“The Non-Gaussian Universe” workshop - 18/06/2026

📍 Heraklion, Greece

“Foreground removal in HI 21cm intensity mapping”

Athanasia GKOGKOU

Institute of Astrophysics/Institute of Computer Science - FORTH, Greece

Collaborators: Victor Bonjean, Jean-Luc Starck, Marta Spinelli, Pauline Gorbatchev

FORTH INSTITUTE OF COMPUTER SCIENCE

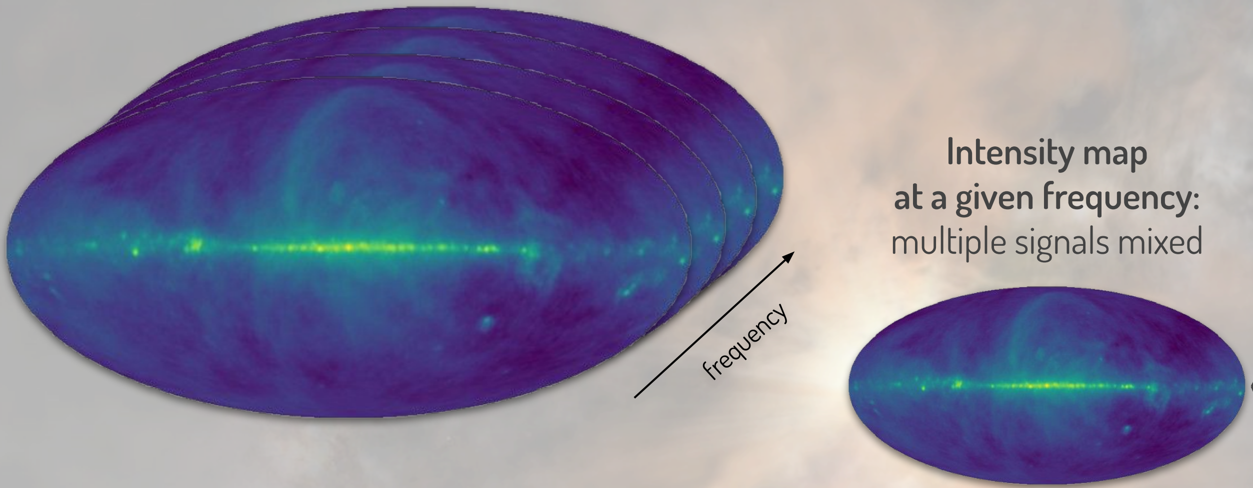
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

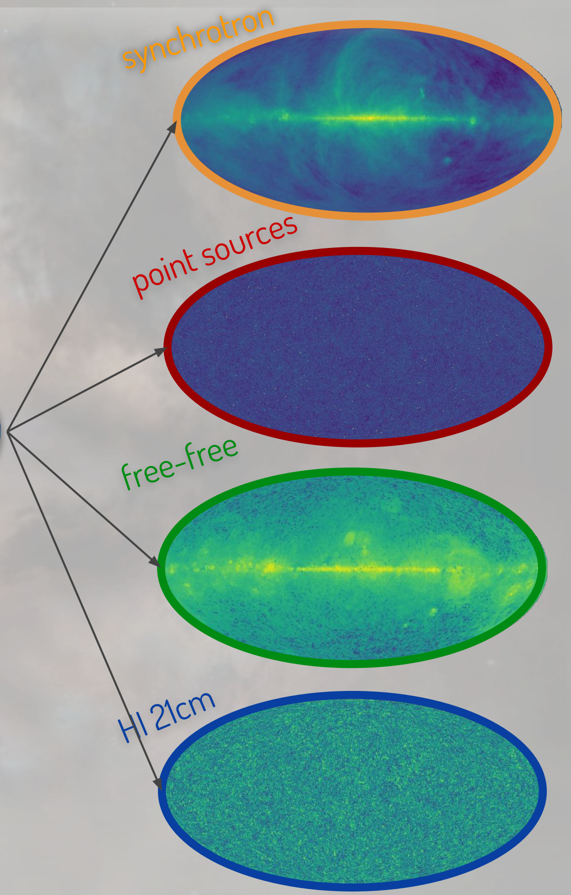
Signal Processing Laboratory

Funded by the European Union

FORTH INSTITUTE OF ASTROPHYSICS

HI 21cm intensity mapping

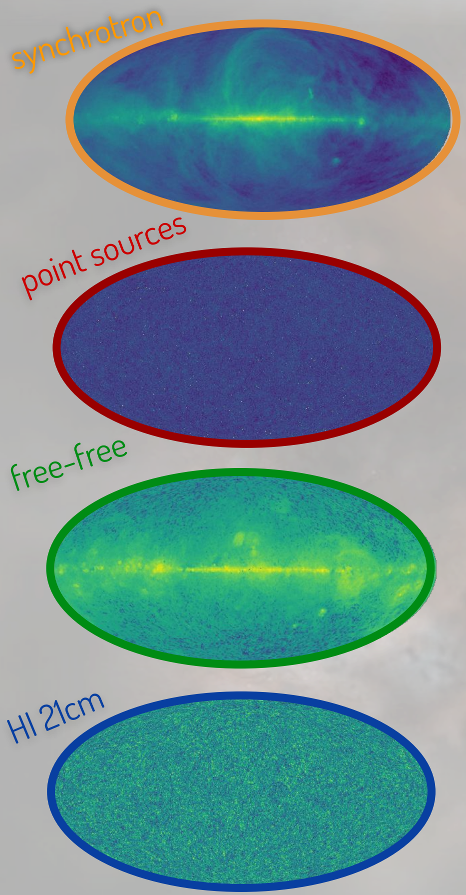


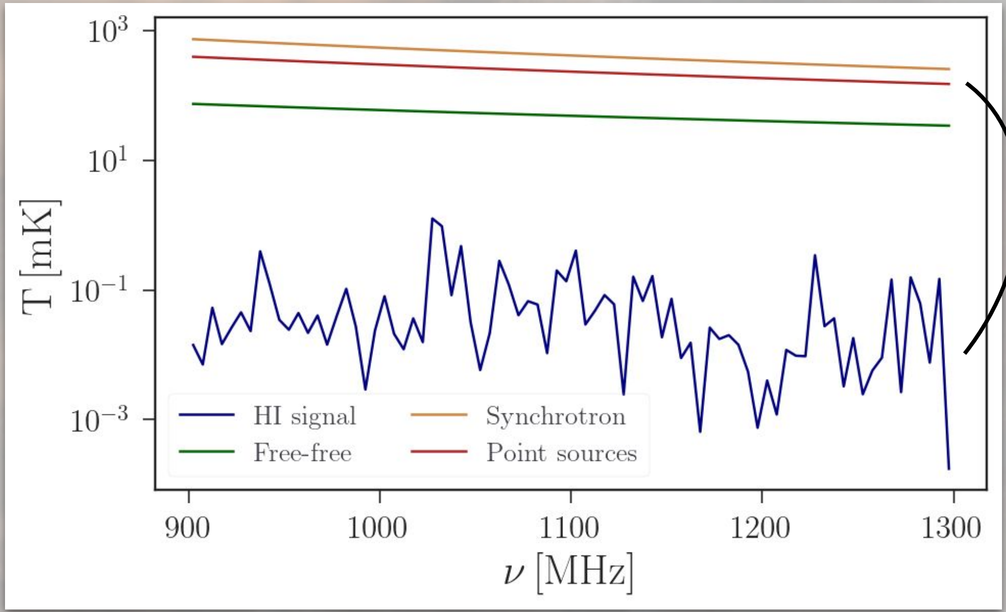


Why <u>intensity mapping</u>?

- Maps large cosmological volumes efficiently
- No need to detect individual galaxies
- Preserves 3D information: angle + frequency/redshift
- Powerful probe of large-scale structure

Foregrounds of the HI 21 cm signal

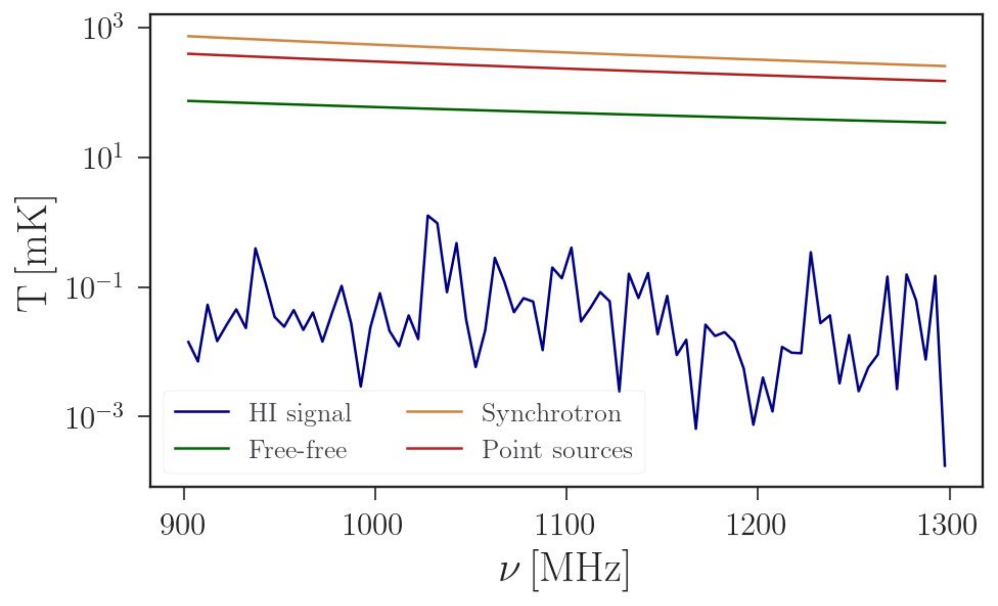




~4 orders of magnitude

Foreground cleaning

What makes foreground cleaning possible?



- Foregrounds are spectrally smooth
- Foregrounds are highly correlated across frequency
- Different components have different spatial morphology / sparsity

Foreground cleaning

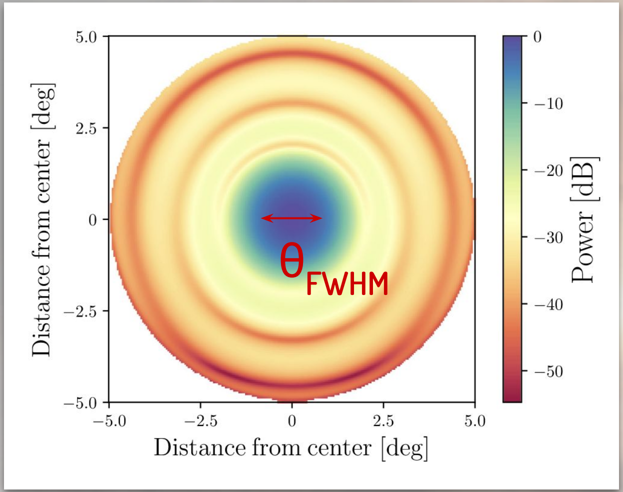
What information do cleaning methods exploit?

- Spectral smoothness → polynomial / parametric fitting
- Low-rank subspace → PCA / ICA
- Sparsity / morphology → GMCA, SDecGMCA, JESSNet
- Known beam model → joint deconvolution + component separation (SDecGMCA, JESSNet)

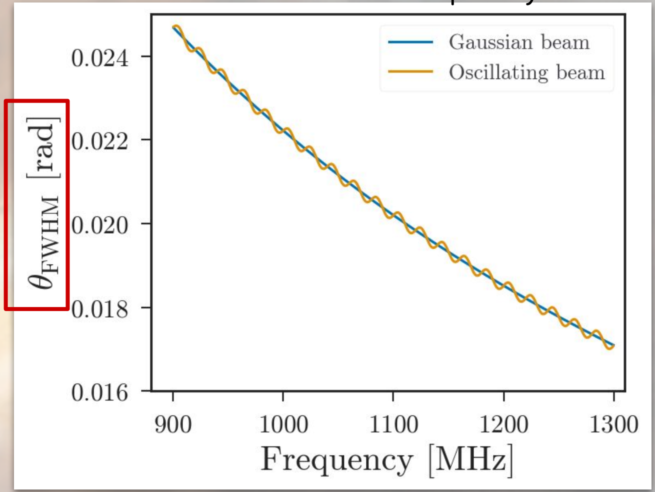
Non smooth frequency evolution of the beam can complicate things further

Chromatic telescope beam

modelled MeerKAT beam behavior



main-lobe width vs frequency

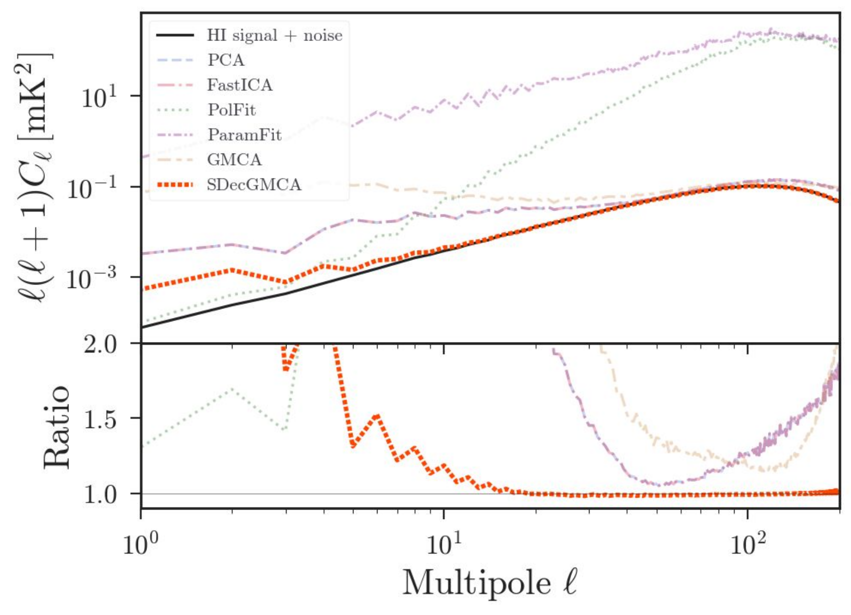


$$\theta_{\mathrm{FWHM}}(\nu) = \frac{\lambda}{D} \longrightarrow \theta_{\mathrm{FWHM}}(\nu) = \frac{\lambda}{D}\left[\sum_{d=0}^{8} a_d\,\nu^d + A\sin\left(\frac{2\pi\nu}{T}\right)\right]$$

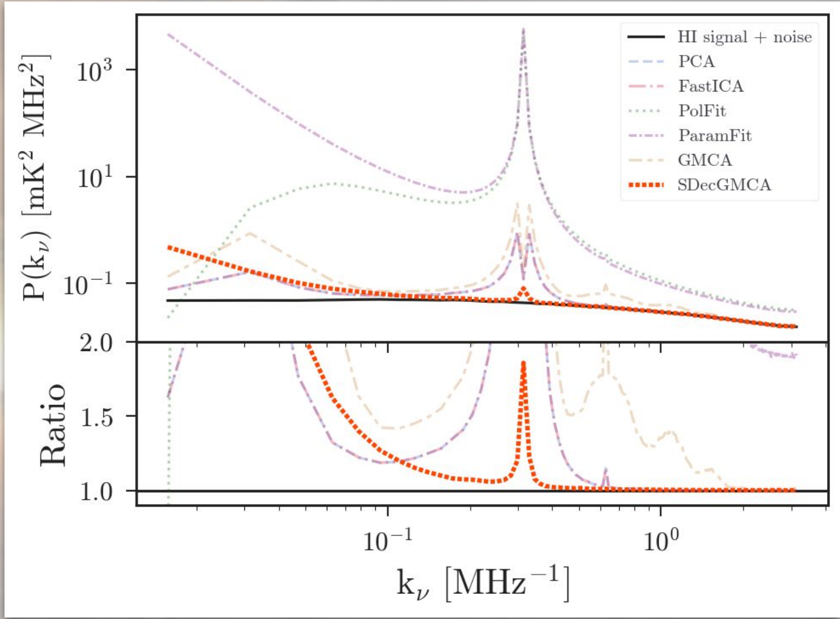
Matshawule et al. (2021)

Need for deconvolution

Angular power spectrum



Frequency power spectrum



➔ The **method** that takes the beam into account has the smallest deviations from the ground truth

➔ Strong indication for a beam-aware foreground cleaning pipeline

Gkogkou et al. (2026)

JESSNet (Joint dEconvolution and Sparse Separation Network)

$$X_\nu = \overbrace{\boxed{(A_\nu S) * H_\nu}}^{\text{foregrounds}} + \overbrace{\boxed{N_\nu}}^{\text{HI+noise}}$$

spherical harmonic space

$$\hat{X}^{\ell m} = \mathrm{diag}(\hat{H}^{\ell})\, A\, \hat{S}^{\ell m} + \hat{N}^{\ell m}$$

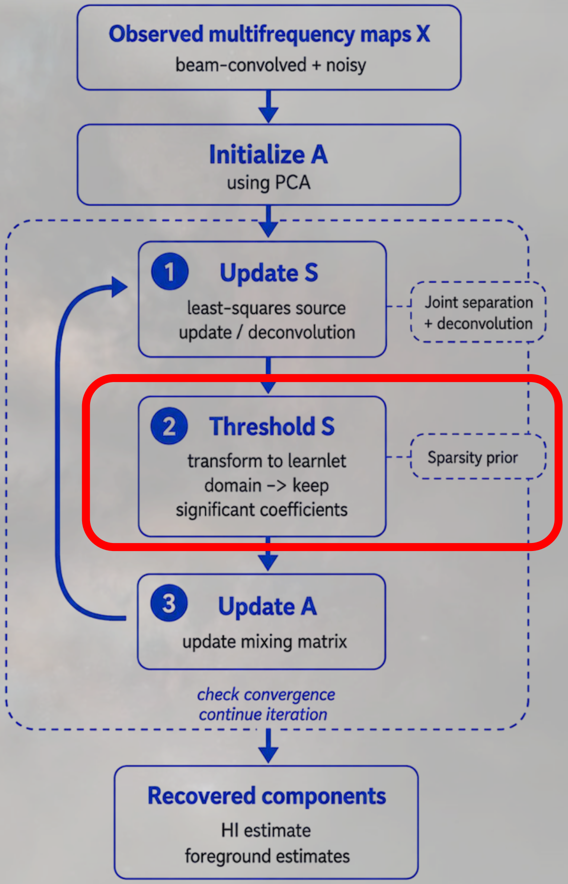
$$\min_{A,\hat{S}} \sum_{\ell m} \left\| \hat{X}^{\ell m} - \mathrm{diag}(\hat{H}^{\ell})\, A \hat{S}^{\ell m} \right\|_2^2 + \lambda\, \mathcal{R}_{\text{sparse}}(S)$$

fit the beam-convolved data **sparsity prior on sources**

$$\boxed{\hat{S}^{\ell m} = \left[A^\top \mathrm{diag}(\hat{H}^{\ell})^2 A + \epsilon_\ell\right]^{-1} A^\top \mathrm{diag}(\hat{H}^{\ell})\, \hat{X}^{\ell m}}$$

$$\boxed{\hat{A}_\nu \leftarrow \left(\sum_{(\ell m)\in D} \hat{X}_\nu^{\ell m} \hat{H}_\nu^{\ell} \hat{S}^{\ell m \dagger}\right)\left(\sum_{(\ell m)\in D} (\hat{H}_\nu^{\ell})^2\, \hat{S}^{\ell m} \hat{S}^{\ell m \dagger}\right)^{-1}}$$

Tikhonov regularization param



JESSNet (Joint dEconvolution and Sparse Separation Network)

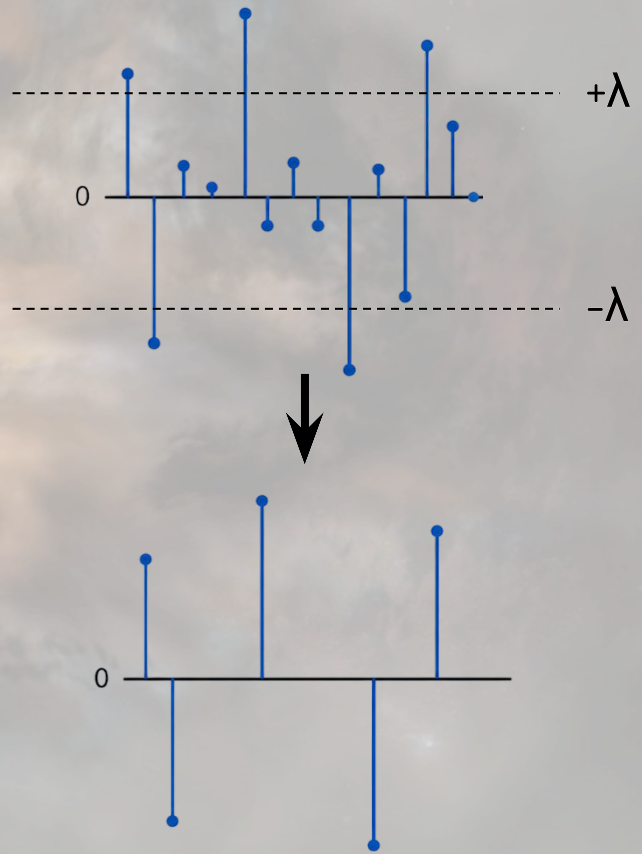
Sparsity and thresholding

Sparse representation

- In the wavelet / learnlet domain, the signal is represented by **a few large coefficients**
- Most coefficients are close to zero
- Noise mainly populates low-amplitude coefficients

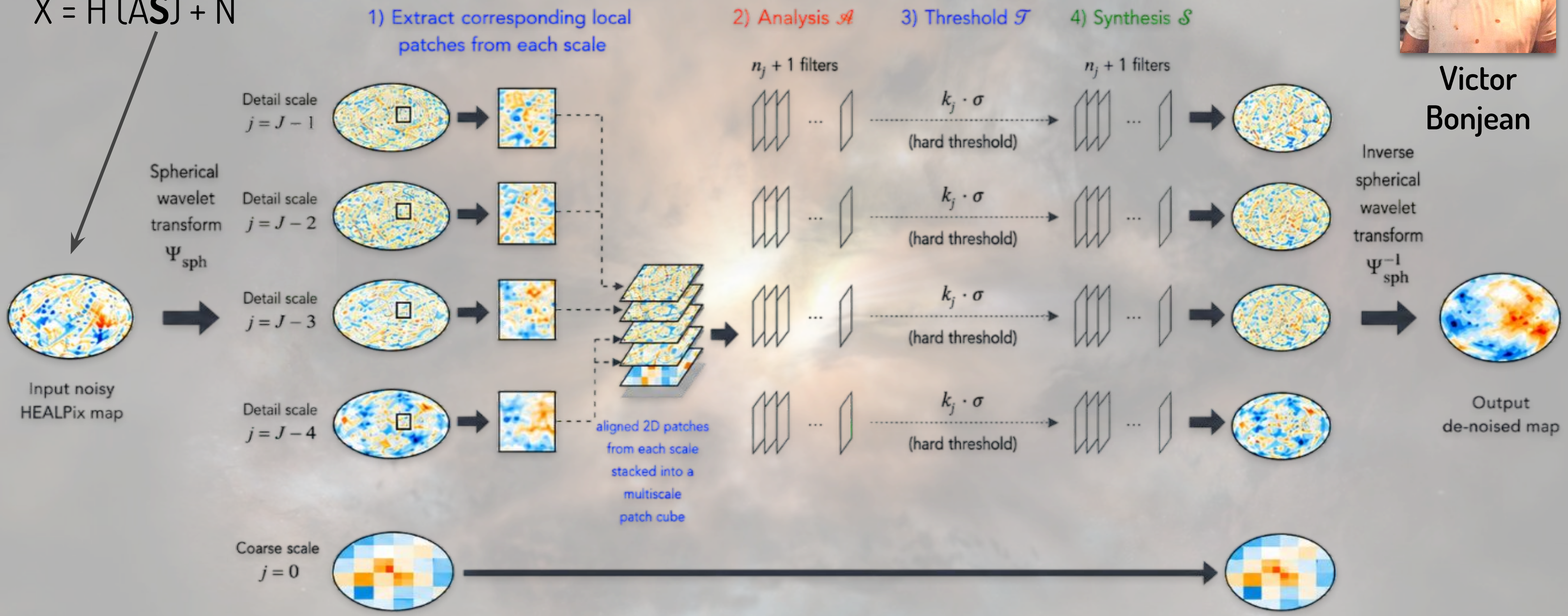
Thresholding

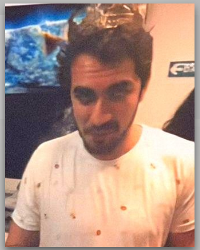
- Keep coefficients with amplitude above a **threshold λ**
- Set small coefficients to zero
- This preserves significant structures and suppresses noise



Learnlets: structure

$X = H(A\mathbf{S}) + N$





Victor Bonjean

Bonjean et al. (2026)
Gkogkou et al. in prep.

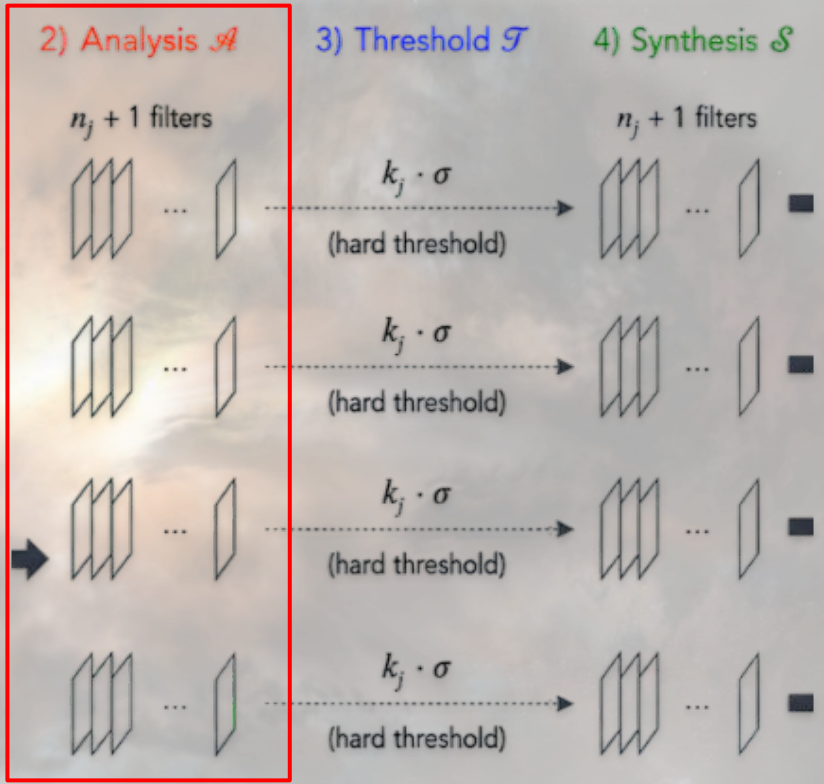
Learnlets: analysis layer

Each detail-scale patch is convolved with a bank of learned filters

The filters are scale- and source-specific

The same filters are applied to all patches from the same scale

The extra (+1) channel is the identity branch for exact reconstruction



Bonjean et al. (2026)
Gkogkou et al. in prep.

Learnlets: thresholding layer

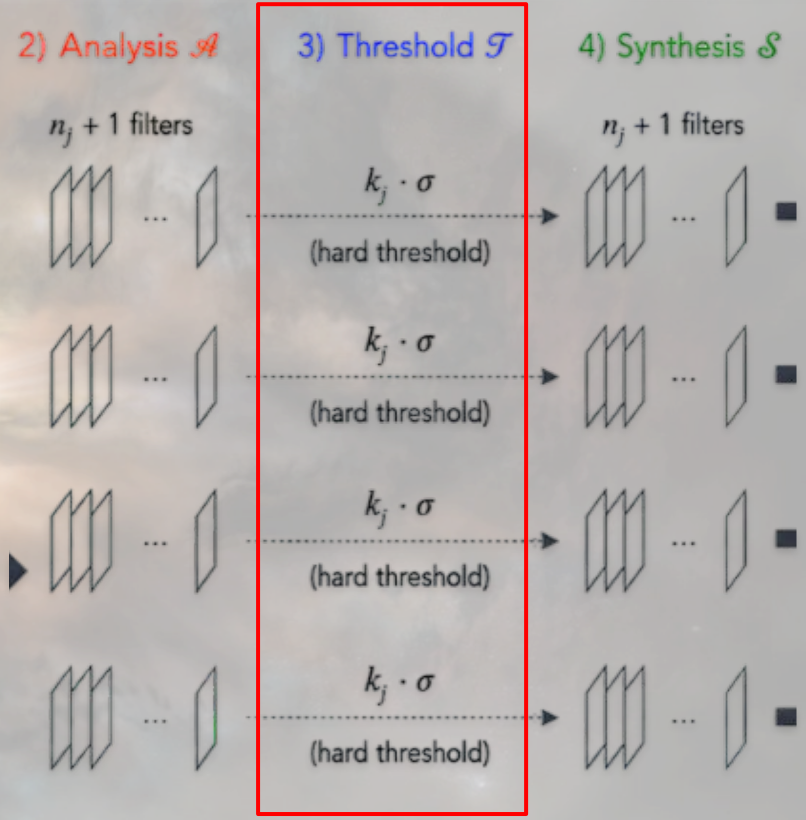
Thresholding promotes sparsity in the learned coefficient domain

Weak coefficients are suppressed

The threshold is computed per source and per scale

k sigma thresholding: $\lambda_{s,j} = k_j \times \sigma_{NMAD}$

k_j comes from the learnlet network through mininet



Bonjean et al. (2026)
Gkogkou et al. in prep.

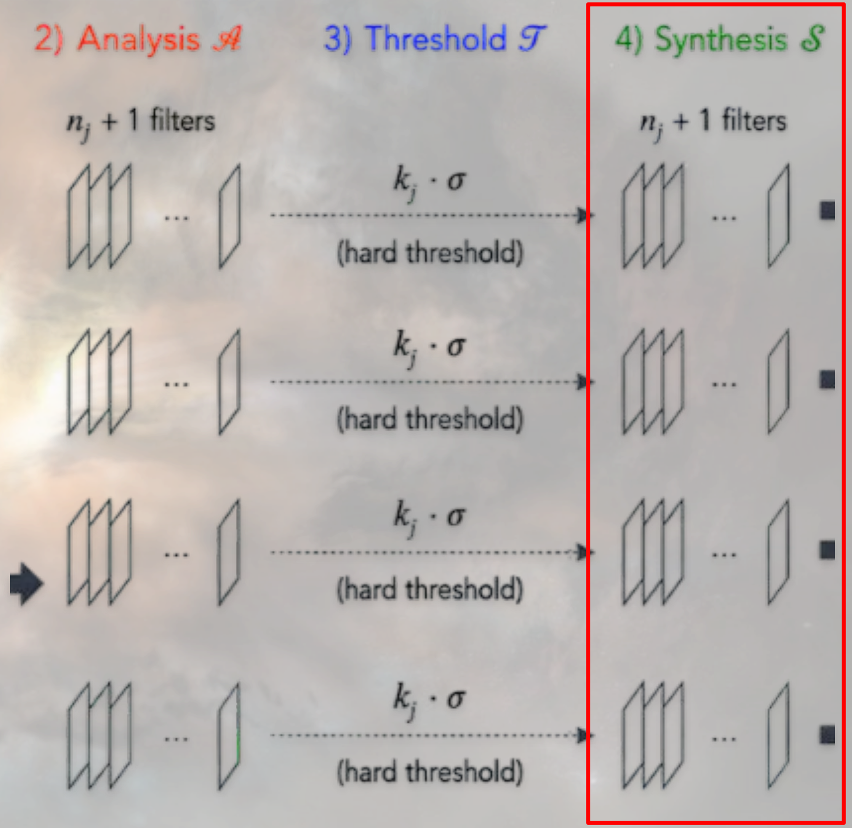
Learnlets: synthesis layer

Converts sparse learned coefficients back into an image-like patch

Uses learned synthesis filters, different from the analysis filters

For each scale and each source

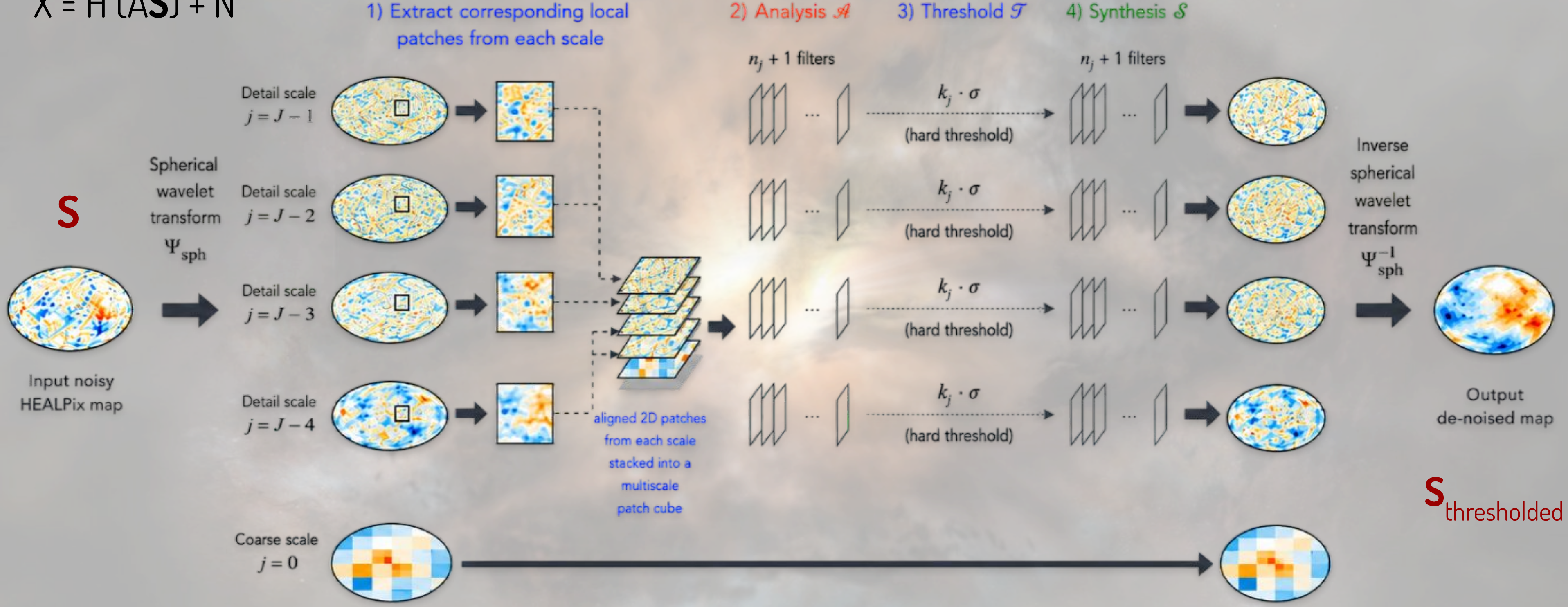
The detail-scale outputs are combined with the preserved coarse scale



Bonjean et al. (2026)
Gkogkou et al. in prep.

Learnlets: structure

$X = H(A\mathbf{S}) + N$

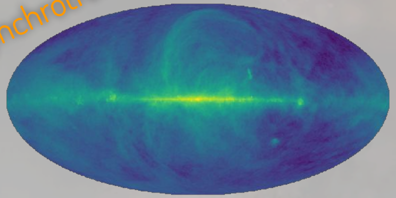


Bonjean et al. (2026)
Gkogkou et al. in prep.

Learnlets: training

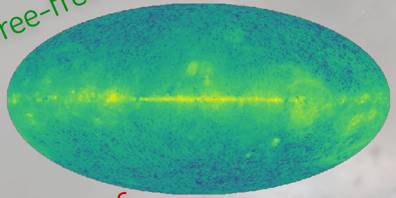
Frequency range: 1000-1100 MHz
Varying spectral indices

synchrotron



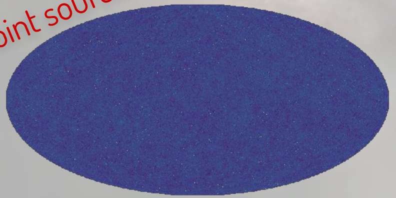
$$T_{\rm sy}(\nu, p) = T_s(p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm sy}(p)}$$

free-free



$$T_{\rm ff}(\nu, p) = T_{\rm ff}(p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm ff}}$$

point sources



$$T_{\rm ps}(\nu, p) = T_{\rm ps}(1.4\,\mathrm{GHz}, p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm ps}(p)}$$

What is learned?

A_j: analysis filters (n_j)

S_j: synthesis filters (n_j)

k_j: scale-dependent threshold multipliers (J-1)

Training target

Noisy multiscale patch cube → clean multiscale patch cube

Purpose

Learn a foreground-adapted denoising / sparse representation

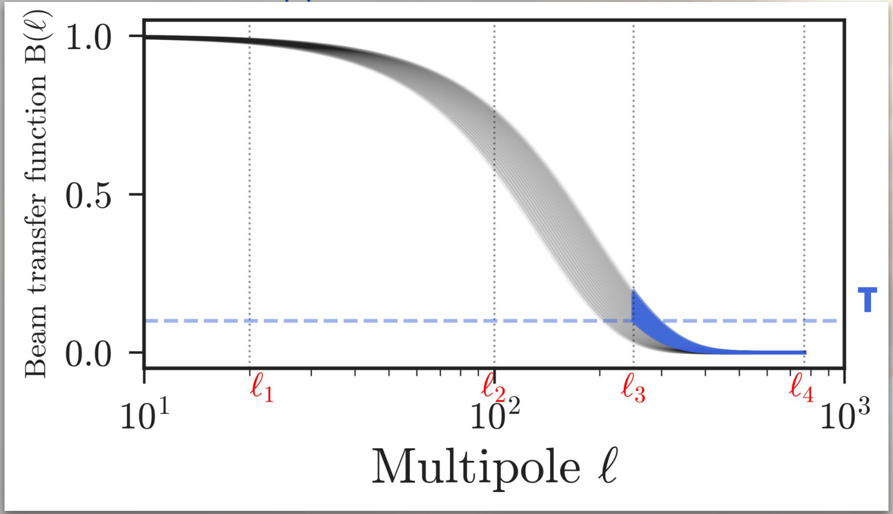
Per scale JESSNet

Windowed data

$$\widehat{X}^{(j)}_{\nu,\ell m} = W_j(\ell)\,\widehat{X}_{\nu,\ell m}$$

ℓ bin: exclude channels with severe beam suppression

$$B_{\nu_b}(\ell_b) \geq \tau$$



HI reconstruction in each ℓ bin & recombination

hard binning

$$\widehat{\mathrm{HI}}_{\ell m} = \begin{cases} \widehat{\mathrm{HI}}^{(1)}_{\ell m}, & \ell < \ell_1 \\ \widehat{\mathrm{HI}}^{(2)}_{\ell m}, & \ell_1 \leq \ell < \ell_2 \\ \vdots & \\ \widehat{\mathrm{HI}}^{(N_b)}_{\ell m}, & \ell \geq \ell_{\mathrm{last}} \end{cases}$$

any other binning

$$\widehat{\mathrm{HI}}_{\ell m} = \frac{\sum_j W_j(\ell)\,\widehat{\mathrm{HI}}^{(j)}_{\ell m}}{\sum_j W_j(\ell)}$$

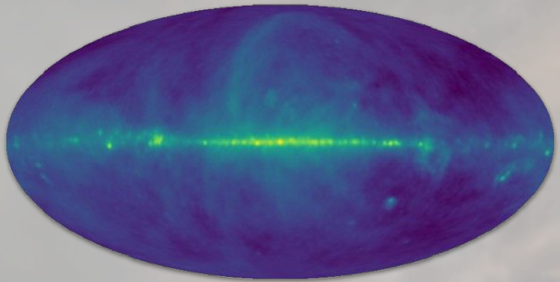
From spherical harmonics back to pixel space

$$\mathrm{HI}(\hat{\mathbf{n}}) = \sum_{\ell m} \widehat{\mathrm{HI}}_{\ell m}\, Y_{\ell m}(\hat{\mathbf{n}})$$

JESSNet: masked

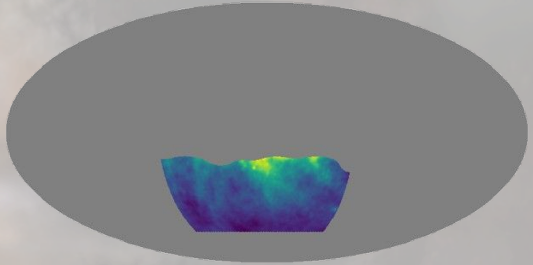
Full sky

$$X_\nu = (A_\nu S) * H_\nu + N_\nu$$

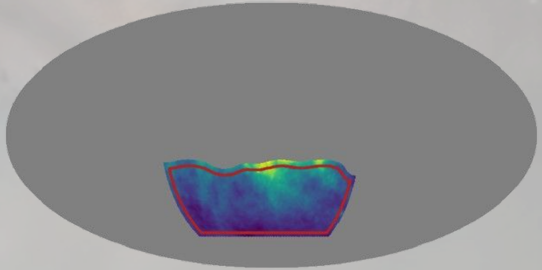


masked region

$$X_\nu^{\text{mask}} = M\left[(A_\nu S) * H_\nu\right] + M N_\nu$$



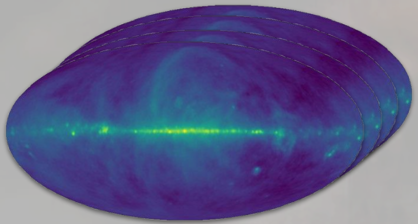
$$X_\nu^{\text{mask}} \simeq (A_\nu M S) * H_\nu + M N_\nu$$



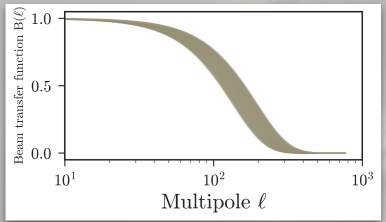
JESSNet: summary

INPUT

Observations (cube)
Full sky / survey footprint



Input the telescope beam



Choose the ℓ binning

Joint **deconvolution** & component separation with JESSNet

Estimating **A** and **S**
Thresholding S with **learnlets**

Apply it **per ℓ scale**

OUTPUT

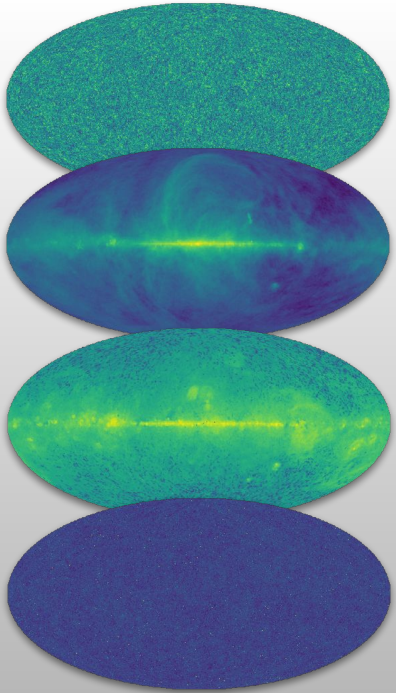
Estimated foregrounds

Residuals = obs - estimated foregrounds

HI+noise = residuals

Foreground cleaning pipeline

Simulate the data



* Beam

Pre-processing

Masking the HI-bright sources

Apply the methods

PCA
FastICA
GMCA
Polinomial Fit
Parametric fit
JESSNet

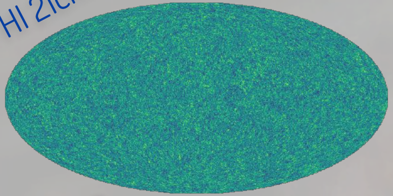
Validation

Radial (frequency) power spectrum **$P(k_\nu)$**

Angular power spectrum **C_ℓ**

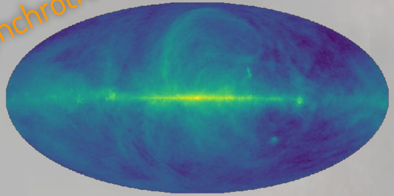
Simulating the data

HI 21cm



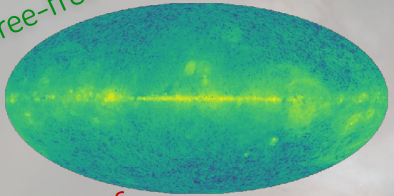
Cosmological simulations
CoLoRe

synchrotron



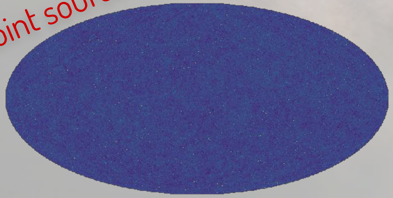
$$T_{\rm sy}(\nu, p) = T_s(p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm sy}(p)}$$

free-free



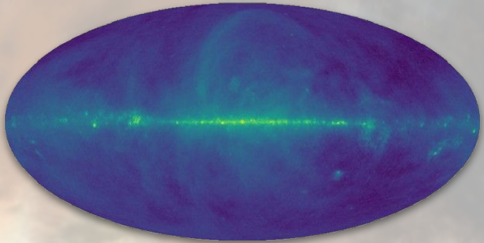
$$T_{\rm ff}(\nu, p) = T_{\rm ff}(p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm ff}}$$

point sources



$$T_{\rm ps}(\nu, p) = T_{\rm ps}(1.4\,{\rm GHz}, p)\left(\frac{\nu}{\nu_0}\right)^{\beta_{\rm ps}(p)}$$

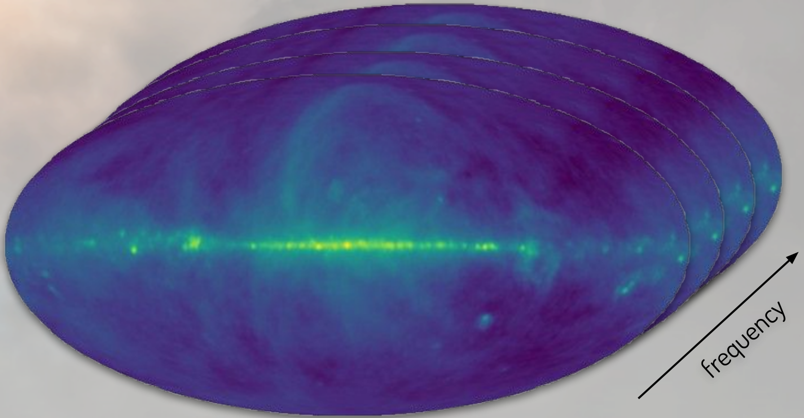
+



* telescope beam

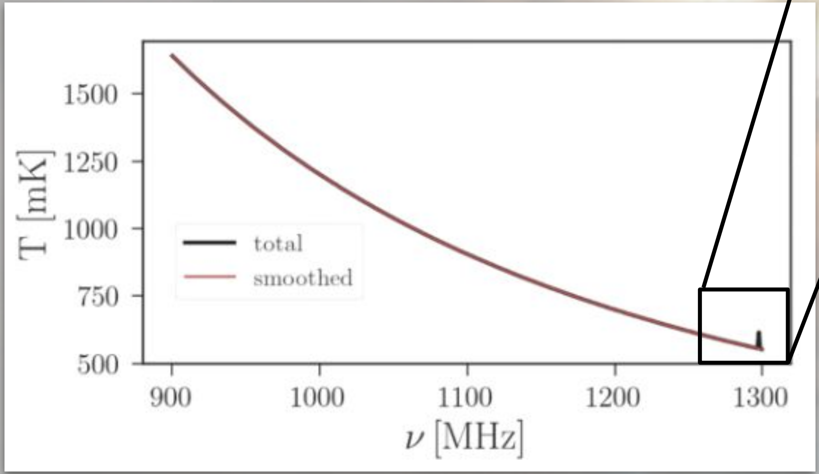
+ instrumental noise

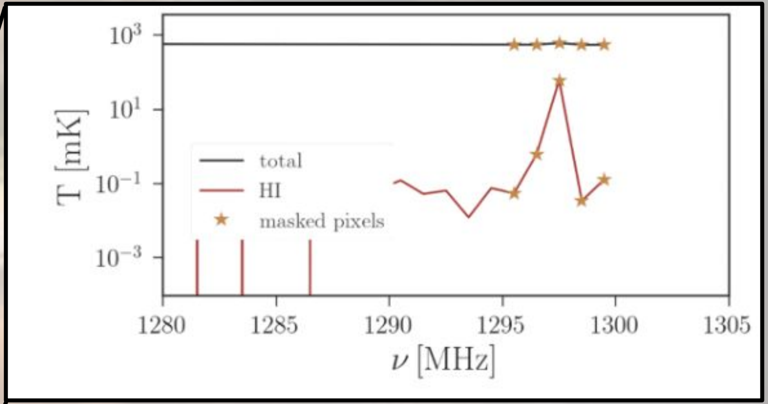
=



Pre-processing

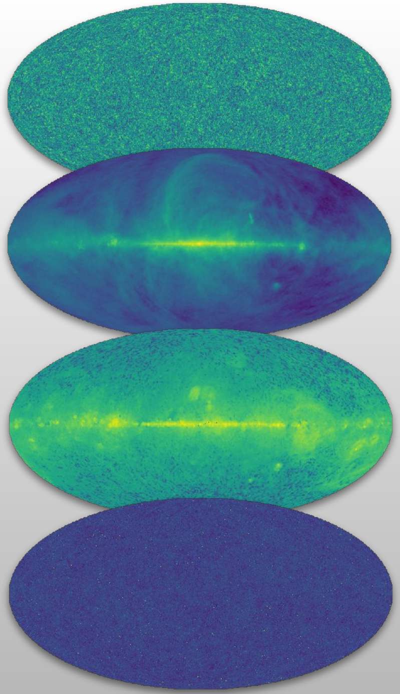
- Detecting lines of sight with peaks (extremely bright HI-sources)
- Assigning the baseline value to this pixel





Foreground cleaning pipeline

Simulate the data



* Beam

Pre-processing

Masking the HI-bright sources

Apply the methods

PCA
FastICA
GMCA
Polinomial Fit
Parametric fit
JESSNet

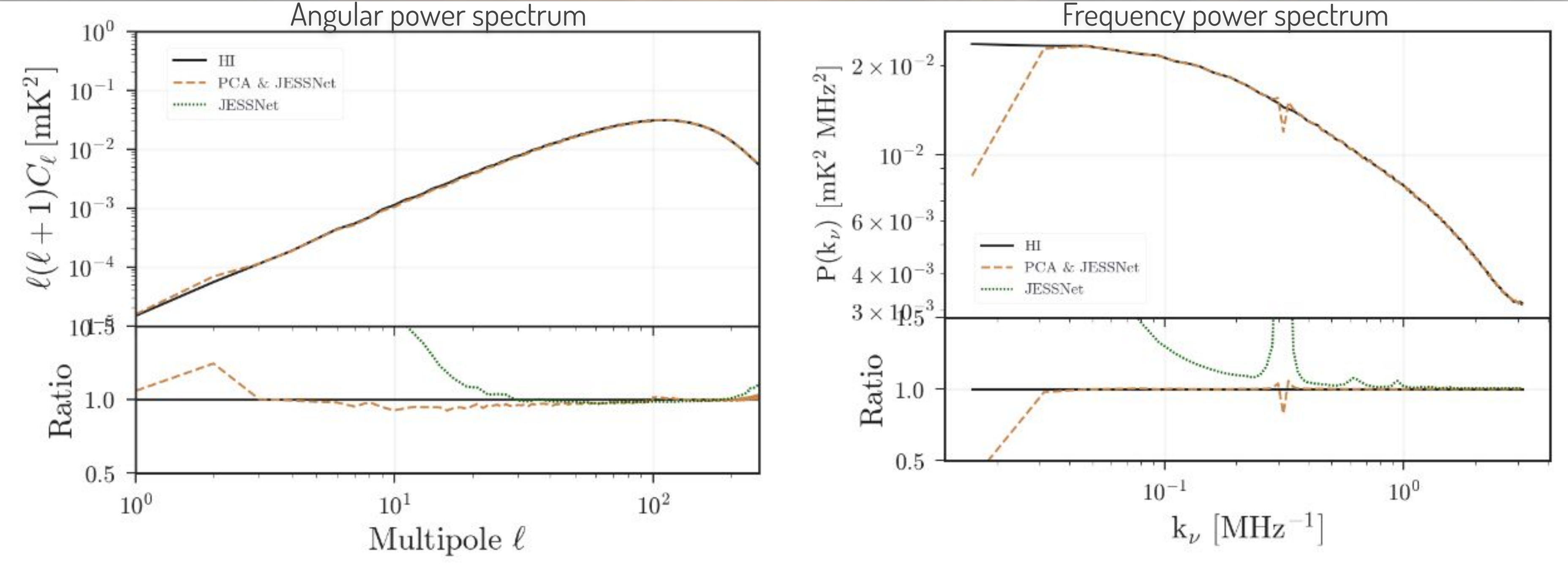
Validation

Radial (frequency) power spectrum $\mathbf{P(k_\nu)}$

Angular power spectrum $\mathbf{C_\ell}$

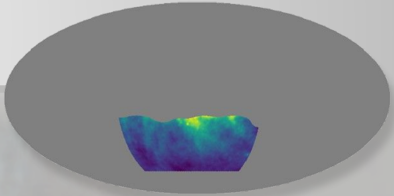
Results on full sky

All scales VS. per-scale cleaning



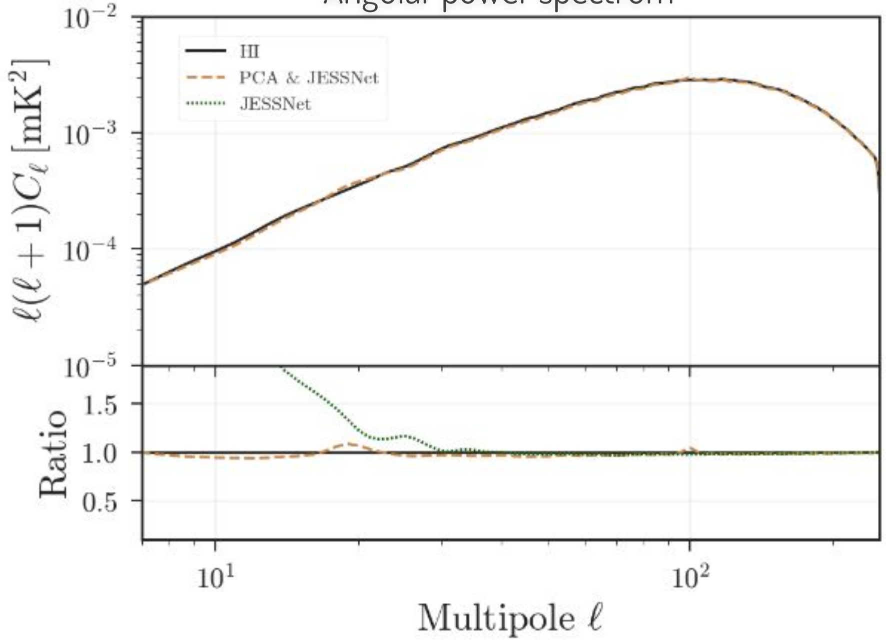
Gkogkou et al. in prep.

Results on a footprint

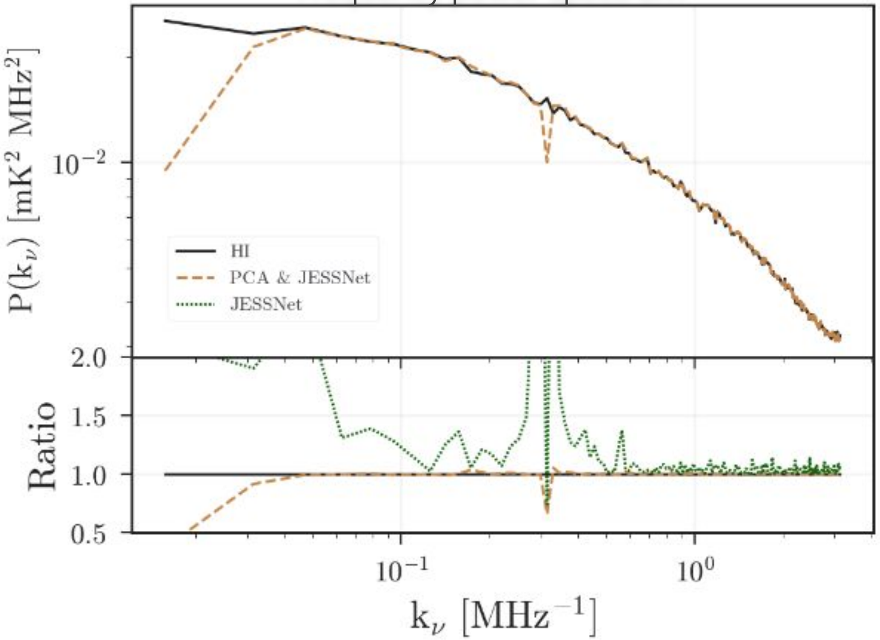


All scales VS. per-scale cleaning

Angular power spectrum



Frequency power spectrum



Gkogkou et al. in prep.

Summary & Conclusions

- **Foreground removal is a central challenge for HI intensity mapping**
- **A non-smooth beam evolution imprints artificial spectral structure on foregrounds**
  They imprint artificial frequency structure on otherwise smooth foregrounds
- **This motivates joint deconvolution and component separation**
  The beam should be included directly in the cleaning model
- **`JESSNet`** provides a beam-aware sparse separation framework
  - Joint beam deconvolution + component separation
  - Learnlet-based sparse regularization
  - Applicable to partial-sky / survey-footprint data
  - Per-scale cleaning improves flexibility and recovery
  - **Per-scale `JESSNet` gives better HI recovery in both full-sky and footprint**