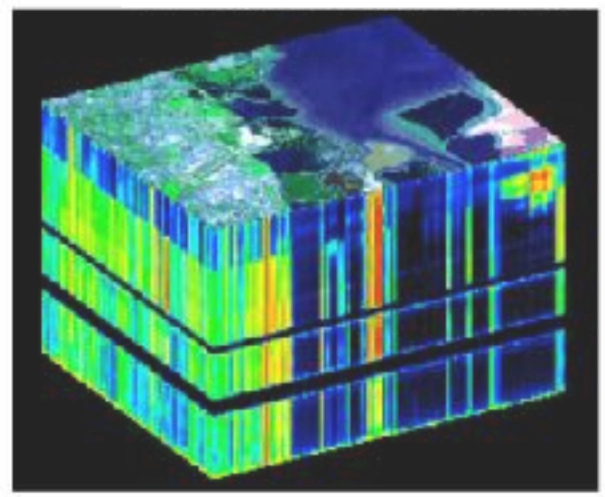
Sparse matrix factorization method and its applications in astrophysics

Jérôme Bobin

joint work with I. El Hamzaoui, A. Picquenot, F.Acero and C.Kervazo

Laboratoire CosmoStat - CEA/Irfu, France

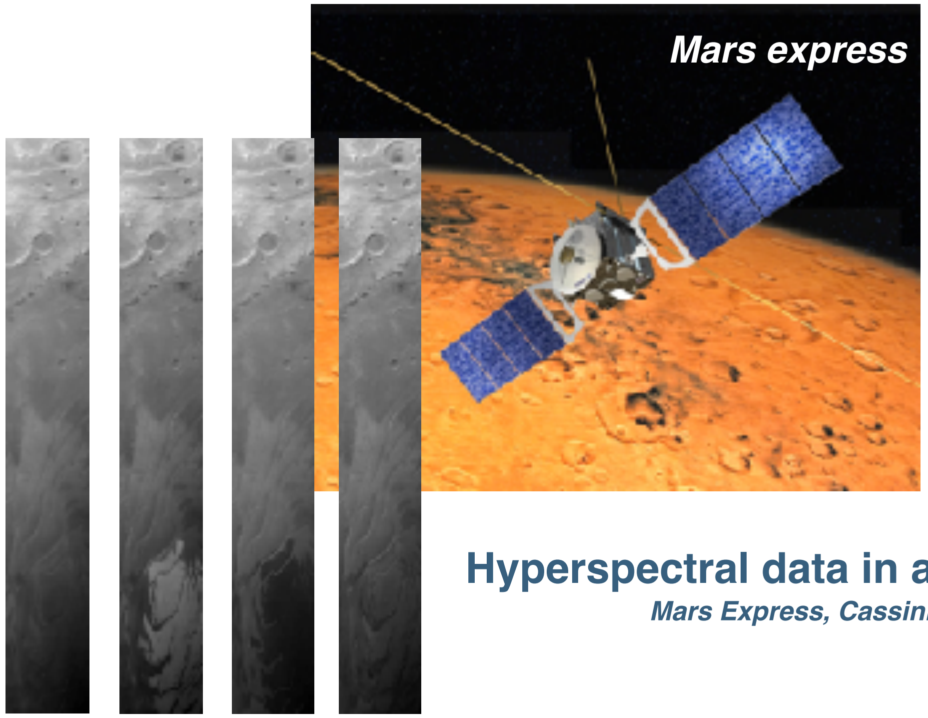
Analysing multispectral data



Courtesy of M. Lennon

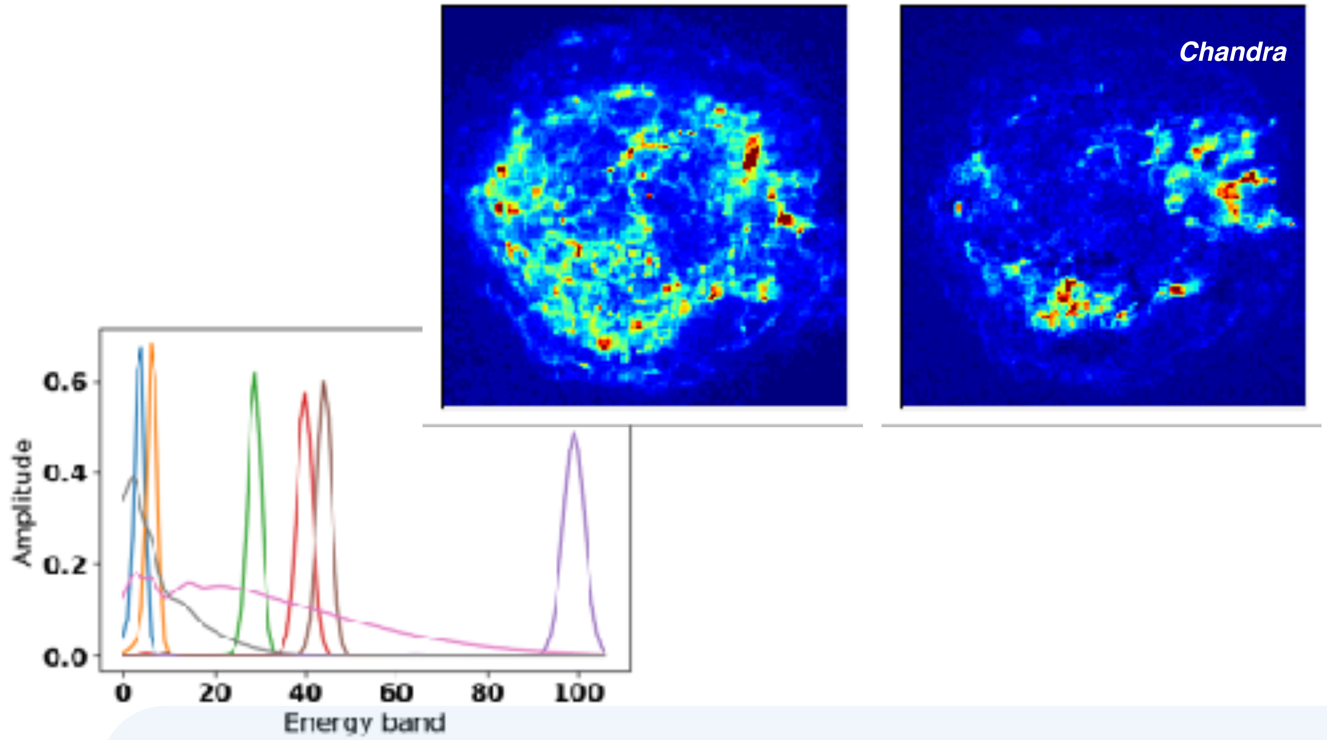
Hyperspectral data

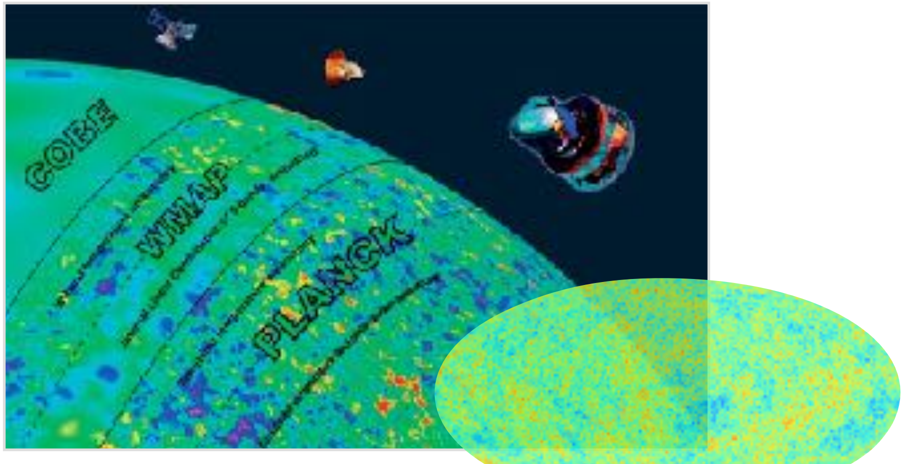
remote sensing, aerial data, etc.



Hyperspectral data in astrophysics

Mars Express, Cassini, etc.





Multispectral data in astrophysics

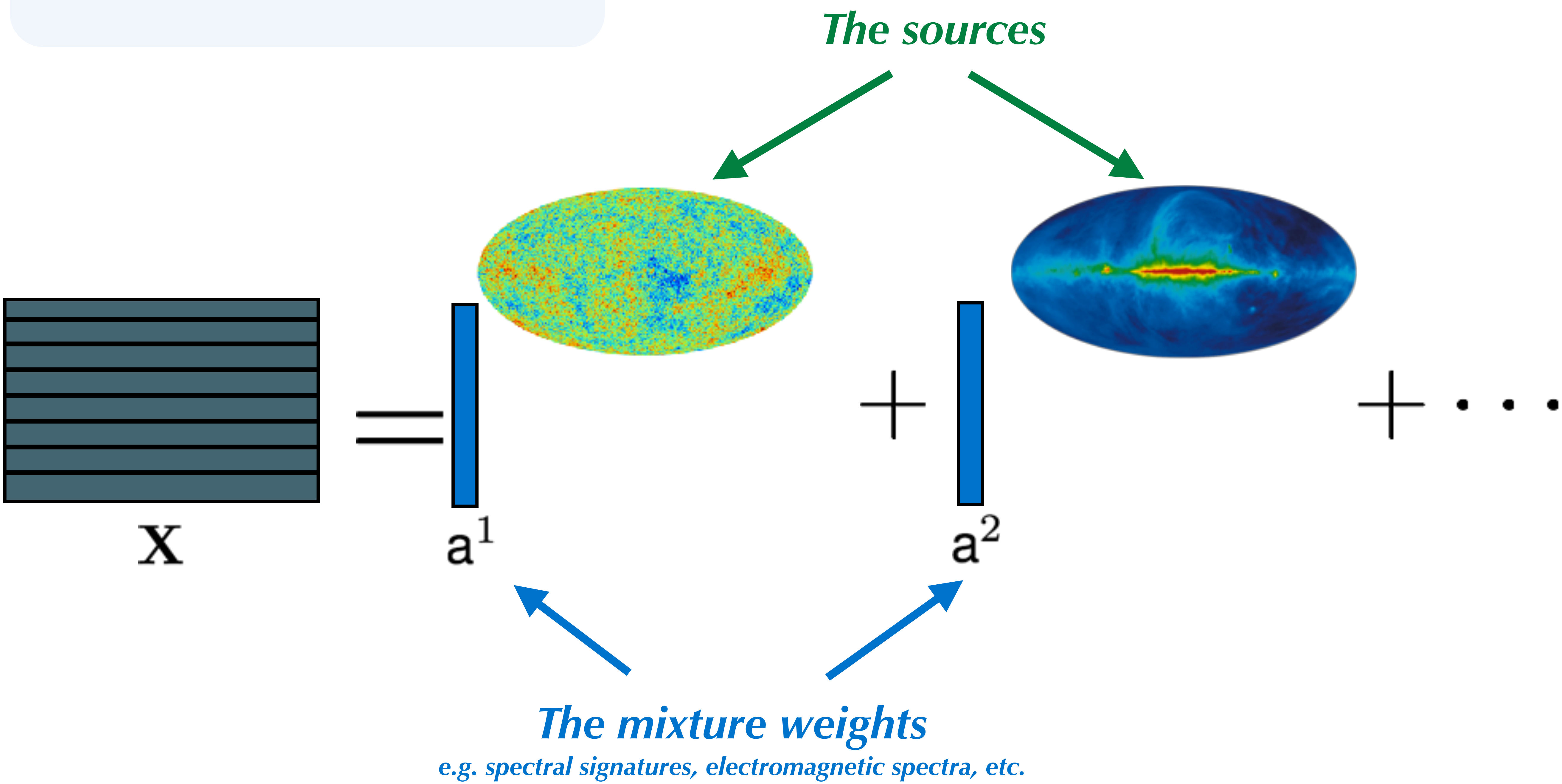
Planck, Fermi, radio-interferometry (Lofar/SKA/…), etc.

Different scientific fields but …

common problems: mixtures of elementary signals or sources

The underlying mixture model

The linear mixture model



Unsupervised matrix factorisation

The source matrix

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

The mixing matrix

Noise

Blind Source Separation:

Estimation both A and S from X only

This is an ill-posed matrix factorization problem

Non-negative Matrix Factorization, Dictionary Learning, …

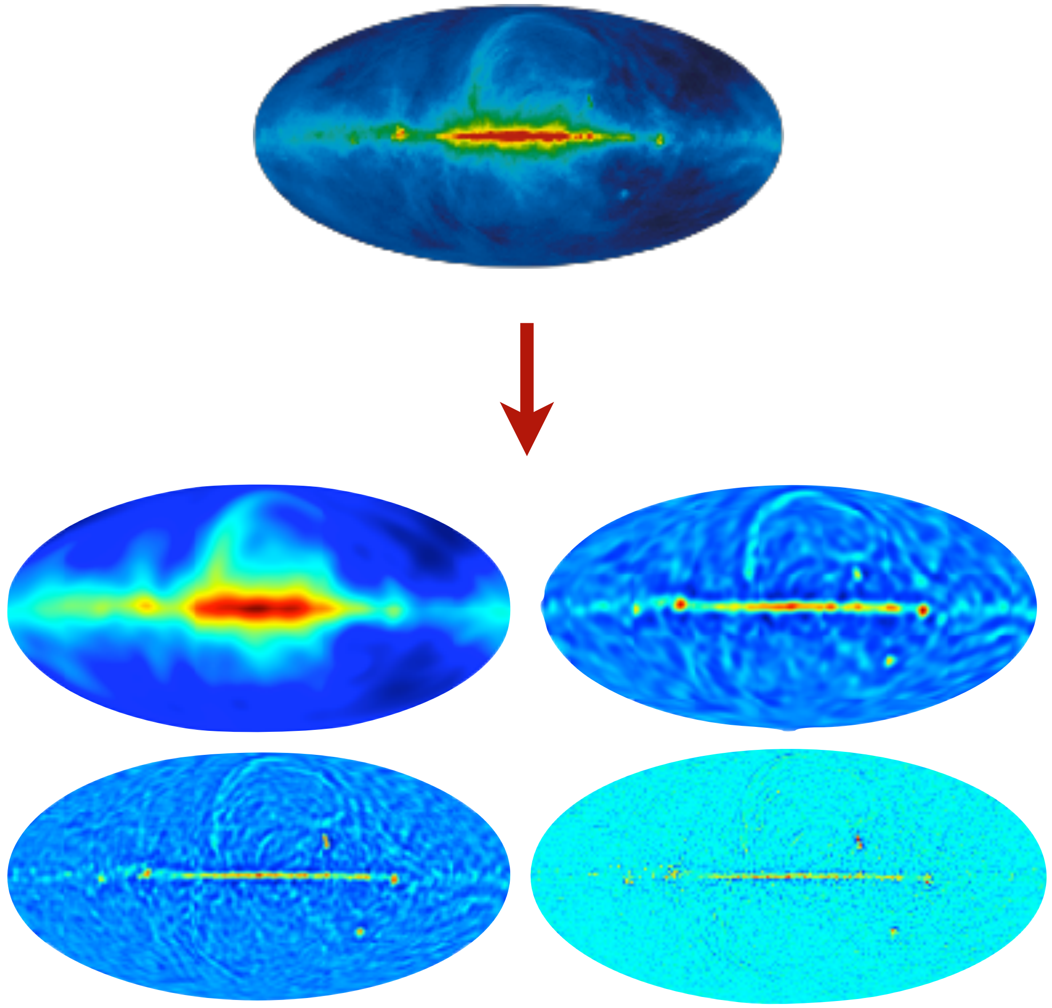
Sparse signal modeling

Prior information on S and/or A

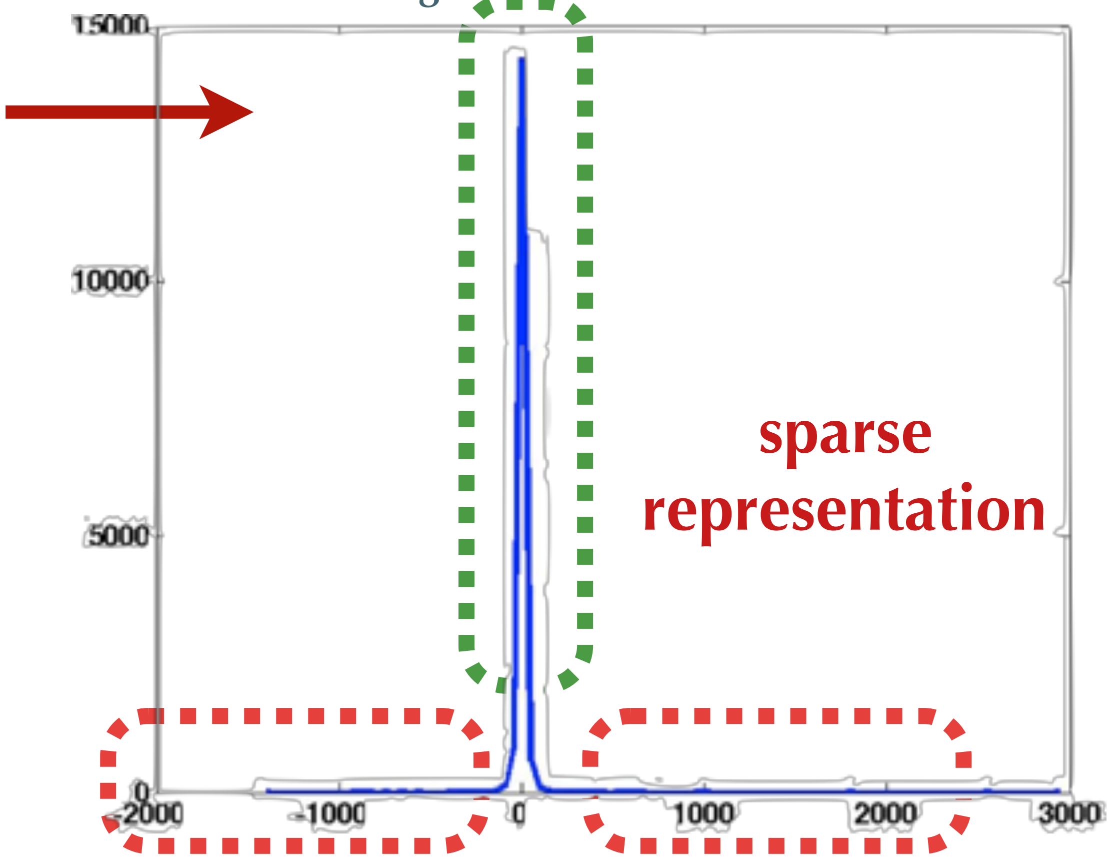
Statistical independence, non-negativity, etc.

Sparse signal modeling

Zibulevsky01, Cichocki06, Bobin07



Wavelet transform for spherical data



Sparse Matrix Factorisation

Gist: looking for the sparsest sources

Regularization params., weight matrix, etc.

$$\min_{\mathbf{A},\mathbf{S}} \|\Lambda \odot \mathbf{SW}\|_p + \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_F^2$$

Sparse regularization — *Data fidelity term*

Generalized Morphological Component Analysis (GMCA):

- *S-BSS with redundant sparse representations*
- *Iterative soft/hard thresholding algorithm*
- *Thresholding strategy, robustness to Gaussian noise/local stationary points*
- *No parameters to tune*

Bobin, Starck, Fadili, and Moudden, Sparsity, Morphological Diversity and Blind Source Separation, IEEE Trans. on Image Processing, Vol 16, No 11, pp 2662 - 2674, 2007.

Bobin, Starck, Fadili, and Moudden, Blind Source Separation: The Sparsity Revolution, Advances in Imaging and Electron Physics , Vol 152, pp 221 -- 306, 2008.

Beyond standard models

- The **global** linear mixture does not hold true

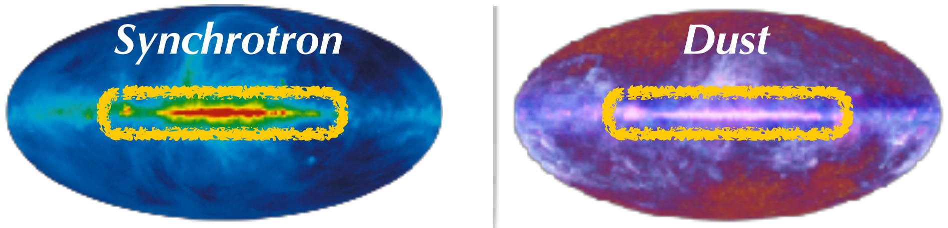
Local-GMCA: local/multiscale mixture model, handles spectral variabilities

Bobin J., Sureau F., Starck, CMB reconstruction from the WMAP and Planck PR2 data, A&A, 2016

- Galactic components are **partially correlated**

AMCA: robustness w/r to partial correlations

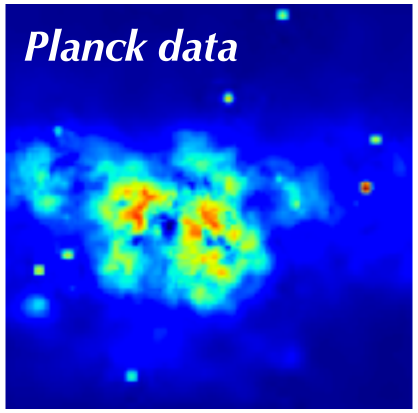
Bobin J., et al., IEEE Tr. on signal processing, 2015



- Many point sources as **outliers**

rGMCA: robustness w/r to outliers, based on morphological diversity

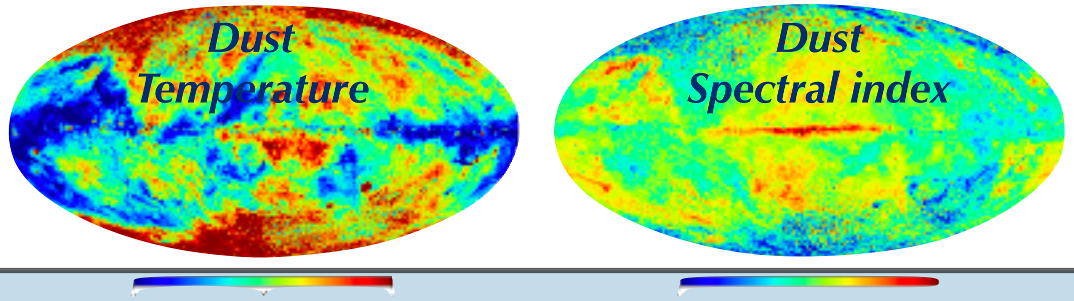
Chenot, et al., SIAM Imaging Sciences, 2018



- Accounting for sparse **parametric non-linear physical models**

premise: include astrophysical models for a more precise estimation of the galactic sources

Irfan, et al., MNRAS, 2018

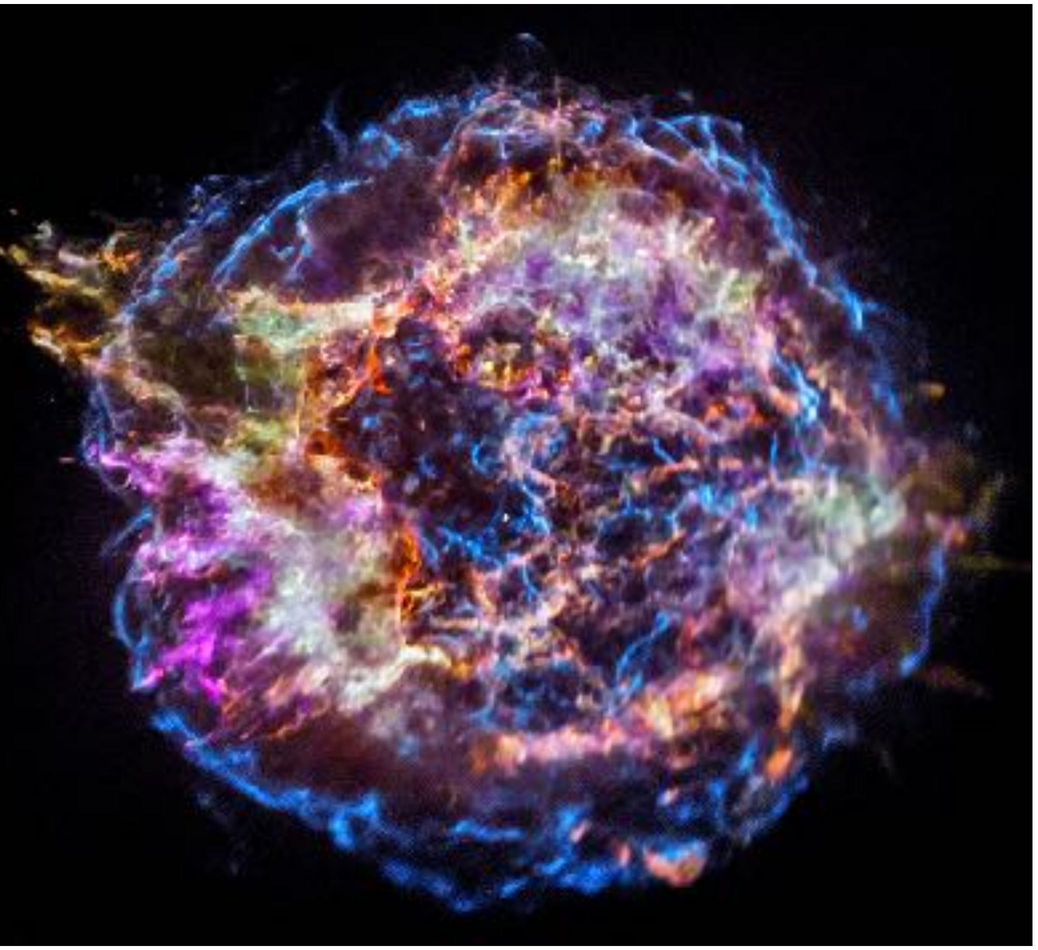


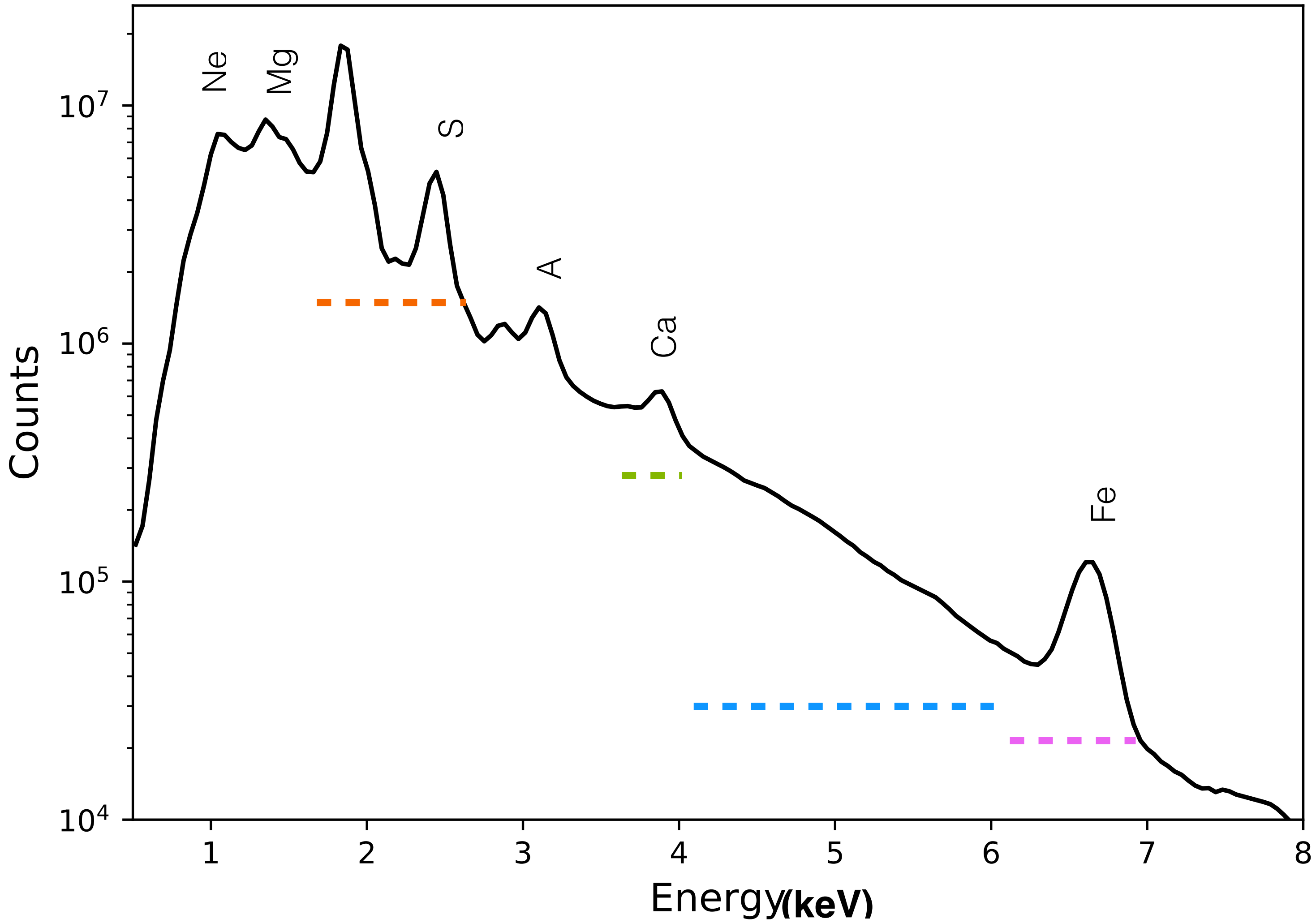
Analyzing X-ray data in high-energy astrophysics

Supernova remnants as seen in X-rays

X-ray observations of SNR

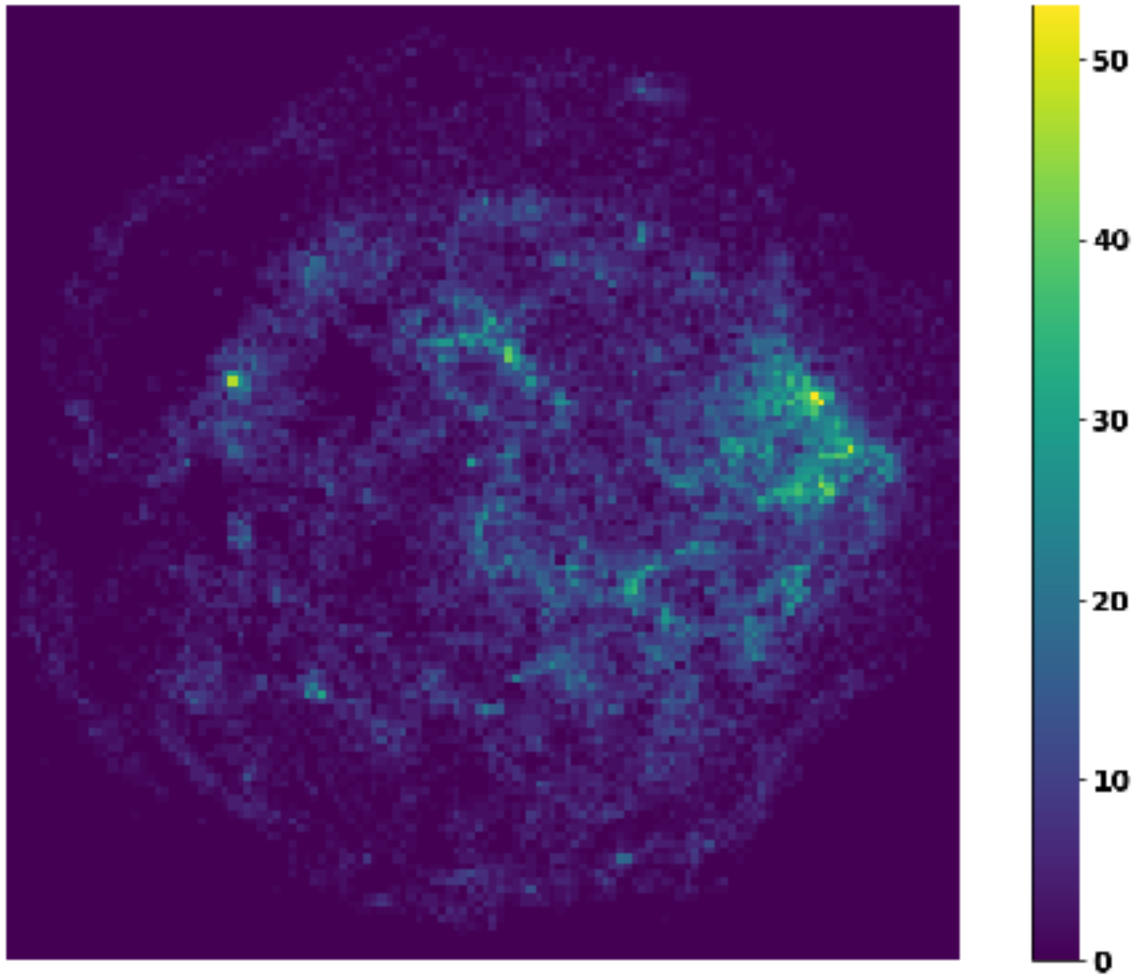
CasA with Chandra
1 Ms observation
~1 billions counts !!





- **Ejecta thermal emission gives insight on :**
  - **Individual elements distribution**
  - **Morphology, asymmetries**
  - **Velocities**

… but the data follow a Poisson distribution



- Each image is built from a sequence of events (photons)

- The measurements follow a Poisson distribution

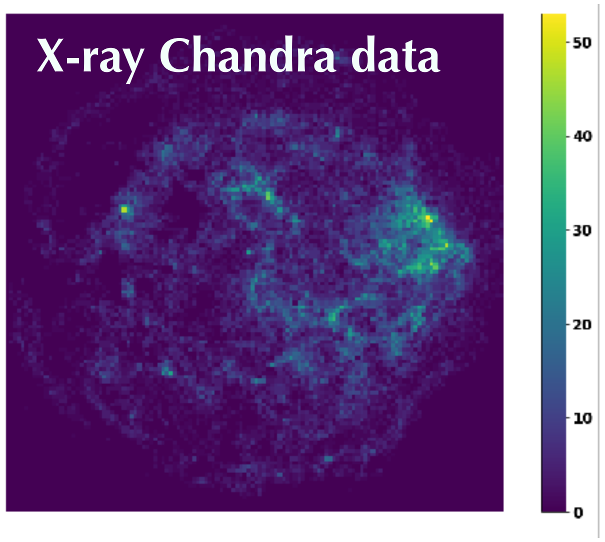
$$\mathcal{P}(\mathbf{X}_i[t] | [\mathbf{AS}]_i[t]) = \frac{e^{-[\mathbf{AS}]_i[t]} \, [\mathbf{AS}]_i[t]^{\mathbf{X}_i[t]}}{\mathbf{X}_i[t]!}$$

- The mixture model is valid only on average $[\mathbf{AS}]_i[t] = \mathbb{E}\left\{\mathbf{X}_i[t]\right\}$

- The noise is data-dependent and larger for features with larger amplitudes, which is likely to hamper sparse BSS methods:

$$\text{Var}\left\{\mathbf{X}_i[t]\right\} = [\mathbf{AS}]_i[t]$$

BSS: switching from Gaussian to Poisson statistics



Extending sparse BSS to account for the Poisson statistics of the measurements

$$\min_{\mathbf{A}\in\mathcal{C},\mathbf{S}\geq 0} \left\|\mathbf{\Lambda}\odot\mathbf{S}\mathbf{\Phi}^T\right\|_{\ell_1} + \mathcal{L}\left(\mathbf{X}|\mathbf{A},\mathbf{S}\right)$$

$$\mathcal{L}\left(\mathbf{X}|\mathbf{A},\mathbf{S}\right) = \mathbf{AS} - \mathbf{X}\odot\log(\mathbf{AS})$$

Poisson neg-loglikelihood

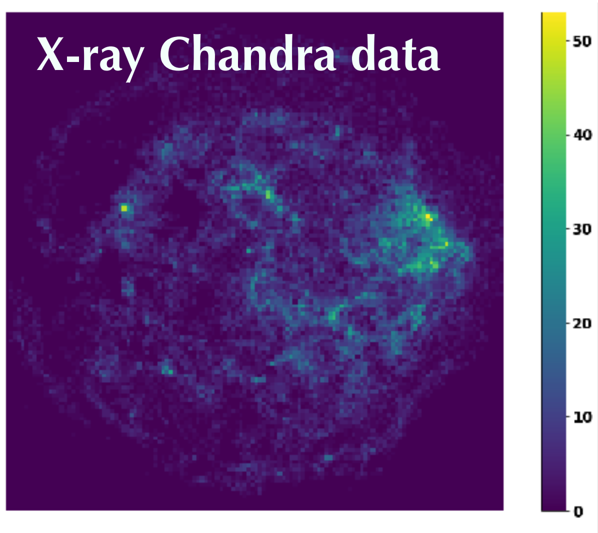
- Multi-convex problem with non-smooth data fidelity term

  standard methods (e.g. PALM, BCD) are not applicable

- The curvature of the data fidelity term soars at the vicinity of 0 $\propto 1 \oslash (\mathbf{AS}\odot\mathbf{AS})$

- How to choose the regularisation parameters $\mathbf{\Lambda}$?

Building a smooth approximation



$$\min_{\mathbf{A}\in\mathcal{C},\alpha\mathbf{\Phi}\geq 0} \|\mathbf{\Lambda}\odot\alpha\|_{\ell_1} + \mathcal{L}_\mu\left(\mathbf{X}|\mathbf{A},\alpha\mathbf{\Phi}\right)$$

Smooth approximation

- Makes use of Nesterov's smoothing to build a smooth approximation

$$\mathcal{L}_\mu(\mathbf{X}|\mathbf{Y}) = \inf_{\mathbf{U}} \langle \mathbf{Y},\mathbf{U}\rangle - \mathcal{L}^\star(\mathbf{X}|\mathbf{U}) - \mu g(\mathbf{U}),$$

Dual of $\mathcal{L}$

Strongly convex proximity term

- Smooth approximation with Lipschitz gradient

$$\nabla\mathcal{L}_\mu(\mathbf{X}|\mathbf{Y}) = \frac{1}{2\mu}(\mathbf{Y}+\mu)\odot\left[1-\sqrt{1-4\mu(\mathbf{Y}-\mathbf{X})\oslash(\mathbf{Y}+\mu)^2}\right]$$

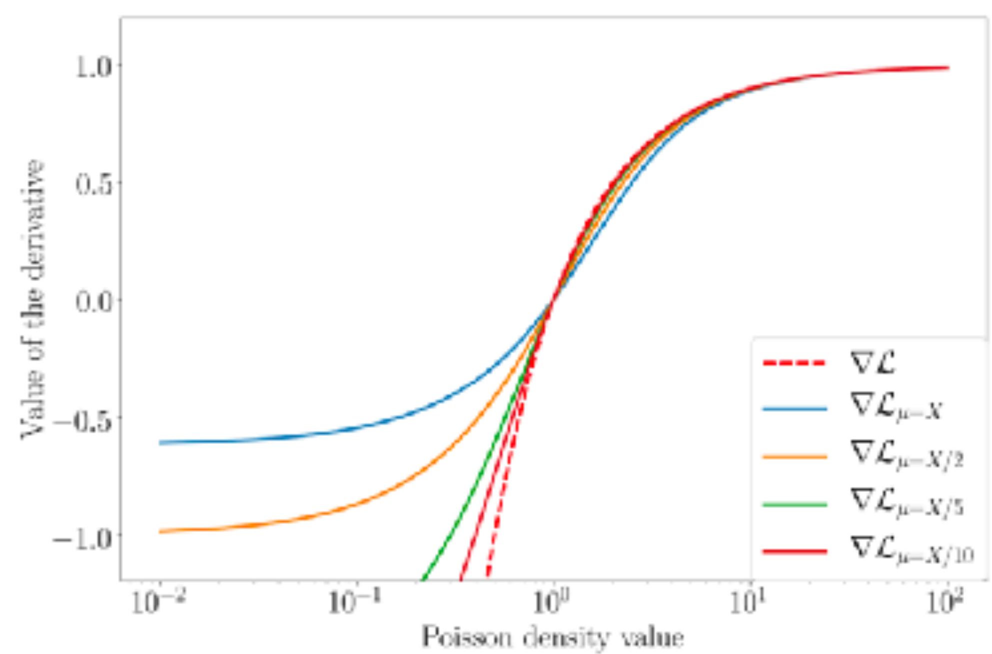


Figure 1: Approximations of $\nabla\mathcal{L}$ for different values of μ.

The pGMCA algorithm

The pGMCA builds upon a Block-Coordinate Descent algorithm:

Initialization:

i) Starts from the solution given by GMCA

ii) Re-weighted l1 parameters derived from the GMCA solution

$$\min_{\alpha} \|\mathbf{\Lambda} \odot \alpha\|_{\ell_1} + \iota_{K^+}(\alpha \mathbf{\Phi}) + \mathcal{L}_\mu(\mathbf{X}|\mathbf{A}, \alpha\mathbf{\Phi})$$

Solved using a G-FBS implementation *(Raguet et al. 2011)*

$$\min_{\mathbf{A}} \iota_{\mathcal{C}}(\mathbf{A}) + \mathcal{L}_\mu(\mathbf{X}|\mathbf{A}, \mathbf{S})$$

Positivity and oblique constraint

Solved using a FISTA implementation

Bobin et al, submitted 2019.

Choosing the regularisation parameter

Standard heuristic in the case of additive Gaussian noise:

$$\alpha^{+} = \mathrm{Prox}_{\gamma\|\mathbf{\Lambda}\odot\,.\,\|_{\ell_1}}\left(\alpha^{-} - \gamma \mathbf{A}^T \nabla \mathcal{L}_\mu(\mathbf{X}|\alpha^{-})\mathbf{\Phi}^T\right)$$

$$\lambda_i \propto \sigma\left\{\gamma\left[\mathbf{A}^T \nabla \mathcal{L}_\mu(\mathbf{X}|\alpha^{-})\mathbf{\Phi}^T\right]_i\right\}$$

Noise term that propagates through the gradient descent

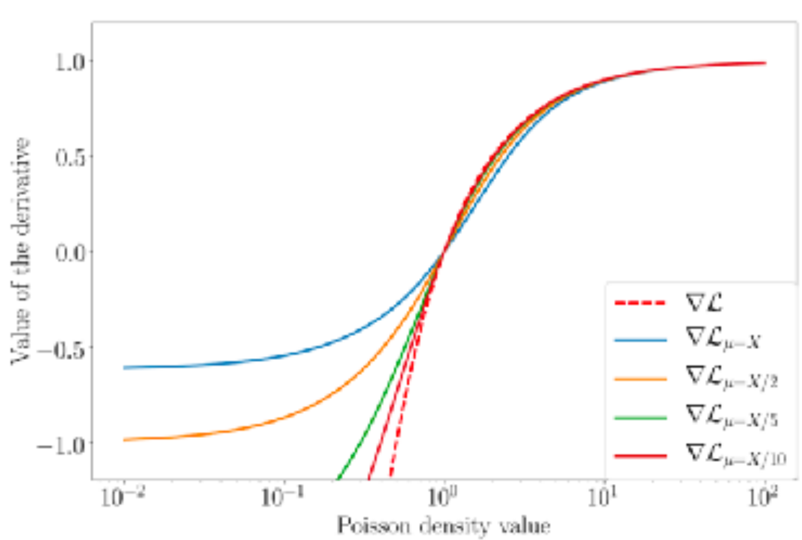
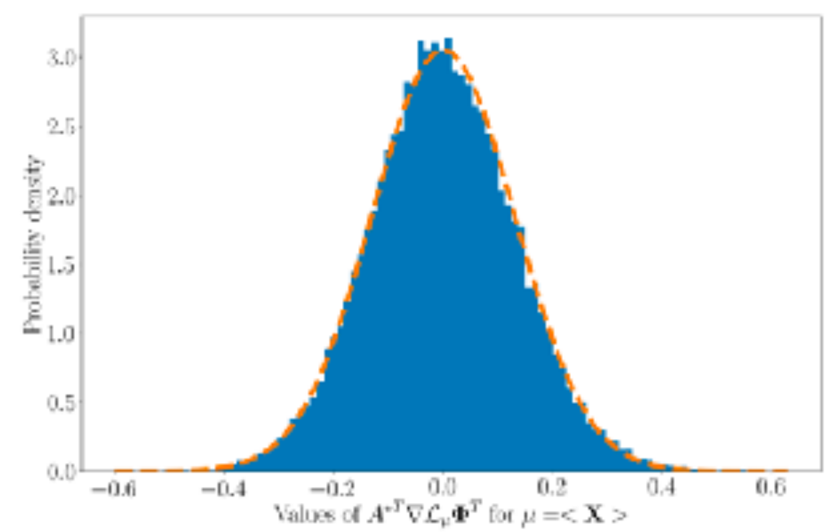


Figure 1: Approximations of $\nabla\mathcal{L}$ for different values of μ.



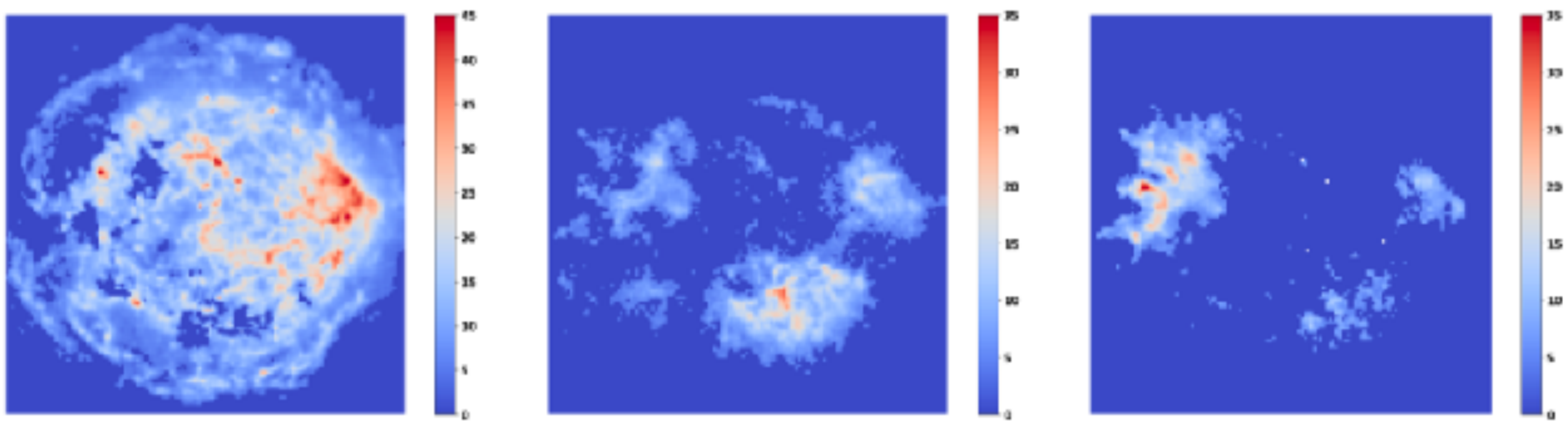
Smoothing tend to "gaussianize" the variance

For $\mu \sim <\mathbf{X}>$ the exact same strategy can be used thanks to smoothing

Illustration on synthetic astrophysical data

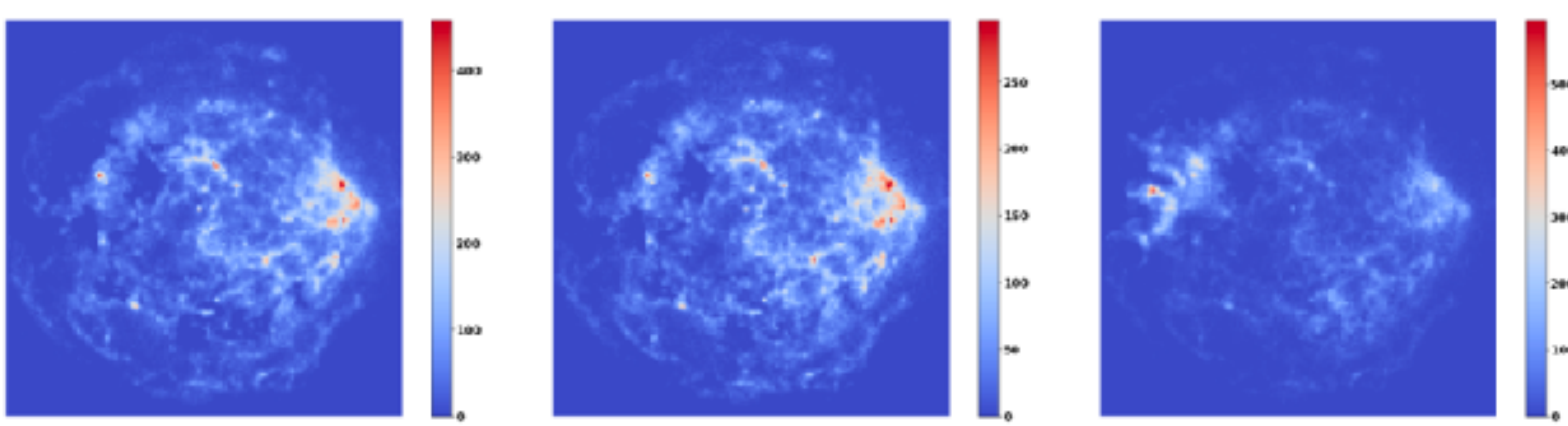
Synthetic Chandra X-ray data with a 3 sources model:

- **synchrotron emission**
- **2 redshifted iron sources**



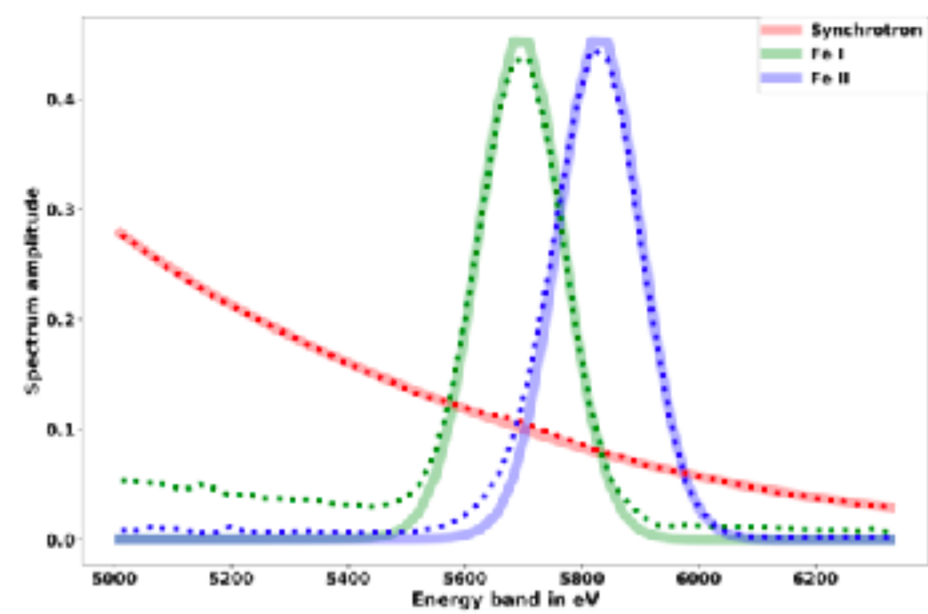
(a) Synchrotron (b) Fe-1 (c) Fe-2

Sources



(a) 5279 eV (b) 5548 eV (c) 5817 eV

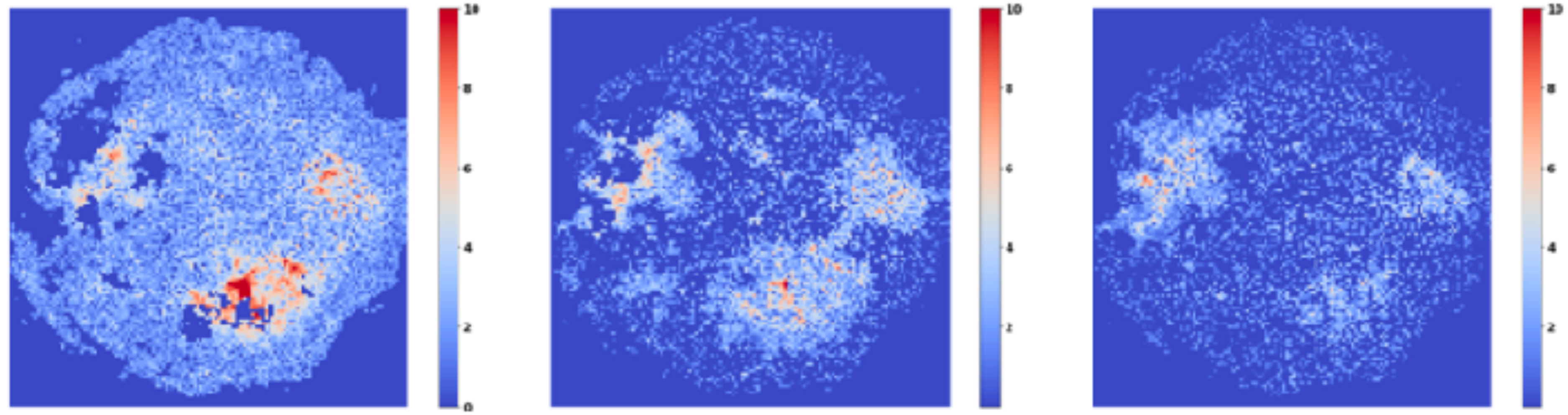
Observations



Spectra

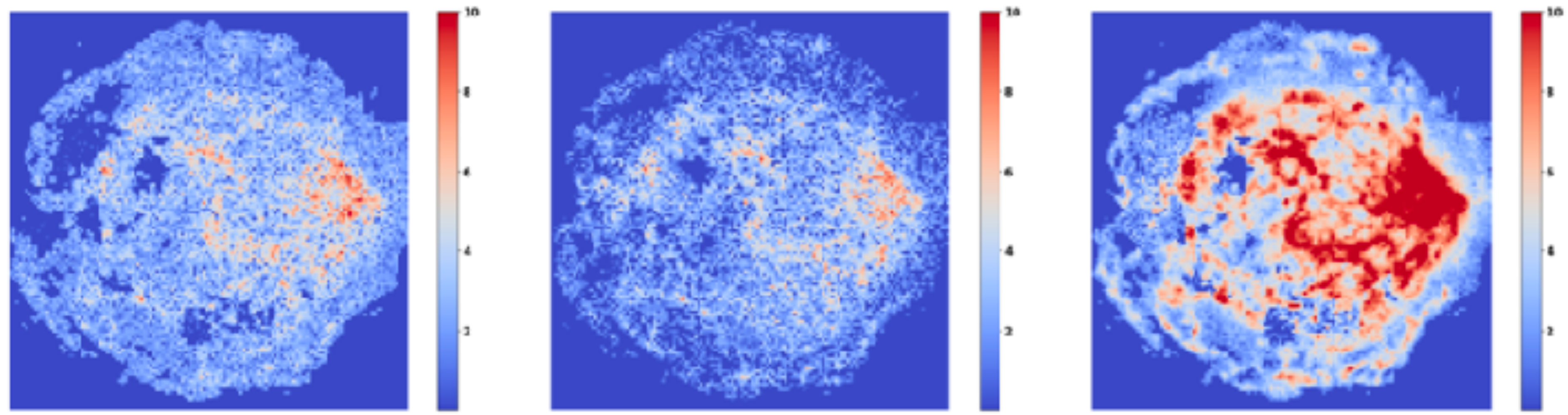
Illustration on synthetic astrophysical data

GMCA



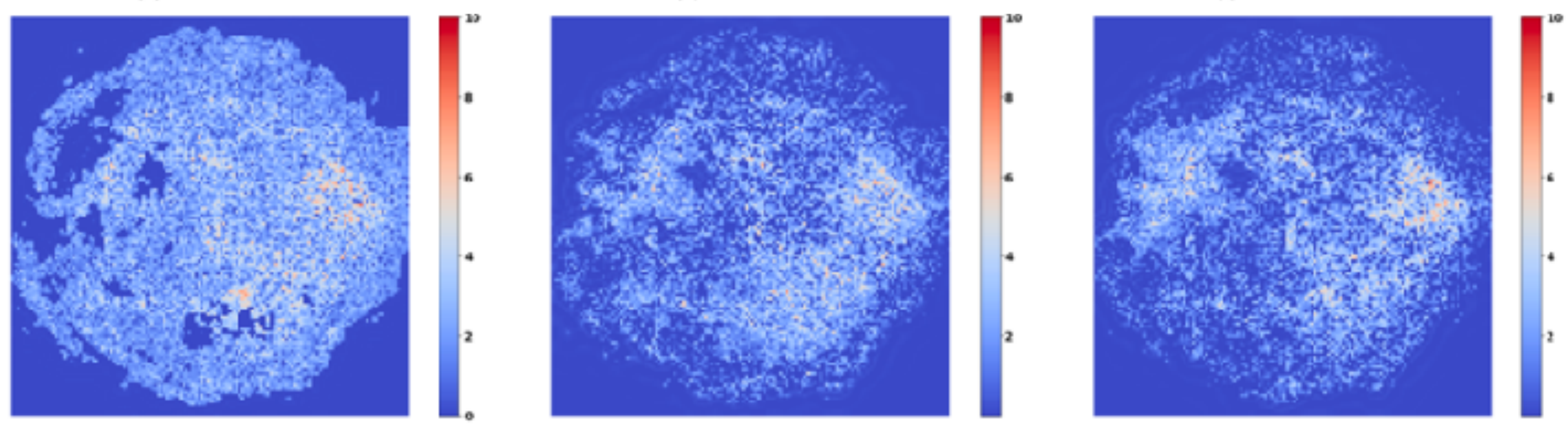
(d) Residual error (e) Residual error (f) Residual error

HALS



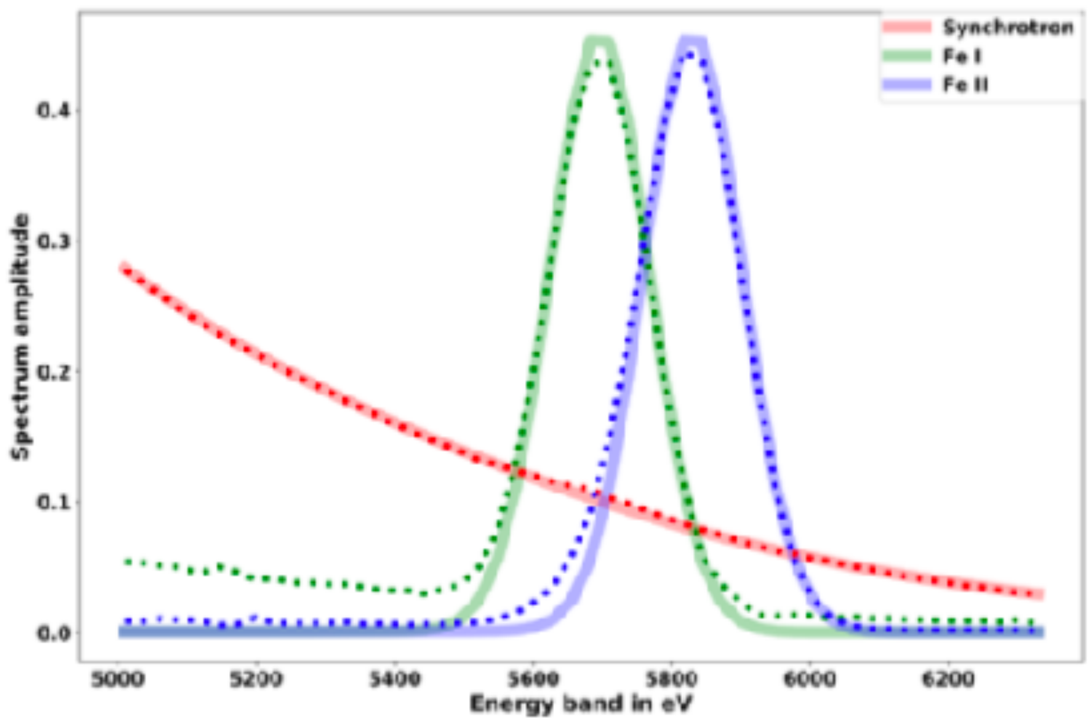
(d) Residual error (e) Residual error (f) Residual error

pGMCA

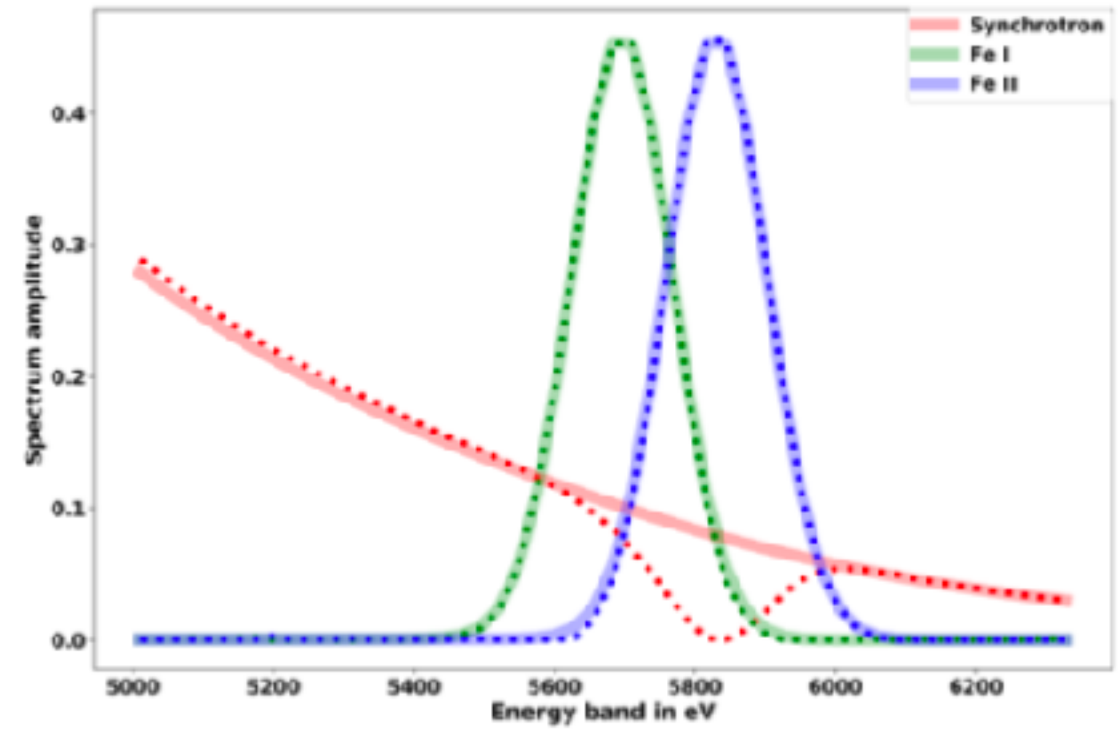


(d) Residual error (e) Residual error (f) Residual error

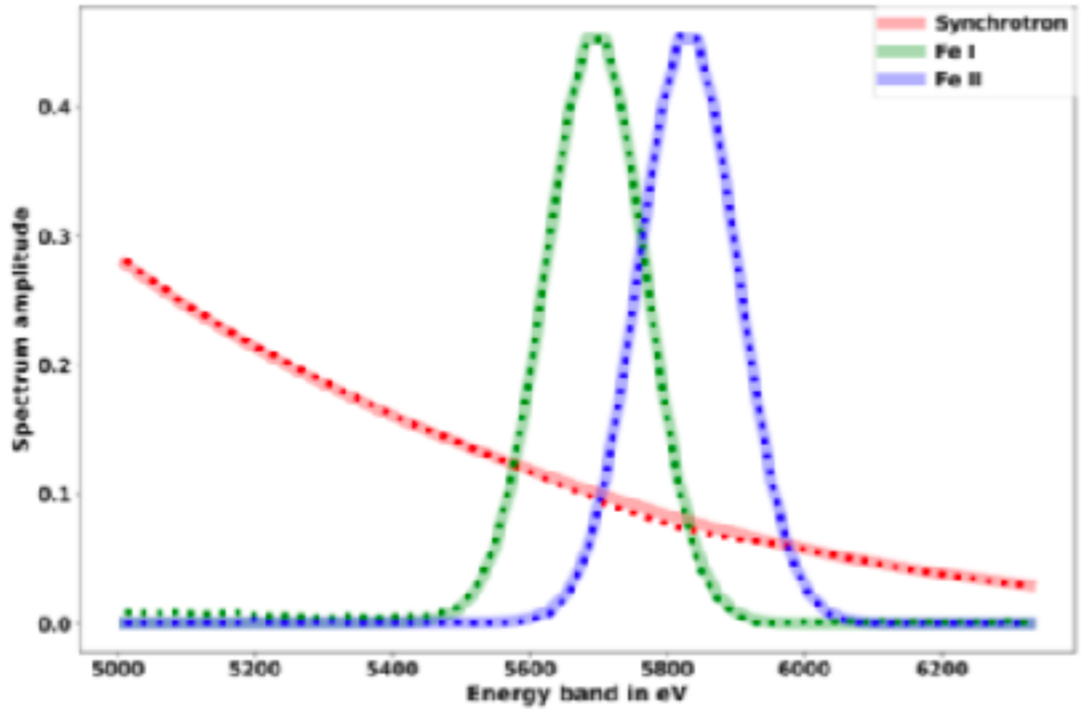
Illustration on synthetic astrophysical data



(a) GMCA



(c) HALS

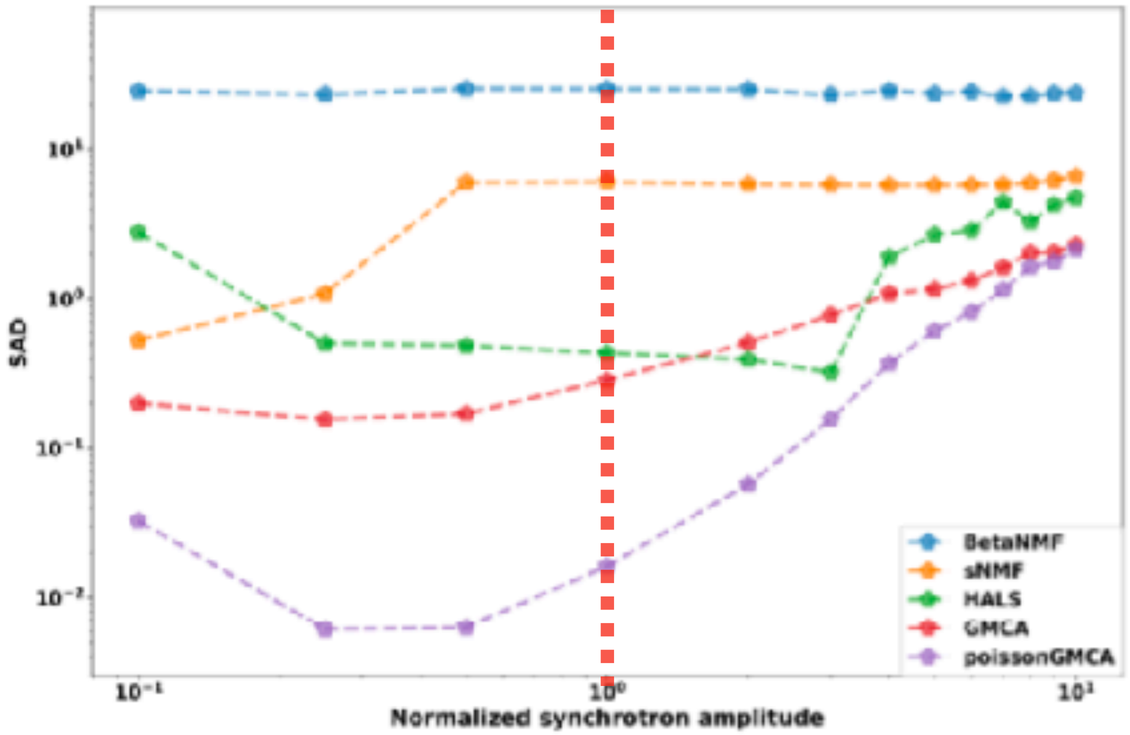


(b) pGMCA

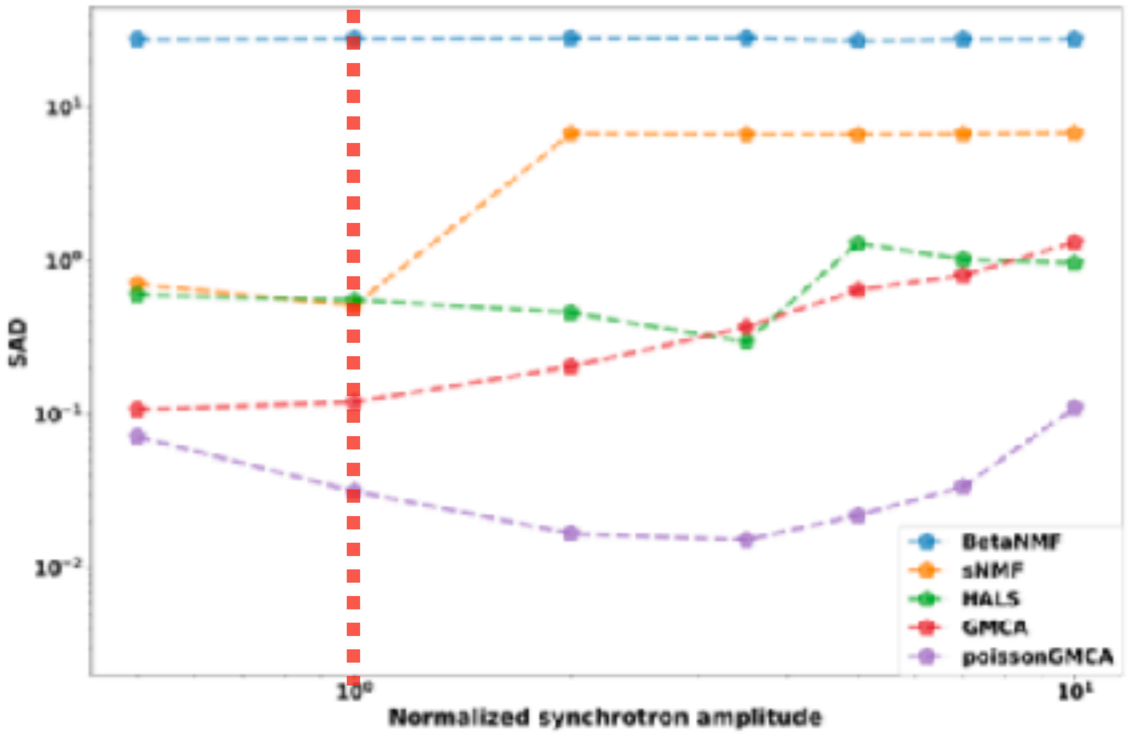
Illustration on synthetic astrophysical data

Scenario: in supernovae remnants, the synchrotron emission is a background w.r.t atomic components

The goal is to evaluate the robustness of the separation process w.r.t to the background level

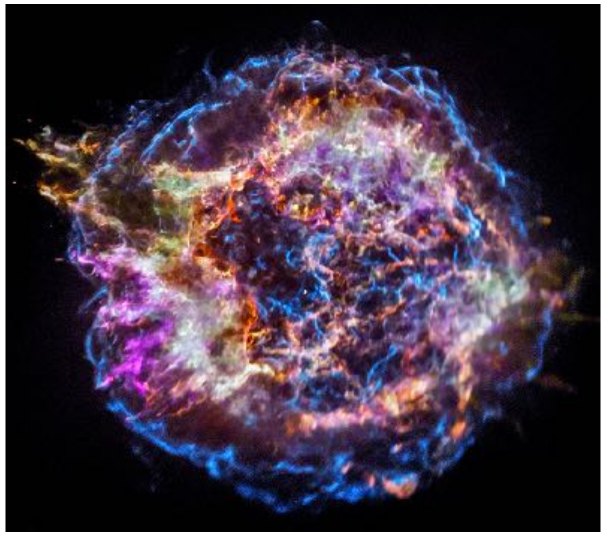


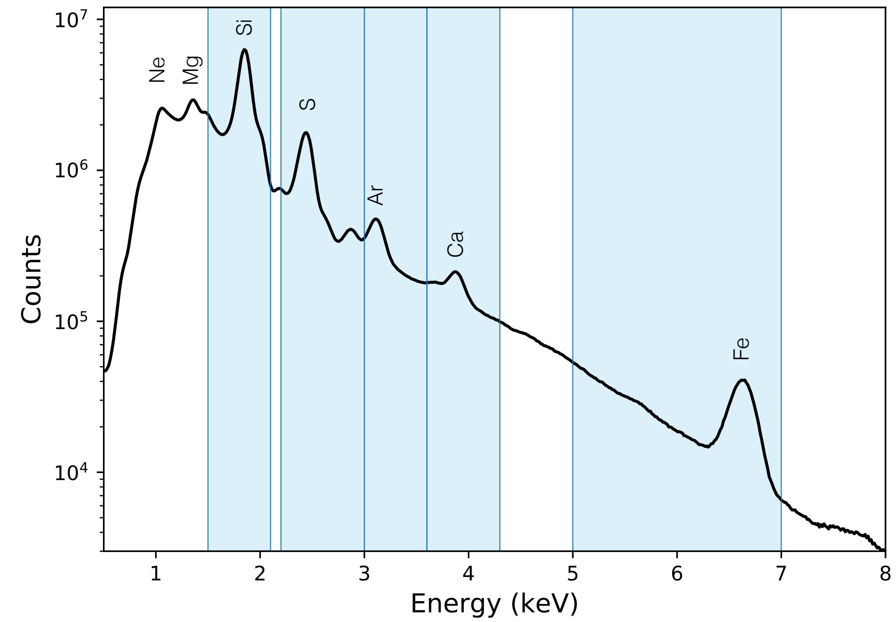
(a) Case $m = 12$

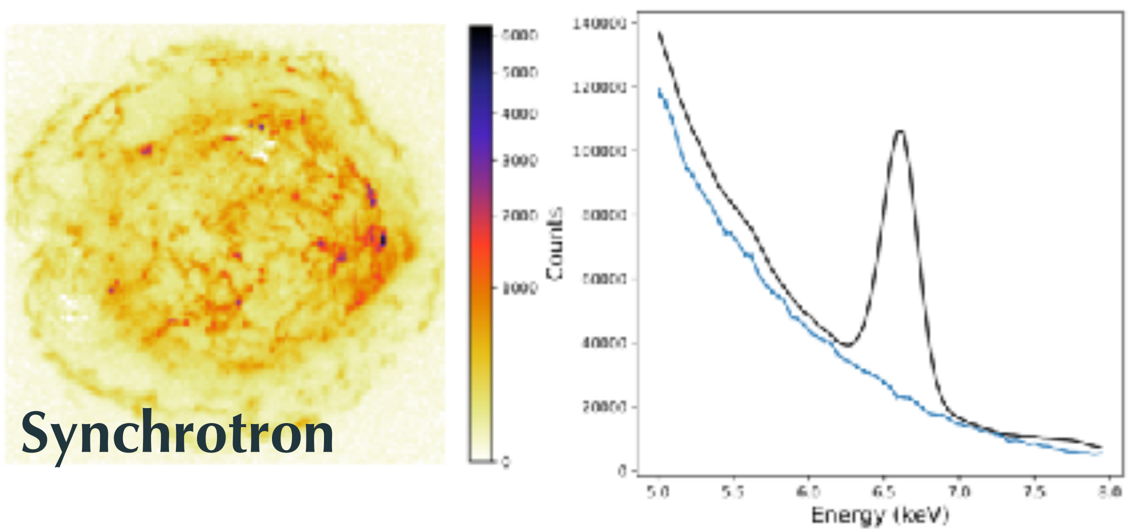


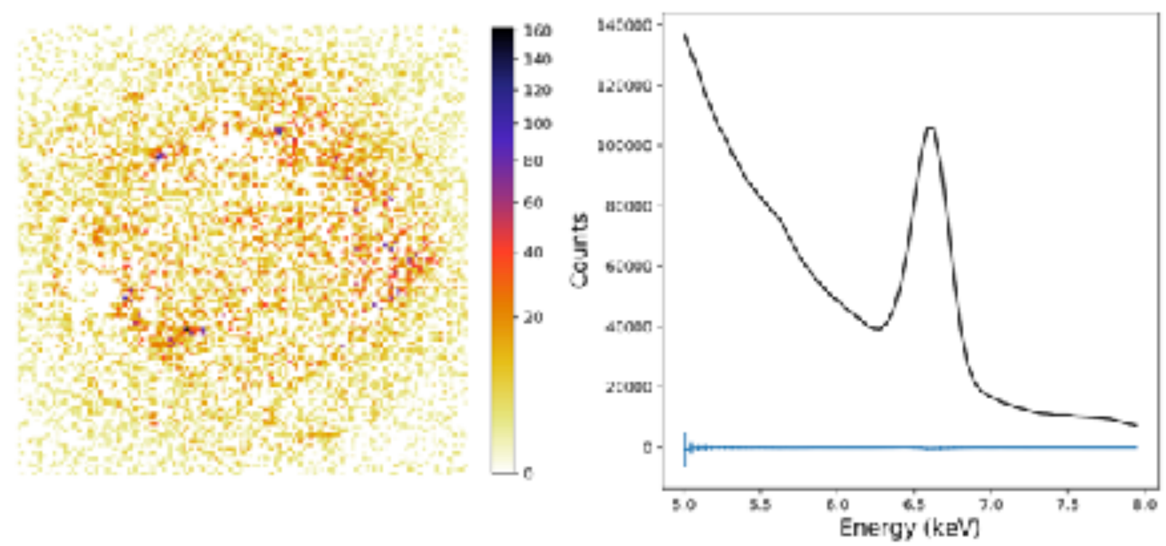
(b) Case $m = 50$

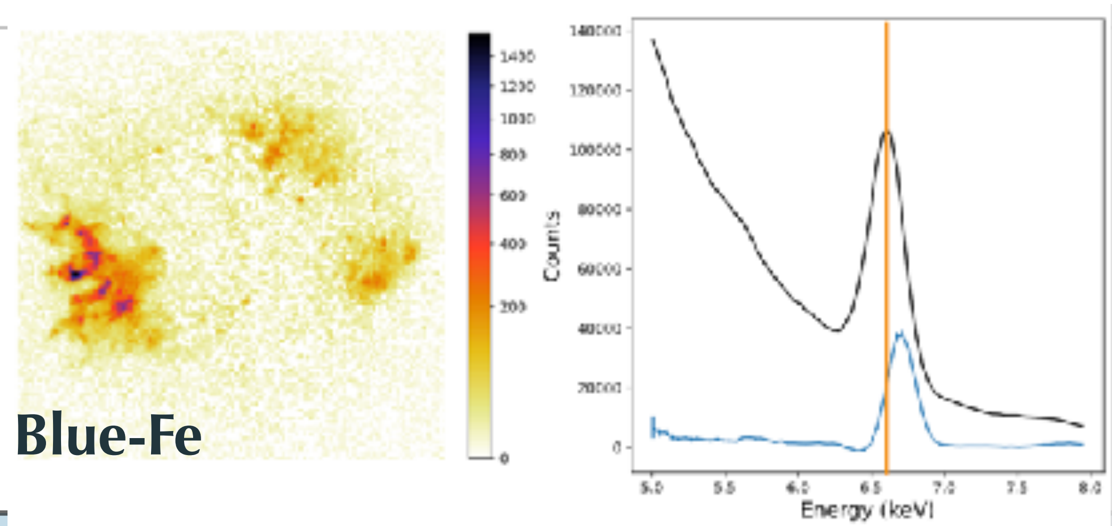
Application to the Chandra data

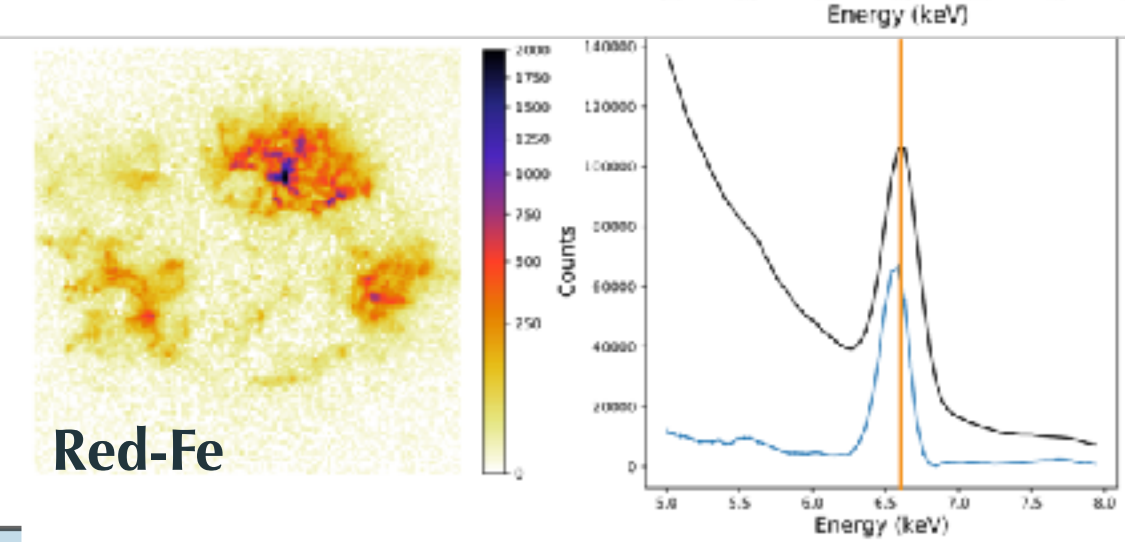




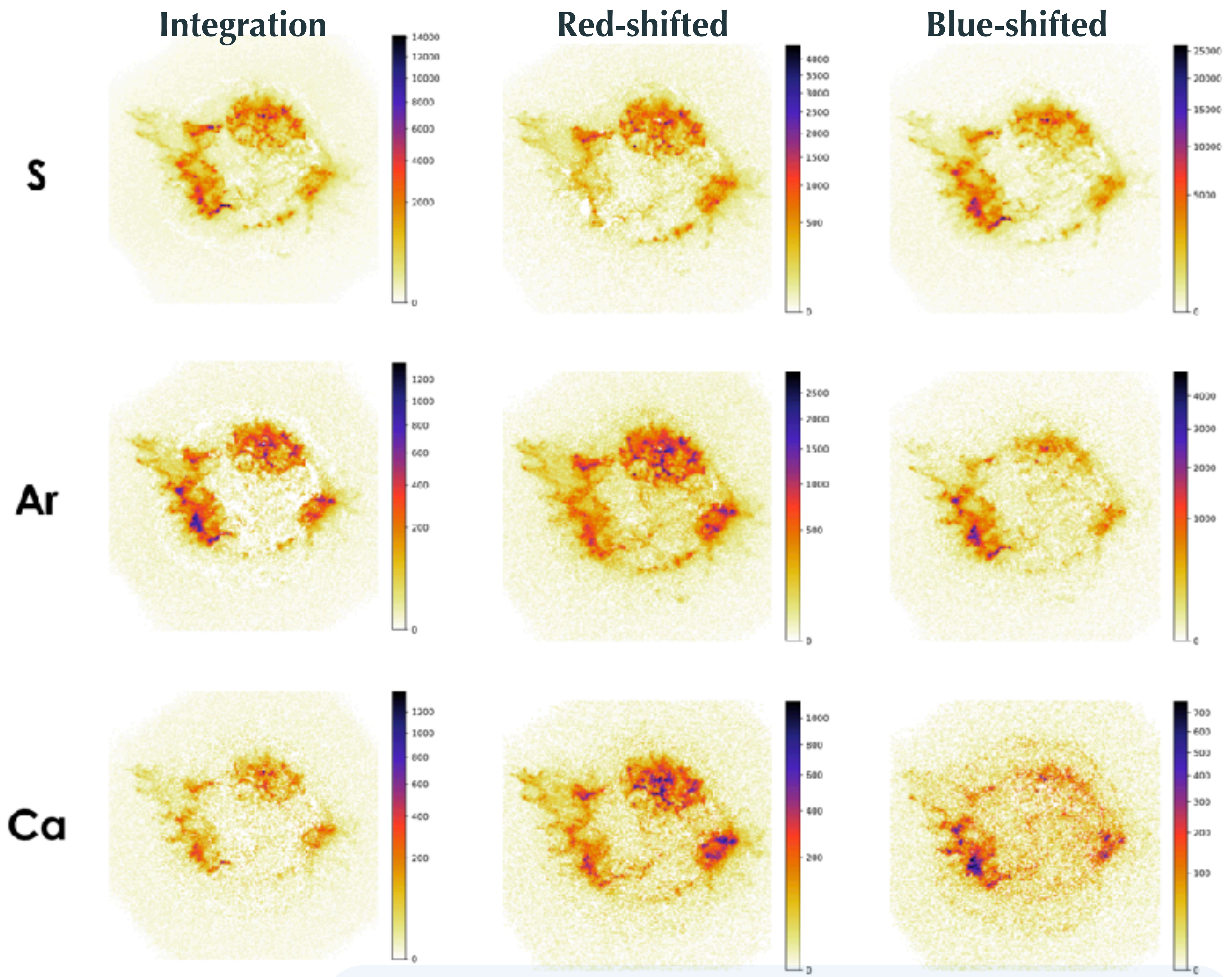








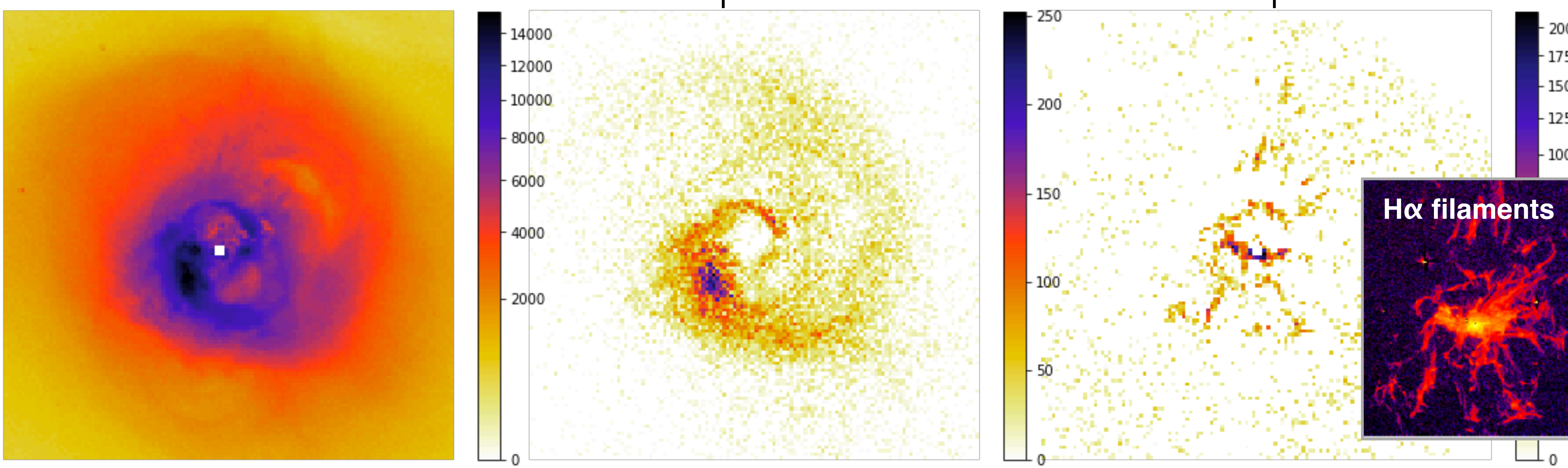
Application to the Chandra data



Picquenot et al, submitted 2019.

Blindly estimates red/blue-shifted atomic components !

Chandra image 0.5-8 keV	GMCA component 2	GMCA component 3



X-ray filaments have~50-100 counts buried under 10^4 counts
Finding features with contrast < 1%

First time detection of X-ray filaments
Filaments features would be impossible to find without a unsupervised approach

Take-away messages

Flexible framework to tackle:

- **More generic statistical models**
    - **Account for Poisson statistics made easier thanks to Nesterov's smoothing + BCD**

Effective on real-world data, opens a new way to analyse X-ray data

pyGMCALab: [***https://github.com/jbobin/pyGMCALab***](https://github.com/jbobin/pyGMCALab)

Also visit the lab's website [**www.cosmostat.org**](http://www.cosmostat.org)