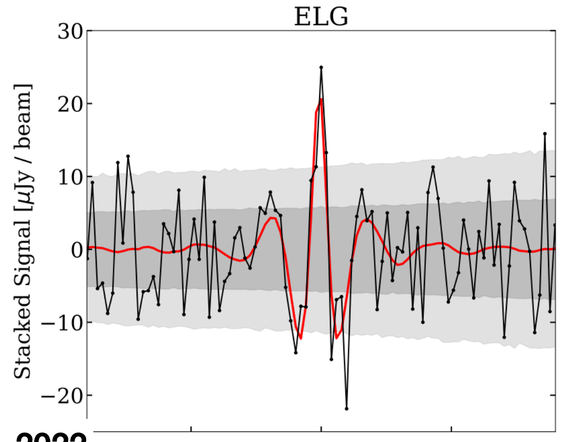
State-of-the-art



also Masui+ 2013, Switzer+ 2013, Wolz+ 2017,2022







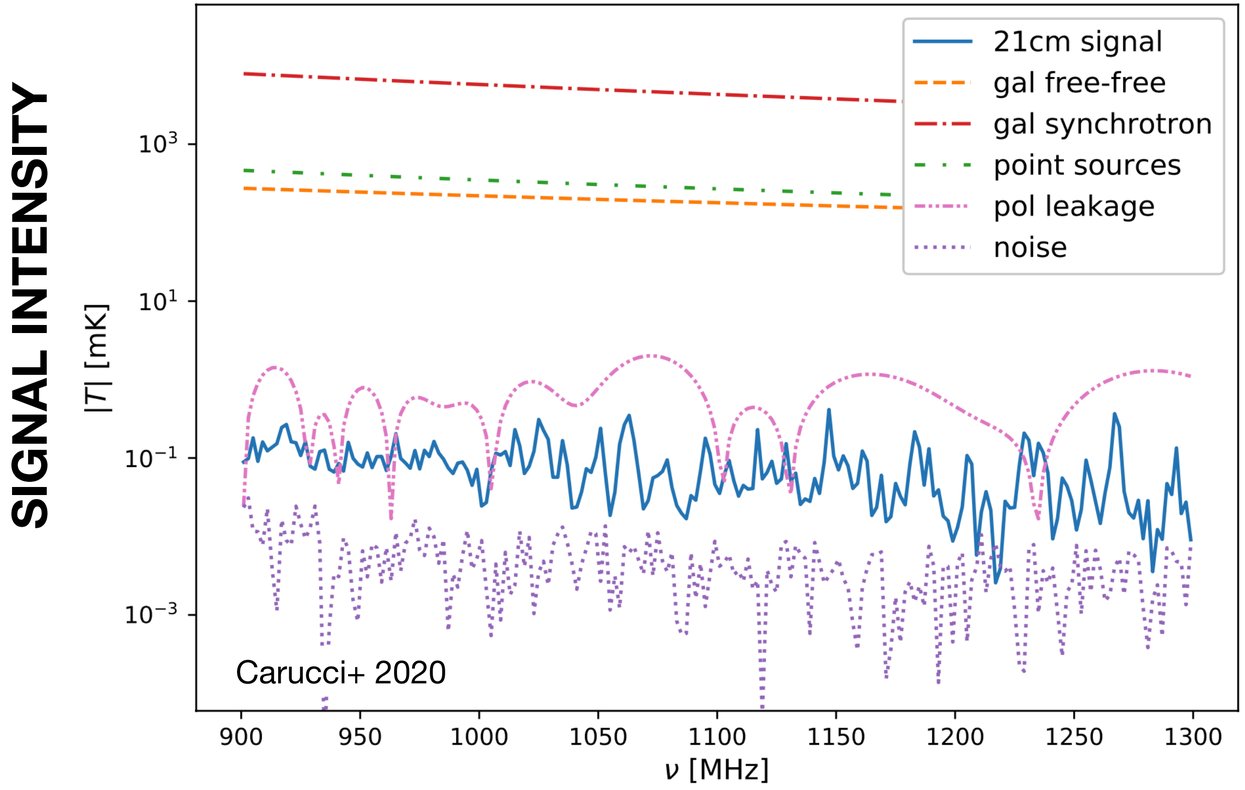
CHIME collaboration, 2022

stacking LRGs, ELG and QSOs from eBOSS
0.8<z<1.5

Contaminants are THE challenge to overcome with HI intensity mapping

HI intensity mapping

buried under the contaminants

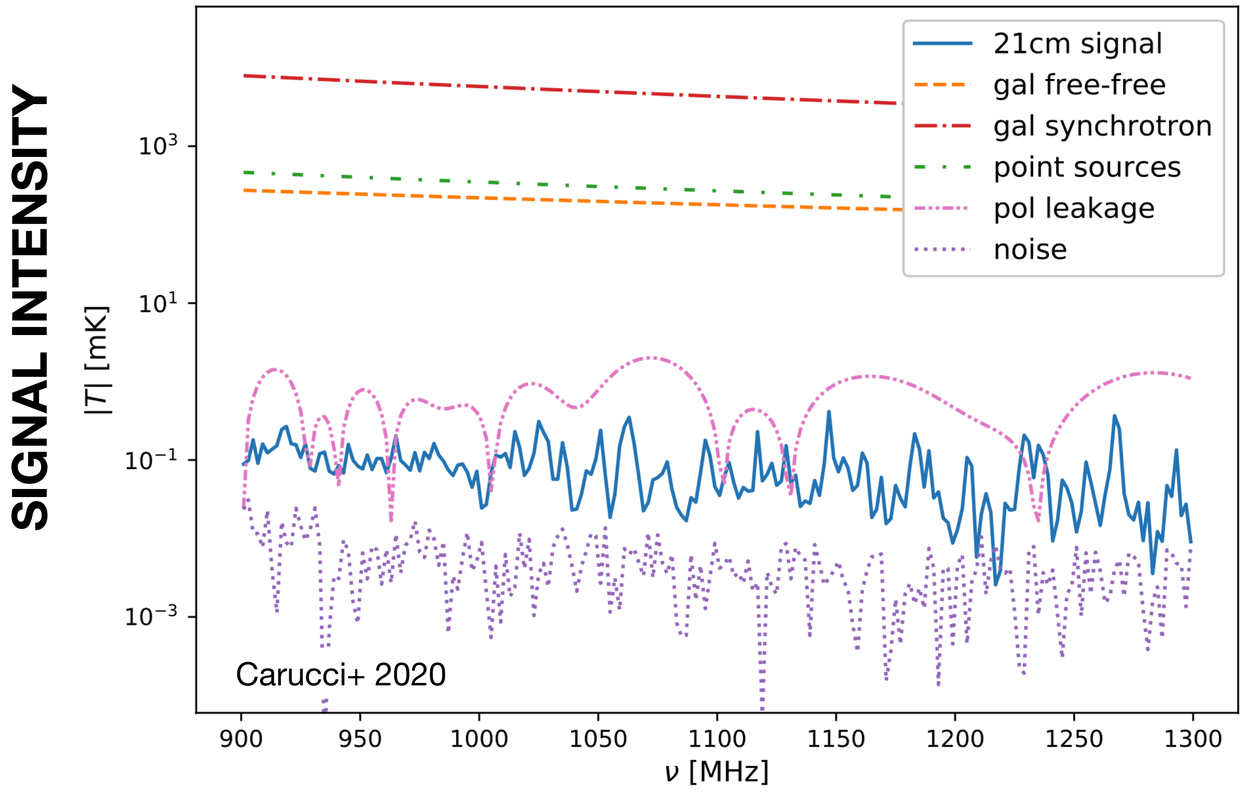


CHALLENGES:

- Foregrounds
- Systematics

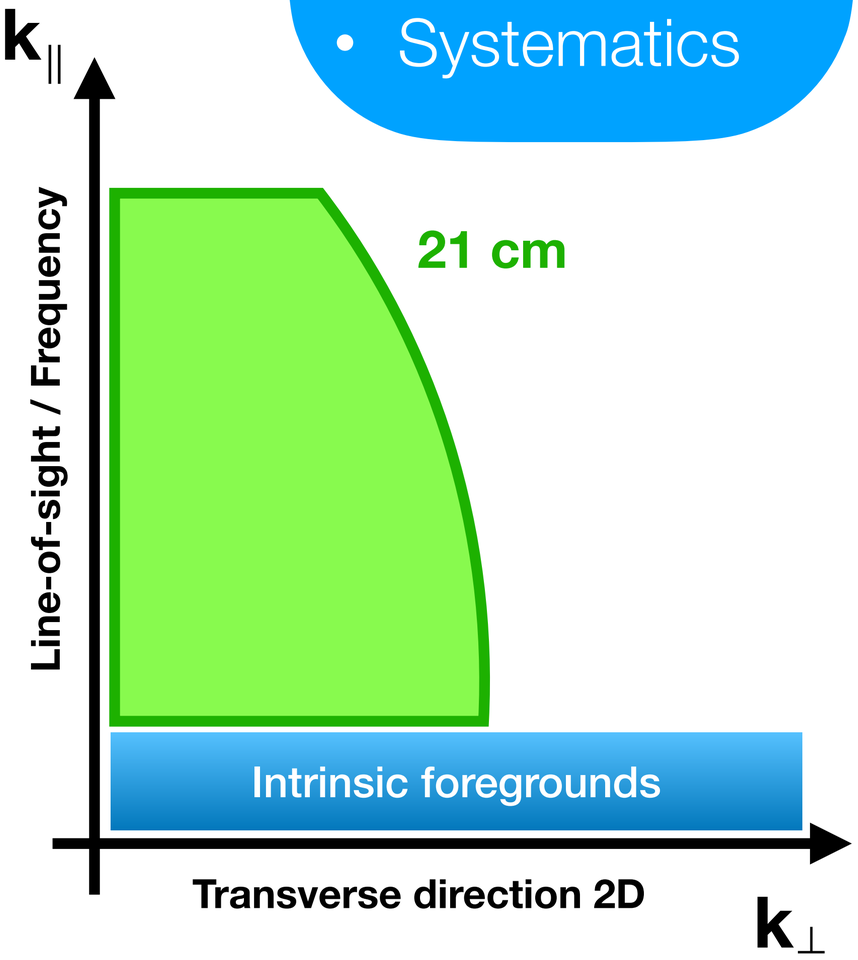
HI intensity mapping

buried under the contaminants



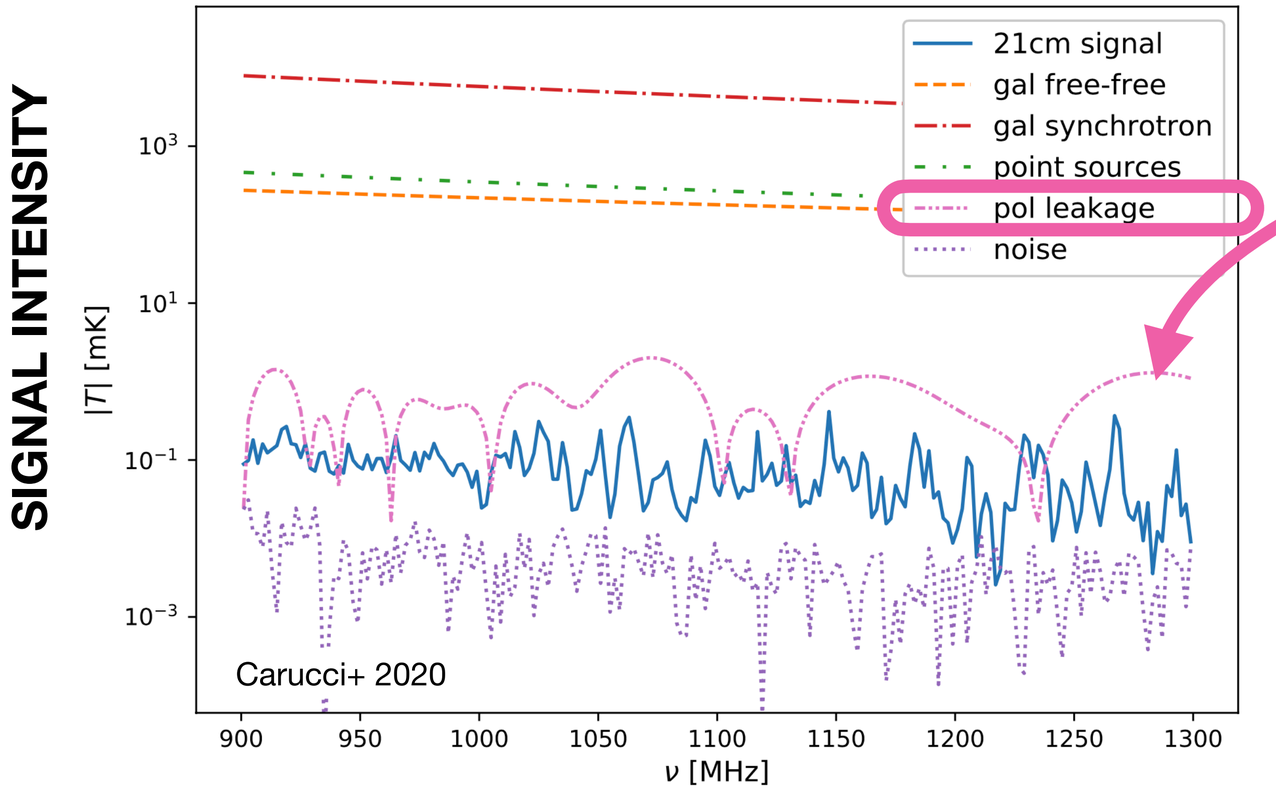
CHALLENGES:

- Foregrounds
- Systematics



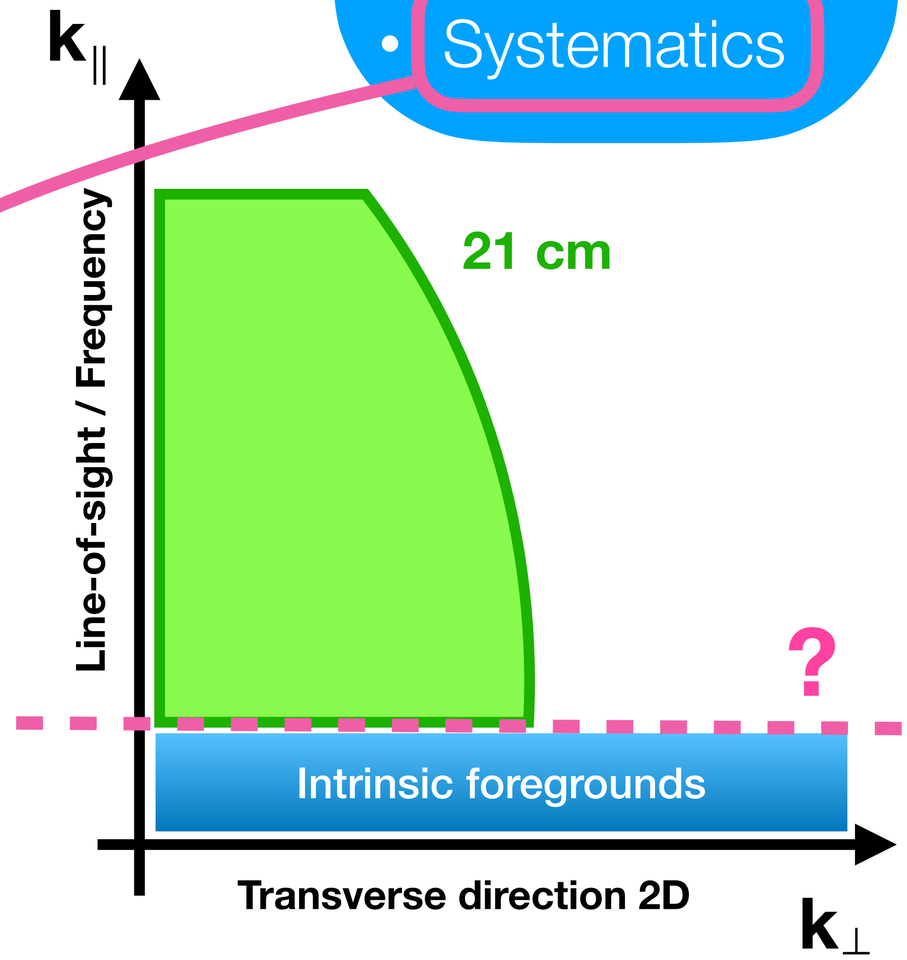
HI intensity mapping

buried under the contaminants



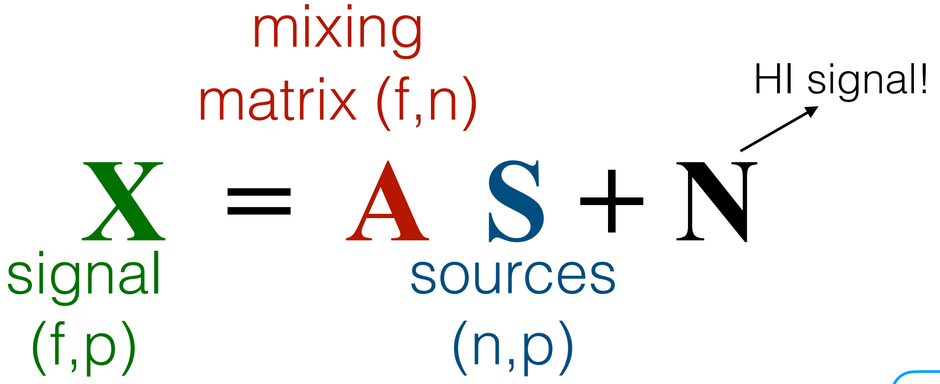
CHALLENGES:

- Foregrounds
- Systematics



Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info about the source signal or the mixing process.



- **Decorrelation** —> diagonalise the covariance matrix
- **Independence** —> as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.

Principal Component Analysis (**PCA**)

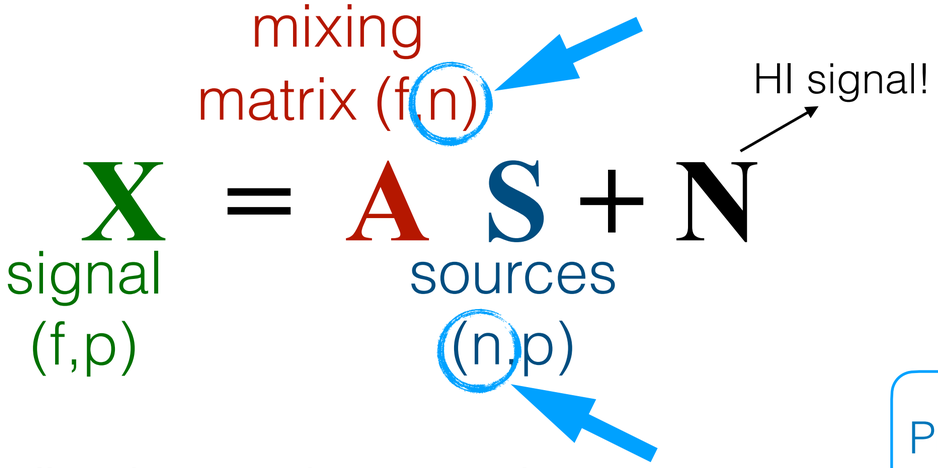
Independent Component Analysis (**ICA**)

Hyvarinen + 1999

Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info about the source signal or the mixing process.

Need to set number *n* of sources!



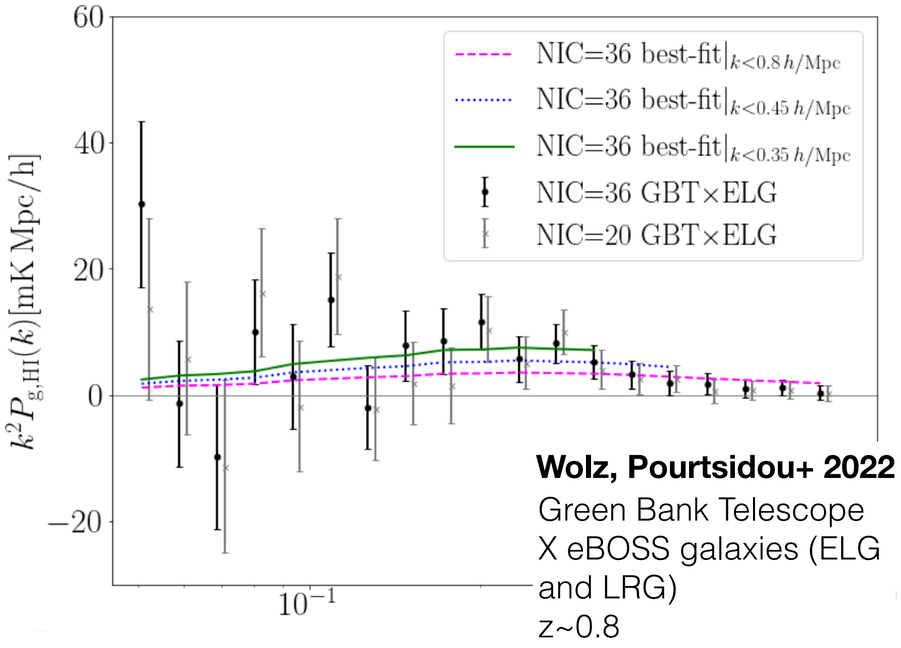
- **Decorrelation** —> diagonalise the covariance matrix
- **Independence** —> as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.

Principal Component Analysis (**PCA**)

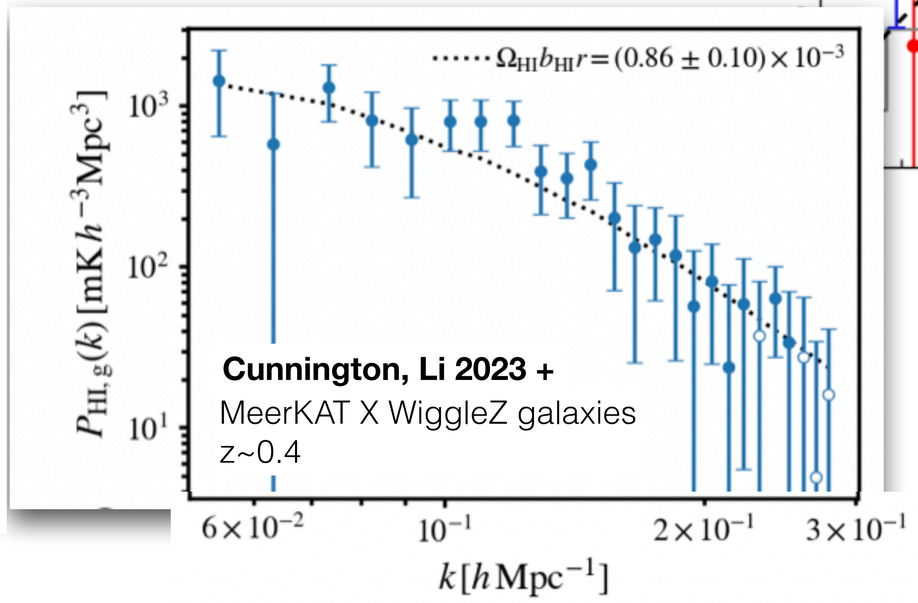
Independent Component Analysis (**ICA**) Hyvarinen + 1999

HI intensity mapping

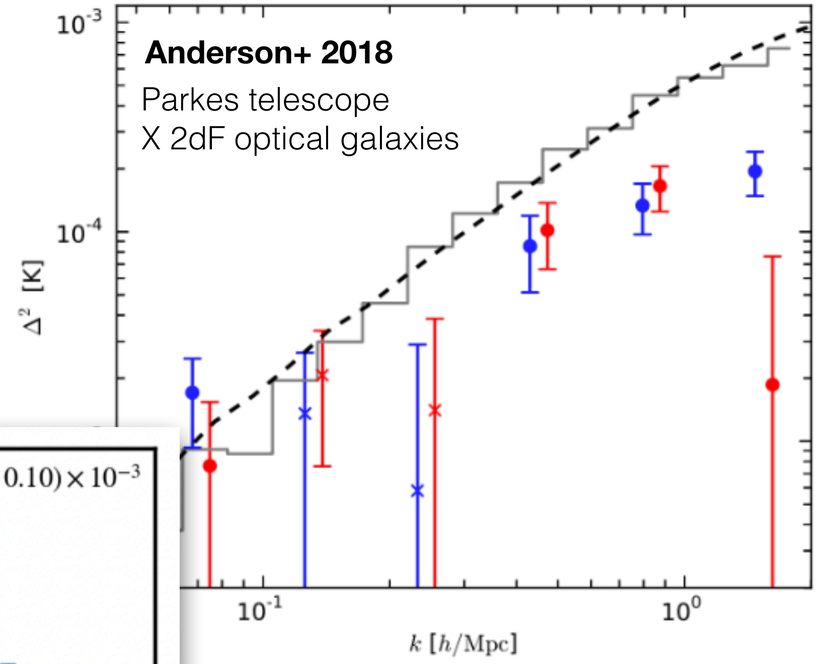
State-of-the-art



Wolz, Pourtsidou+ 2022
Green Bank Telescope
X eBOSS galaxies (ELG and LRG)
z~0.8



Cunnington, Li 2023 +
MeerKAT X WiggleZ galaxies
z~0.4

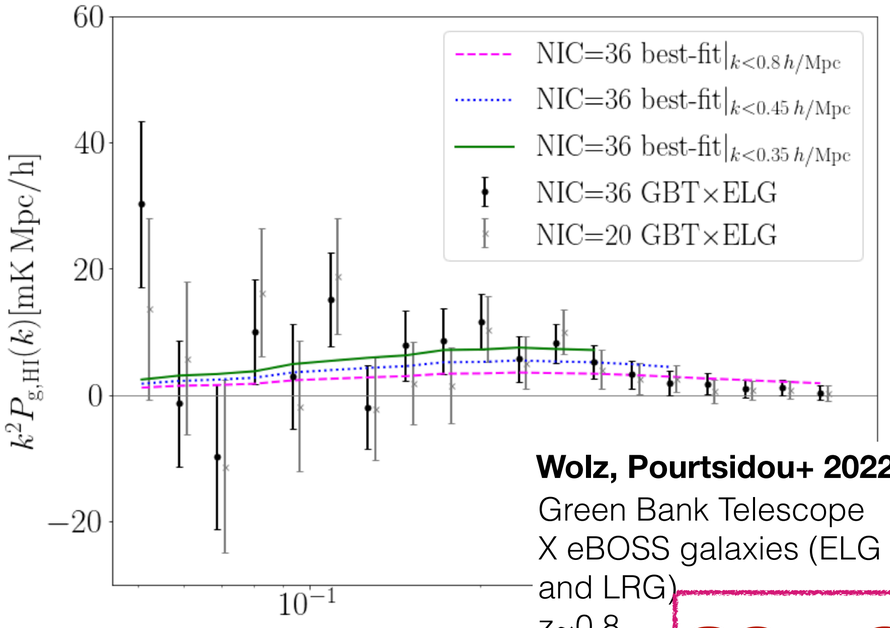


Anderson+ 2018
Parkes telescope
X 2dF optical galaxies

See also Masui+ 2013, Switzer+ 2013, Wolz+ 2017

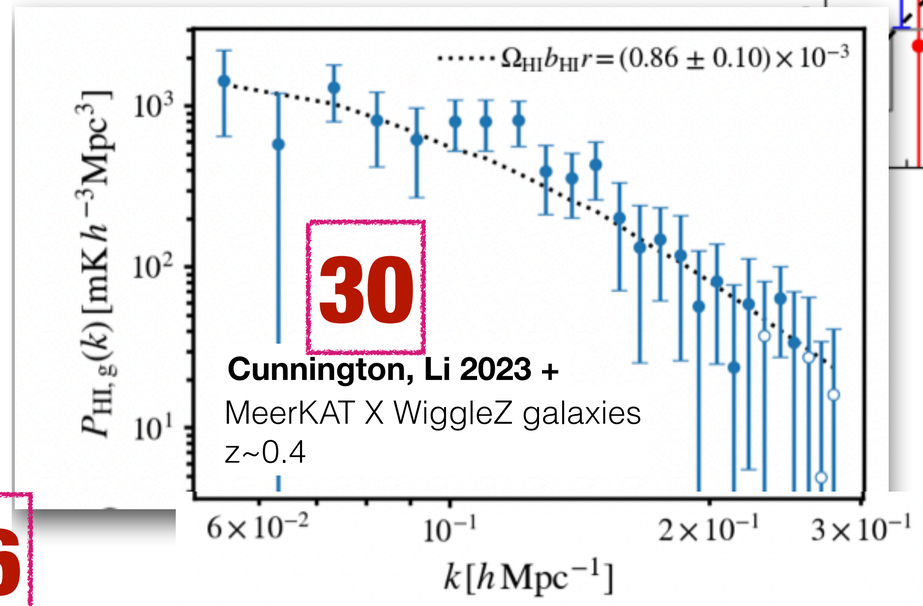
HI intensity mapping

State-of-the-art



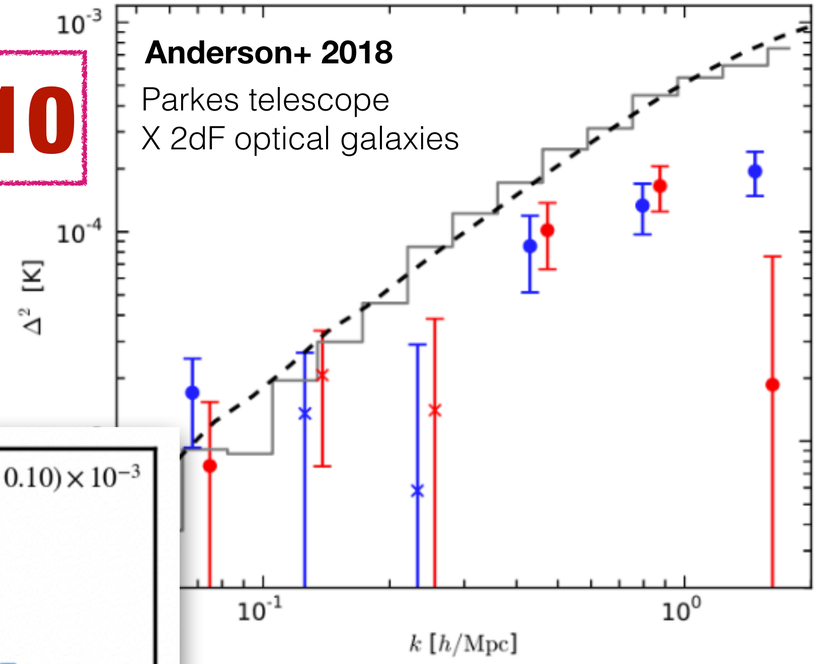
Wolz, Pourtsidou+ 2022
Green Bank Telescope
X eBOSS galaxies (ELG and LRG)
z~0.8

20 - 36



30

Cunnington, Li 2023 +
MeerKAT X WiggleZ galaxies
z~0.4



10

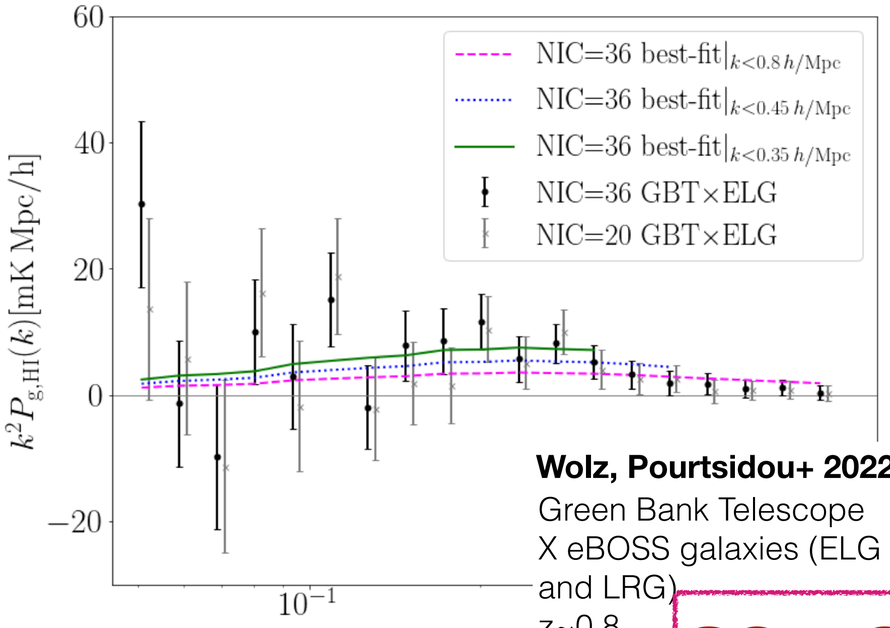
Anderson+ 2018
Parkes telescope
X 2dF optical galaxies

See also Masui+ 2013, Switzer+ 2013, Wolz+ 2017

10 - 20

HI intensity mapping

State-of-the-art

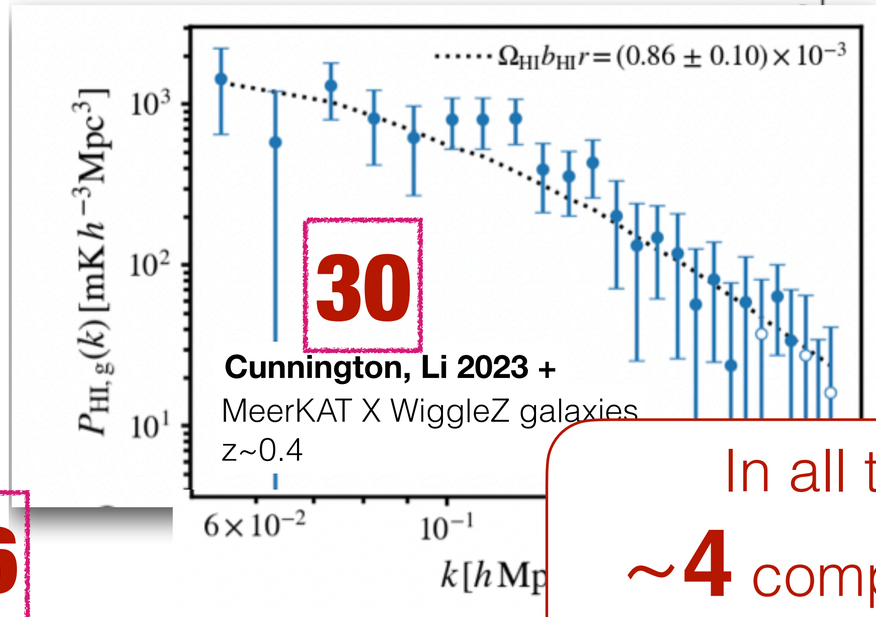


Wolz, Pourtsidou+ 2022
Green Bank Telescope
X eBOSS galaxies (ELG and LRG)
z~0.8

20 - 36

See also Masui+ 2013, Switzer+ 2013, Wolz+ 2017

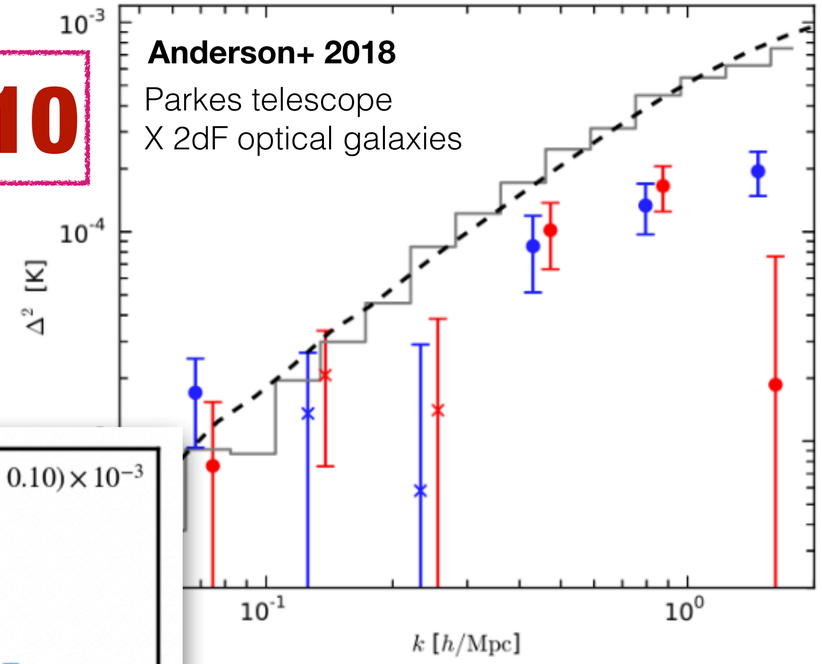
10 - 20



30

Cunnington, Li 2023 +
MeerKAT X WiggleZ galaxies
z~0.4

10



Anderson+ 2018
Parkes telescope
X 2dF optical galaxies

In all theoretical works:

~4 components removed are enough

(e.g., Wolz+ 2014, Alonso+ 2015, Cunnington+ 2019, …)

1. **Hydrogen Intensity Mapping (IM):** what is it and why to do it

2. **IM is hard!** Biggest challenge: weakness of the IM signal compared to contaminants

3. We are getting there. **MeerKLASS**

Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info about the source signal or the mixing process.

mixing matrix (f,n)

HI signal!

$$\mathbf{X} = \mathbf{A}\ \mathbf{S} + \mathbf{N}$$

signal (f,p)

sources (n,p)

- **Decorrelation** —> diagonalise the covariance matrix
- **Independence** —> as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.
- **Sparsity** —> mixtures are less sparse than sources!

Principal Component Analysis (**PCA**)

Independent Component Analysis (**ICA**)

Generalised Morphological Component Analysis (**GMCA**)

Bobin + 2007, 2008, 2012

Blind Source Separation algorithms

The separation of a set of source signals (contaminants) from a set of mixed signals (the maps), with little or no info ab

mixing matrix (f,n)

$$\underset{\text{signal (f,p)}}{\mathbf{X}} = \underset{\text{sources (n,p)}}{\mathbf{A}\ \mathbf{S}} + \mathbf{N}$$

HI signa

Tested on **CMB** data (e.g. Bobin+ 2016)

on **EoR** signal (Patil+ 2017)

on **X-ray** images of **Supernova remnants** (Picquenot+ 2019)

- **Decorrelation** —> diagonalise the covariance matrix
- **Independence** —> as more independent sources are mixed the signal becomes more Gaussian (central limit theorem). So, let's maximise the non-gaussianity of the sources to *unmix* them.
- **Sparsity** —> mixtures are less sparse than sources!

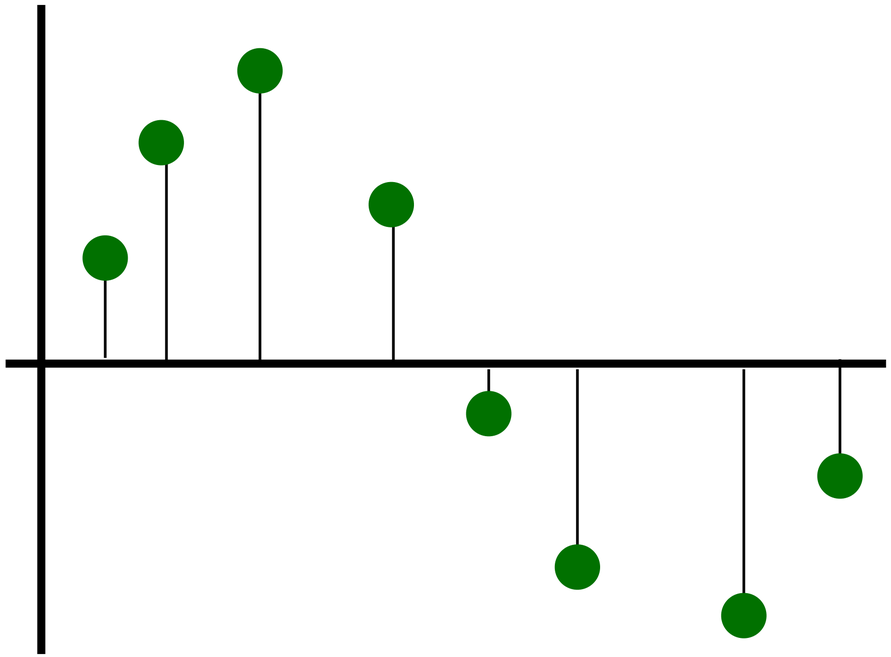
Principal Component Analysis (**PCA**)

Independent Component Analysis (**ICA**)

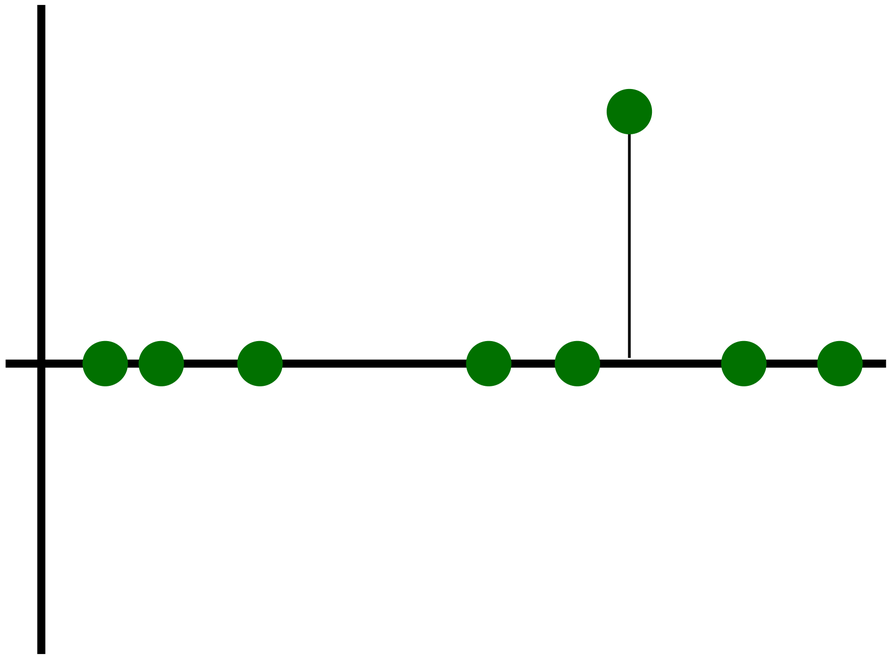
Generalised Morphological Component Analysis (**GMCA**)

Bobin + 2007, 2008, 2012

sparsity

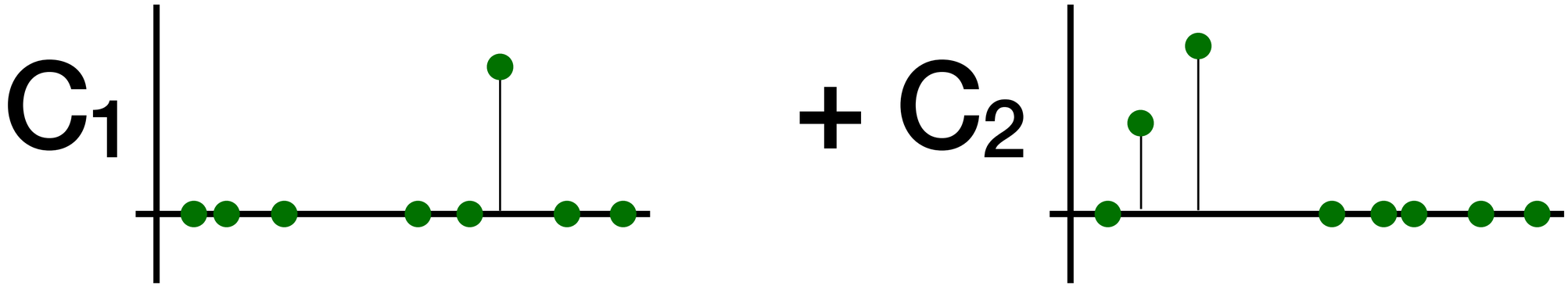


sparsity



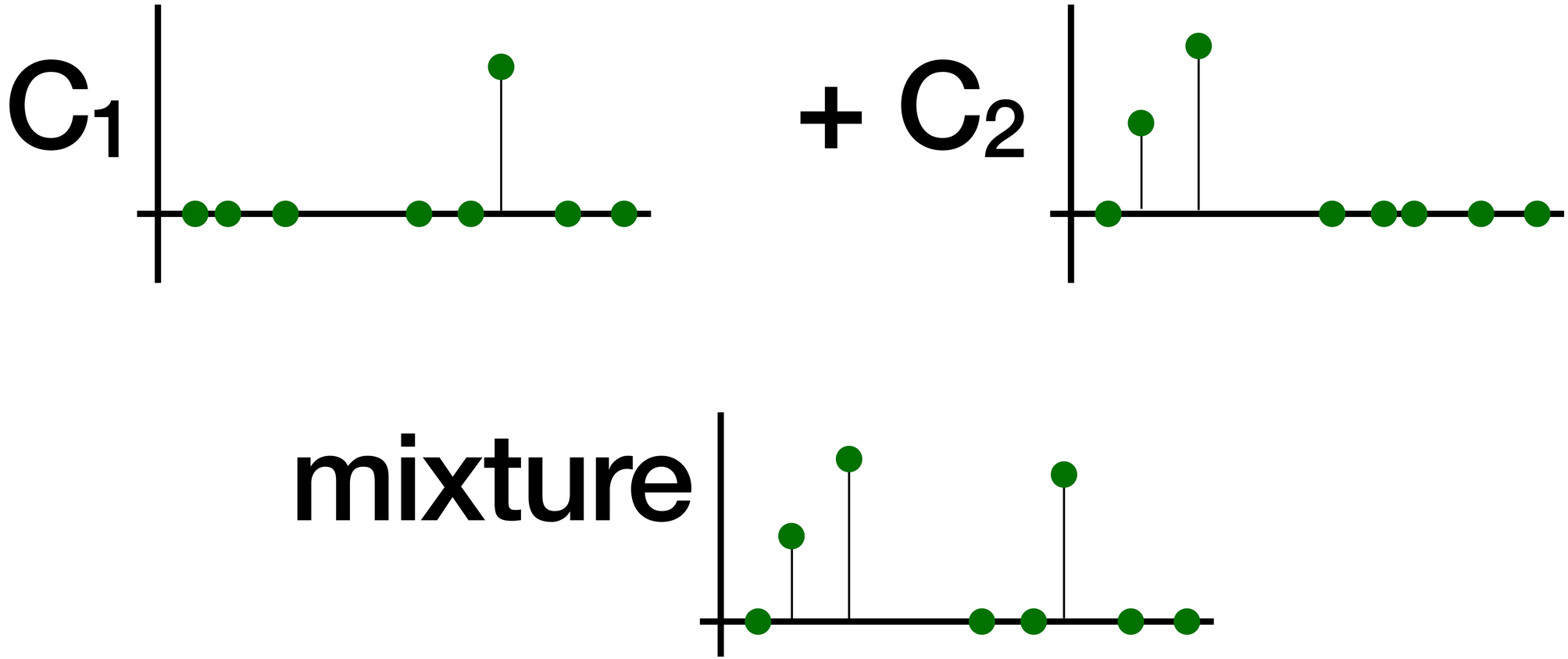
why sparsity?

mixtures are less sparse than components

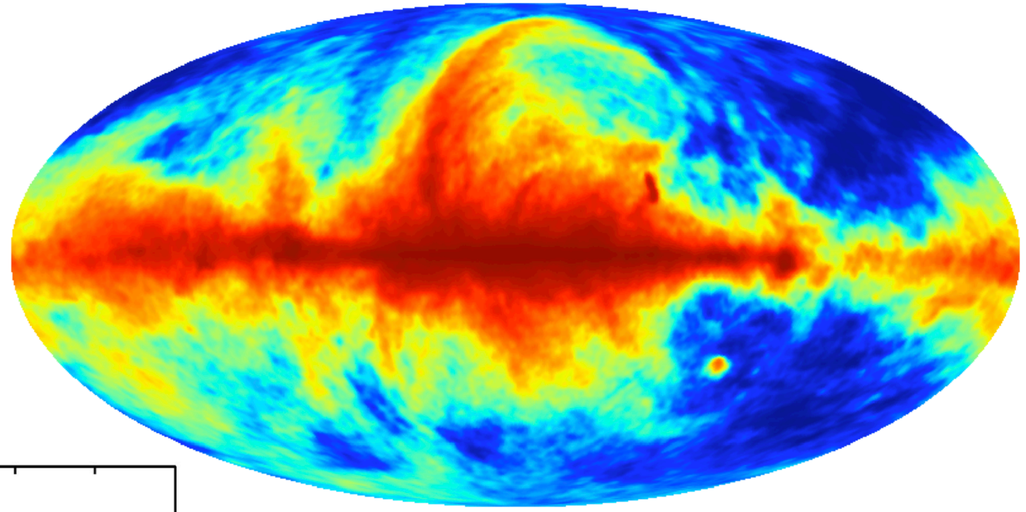


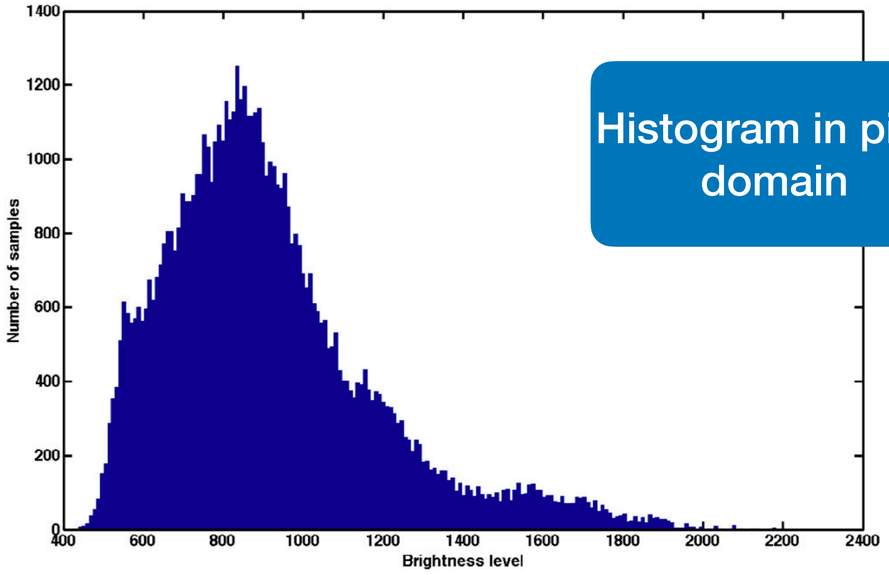
why sparsity?

mixtures are less sparse than components



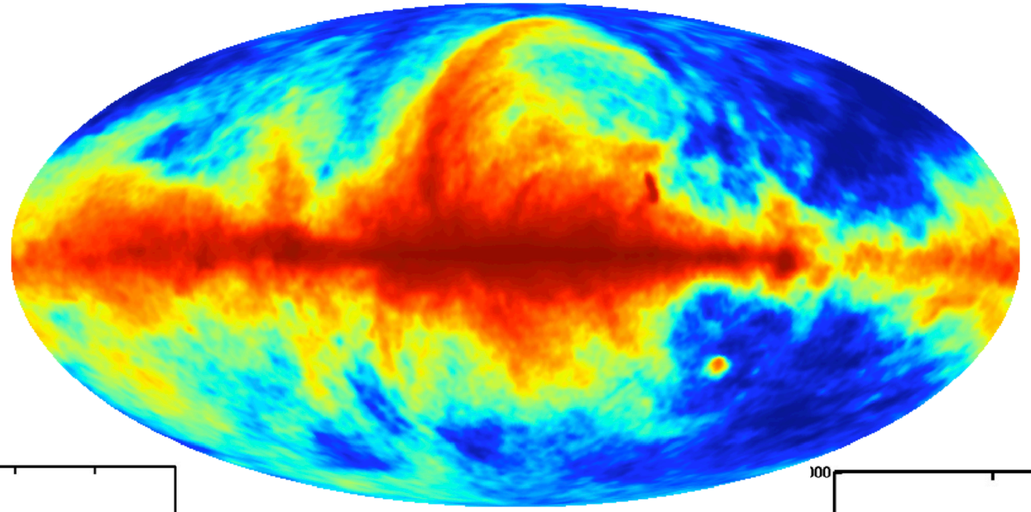
Enforcing sparsity: in which domain?

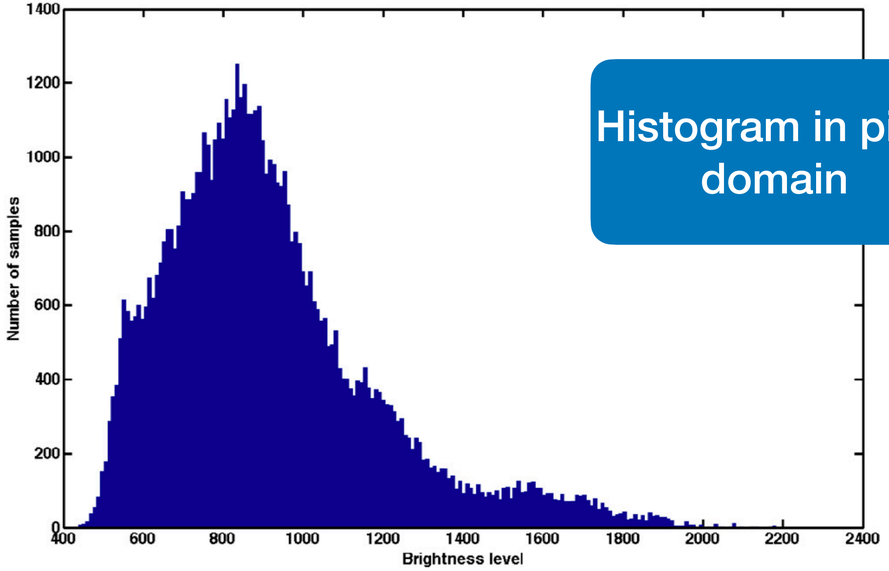




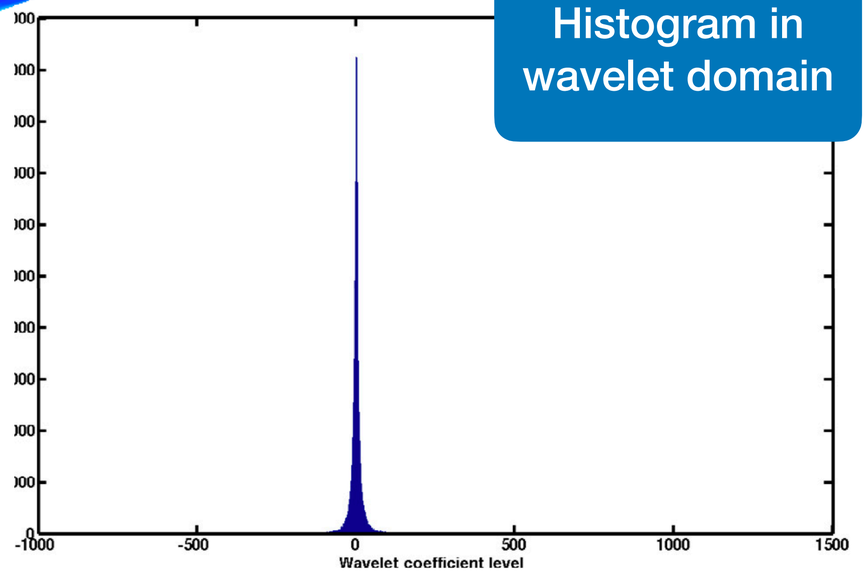
Histogram in pixel domain

Enforcing sparsity: in which domain?





Histogram in pixel domain



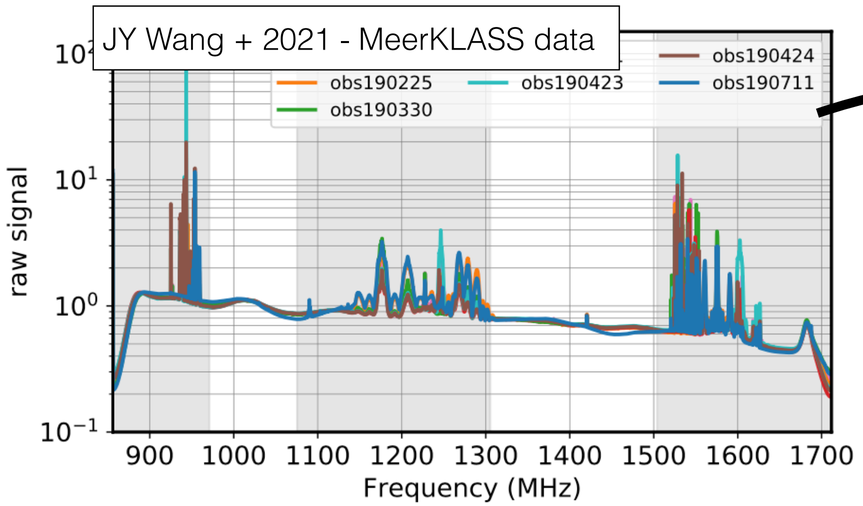
Histogram in wavelet domain

Sparsity-based component-separation for 21-cm IM

GMCA: Generalised Morphological Component Analysis

Bobin+ 2007, 2008, 2012,.. Applied on data in different astro-context: CMB (e.g. Bobin+2016), EoR (e.g. Hothi+2020), X-ray (Picquenot+2019), …

- wavelet decomposition —> **multi-scale** approach
- **No priors on signal**

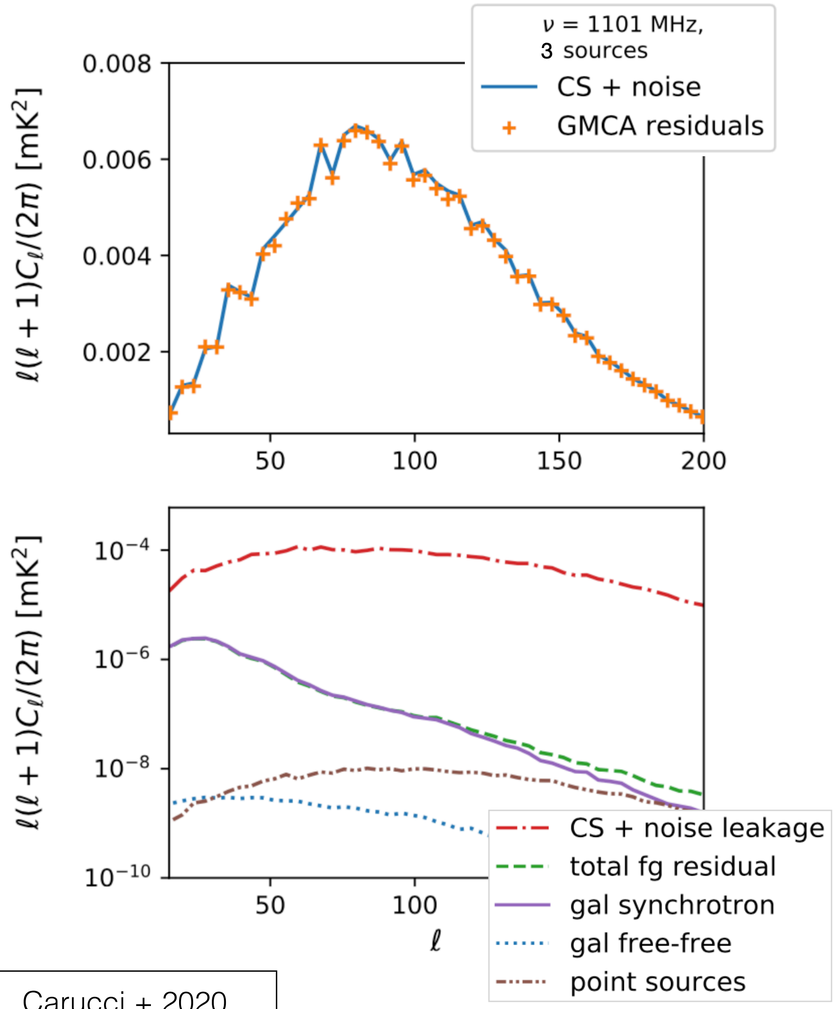


in Carucci+ 2020,
for the fist time in the literature:

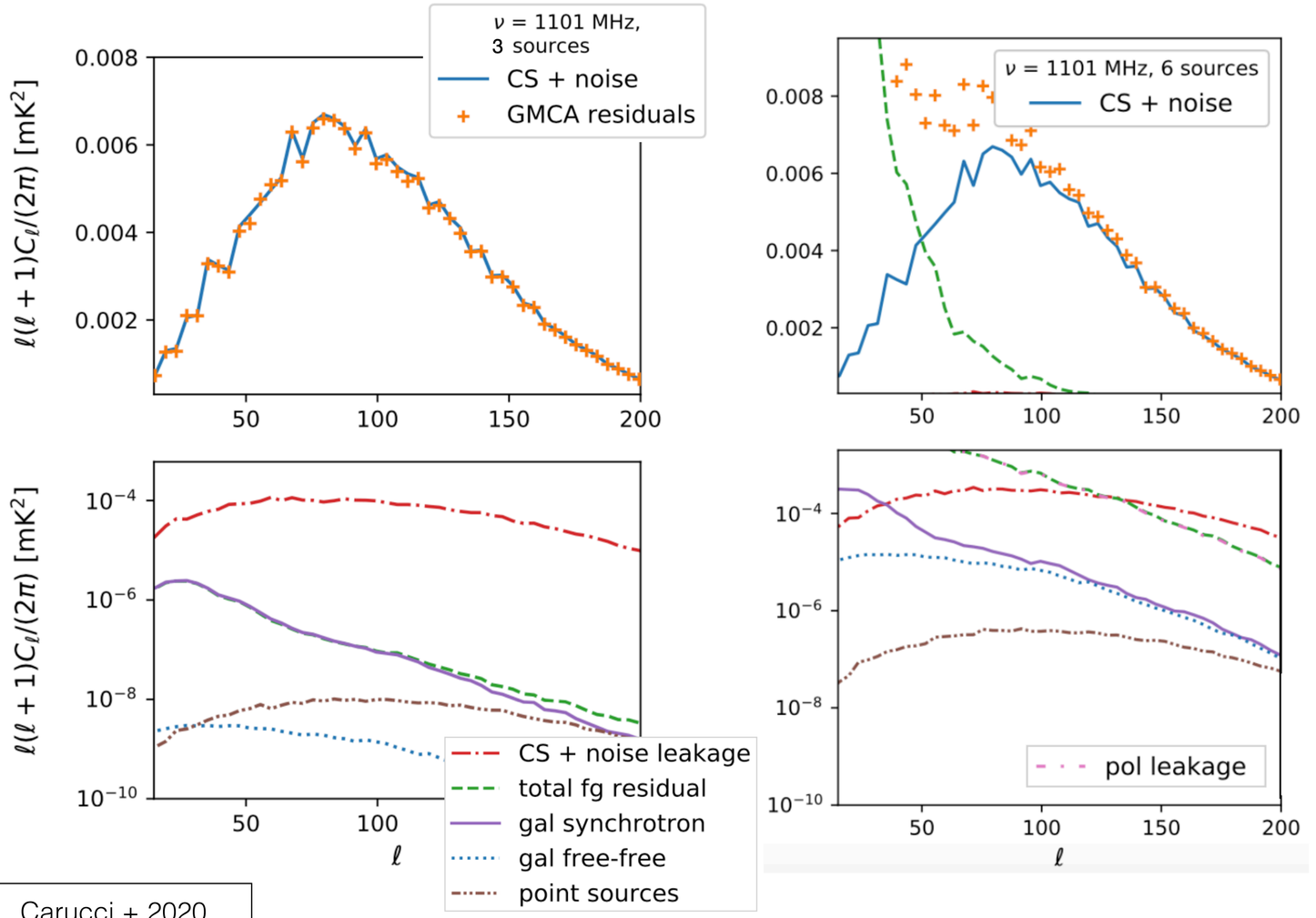
1. Good performance also with **RFI-flagged** data cubes!
   (TV stations, telecommunication, satellites,..)
2. **Pol leakage:** greater complexity of data
   (higher number of sources needed, convergence not assured, mode-mixing assured)

To reproduce these results: codes and sims available online

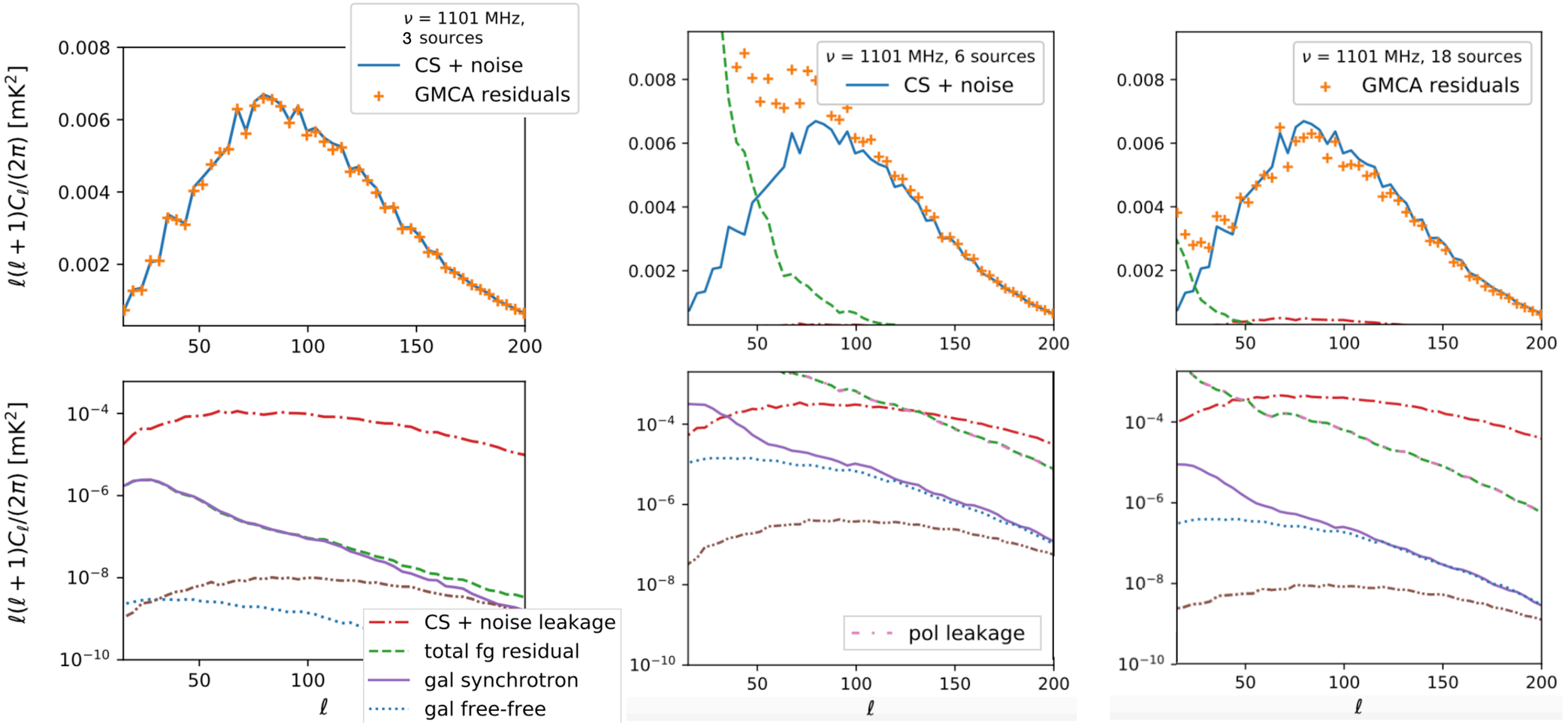
a quick interlude on mixGMCA



Carucci + 2020

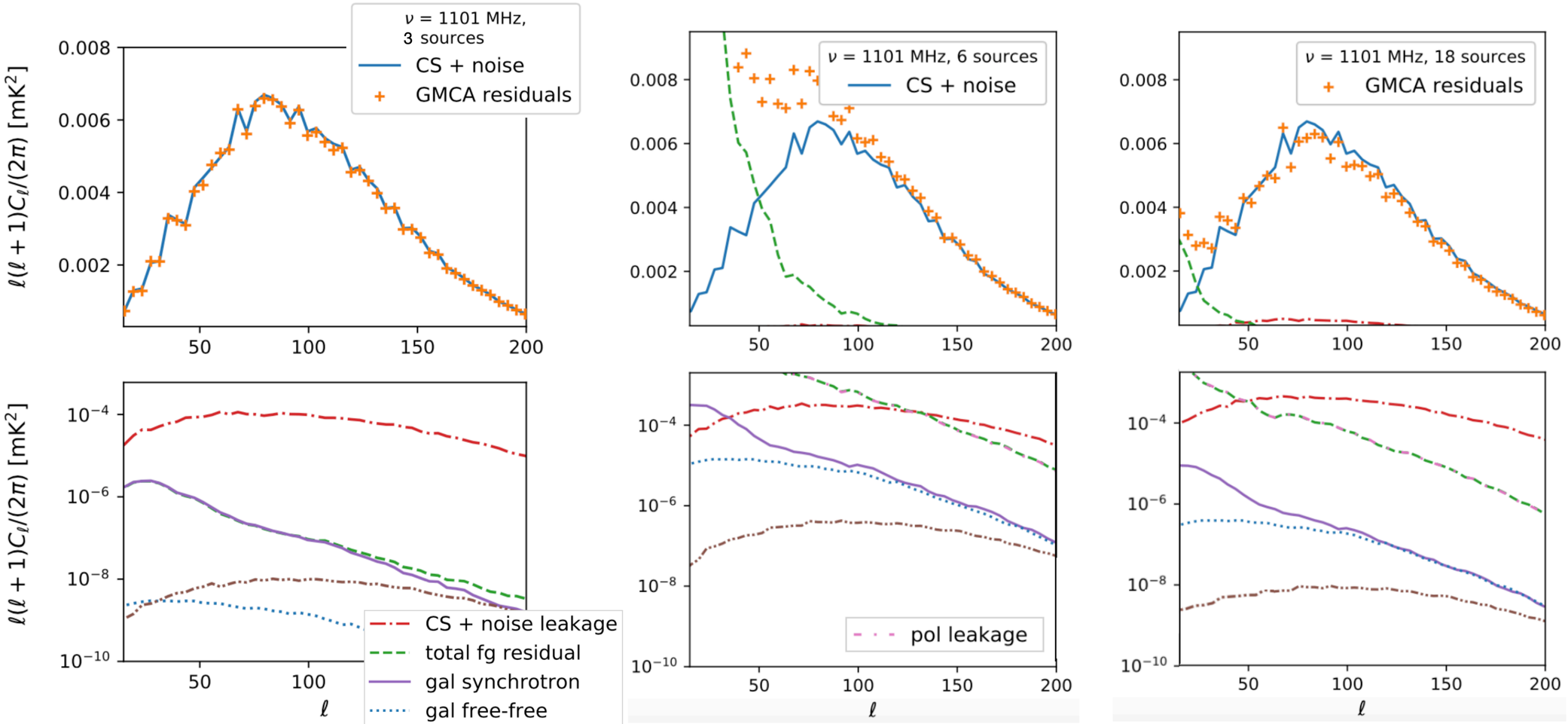


Carucci + 2020



Carucci + 2020

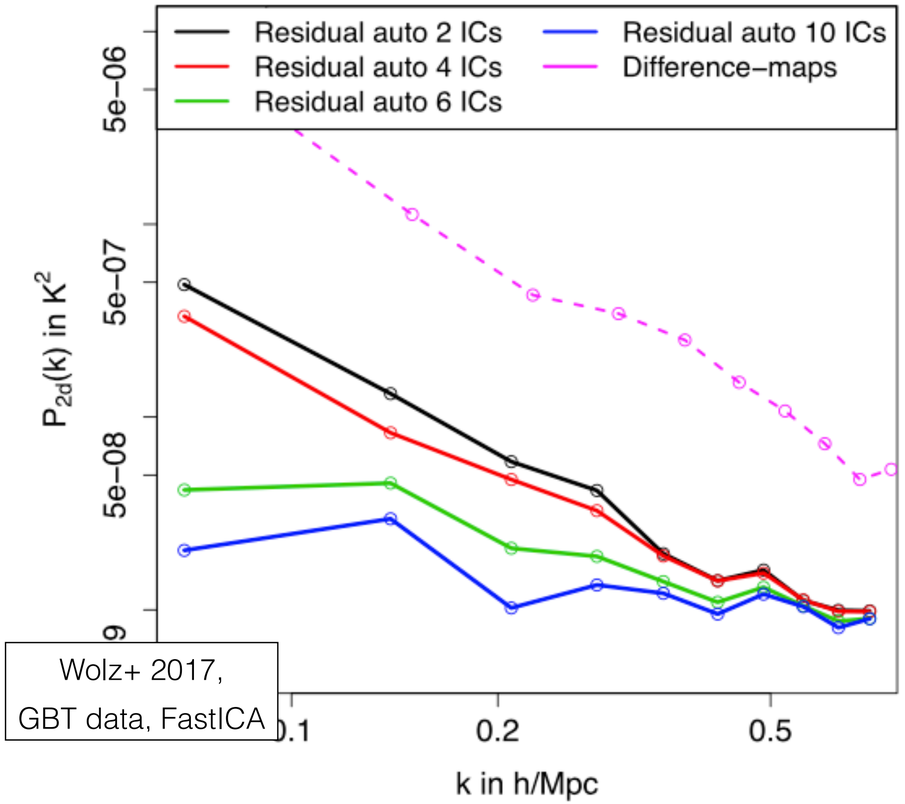
Different scales need different care



Carucci + 2020

Different scales need different care

The wavelet domain is a multi-scale framework!



Wolz+ 2017,
GBT data, FastICA

- GMCA performs very well on small scales
- PCA / ICA —> overfit the large scales

PCA on the large scale

+

GMCA on the small scales

mixGMCA

See also Hothi+2020 with LOFAR data

MeerKAT

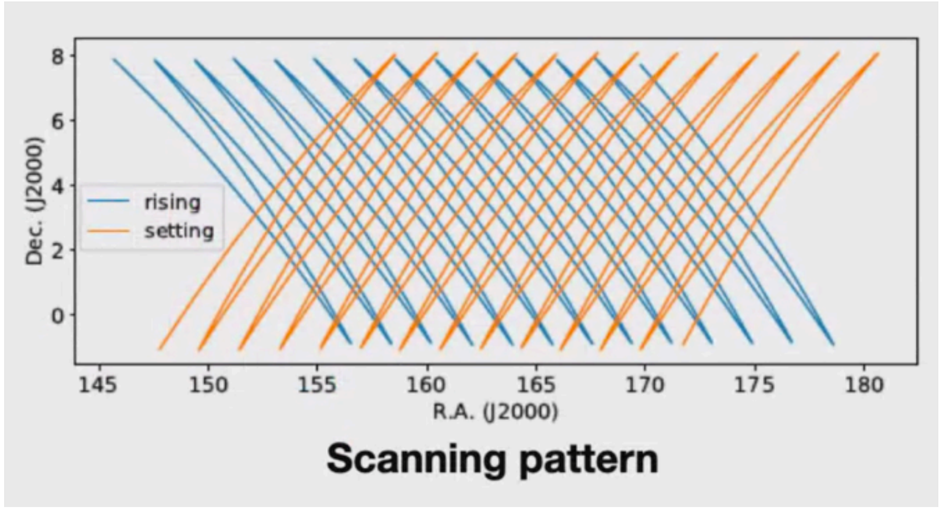
MeerKLASS: MeerKAT Large Area Synoptic Survey

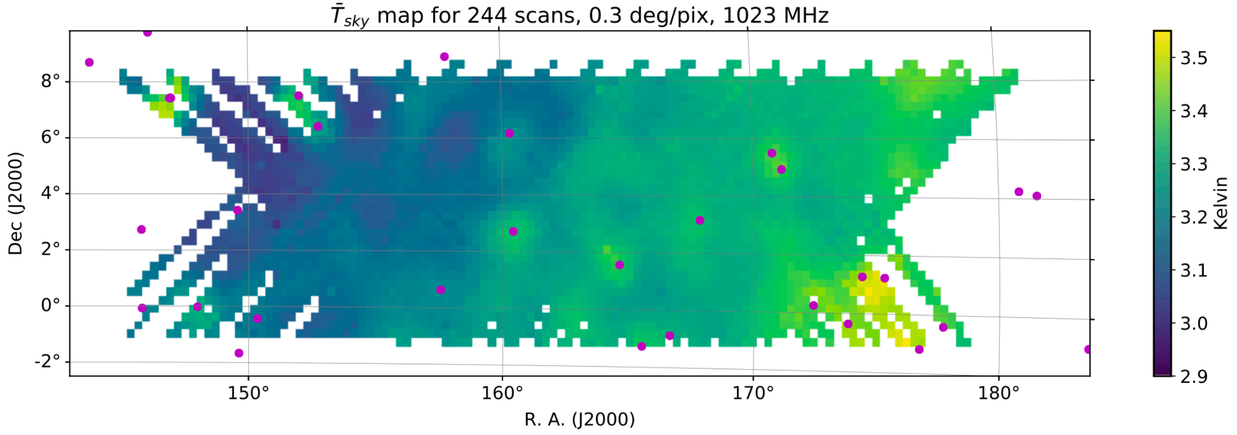
ArXiv: 1709:06099

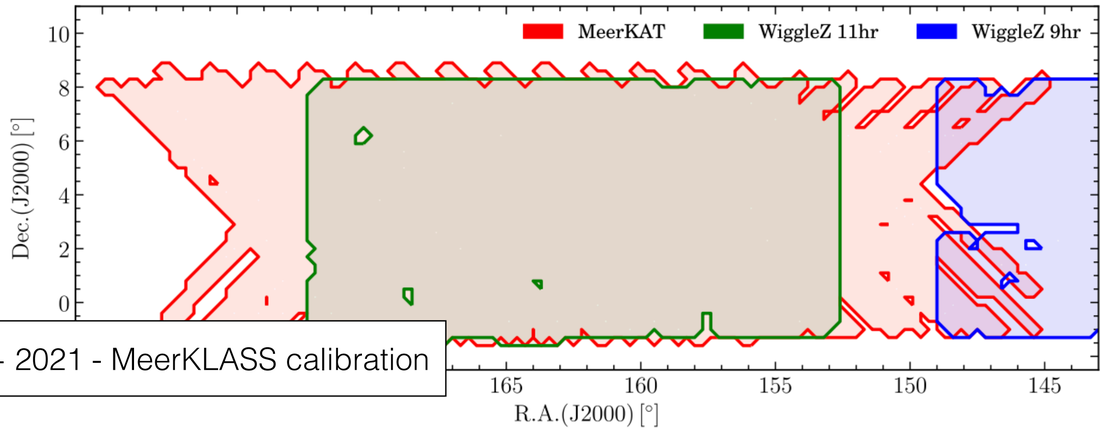
Alkistis Pourtsidou, Amadeus Wild, Brandon Engelbrecht, Isabella Carucci, Jingying Wang, Keith Grainge, Laura Wolz, Mario Santos, Marta Spinelli, Mel Irfan, Phil Bull, Stefano Camera, Steve Cunnington, Tamirat Gebeyehu, Zé Fonseca, …

Pilot survey data (2019):

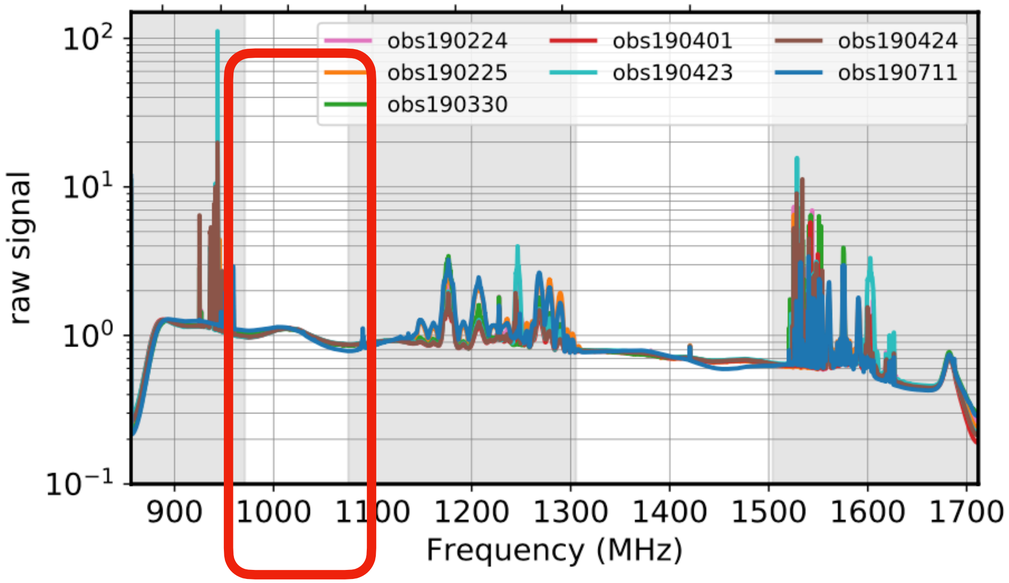
- 10.5 hours of data from six nights of observations
- Overlapping with the WiggleZ11hr field (~200 deg2)
- We use data in range 973-1015 MHz (0.40 < z < 0.46)





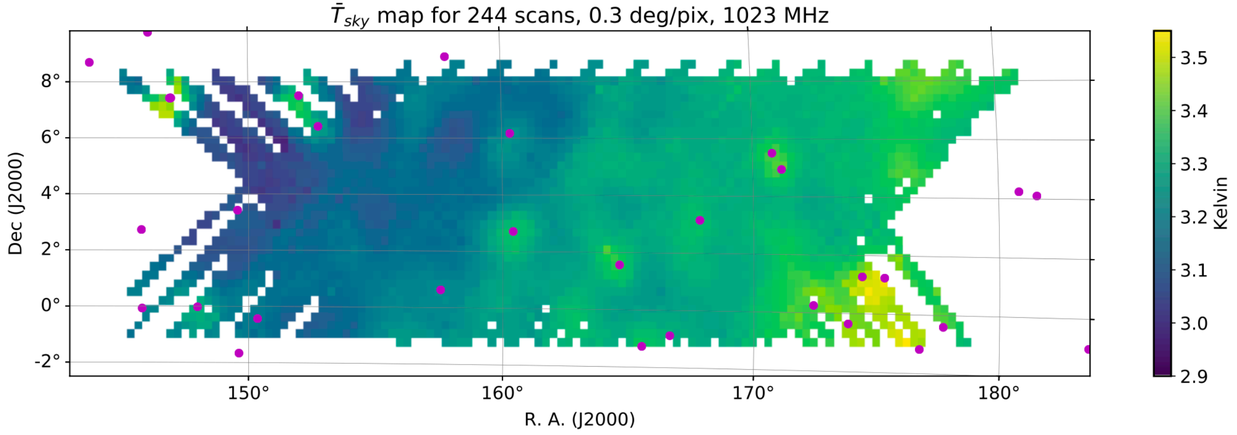


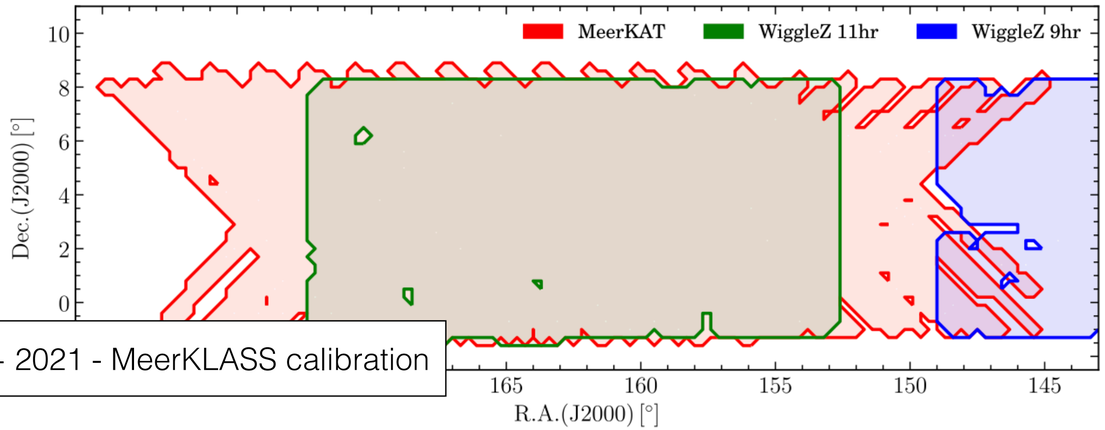
JY Wang + 2021 - MeerKLASS calibration



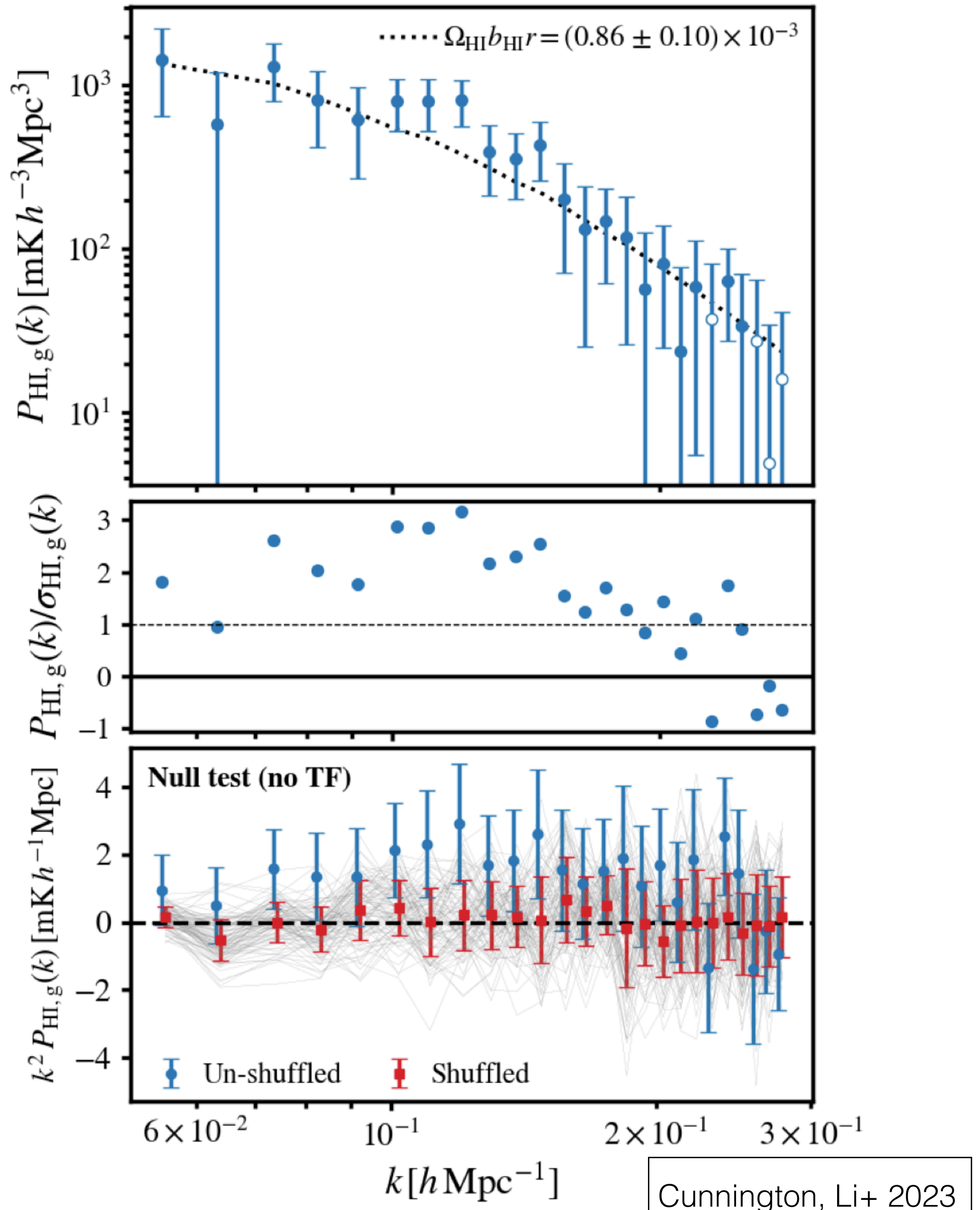
Pilot survey data (2019):

- 10.5 hours of data from six nights of observations
- Overlapping with the WiggleZ11hr field (~200 deg2)
- We use data in range 973-1015 MHz (0.40 < z < 0.46)



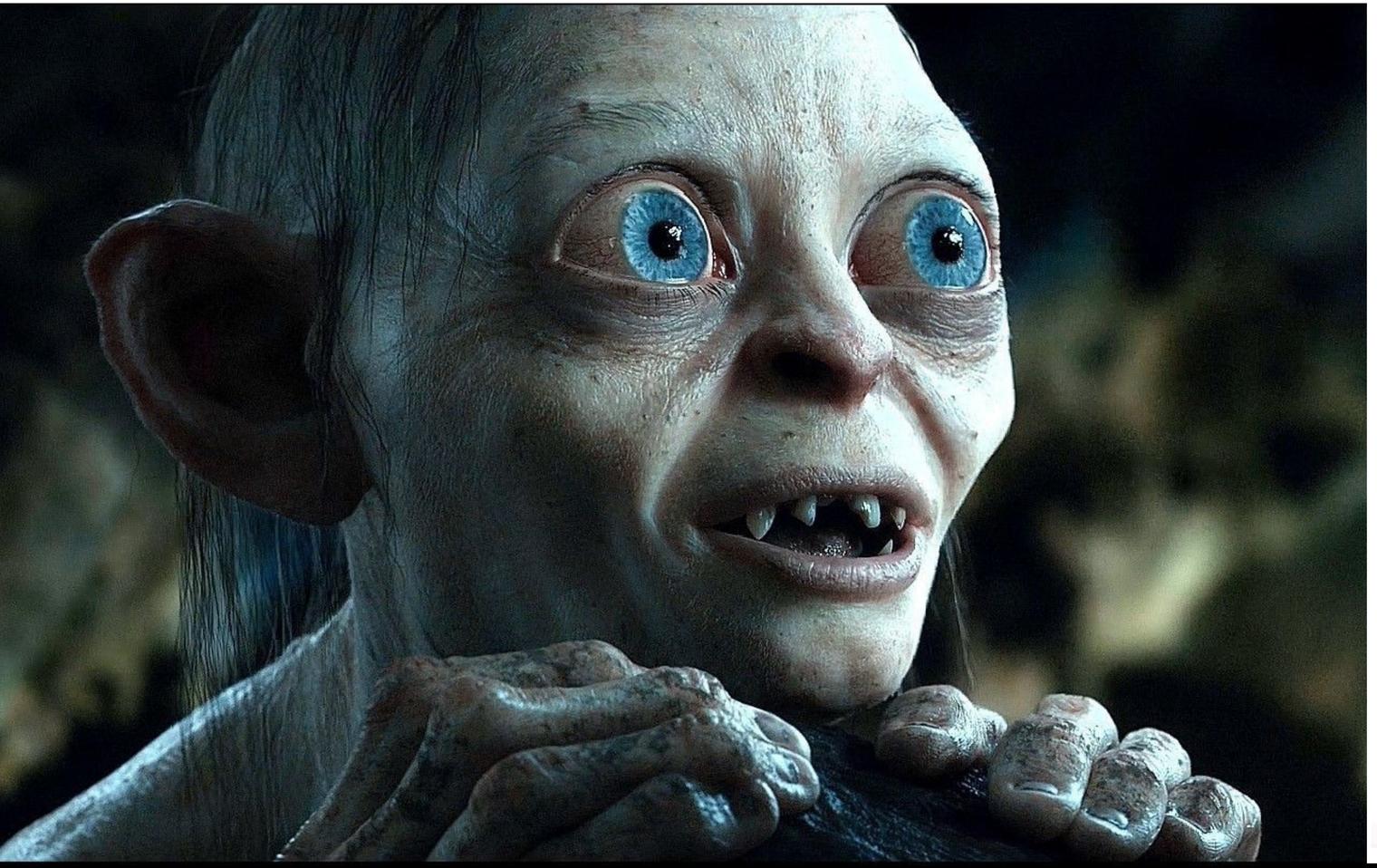


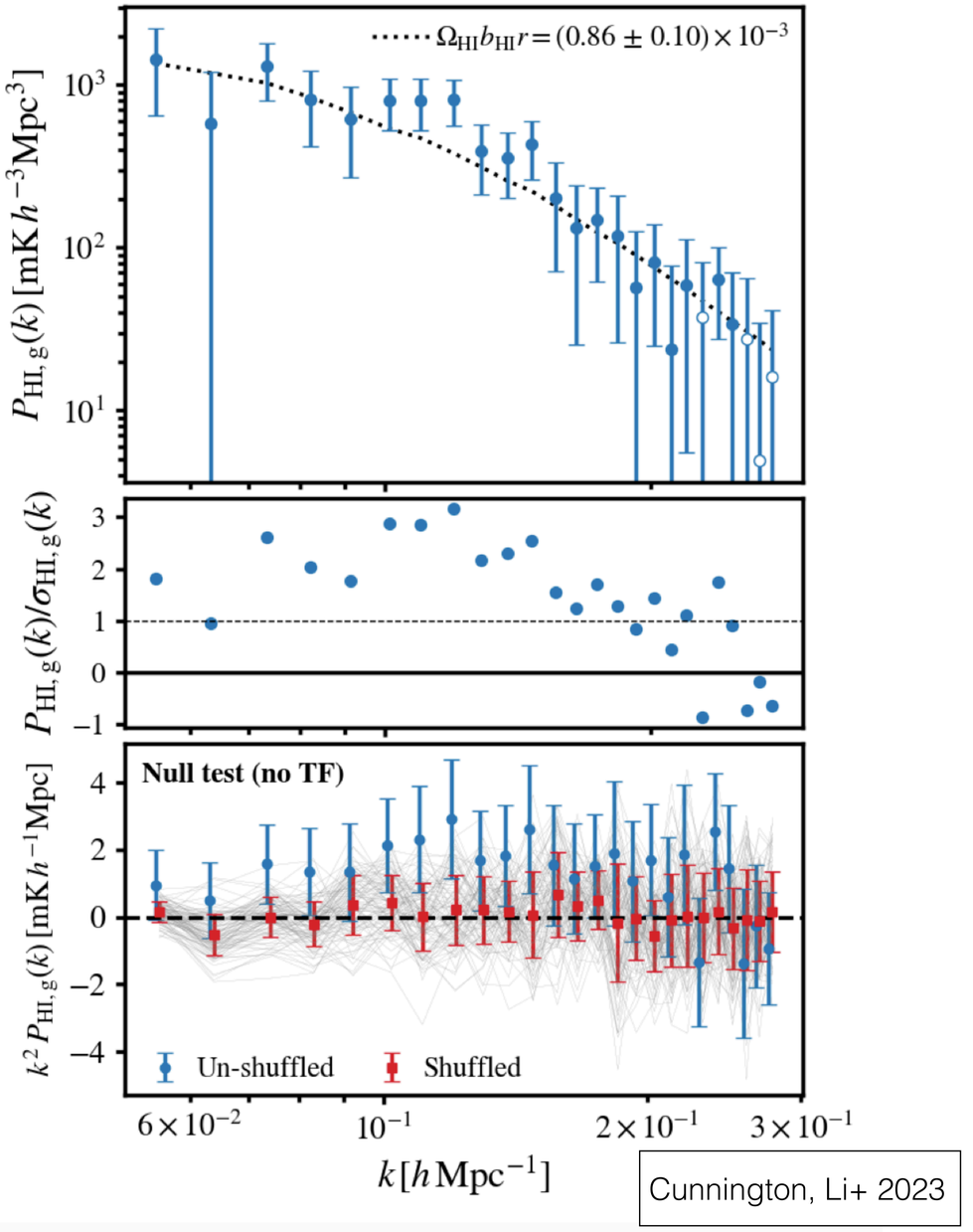
JY Wang + 2021 - MeerKLASS calibration



Cunnington, Li+ 2023

Our precious!!

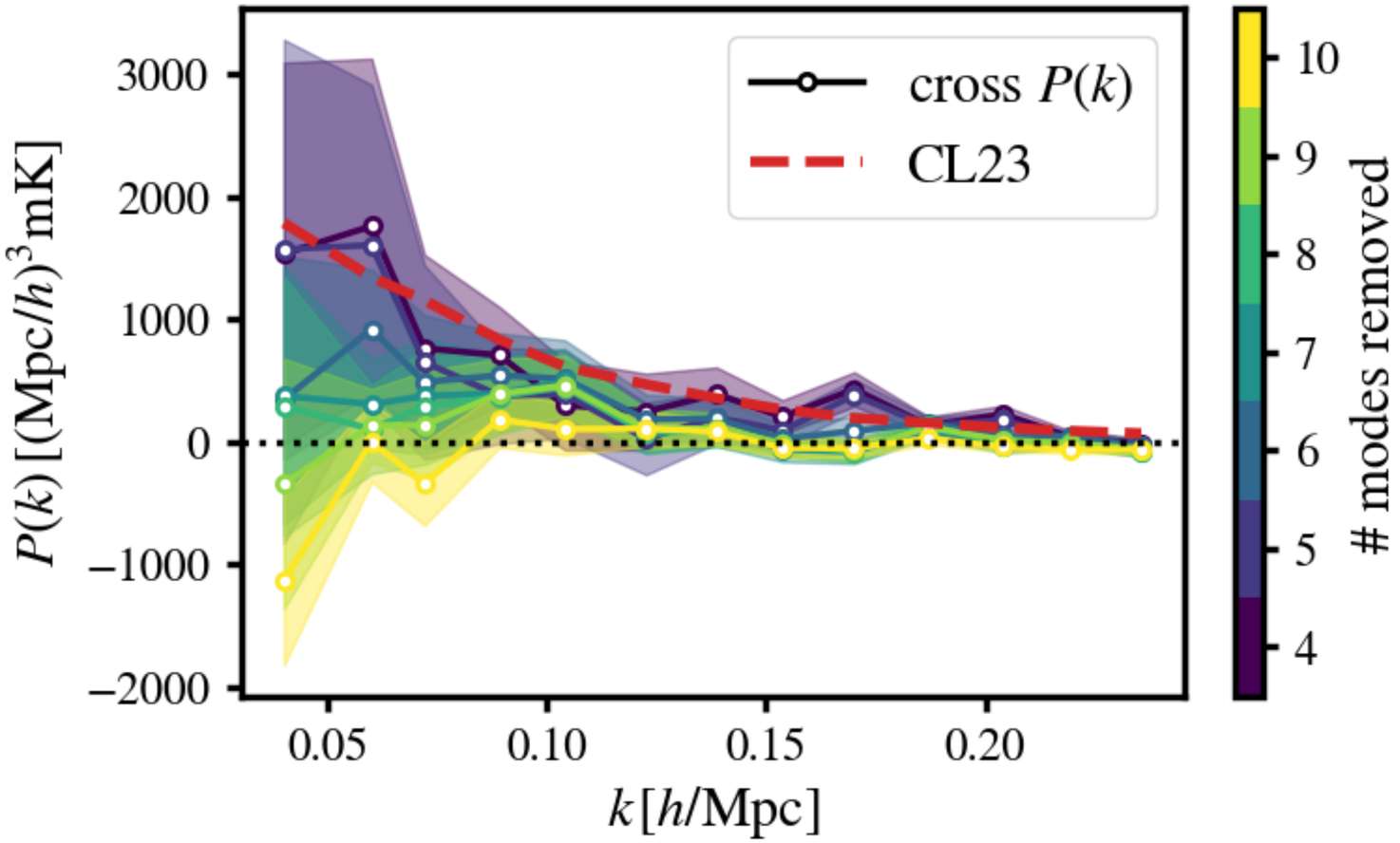




Cunnington, Li+ 2023

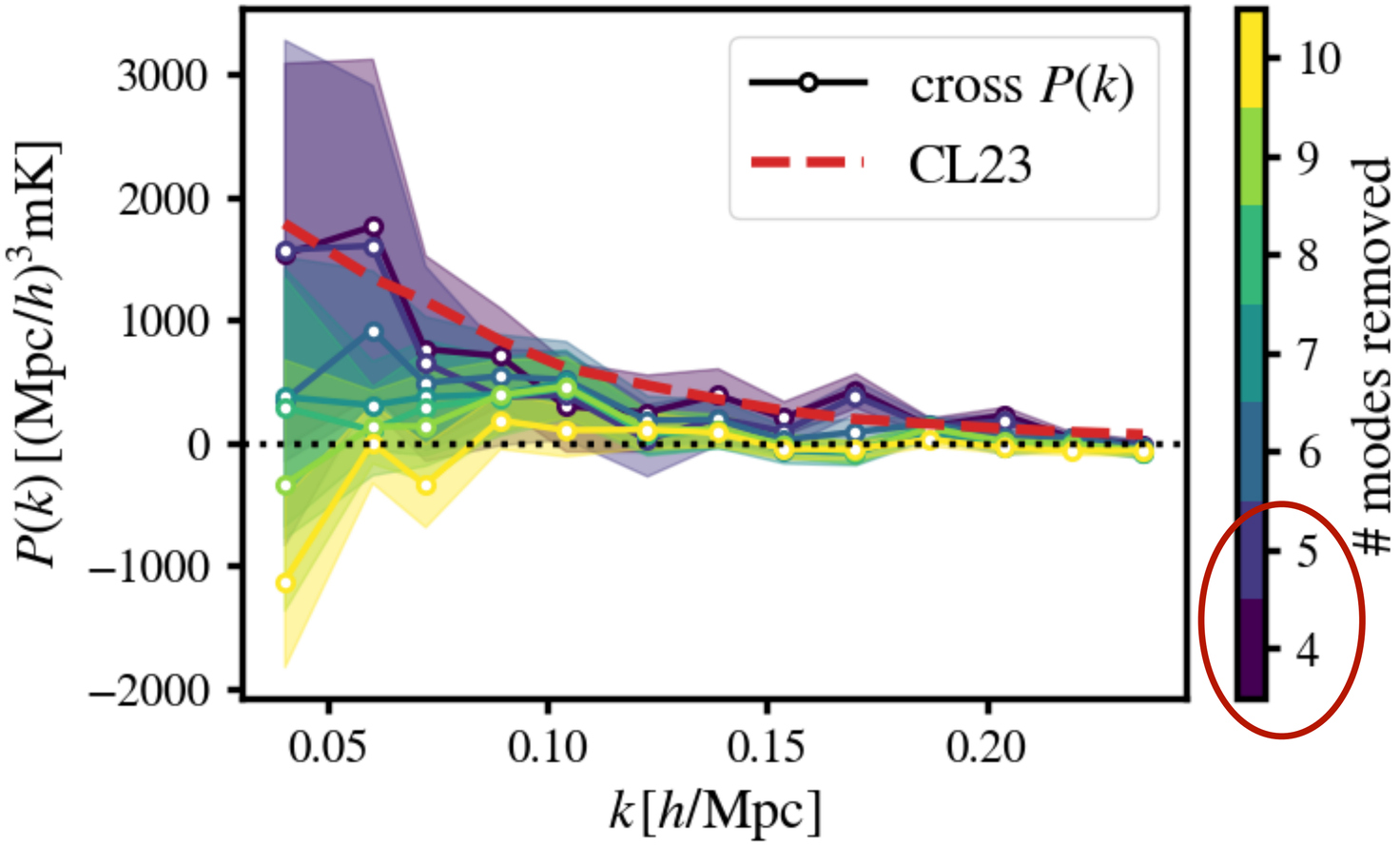
Preliminary

Use the cross-correlation as a benchmark



Preliminary

Use the cross-correlation as a benchmark

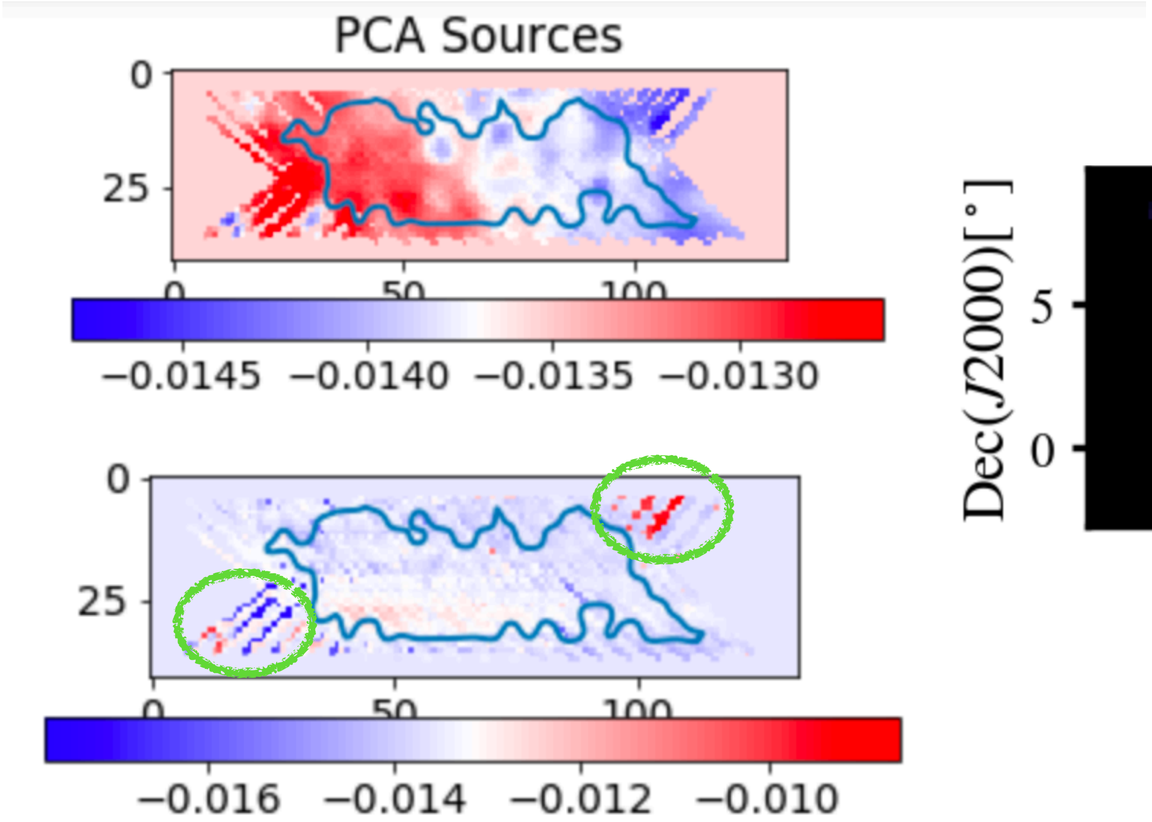


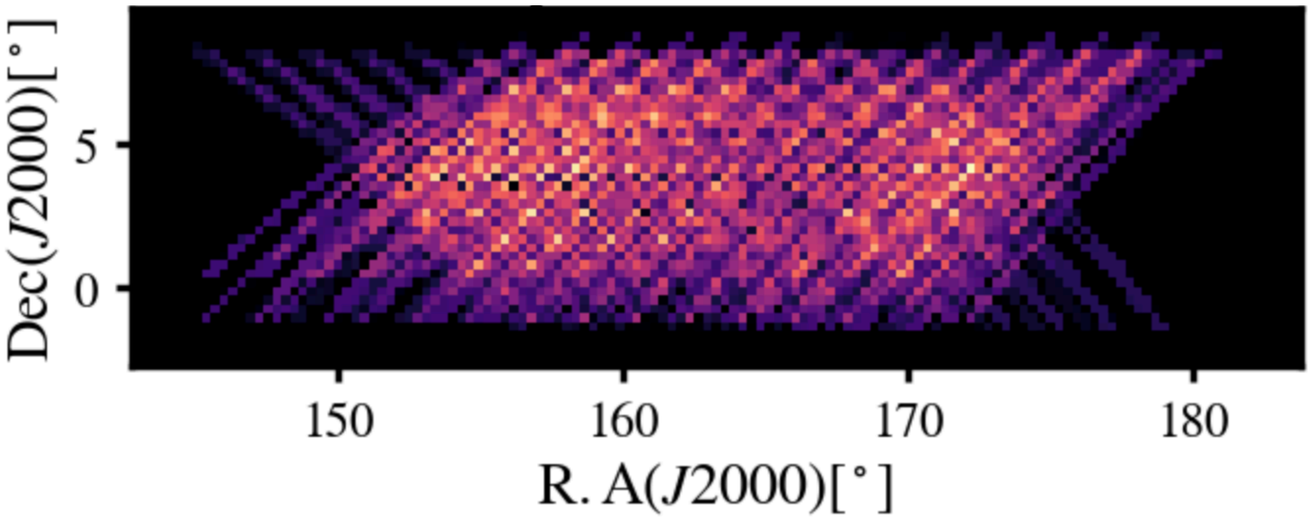
To be compared with 30 !

Re-analyis of 2019 data

Preliminary

$$X = A\ S$$



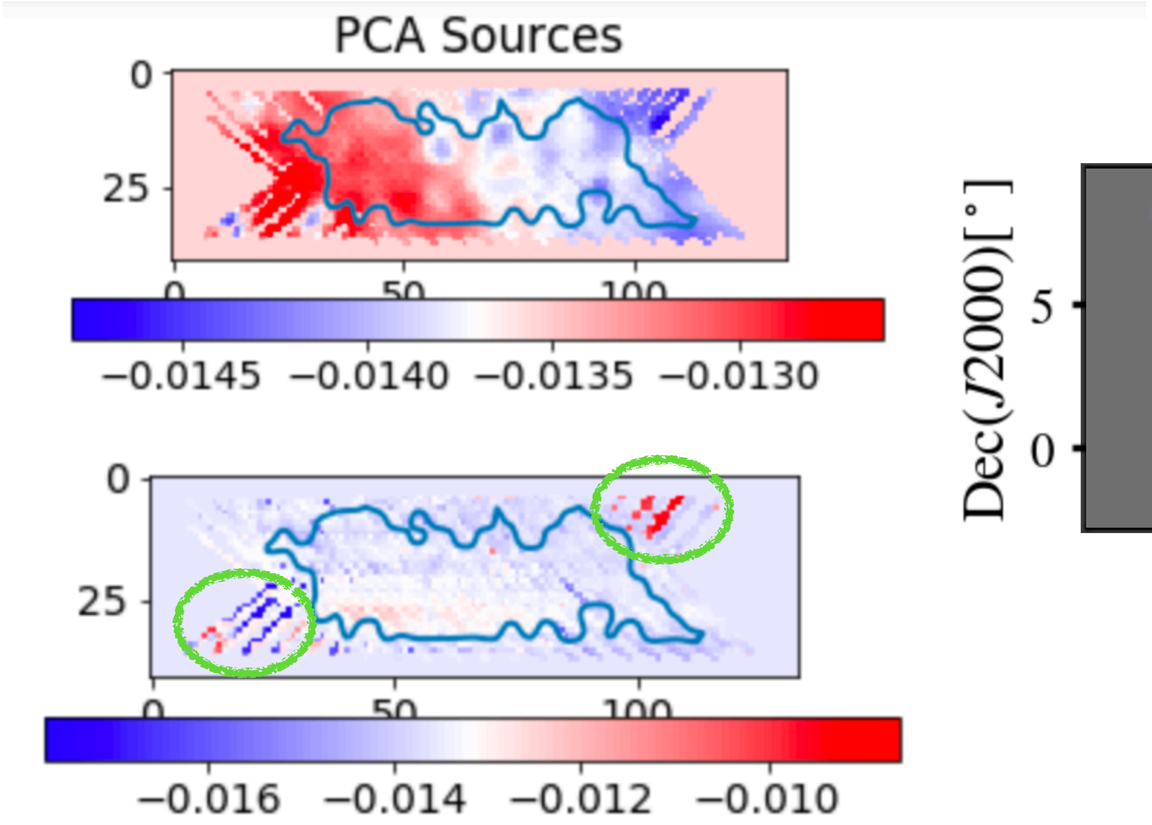


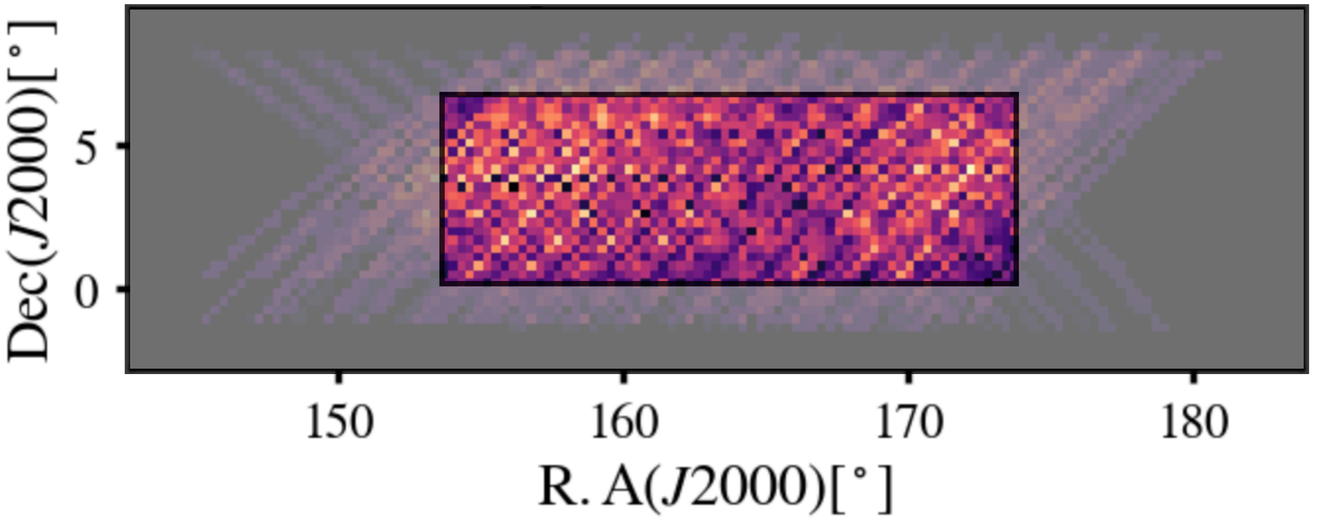
Preliminary

Re-analyis of 2019 data

1. PCA-informed pixel flagging

$$\mathbf{X} = \mathbf{A}\ \mathbf{S}$$



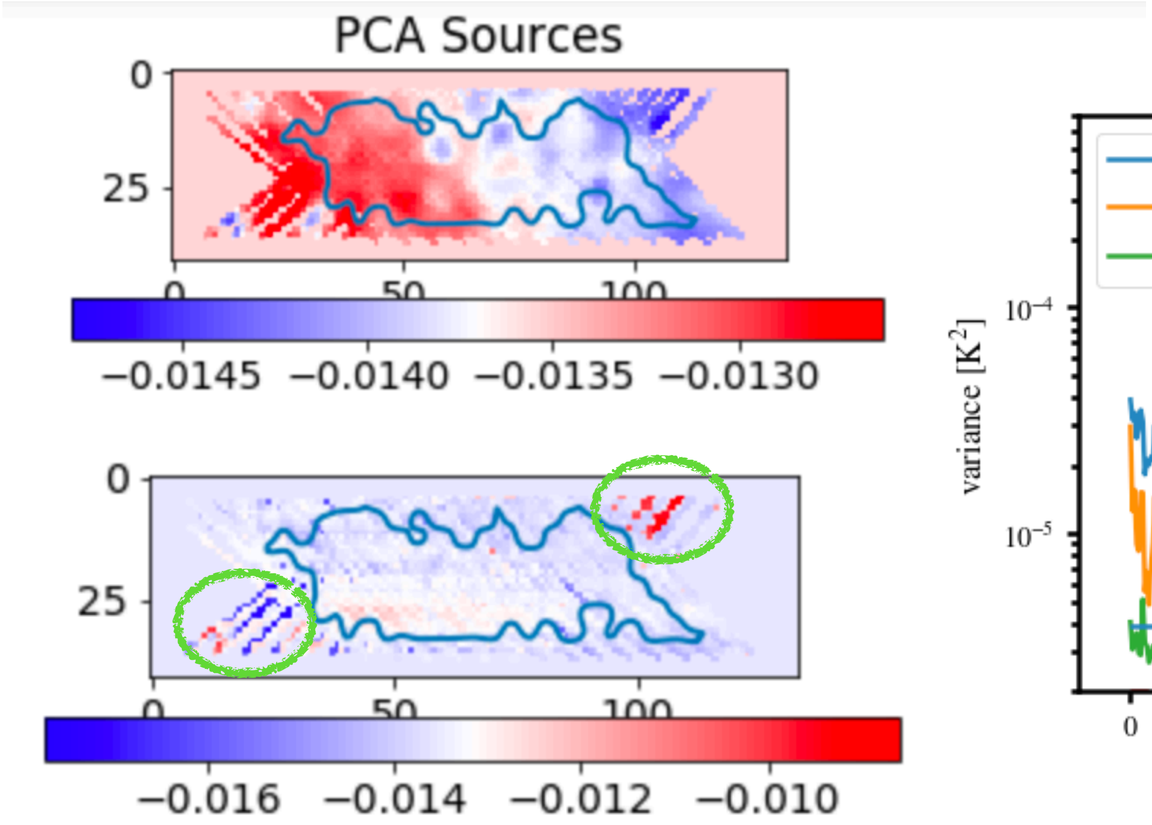


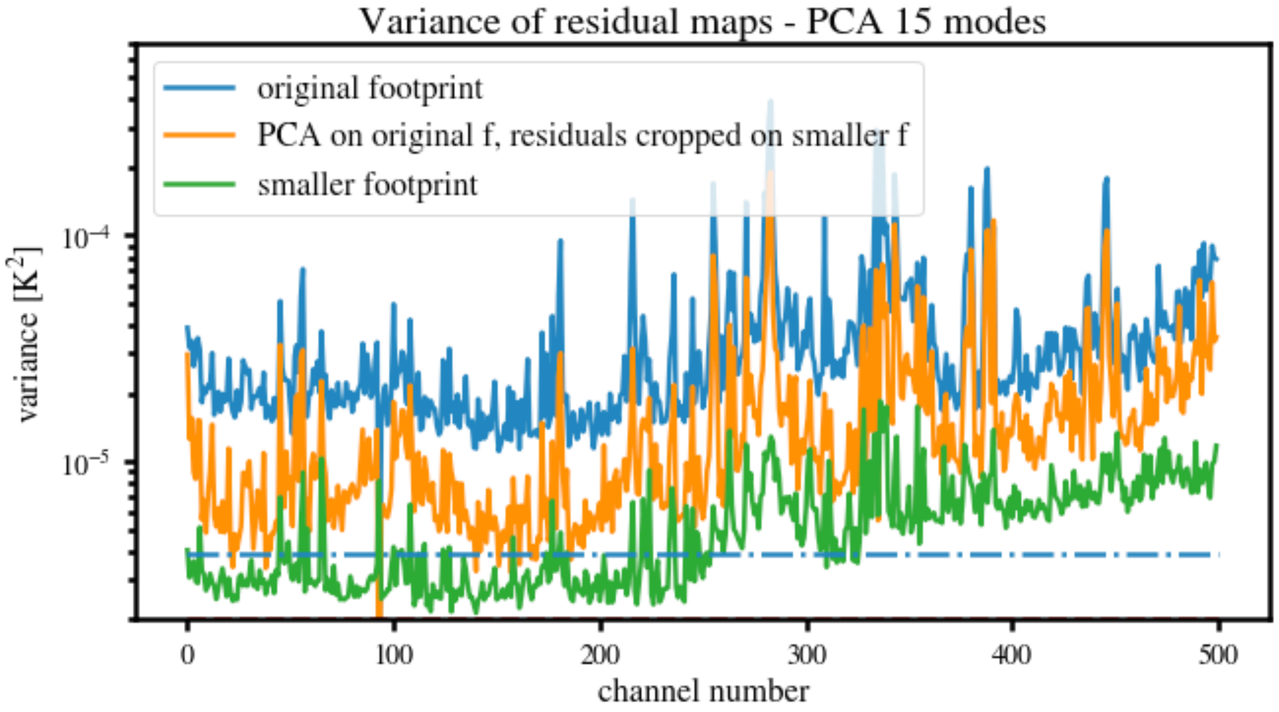
Re-analyis of 2019 data

1. PCA-informed pixel flagging

$$\mathbf{X} = \mathbf{A}\ \mathbf{S}$$

Preliminary



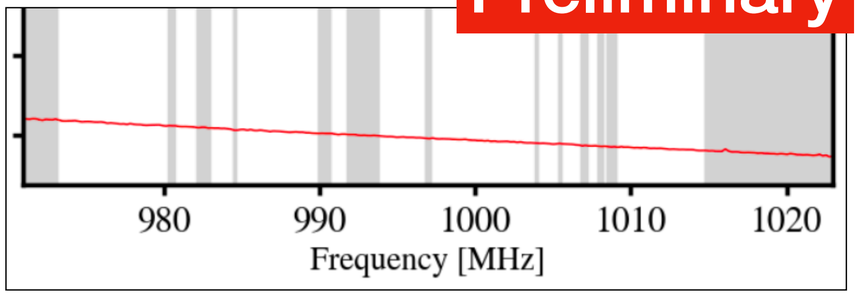


Re-analyis of 2019 data

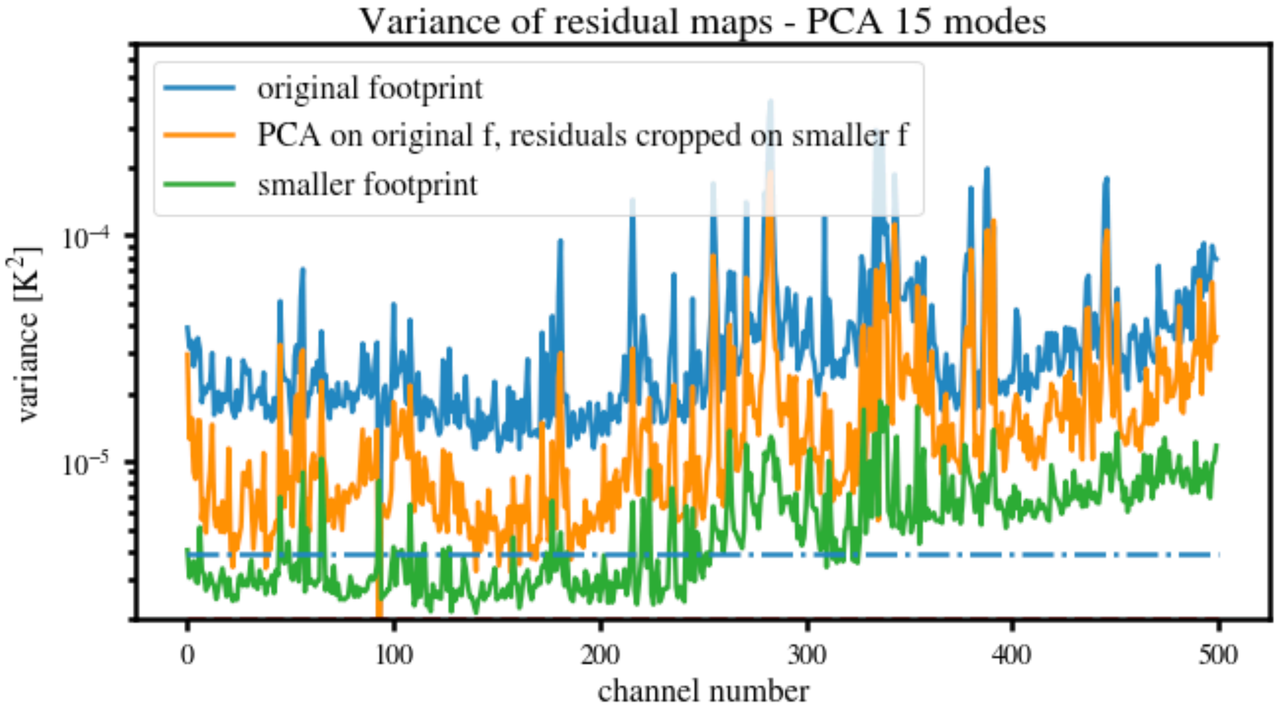
1. PCA-informed pixel flagging
2. Keep *bad* channels

see also discussion in Carucci+ 2020

$$\mathbf{X} = \mathbf{A}\,\mathbf{S}$$



Preliminary

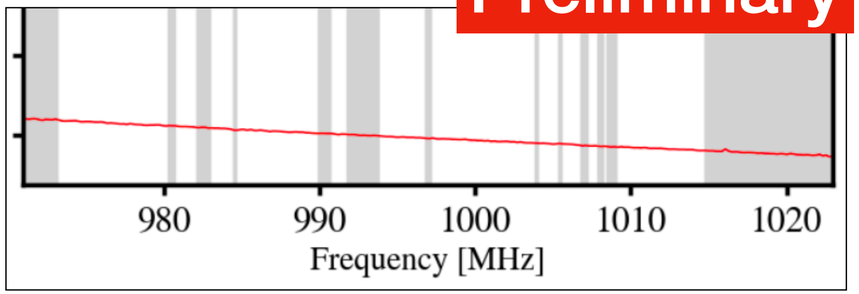


Re-analyis of 2019 data

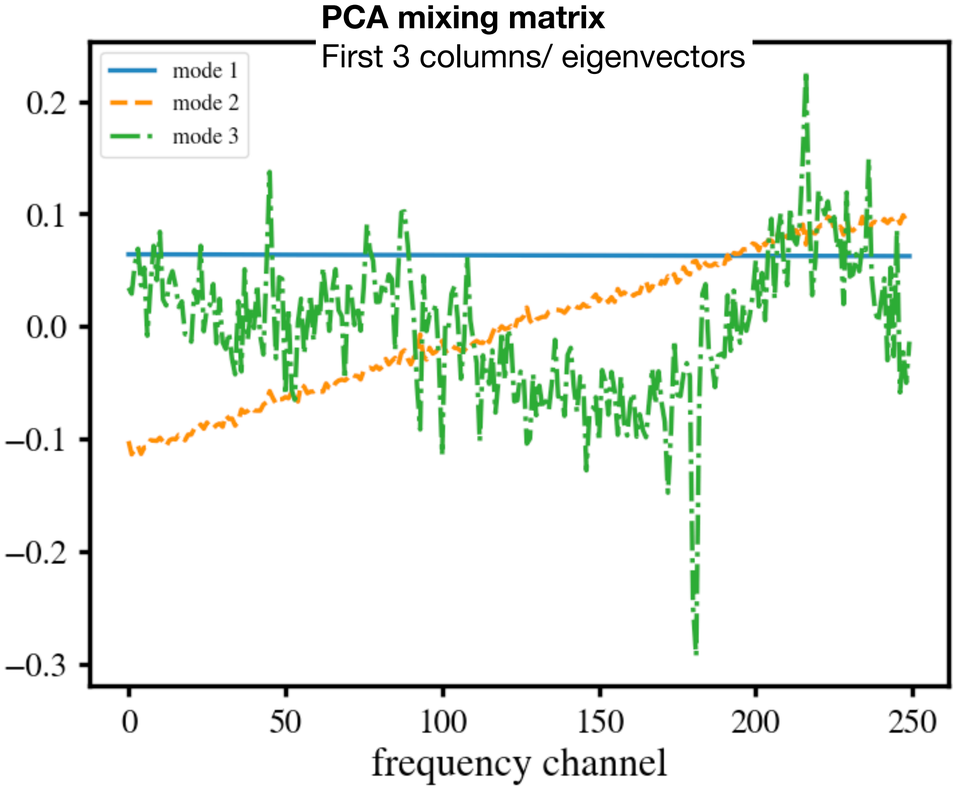
1. PCA-informed pixel flagging
2. Keep *bad* channels

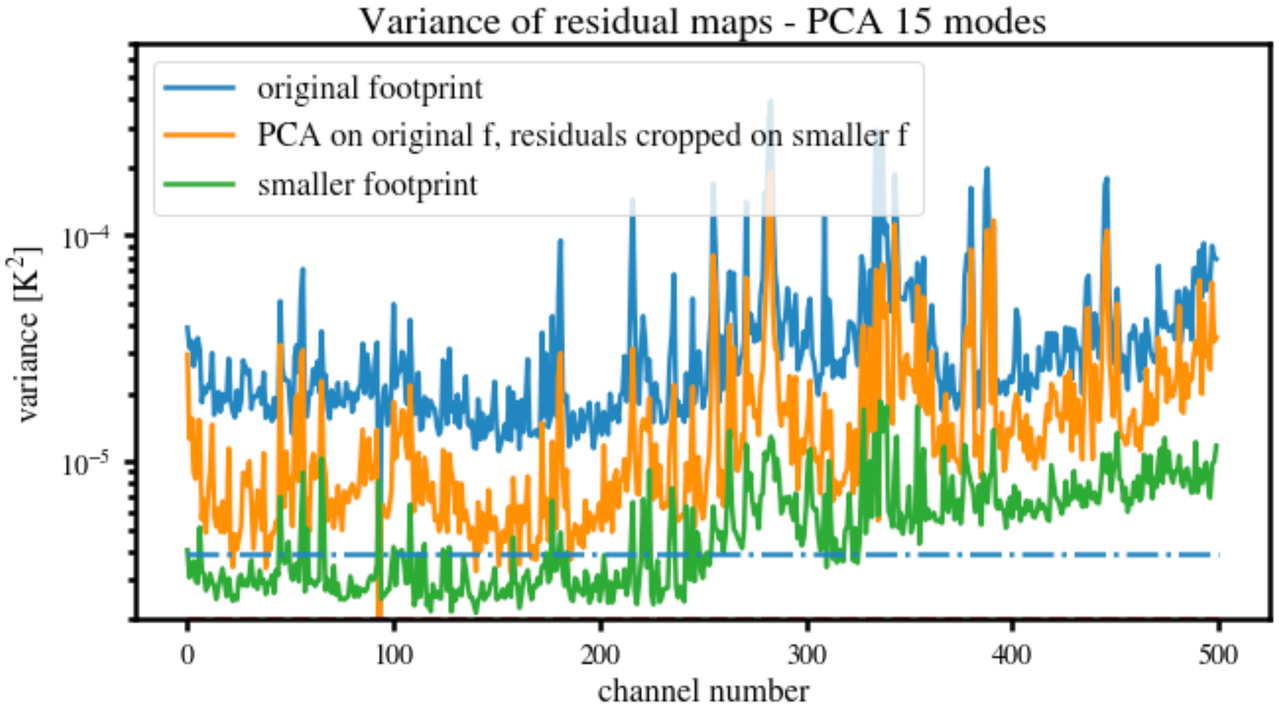
see also discussion in Carucci+ 2020

$$\mathbf{X} = \mathbf{A}\,\mathbf{S}$$



Preliminary

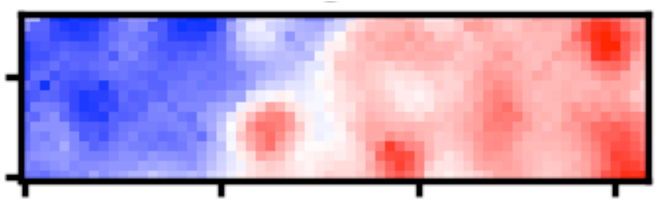


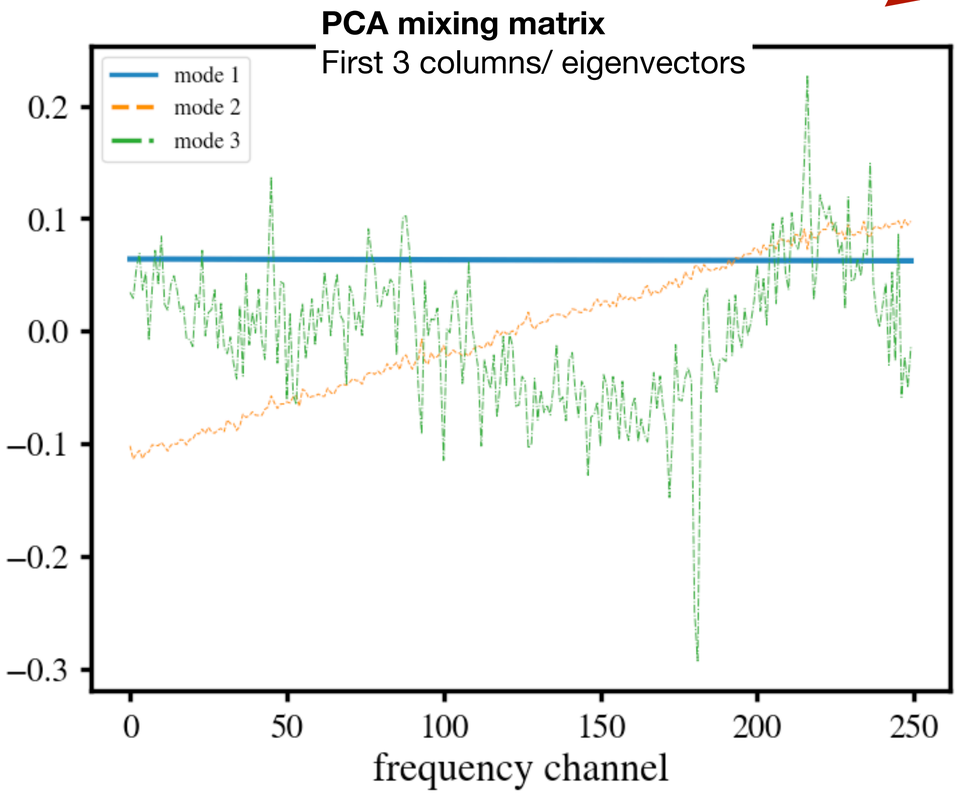


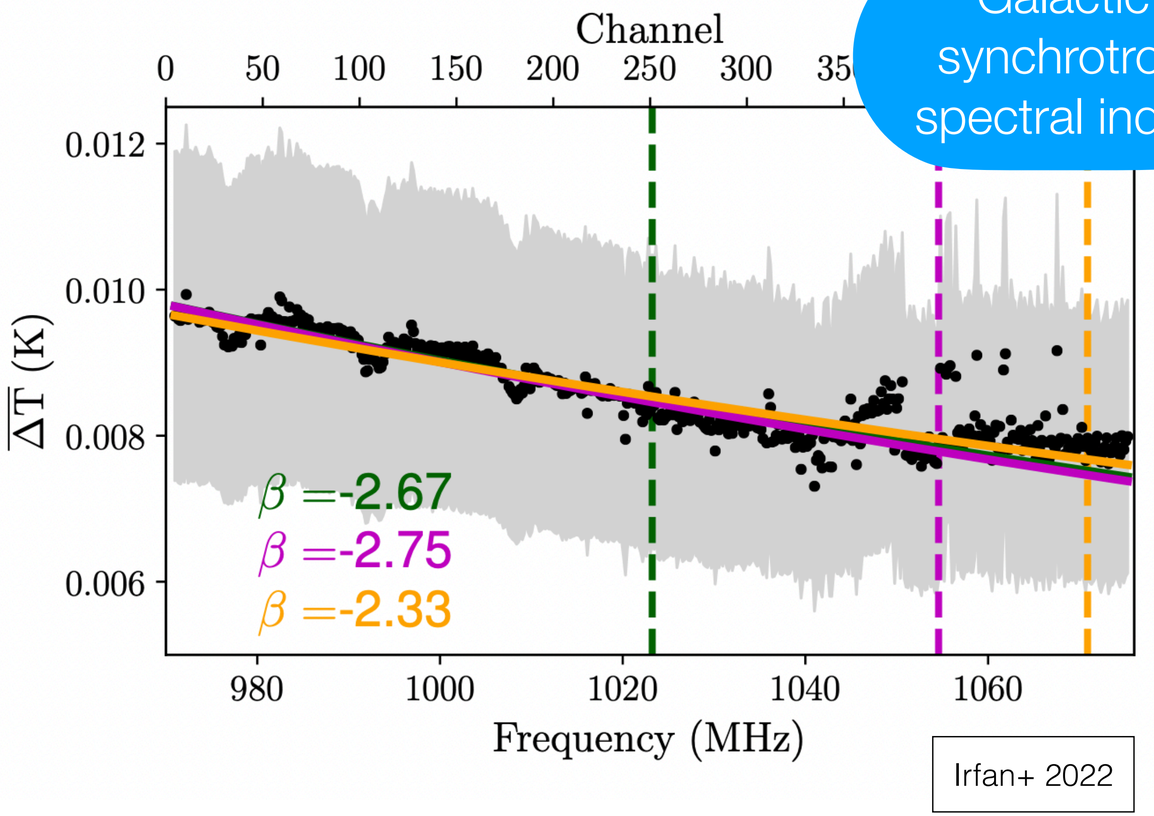
Re-analyis of 2019 data

1. PCA-informed pixel flagging
2. Keep *bad* channels

$$\mathbf{X} = \mathbf{A}\ \mathbf{S}$$





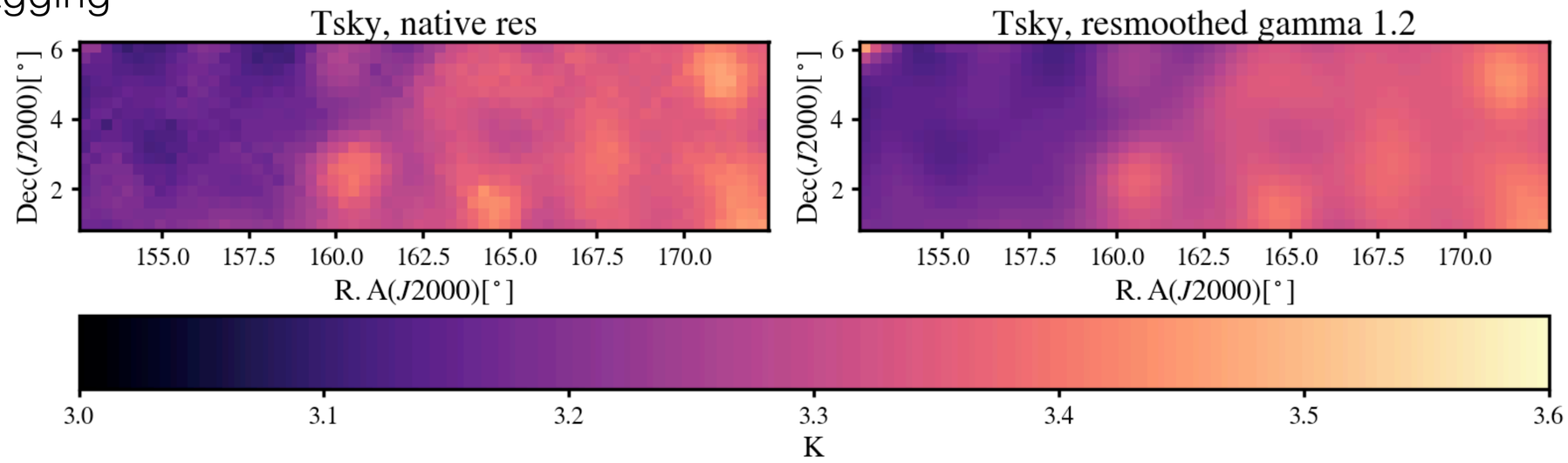


Galactic synchrotron spectral index

Irfan+ 2022

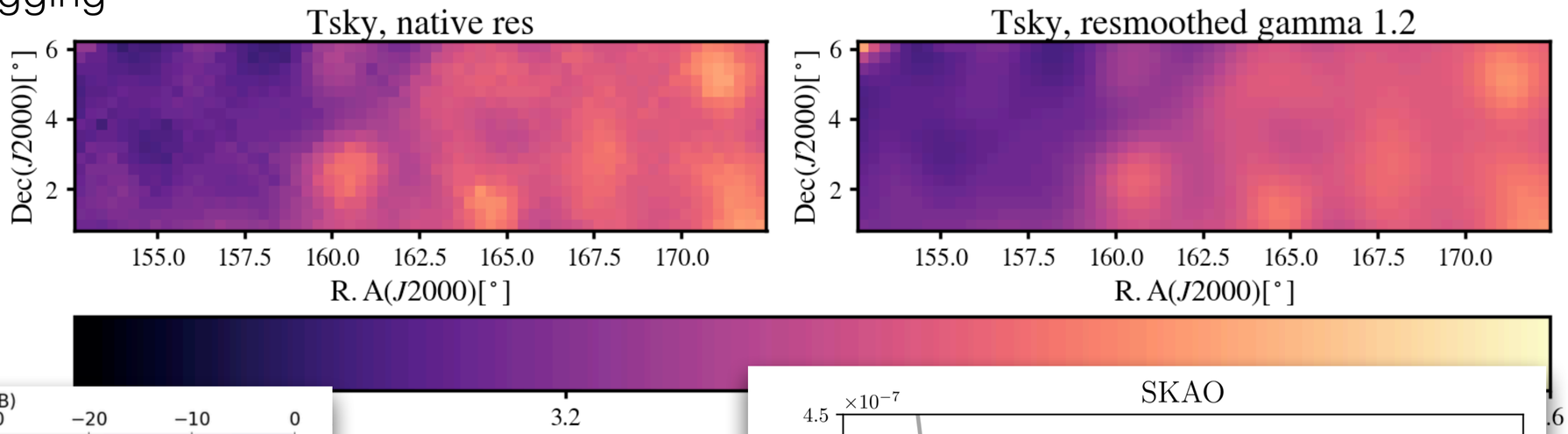
Re-analyis of 2019 data

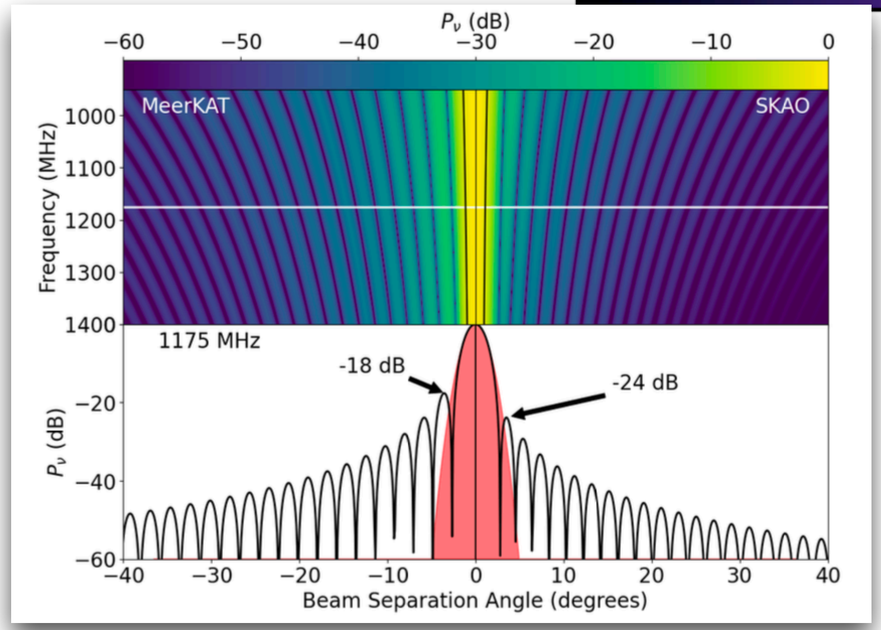
1. PCA-informed pixel flagging
2. Keep *bad* channels
3. No re-smoothing



Re-analyis of 2019 data

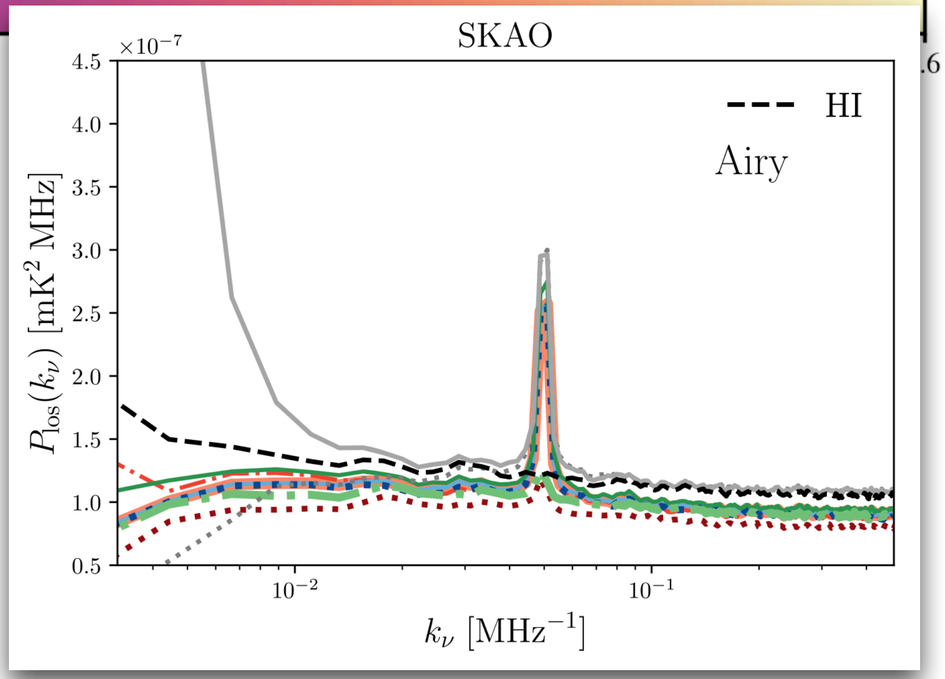
1. PCA-informed pixel flagging
2. Keep *bad* channels
3. No re-smoothing





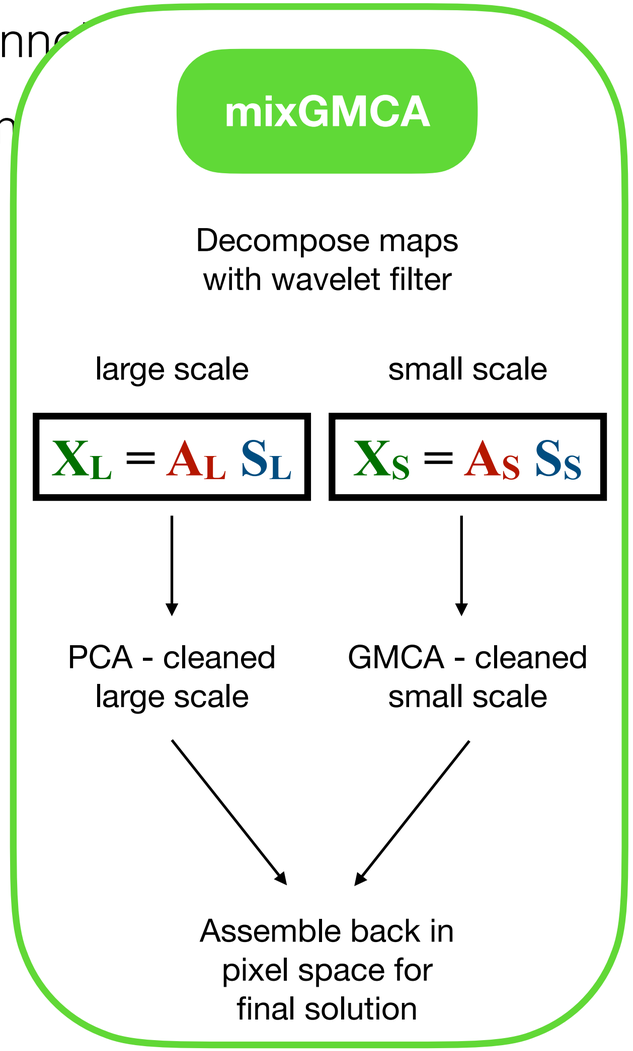
Blind foreground subtraction challenge

Spinelli, Carucci + 2022



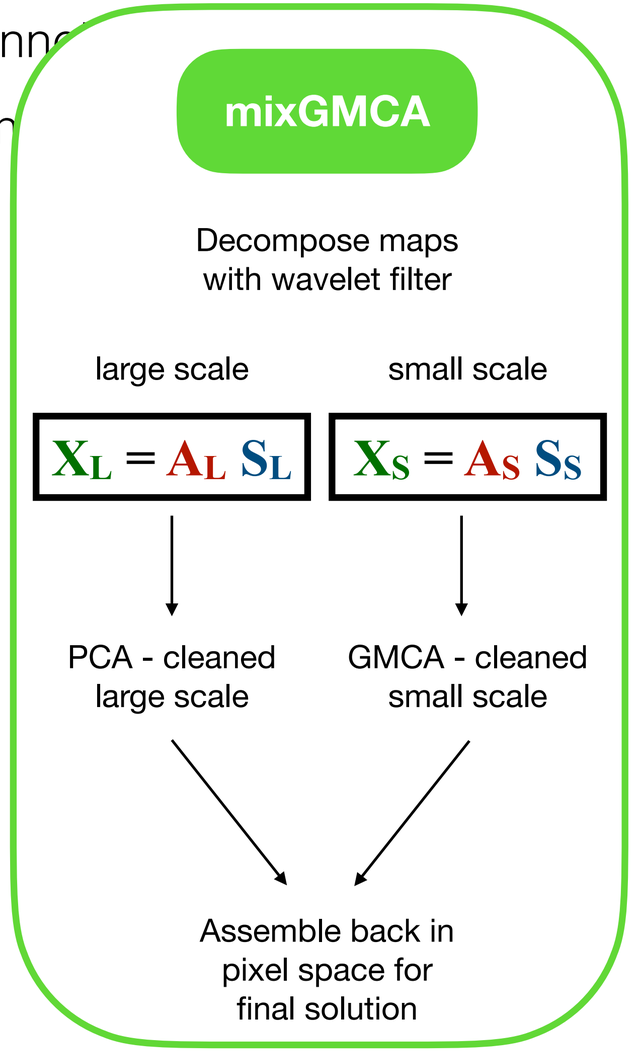
Re-analyis of 2019 data

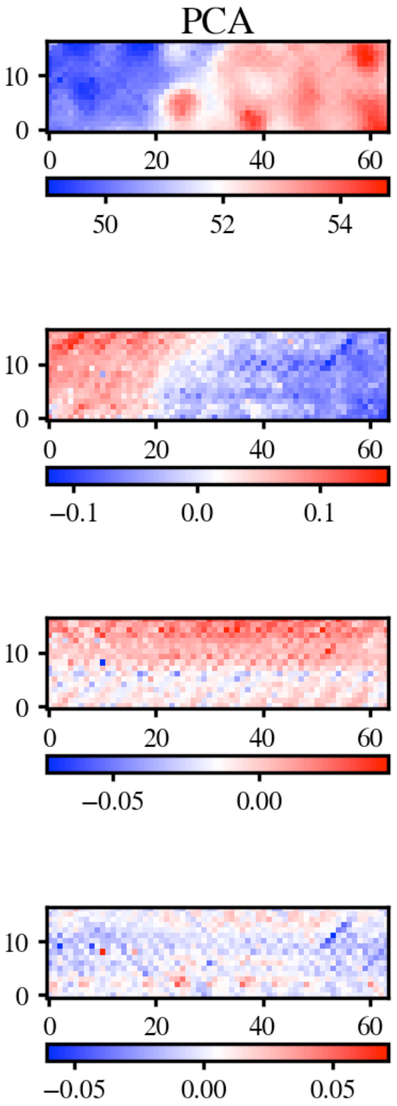
1. PCA enforced pixel-flagging
2. Keep *bad* channe
3. No re-smoothin
4. Beyond PCA?



Re-analyis of 2019 data

1. PCA enforced pixel-flagging
2. Keep *bad* channe
3. No re-smoothin
4. Beyond PCA?





Re-analyis of 2019 data

1. PCA enforced pixel-flagging
2. Keep *bad* channe
3. No re-smoothin
4. Beyond PCA?

mixGMCA

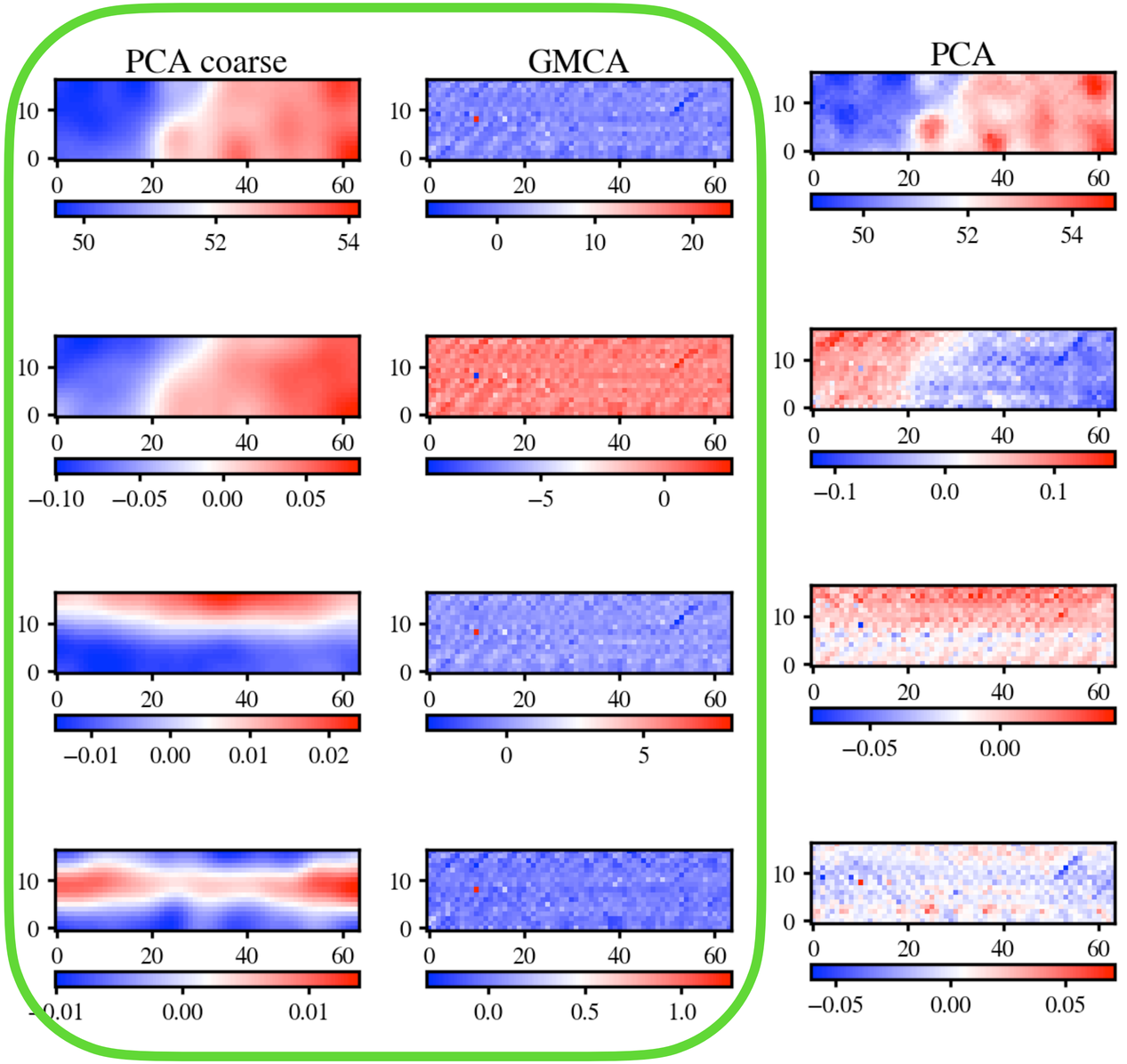
Decompose maps with wavelet filter

large scale | small scale

$X_L = A_L S_L$ | $X_S = A_S S_S$

PCA - cleaned large scale | GMCA - cleaned small scale

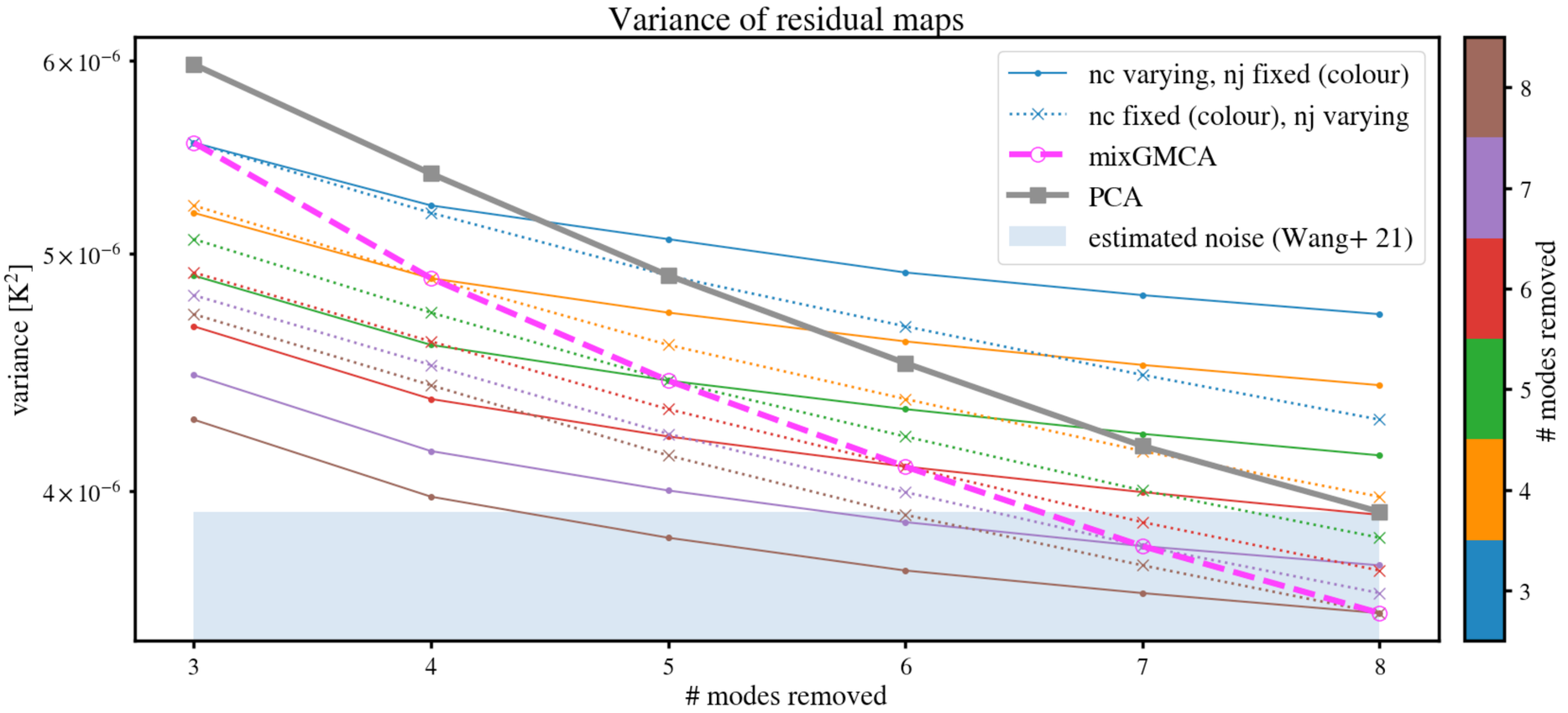
Assemble back in pixel space for final solution



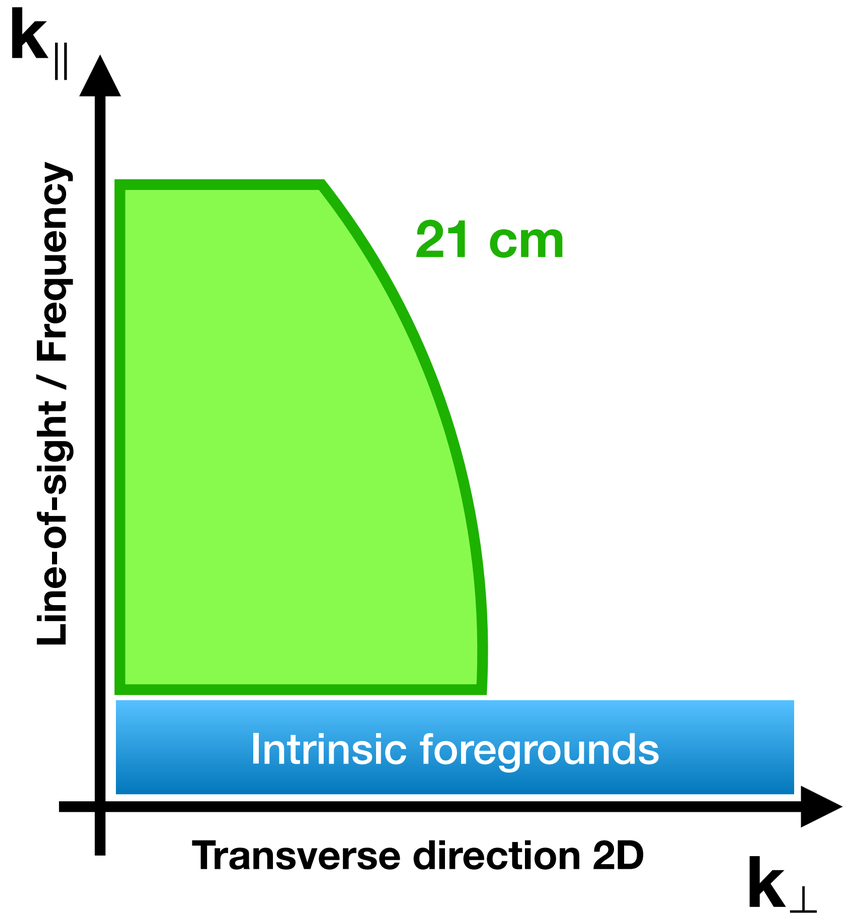
Re-analyis of 2019 data

Preliminary

1. PCA enforced pixel-flagging
2. Keep *bad* channels
3. No re-smoothing
4. Beyond PCA?



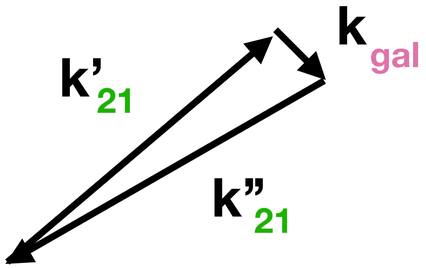
Beat systematics: **cross correlations**



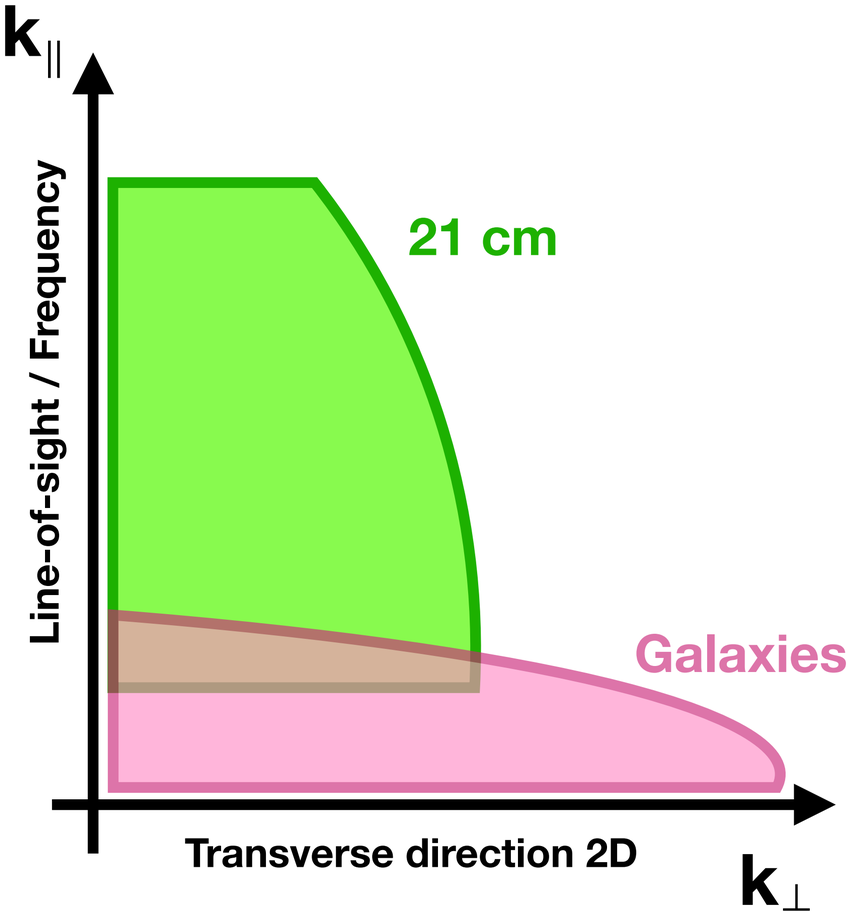
Beat systematics: **cross correlations**

Direct 21cm x galaxies signal vanishes due to foregrounds in long wavelength line-of-sight modes. Need to use higher order correlations.

- e.g., a *squeezed* bispectrum estimator:
  1 low-k mode from galaxy survey X 2 high-k 21 cm modes.

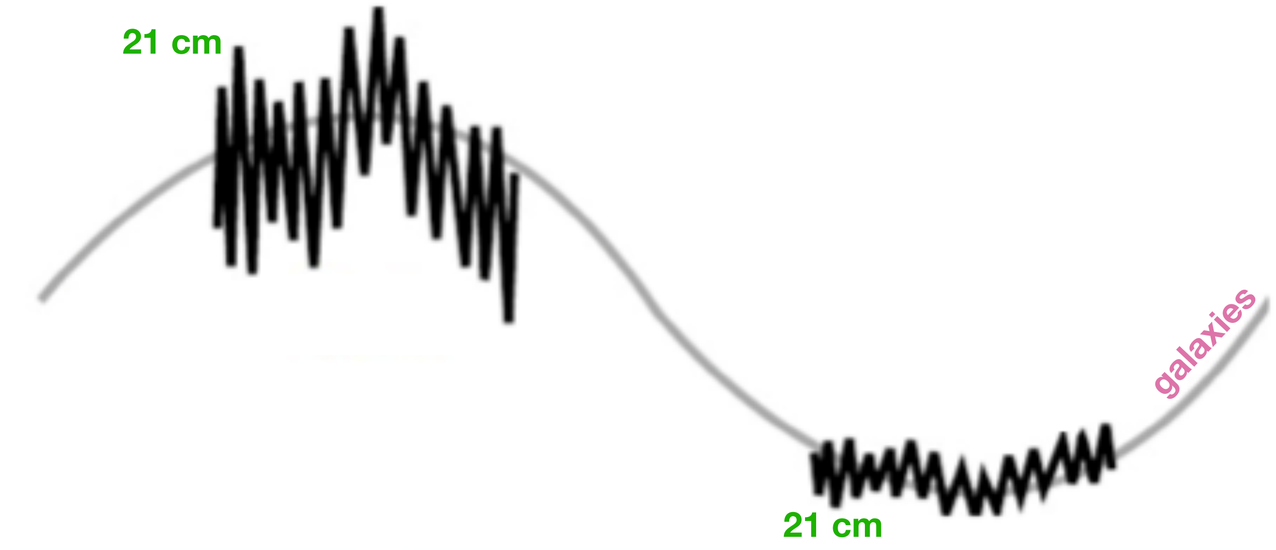


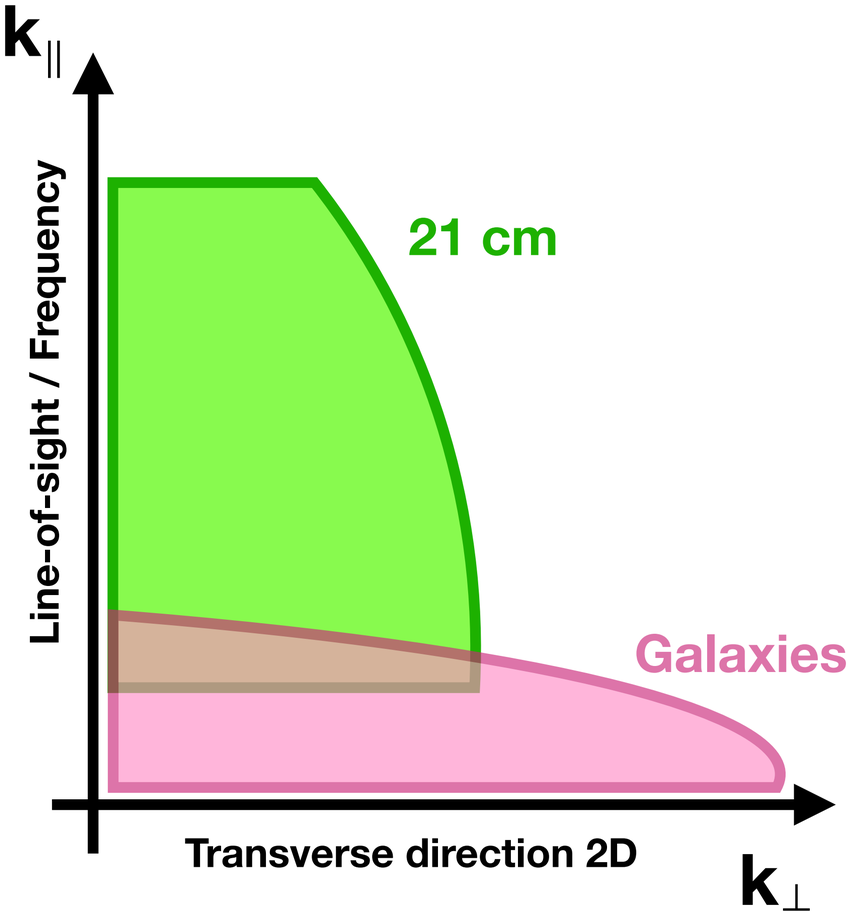
$$\langle \delta(\mathbf{k})\, \delta(\mathbf{k'})\, \delta(\mathbf{k''}) \rangle = \delta_D(\mathbf{k}+\mathbf{k'}+\mathbf{k''})\, B(\mathbf{k}+\mathbf{k'})$$



Beat systematics: **cross correlations**

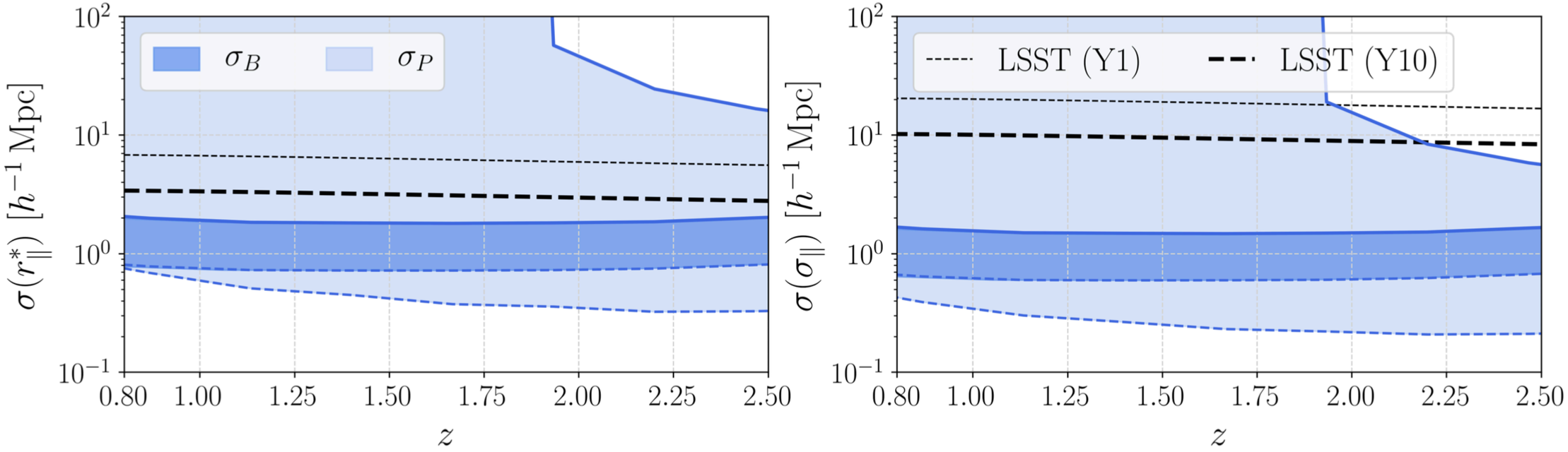
Direct 21cm x galaxies signal vanishes due to foregrounds in long wavelength line-of-sight modes. Need to use higher order correlations.





Beat systematics: **cross correlations**

Direct 21cm x galaxies signal vanishes due to foregrounds in long wavelength line-of-sight modes.
Need to use higher order correlations.



Guandalin+ 2022

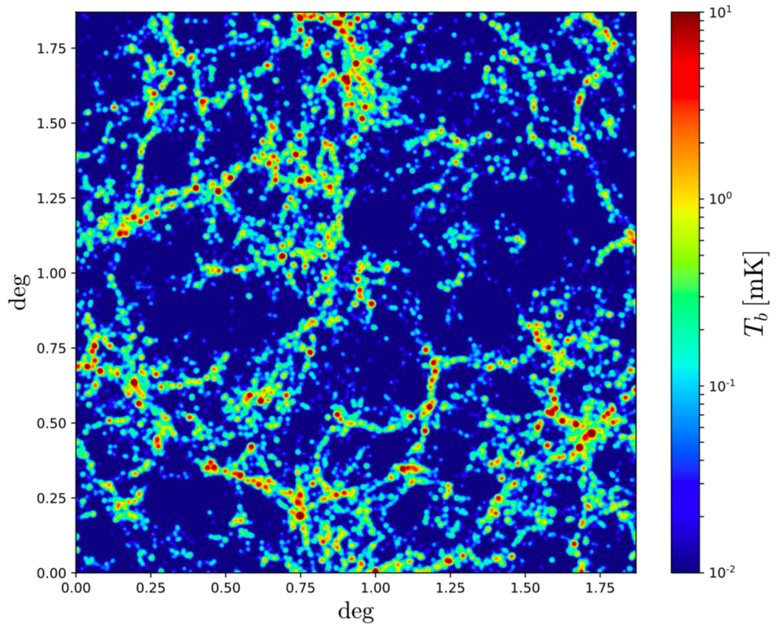
Summary

- **HI IM** will be game changer in cosmology
- Contaminants-removal is the biggest problem, lots of efforts devoted to this
- **MeerKLASS** ongoing!
- We are detecting (again!) the cross signal with WZ galaxies to test different pre-processing steps and cleaning algorithms
- Cross-Pk as sweet-spot as we play with # of components removed
- Cross-Pk in agreement among different cleaning methods
- Separating scales for the cleaning is more efficient at reducing the cube variance (mixGMCA)
- Smart estimators + cleaver cross possibilities: we can get plenty of new/complementary info out of the maps

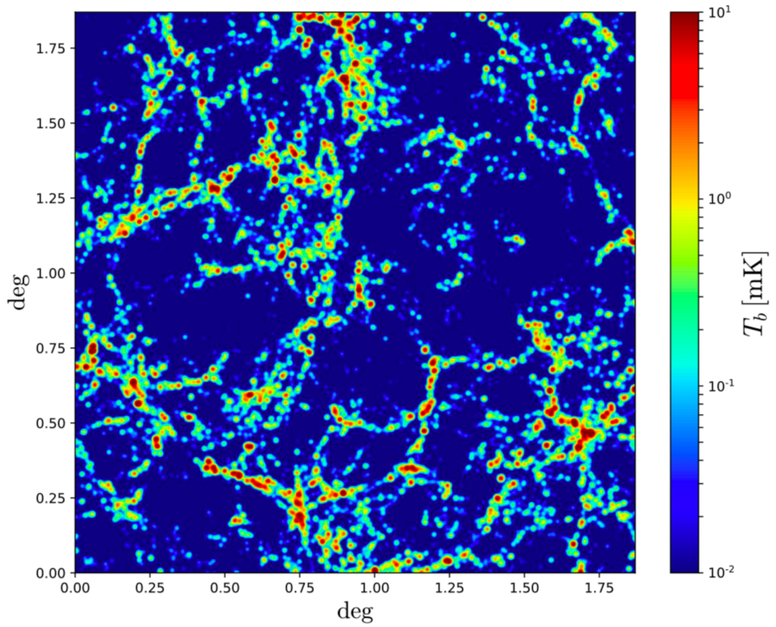
Getting ready for the SKAO HI IM science

Distribution of HI in the Universe

cheap Nbody

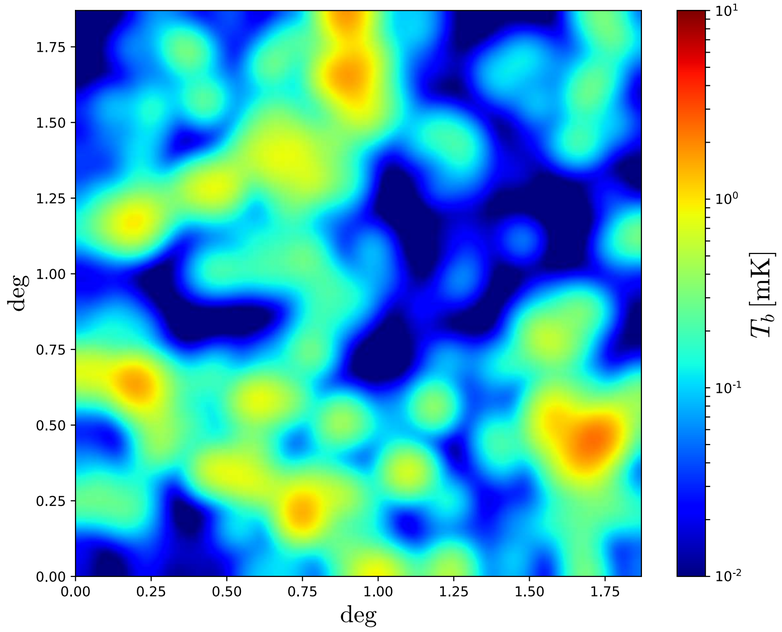


IllustrisTNG



Villaescusa-Navarro, .. , IPC + 2018

forecasts



HI intensity mapping

plot from Laura Wolz