New methods for fast X-ray inference

Simon Dupourqué (IRAP, Toulouse)

Lumières Workshop (2026)

IFU

irap

astrophysique & planétologie

Bayesian or frequentist?

$$\hat{\theta} = \min_{\theta} \mathscr{L}(\theta, X)$$

$$P(\theta|X) = \frac{P(X|\theta)}{P(X)} P(\theta)$$

Bayesian or frequentist?

$$\hat{\theta} = \min_{\theta} \mathscr{L}(\theta, X)$$

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

This is a **philosophical** question, rather than a scientific one.

Bayesian or frequentist?

$$\hat{\theta} = \min_{\theta} \mathscr{L}(\theta, X)$$

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

This is a **philosophical** question, rather than a scientific one.

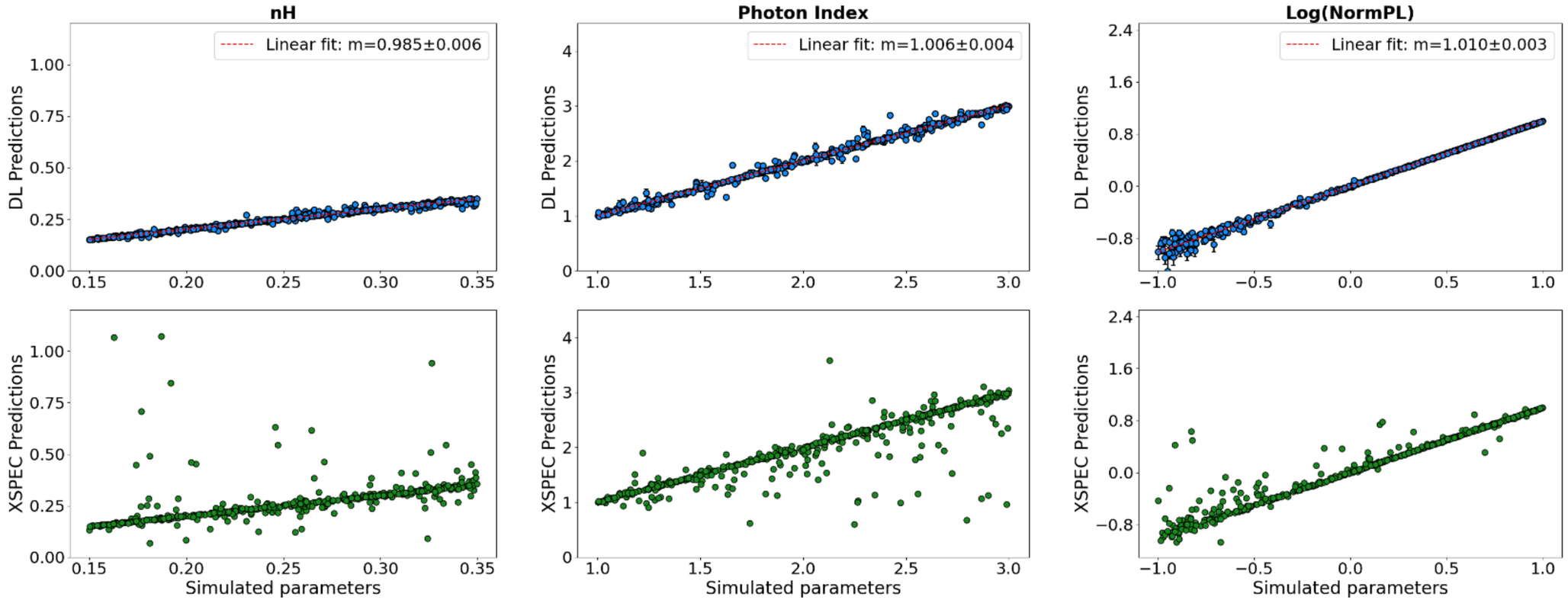
Some methods are more robust than others, and you can use them wether you're Bayesian or frequentist

Not the Frequentist vs Bayesian debate

Adapted from Tutone & al. (2025)

Local minima robust approach

XSPEC vanilla minimization scheme



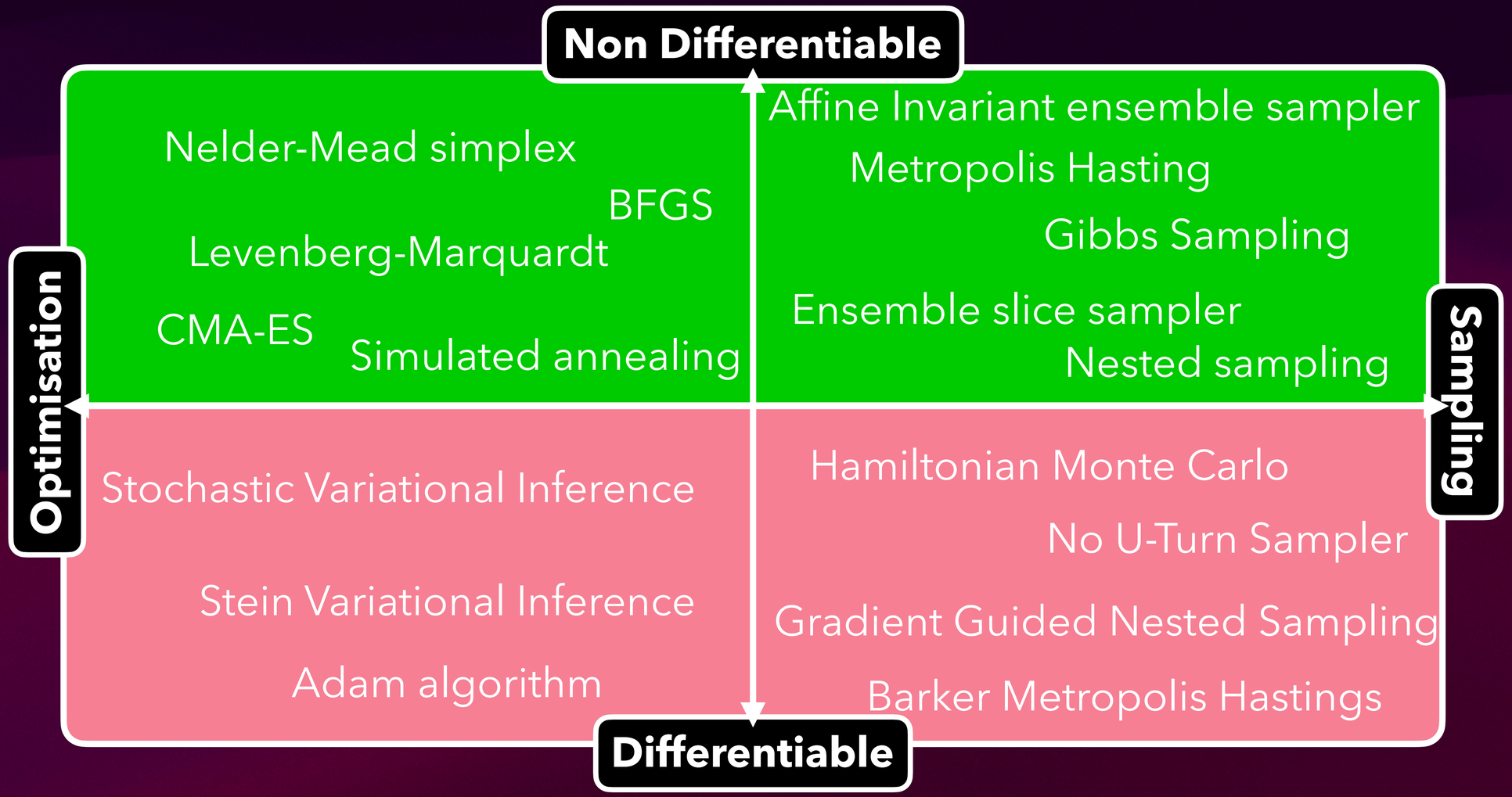
NICER spectrum simulated for an absorbed blackbody and powerlaw

Garbage IN ⟷ Garbage OUT on simulated datasets

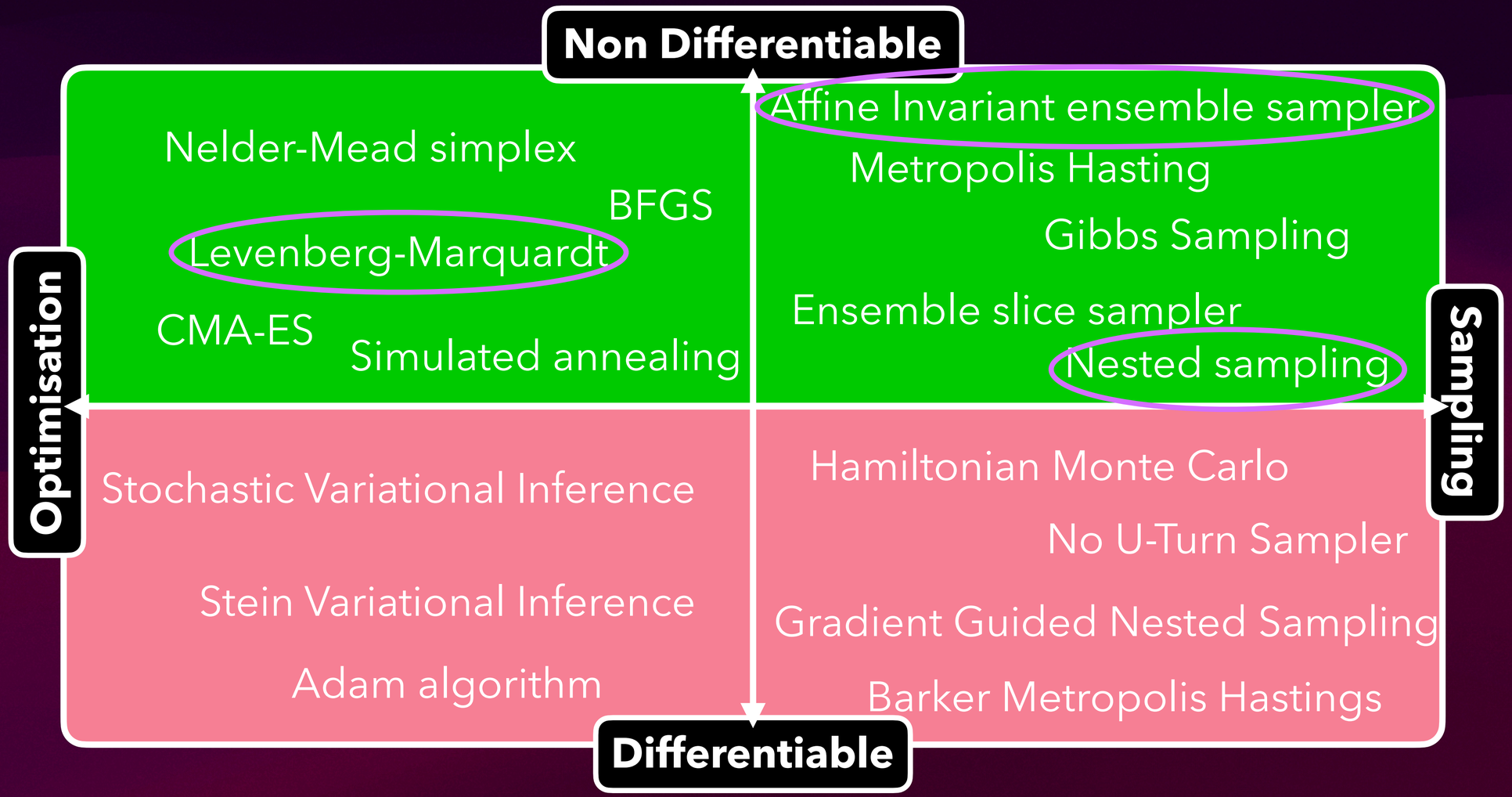
↔

At least 20% of local minima for a simple case

Landscape of exact inference methods

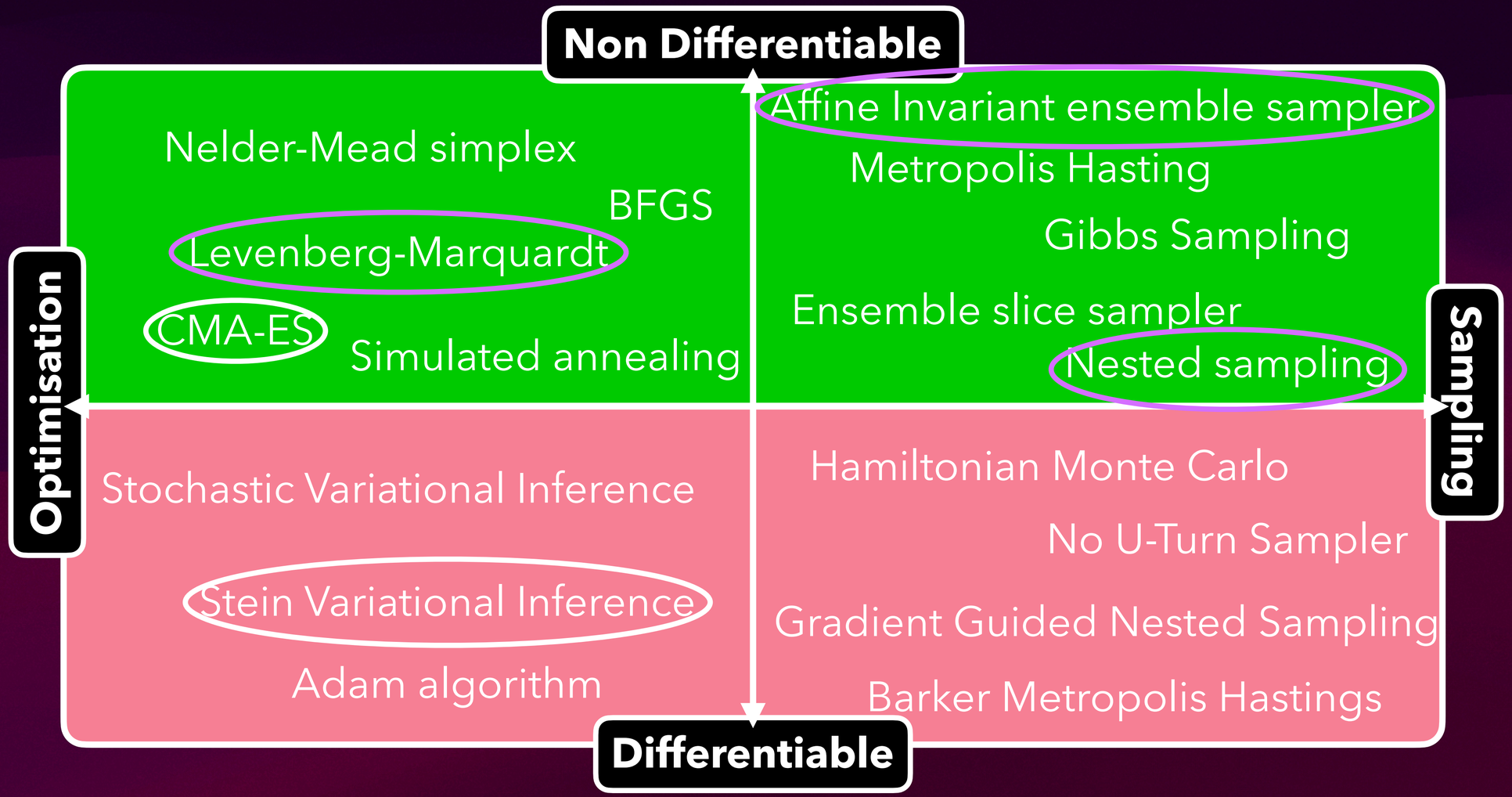


Landscape of exact inference methods



Methods available to X-ray community

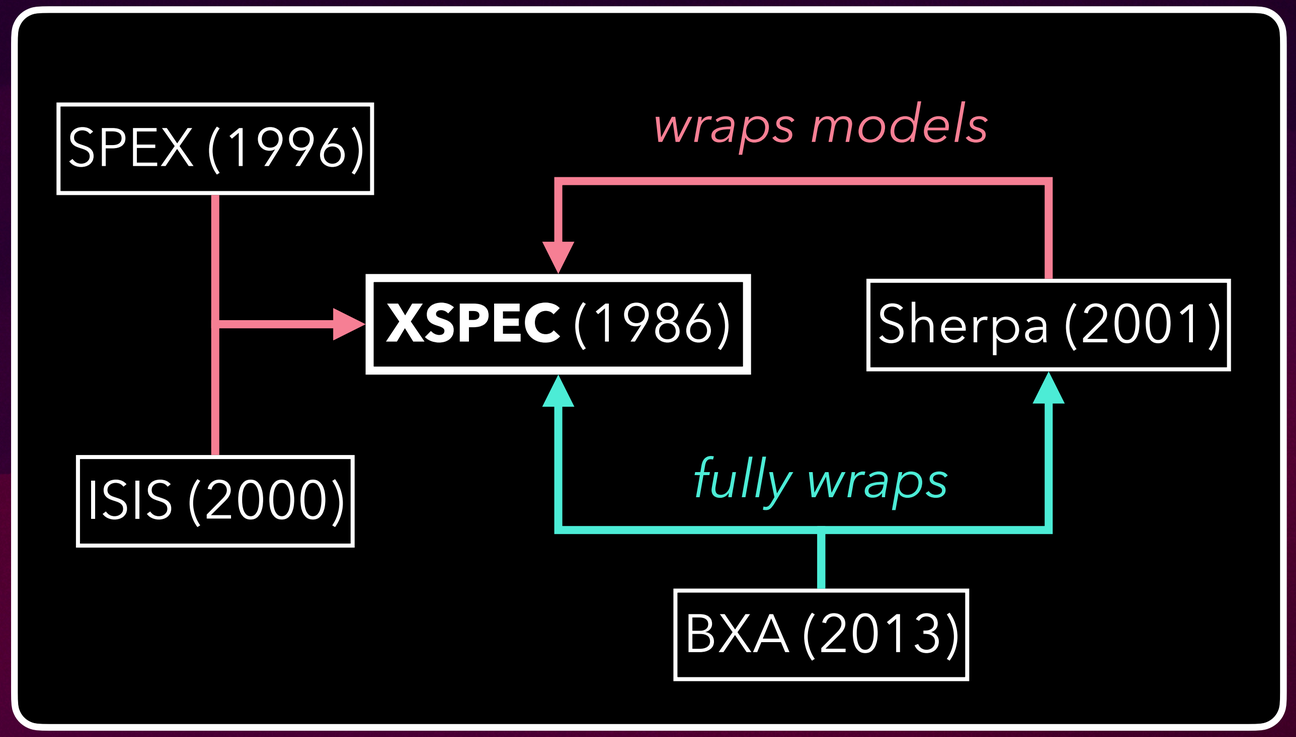
Landscape of exact inference methods

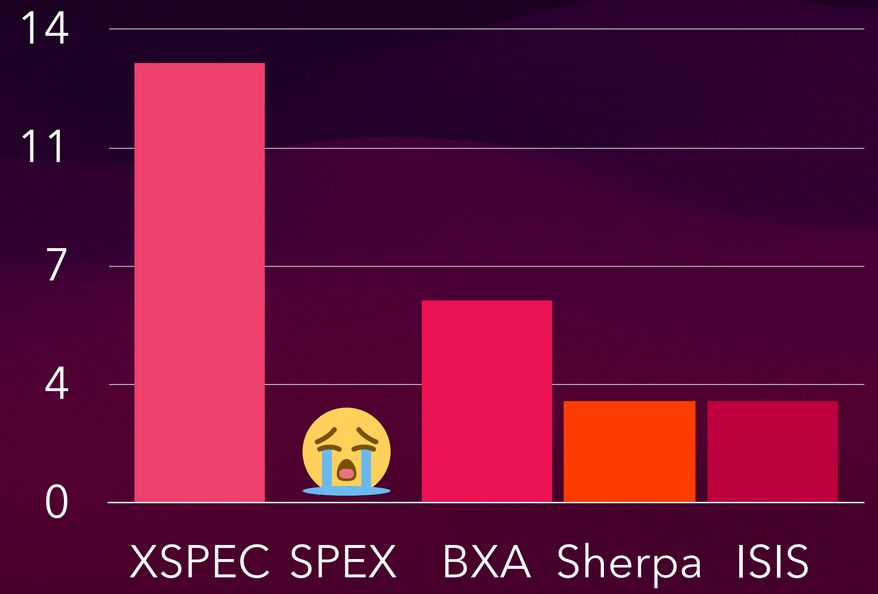


Methods available to X-ray community

Methods I personally find funny

Landscape of X-ray Spectral fitting





XSPEC is an extremely powerful and convenient software, but 40 years of monolithic development make it hard to maintain and extend

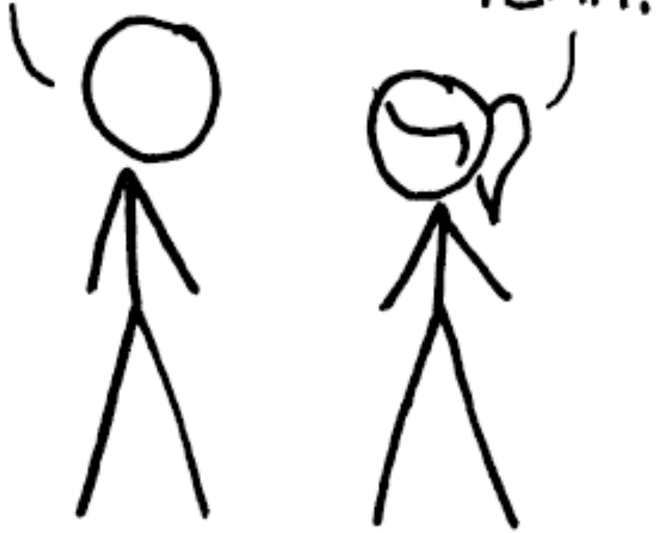
If you want to contribute to the survey (N=18 currently)

« Statistiques de Sardine » (D. Barret)

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

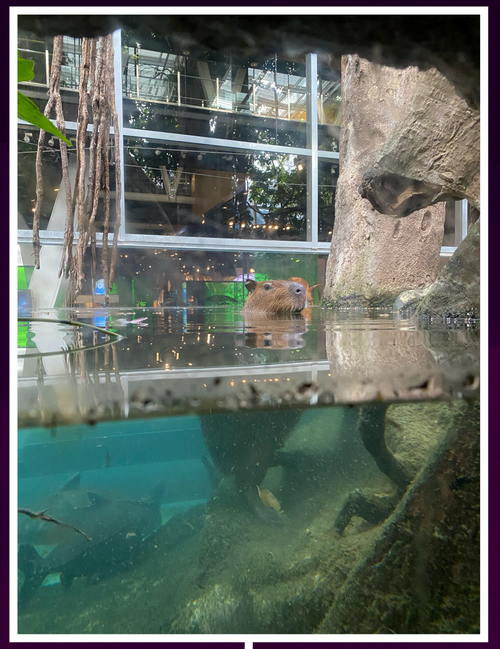
SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.

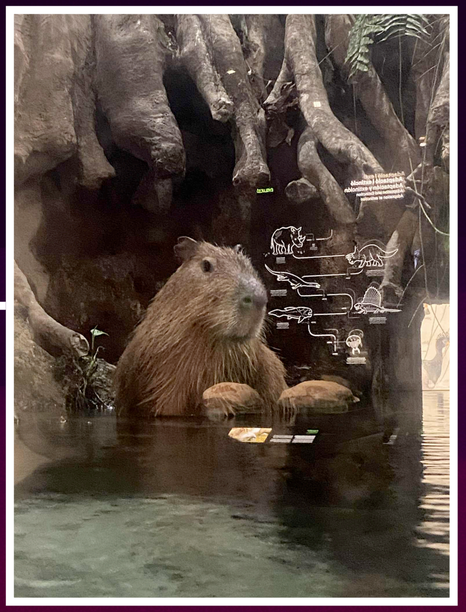


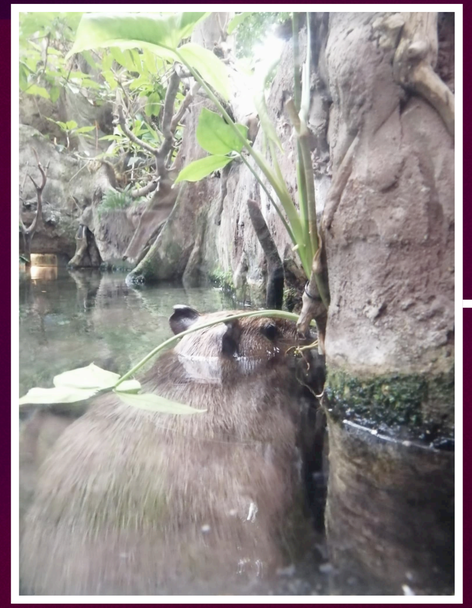
SOON:

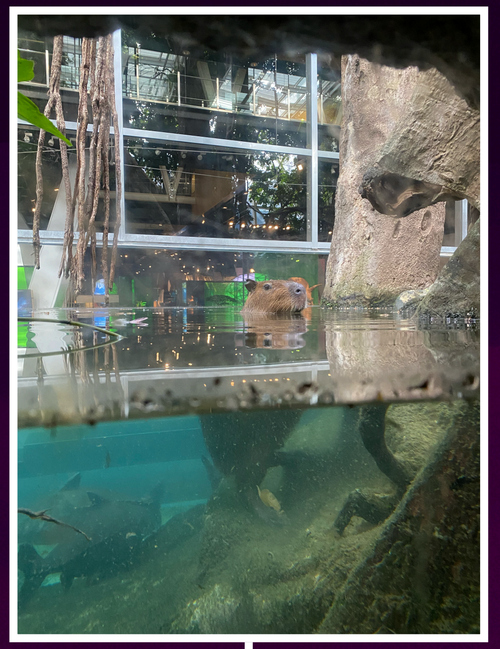
SITUATION:
THERE ARE
15 COMPETING
STANDARDS.



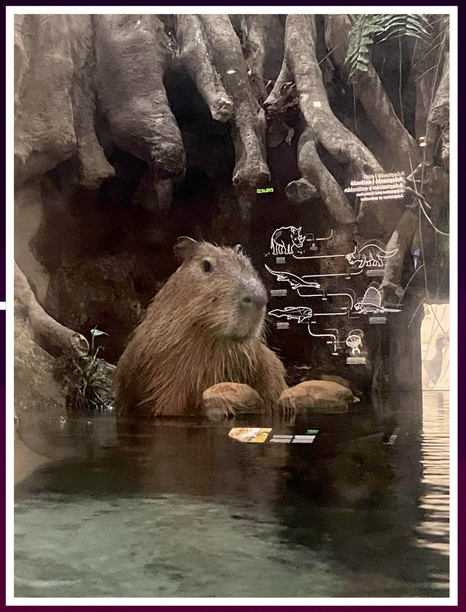
JAXSPEC

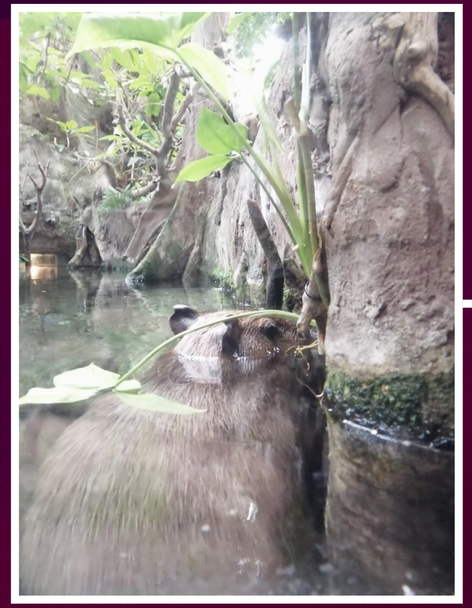




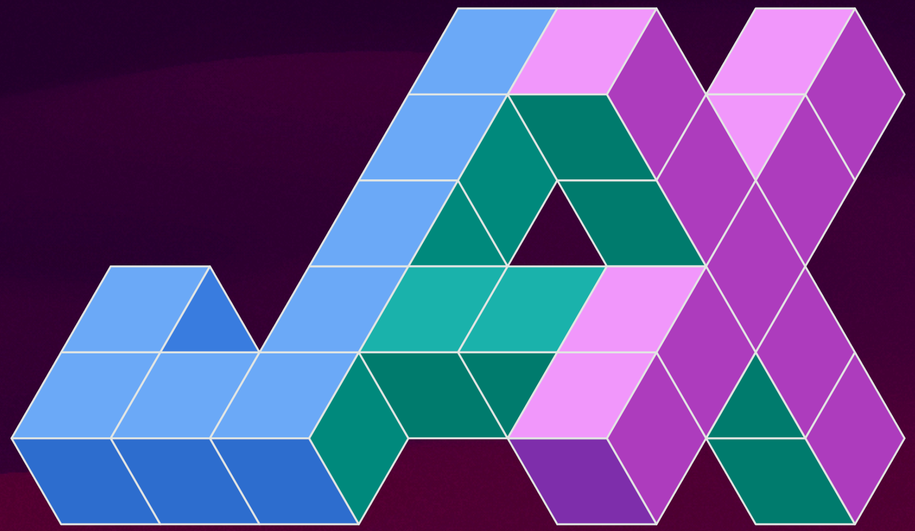


JAXSPEC





A brief introduction to JAX



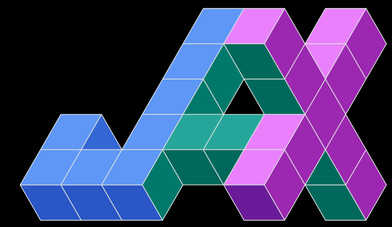
A numpy-like library with

- Just-in-time compilation
- Automatic differentiation
- Accelerators (GPUs, TPUs)

JAX is widely use by Google, DeepMind, and grows in the machine learning community. *AlphaFold*, *Gemini* or *midjourney* are built on top of it.

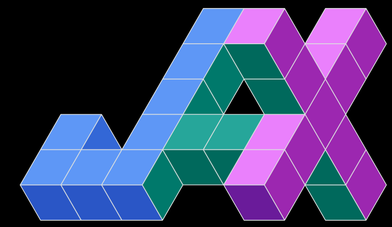
Software and hardware acceleration

Rewrite the existing stack using JAX



Software and hardware acceleration

Rewrite the existing stack using JAX



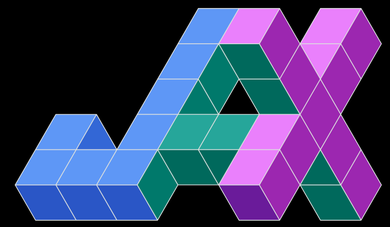
Autodifferentiation

Differentiable algorithms

- Adam for minimization
- HMC/NUTS for MCMC
- Variational Inference

Software and hardware acceleration

Rewrite the existing stack using JAX



Autodifferentiation

Transparent on accelerators

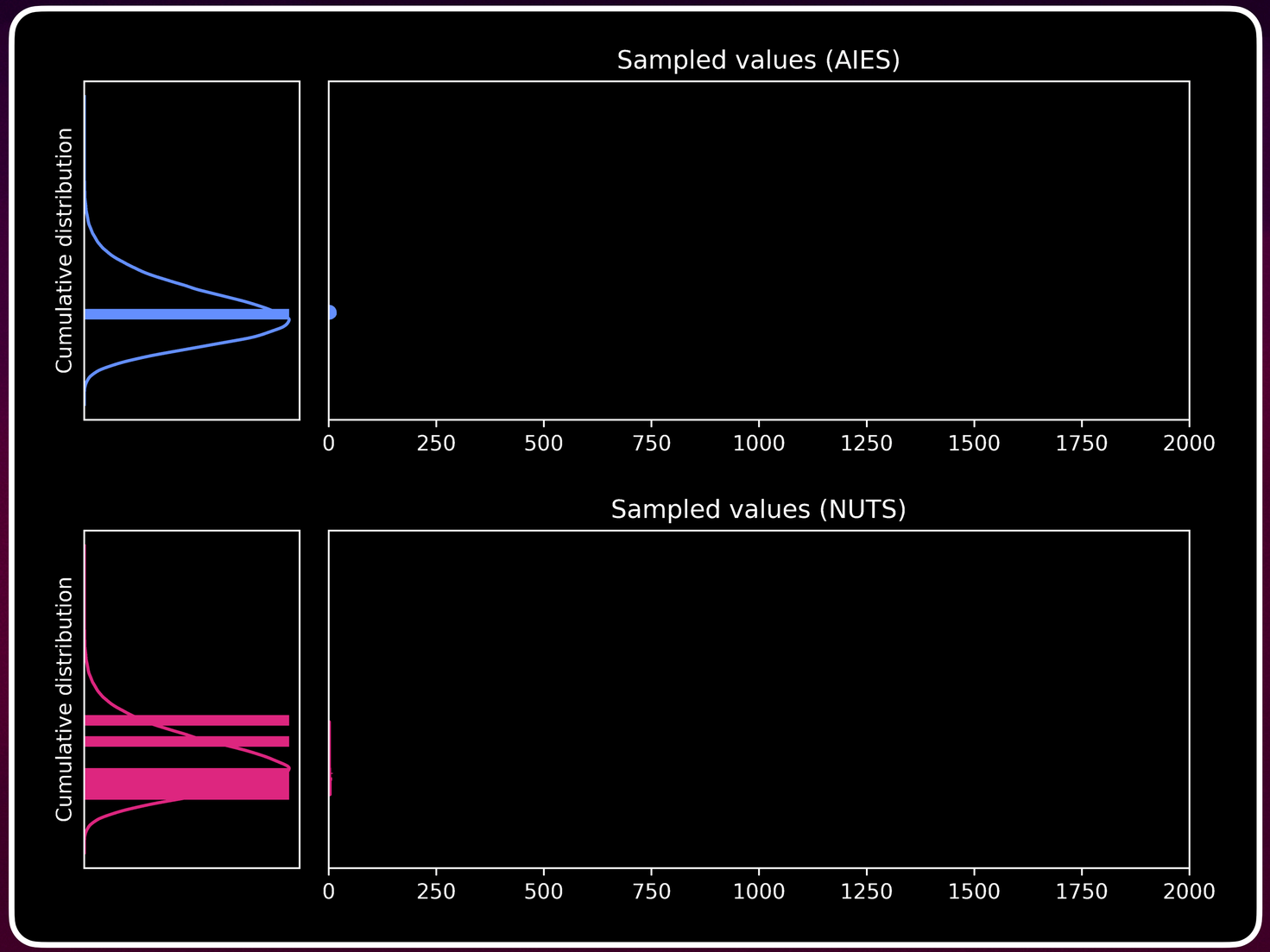
Differentiable algorithms

- Adam for minimization
- HMC/NUTS for MCMC
- Variational Inference

GPU/TPU acceleration

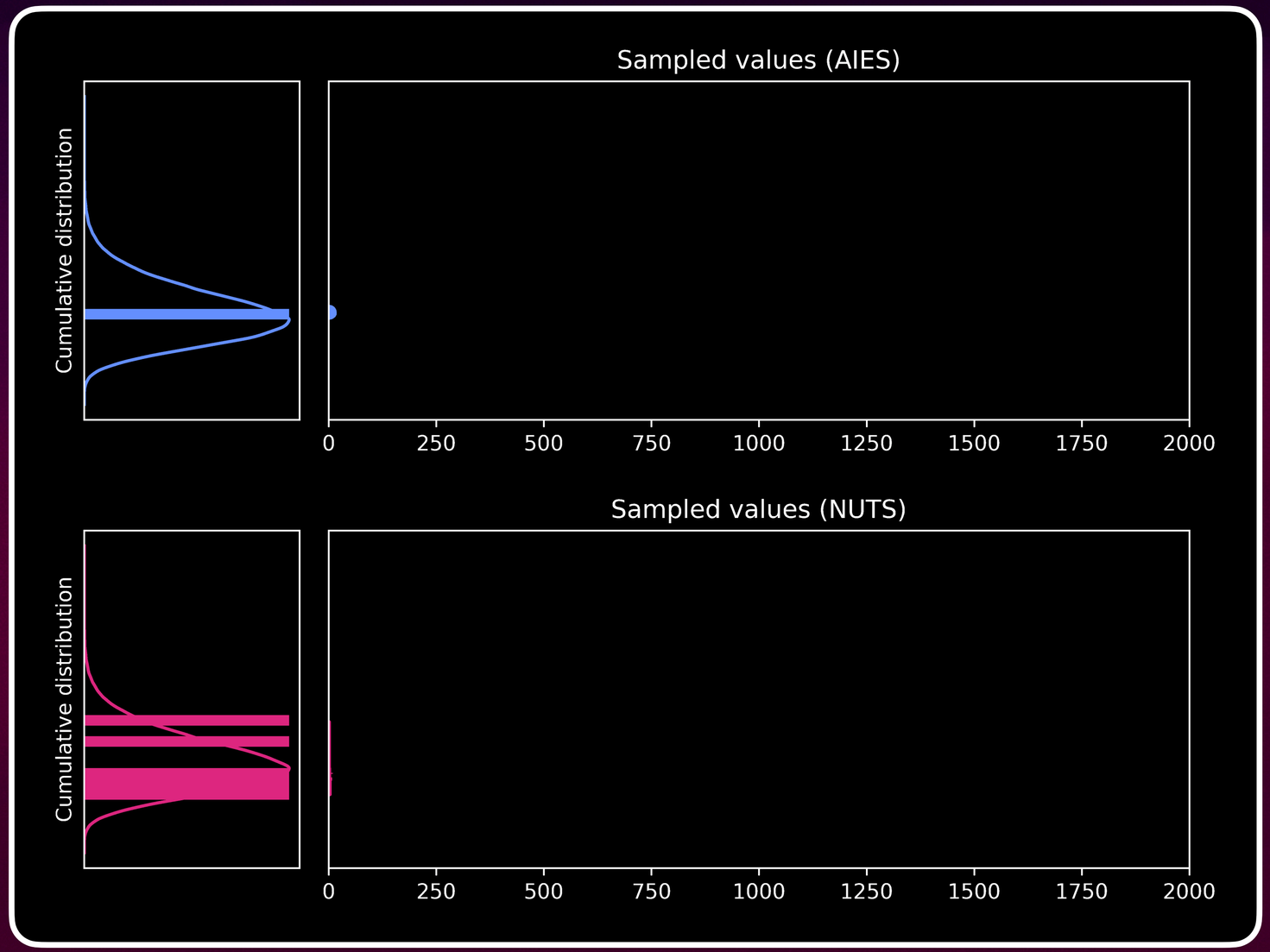
What does jaxspec bring?

- Inference on the GPU
- New algorithms
- A promise of full Python development



What does jaxspec bring?

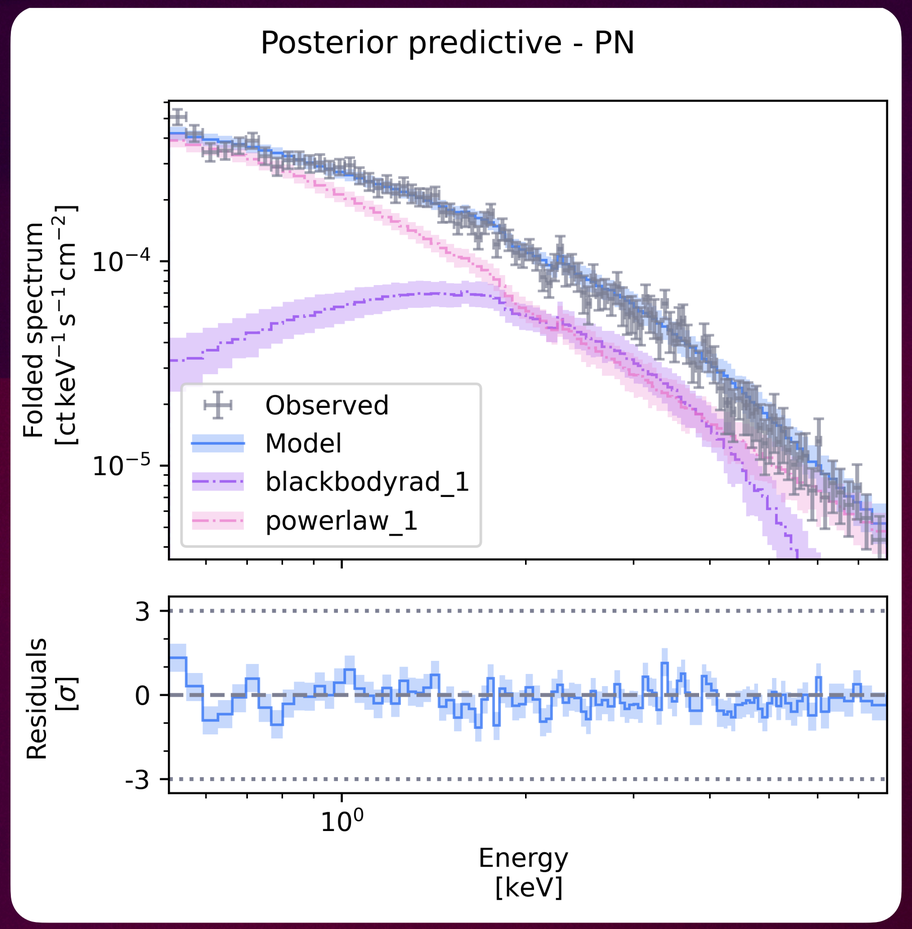
- Inference on the GPU
- New algorithms
- A promise of full Python development



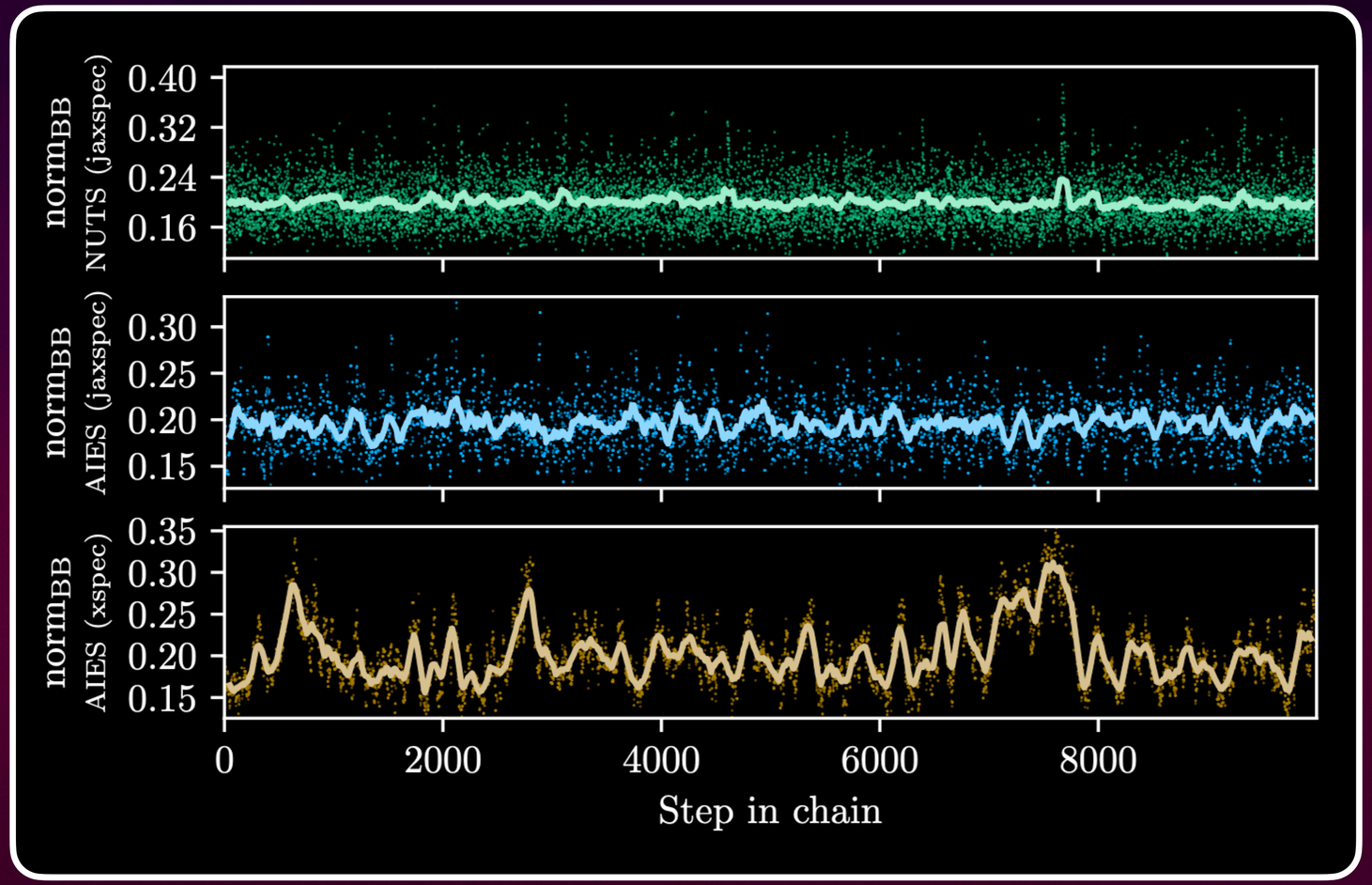
Live demo : NUTS

Some performances from the jaxspec paper

- NGC 7793 ULX-4 (XMM-Newton)
- Absorbed powerlaw and blackbody emission model

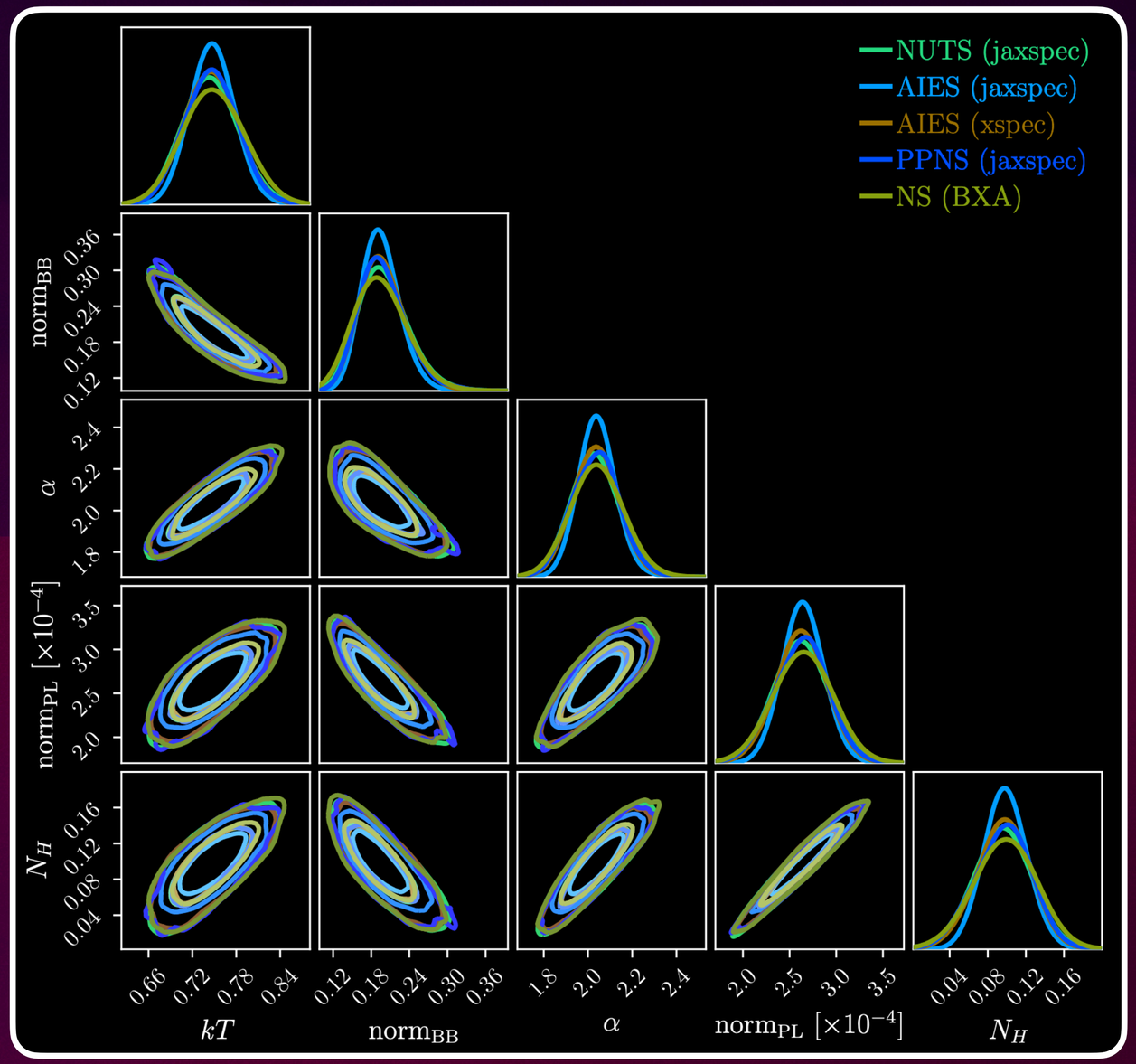


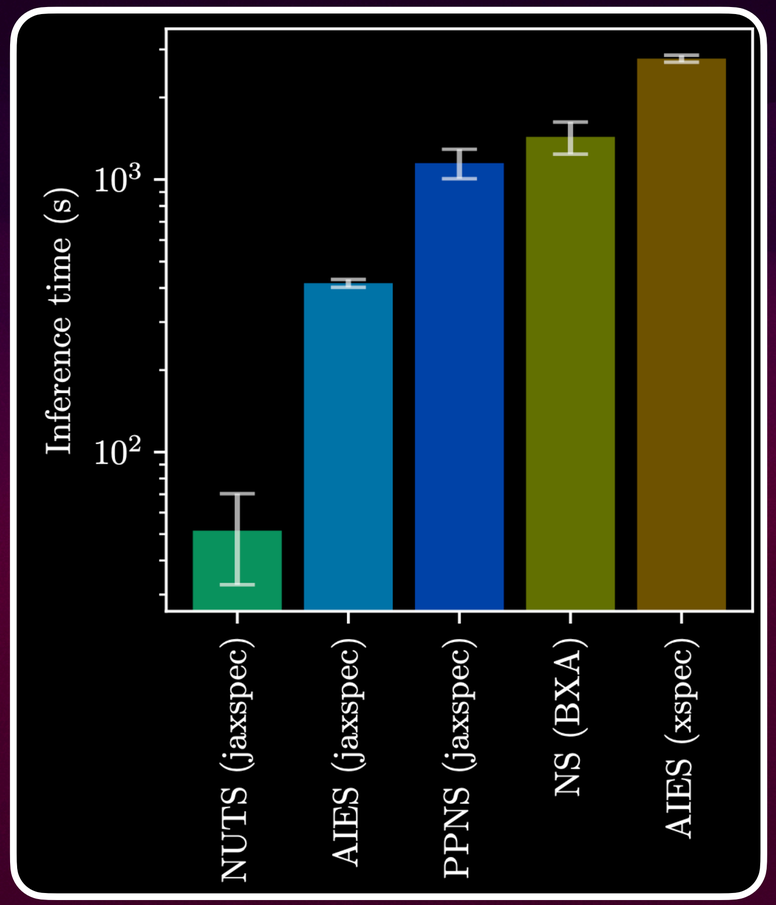
Increasing autocorrelation



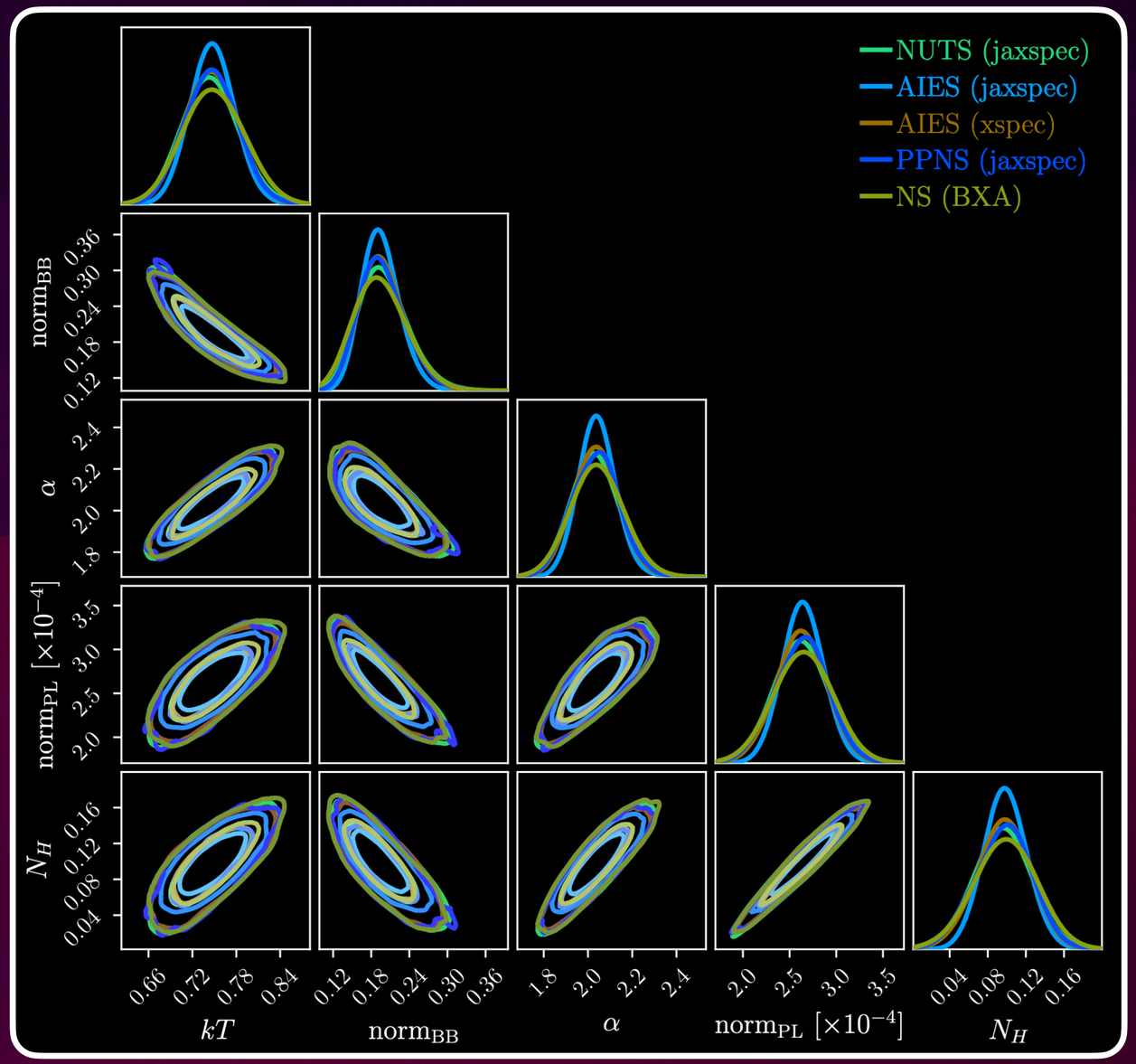
jaxspec with **NUTS (i)** or **AIES (ii)** gives much better mixing than XSPEC with **AIES (iii)**

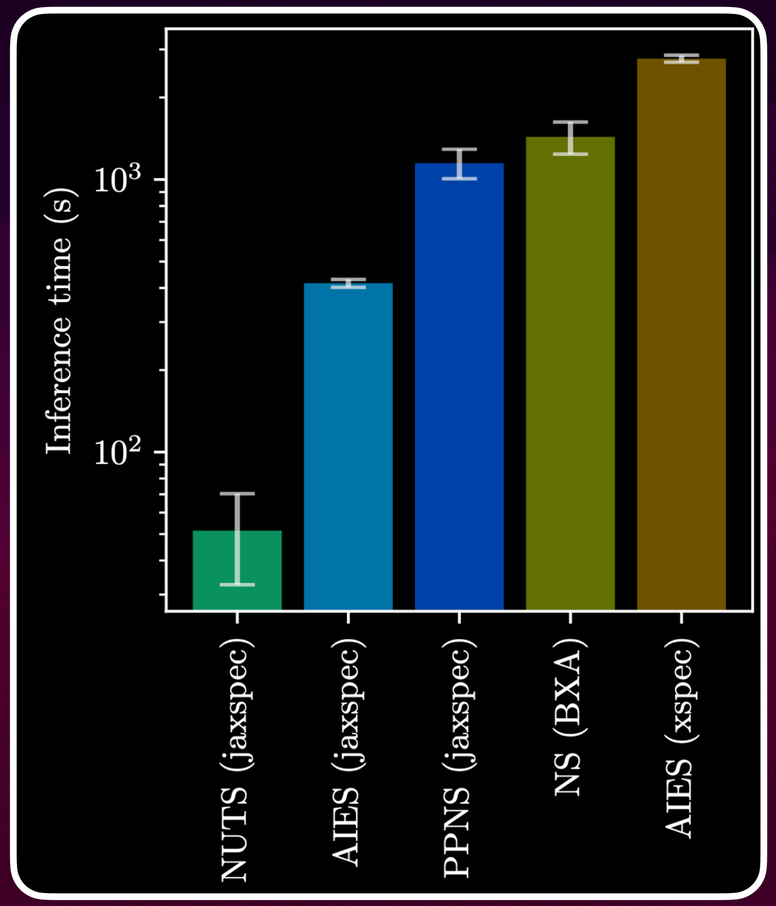
Some performances from the jaxspec paper





Some performances from the jaxspec paper

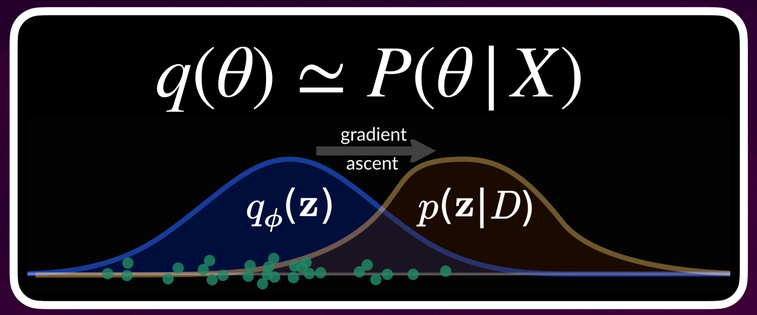


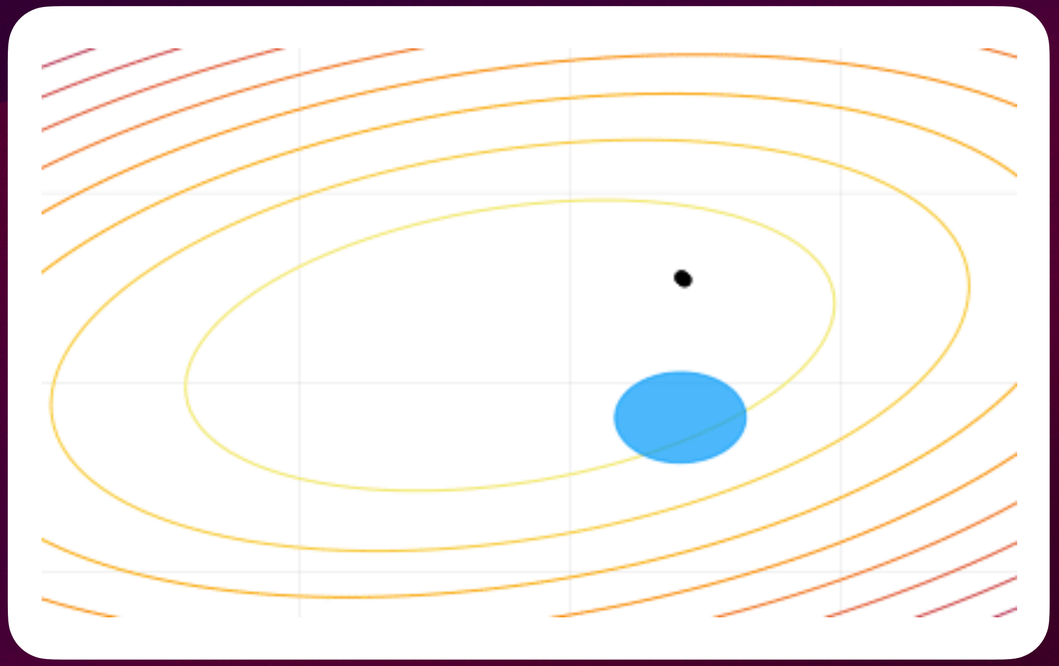


Time to reach convergence $\hat{=} \hat{R} < 1.01$ for MCMC samplers

- 10 order of magnitude speedup
- Equivalent results, no approximation
- **Run on equivalent material**, no GPU

Intro to Variational inference

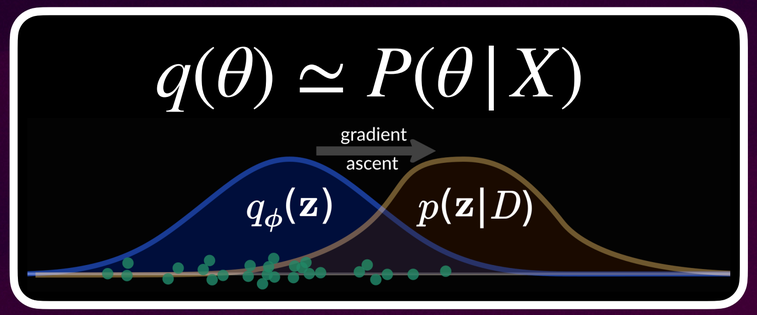


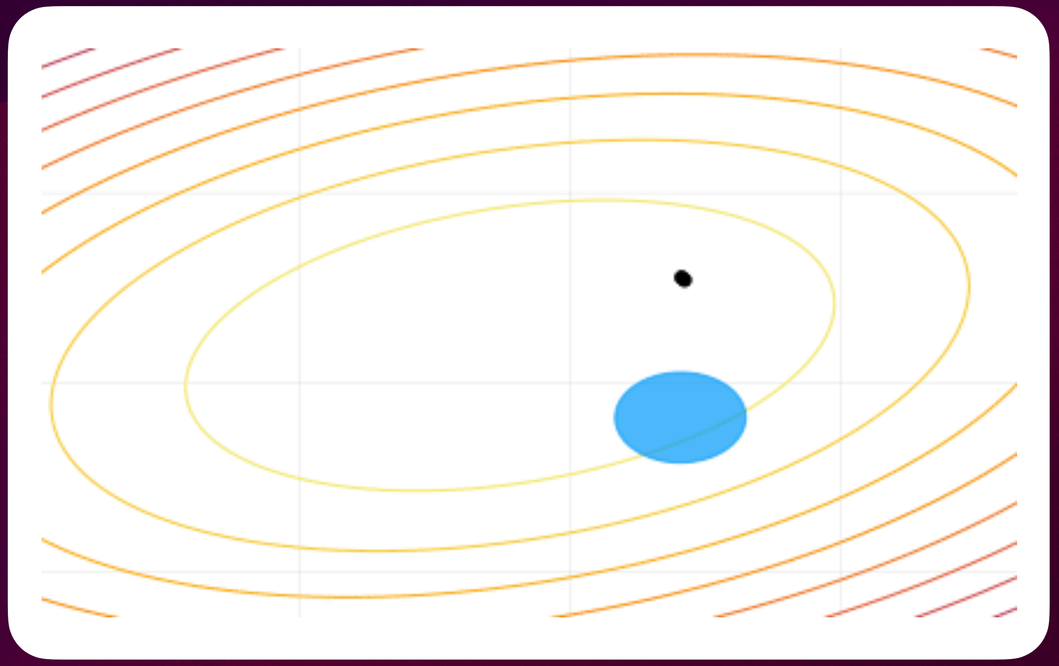


1. **Chose a variational distribution** $q(\theta)$
   e.g. "my posterior could be approximated with a multivariate Gaussian distribution"

2. **Minimize the Evidence Lower Bound**
   Turn Bayesian inference into an optimisation problem. Best if performed with gradient guided minimiser (Adam)

Intro to Variational inference





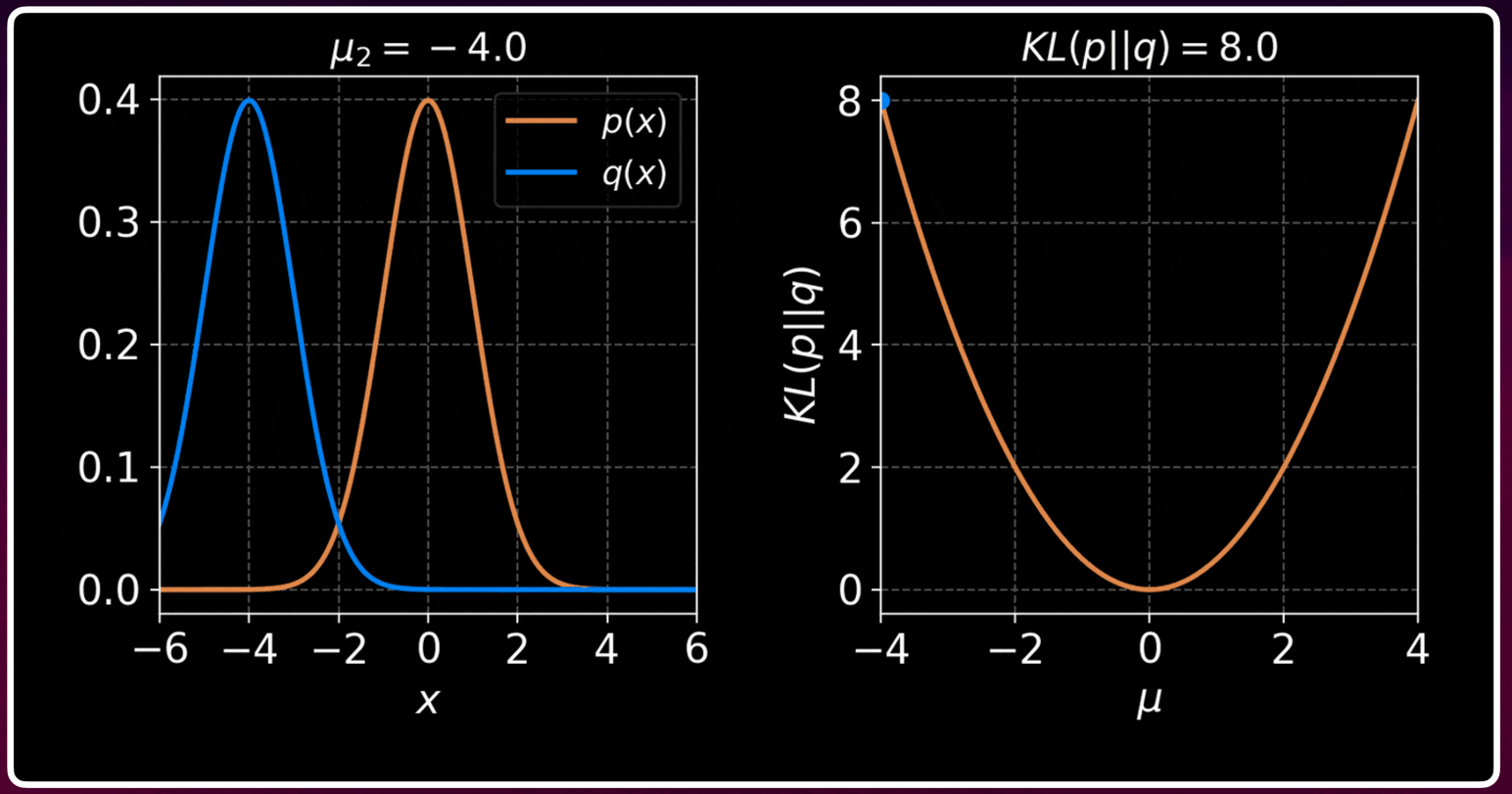
1. **Chose a variational distribution** $q(\theta)$
   e.g. "my posterior could be approximated with a multivariate Gaussian distribution"

2. **Minimize the Evidence Lower Bound**
   Turn Bayesian inference into an optimisation problem. Best if performed with gradient guided minimiser (Adam)

Kullback–Leibler divergence

Divergence (not distance!) between two probability distributions

$$D_{\mathrm{KL}}(p \parallel q) = \int p(x) \log \frac{p(x)}{q(x)}$$



Minimise the divergence between the **variational distribution** and the **posterior distribution**

$$D_{\mathrm{KL}}(q(\theta) \parallel p(\theta|x))$$

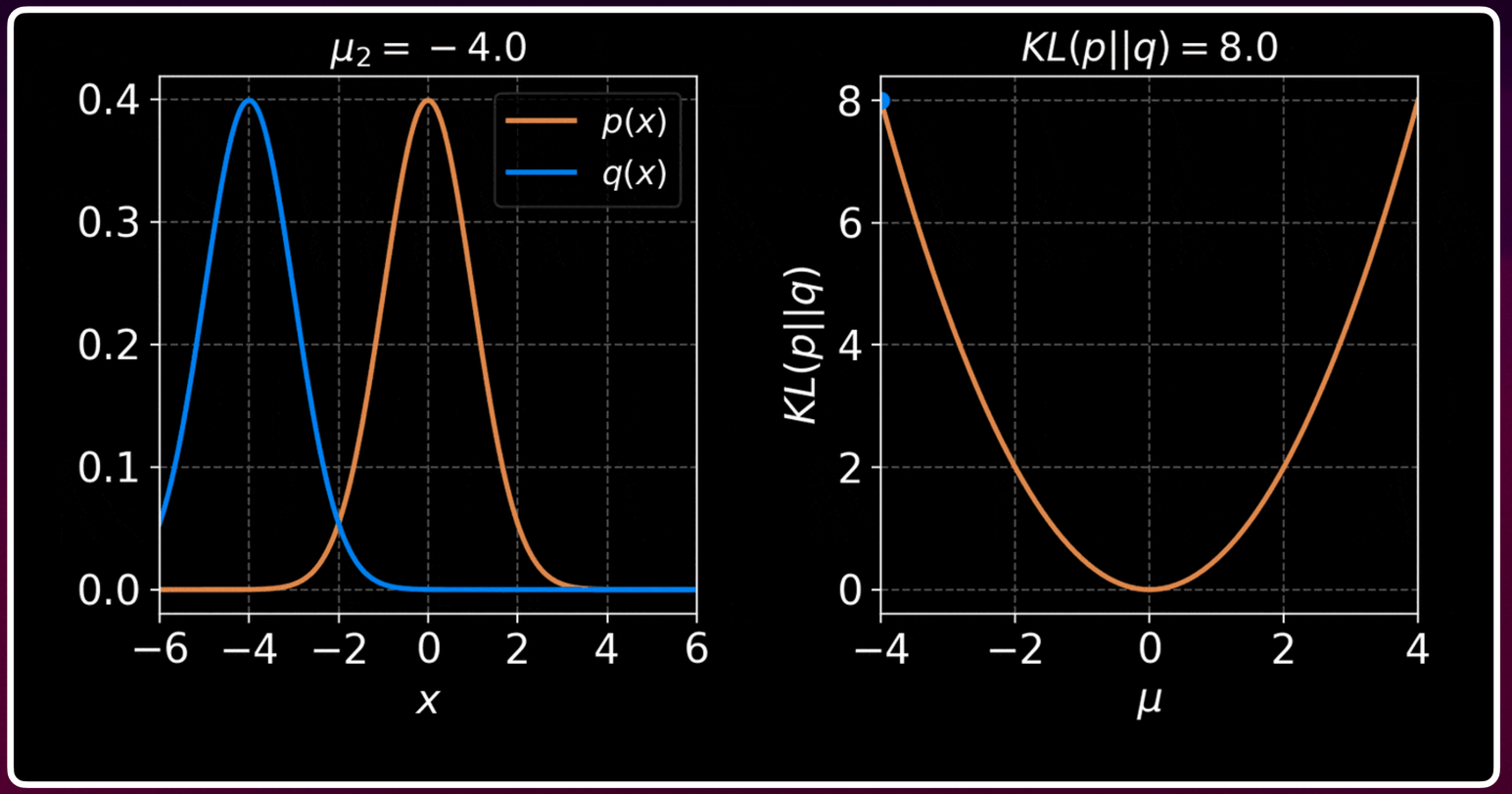
Extra maths

We maximise the **Evidence Lower Bound** (ELBO)

Kullback–Leibler divergence

Divergence (not distance!) between two probability distributions

$$D_{\mathrm{KL}}(p \parallel q) = \int p(x) \log \frac{p(x)}{q(x)}$$



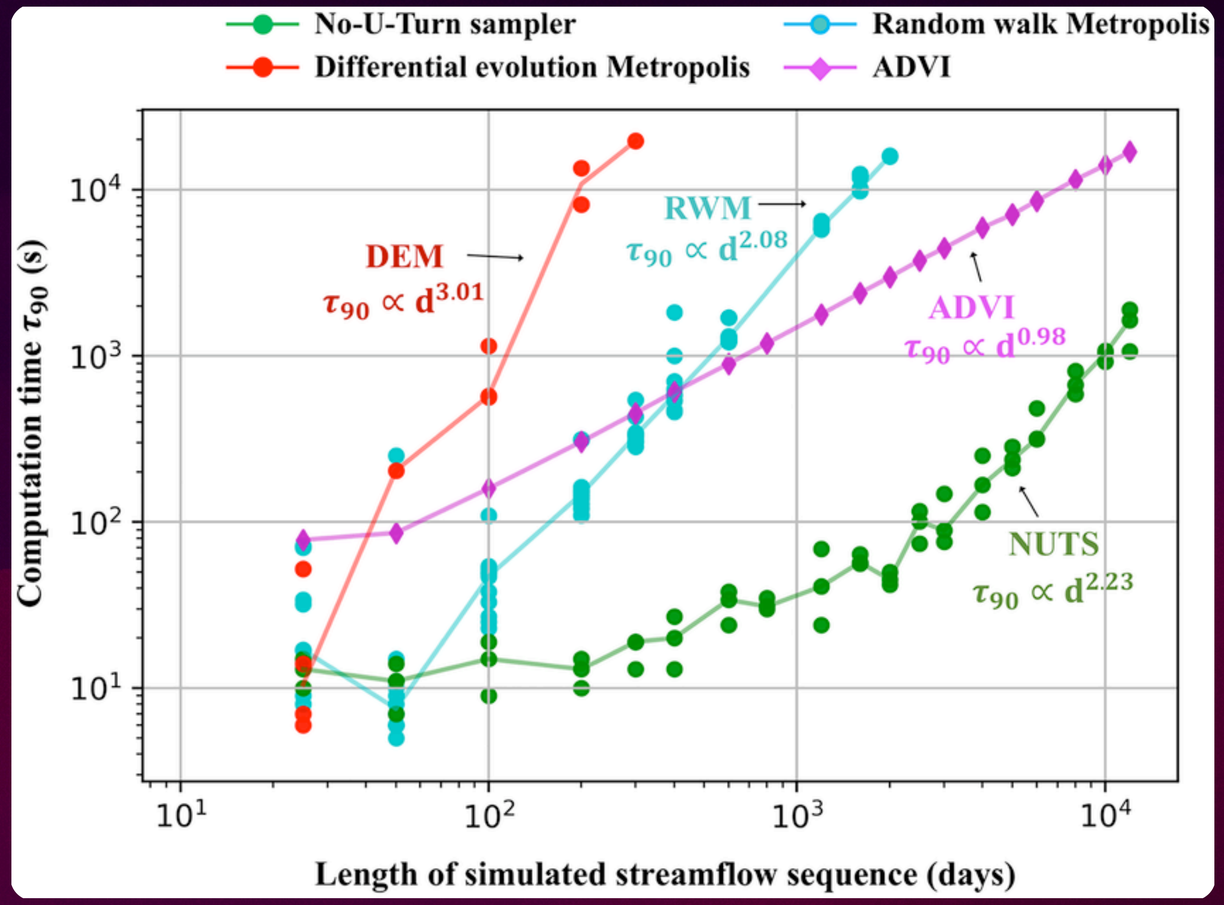
Minimise the divergence between the **variational distribution** and the **posterior distribution**

$$D_{\mathrm{KL}}(q(\theta) \parallel p(\theta|x))$$

Extra maths

We maximise the **Evidence Lower Bound** (ELBO)

VI for high-resolution spectroscopy



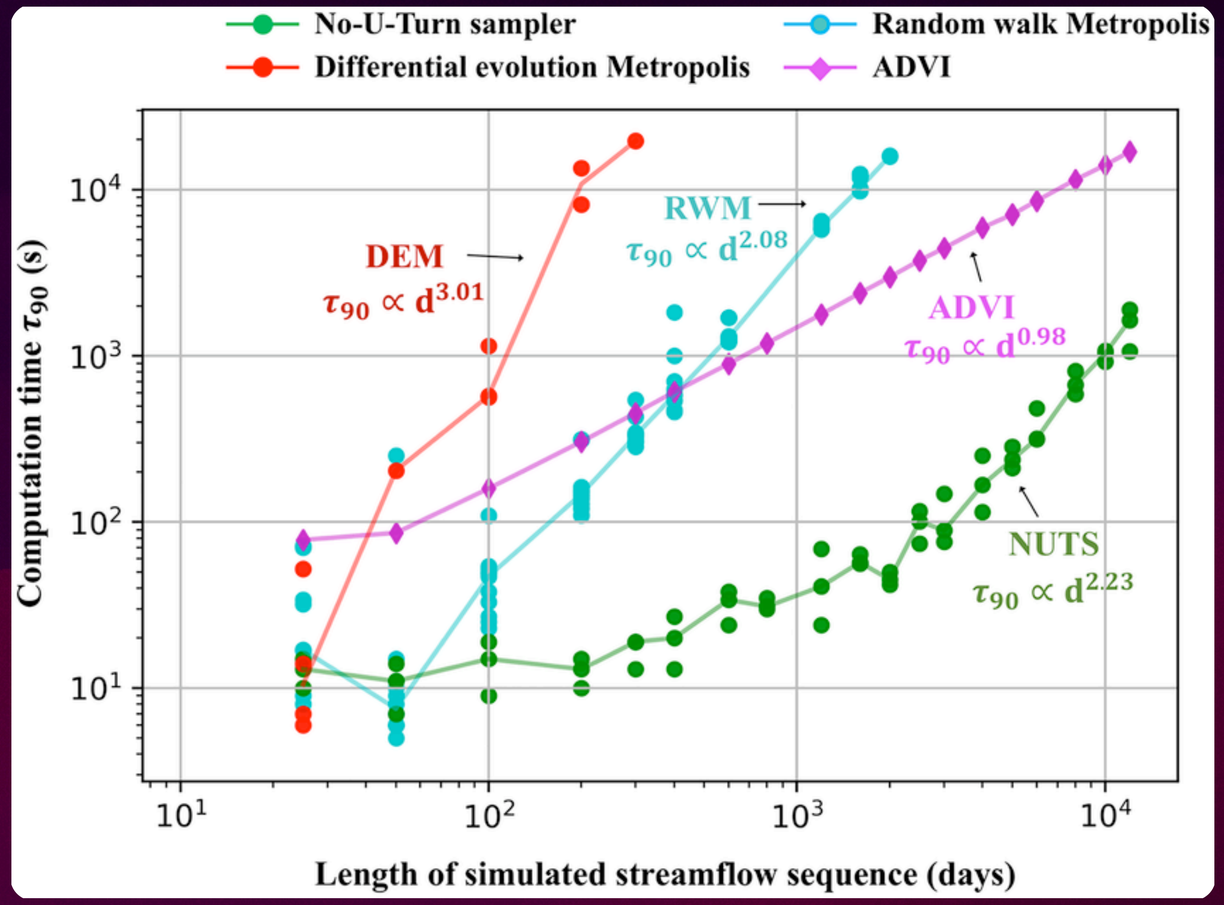
Adapted from Krapu & al. (2019)

Suited for high-resolution X-ray spectroscopy

- Requires less gradient evaluation compared to NUTS
- Scales better with high number of parameters
- Fully profits from GPU acceleration, where NUTS is less efficient for multiple chains on the same device

$\tau \propto d^{0.98}$ **for VI** vs $\tau \propto d^{2.23}$ **for NUTS**

VI for high-resolution spectroscopy



Adapted from Krapu & al. (2019)

Suited for high-resolution X-ray spectroscopy

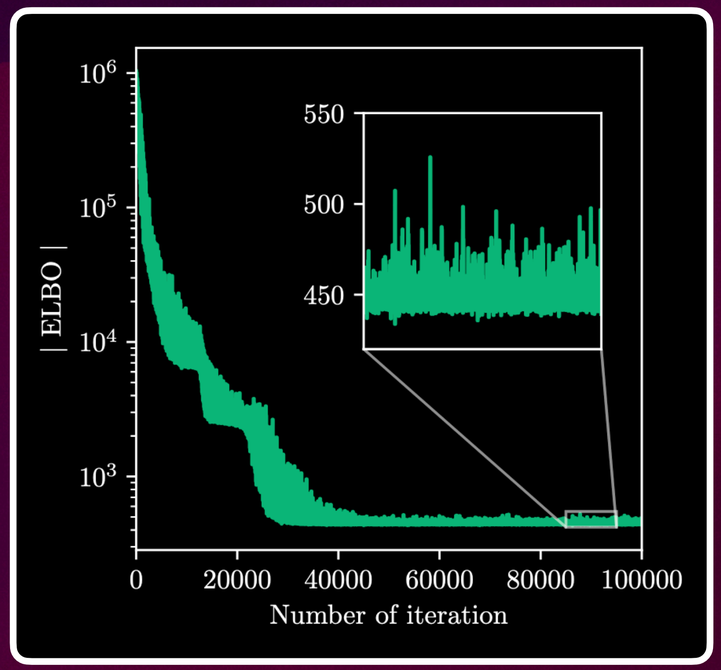
- Requires less gradient evaluation compared to NUTS
- Scales better with high number of parameters
- Fully profits from GPU acceleration, where NUTS is less efficient for multiple chains on the same device

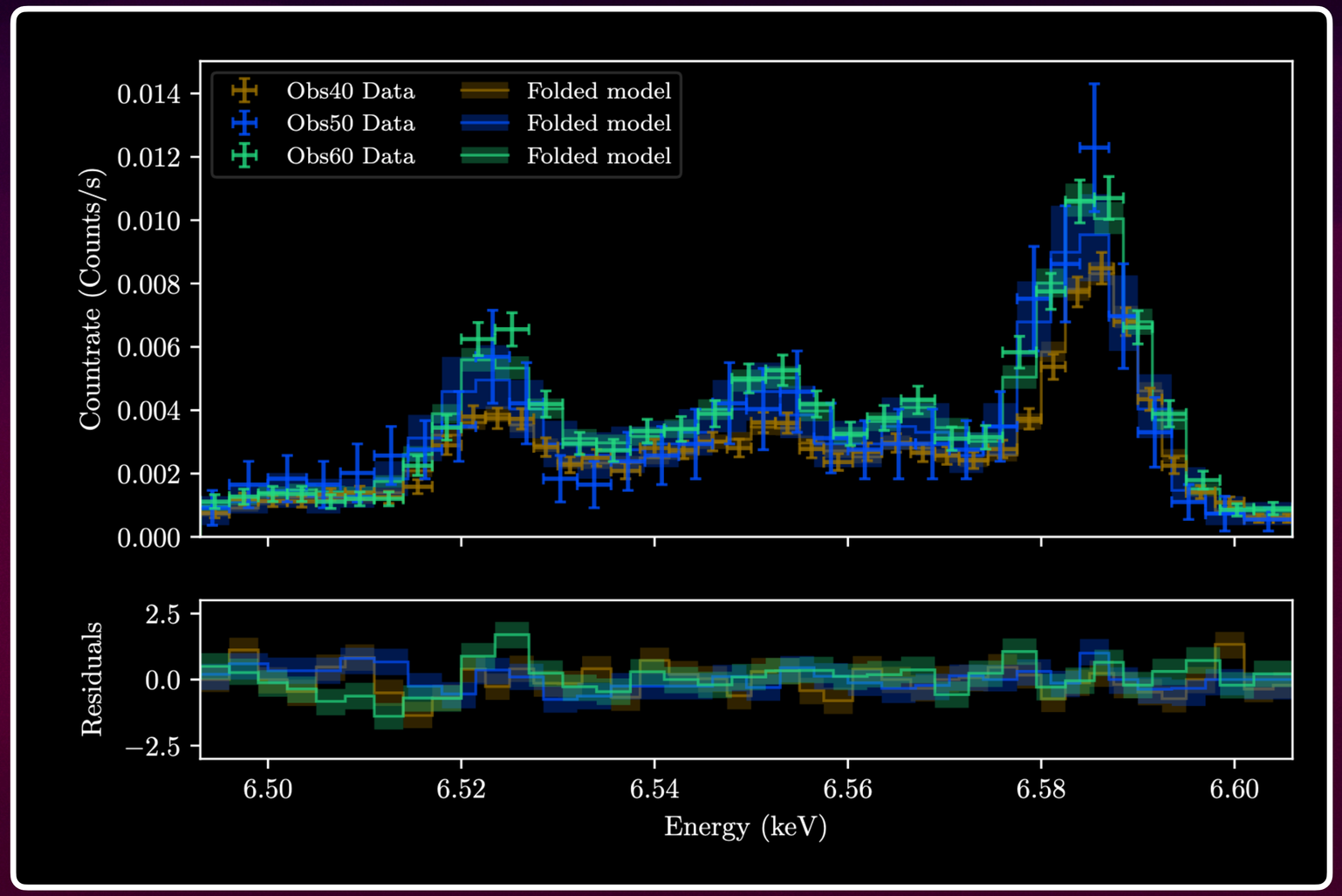
$\tau \propto d^{0.98}$ **for VI** vs $\tau \propto d^{2.23}$ **for NUTS**

Example : application to Hitomi

See Dupourqué & al . 2024

- 28 parameter model
- Iron complex with 3 *Hitomi* pointings on Perseus
- 10^5 steps of *adamw*
- 7 minutes on a A2 GPU

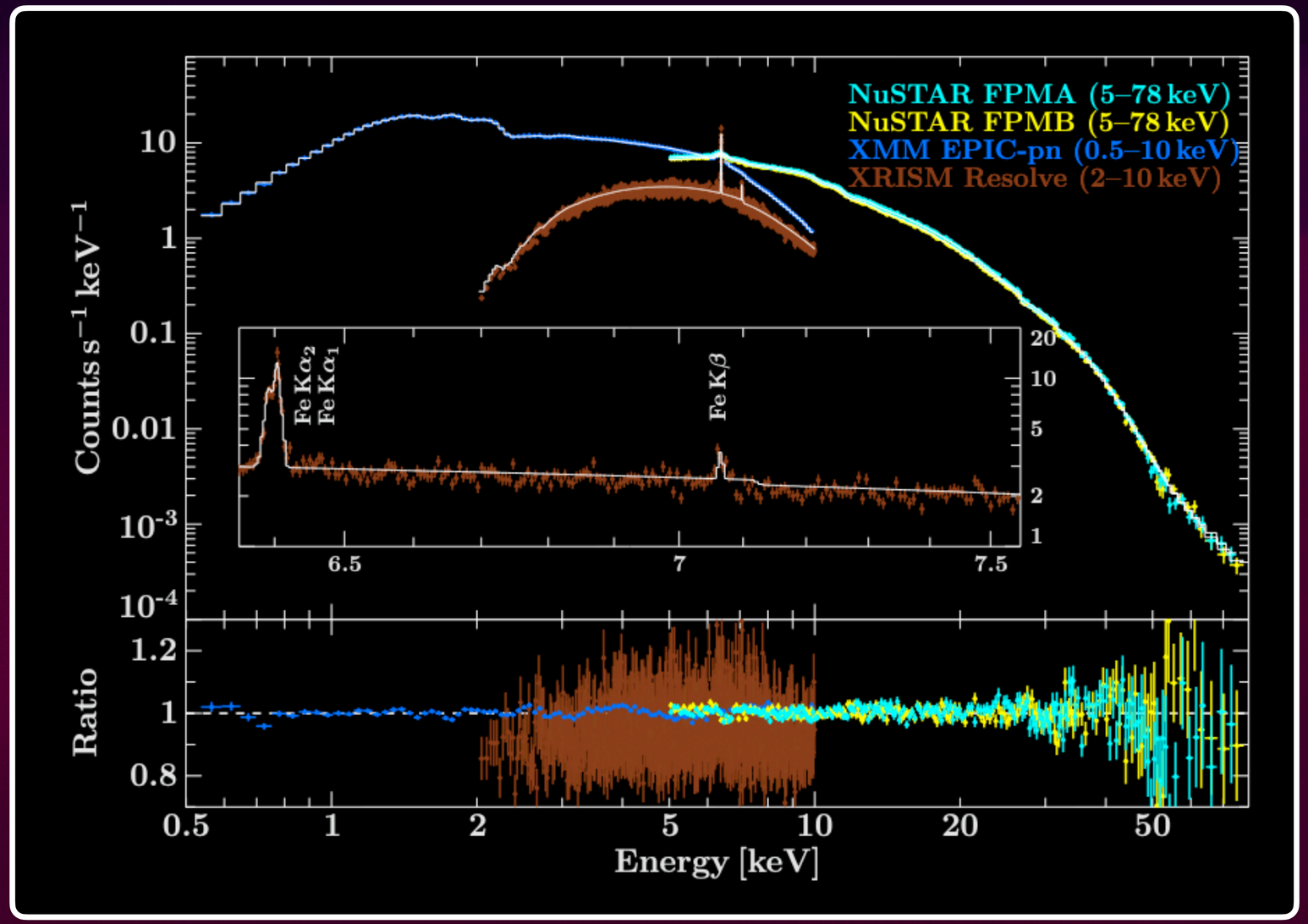




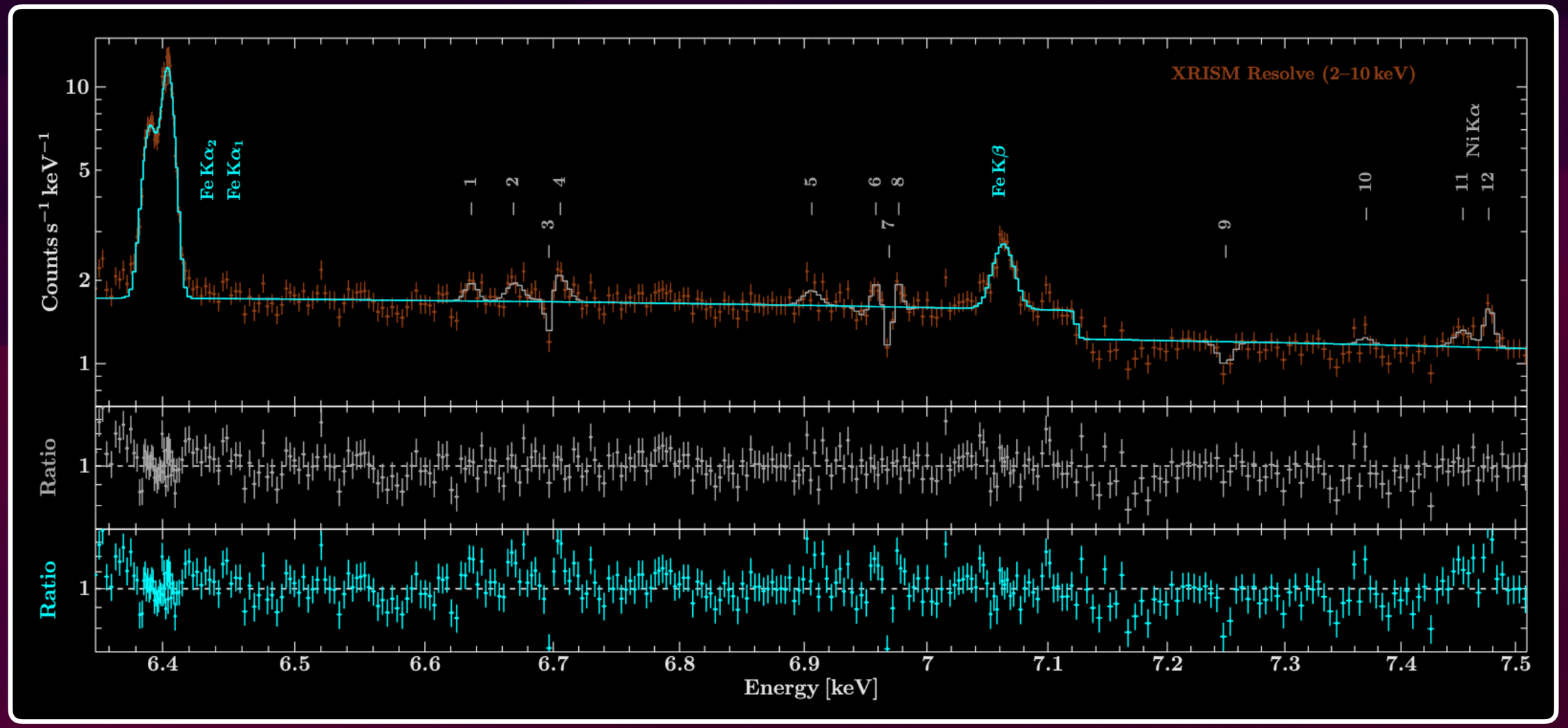
Example : application to Vela X-1

See Diez & al . 2025

- 38 parameter model + thousands of background parameters
- Powerlaw continuum, cyclotron lines, 10 keV features, emission lines …
- 4 instruments (*XRISM*/Resolve, *NuSTAR*/FPMA, *NuSTAR*/FPMB, *XMM*/Epic-PN)
- Cross calibration
- 50 minutes inference on a H100 GPU



Example : application to Vela X-1



- Nothing here is magic, **every algorithm can fail**
- There is no single test or statistic that will give you 100% confidence
- There must be a **human in the loop** questioning the results

- Nothing here is magic, **every algorithm can fail**
- There is no single test or statistic that will give you 100% confidence
- There must be a **human in the loop** questioning the results

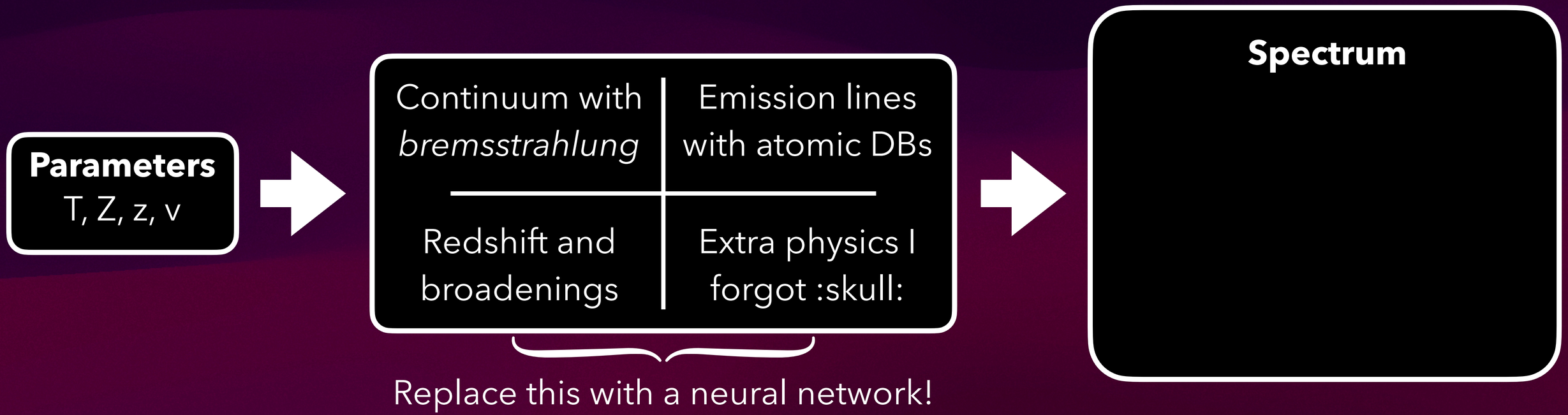
Conclusion

- jaxspec is in early development stage
- Yet it starts to be science ready, as it passes cross calibration tests with existing software with increasingly complex case studies
- Extending jaxspec with complex physical models (APEC, relxill) is horrendously hard
- Is surrogate modeling a good solution ?

Backup slides

Surrogate modelling in a nutshell

Example : fully ionized plasma with APEC model



- Same wild machinery with reflexion models, obscuration models etc.
- Models of this kind are usually tabulated to speed things up
- Tabulation has its own issues and scales poorly with the number of parameters

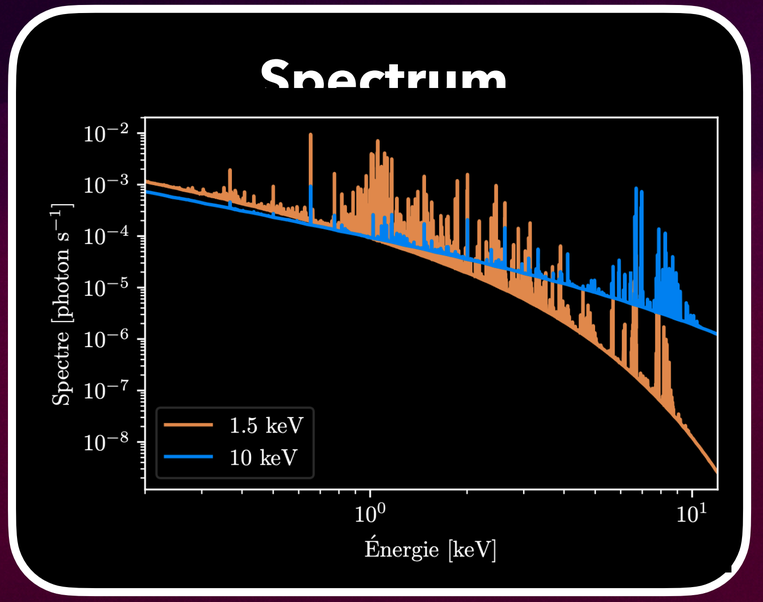
Surrogate modelling in a nutshell

Example : fully ionized plasma with APEC model

Parameters
T, Z, z, v

Continuum with *bremsstrahlung*	Emission lines with atomic DBs
Redshift and broadenings	Extra physics I forgot :skull:

Replace this with a neural network!

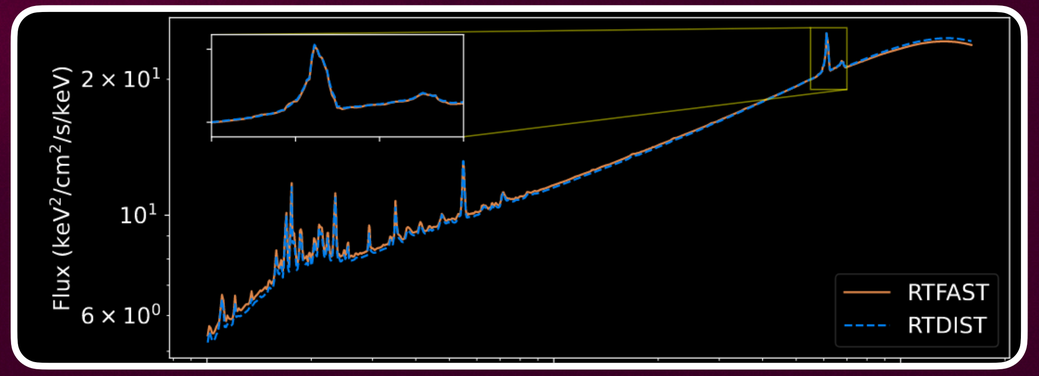


- Same wild machinery with reflexion models, obscuration models etc.
- Models of this kind are usually tabulated to speed things up
- Tabulation has its own issues and scales poorly with the number of parameters

Example : RTFAST

Parameters
[17 dims]

256	256	256	256	256	256	256	256	256	256	256	256	200
256	256	256	256	256	256	256	256	256	256	256	256	200
256	256	256	256	256	256	256	256	256	256	256	256	200

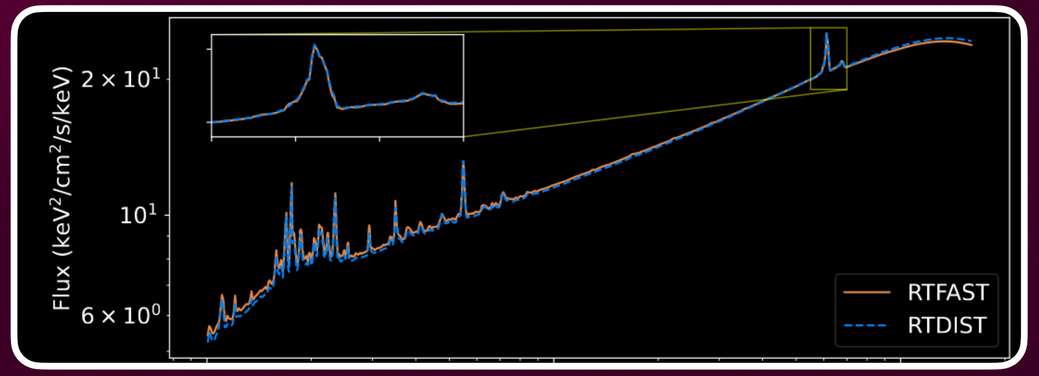


Example : RTFAST

Parameters
[17 dims]

Standard Scaling

256	256	256	256	256	256	256	256	256	256	256	256	200
256	256	256	256	256	256	256	256	256	256	256	256	200
256	256	256	256	256	256	256	256	256	256	256	256	200



Example : RTFAST

Emulation for X-ray reverberation spectra : check Ricketts & al. (2024)

Parameters
[17 dims]

Standard Scaling

Ensemble of 7 Dense MLP (GeLU activation)

256 256 256 256 256 256 256 256 256 256 256 256 200

256 256 256 256 256 256 256 256 256 256 256 256 200

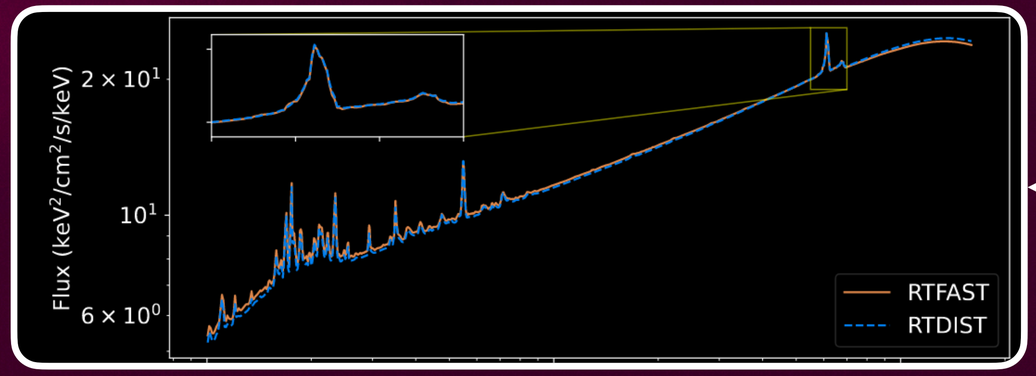
•••

256 256 256 256 256 256 256 256 256 256 256 256 200

Averaging

Inverse PCA
Linear combination of **200** fixed PCA components

Standard Unscaling + Exponential transform



- Emulation with $\mathcal{O}(3\%)$ accuracy
- 200 times faster calls
- Limited to XMM resolution

From KL to ELBO

Let's compute the KL-divergence between our variational distribution and the target posterior

$$D_{\mathrm{KL}}(q(\theta) \parallel p(\theta|x) = \int_\theta q(\theta) \log \frac{q(\theta)}{p(\theta|x)}$$

$$= \int_\theta q(\theta) \log \frac{q(\theta)p(x)}{p(x,\theta)}$$

$$= \int_\theta \underbrace{q(\theta) \log q(\theta)} + \int_\theta \underbrace{q(\theta) \log p(x)} - \int_\theta \underbrace{q(\theta) \log p(x,\theta)}$$

Negative entropy of the variational distribution

Constant wrt θ

Expectation of the joint probability

From KL to ELBO

$$D_{\mathrm{KL}}(q(\theta) \parallel p(\theta|x)) = \mathbb{E}_{q(\theta)}\left[\log q(\theta)\right] + \log p(x) - \mathbb{E}_{q(\theta)}\left[\log p(x,\theta)\right]$$

Rearrange the terms

$$\log p(x) = \underbrace{\mathbb{E}_{q(\theta)}\left[\log p(x,\theta)\right] - \mathbb{E}_{q(\theta)}\left[\log q(\theta)\right]}_{\mathcal{L}(q)} - D_{\mathrm{KL}}(q(\theta) \parallel p(\theta|x))$$

$$\log p(x) = \mathcal{L}(q) - D_{\mathrm{KL}}(q(\underbrace{\theta) \parallel p}(\theta|x))$$

Divergence > 0

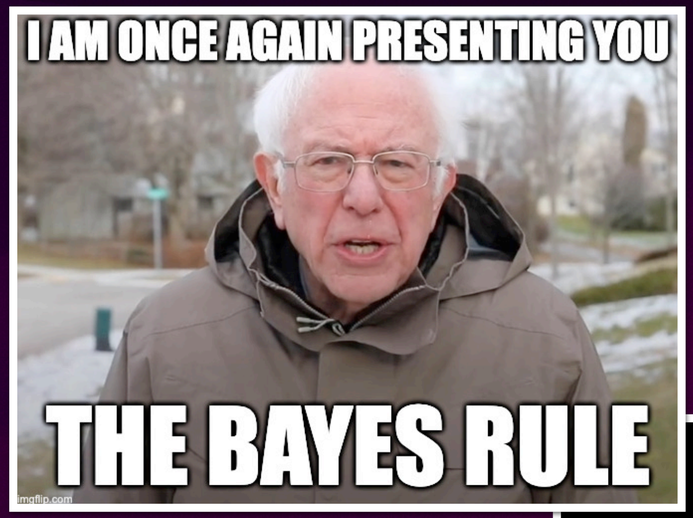
$\log p(x) \geq \mathcal{L}(q)$ so maximising $\mathcal{L}(q)$ maximises the evidence

Bayesian inference

θ : parameters

X: observation(s)

Bayesian inference

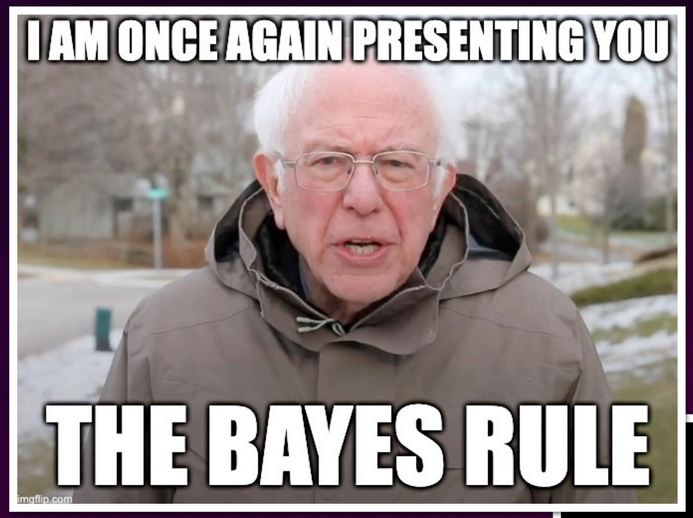


θ : parameters

X: observation(s)

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

Bayesian inference



θ : parameters

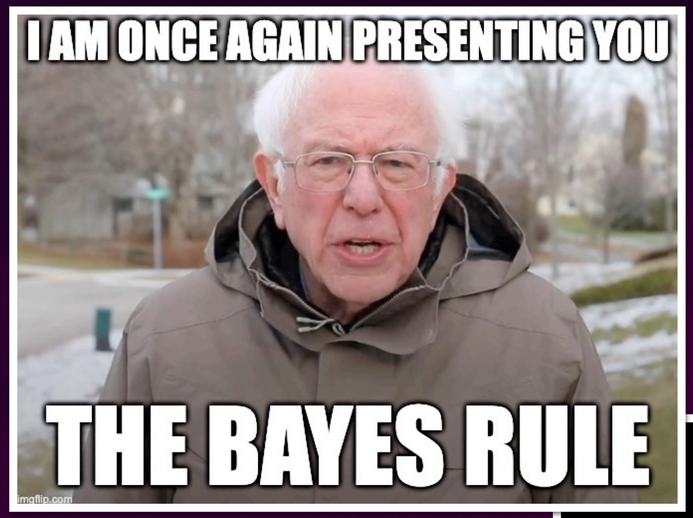
X: observation(s)

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

Prior

A *priori* probability of the parameters

Bayesian inference



θ : parameters

X: observation(s)

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

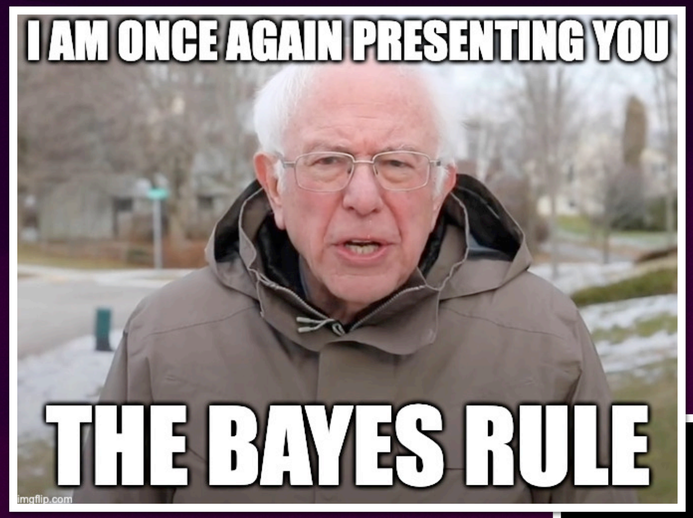
Likelihood

Probability of the observation(s) given the parameters

Prior

A *priori* probability of the parameters

Bayesian inference



θ : parameters

X: observation(s)

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

Likelihood

Probability of the observation(s) given the parameters

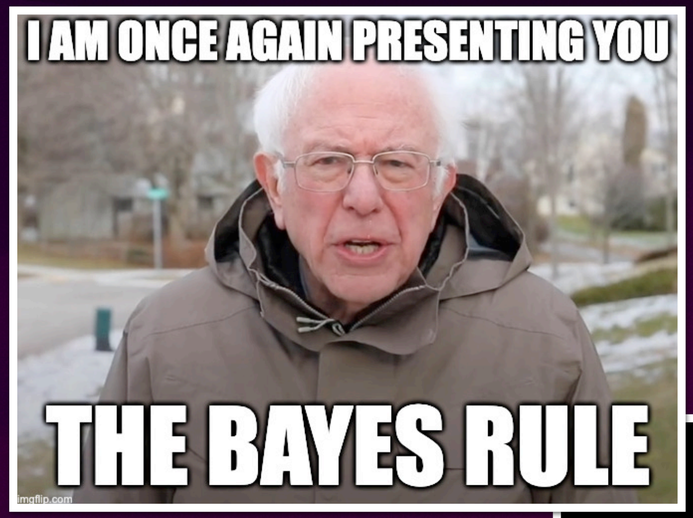
Posterior

A *posteriori* of the parameters given the observation(s)

Prior

A *priori* probability of the parameters

Bayesian inference



θ : parameters

X: observation(s)

$$P(\theta \mid X) = \frac{P(X \mid \theta)}{P(X)} P(\theta)$$

Likelihood

Probability of the observation(s) given the parameters

Posterior

A *posteriori* of the parameters given the observation(s)

Prior

A *priori* probability of the parameters

Reason why Bayesian inference is hard to perform

Reminders of X-ray spectroscopy

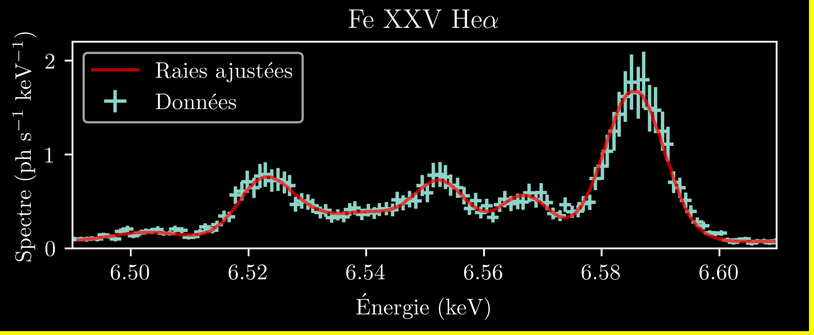
$$\mathcal{S}(E,\theta) = \mathcal{R} * \mathcal{M}(E,\theta)$$

Reminders of X-ray spectroscopy

$$S(E,\theta) = \mathcal{R} * \mathcal{M}(E,\theta)$$

Observed data

Spectra measured by the instrument

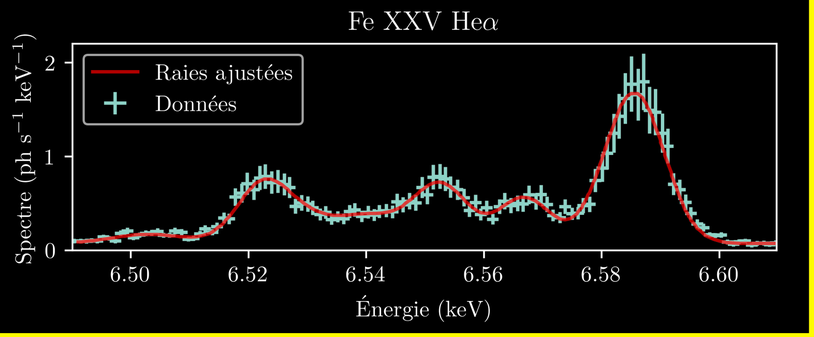


Reminders of X-ray spectroscopy

$$S(E,\theta) = \mathcal{R} * \mathcal{M}(E,\theta)$$

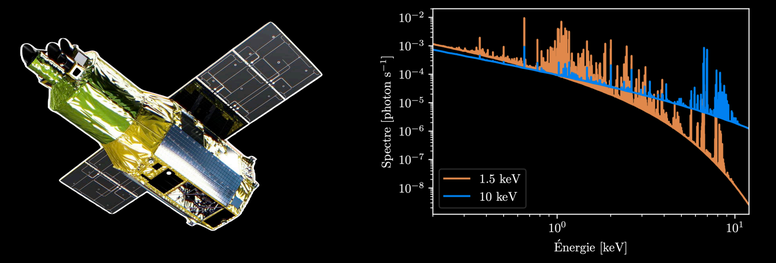
Observed data

Spectra measured by the instrument



Model

Spectral model and instrumental convolution



Reminders of X-ray spectroscopy

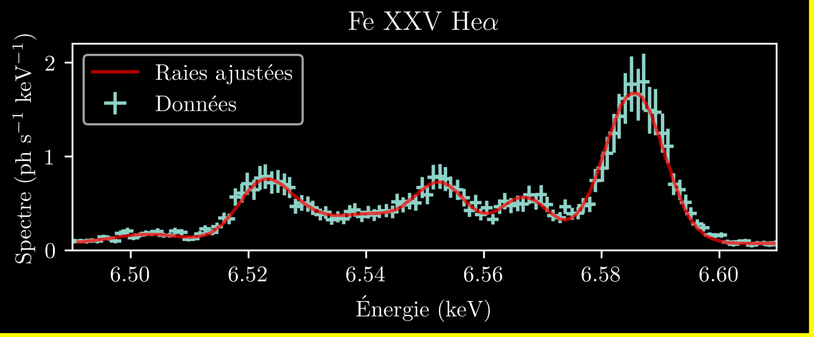
$$S(E, \theta) = \mathcal{R} * \mathcal{M}(E, \theta)$$

Parameters

Temperature, Abundance, Redshift…

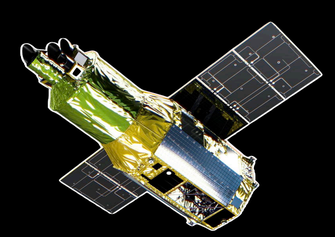
Observed data

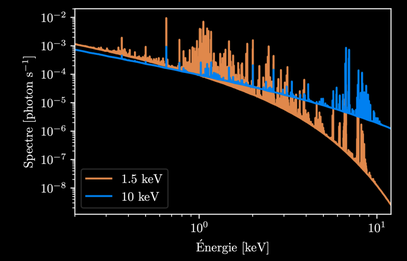
Spectra measured by the instrument



Model

Spectral model and instrumental convolution





Reminders of X-ray spectroscopy

$$S(E,\theta) = \mathcal{R} * \mathcal{M}(E,\theta)$$

Parameters

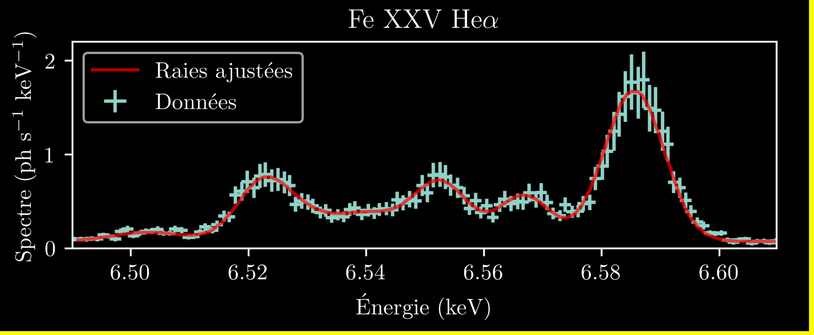
Temperature, Abundance, Redshift…

Expected photons in each channel

$$\lambda \equiv S(E,\theta)$$

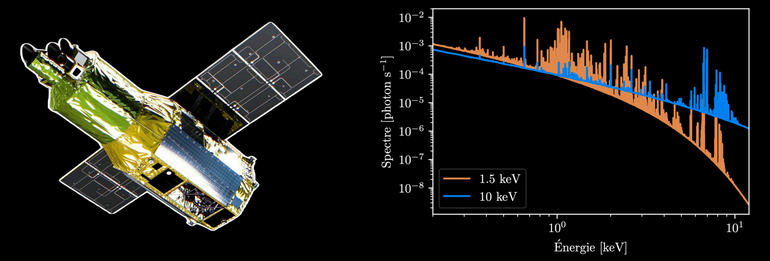
Observed data

Spectra measured by the instrument



Model

Spectral model and instrumental convolution



Reminders of X-ray spectroscopy

$$S(E,\theta) = \mathcal{R} * \mathcal{M}(E,\theta)$$

Parameters

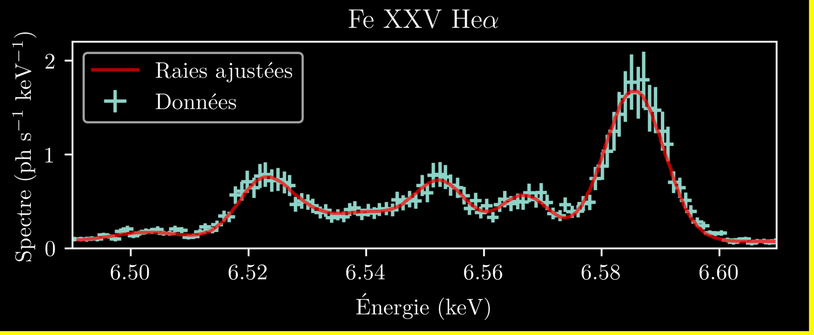
Temperature, Abundance, Redshift…

Expected photons in each channel

$$\lambda \equiv S(E,\theta)$$

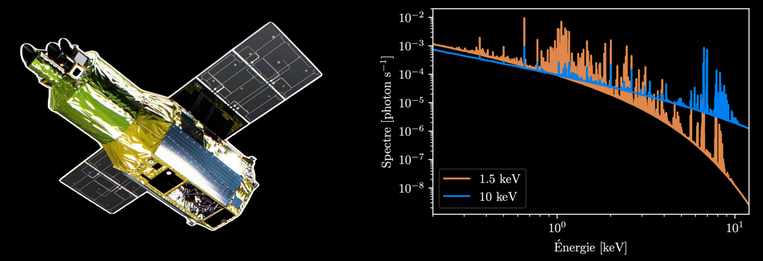
Observed data

Spectra measured by the instrument



Model

Spectral model and instrumental convolution



Counting process : Poisson likelihood

$$P(X = k \mid \theta) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Turn the likelihood into a posterior distribution

Traditional Bayesian inference

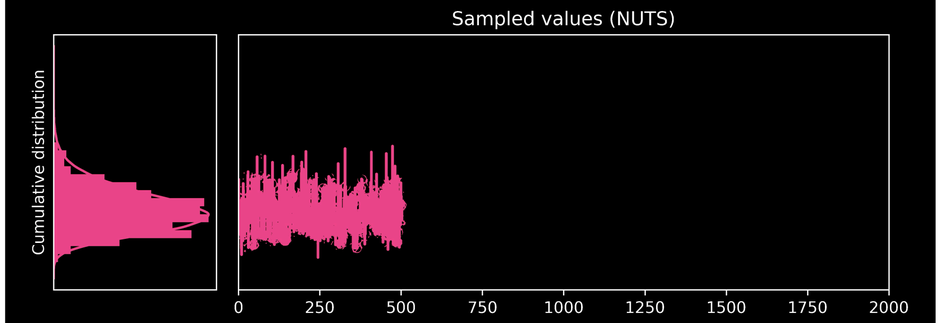
Evaluate

$$P(X \mid \theta)$$

Traditional Bayesian inference

Sampling $\{\theta\}_i \sim P(\theta|X)$

Monte Carlo Markov Chain (HMC, NUTS, AIES), Nested Sampling



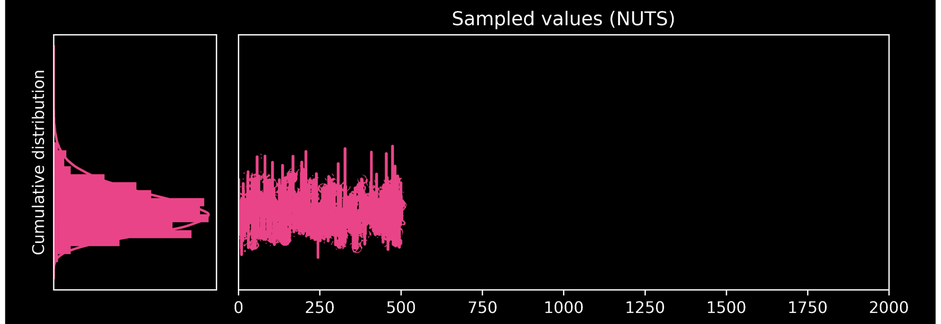
Evaluate

$P(X|\theta)$

Traditional Bayesian inference

Sampling $\{\theta\}_i \sim P(\theta|X)$

Monte Carlo Markov Chain (HMC, NUTS, AIES), Nested Sampling



Evaluate

$P(X|\theta)$

Variational $q(\theta) \simeq P(\theta|X)$

Minimize Evidence Lower Bound for a parametric and analytical approximation of the posterior distribution