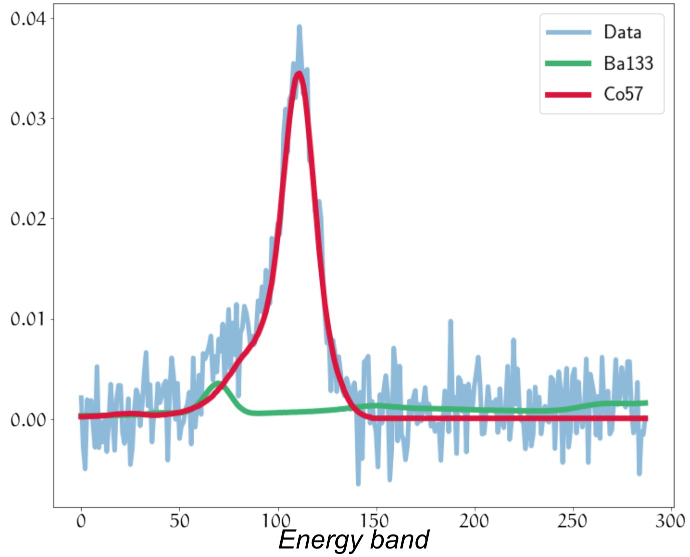
Data frugal machine learning approaches for unmixing problems in Physics

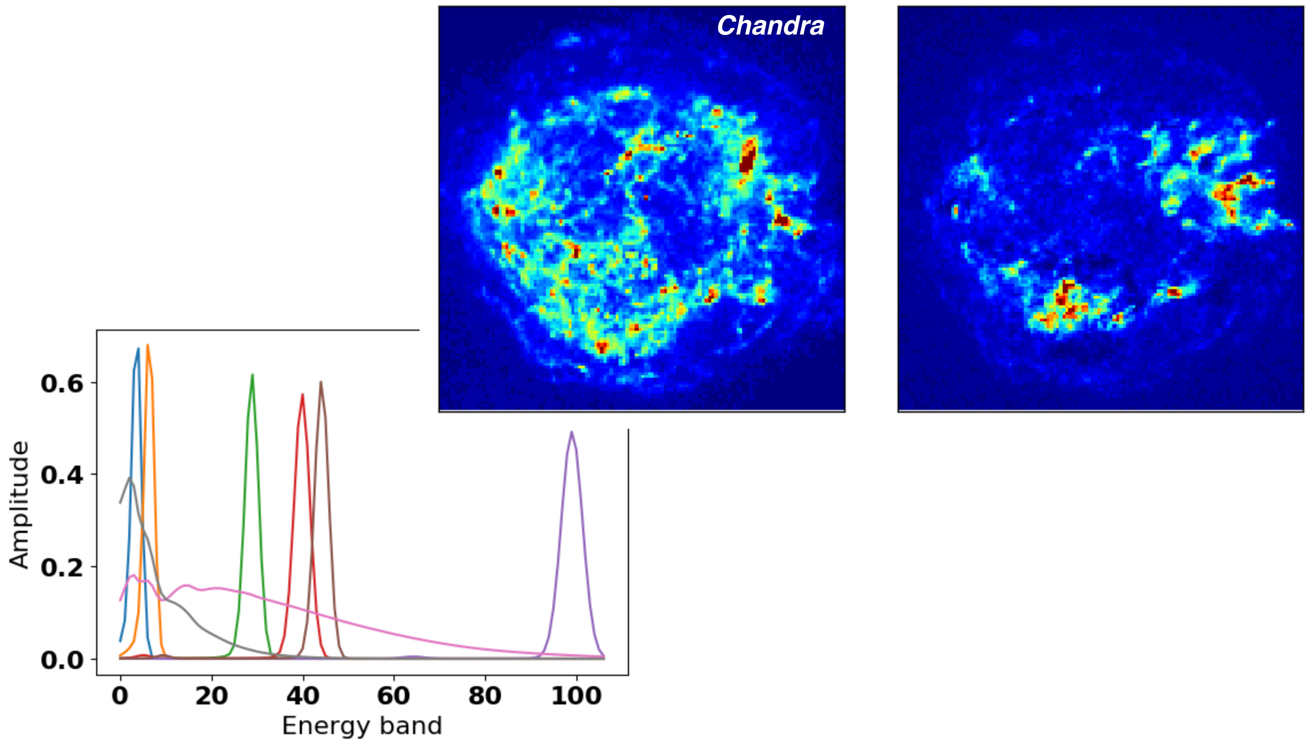
J.Bobin

With R. Carloni - F.Acero - J.Lascar - T. Pham - C.Bobin

Some unmixing problems in physics



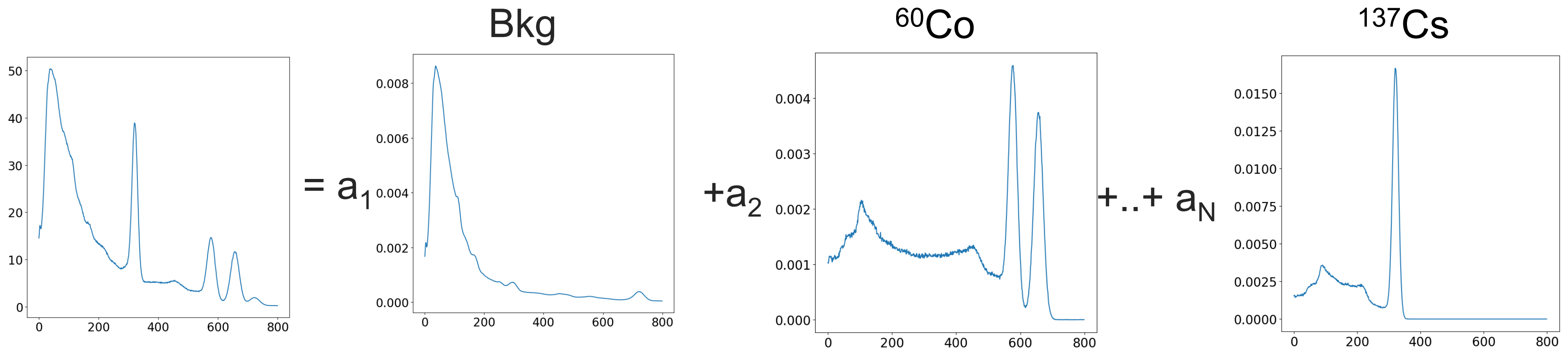
Unmixing γ-ray spectra
to recover radionuclides'activities



Unmixing X-ray multispectral images
to recover physically relevant components
e.g. synchrotron, thermal, etc.

And many others: radio-astronomy, gravitational wave astro., etc.

Unmixing, what's at stake ?



The mixing matrix

The source matrix

$$\mathbf{X} = \mathcal{F}(\mathbf{AS})$$

Contamination

e.g. additive Gaussian noise
Poisson stats.

Blind Source Separation/unmixing

Estimating both A and S from X only

Unmixing, what's at stake ?

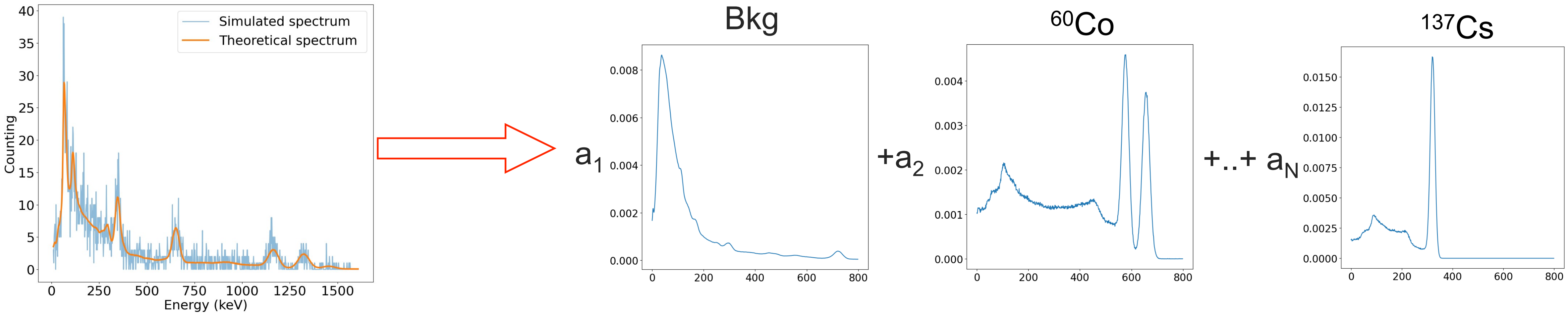
$$\min_{\mathbf{A},\mathbf{S}} \mathcal{R}(\mathbf{A}) + \mathcal{J}(\mathbf{S}) + \frac{1}{2}\|\mathbf{X} - \mathbf{AS}\|_F^2$$

Regularization Terms — *Data fidelity term*

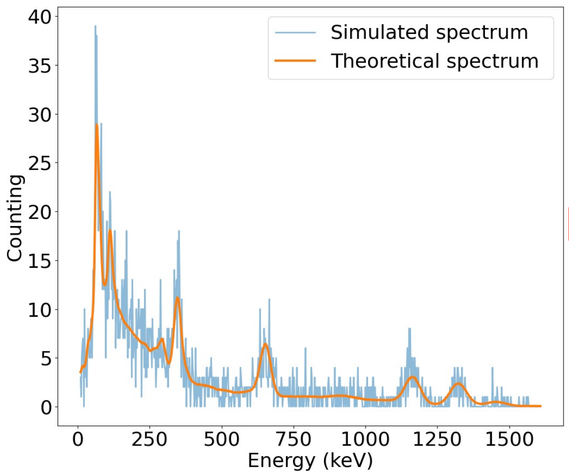
- Allows great **flexibility** to include information about the observation model/prior information about the factors
- BSS is a non-convex problem particularly ill-posed: the **regularization** is **crucial** *(non-negativity, smoothness, sparsity, etc.)*
- But generally ill-posed/badly-posed, requires physics-enforcing regularisations

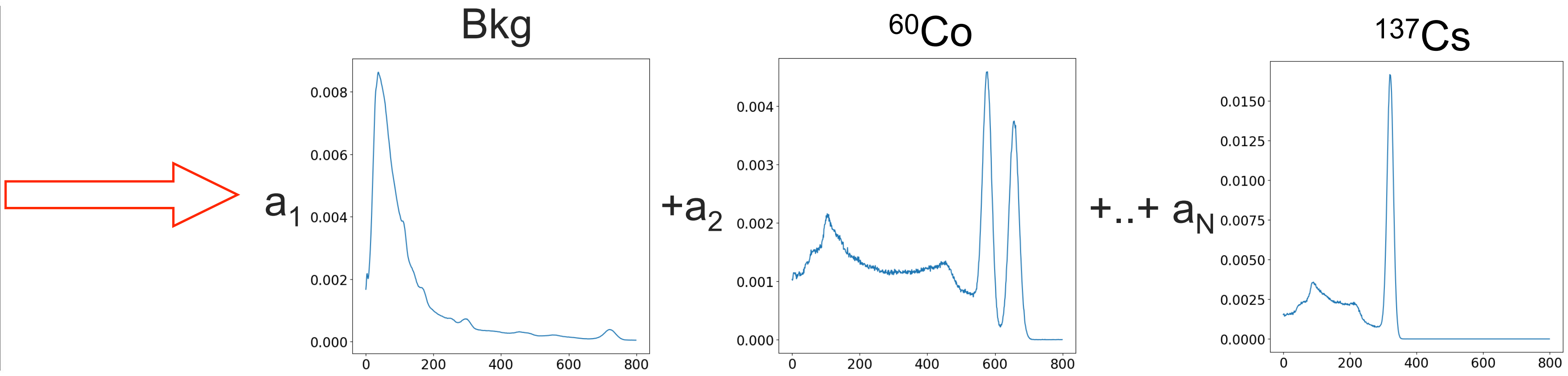
Zibulevsky 04, Comon 10, Bobin 15, Carloni 22, etc.

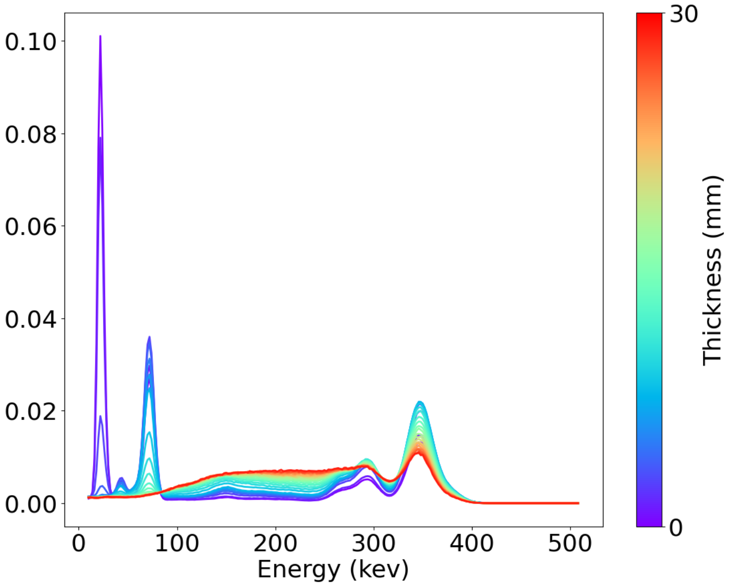
Focus on the spectrometry case



Focus on the spectrometry case



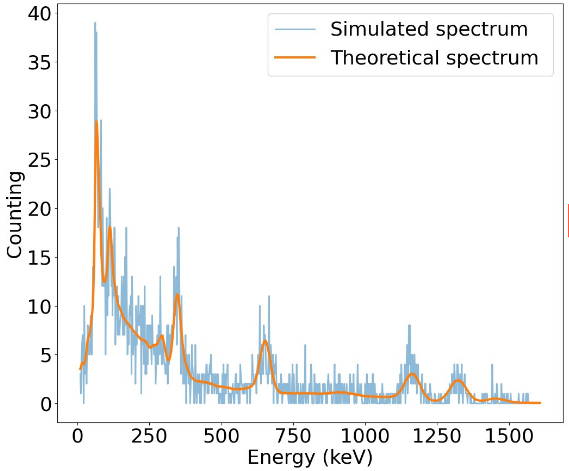


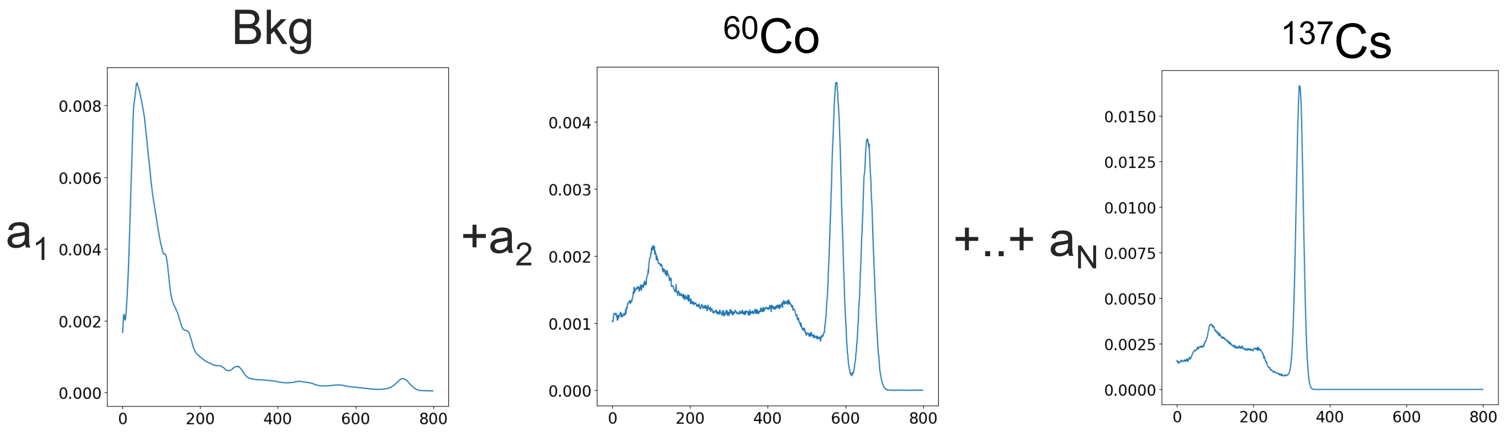


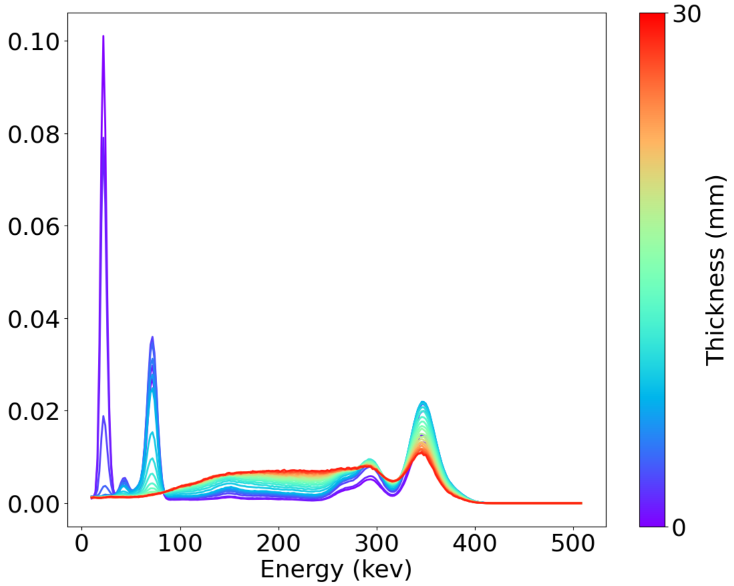
Accounting for the spectral variabilities

*^{133}**Ba** spectral signature as a function of thickness of the container*

Focus on the spectrometry case







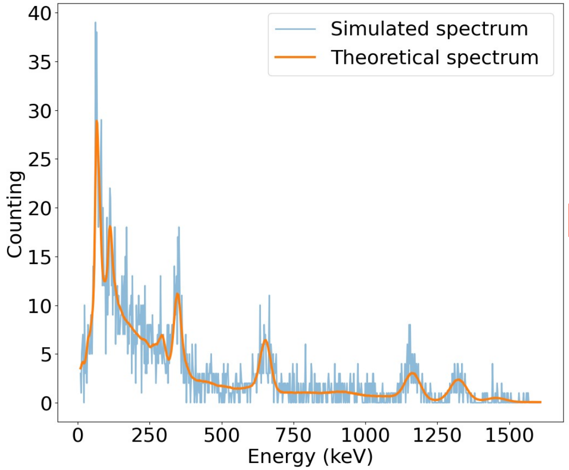
Accounting for the spectral variabilities

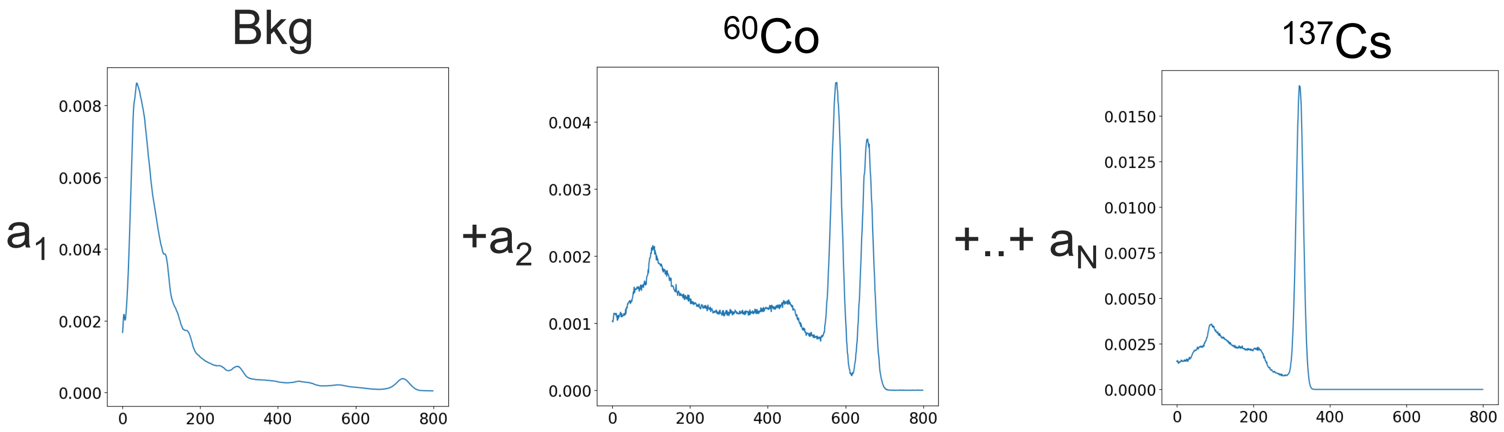
*^{133}**Ba** spectral signature as a function of thickness of the container*

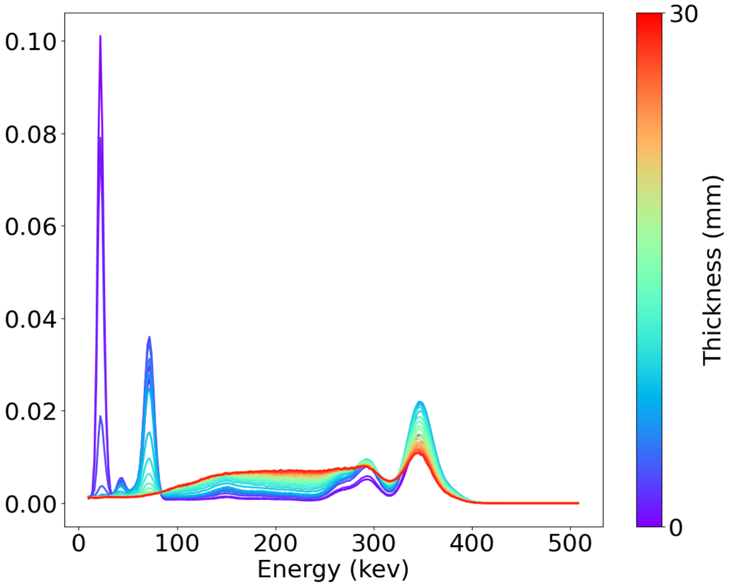
The spectra live on a unknown low-dimensional manifolds

Let's learn a representation for the spectra

Focus on the spectrometry case

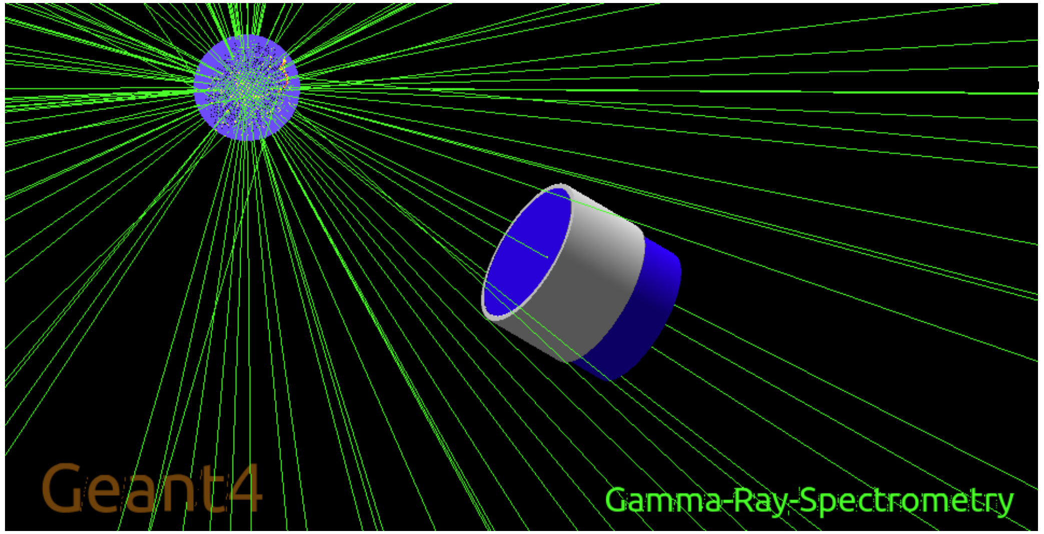






Accounting for the spectral variabilities

*^{133}**Ba** spectral signature as a function of thickness of the container*

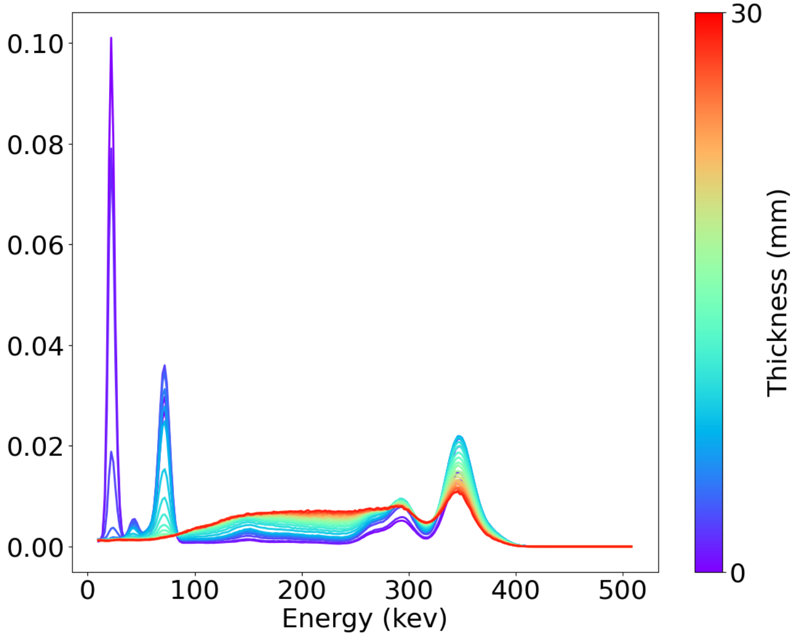


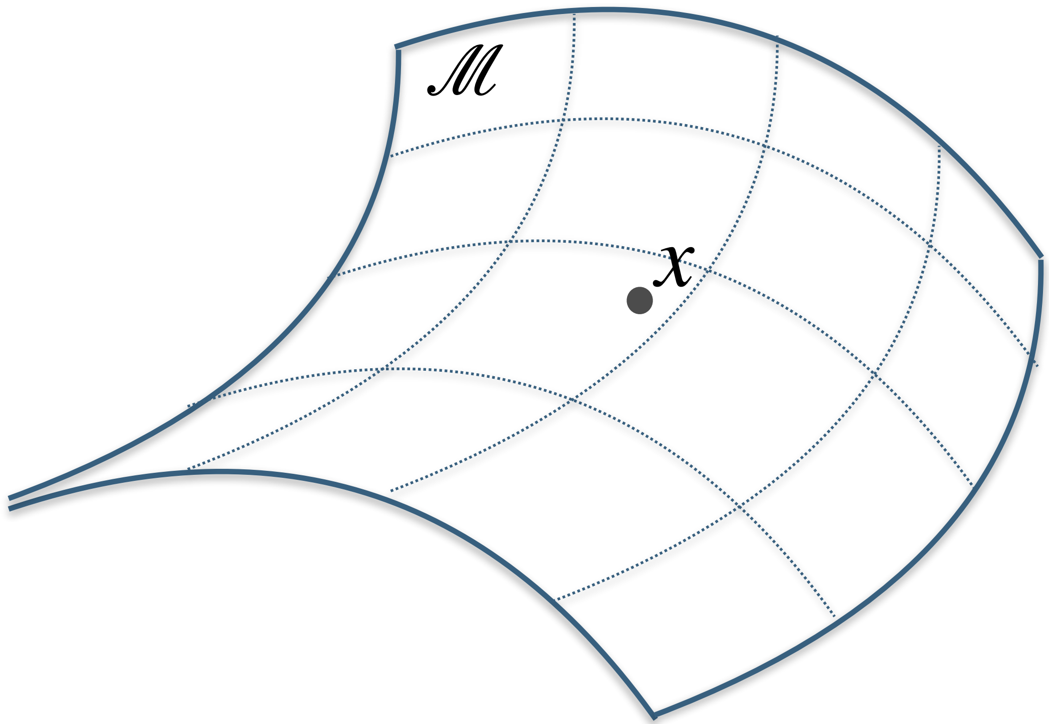
Spectra can be simulated with Monte-Carlo simulations …

1 spectrum in 3 days on a single CPU

Sketch of a data-frugal ML for learning representations

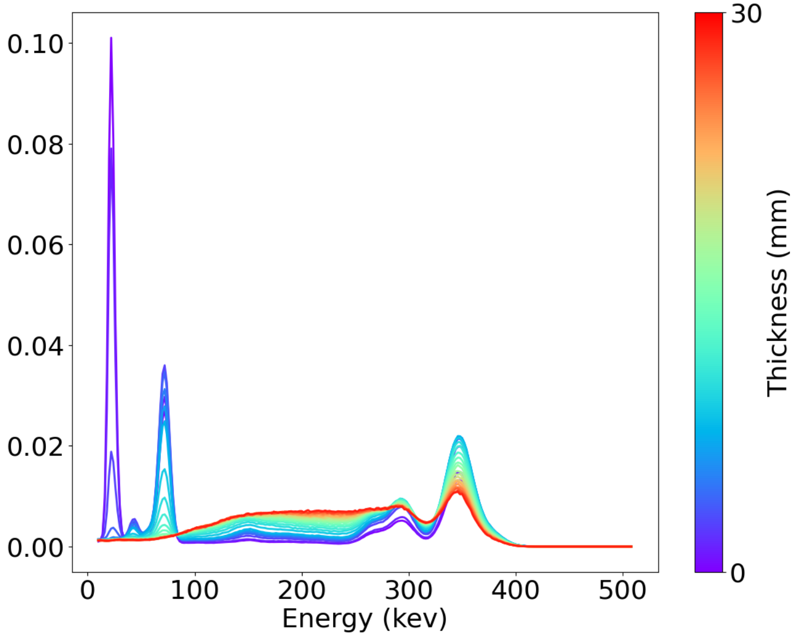
Learn how to transport points on the manifold from **anchor points**

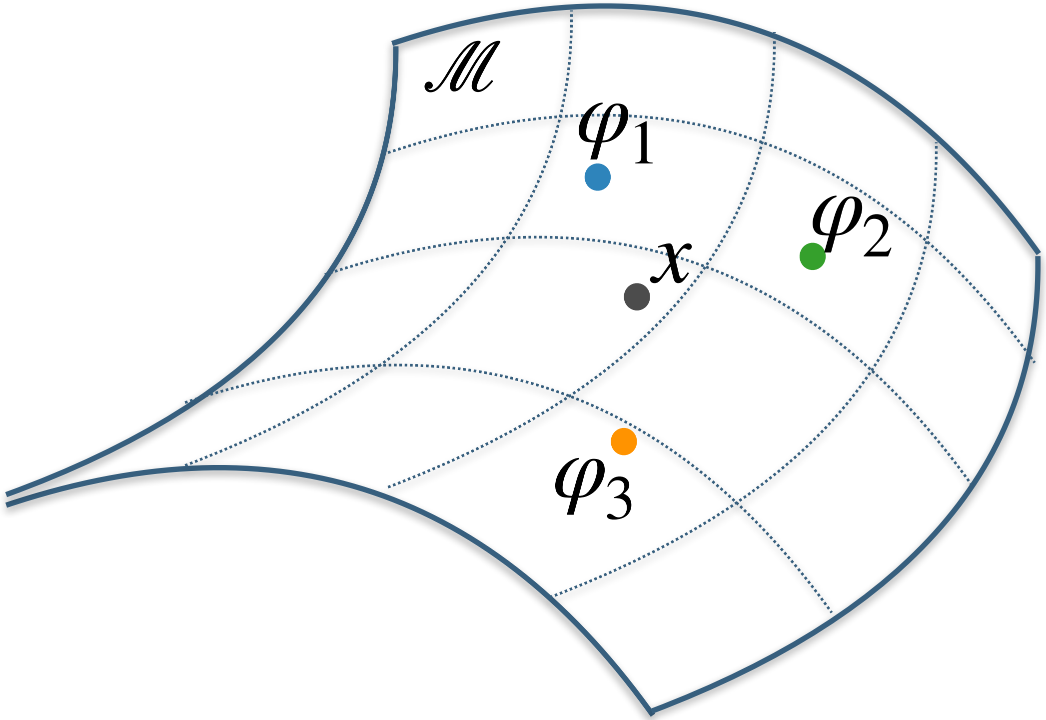




Sketch of a data-frugal ML for learning representations

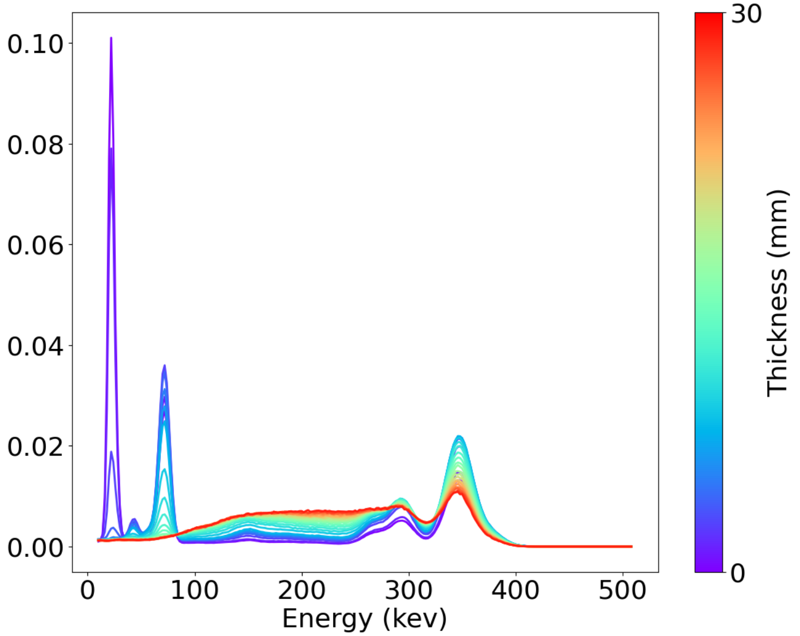
Learn how to transport points on the manifold from **anchor points**

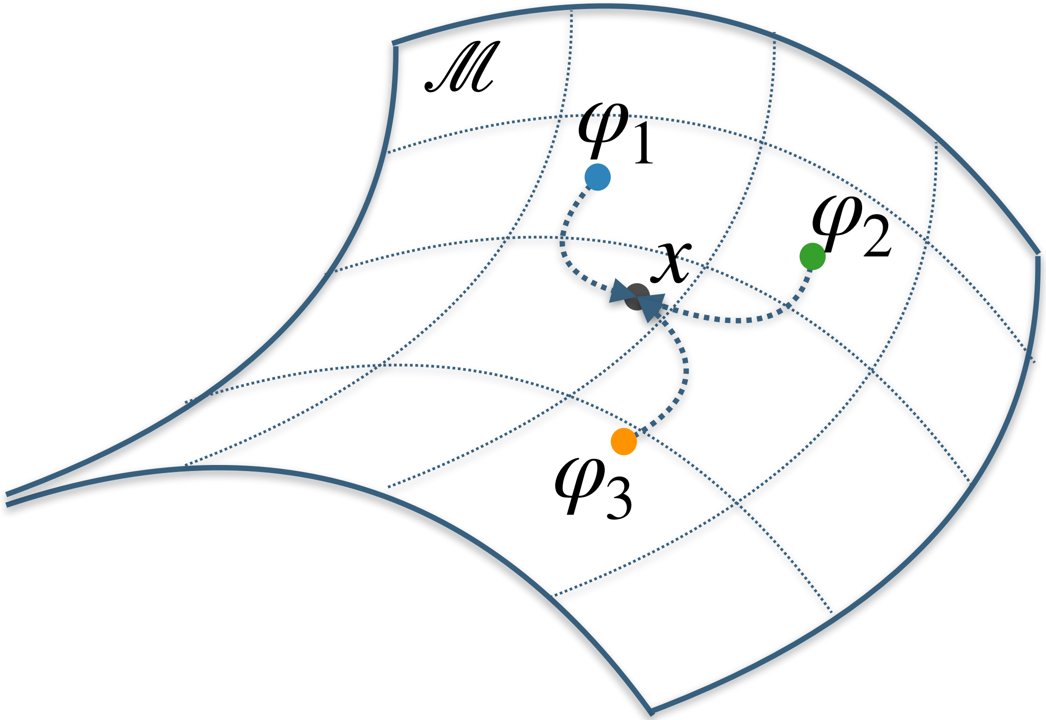




Sketch of a data-frugal ML for learning representations

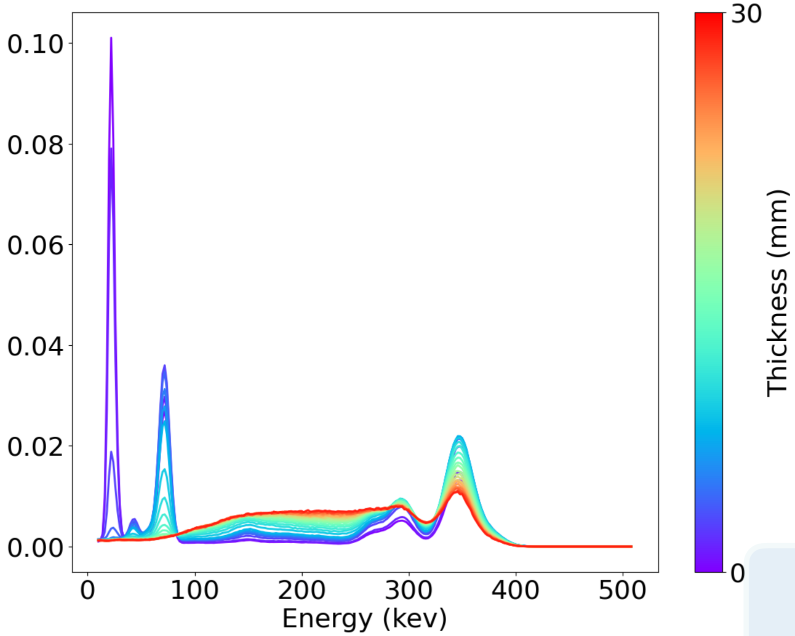
Learn how to transport points on the manifold from **anchor points**

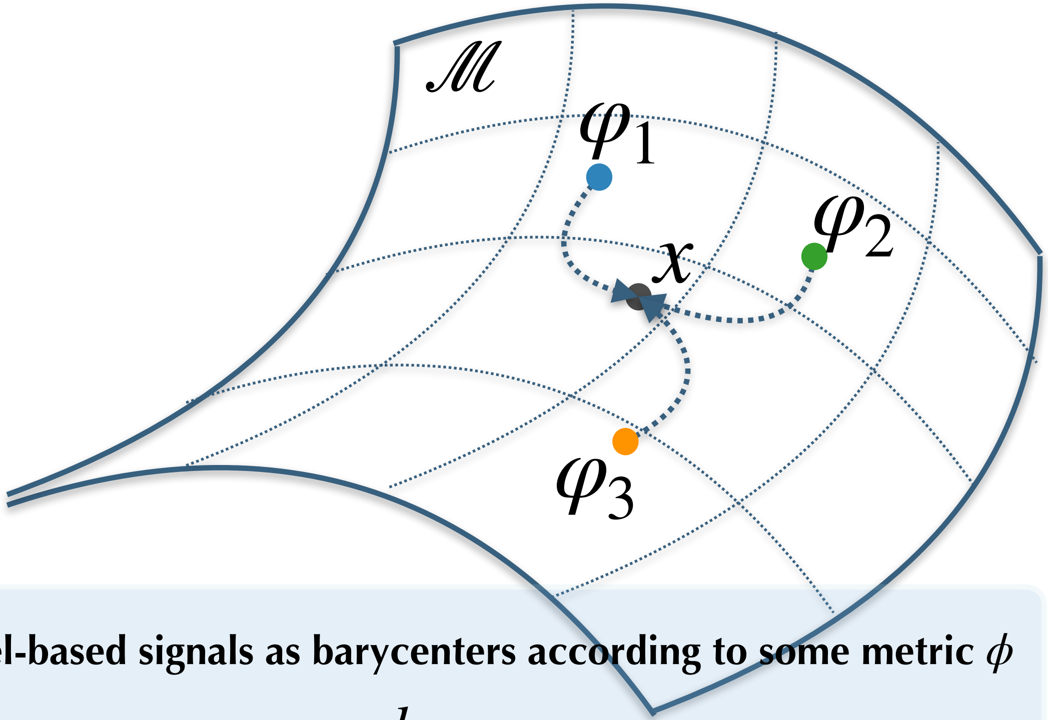




Sketch of a data-frugal ML for learning representations

Learn how to transport points on the manifold from **anchor points**





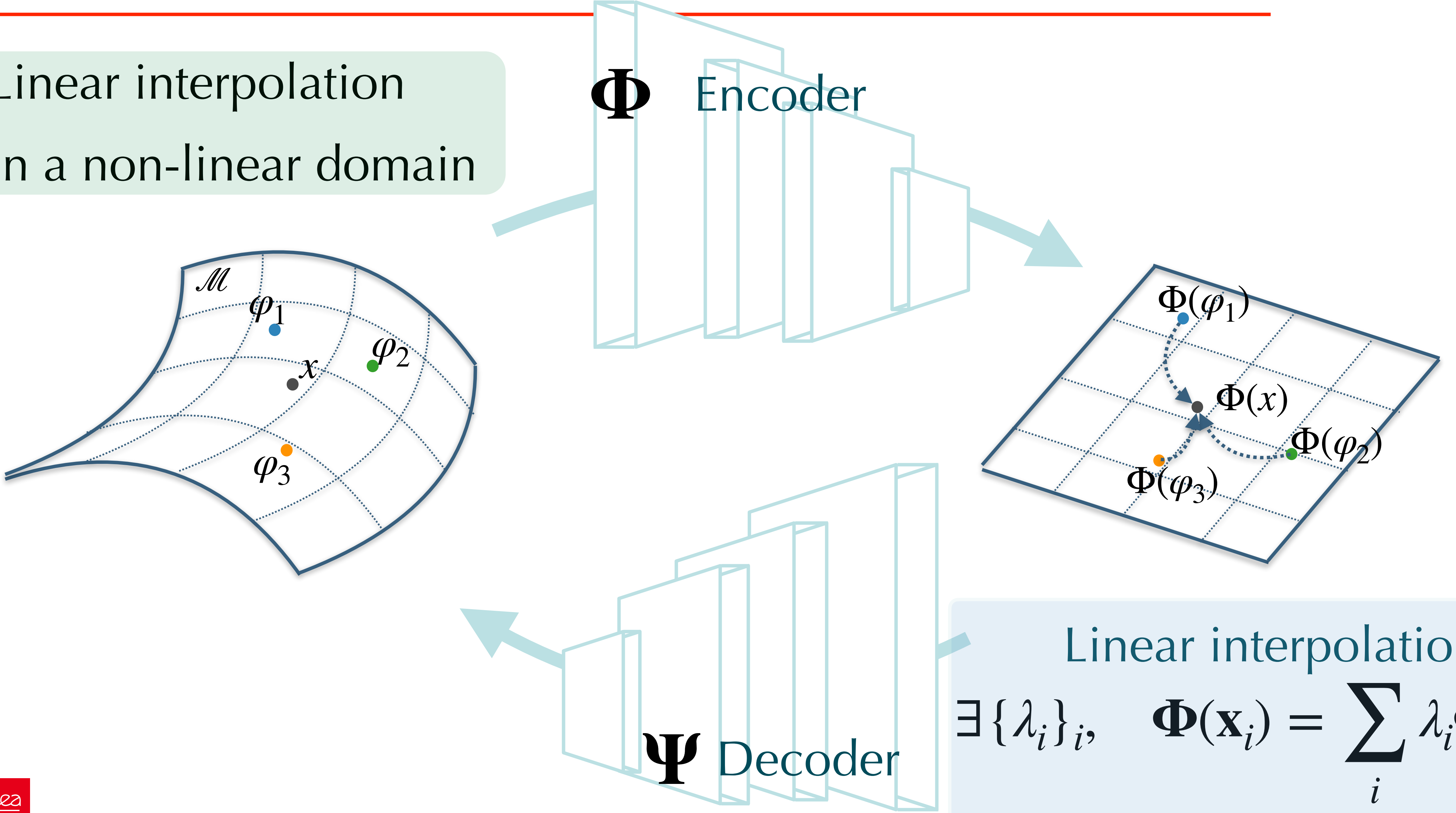
Define model-based signals as barycenters according to some metric ϕ

$$x = \operatorname{argmin}_{\mathbf{z}} \sum_{i=1}^{d} \lambda_i \phi(\mathbf{z}, \varphi_i)$$

Data-frugal AutoEncoder

Linear interpolation

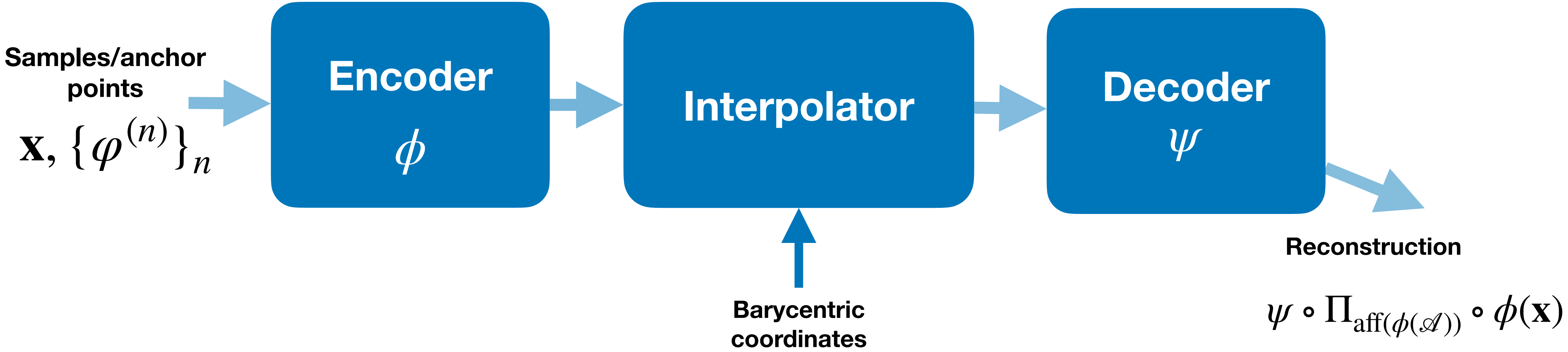
In a non-linear domain



Linear interpolation

$$\exists \{\lambda_i\}_i, \quad \mathbf{\Phi}(\mathbf{x}_i) = \sum_i \lambda_i \mathbf{\Phi}(\varphi_i)$$

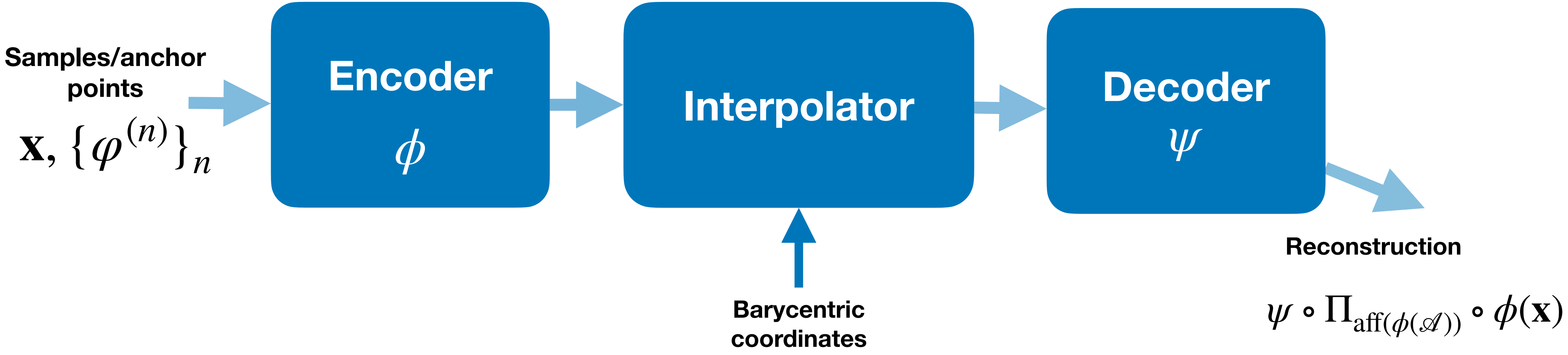
Data-frugal AutoEncoder



Training: $$\min_{\phi,\psi} \sum_{\mathbf{x}\in\mathscr{T}} \left\| \mathbf{x} - \psi \circ \Pi_{\mathrm{aff}(\phi(\mathscr{A}))} \circ \phi(\mathbf{x}) \right\|^2 + \mu \sum_{\mathbf{x}\in\mathscr{T}} \left\| \phi(\mathbf{x}) - \Pi_{\mathrm{aff}(\phi(\mathscr{A}))} \circ \phi(\mathbf{x}) \right\|^2$$

Reconstruction error *Interpolation error*

Data-frugal AutoEncoder



Training:

$$\min_{\phi,\psi} \sum_{\mathbf{x}\in\mathscr{T}} \left\| \mathbf{x} - \psi \circ \Pi_{\mathrm{aff}(\phi(\mathscr{A}))} \circ \phi(\mathbf{x}) \right\|^2 + \mu \sum_{\mathbf{x}\in\mathscr{T}} \left\| \phi(\mathbf{x}) - \Pi_{\mathrm{aff}(\phi(\mathscr{A}))} \circ \phi(\mathbf{x}) \right\|^2$$

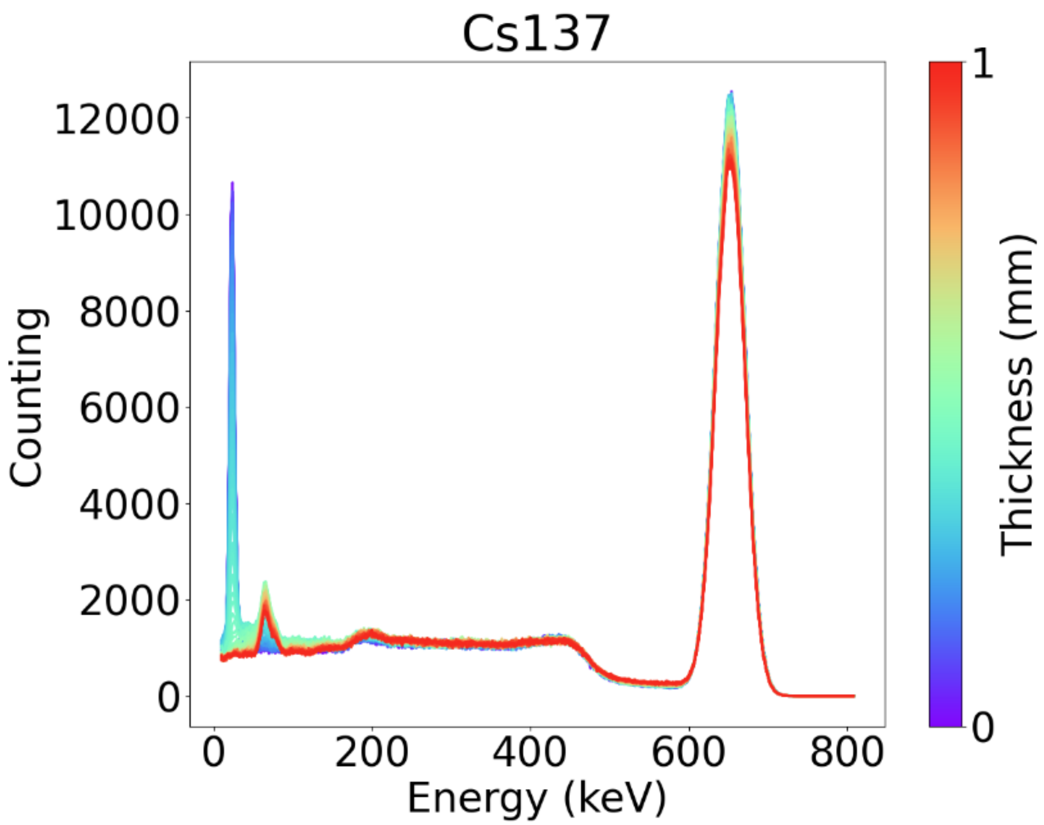
Reconstruction error — *Interpolation error*

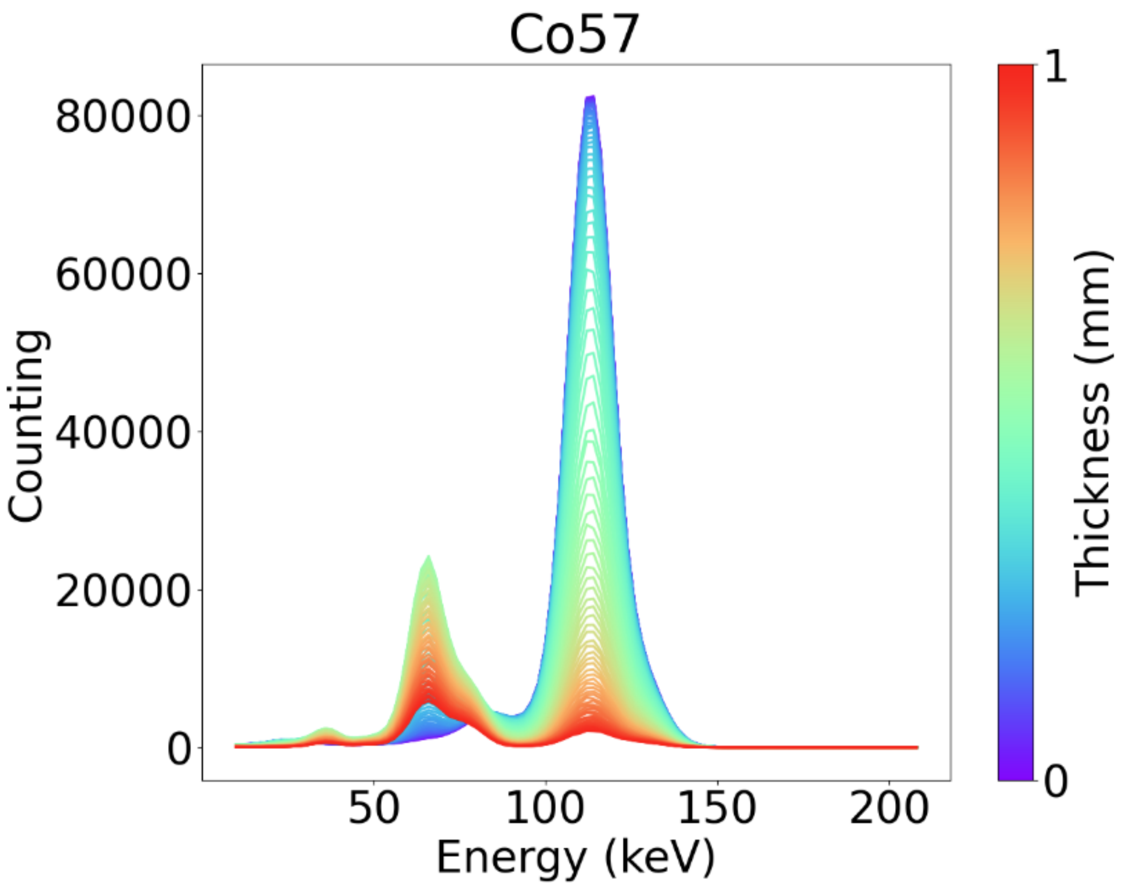
Ideally, all elements of the manifolds can be expressed as the **decoding** of *a* ***linear combination*** *of the* ***encoded anchor points*** :

$$\forall \mathbf{x} \in \mathscr{V}, \exists \{\lambda_n\}_n, \mathbf{x} \approx \psi\left(\sum_n \lambda_n \phi(\varphi^{(n)})\right)$$

Results

Modelling attenuation and Compton scattering by a lead sphere

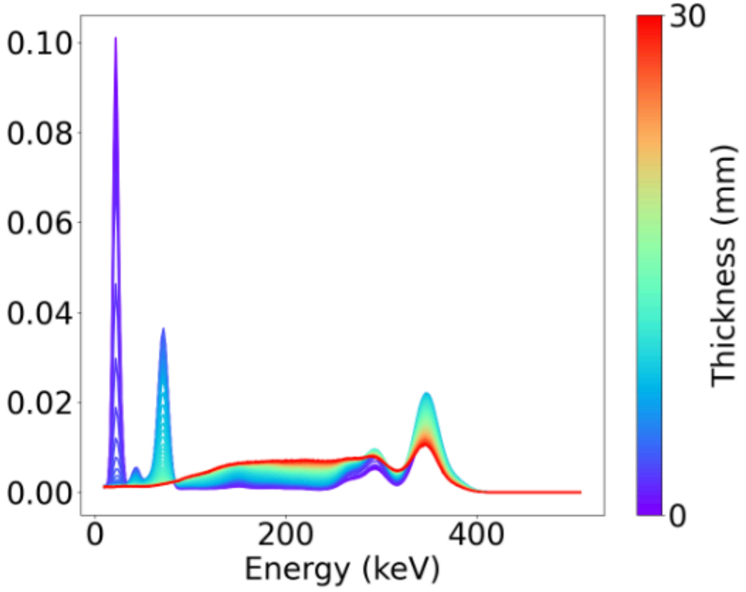




Representation examples

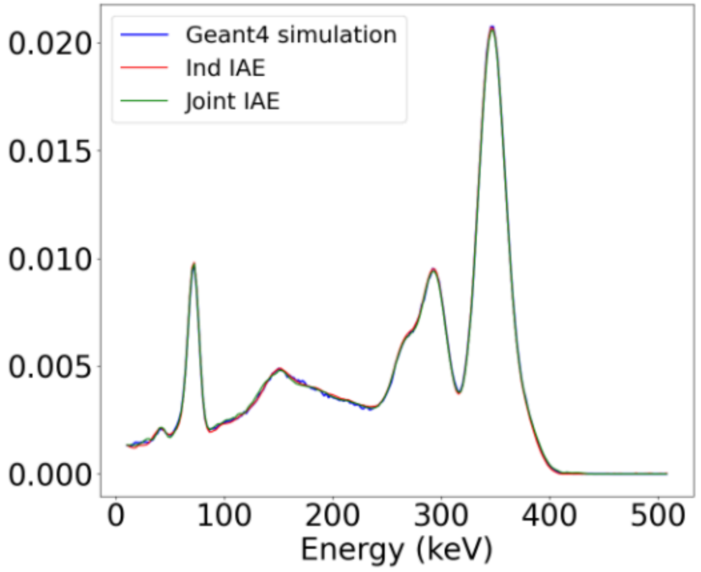
Set-up :

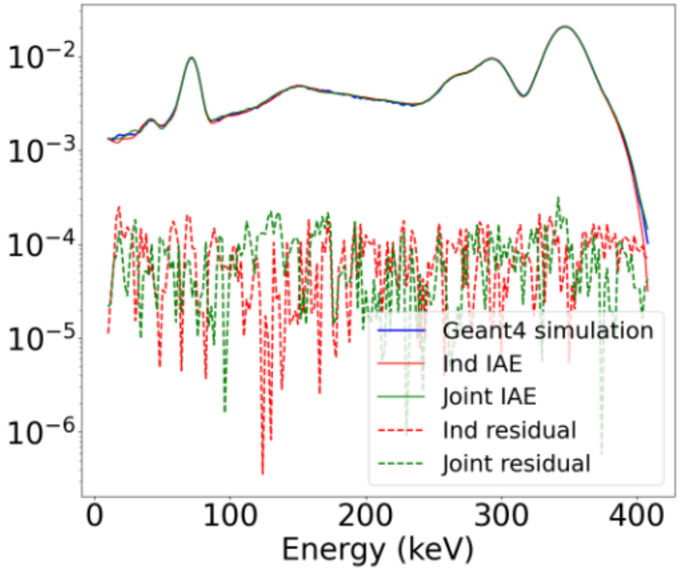
- Radioactive source in a lead sphere
- #Geant 4 simulations: 90
- 2 anchorpoints
- 4 radionuclides

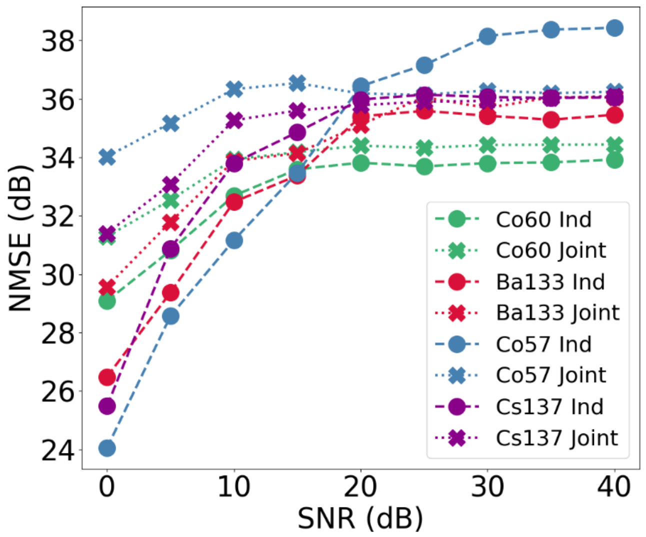


Variabilities as a function of the sphere thickness

Reconstructed ^{133}Ba







Unmixing with a plug-and-play approach

- Hybrid approach: combination with standard statistical inference

  Allows to account for the exact mixture model

  Built on the measurement statistics

Unmixing with a plug-and-play approach

- Hybrid approach: combination with standard statistical inference

  Allows to account for the exact mixture model

  Built on the measurement statistics

Positivity of the weights

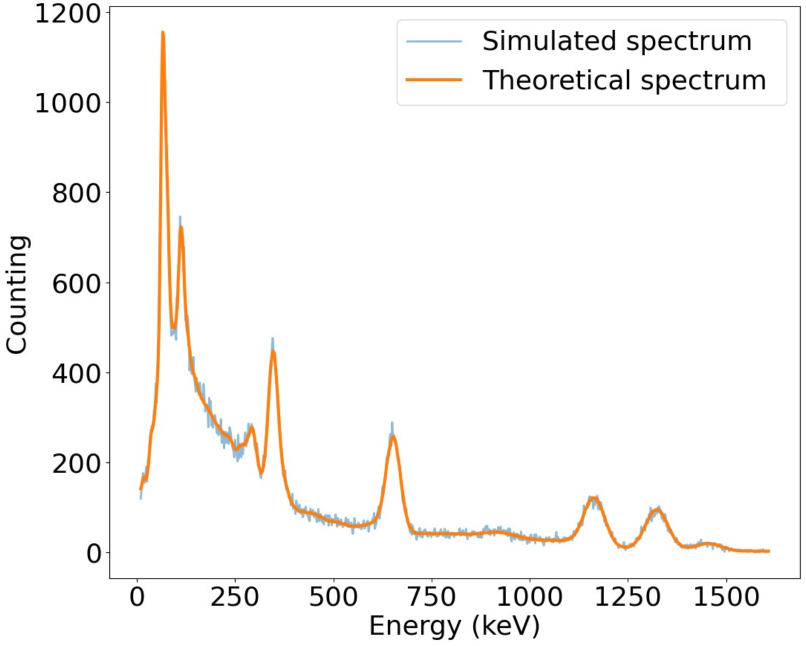
$$\min_{\{\Lambda_i\},\{a_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i a_i \Psi_i(\Lambda_i)\right) + \sum_i \chi_{\geq 0}(a_i)$$

Poisson neg-loglikelihood *Latent space parameter*

- SEMSUN algorithm: block-coordinate descent *(Phan et al, 23)*

Results - high statistics case

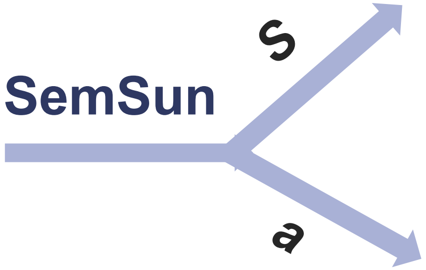
Theoretical signatures: 0.5mm thickness
1000 Monte Carlo simulations

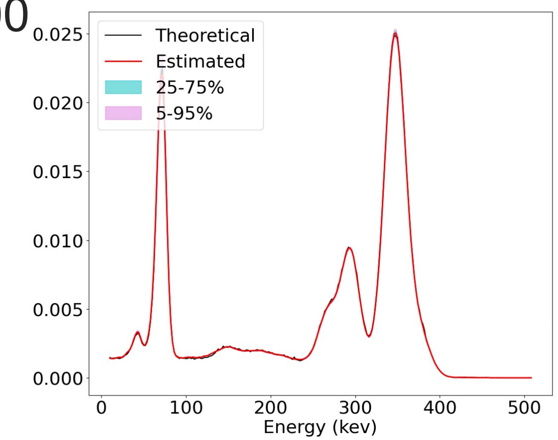


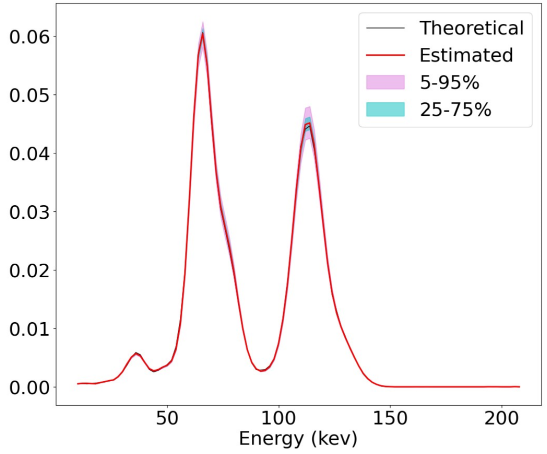
Simulated spectrum from a Poisson distribution

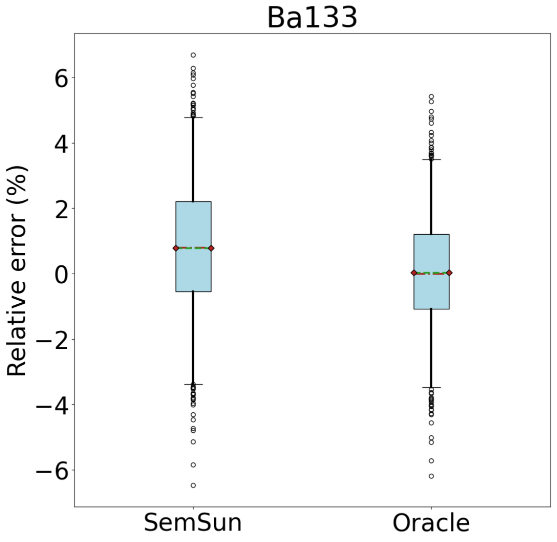
Mixture:
- Bkg : 50000
- ^{60}Co : 20000
- ^{133}Ba : 12000
- ^{57}Co : 8000
- ^{137}Cs : 10000

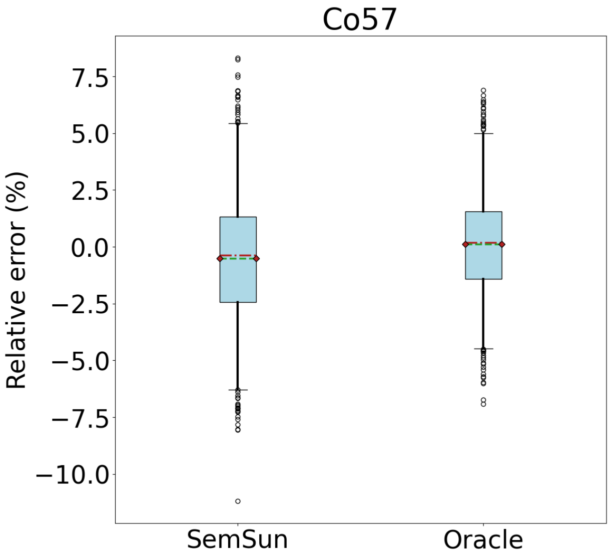
Oracle: estimate ***a*** when ***S*** is known, the best possible result











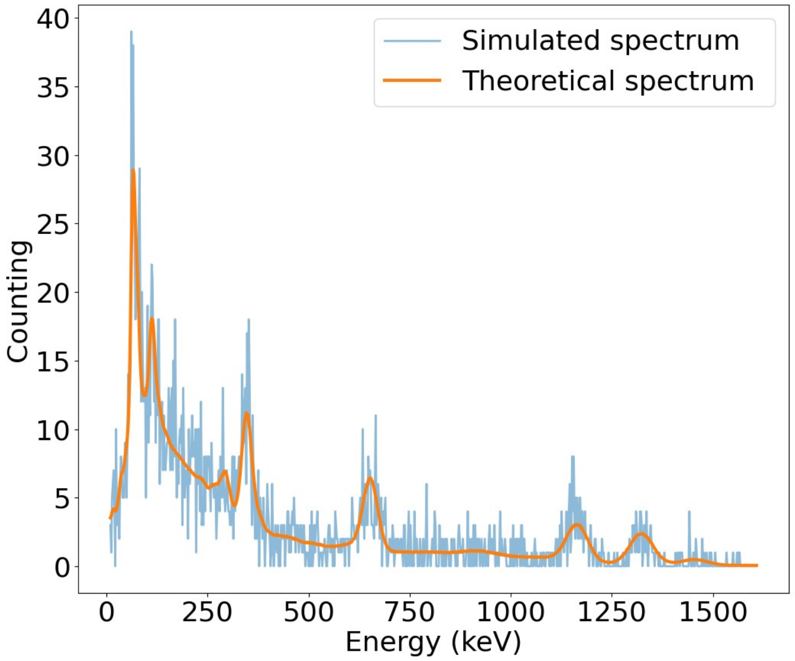
Results - low statistics case

Mixture:

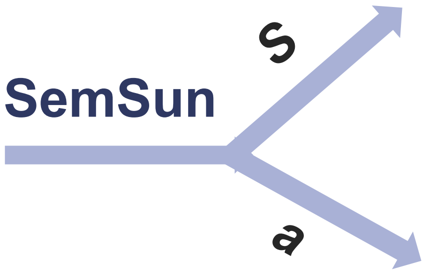
- Bkg : 1250
- ^{60}Co : 500
- ^{133}Ba : 300
- ^{57}Co : 200
- ^{137}Cs : 250

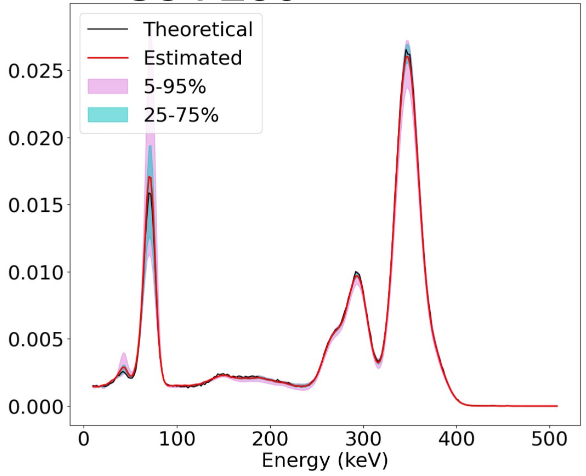
Theoretical signatures: 0.5mm thickness
1000 Monte Carlo simulations

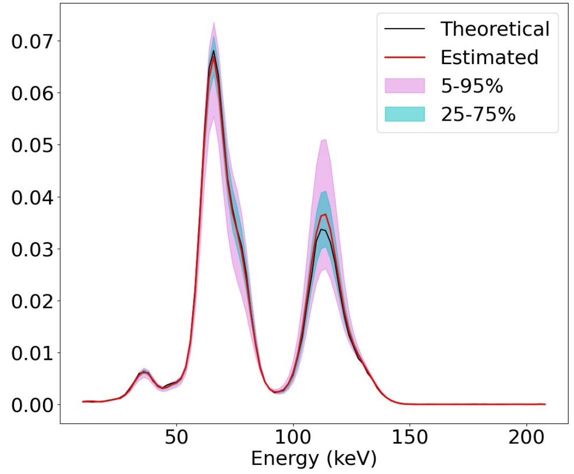
Oracle: estimate ***a*** when ***S*** is known, the best possible result

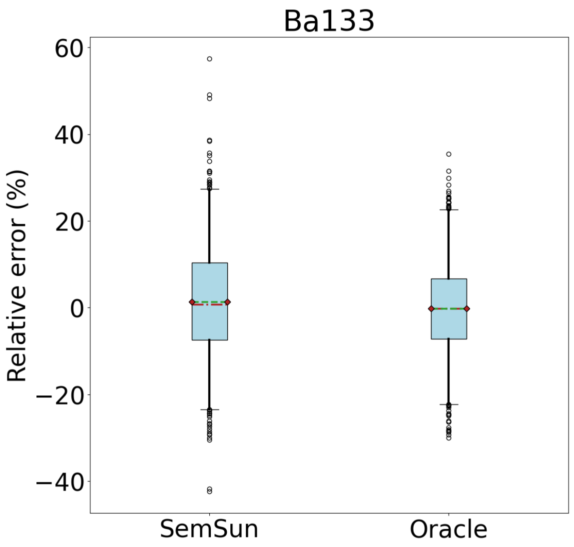


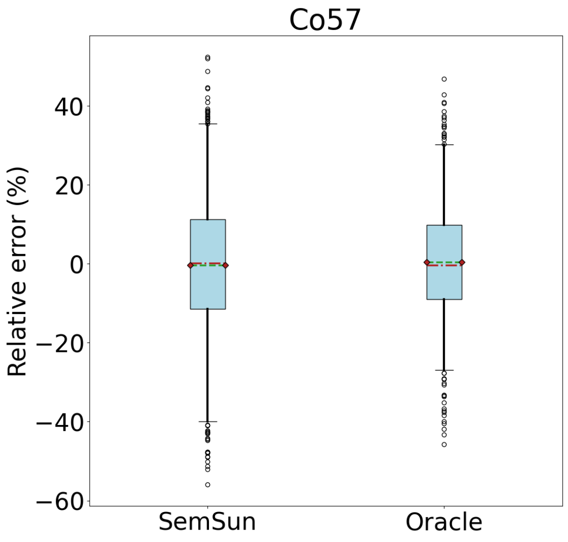
Simulated spectrum from a Poisson distribution







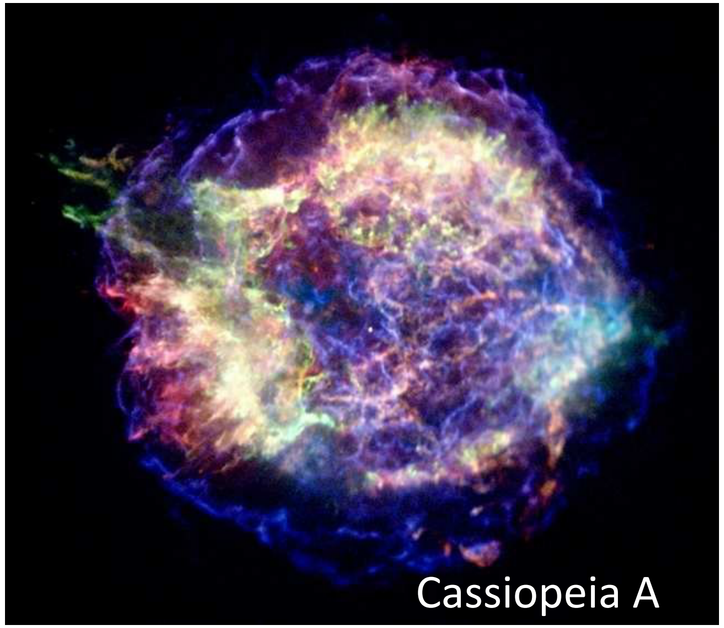


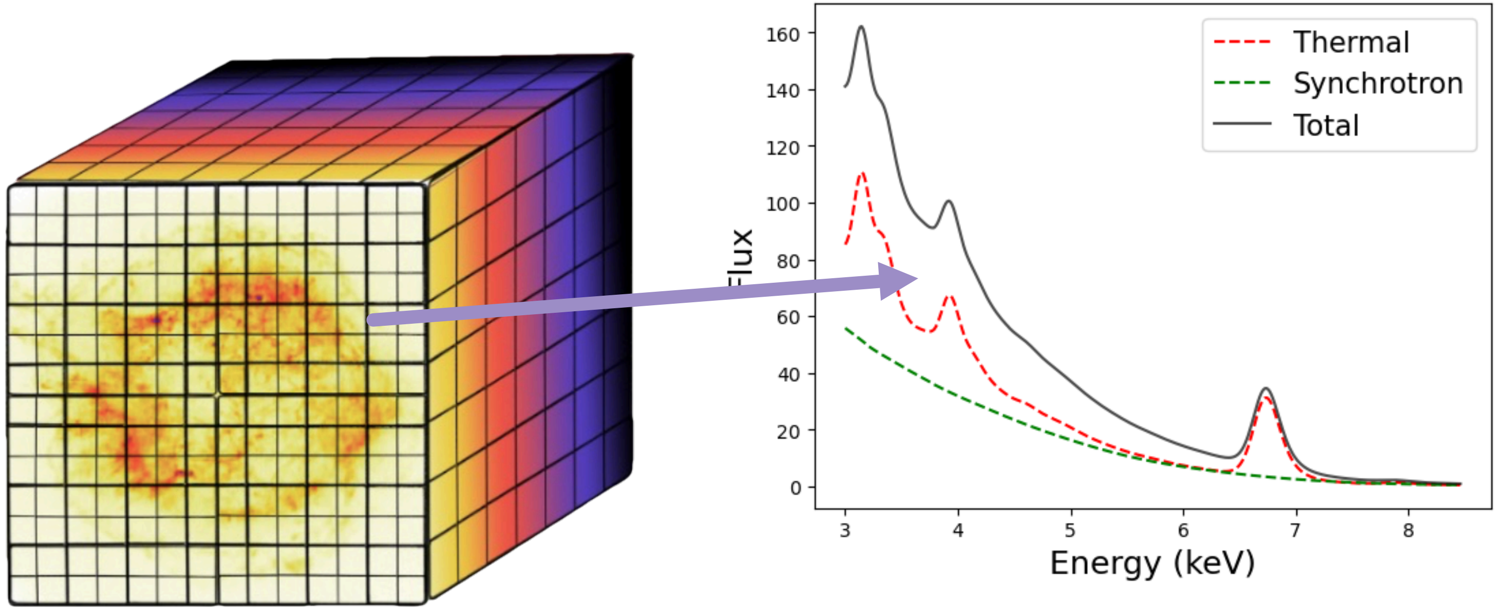


Unmixing X-ray images in astrophysics

Case Study: Supernova Remnants in X-ray multispectral data

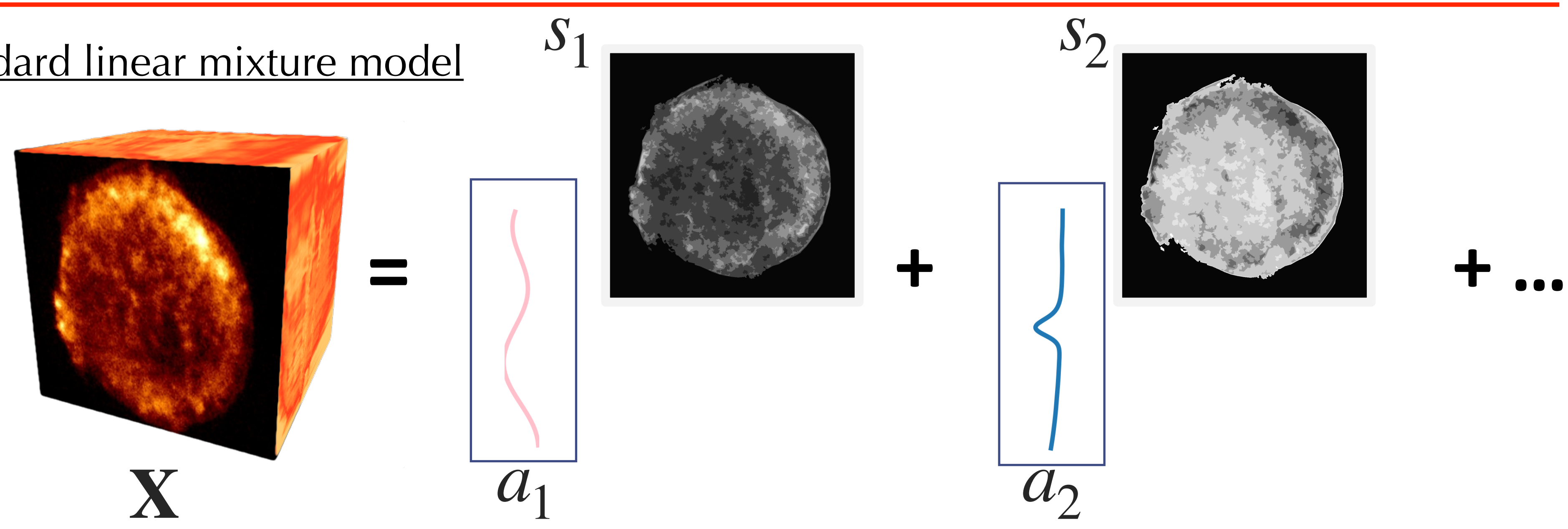
- ► Poisson noise, low signal/noise
- ► Entangled physical components
- ► Variabilities described by non-analytical models





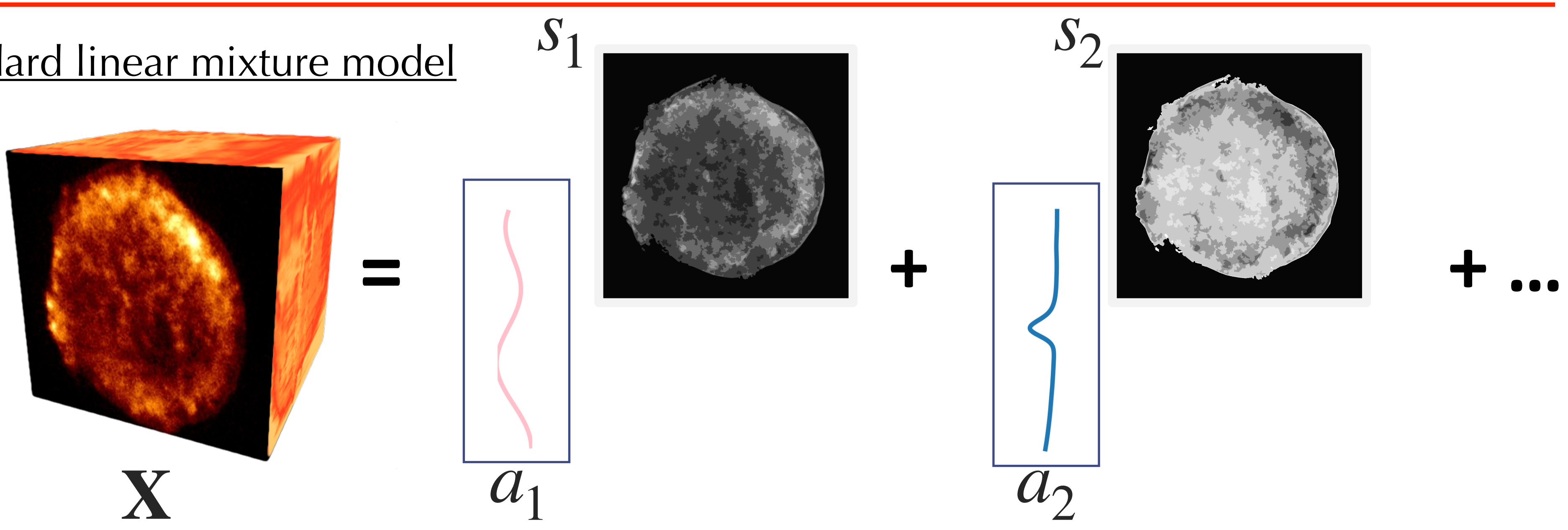
A different mixture model

Standard linear mixture model

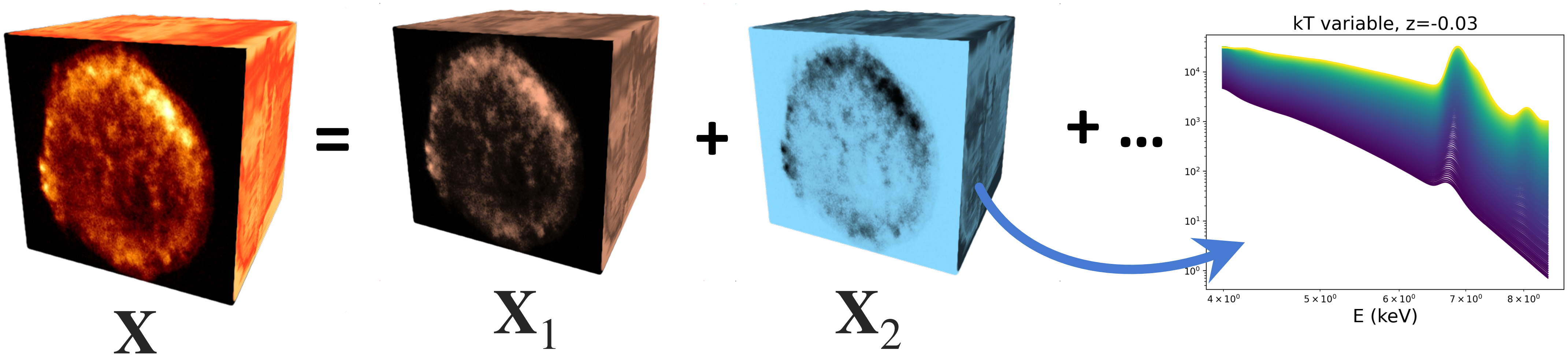


A different mixture model

Standard linear mixture model

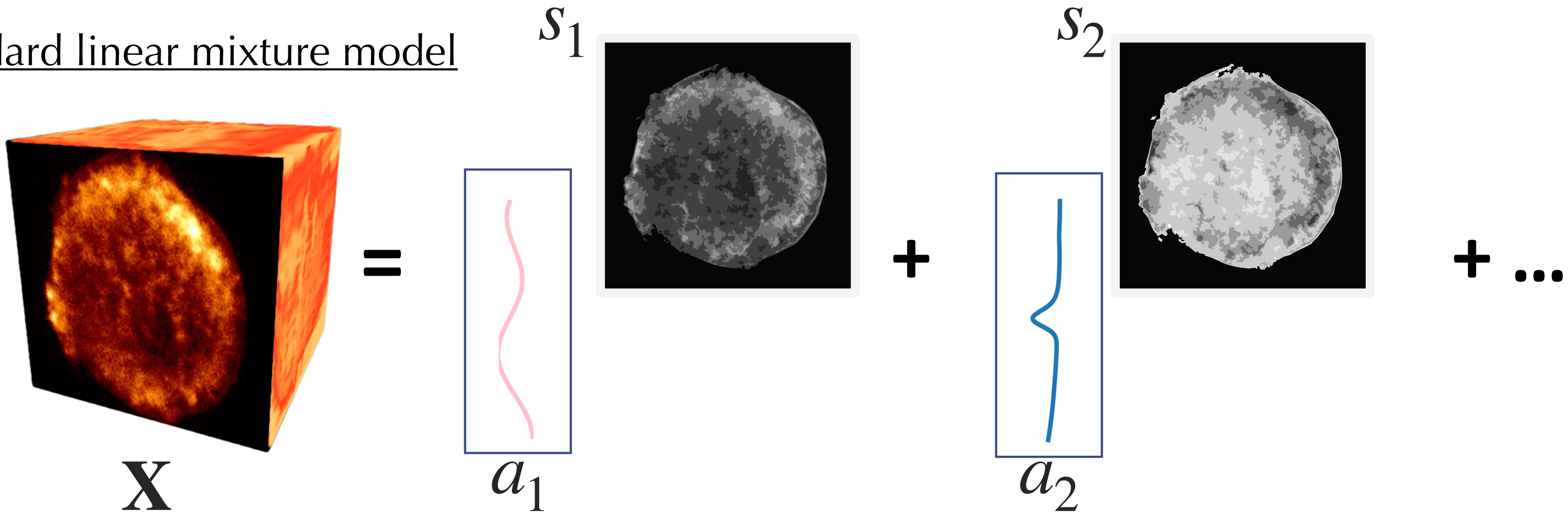


Non-stationary linear mixture model

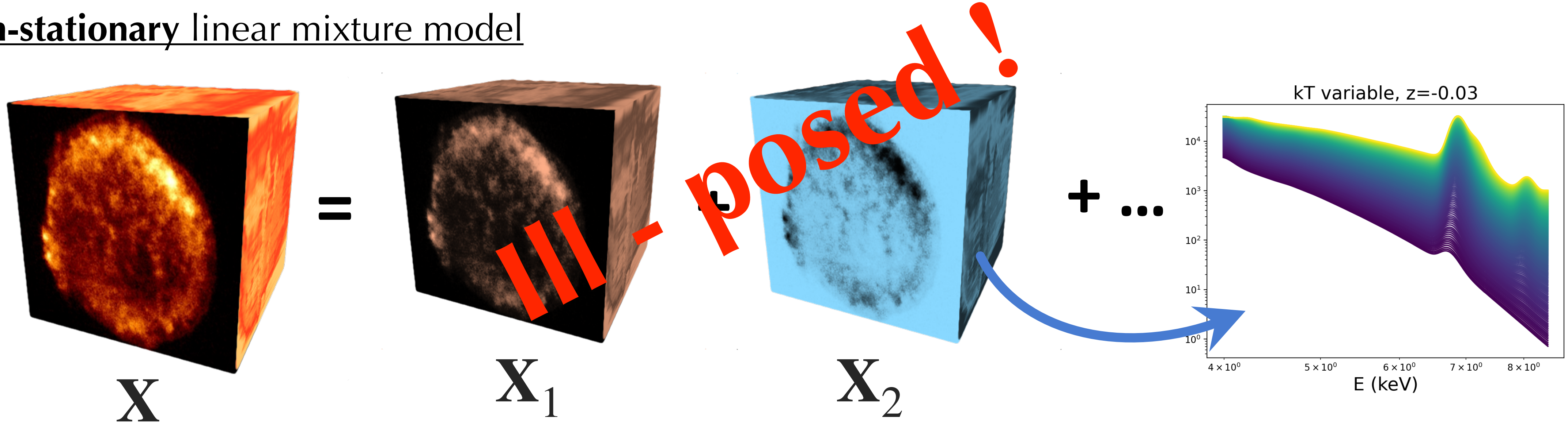


A different mixture model

Standard linear mixture model



Non-stationary linear mixture model



Non-stationary mixture model

Non-stationary mixture model *(noiseless)*

$$\mathbf{X} = \sum_i \mathbf{A_i} \odot s_i$$

Amplitude

Spectral cube

Non-stationary mixture model

Non-stationary mixture model *(noiseless)*

Amplitude

$$\mathbf{X} = \sum_i \mathbf{A_i} \odot s_i$$

Spectral cube

Spectral parametric models exist for the spectra but

Costly … to be plugged into unmixing algorithms

Non-differentiable … cannot be plugged into unmixing algorithms

Non-stationary mixture model

Non-stationary mixture model *(noiseless)*

$$\mathbf{X} = \sum_i \mathbf{A_i} \odot s_i$$

Amplitude (s_i)

Spectral cube ($\mathbf{A_i}$)

Spectral parametric models exist for the spectra but

Costly … to be plugged into unmixing algorithms

Non-differentiable … cannot be plugged into unmixing algorithms

AE-based surrogates are not costly *(at inference time)* and differentiable

They are good candidates for hybrid unmixing solvers

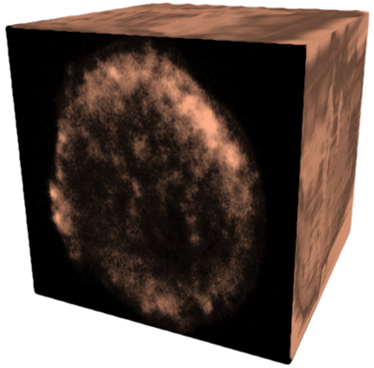
More formally - spectral regularisation

$$\min_{\{\mathbf{A}_i\},\{s_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i \mathbf{A}_i \odot s_i\right)$$

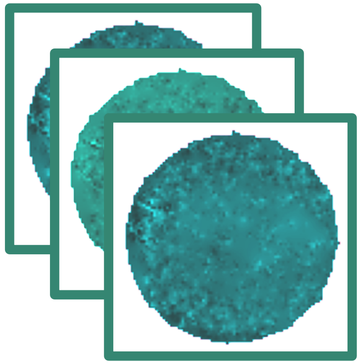
More formally - spectral regularisation

$$\min_{\{\mathbf{A}_i\},\{s_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i \mathbf{A}_i \odot s_i\right)$$

- The spectra can be described by an AE-based model



$$= \Psi_i(\quad) \odot s_i$$

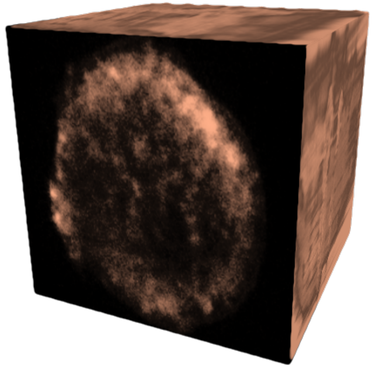


Latent space

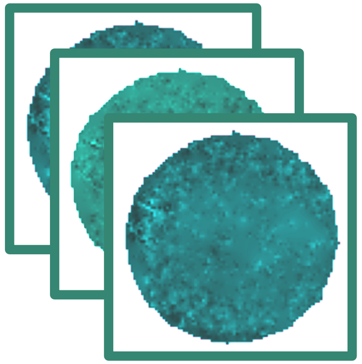
More formally - spectral regularisation

$$\min_{\{\mathbf{A}_i\},\{s_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i \mathbf{A}_i \odot s_i\right)$$

- The spectra can be described by an AE-based model



$$= \Psi_i\Big($$



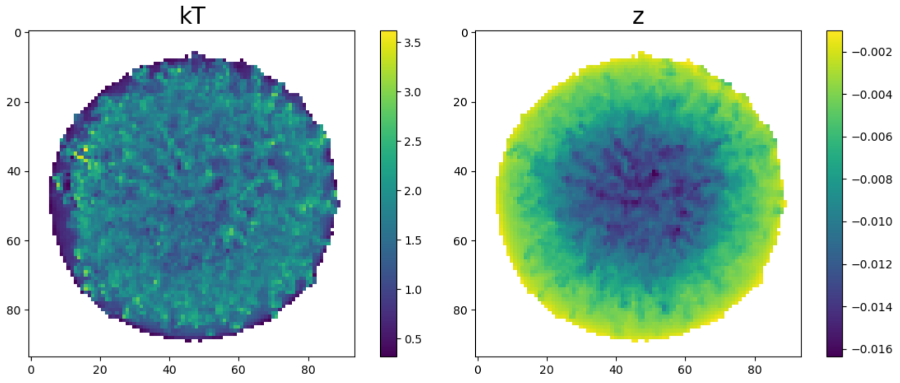
$$\Big) \odot s_i$$

Latent space

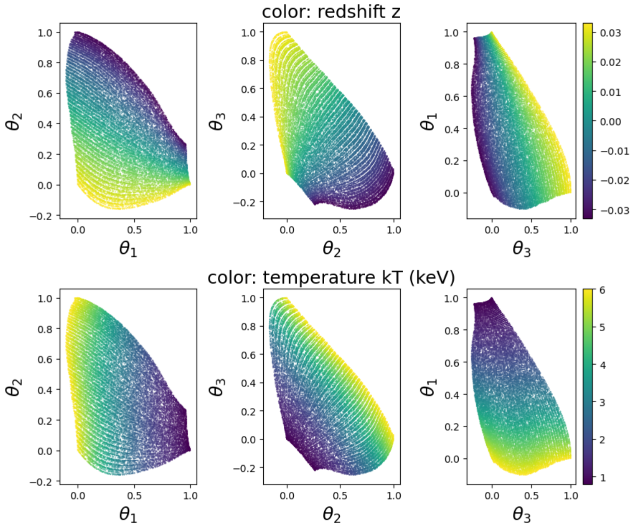
$$\min_{\{\boldsymbol{\Lambda}_i\},\{s_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i \Psi_i(\boldsymbol{\Lambda}_i) \odot s_i\right)$$

More formally - spatial regularisation

- The spectra evolve smoothly across the sky



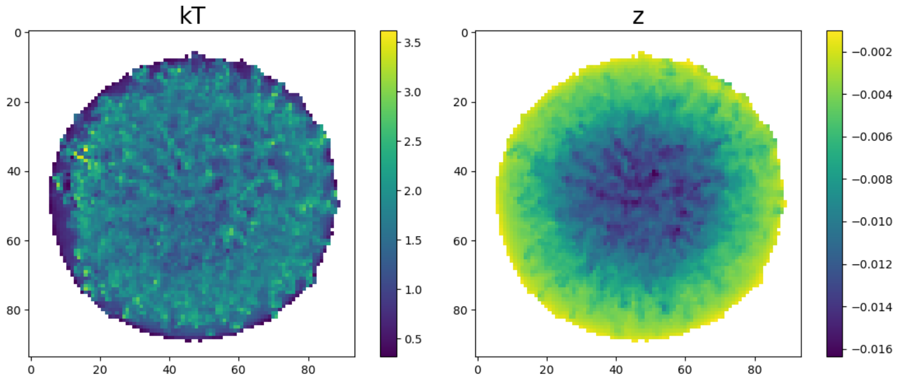
Physical parameters



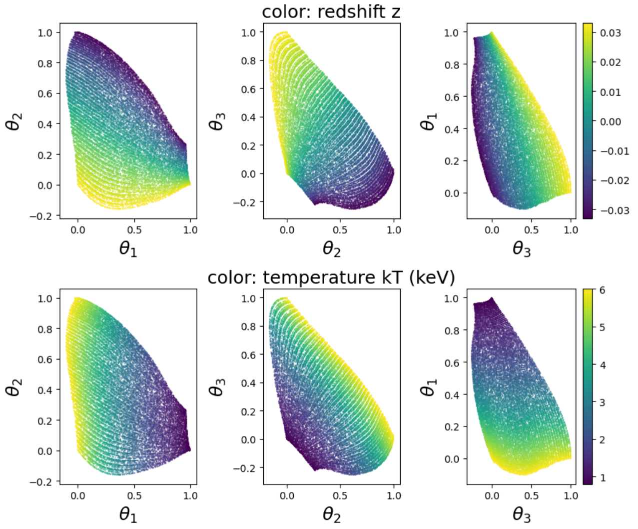
Latent space

More formally - spatial regularisation

- The spectra evolve smoothly across the sky



Physical parameters



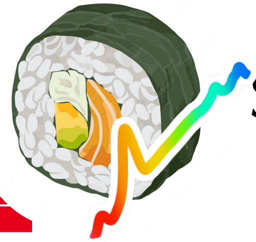
Latent space

Positivity of the amplitude

$$\min_{\{\boldsymbol{\Lambda}_i\},\{s_i\}_i} \mathscr{L}\left(\mathbf{X}, \sum_i \Psi_i(\boldsymbol{\Lambda}_i) \odot s_i\right) + \sum_i \chi_{\geq 0}(s_i) + \mu \|\mathbf{W}\boldsymbol{\Lambda}_i\|_{\ell_1}$$

Sparsity-enforcing regularisation in the domain W

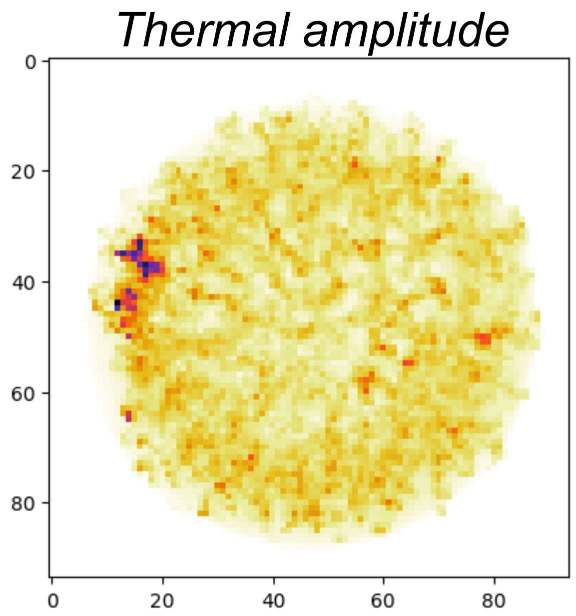
(e.g. wavelets, etc)

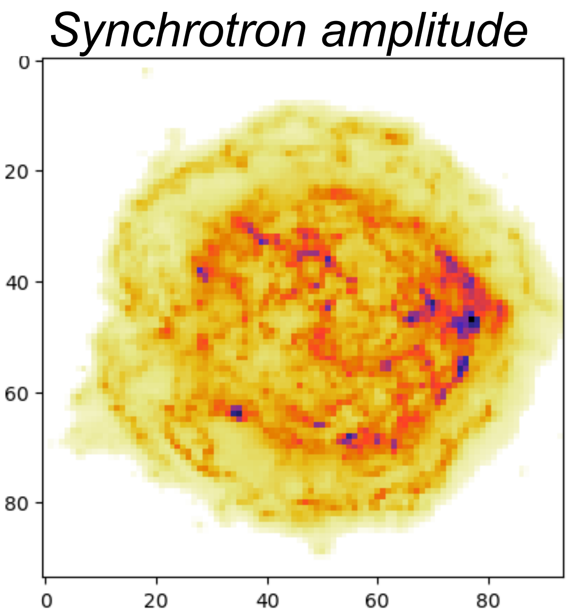


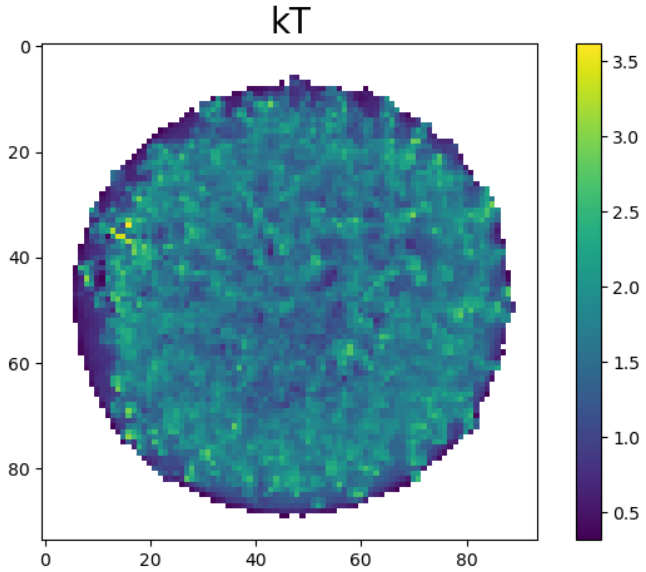
Sushi: PALM-based solver *(Lascar et al, 23)*

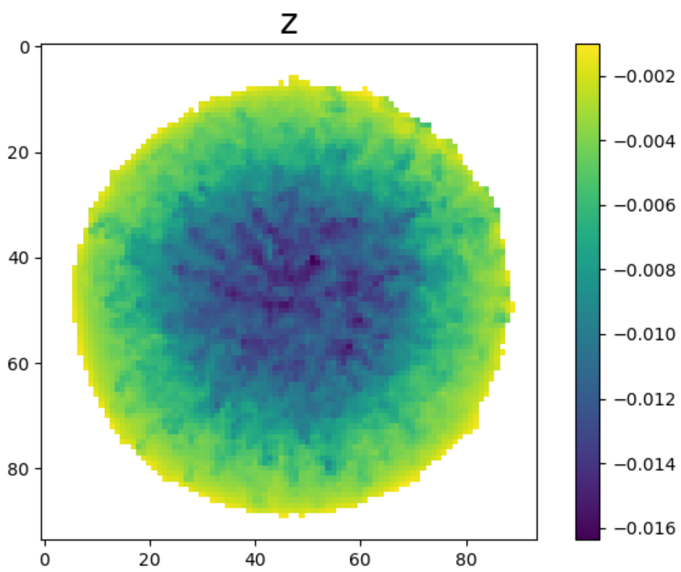
Results on synthetic data

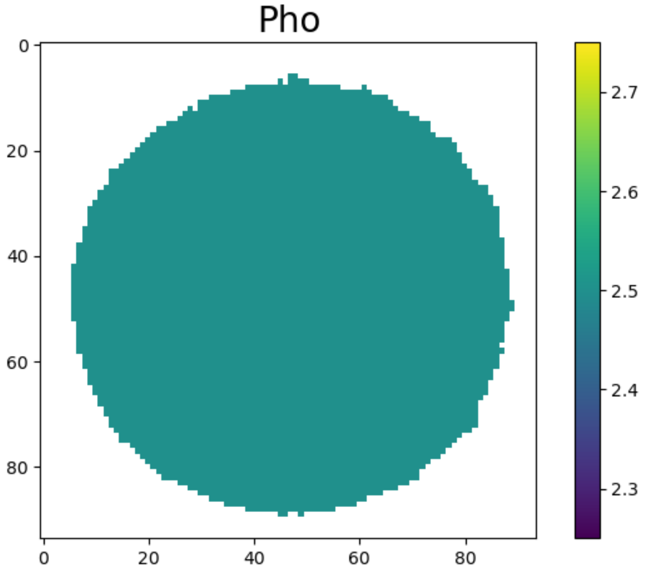
- ► From real images + numerical simulations of CasA
- ►Thermal Component: Varying redshift, temperature
- ►Synchrotron Component: Constant Photon Index
- ►#Simulated spectra ~400
- ►3 anchorpoints





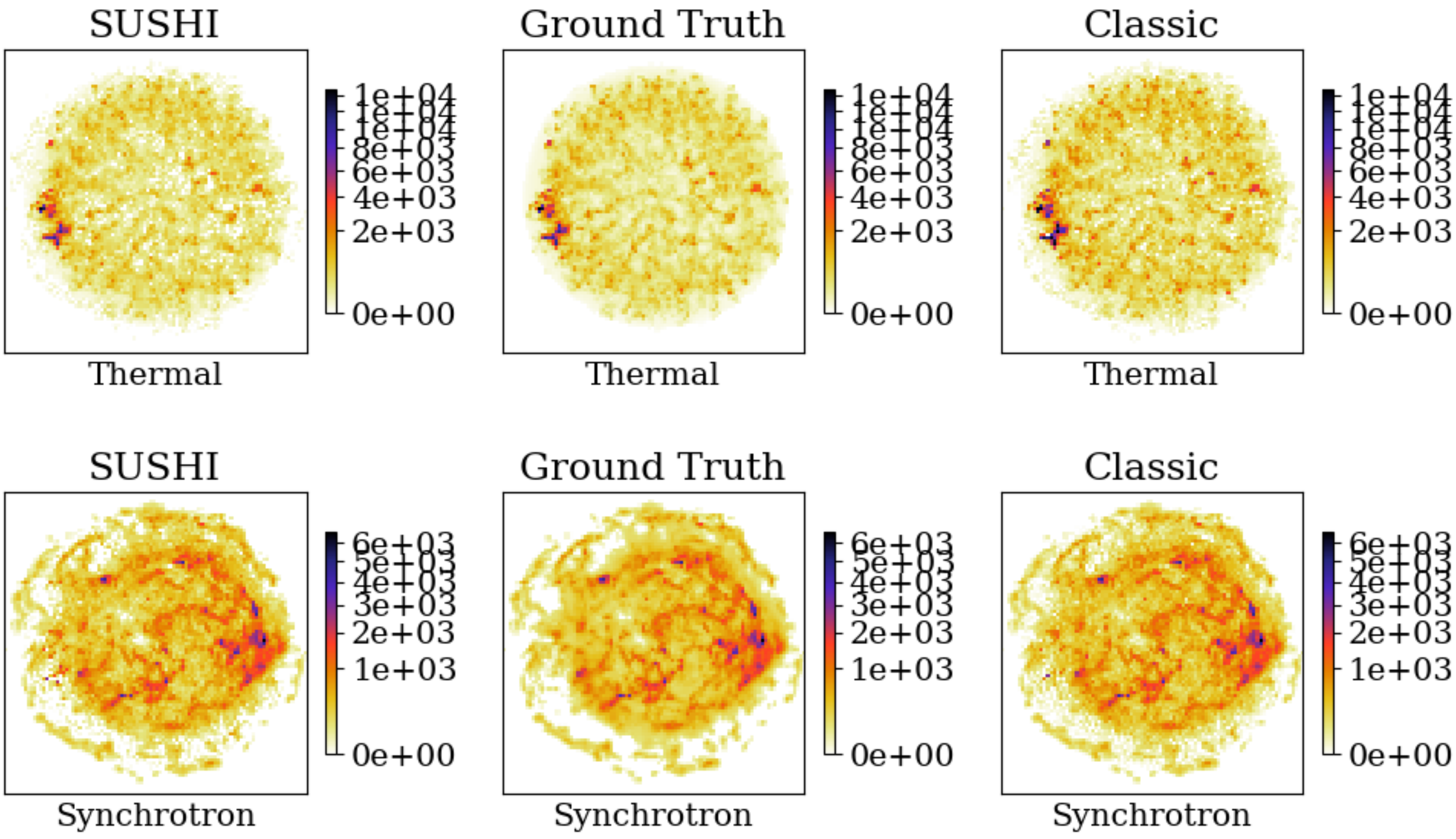




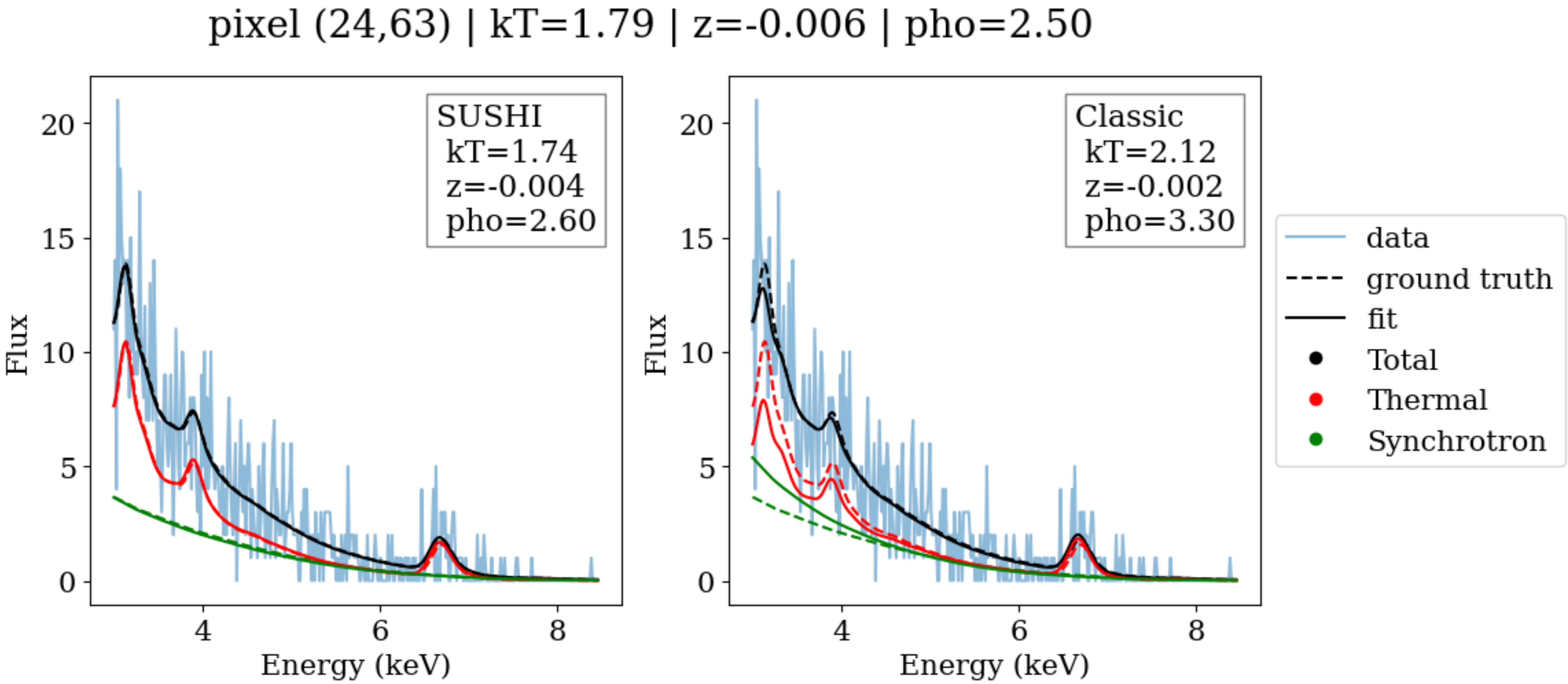


Estimated amplitude map

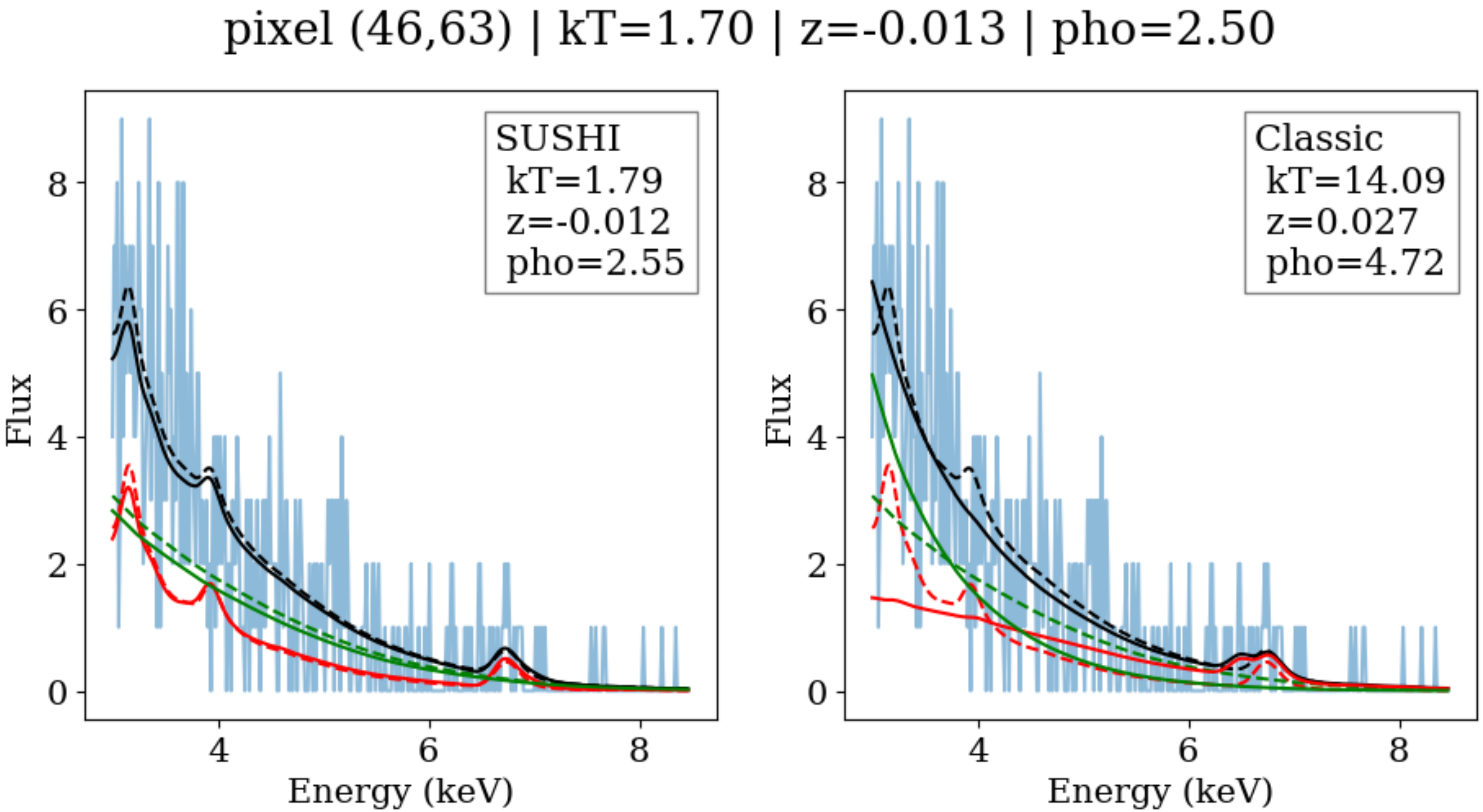
Fit 1D pixel-per-pixel



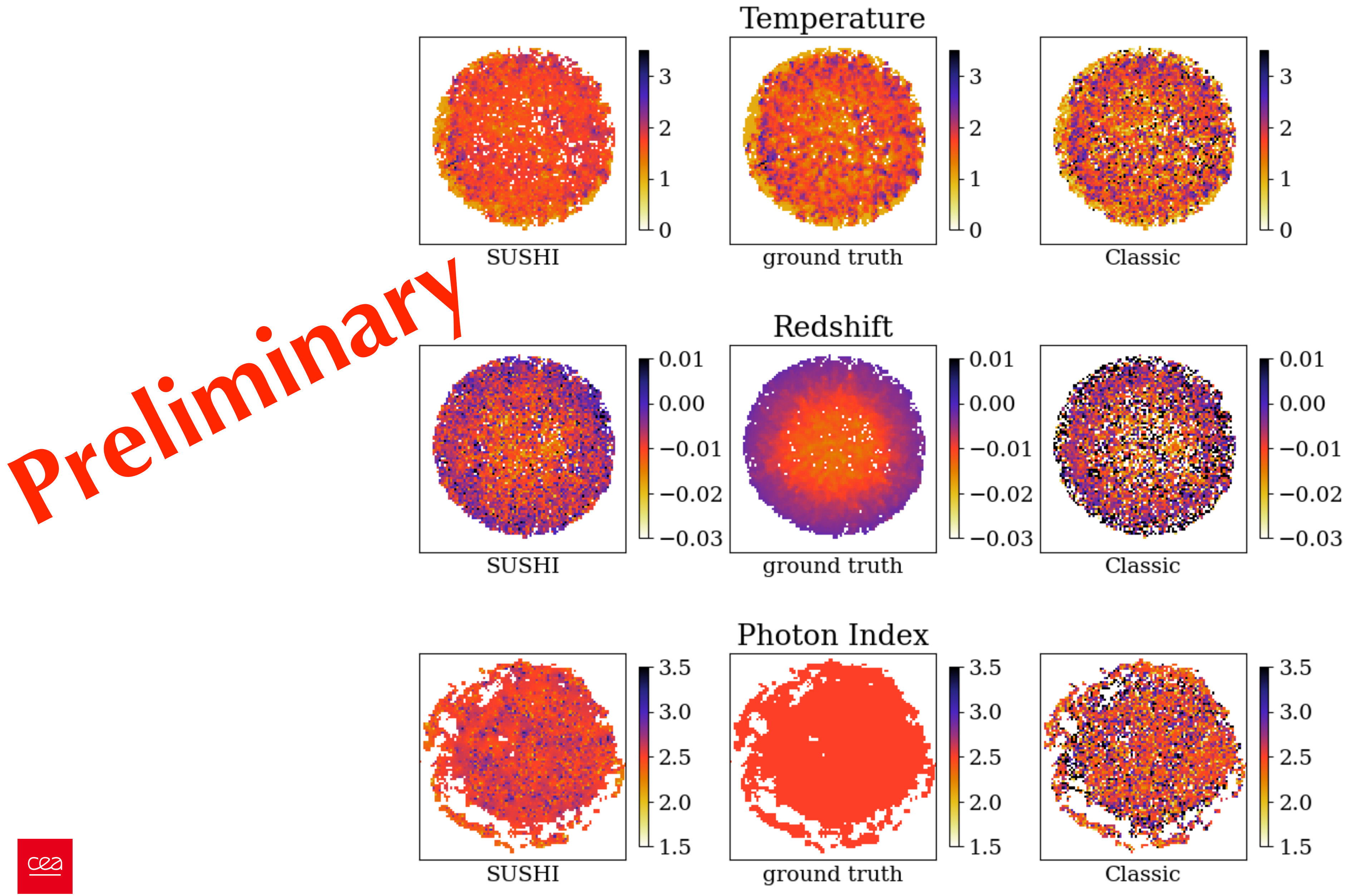
Results on synthetic data



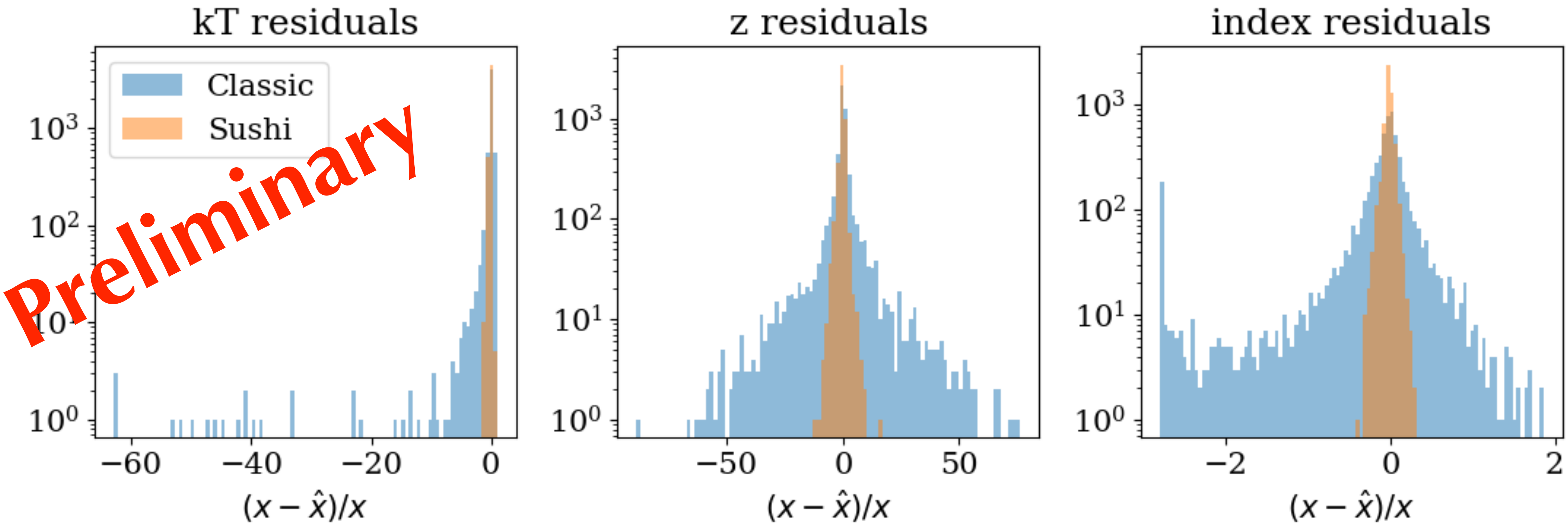
Results on synthetic data



Estimated physical parameters

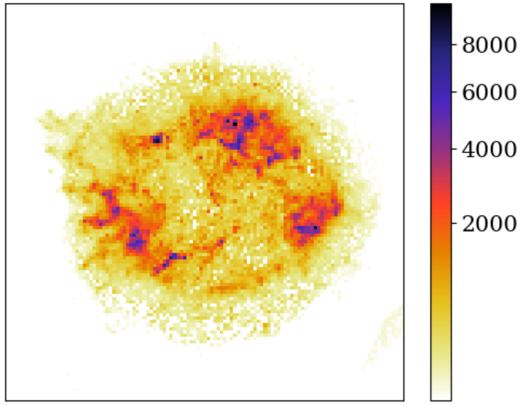


Estimated physical parameters

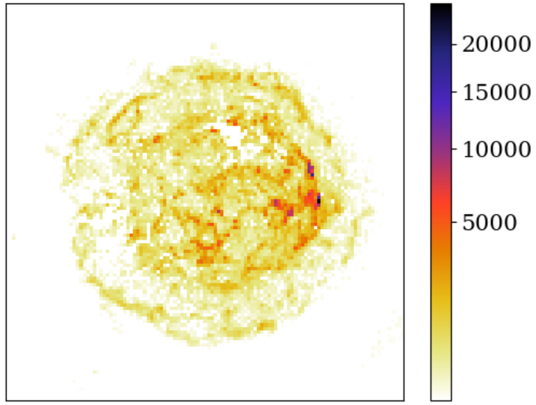


Results from real data - preliminary results !

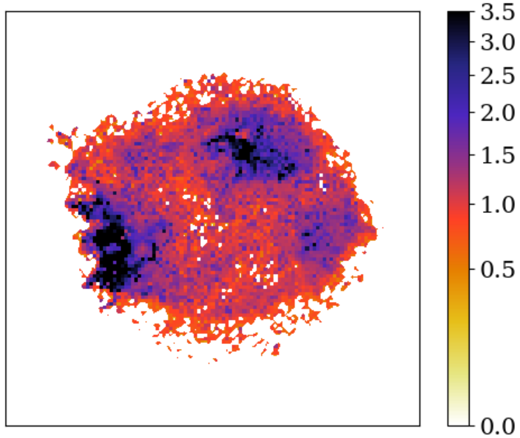
Preliminary



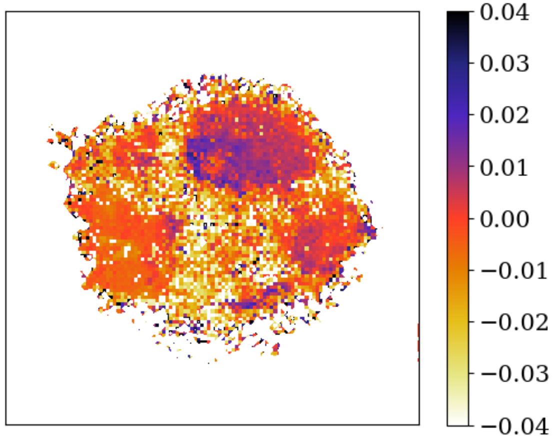
Thermal amplitude



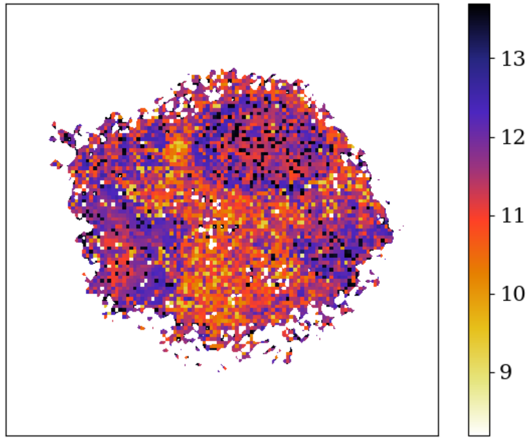
Synchrotron amplitude



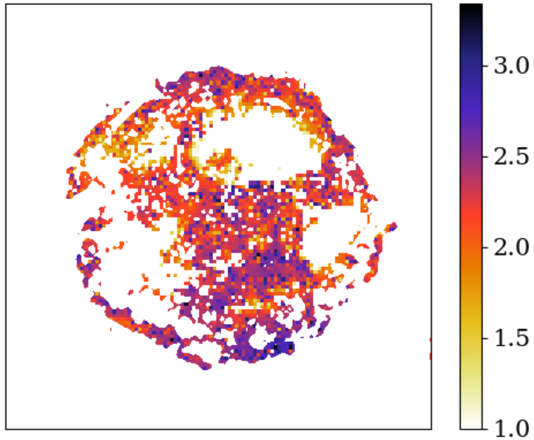
Thermal index



Thermal redshift



Thermal ionisation



Synchrotron index

Take-away messages

- IAE: a flexible model to learn representations when training samples are scarce.
- Deployable as surrogates in standard solvers to tackle complex/ill-posed unmixing problems.
- Unmixing is costly but can be accelerated using deep unrolling *(Fahes22)*
- Quantifying uncertainties is key but complex; under investigation !

https://github.com/jbobin/IAE

https://github.com/JMLascar/SUSHI

SEMSUN to come soon

Back-up slides

Unmixing with a plug-and-play approach

Unmixing with a plug-and-play approach

Joint estimation of ***X*** and ***a***

$$\hat{X}, \hat{a} = Argmin_{X,a} \sum_{i=2}^{p} c_i(X_i) + \chi_{(.\geq 0)}(a) + L(a, X) \quad (3)$$

- Constraints for each radionuclide i : $c_i(X_i)$
- The spectral signature is the decoding of the latent variable of IAE $X_i = g_i(\lambda_i)$

Complex problem, non-convex, multiple local minima.

Unmixing with a plug-and-play approach

Joint estimation of ***X*** and ***a***

$$\hat{X}, \hat{a} = Argmin_{X,a} \sum_{i=2}^{p} c_i(X_i) + \chi_{(.\geq 0)}(a) + L(a, X) \quad (3)$$

- Constraints for each radionuclide i : $c_i(X_i)$
- The spectral signature is the decoding of the latent variable of IAE $X_i = g_i(\lambda_i)$

Complex problem, non-convex, multiple local minima.

Block coordinate descent (BCD) (Y.XU, 2017)

X is fixed, Estimate *a*

1. Update a: $a^{k+1} = Argmin_a \chi(. \geq 0)(a) + L(a, X^k)$

Multiplicative update algo NNPU

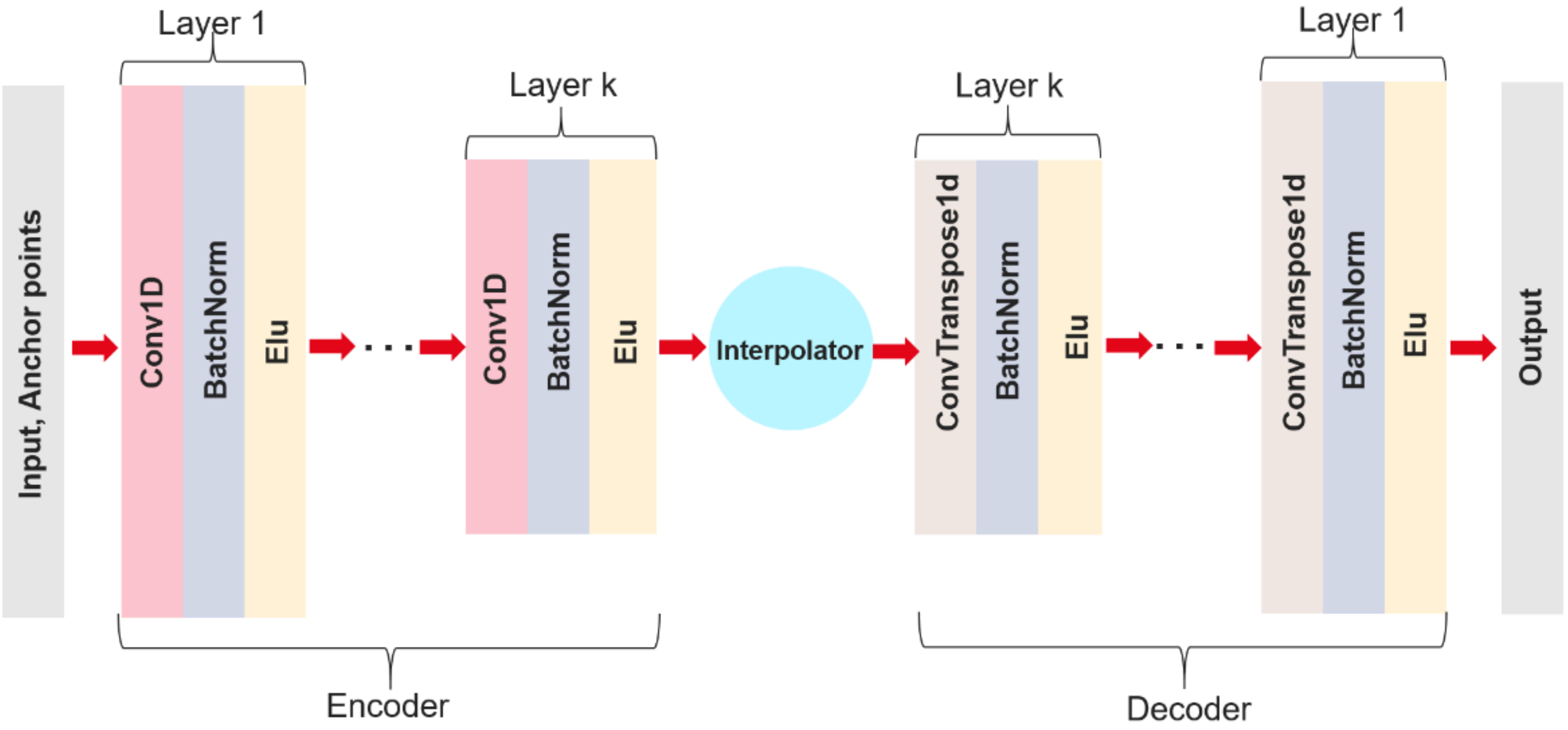
a is fixed, Estimate *X*

2. Update X: $X^{k+1} = Argmin_X \sum_{i=2}^{p} c_i(X_i) + L(a^{k+1}, X)$

Sequential Least Squares Programming (SLSQP) (D. Kraft, 1988)

SEMSUN - network description

- CNN-based networks



SEMSUN - network description

Hyperparameters	Co60	Ba133	Co57	Cs137	Joint
Maximum channel	800	250	100	400	800
Solver	Adam	Adam	Adam	Adam	Adam
Learning rate	0.001	0.001	0.001	0.001	0.001
Batch size	36	36	36	36	36
Number of epochs	20000	20000	20000	20000	20000
Regulisation paramater	0.001	0.001	0.001	0.001	0.001
Encoder: numbers of layers	6	6	6	6	6
Activation	Elu(alpha=1)	Elu(alpha=1)	Elu(alpha=1)	Elu(alpha=1)	Elu(alpha=1)
Encoder 1 : Conv1D (in_channels, out_channels, kernel_size, stride)	1, 12, 4, 1	1, 12, 4, 1	1, 12, 4, 1	1, 12, 4, 1	4, 12, 4, 1
Encoder 2 : Conv1D	12, 12, 4, 1	12, 12, 4, 1	12, 12, 4, 1	12, 12, 4, 1	12, 12, 4, 1
Encoder 3 : Conv1D	12, 12, 6, 2	12, 12, 6, 2	12, 12, 3, 1	12, 12, 6, 2	12, 12, 6, 2
Encoder 4 : Conv1D	12, 16, 6, 2	12, 16, 6, 2	12, 16, 3, 1	12, 16, 6, 2	12, 16, 6, 2
Encoder 5 : Conv1D	16, 16, 6, 2	16, 16, 6, 2	16, 16, 3, 1	16, 16, 6, 2	16, 16, 6, 2
Encoder 6 : Conv1D	16, 16, 4, 2	16, 16, 4, 2	16, 16, 3, 1	16, 16, 4, 2	16, 16, 4, 2
cost function	log	log	log	log	mean log of each radionuclide

Sushi - network

- Dense networks

	Thermal (toy model)	Thermal (Cassopeia A data)	Synchrotron (toy model)	Synchrotron (Cassopeia A data)	Synchrotron (Crab data)
Physical model	Equilibrium collisional ionized plasma emission (APEC)	Non-equilibrium collisional ionized plasma emission	Power Law		
Number of anchor points	4	6	2	2	2
Number of layers	4	4	4	2	2
Step size	6×10^{-4}	4×10^{-4}	8×10^{-4}	10^{-3}	10^{-3}
Optimizer	Adaptive Gradient Algorithm (Adagrad)				
Activation function	Leaky Rectified Linear Activation (LReLU)				

Sushi - algorithm

Algorithm 1 SUSHI: Semi-blind Unmixing with Sparsity for Hyperspectral Images

---

input data X, trained IAE models $\{\mathcal{M}^0, ..., \mathcal{M}^{n_C}\}$, number of wavelet scales J, sparsity threshold factor k, cost function $\mathcal{L}$.

initialisation $\{\theta_0^0, ..., \theta_0^{n_C}\} \leftarrow \{\nVdash/N_A^0, ..., \nVdash/N_A^{n_C}\}$

$\{A_0^0, ..., A_0^{n_C}\} \leftarrow \sum_e^{n_E} X(., e)/n_C$

$\alpha_\theta \leftarrow 0.1/\max(A_0^0)$

$t \leftarrow 0$

while stopping criterion is not met **do**

  for component c in $\{0, ..., n_C\}$ **do**

    Gradient descent step on θ^c

    $\theta_{t+1/2}^c \leftarrow \theta_t^c - \alpha_\theta \nabla_{\theta^c} \mathcal{L}(\theta^c | X, A^c, \theta^{C \neq c})$

    Sparsity step on θ^c

    $\theta_{t+1}^c \leftarrow \mathbf{prox}_{l_1, J, k}(\theta_{t+1/2}^c)$

    Gradient Descent step on A^c

    $H \leftarrow \nabla_{A^c}^2 (\mathcal{L}(A^c | X, \theta_{t+1}^c))$

    $A_{t+1}^c \leftarrow A_t^c - 1/H \nabla_{A^c} \mathcal{L}(A^c | X, \theta_{t+1}^c)$

  end for

  $t \leftarrow t + 1$

end while

$\hat{X}^c \leftarrow A_t^c \mathcal{M}^c(\theta_t^c)$

$\hat{X} \leftarrow \sum_{c=0}^{n_C} \hat{X}^c$

return $\hat{X}, \{\hat{X}^0, ... \hat{X}^C\}$
