Deep CosmoStat Days

17/01/2026

CEA Paris-Saclay, France

Denoising and Signal Restoration for (high-redshift) extragalactic data: from simulations to observations

ARNAB LAHIRY

PhD Student

IA / ICS -FORTH

Supervised by:

Dr. Jean-Luc Starck

CEA Paris-Saclay / IA-FORTH

Dr. Tanio Diaz-Santos

IA - FORTH

Funded by the European Union

Current project: Signal restoration, flux preservation and denoising in spectral cubes

Challenges with high-redshift IFUs

- **Low SNR**: Weak signals dominated by noise due to large cosmological distances.
- **Instrumental Noise**: Detector artifacts and sky subtraction errors.
- **Convolution Effects**: PSF and beam smearing may distort flux values affecting flux conservation.
- **Denoising Bias**: Aggressive methods may lead to over-smoothing and can suppress real signals

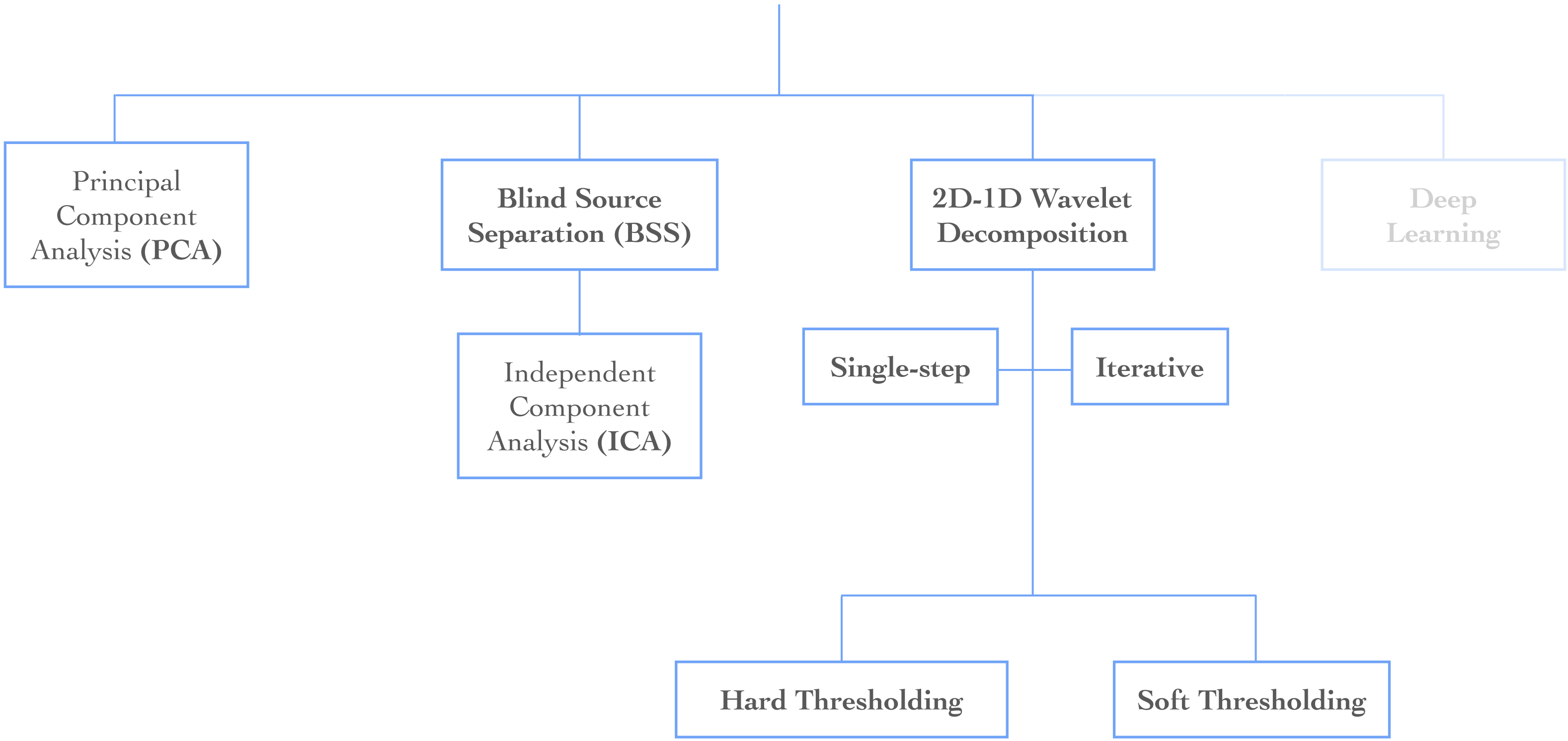
A I M S

To **analyse and compare denoising methods** for **spectral cubes** across a **broad parameter space**, focusing on:

- **Noise characteristics**: Gaussian noise (ALMA), cosmic ray artifacts (JWST), vertical/horizontal stripes (MUSE)
- **Noise levels**: Varying signal-to-noise ratios
- **Spatial resolutions**: Different beam sizes affecting the resolution

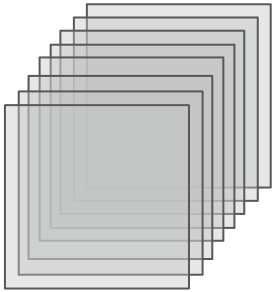
And understand how each method performs under different conditions and identify the **optimal approach** for **flux conservation** and **denoising,** for specific datasets and noise characteristics, with application to observational data

Denoising Methods



Datasets

- Toy models of rotating galaxies
- Mock IFU data from cosmological simulations
- Observational data (ALMA, JWST, MUSE)



Spectral cubes

multiple spectral observation of the same spatial area, where each *(x,y)* spatial point corresponds to a spectrum

Constructing toy data of rotating galaxies

Step 1: Describing the flux density profile in 3D space

$$F(x, y, z) = F_e \exp\left[-b_n\left(\left(\frac{\sqrt{x^2 + y^2}}{R_e}\right)^{1/n} - 1\right)\right] \cdot \exp\left(-\frac{|z|}{h_z}\right)$$

Sérsic profile

Exponential profile

Constructing toy data of rotating galaxies

Step 1: Describing the flux density profile in 3D space

Sérsic profile

$$F(x,y,z) = F_e \exp\left[-b_n\left(\left(\frac{\sqrt{x^2+y^2}}{R_e}\right)^{1/n} - 1\right)\right] \cdot \exp\left(-\frac{|z|}{h_z}\right)$$

Exponential profile

Flux density at a given location in 3D space

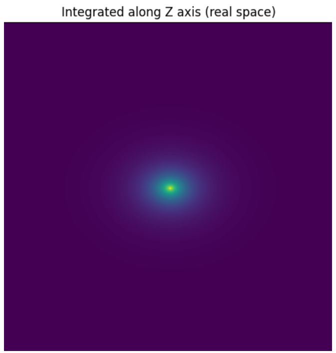
Constant depending on n, ensuring that the effective radius encloses half of the total light

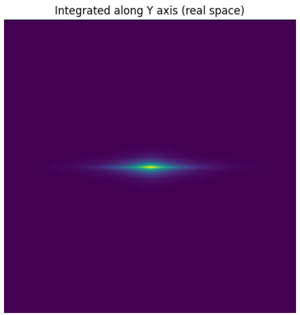
Sérsic index: determines the shape of the profile

Scale height: determines how the flux density varies above or below the galactic mid-plane.

Effective flux density: flux contained within the half-light radius

Effective/half-light radius: radius at which half of the total light of the galaxy is contained





Constructing toy data of rotating galaxies

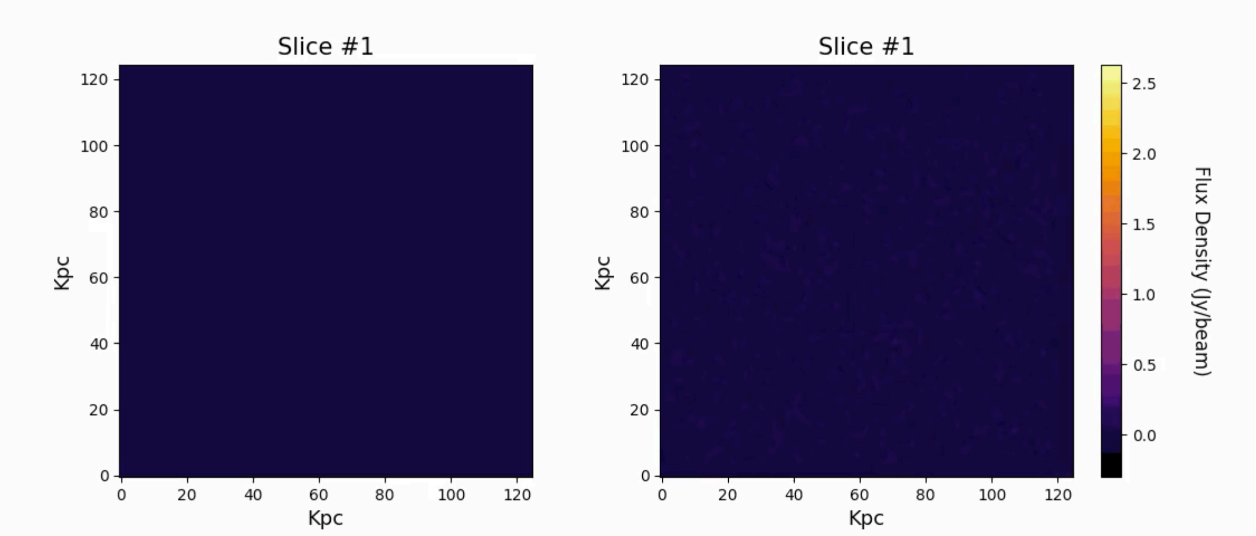
- Rotation velocity vectors (v_x, v_y, v_z) calculated **in the plane of the disk**

$$v = v_0 \times 1.022 \times \left(\frac{R}{R_0}\right)^{0.0803}$$

- The entire system is rotated and ***Z* axis is chosen as the line of sight**

- $\mathbf{n_z}$ 2D projections are made along line of sight based on $\mathbf{v_z}$ bins

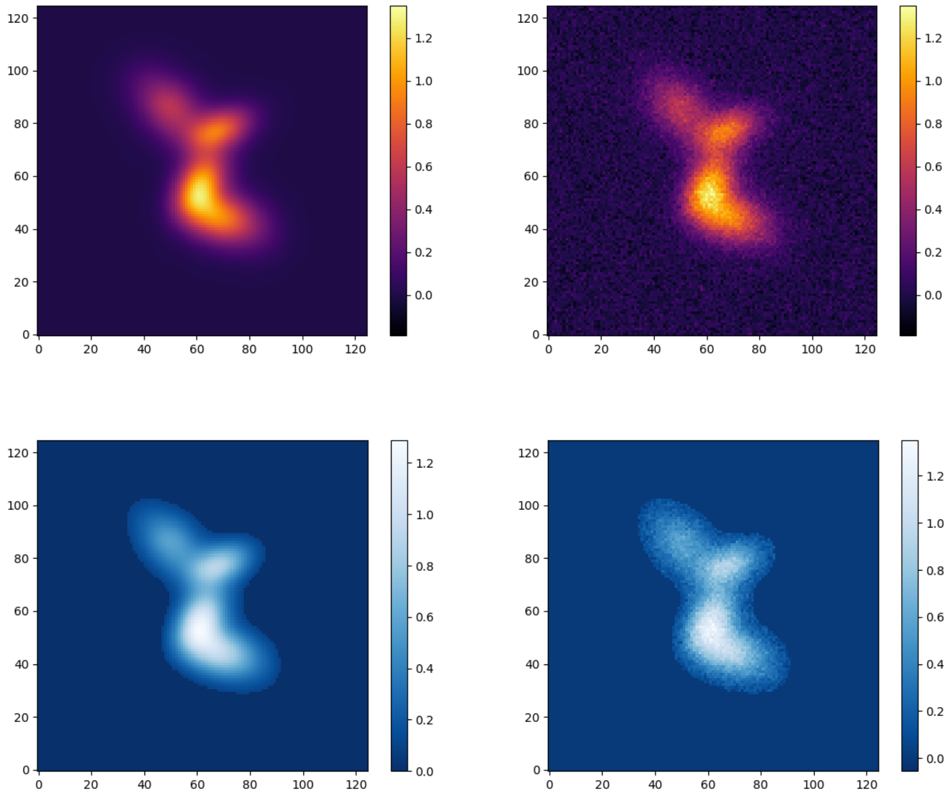
- **Gaussian noise** is overlaid onto the cube



Noise-free
(ground truth)

Noisy data
(input)

Identifying emission regions



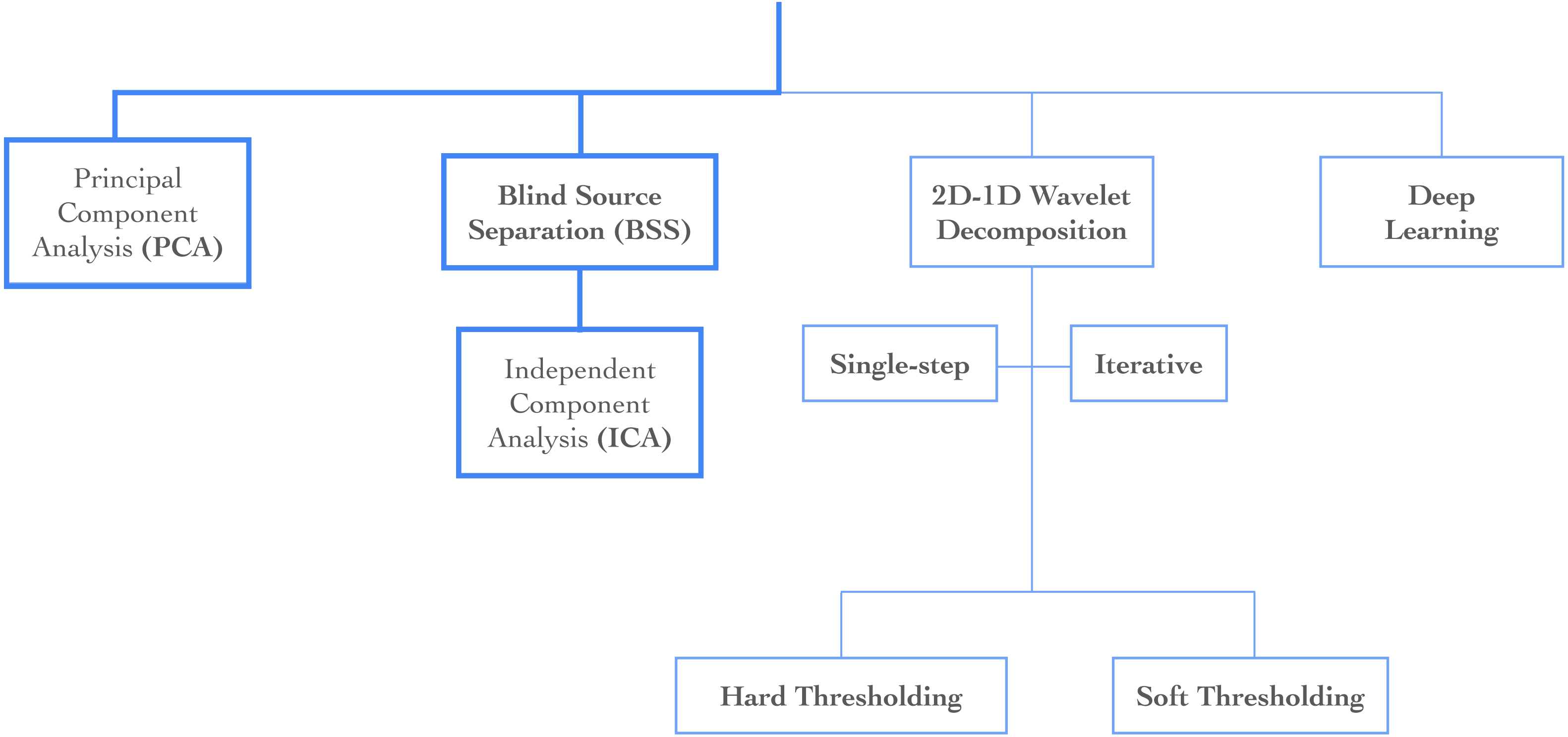
Slice #10 of the clean and noisy toy cube and flux emission regions (masked)

Masks constructed using **`astrodendro`**, which identifies regions with strong emission in the whole cube

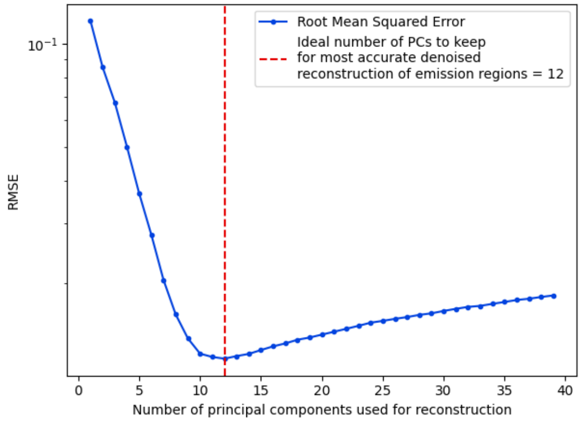
I N I T I A L

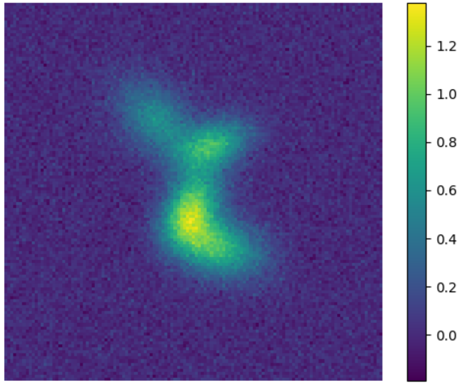
Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Denoising Methods

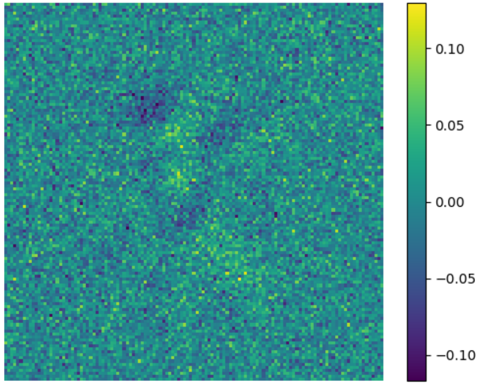


PCA based denoising

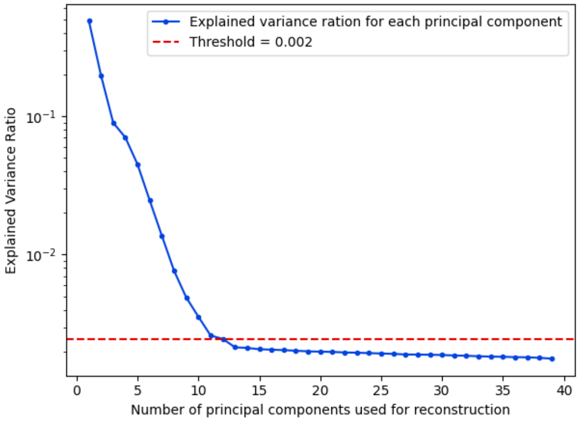


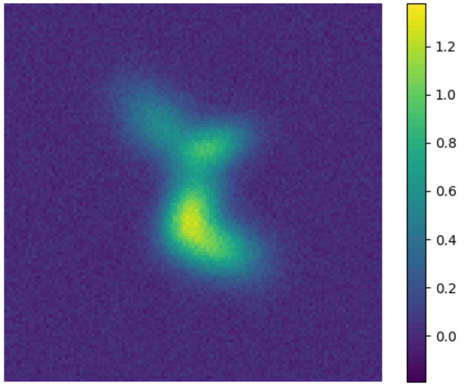


Noisy input



Residual





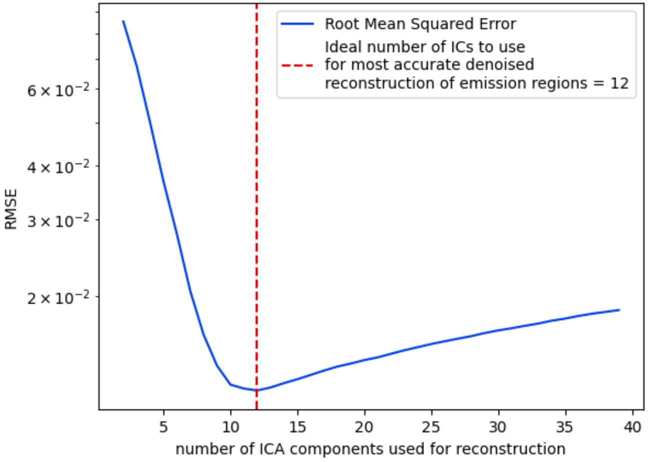
PCA denoised

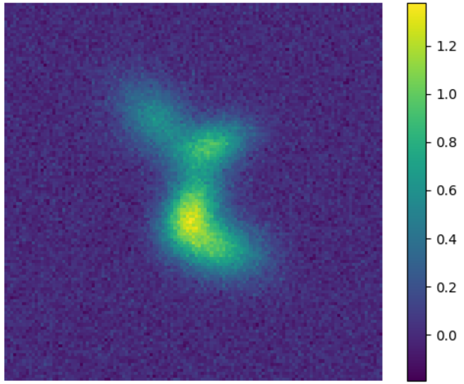
I N I T I A L

Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

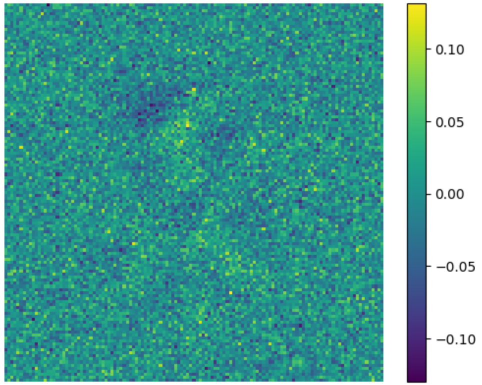
Noise std. deviation: 3.016e-02
Emission RMSE: 1.216e-02
Average flux: 5.417-02

BSS/ICA based denoising

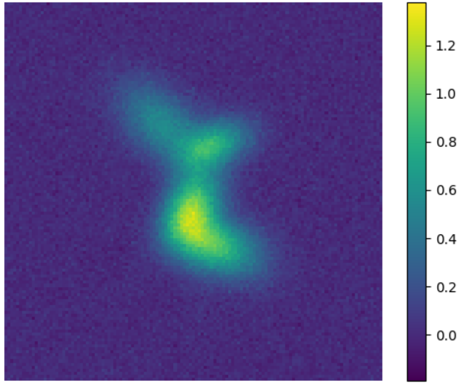




Noisy input



Residual



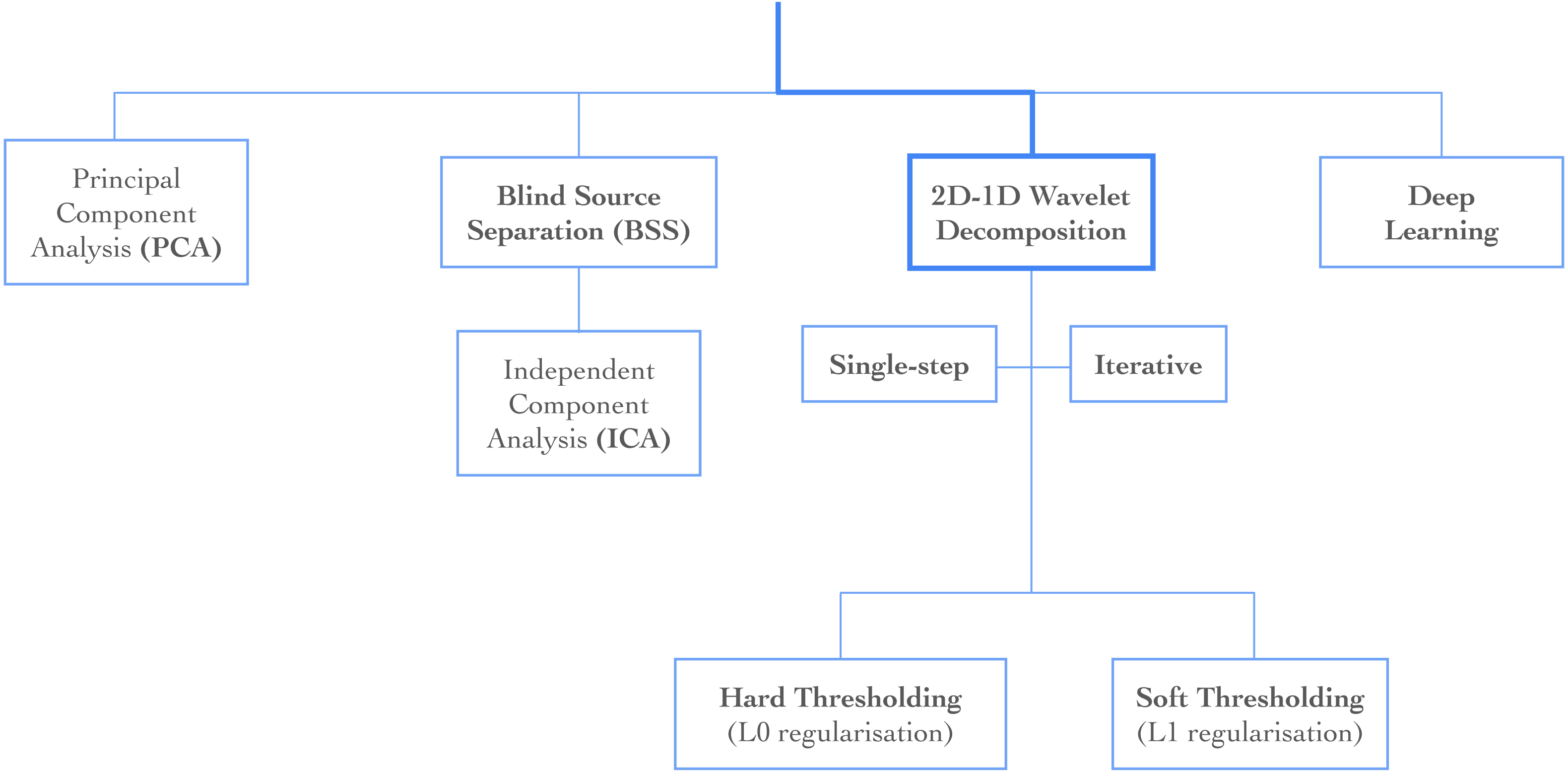
ICA denoised

I N I T I A L

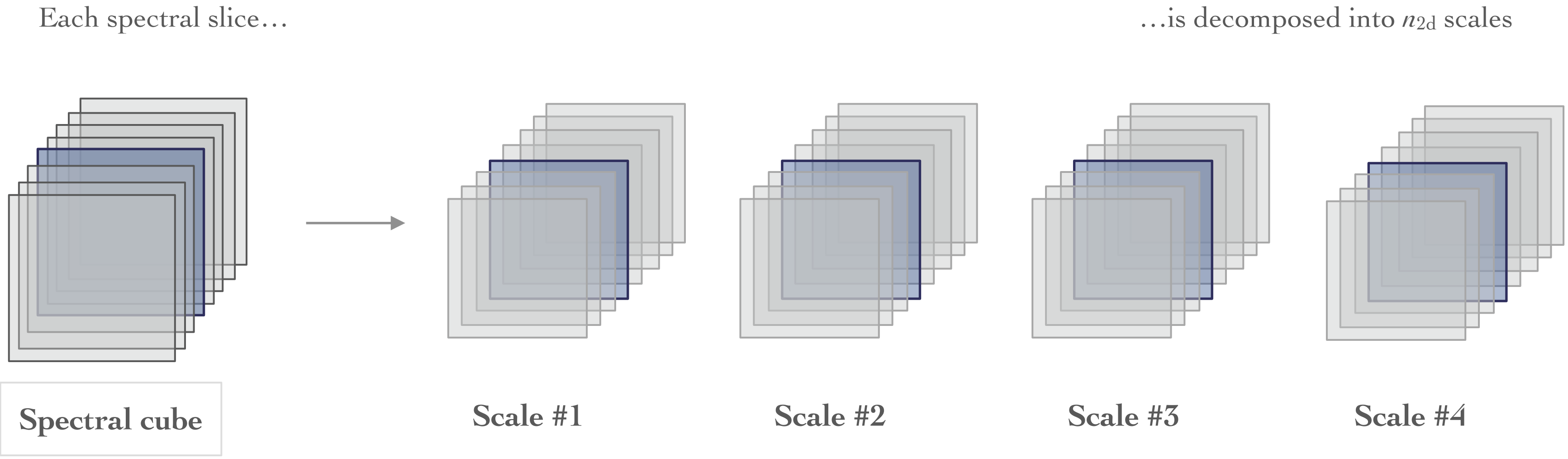
Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Noise std. deviation: 3.012e-02
Emission RMSE: 1.212e-02
Average flux: 5.420-02

Denoising Methods

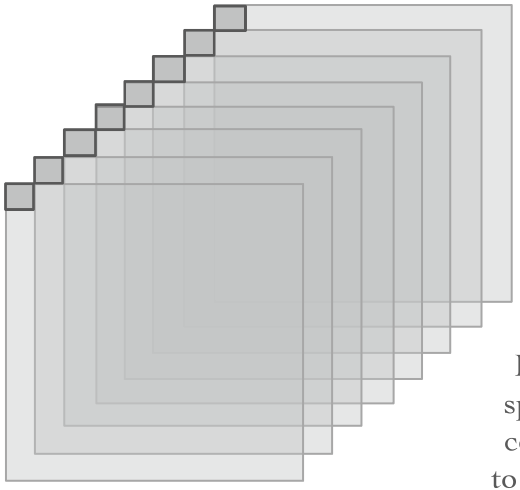


2D-1D wavelet decomposition



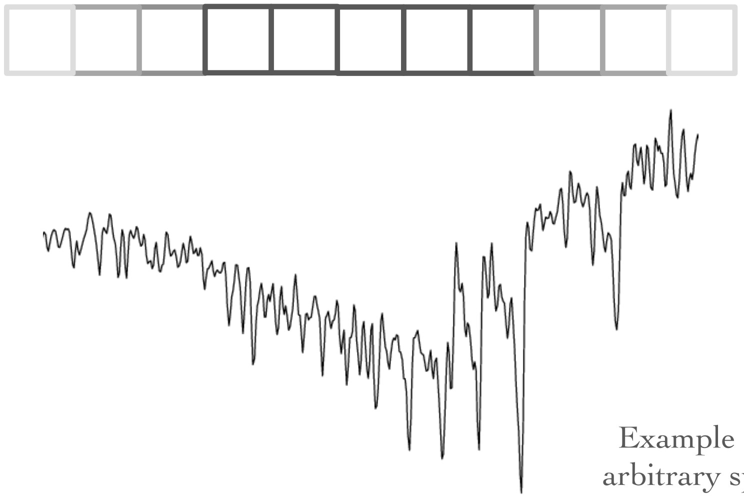
2D decomposition: Starlet transform on each spectral slice : undecimated and non orthogonal (dimensions of each slice are preserved)

2D-1D wavelet decomposition



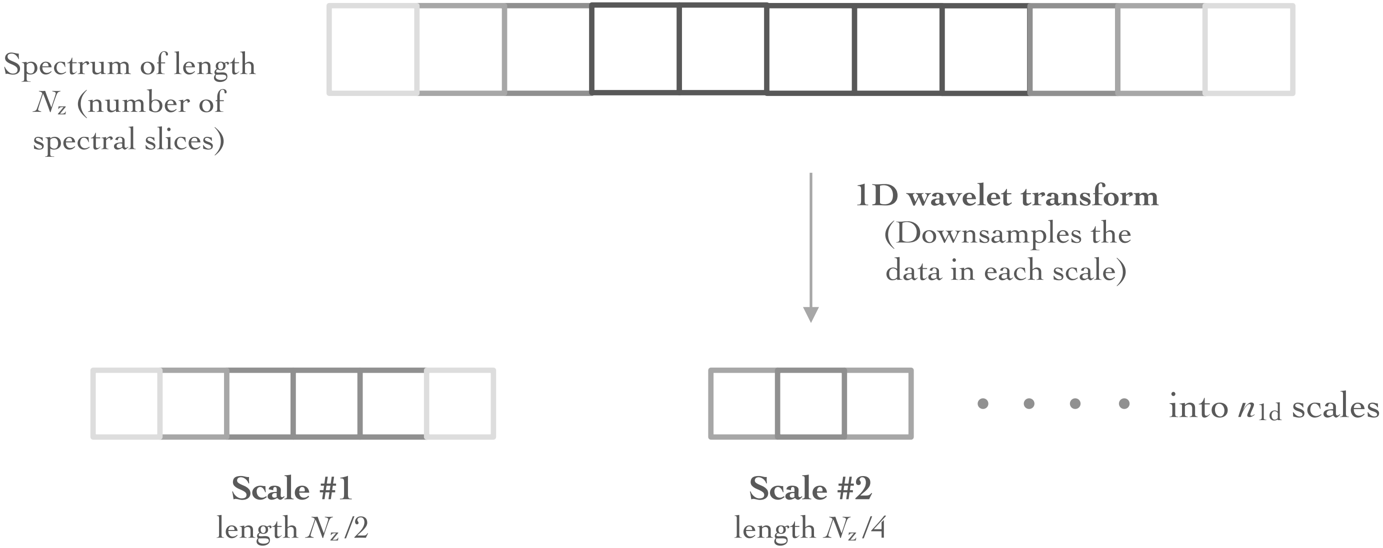
Each *(x,y)* spatial point corresponds to a spectrum.

Cube at one 2D scale



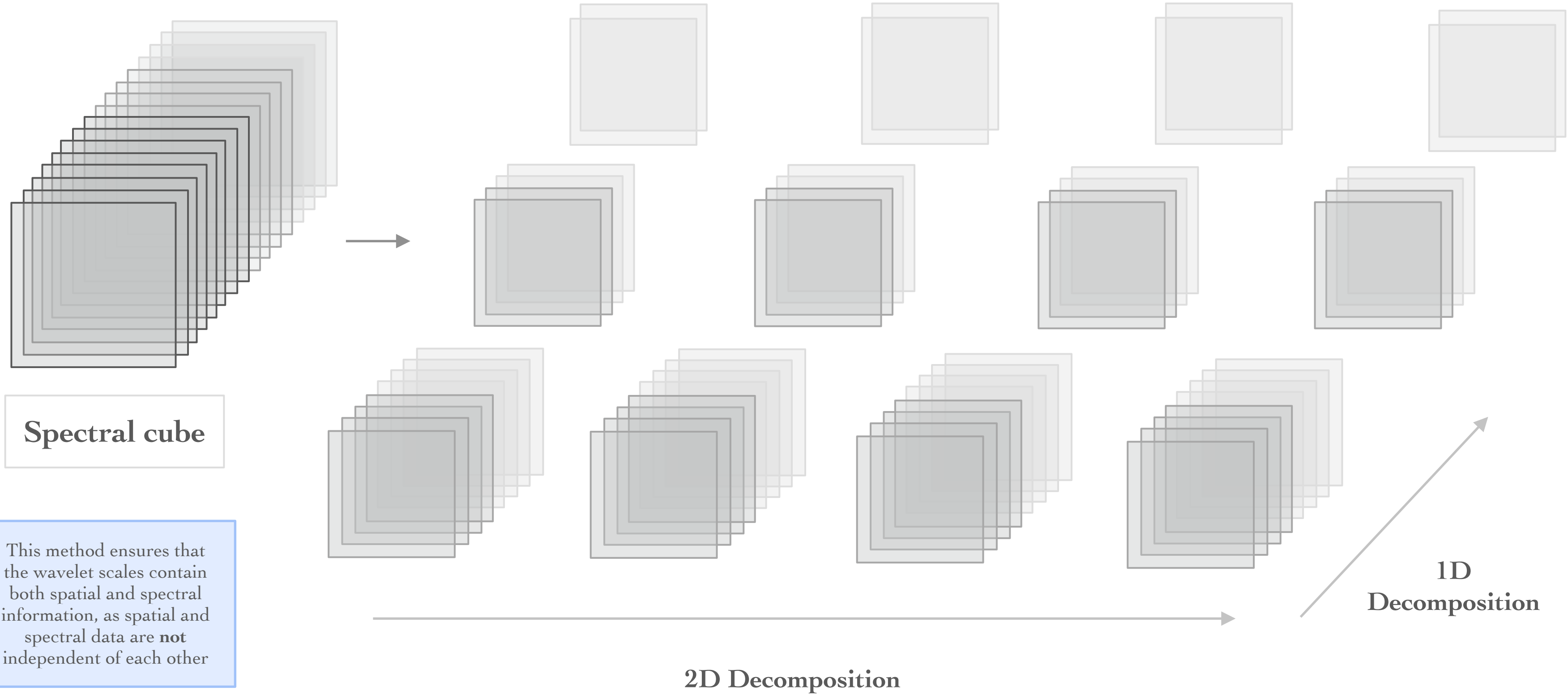
Example of an arbitrary spectra corresponding to a 1D array for one spaxel

2D-1D wavelet decomposition



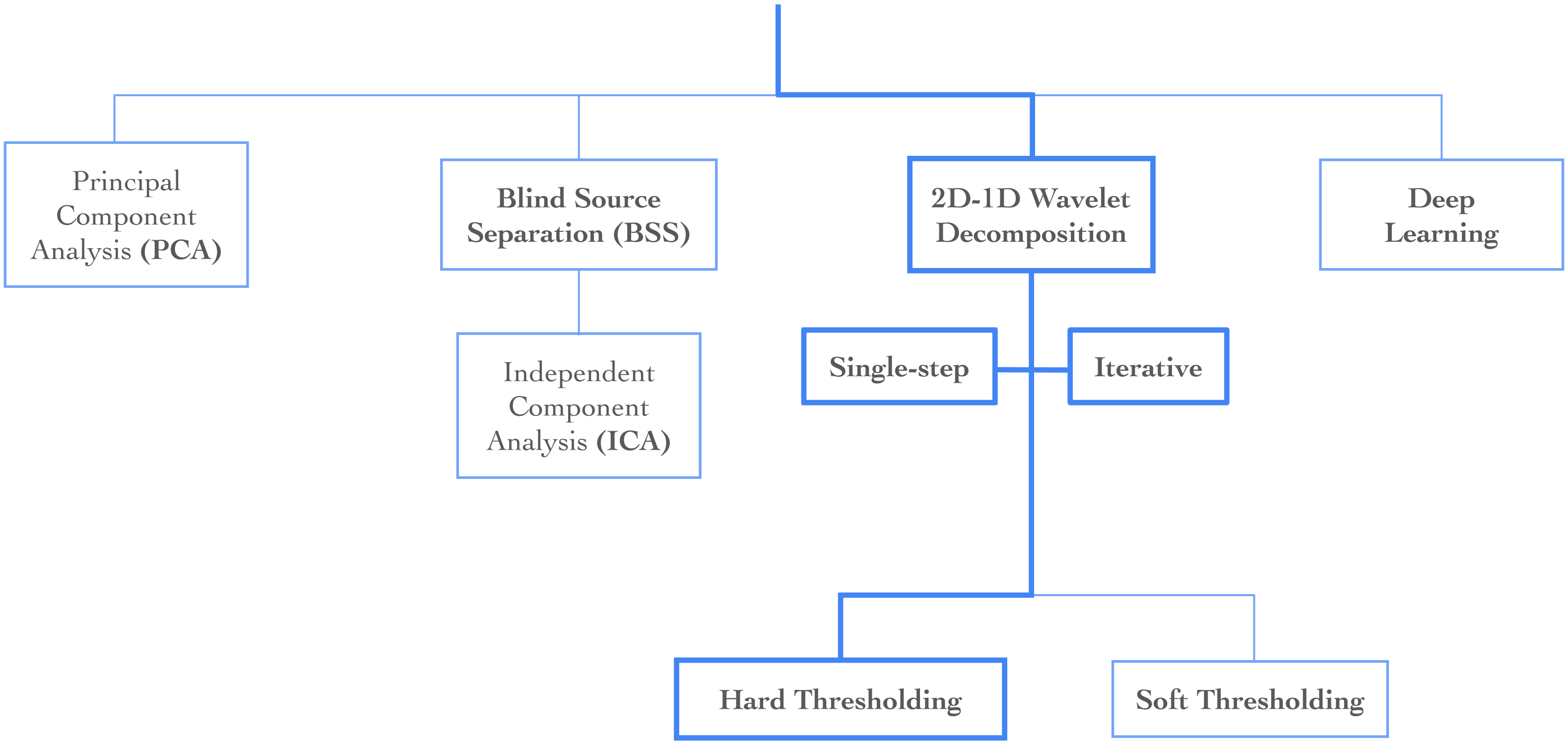
1D decomposition: The spectra associated with each spaxel is decomposed with a 1D wavelet transform

2D-1D wavelet decomposition

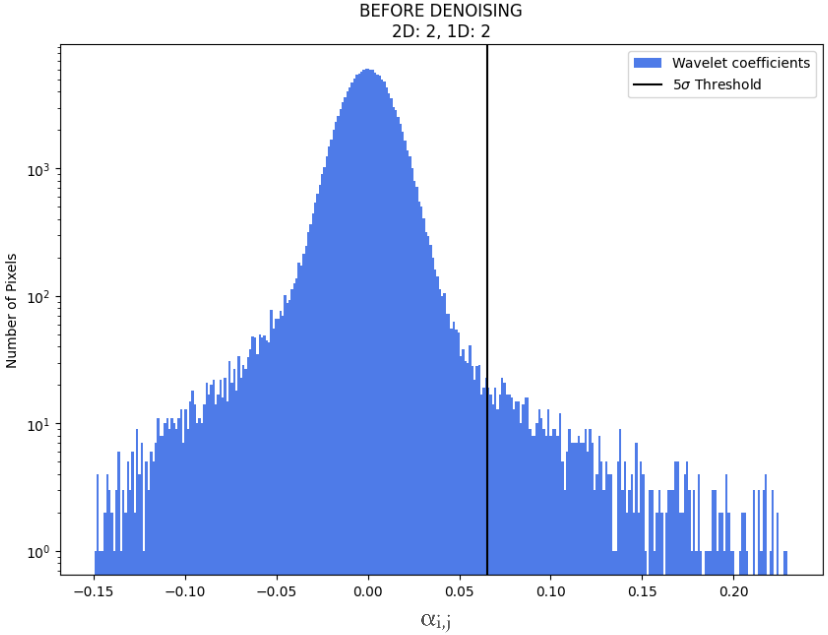


This method ensures that the wavelet scales contain both spatial and spectral information, as spatial and spectral data are **not** independent of each other

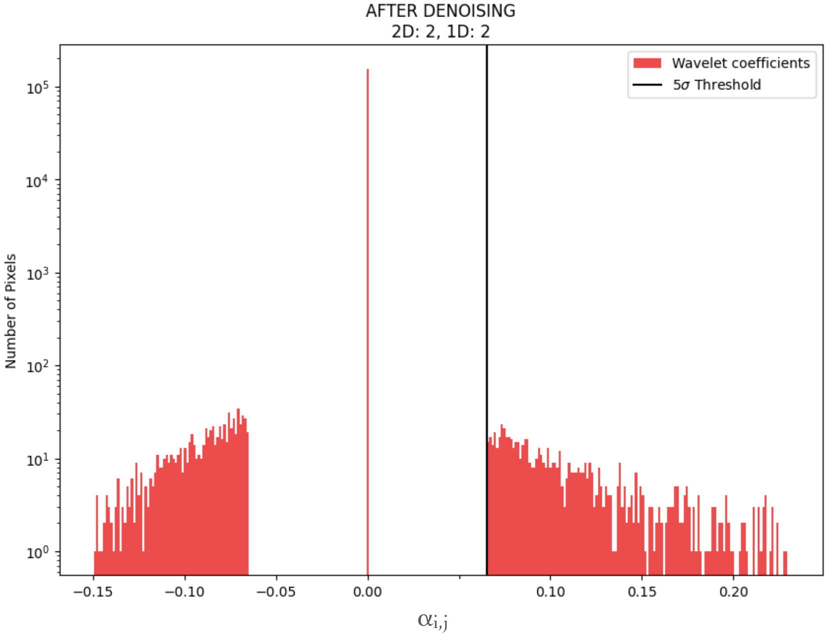
Denoising Methods



Single-step Hard Thresholding



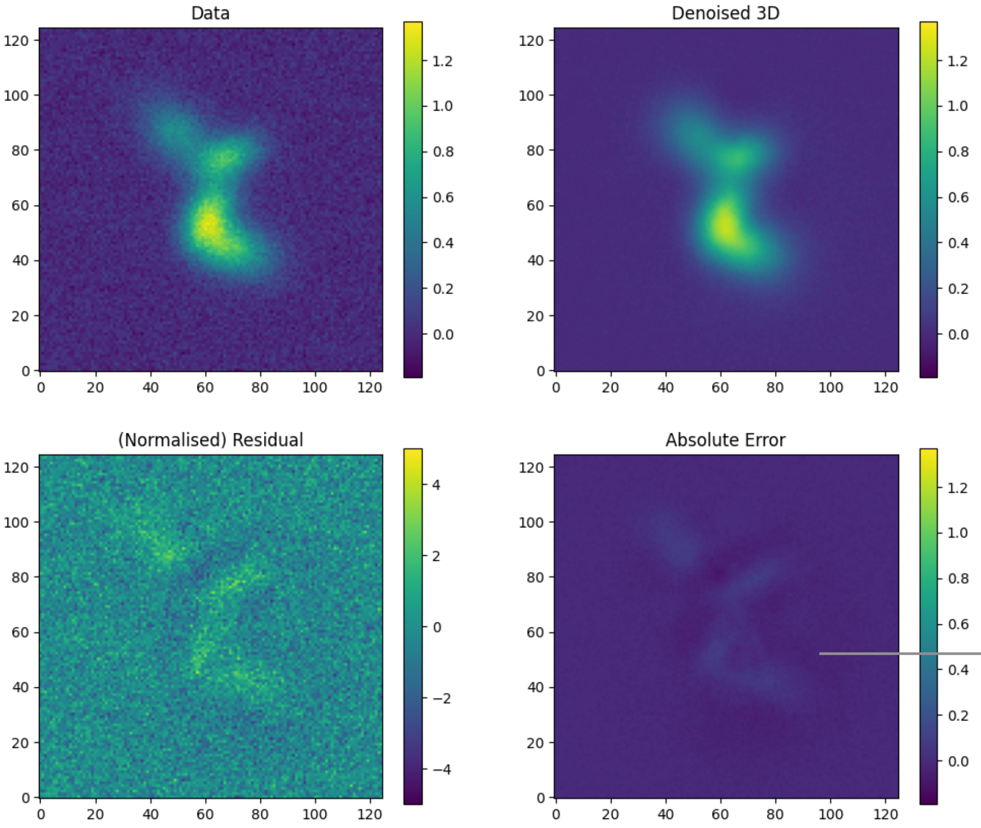
Distribution of wavelet coefficients in **one (2D, 1D) scale**



After thresholding at 5σ using

$$\mathcal{T}_{\text{hard}}(\alpha, \lambda) = \begin{cases} \alpha, & \text{if } |\alpha| \geq \lambda \\ 0, & \text{if } |\alpha| < \lambda \end{cases}$$

Single-step Hard Thresholding



INITIAL

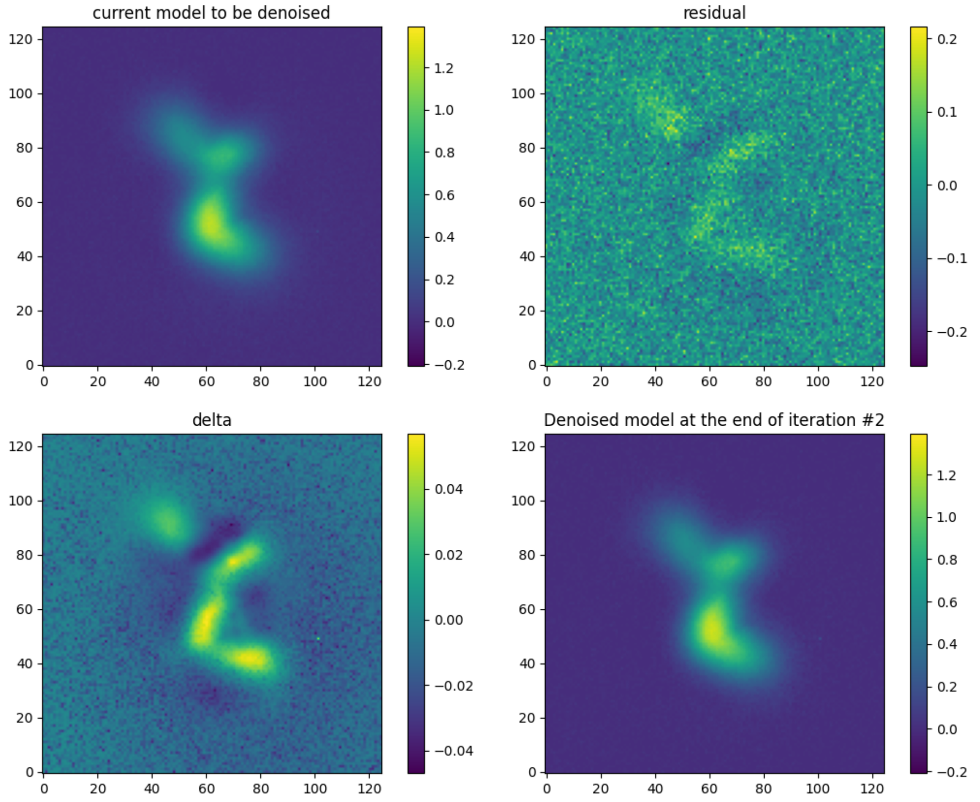
Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Noise std. deviation: 2.000e-02
Emission RMSE: 1.253e-02
Average flux: 5.368e-02

Un-obtained signal in the residual!

Signal extraction can be further refined!

Iterative Hard Thresholding



In subsequent iterations, **wavelet thresholding is applied on the residual** of the current iteration's input with the noisy data

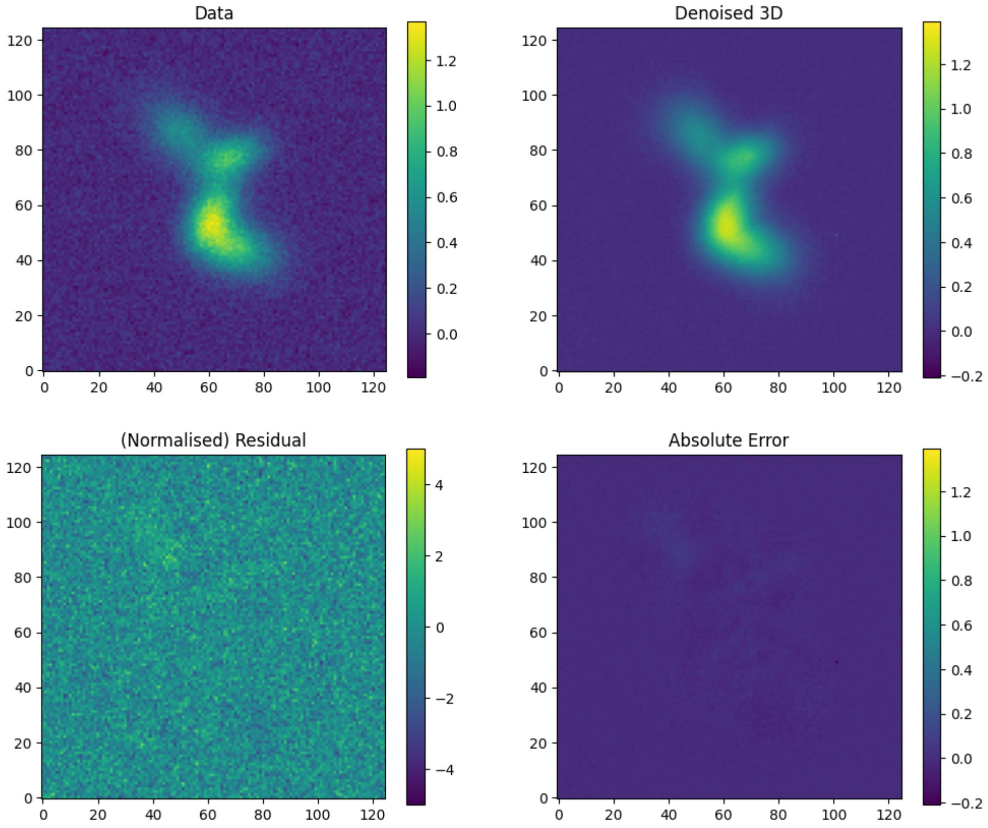
The **previously undetected signal from the residual is added onto the initial denoised data**, thus improving flux conservation and SNR improvement

INITIAL

Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Noise std. deviation: 1.666e-02
Emission RMSE: 8.276e-03
Average flux: 5.368e-02

Iterative Hard Thresholding



After **3 iterations**, final result:

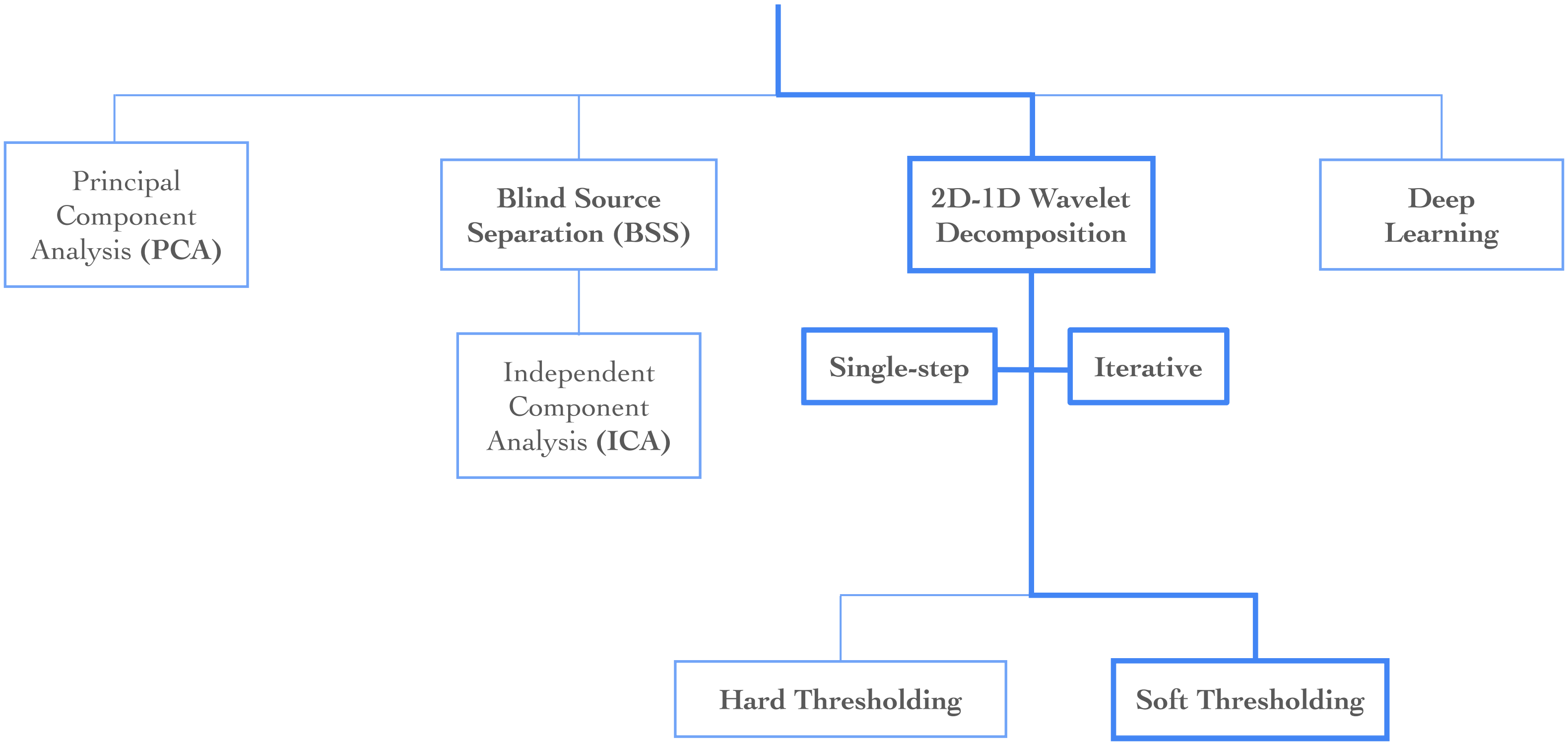
INITIAL

Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

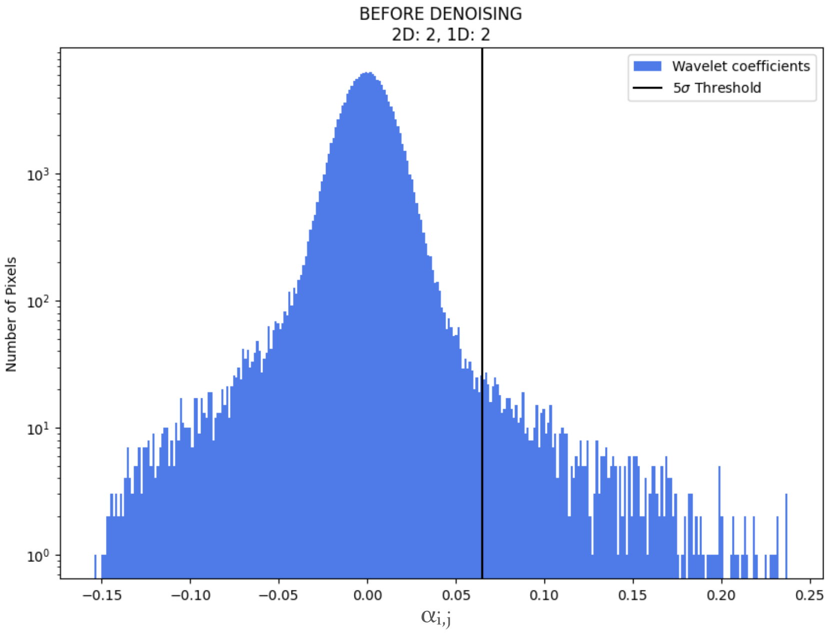
Noise std. deviation: 1.624e-02
Emission RMSE: 7.827e-03
Average flux: 5.411e-02

Flux is conserved well, the noise level and emission RMSE are noticeably lower than PCA/ICA or single-step wavelet case

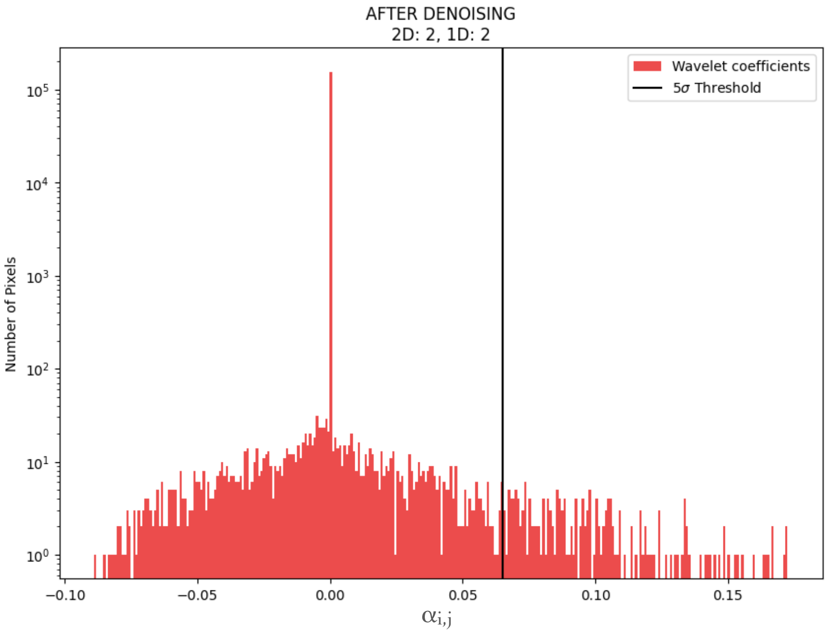
Denoising Methods



Single-step Hard Thresholding



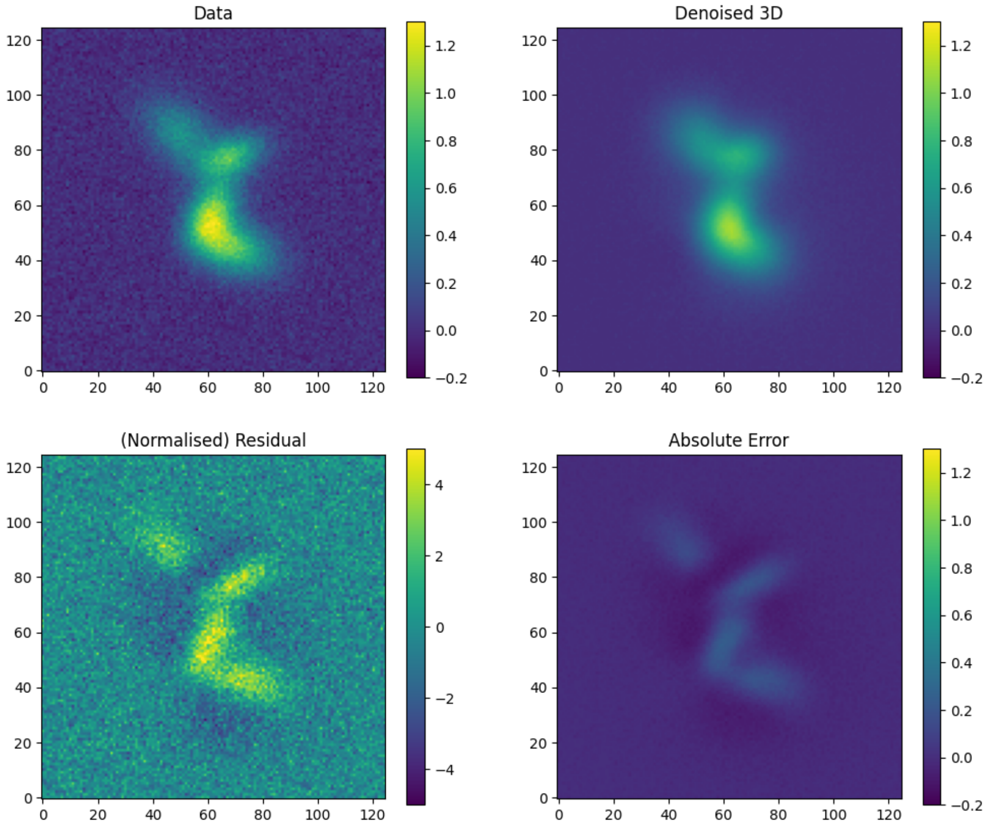
Distribution of wavelet coefficients in **one (2D, 1D) scale**



After thresholding at 5σ using

$$\mathcal{T}_{\text{soft}}(\alpha, \lambda) = \text{sign}(\alpha) \cdot \max(|\alpha| - \lambda, 0)$$

Single-step Soft Thresholding



INITIAL

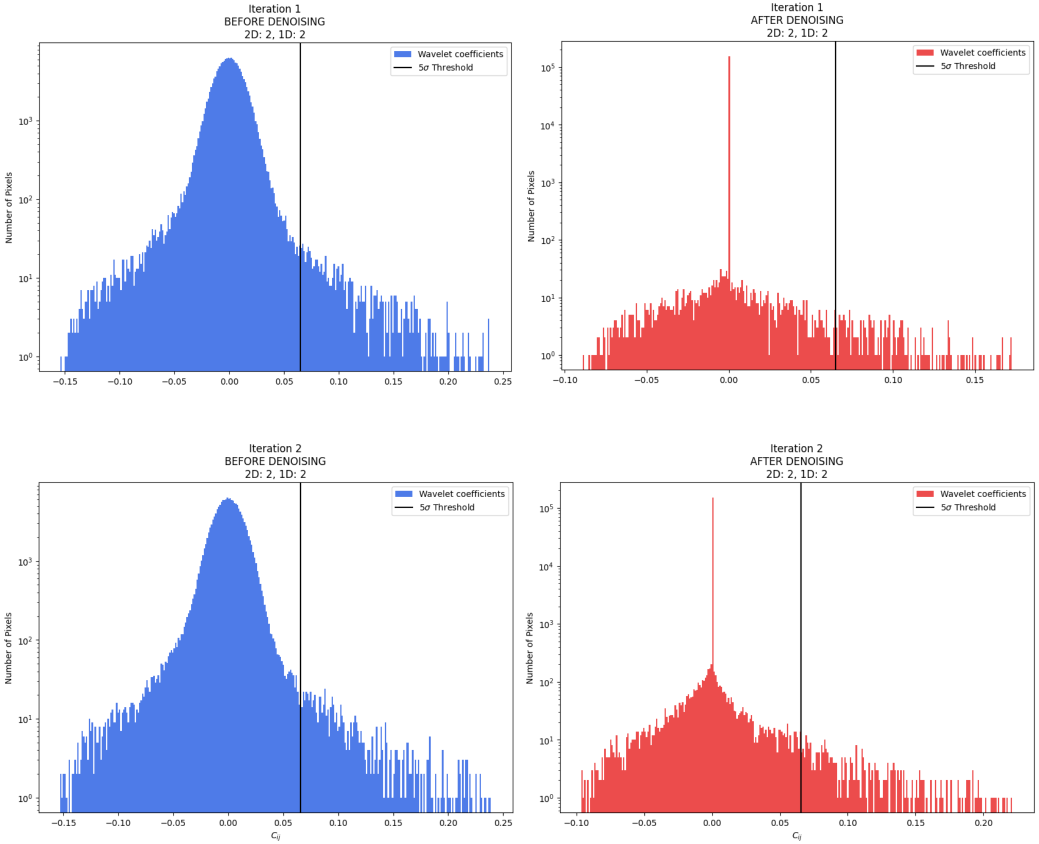
Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Noise std. deviation: 2.943e-02
Emission RMSE: 3.042e-02
Average flux: 5.157e-02

Flux is not conserved well and significant signal residuals due to **bias induced by shrinking all wavelet coefficients** smoothly

Attempting to iteratively refine

Iterative Soft Thresholding



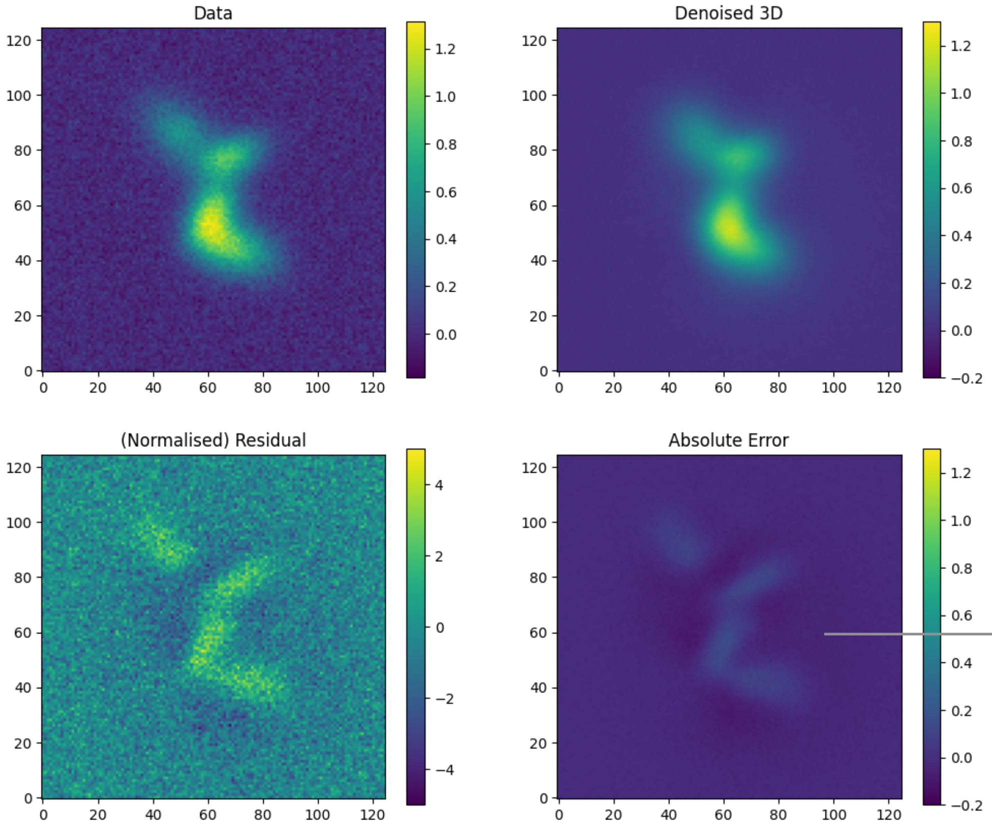
In subsequent iterations, **weights are calculated** as a function of the closeness of the coefficient magnitudes of the previous iteration to the threshold value

$$w_{ij} = \frac{1}{1 + \exp\left(10|\alpha_{ij} - \lambda|^2\right)}$$

Re-weighted thresholding is applied in the next iteration after a gradient step, which pushes the data closer to the input

Coefficients that are closer to the threshold are thresholded more aggressively (higher w) and shrunken more towards 0 and the remaining coefficients are shrunken less

Iterative Soft Thresholding



INITIAL

Noise std. deviation: 5.088e-02
Emission RMSE: 1.854e-02
Average flux: 5.418e-02

Noise std. deviation: 2.943e-02
Emission RMSE: 3.042e-02
Average flux: 5.157e-02

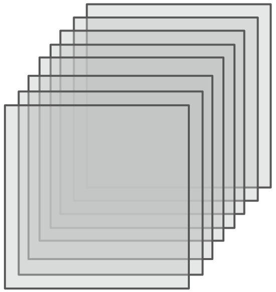
Noise std. deviation: 2.796e-02
Emission RMSE: 2.330e-02
Average flux: 5.252e-02

Flux is still not conserved very well as there are significant residuals and RMSE is still large, however the reweighing method slightly improves the reconstruction

Further improvements are needed

Datasets

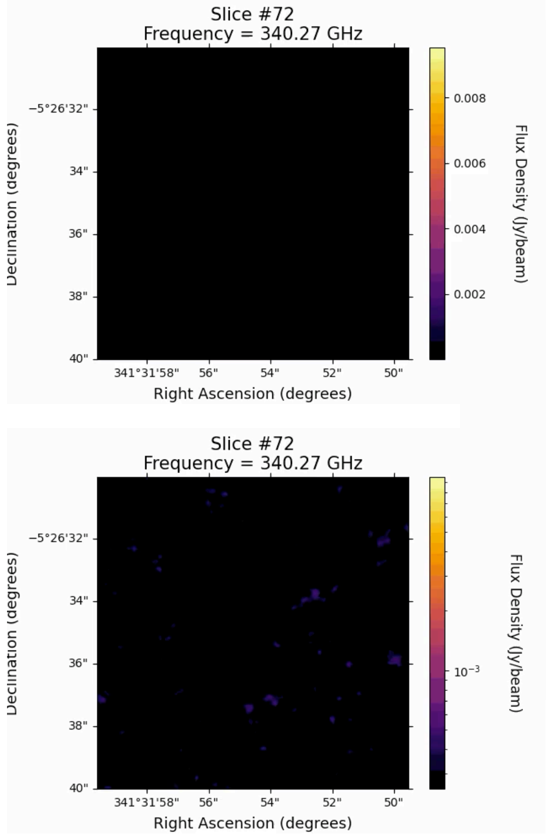
- Toy models of rotating galaxies
- **Mock IFU data from cosmological simulations**
- Observational data (ALMA, JWST, MUSE)



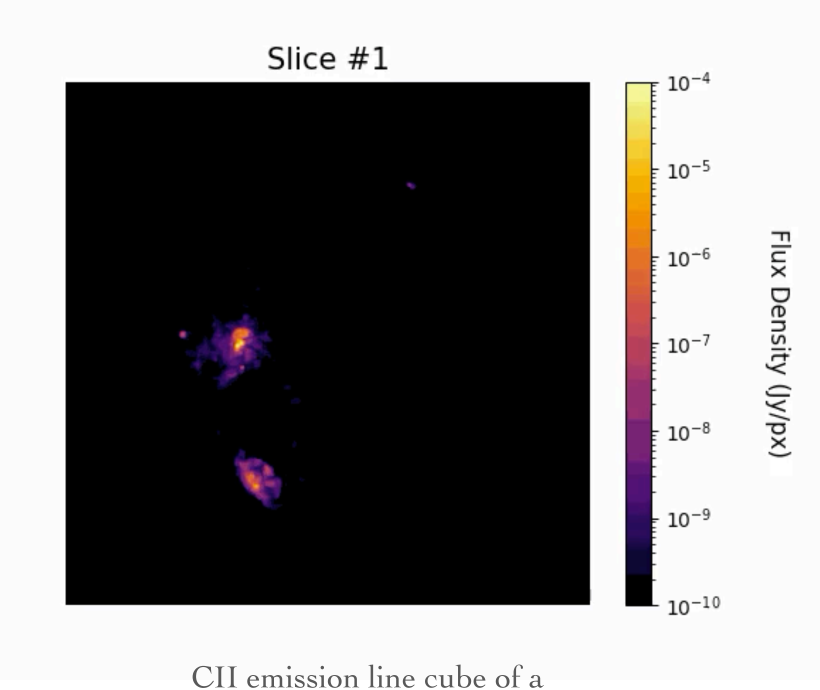
Spectral cubes

multiple spectral observation of the same spatial area, where each *(x,y)* spatial point corresponds to a spectrum

Pre-processing mock IFU cubes from FIRE



ALMA cubes of real W2246

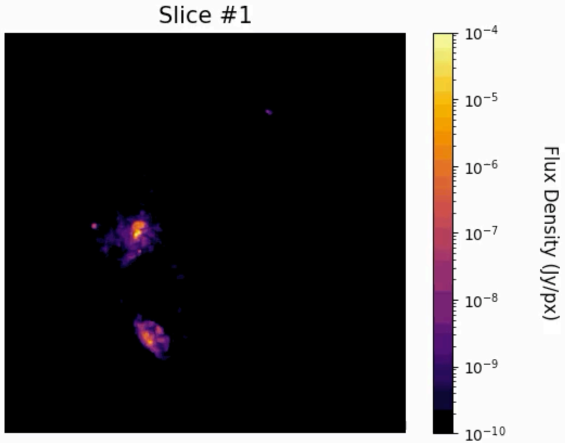


CII emission line cube of a highly-resolved multiple merger system from FIRE simulations

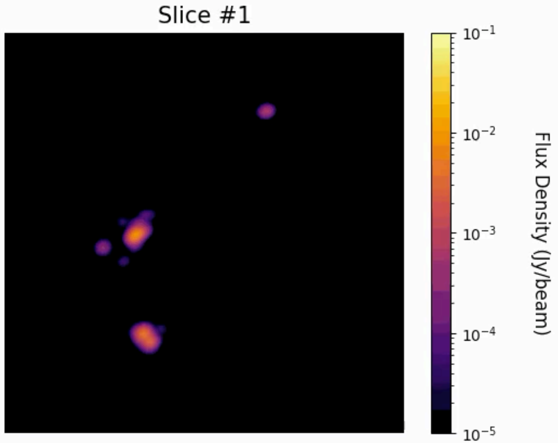
W2246-like system

A multiple-merger system was discovered in the FIRE simulations - simulated at redshift 4.5, central galaxy with multiple satellites and visible bridges and streams of gas

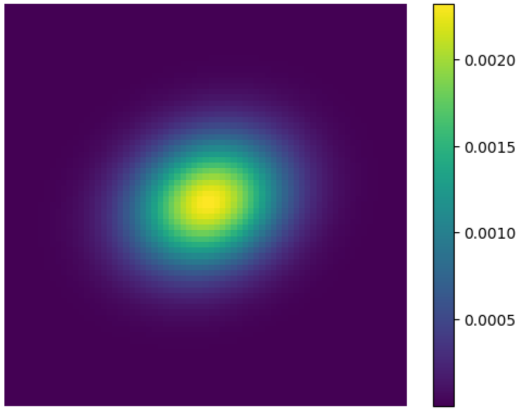
Aim: To pre-process the raw, highly resolved mock IFU cube and convolve it with an appropriate beam to make the cube as close as possible to real observations (ALMA)



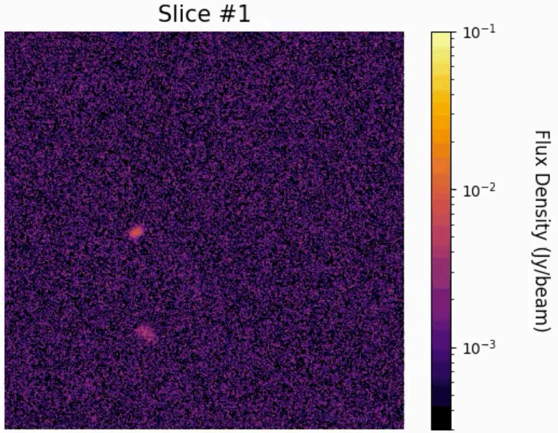
Resolved cube



Convolved cube



ALMA-like PSF



Noisy convolved cube

The resolved cube is **convolved with an ALMA-like PSF** (approximated from observational cube of W2246), and **Gaussian noise is overlaid** on the cube with manually chosen SNR

Same trends observed as in the results with the toy data

Denoising Methods

- Principal Component Analysis (PCA)
- Blind Source Separation (BSS)
  - Independent Component Analysis (ICA)
- 2D-1D Wavelet Decomposition
  - Single-step
  - Iterative
  - Hard Thresholding
  - Soft Thresholding
- Deep Learning

Aim: To develop a **U-net like architecture** (maybe tensor based models) and train the network to de-noise in a supervised way

First with **toy data**, and subsequently with **mock IFUs** with the goal of **applying to real data** the mock cubes are based on

Future/Ongoing Work

Refining the re-weighting step of iterative soft thresholding

Deep learning denoising application on toy data and mock IFU cubes

Application of these methods to different noise characteristics (JWST, MUSE)

Generating statistics for the methods for different resolutions and initial SNRs

Testing different method performances on real W2246 observations from ALMA, MUSE and JWST

Next project on high-redshift galaxy property estimation using spectral data from(de-noised) mock and real cubes

Thank you for your time!