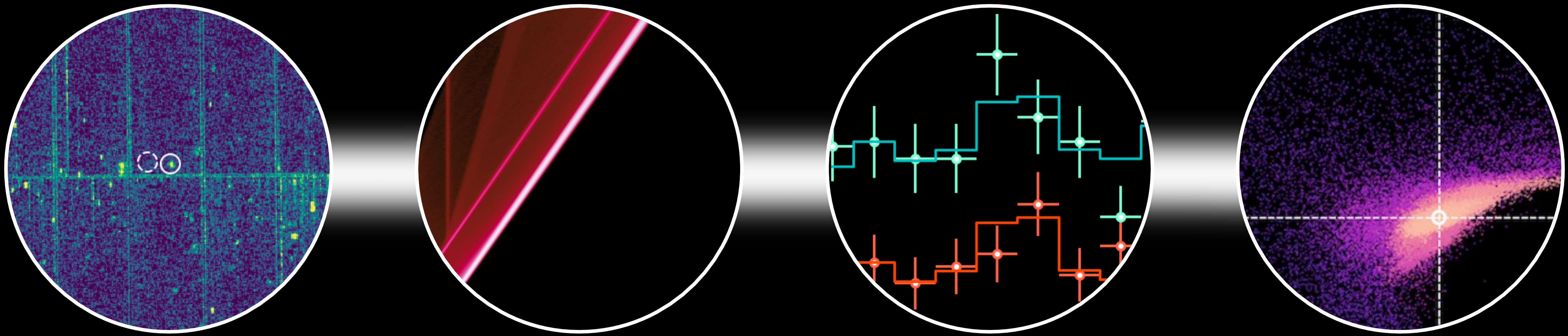
Statistical Aspects of X-ray Spectral Analysis



Johannes Buchner

jbuchner@mpe.mpg.de

astrost.at/istics

Peter Boorman

boorman@mpe.mpg.de

peterboorman.com

X-ray spectral fitting workshops

MPE 2019, lectures by JB,
J. Michael Burgess, Joern Wilms

BiD4BESt School of Astro-Statistics

Online school, 2021
Lectures by JB

CLEMSON UNIVERSITY

Workshop 2023, by PB

Chandra Data Science, online 2021, lectures by JB, PB

XSF workshop, Prague 2022, PB, JB

SILESIAN UNIVERSITY IN OPAVA

Workshop 2021,
Lectures by PB

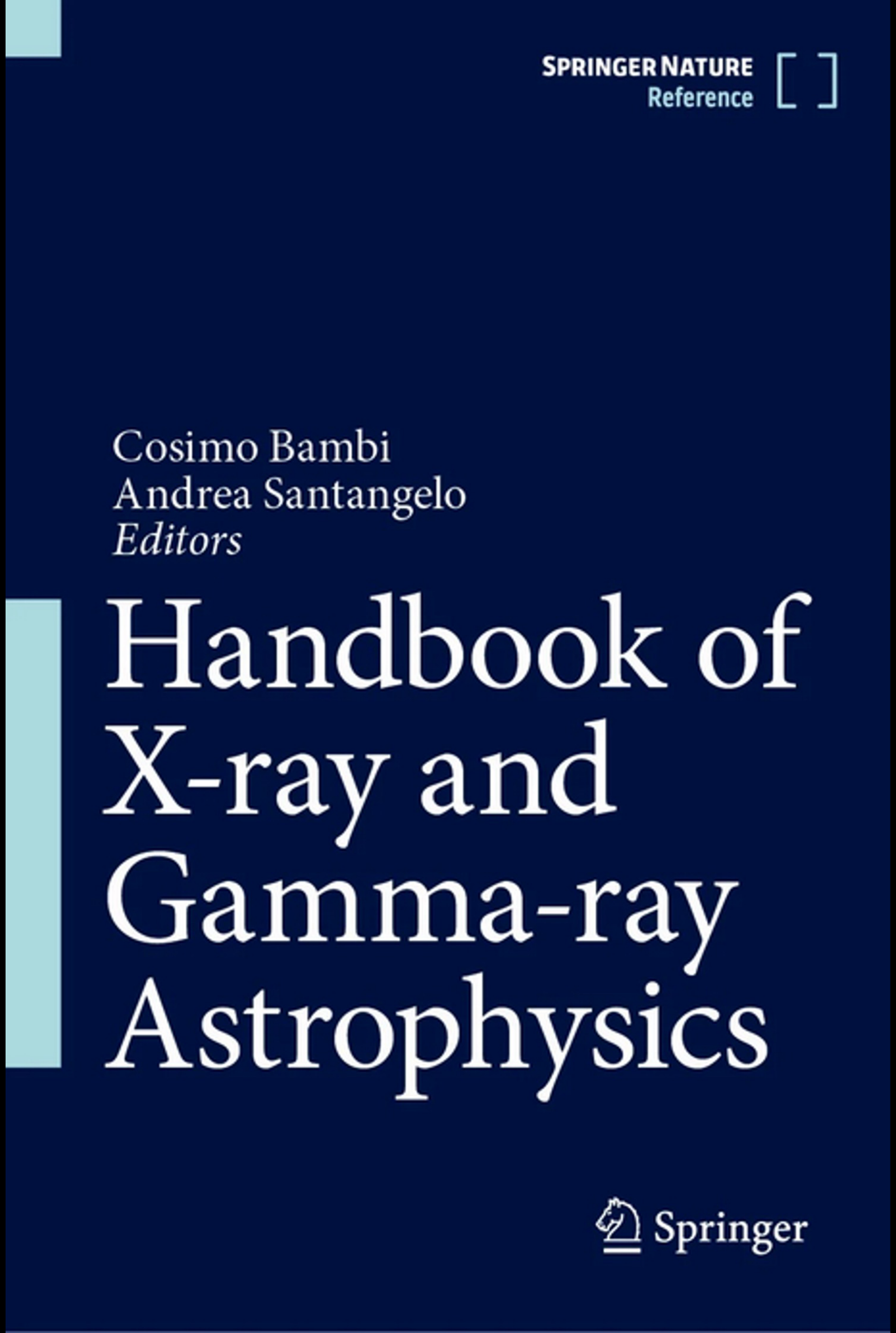
Book chapter

Statistical Aspects of X-ray Spectral Analysis

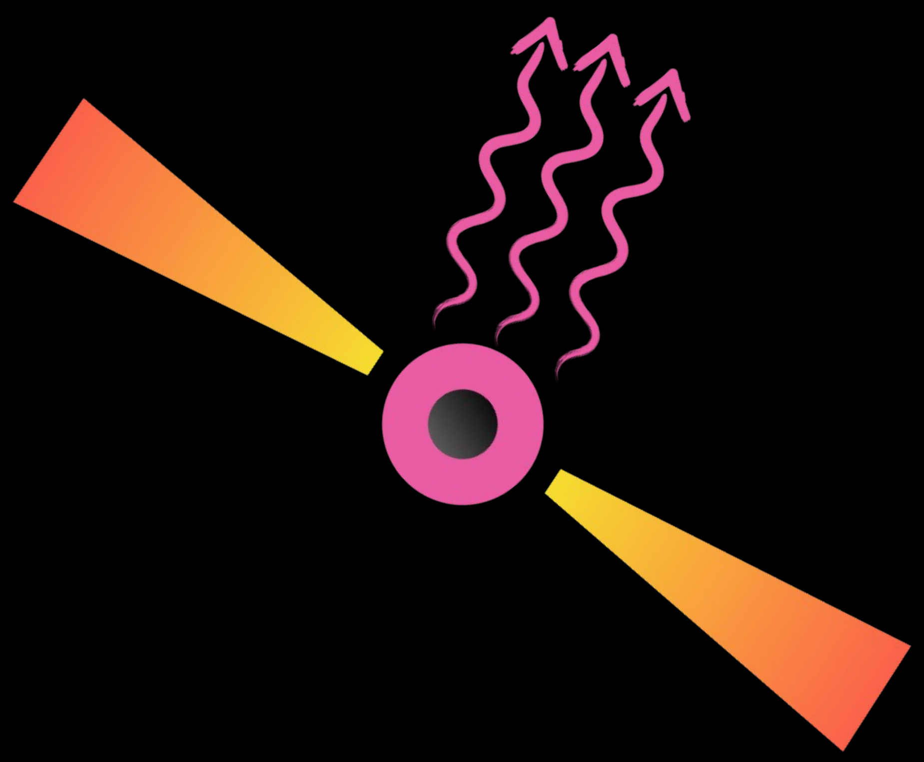
Johannes Buchner & Peter Boorman

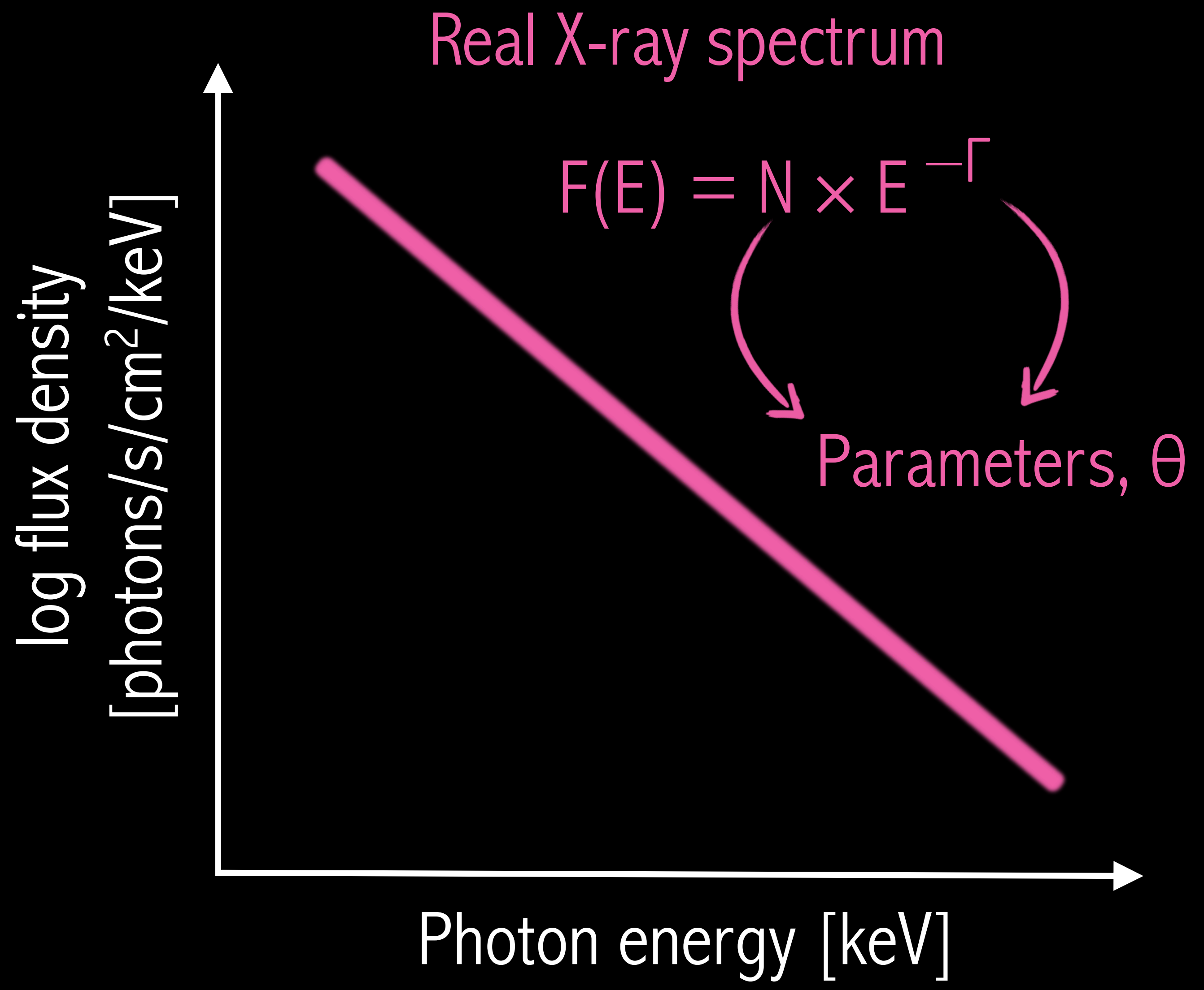
Freely available at:
https://arxiv.org/abs/2309.05705

Includes hands-on exercises for both Sherpa & Xspec

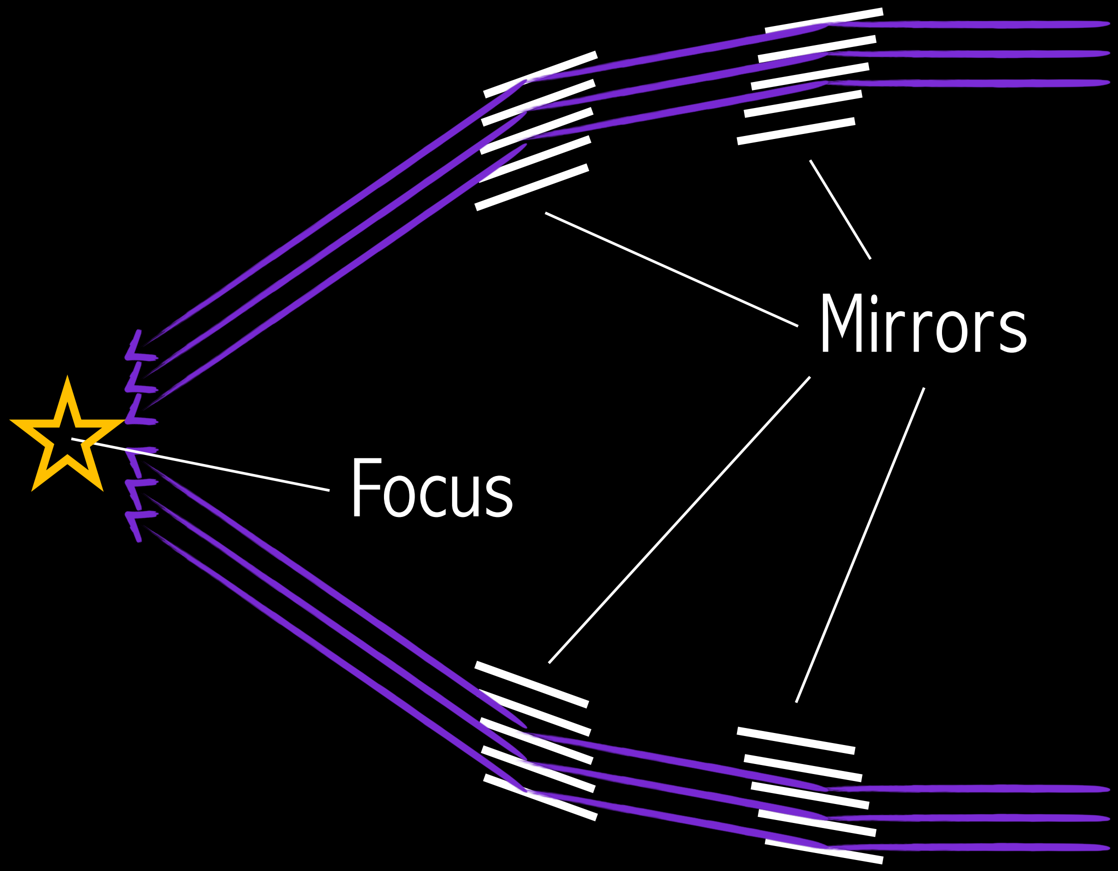


An interesting astrophysical source

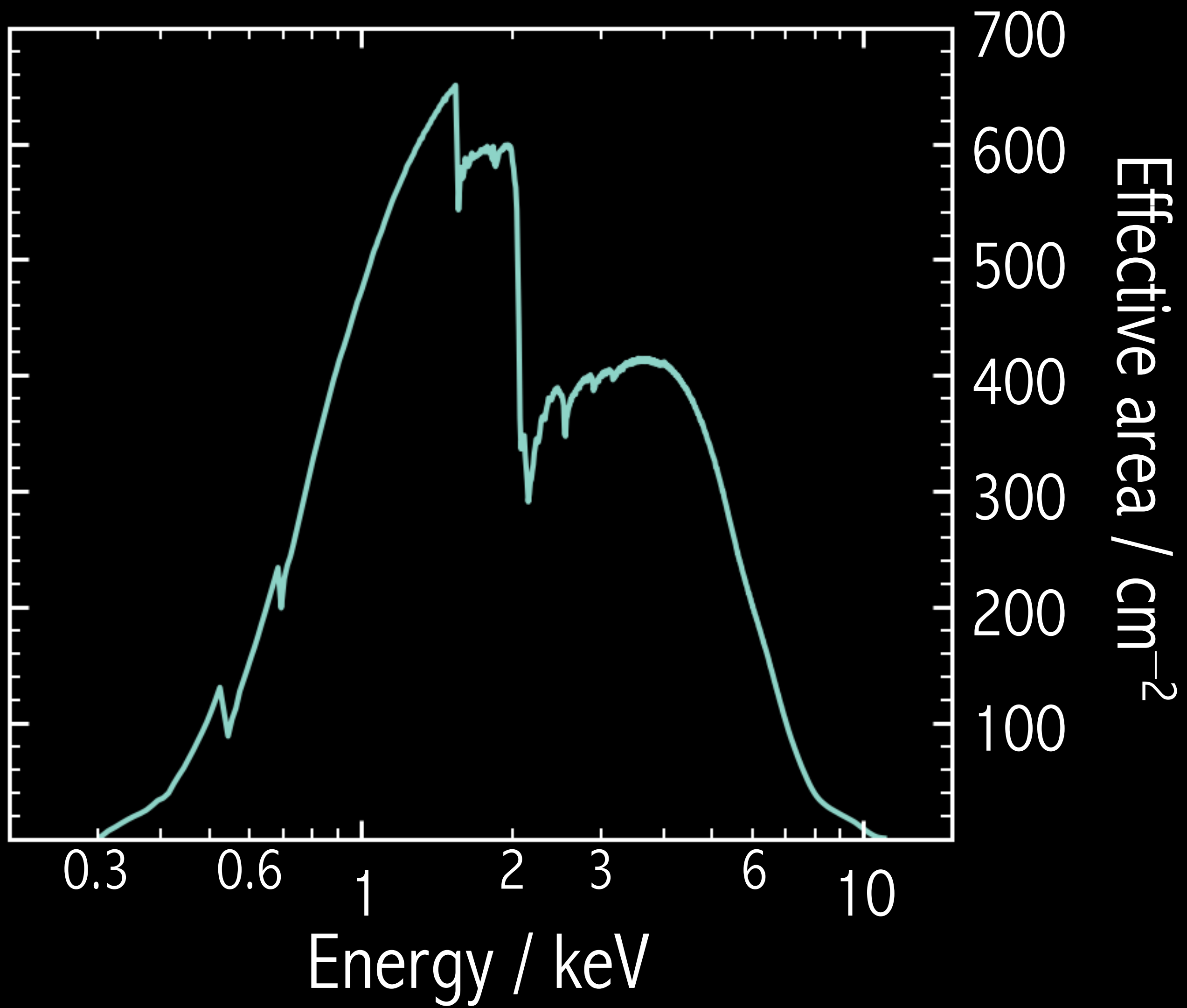




Focus X-ray photons onto detector with mirrors



Paraboloid + hyperboloid
Wolter+52

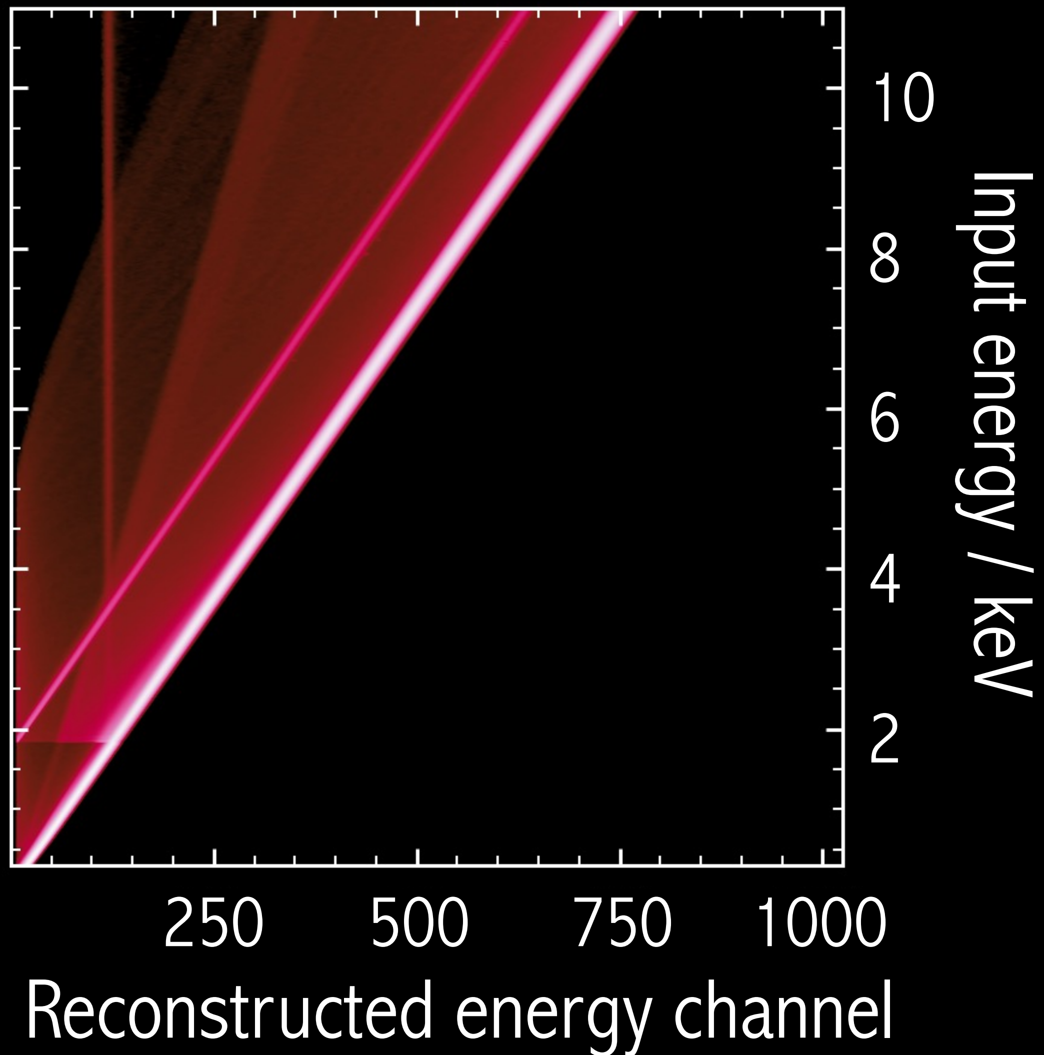


Detector current converts energy into channels

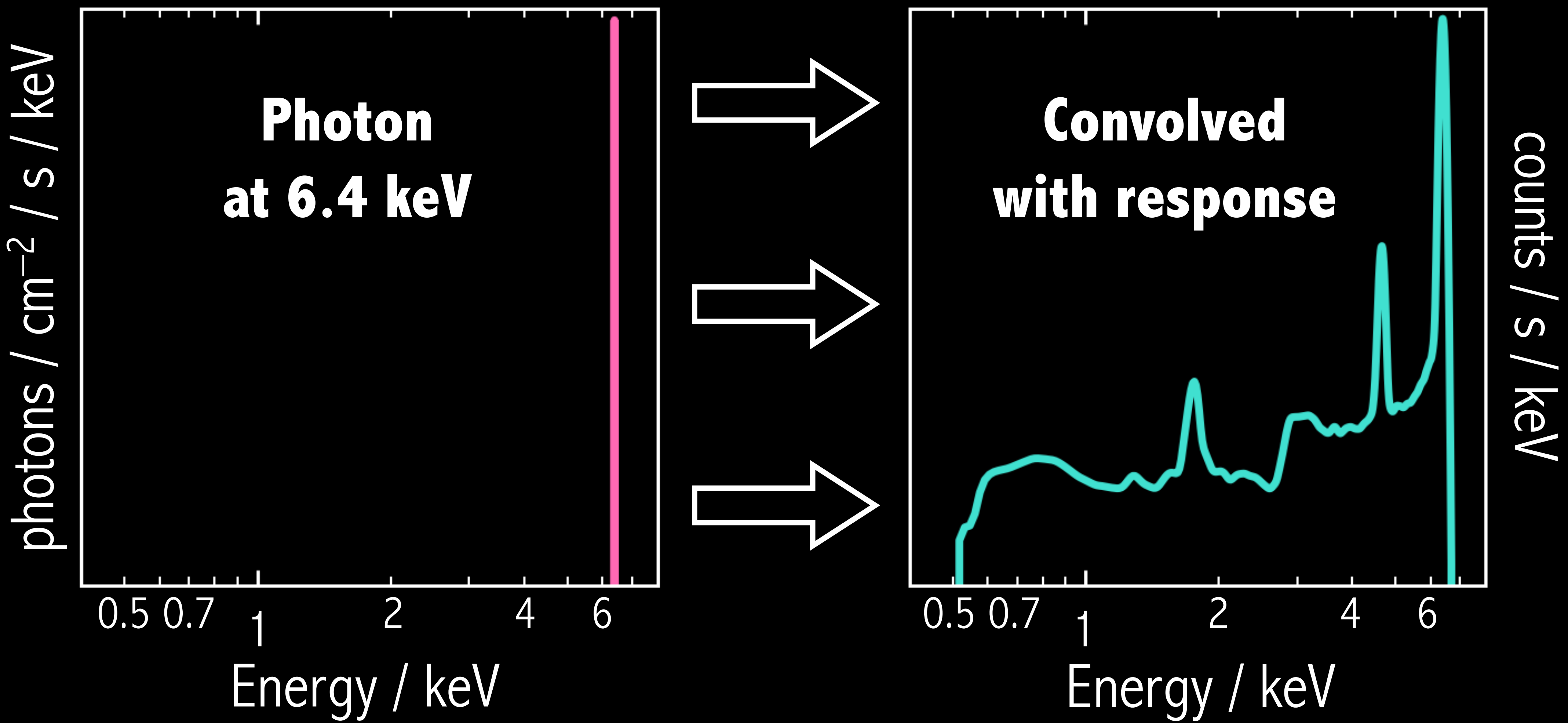
- Detect current proportional to energy of incoming photon

- Diagonal would be a perfect detector

- Secondary effect from incoming photon ionising part of the detector

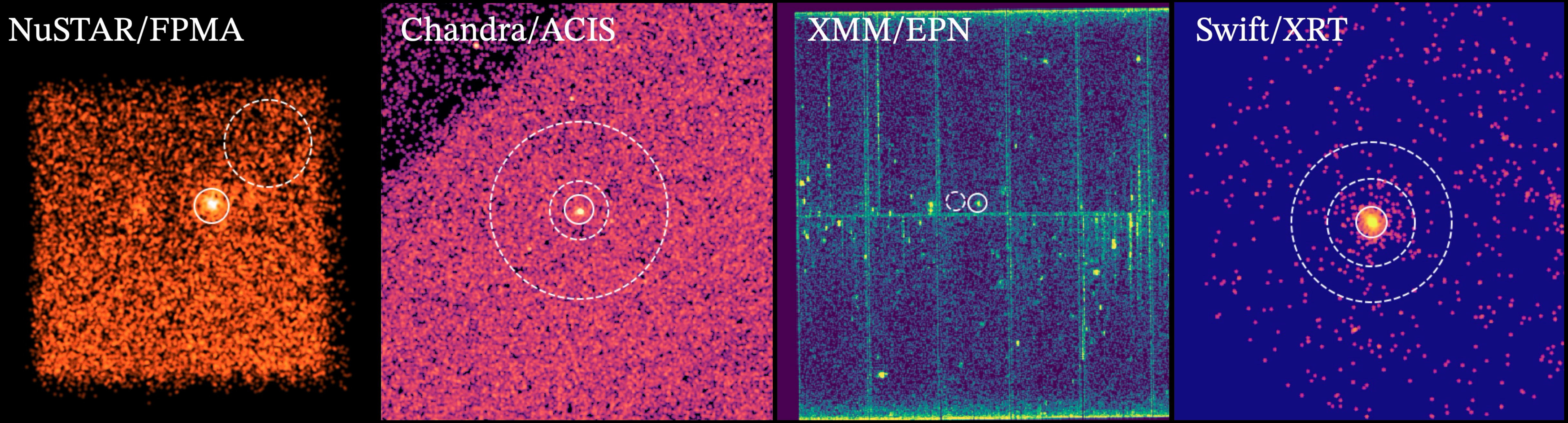


A single photon at 6.4 keV converted to counts



Added complications from background

Buchner & Boorman 23

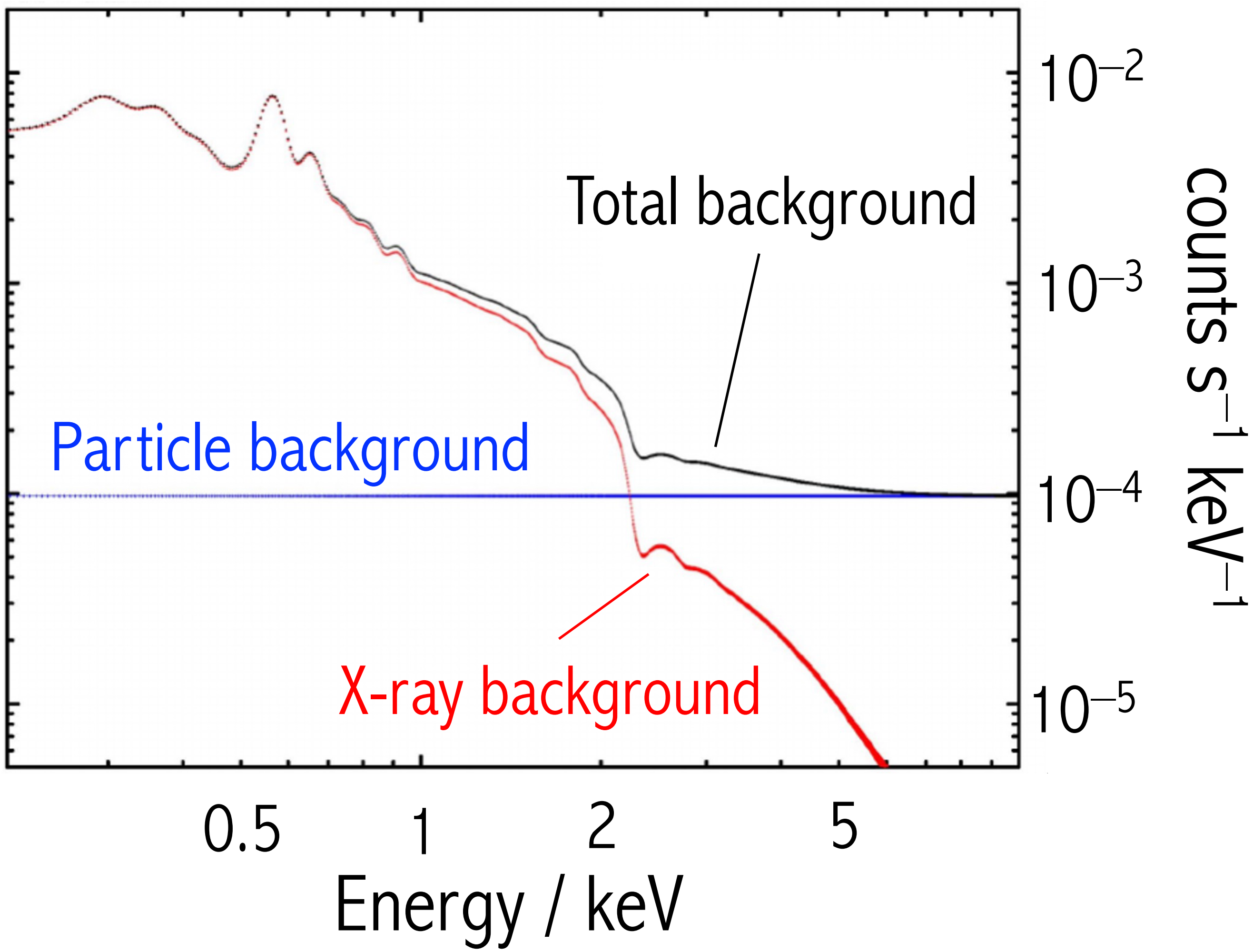


“on” region (source + background) & “off” region (background only)

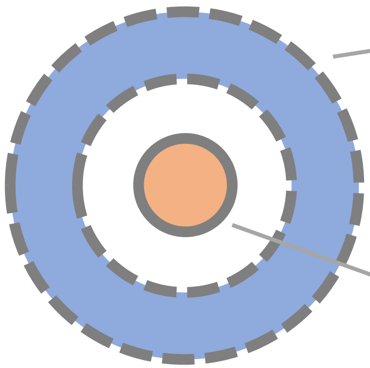
Example: eROSITA background

wiki.mpe.mpg.de/eRosita/ScienceRelatedStuff/Background

- **Diffuse emission**
  - Local hot bubble
  - Galactic disk
  - Galactic halo
- **Cosmic background**
  - Unresolved AGN
- **High-energy particle background**

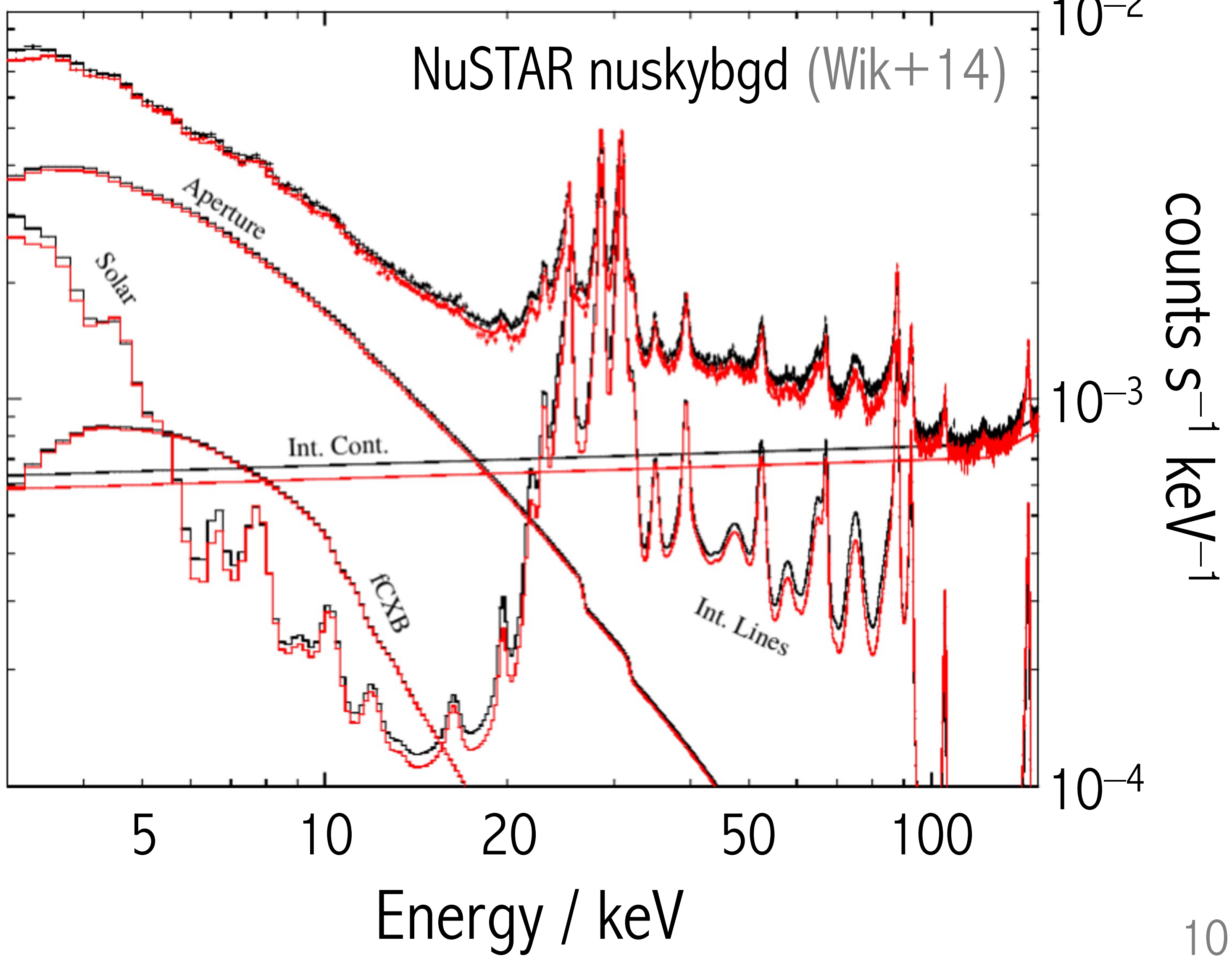


(Semi-)physical background models

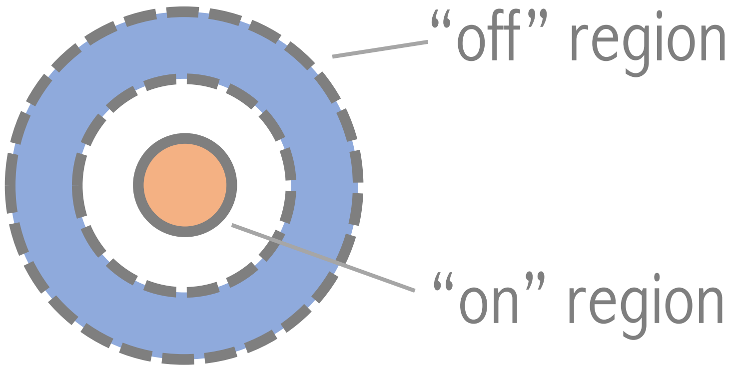


Particle background
Cosmic background
Instrumental background
…
Location & time-dependent
⟶ especially important for
extended sources

github.com/achronal/nuskybgd-py

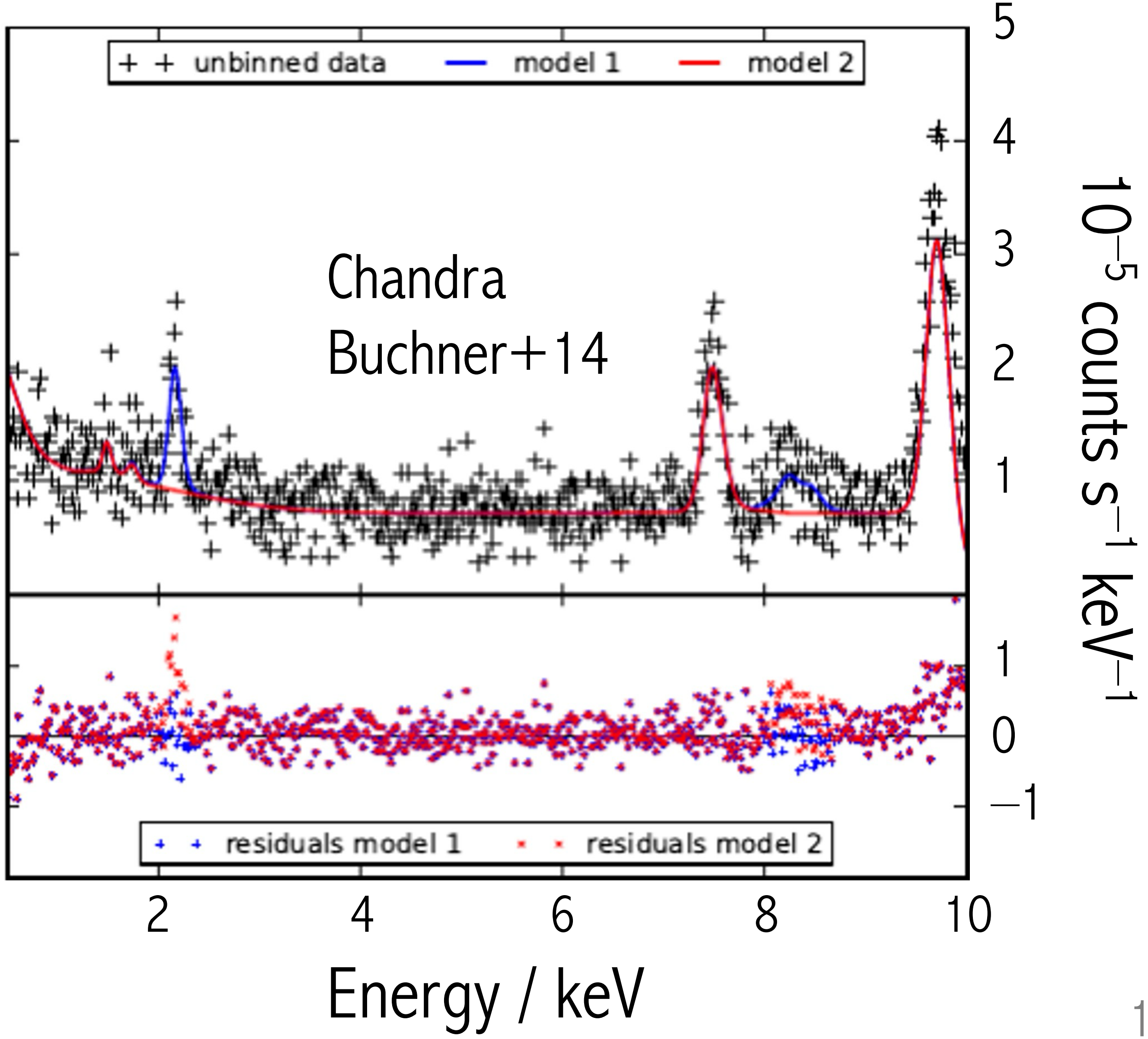


Empirical parametric background models

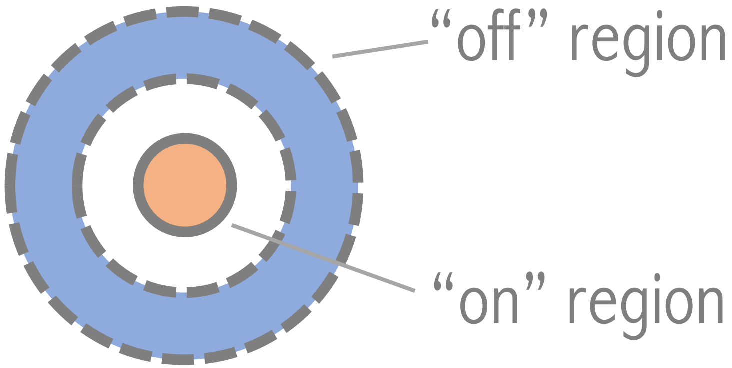


Pros
- Can contain physical knowledge & smoothness
- Small uncertainties
- 0 bin counts ok

Cons
- Need to specify model
- Fit can be poor



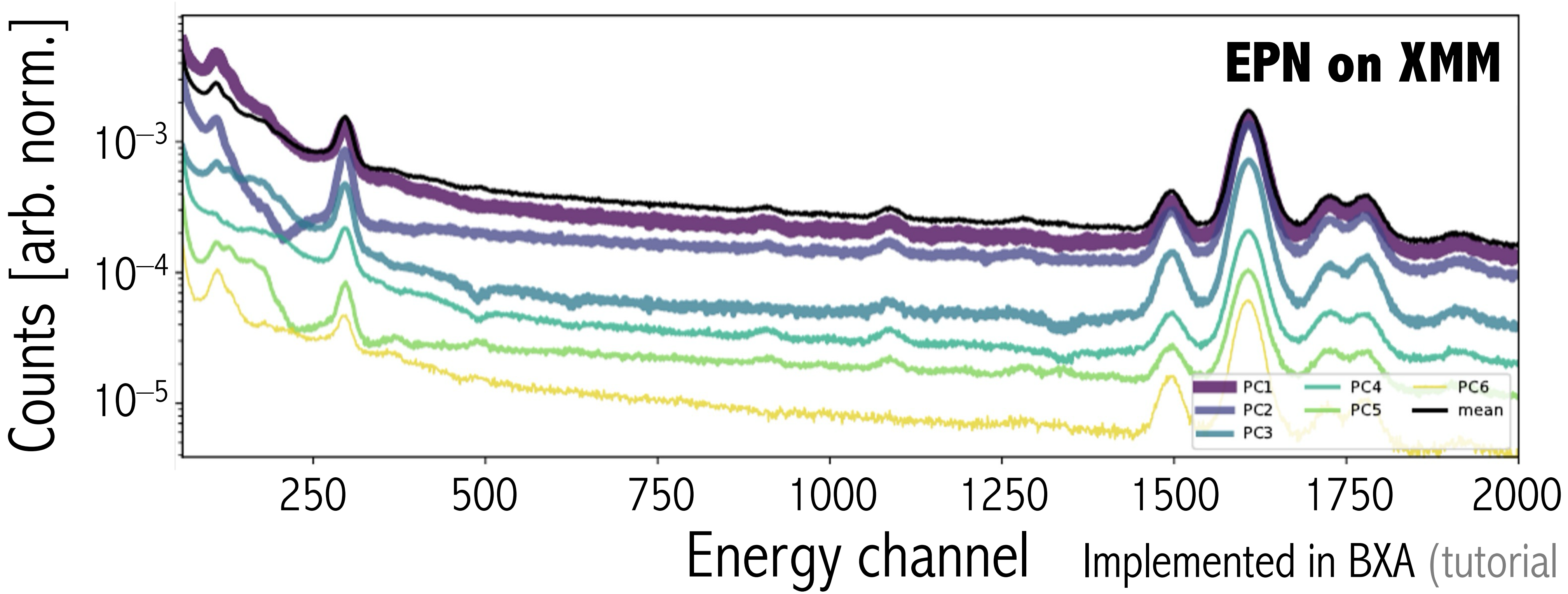
Empirical non-parametric background models (PCA)



Automated shape finding

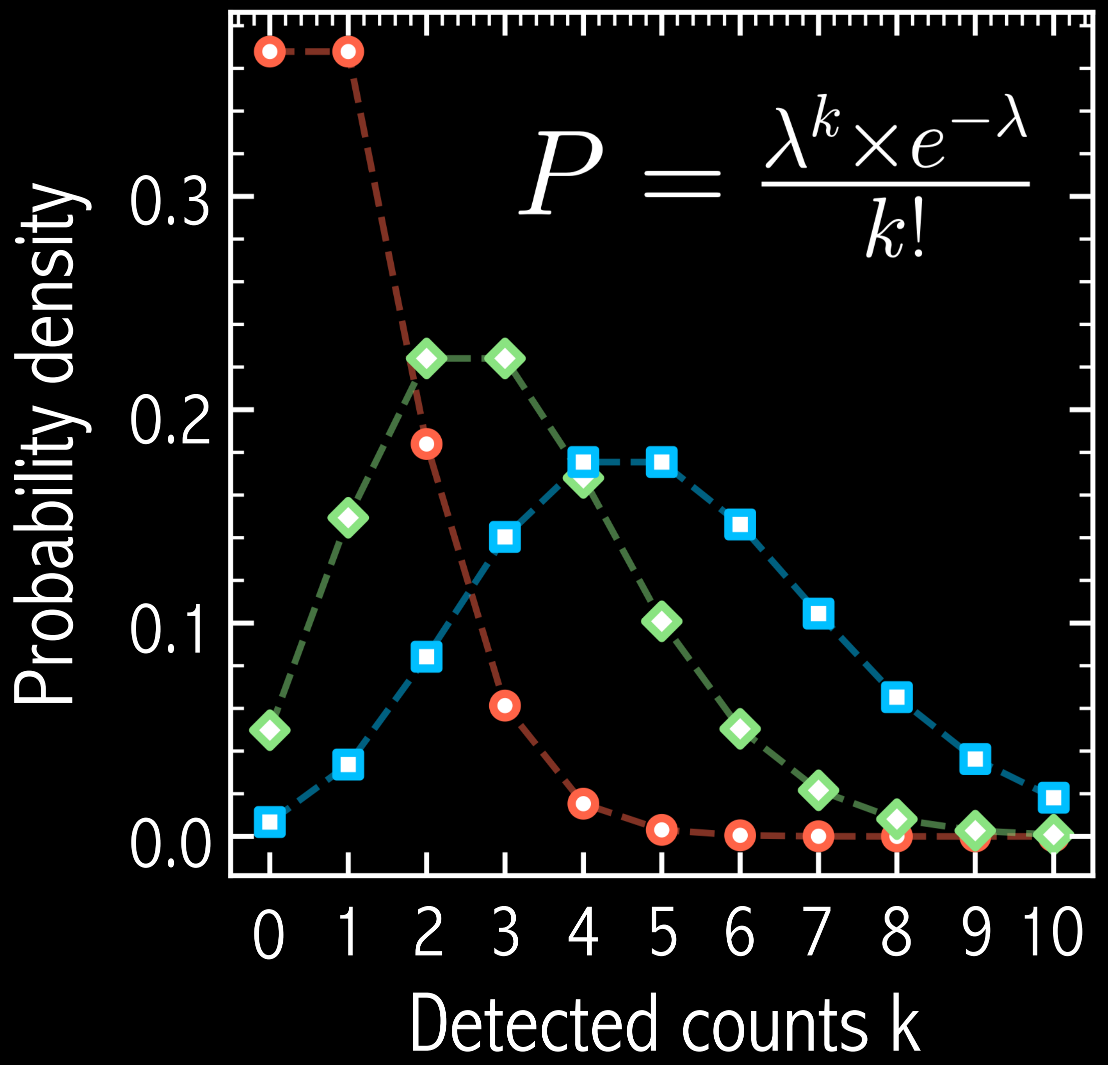
Simmonds, Buchner et al., (2017)

Includes XMM/PN, XMM/MOS, Chandra/ACIS, NuSTAR, Suzaku, RXTE, Swift/XRT



Implemented in BXA (tutorial Exercise 1.6)

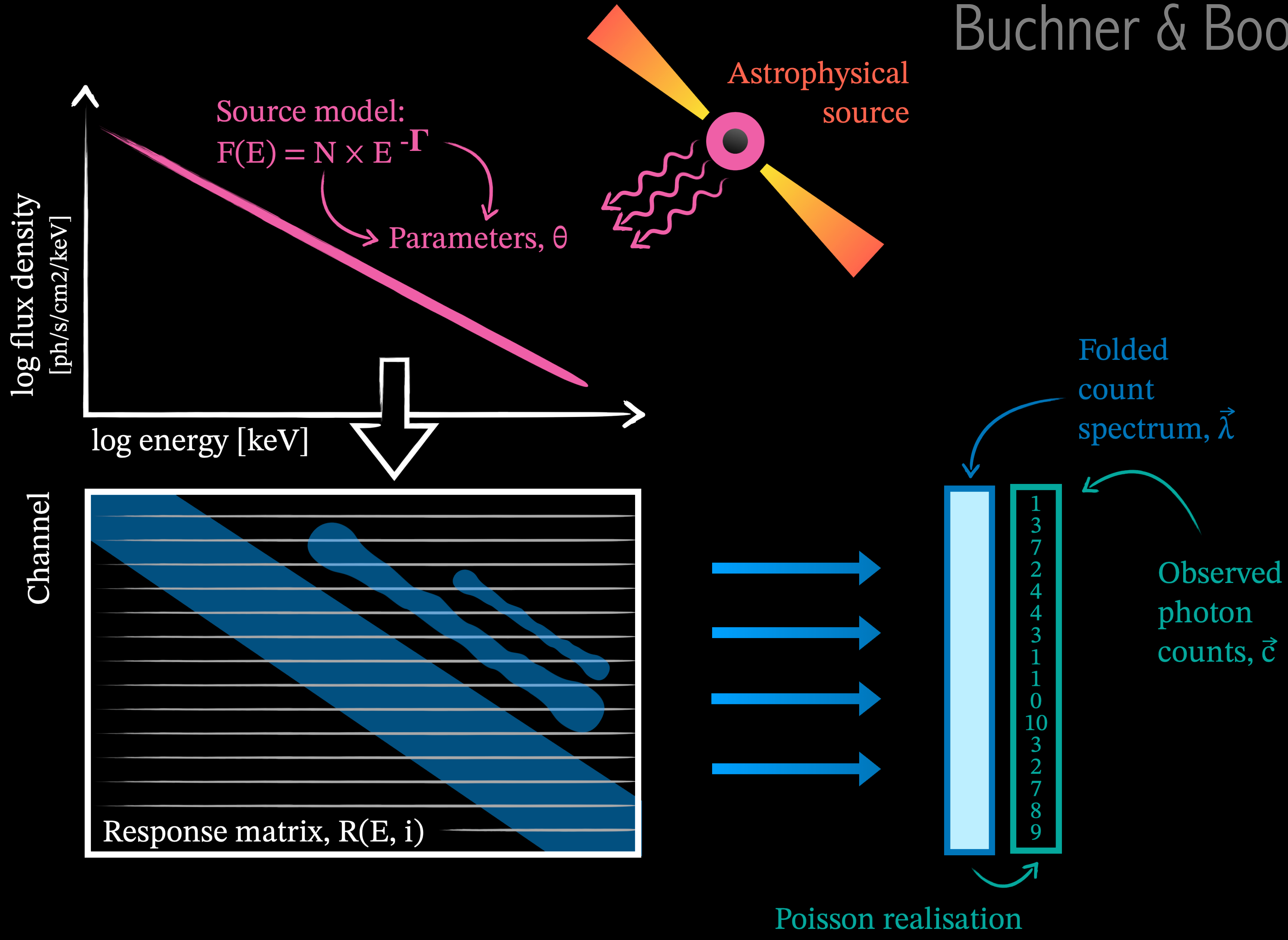
We detect a Poisson realisation of the count spectrum



Poisson distribution

- Detected counts k, integer
- Expected counts λ, real
- Asymmetric
- Non-negative

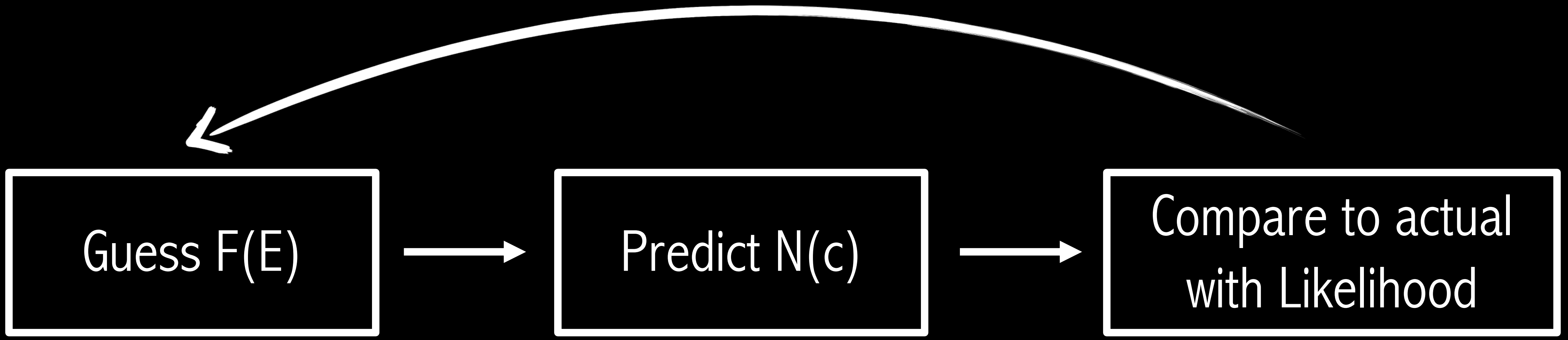
Likelihood is a probability distribution of the data


Astrophysical
source
Source model:
F(E) = N × E -Γ
Parameters, θ
log flux density
[ph/s/cm2/keV]
log energy [keV]
Channel
Response matrix, R(E, i)
Folded
count
spectrum, λ⃗
Observed
photon
counts, c⃗
1
3
7
2
4
4
3
1
1
0
10
3
2
7
8
9
Poisson realisation

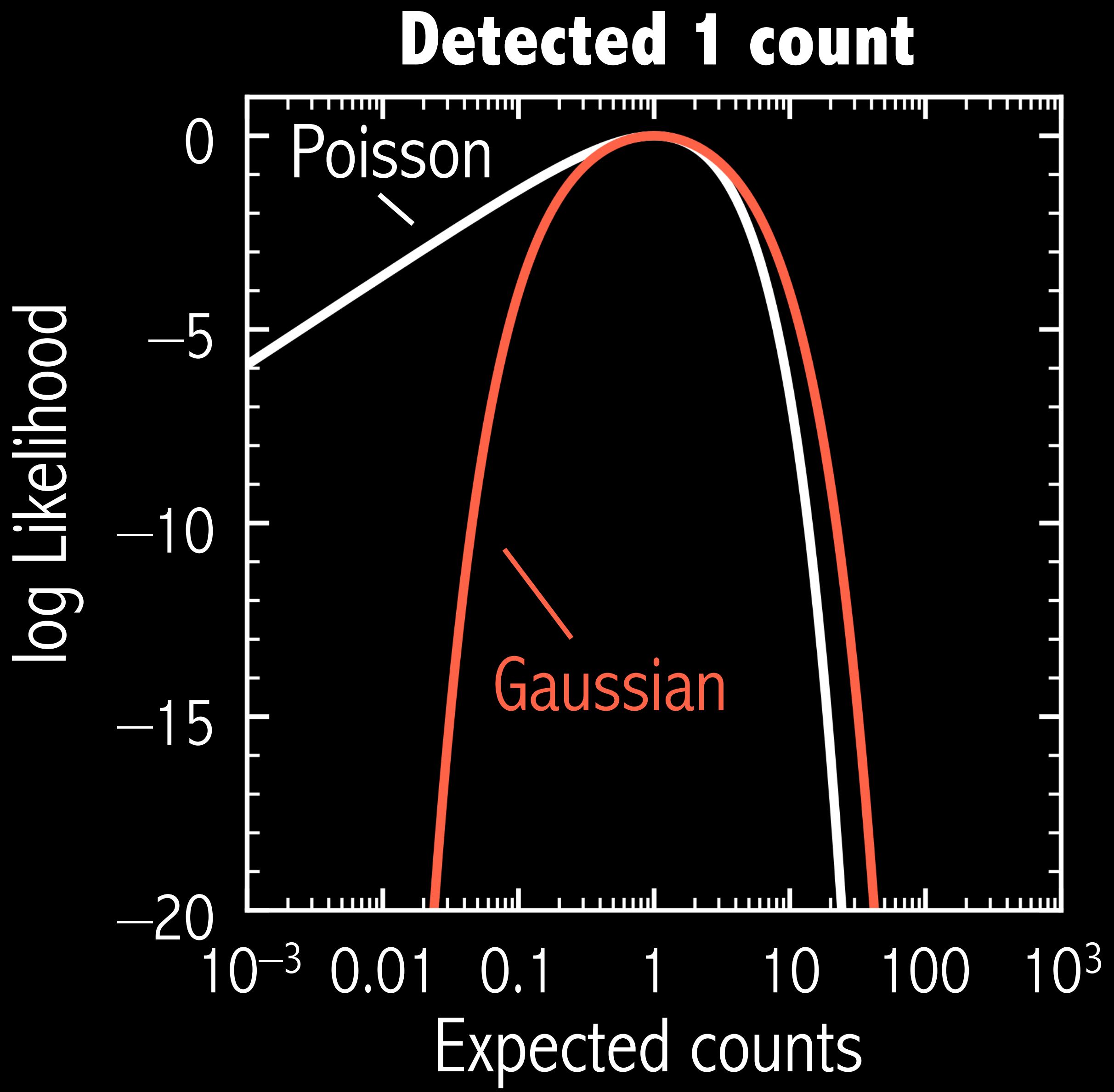
X-ray spectral fitting with forward folding

Detected count rate — Response — Effective area — Astrophysics — background

$$N(c) = \sum R(c,E) \times A(E) \times F(E)dE + b(c)$$



Likelihood shapes (fit statistic = $-2\ \log\mathcal{L}$)



Gaussian high end more permissive

$$\mathcal{L}(k|\lambda,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2}\left(\frac{k-\lambda_i}{\sigma}\right)^2\right\}$$

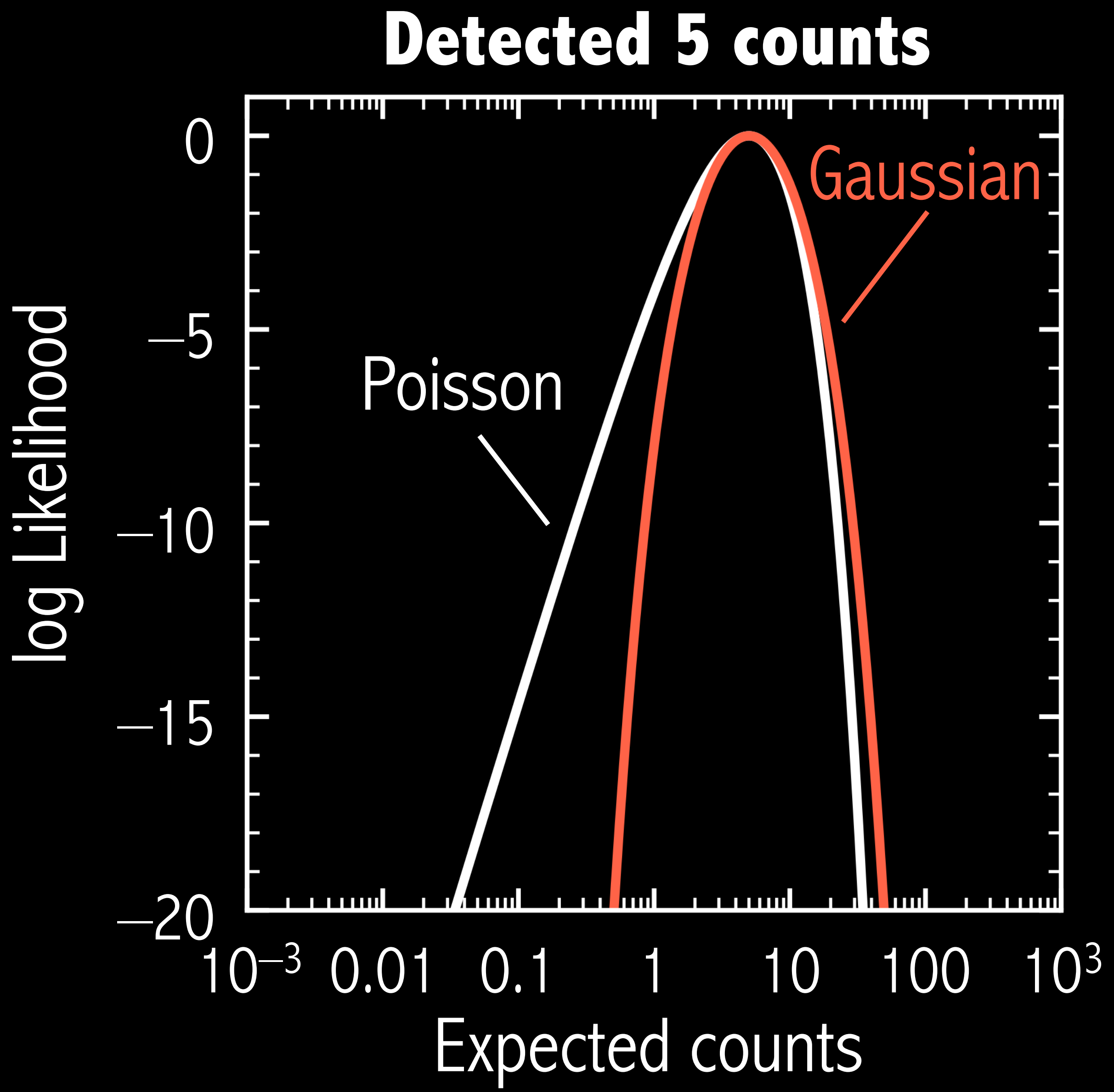
$$-2\ln\mathcal{L}(k|\lambda,\sigma) \propto (k-\lambda_i)/\sigma^2$$

Poisson low end more permissive

$$\mathcal{L}(k|\lambda) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

$$-2\ln\mathcal{L}(k|\lambda) \propto 2\lambda - 2k\ln\lambda$$

Likelihood shapes (fit statistic = $-2\ \log\mathcal{L}$)



Gaussian high end more permissive

$$\mathcal{L}(k|\lambda,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\left(\frac{k-\lambda_i}{\sigma}\right)^2\right\}$$

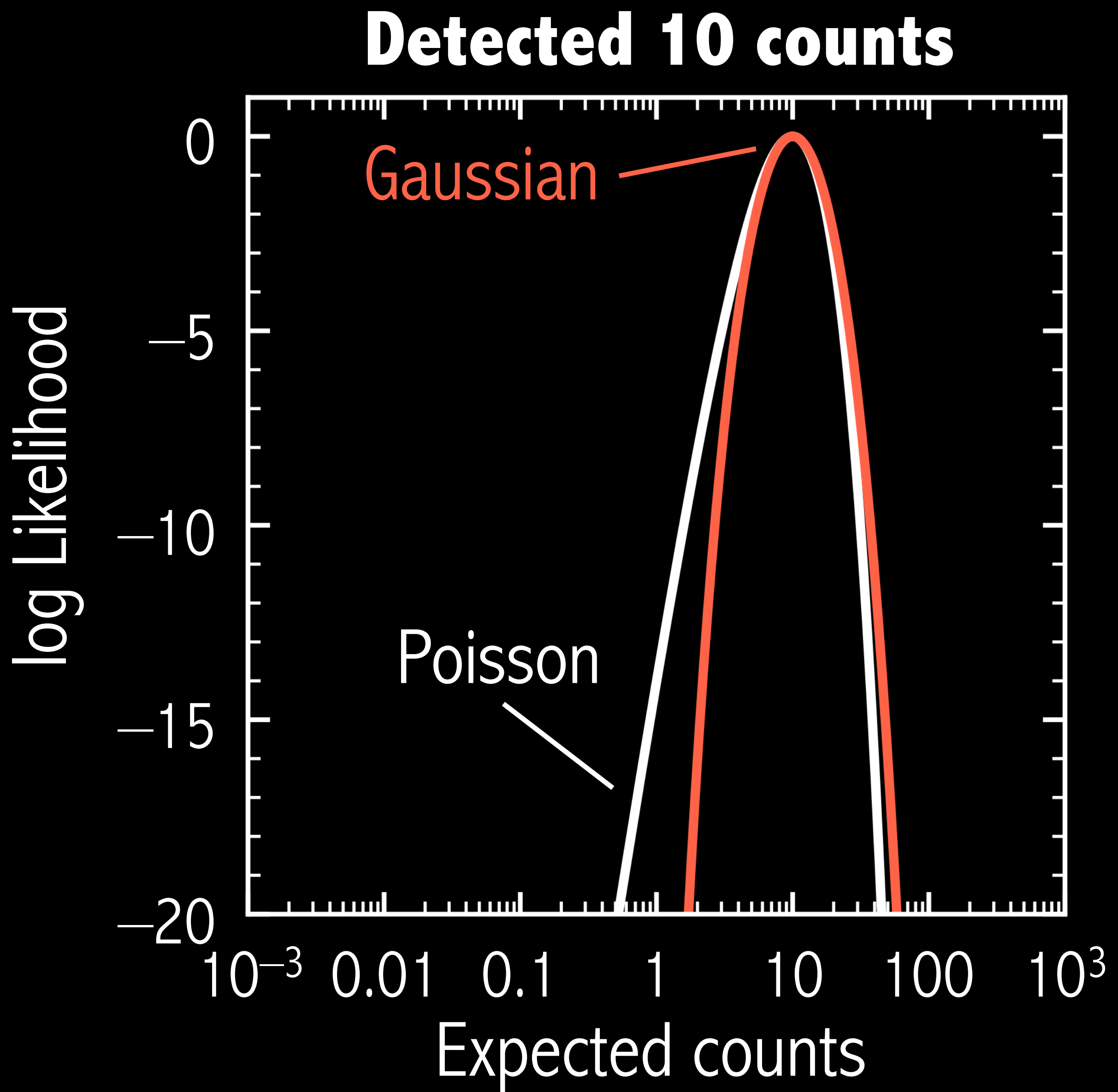
$$-2\ln\mathcal{L}(k|\lambda,\sigma) \propto (k-\lambda_i)/\sigma^2$$

Poisson low end more permissive

$$\mathcal{L}(k|\lambda) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

$$-2\ln\mathcal{L}(k|\lambda) \propto 2\lambda - 2k\ln\lambda$$

Likelihood shapes (fit statistic = $-2\ \log\mathcal{L}$)



Gaussian high end more permissive

$$\mathcal{L}(k|\lambda,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\left(\frac{k-\lambda_i}{\sigma}\right)^2\right\}$$

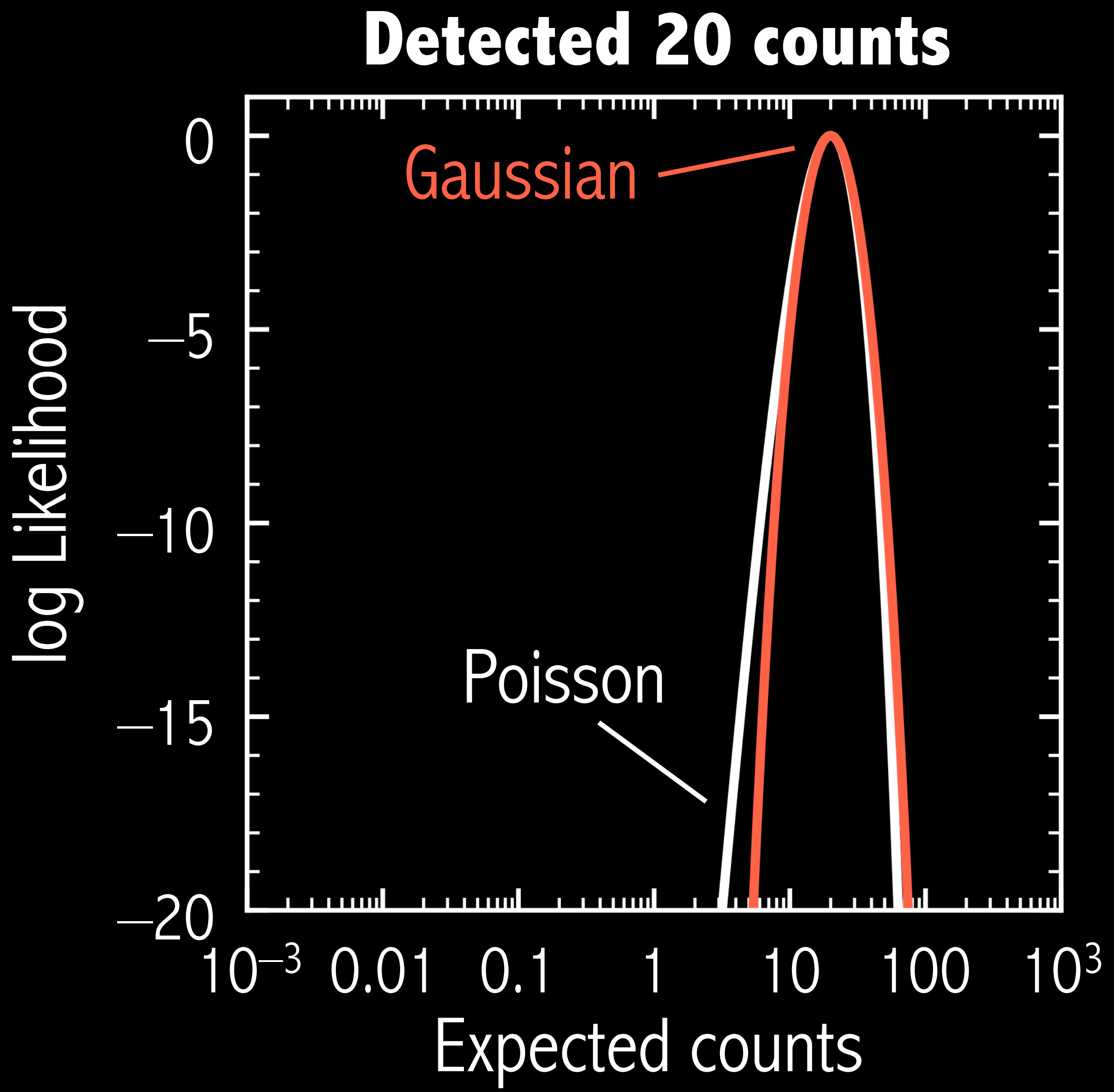
$$-2\ln\mathcal{L}(k|\lambda,\sigma) \propto (k-\lambda_i)/\sigma^2$$

Poisson low end more permissive

$$\mathcal{L}(k|\lambda) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

$$-2\ln\mathcal{L}(k|\lambda) \propto 2\lambda - 2k\ln\lambda$$

Likelihood shapes (fit statistic = $-2\ \log\mathcal{L}$)



Gaussian high end more permissive

$$\mathcal{L}(k|\lambda,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\left(\frac{k-\lambda_i}{\sigma}\right)^2\right\}$$

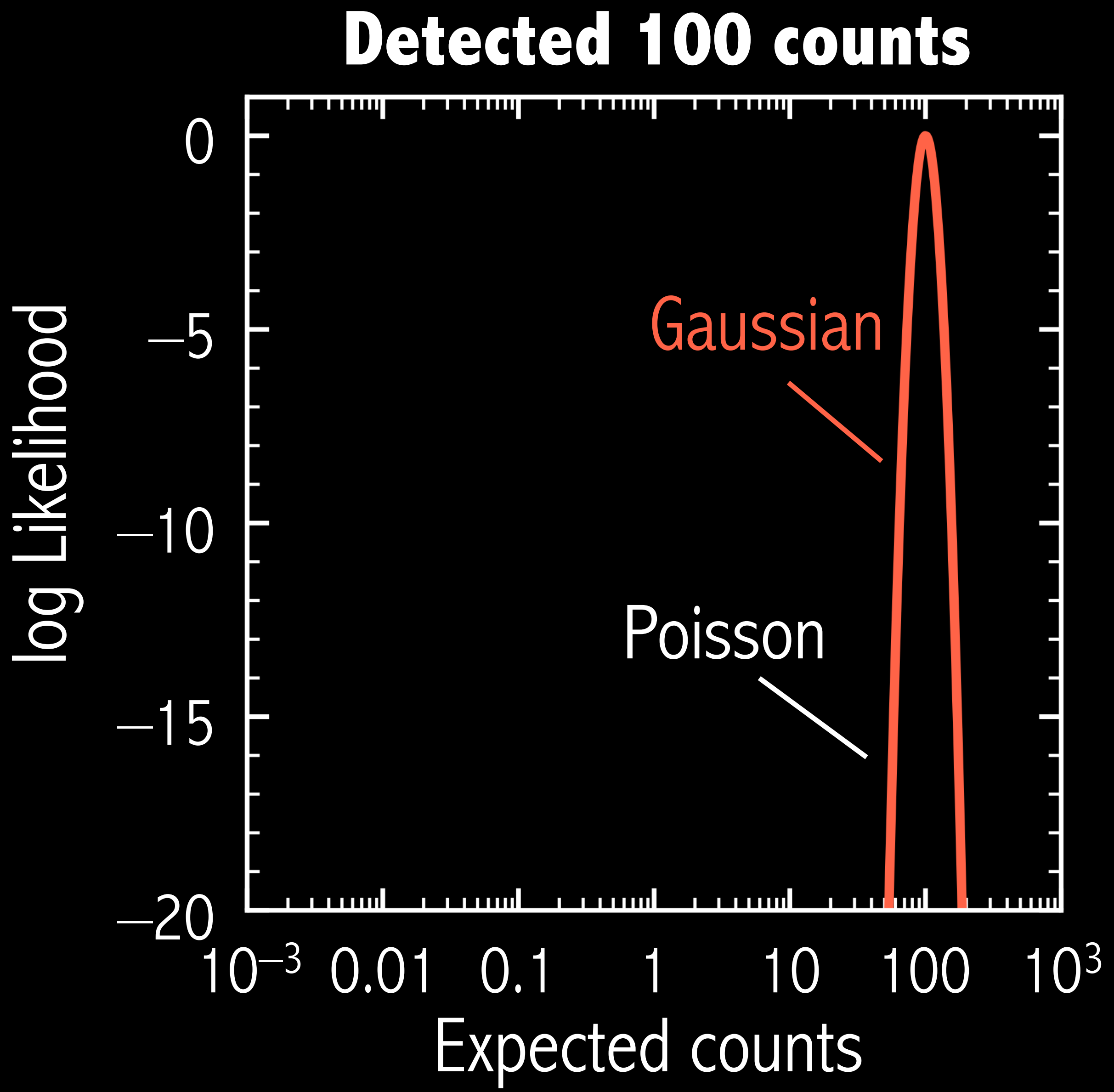
$$-2\ln\mathcal{L}(k|\lambda,\sigma) \propto (k-\lambda_i)/\sigma^2$$

Poisson low end more permissive

$$\mathcal{L}(k|\lambda) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

$$-2\ln\mathcal{L}(k|\lambda) \propto 2\lambda - 2k\ln\lambda$$

Likelihood shapes (fit statistic = −2 log$\mathcal{L}$)



Gaussian high end more permissive

$$\mathcal{L}(k|\lambda,\sigma) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2}\left(\frac{k-\lambda_i}{\sigma}\right)^2\right\}$$

$$-2\ln\mathcal{L}(k|\lambda,\sigma) \propto (k-\lambda_i)/\sigma^2$$

Poisson low end more permissive

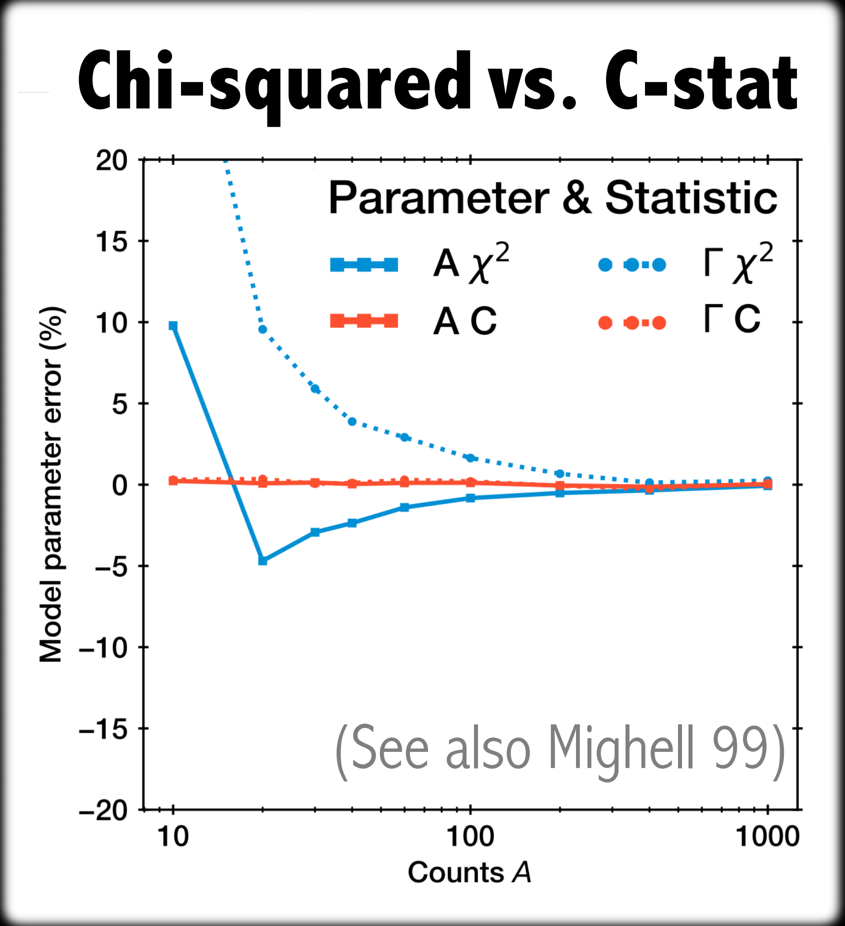
$$\mathcal{L}(k|\lambda) = \frac{\lambda^k \times e^{-\lambda}}{k!}$$

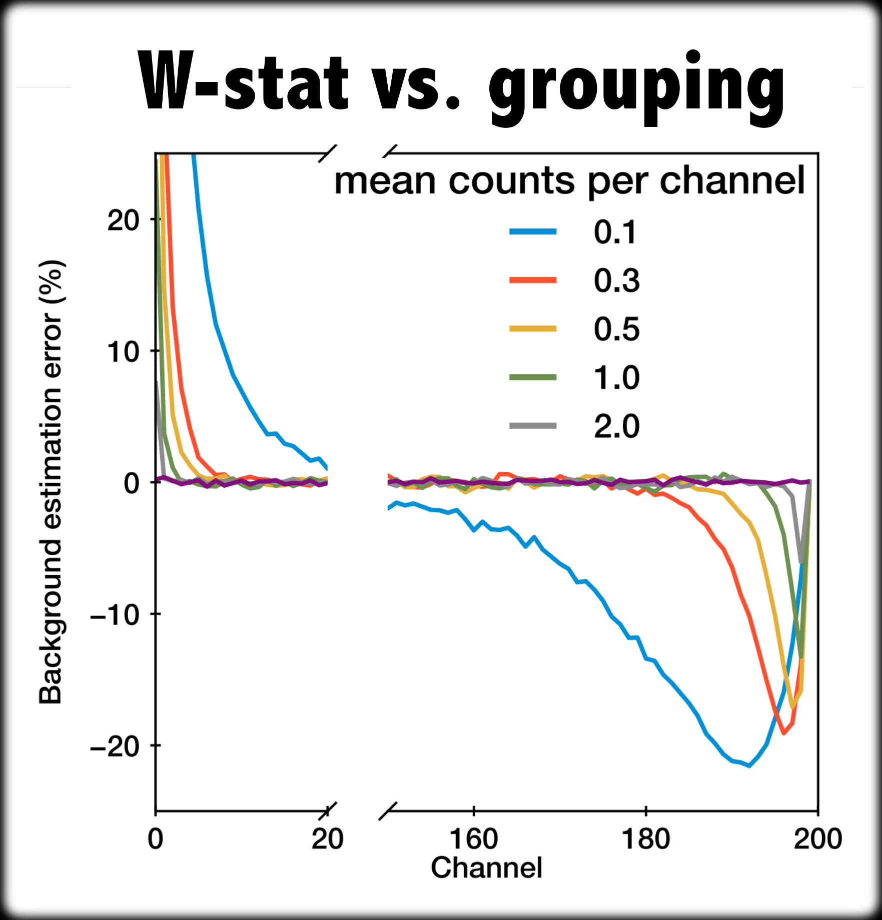
$$-2\ln\mathcal{L}(k|\lambda) \propto 2\lambda - 2k\ln\lambda$$

Chi-squared and modified C-stat (W-stat) Buchner & Boorman 23

Chi-squared is biased at **low and high counts** (e.g., Humphrey+09)

Note W-stat typically requires grouping to avoid biases!





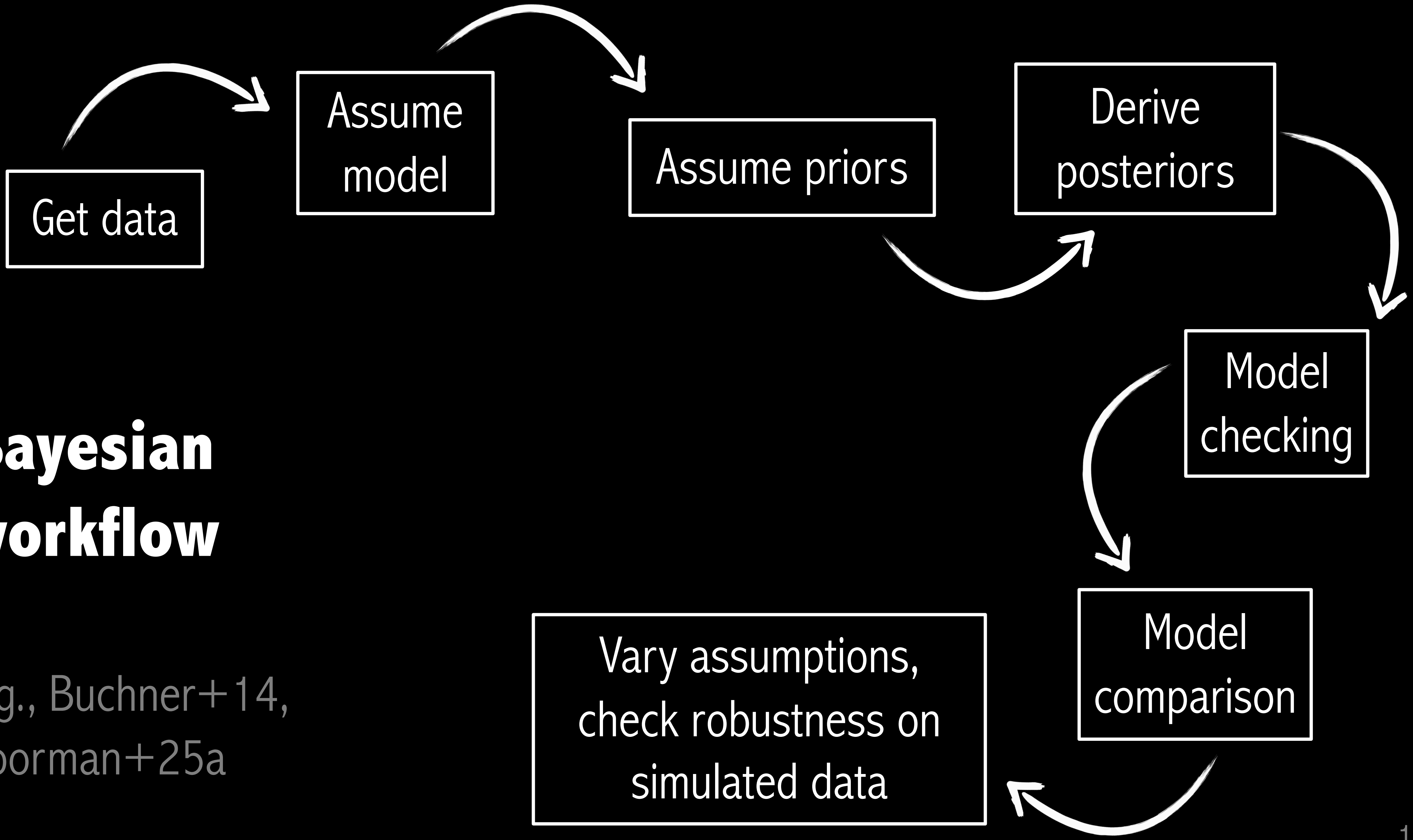
X-ray spectral fitting with forward folding

Detected count rate — Response — Effective area — Astrophysics — background

$$N(c) = \sum R(c, E) \times A(E) \times F(E)dE + b(c)$$

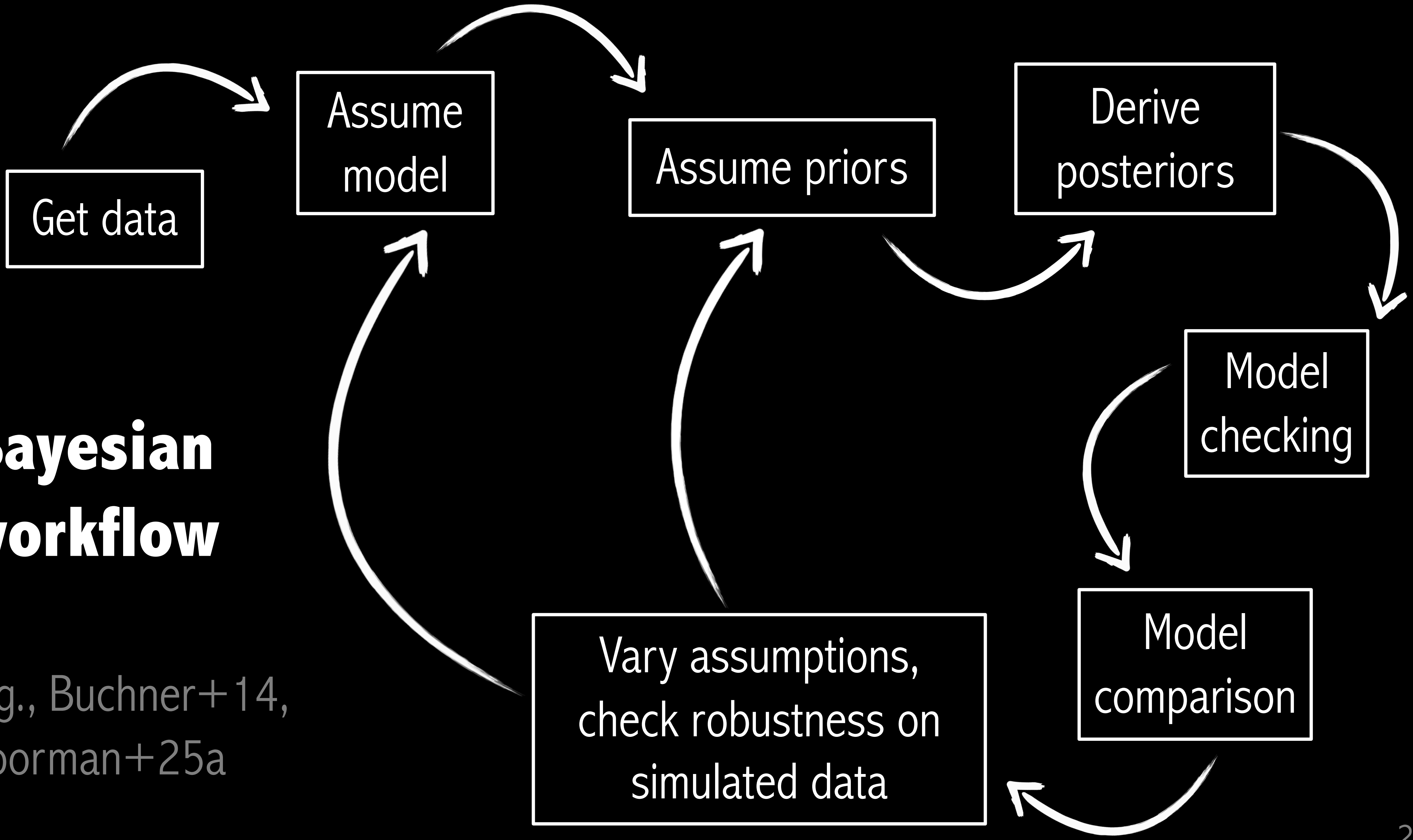
Guess F(E) → Predict N(c) → Compare to actual with Likelihood → (back to Guess F(E))

Bayesian workflow

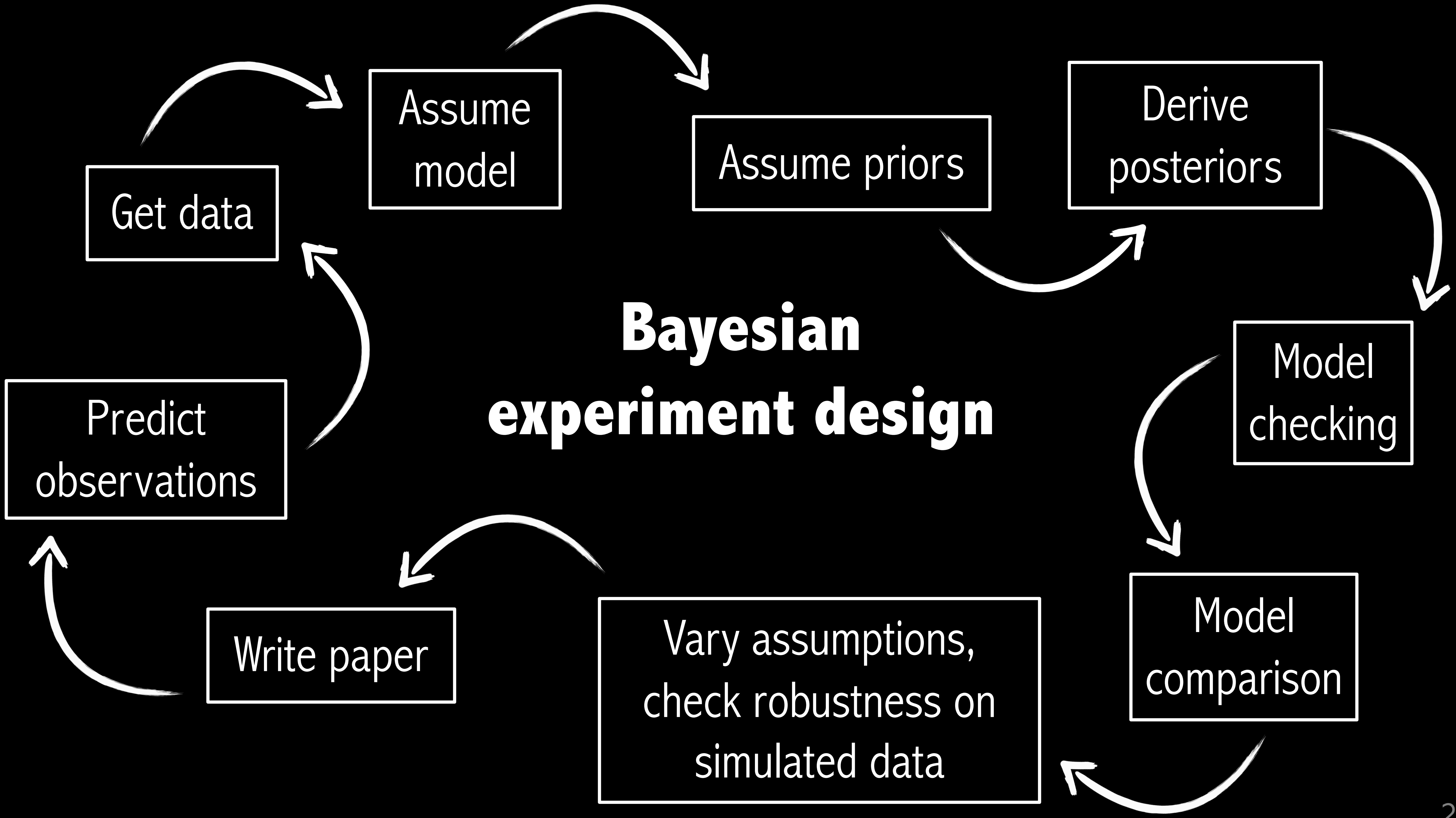


E.g., Buchner+14, Boorman+25a

Bayesian workflow

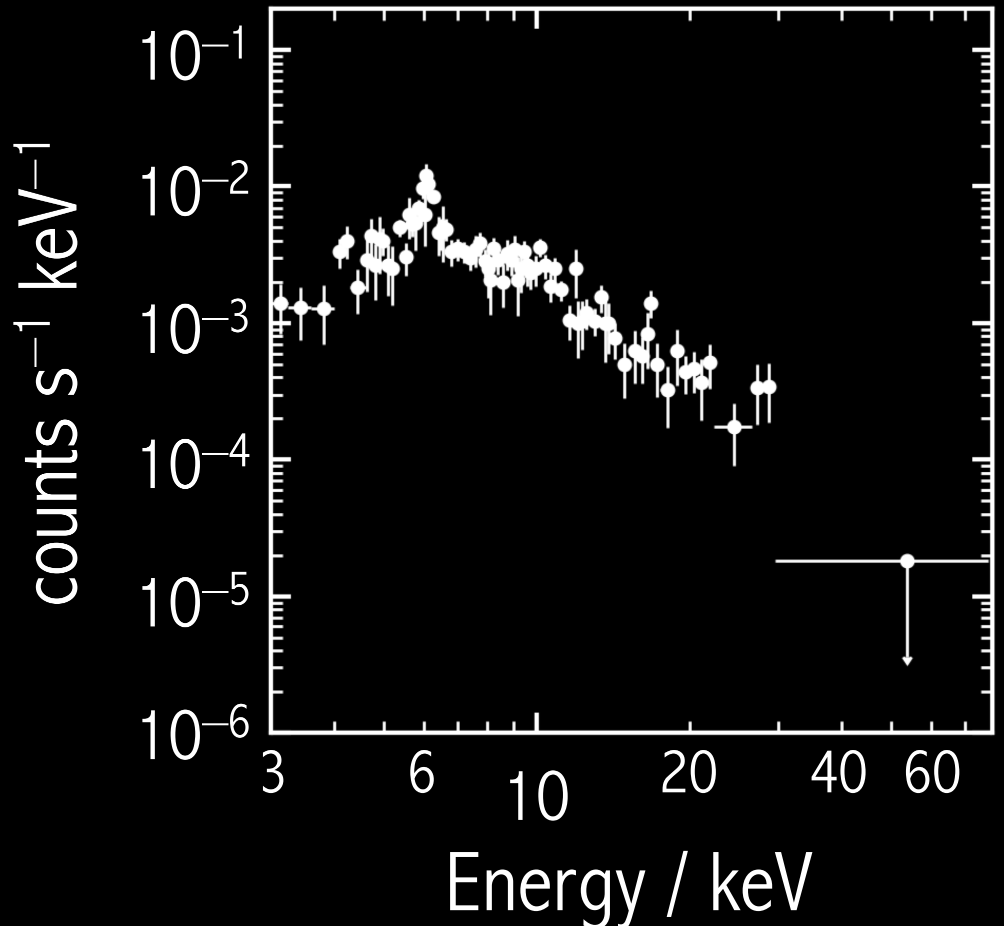


E.g., Buchner+14, Boorman+25a


Bayesian experiment design
Get data
Assume model
Assume priors
Derive posteriors
Model checking
Model comparison
Vary assumptions, check robustness on simulated data
Write paper
Predict observations

An example NuSTAR spectrum

All exercises available through tutorial: **peterboorman.com/tutorial_bxa**



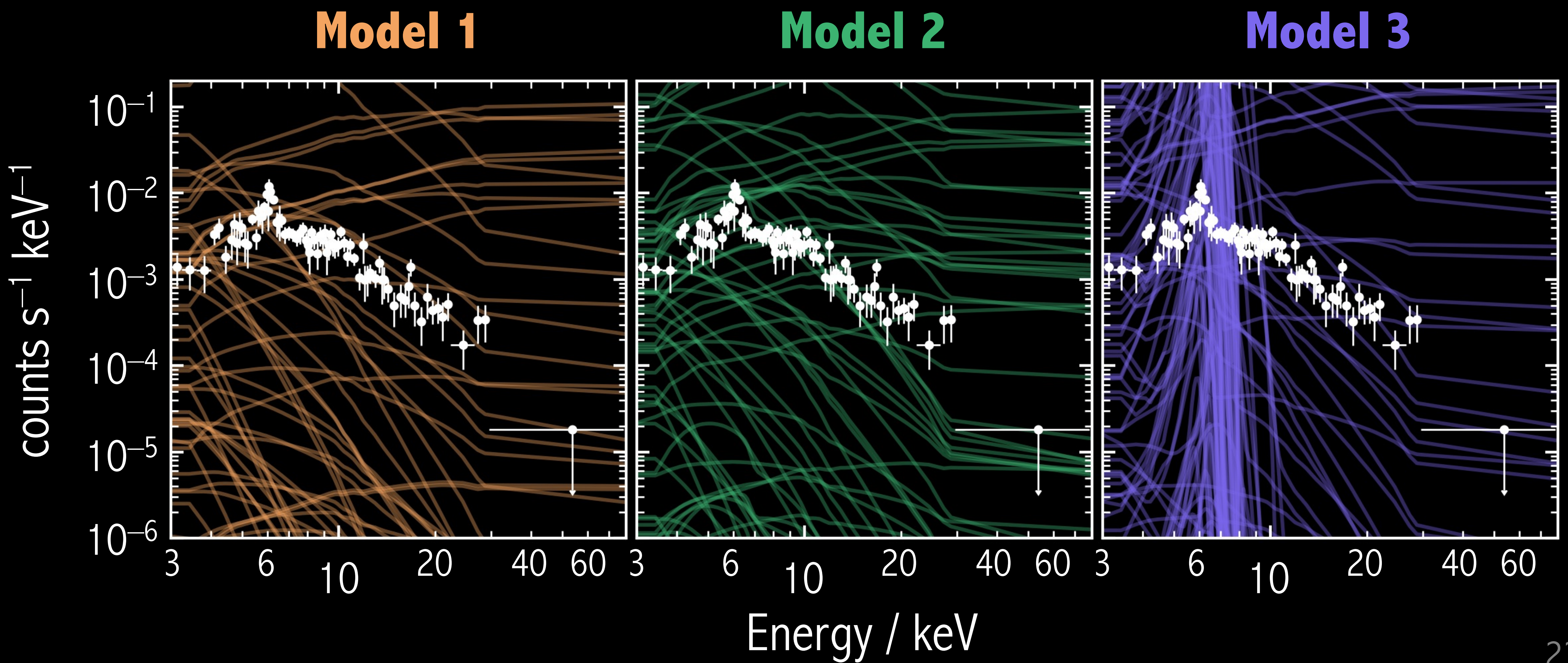
Model 1 = powerlaw

Model 2 = zTBabs * powerlaw

Model 3 = zTBabs * powerlaw + zGauss

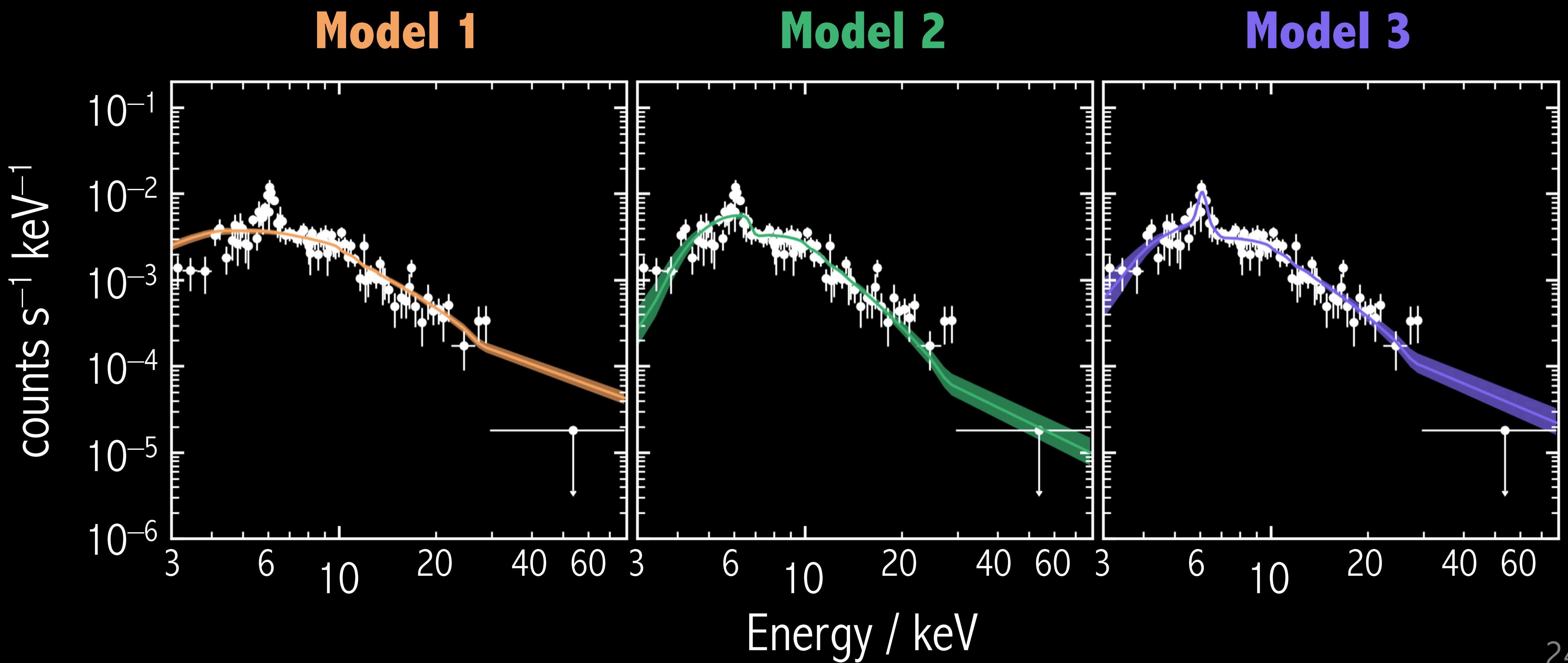
Prior predictive checks (see tutorial Exercise 1.1)

Constrain parameter priors with information **prior to the observation**



Deriving posteriors (see tutorial Exercises 1.3 & 2.1)

Using Monte Carlo sampling to learn from the data



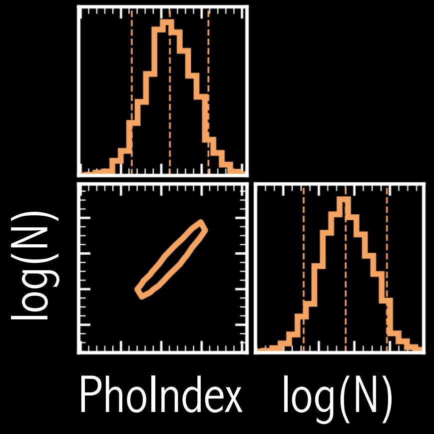
Corner plots

Marginal and conditional posterior distributions

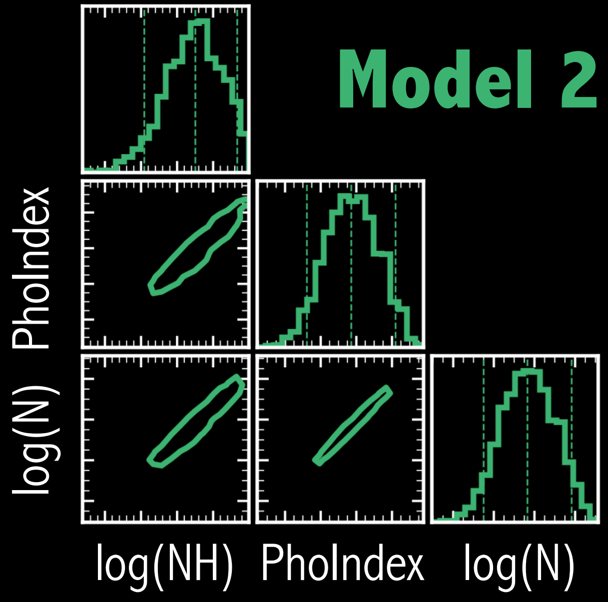
Useful for visualization the posterior

Does not provide a goodness-of-fit, nor proof that the sampling algorithm has worked

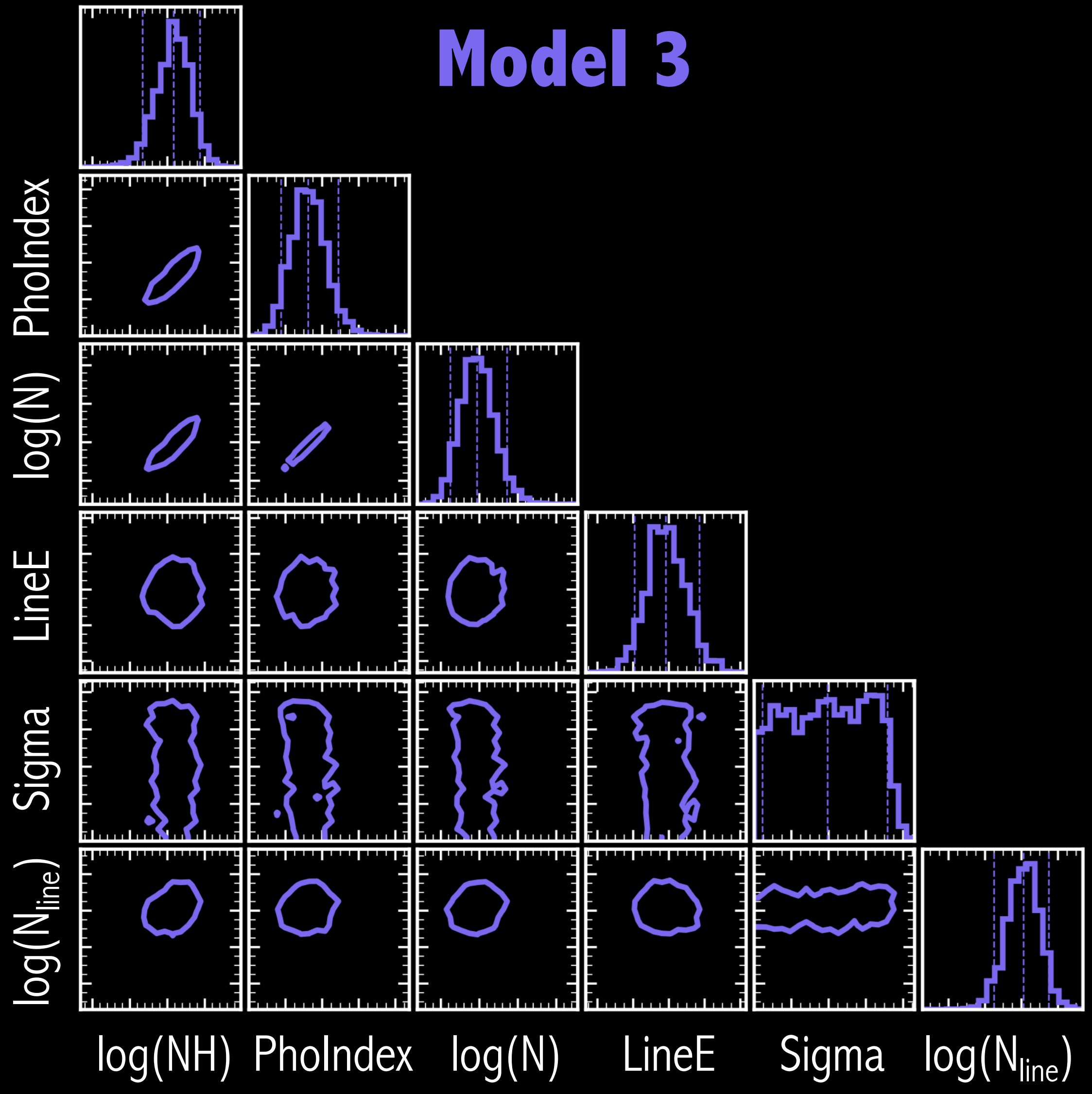
Model 1



Model 2

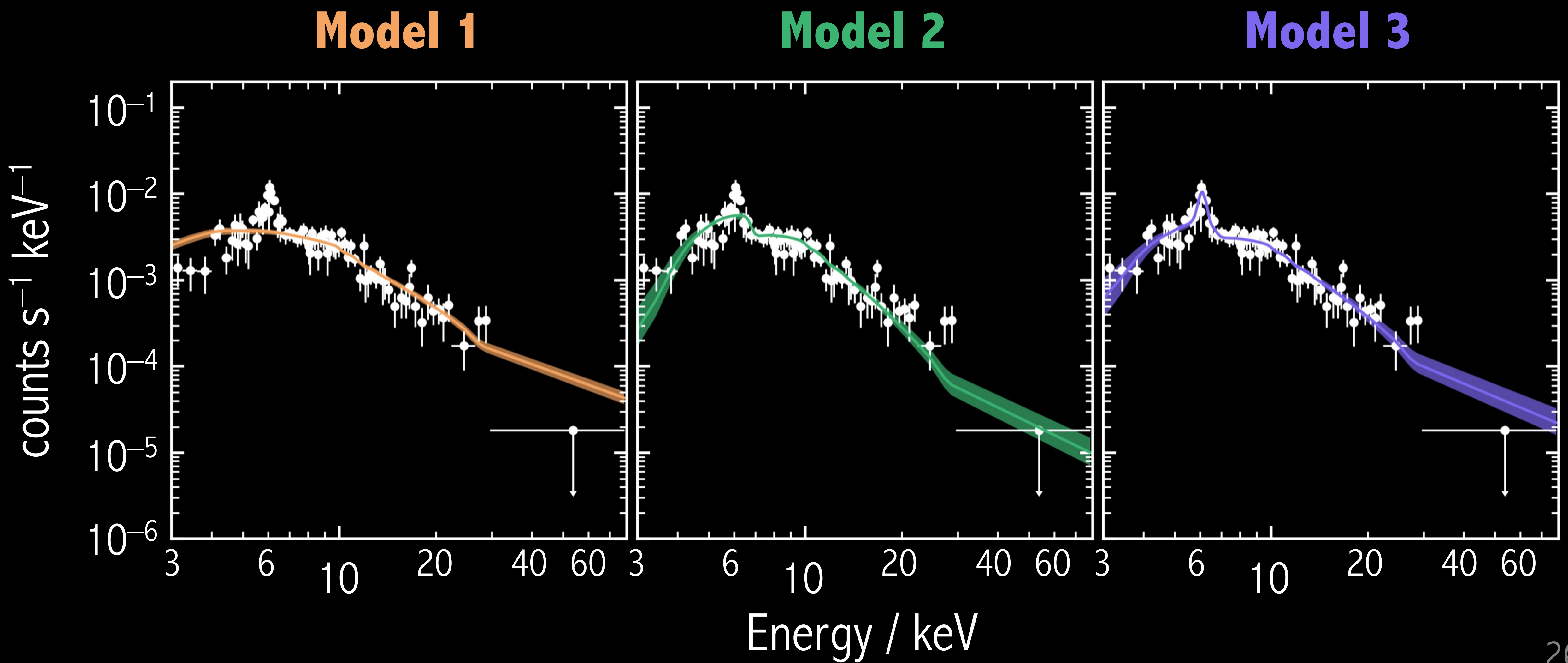


Model 3



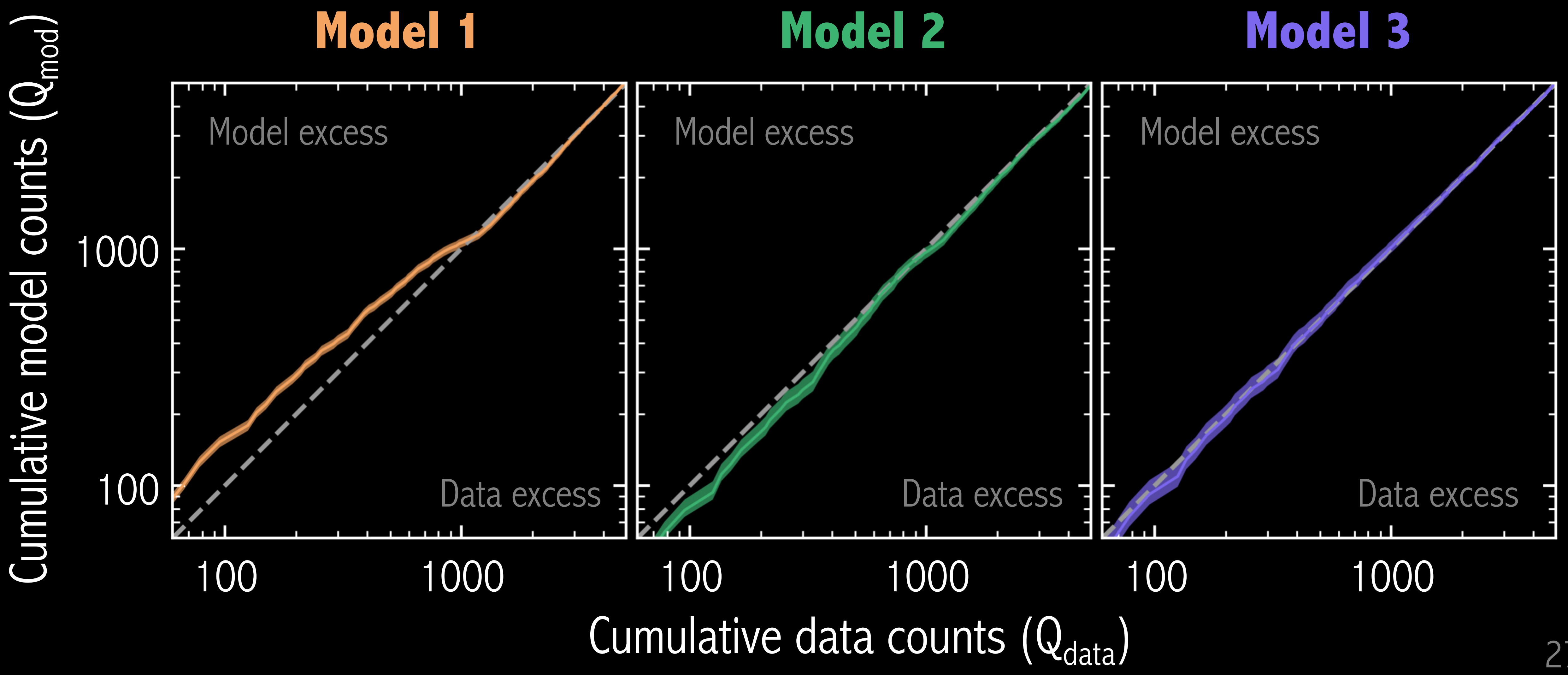
Deriving posteriors (see tutorial Exercises 1.3 & 2.1)

Using Monte Carlo sampling to learn from the data



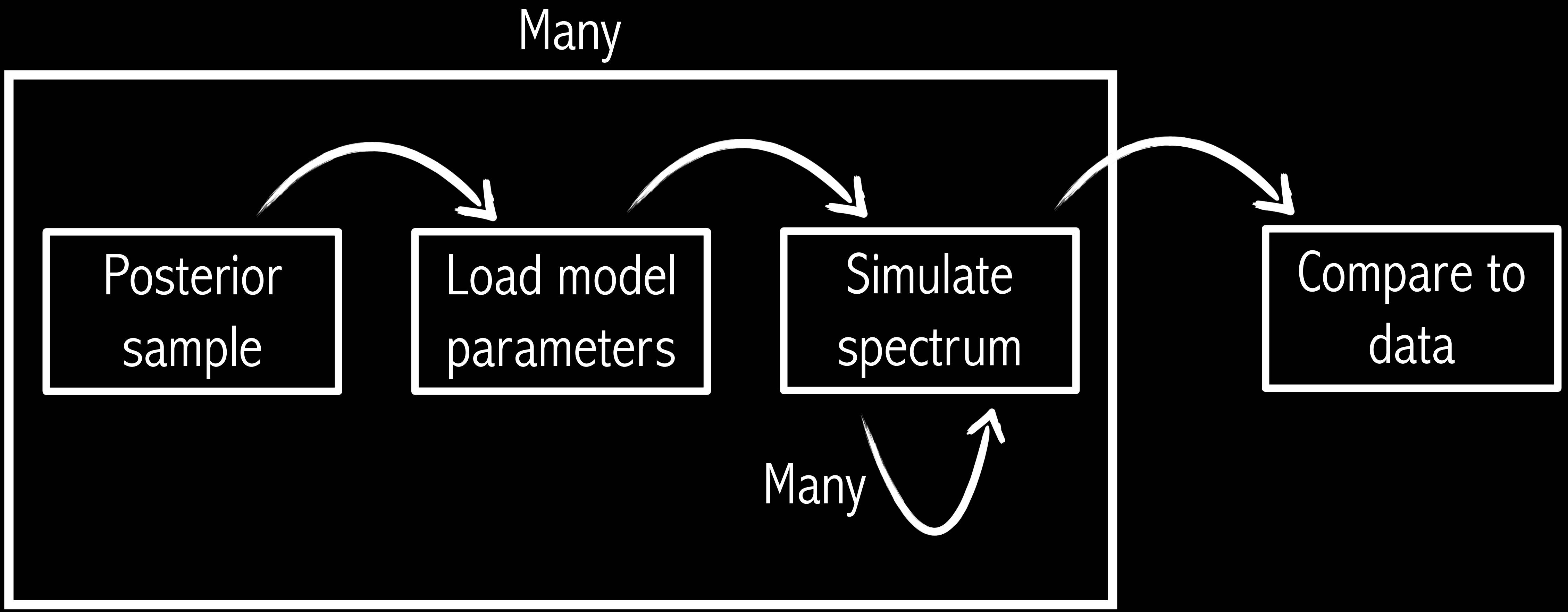
Quantile-Quantile plots (see tutorial Exercise 2.2)

A way to search for missing components from entirely ungrouped spectra



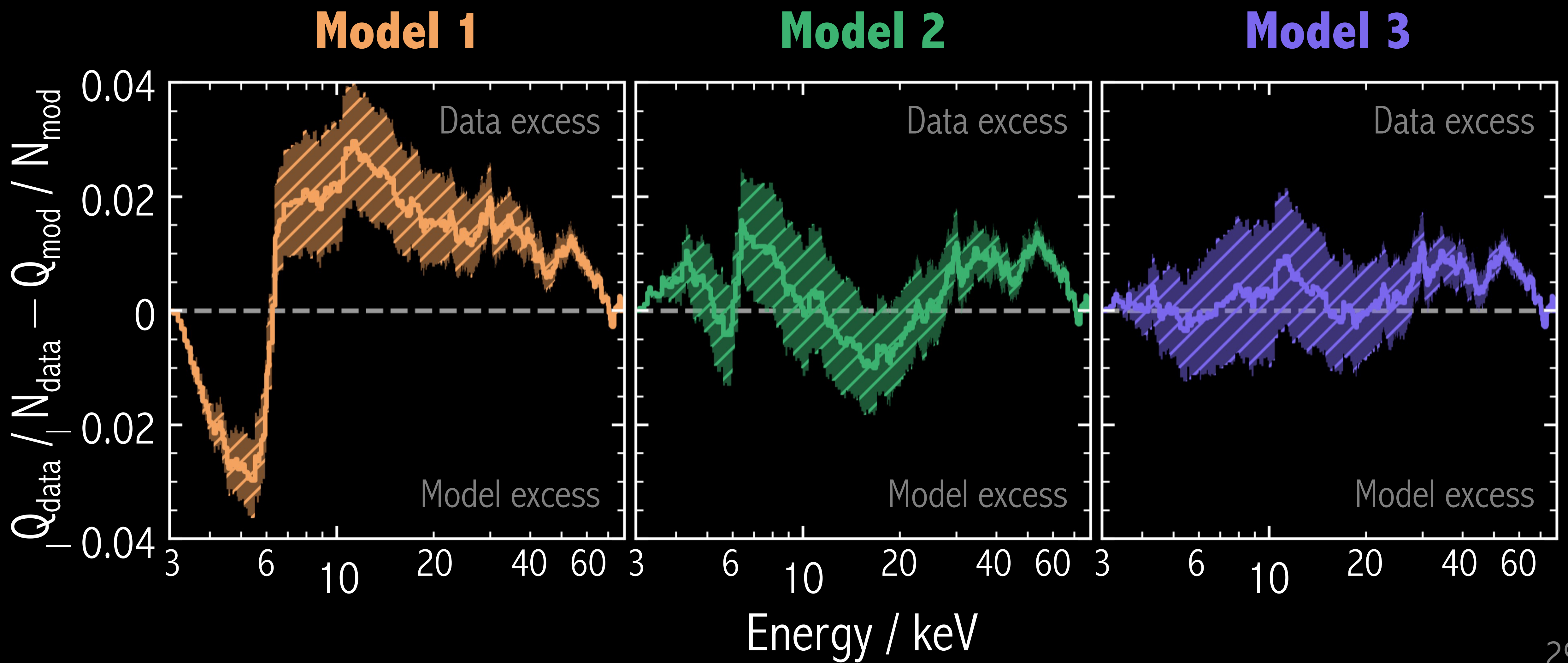
Model checking (see tutorial Exercise 2.3)

Posterior predictive checks quantify the goodness-of-fit and can be useful in the search for missing model components



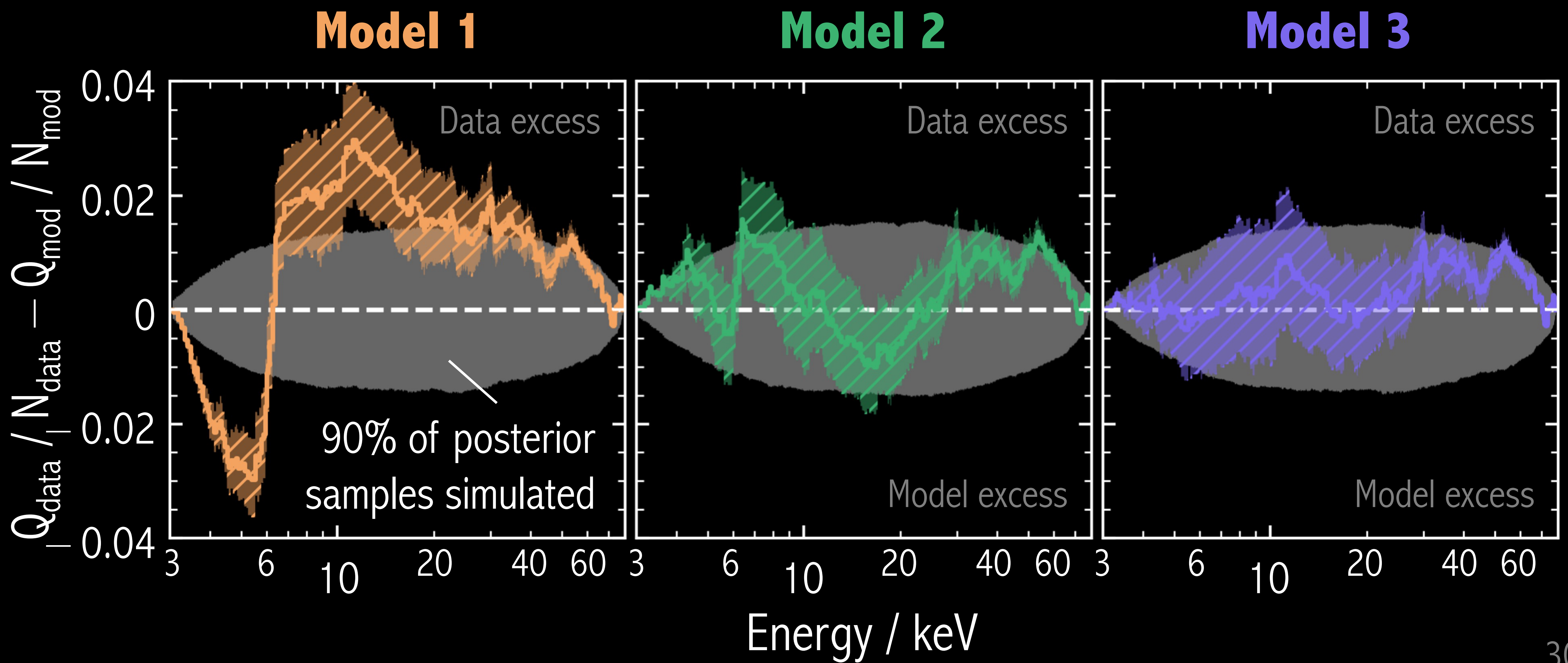
Quantile-Quantile difference (see tutorial Exercise 2.2)

Reproject Quantile-Quantile plots vs. channel energy (Buchner & Boorman 23)



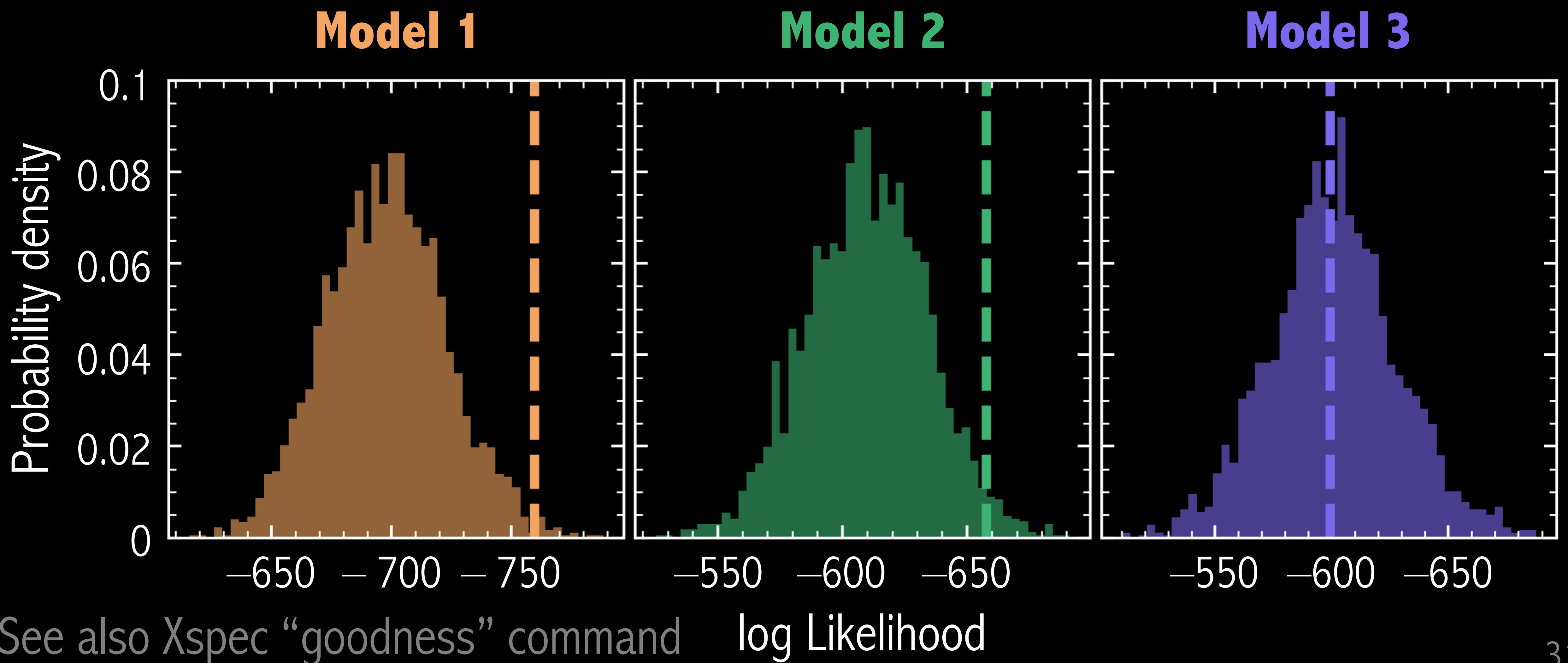
Posterior predictive checks (see tutorial Exercise 2.3)

Useful to quantify the goodness-of-fit and search for missing model components



Profiling the Likelihood (see tutorial Exercise 2.3)

Comparing best-fit to best-fits of many generated data as a goodness-of-fit test

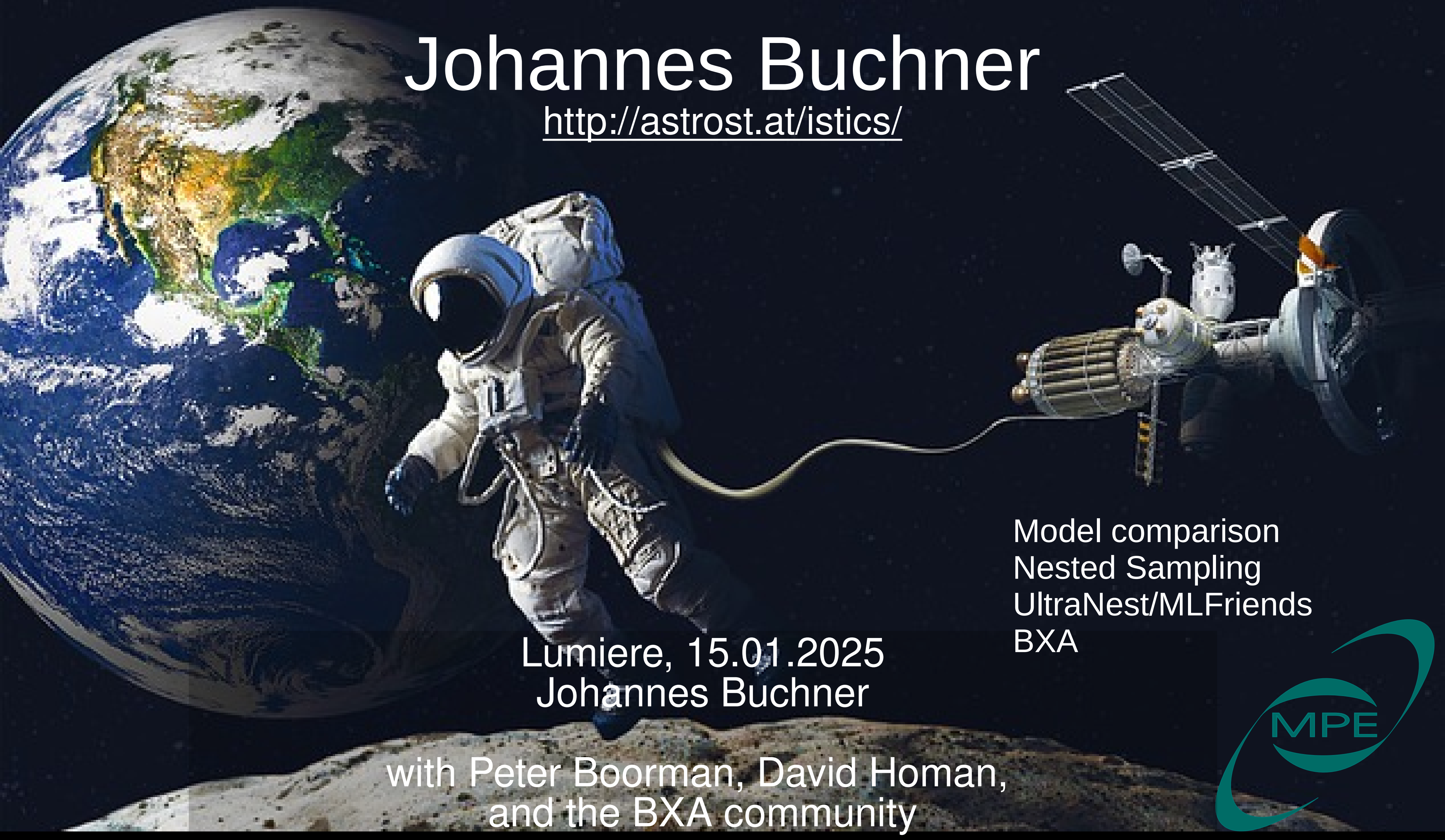


See also Xspec "goodness" command

Traversing the space of parameter spaces in the space sciences

Johannes Buchner

http://astrost.at/istics/



Model comparison
Nested Sampling
UltraNest/MLFriends
BXA

Lumiere, 15.01.2025
Johannes Buchner

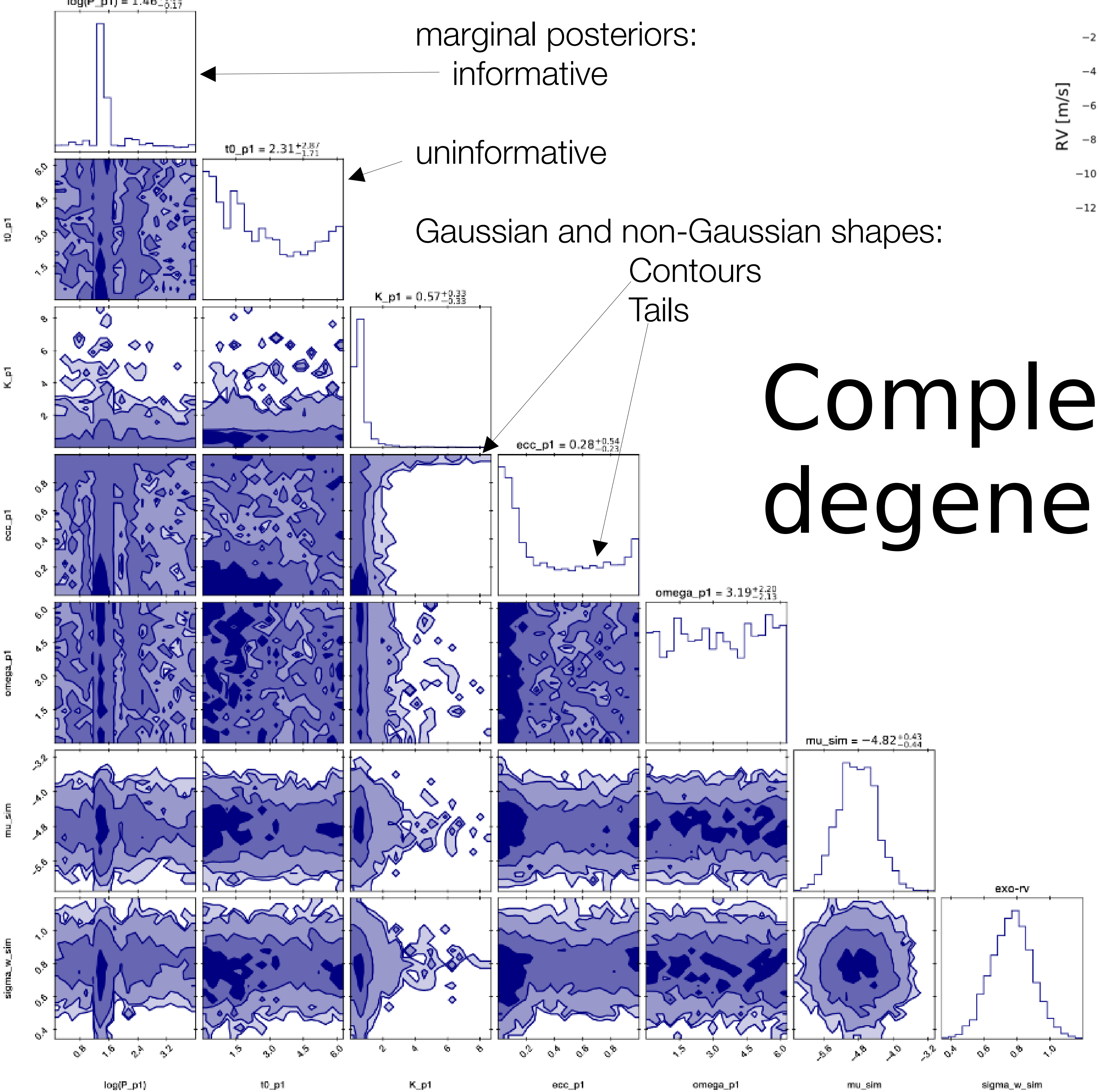
with Peter Boorman, David Homan,
and the BXA community

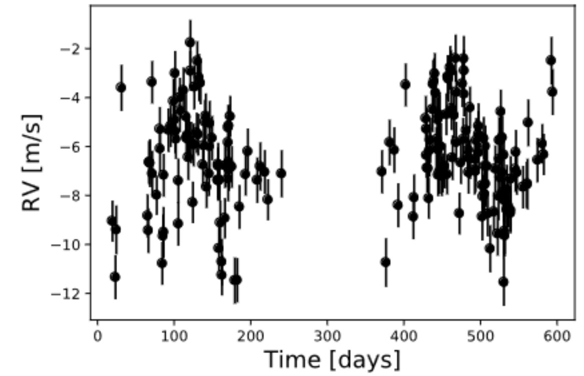
Complex degeneracies

marginal posteriors:
informative

uninformative

Gaussian and non-Gaussian shapes:
Contours
Tails



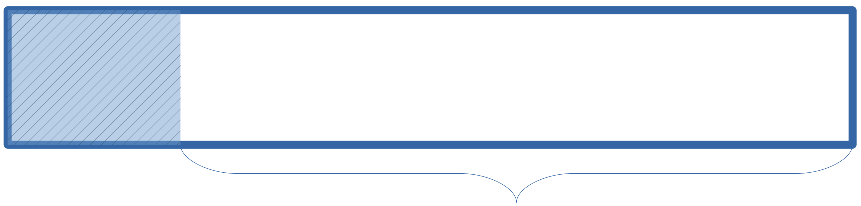


Model comparison

Model comparison

Buchner+14

- Empirical models
  - Information content
  - Prediction quality
- Component presence
  - Regions of practical equivalence
- Physical effects
  - Bayesian model comparison
  - Priors often well-justified



https://arxiv.org/abs/1506.02273
Betancourt (2015)

Information criteria

Akaike (1973)

- Akaike information criterion
- Is more complex worth storing?

$$AIC = 2 * d - 2 * L_{max}$$

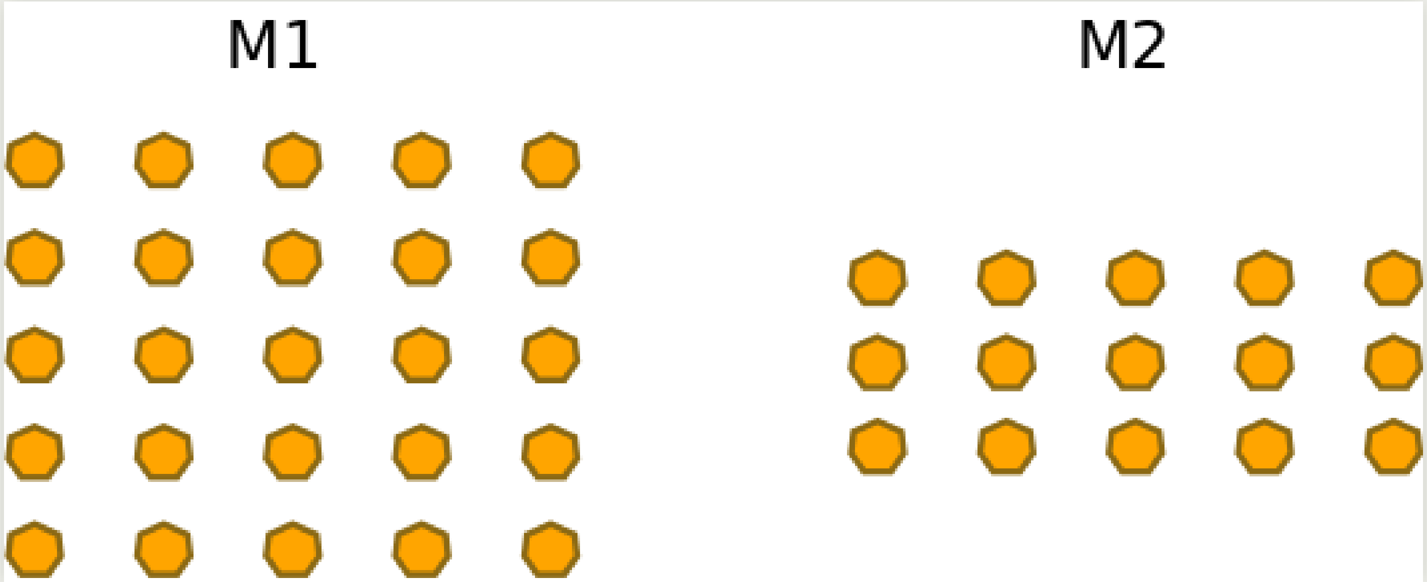
$$AIC = 2 * d + CStat$$

Advantages:
- rooted in information theory
- independent of prior

Disadvantages:
- No uncertainties, thresholds unclear
- ...

Bayesian model comparison

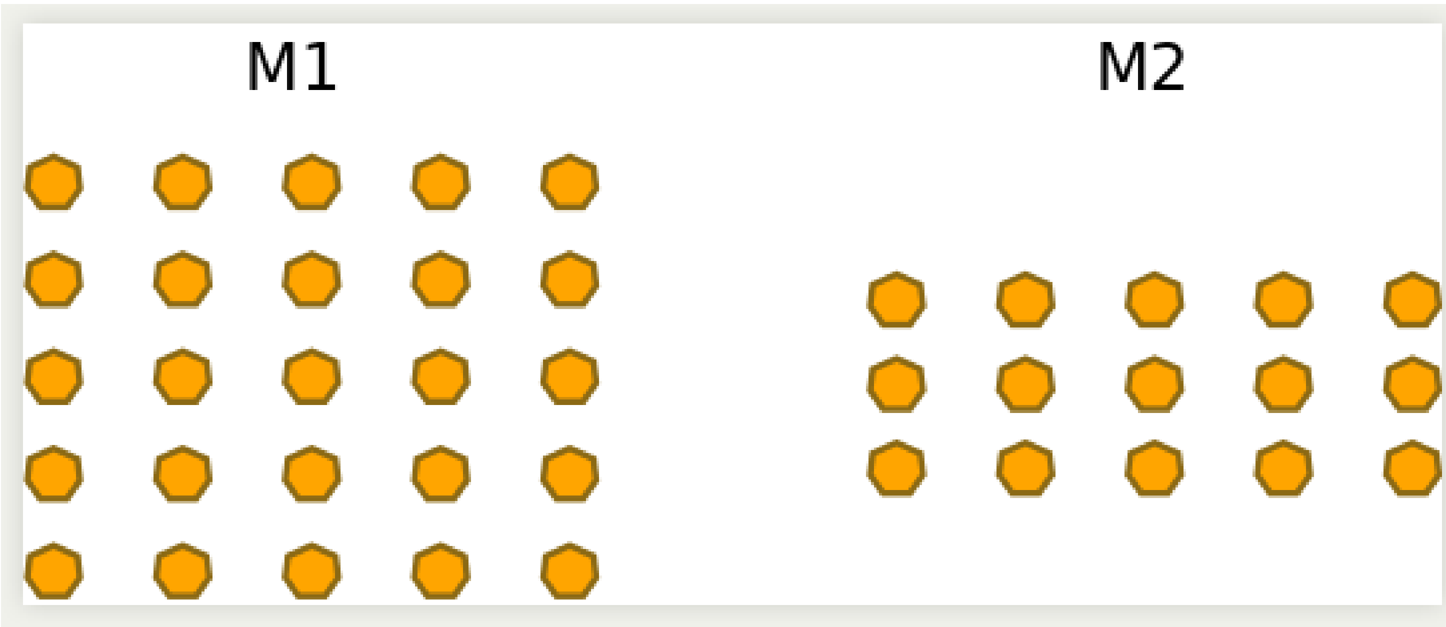
Two models



- Compare two parameter spaces by
$$\sum \mathcal{L}\Big|_{M1} / \sum \mathcal{L}\Big|_{M2}$$
- How many coins to put in M1, M2?
- model prior

Punishing prediction diversity

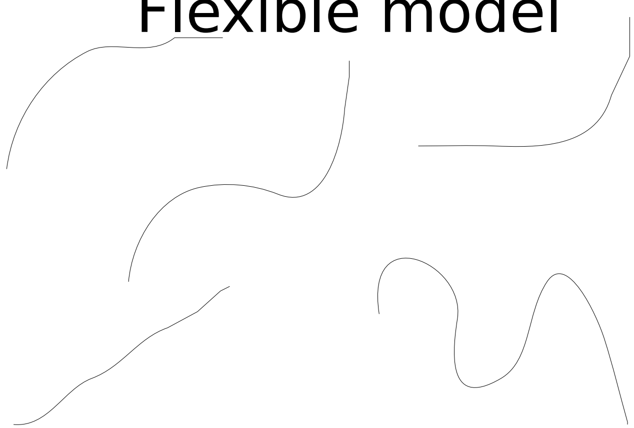
(not number of parameters)

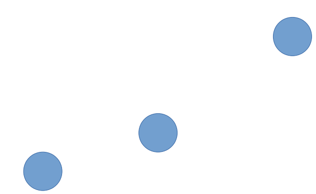


Flexible model

Inflexible model

Data





L high, V tiny

L medium, V medium

ML training?

What to do with Z

- $Z1, Z2$

$$\underbrace{\frac{p(M1|D)}{p(M2|D)}}_{\text{Posterior odds ratio}} = \underbrace{\frac{Z1}{Z2}}_{\text{Bayes factor}} \cdot \underbrace{\frac{p(M1)}{p(M2)}}_{\text{Prior odds ratio}}$$

$$\frac{p(M1|D)}{p(M2|D)} = \frac{Z1 \cdot p(M1)}{Z2 \cdot p(M2)}$$

Buchner+14

What to do with Z

- Z1, Z2

$$\frac{p(M1|D)}{p(M2|D)} = \frac{Z1 \cdot p(M1)}{Z2 \cdot p(M2)}$$

$$\frac{p(M_1|D)}{\sum p(M_i|D)} = \frac{Z_1 \cdot p(M_i)}{\sum_i Z_i \cdot p(M_i)}$$

- model priors: leave to reader or motivated by theory
- Discard highly improbable model or marginalise
- Does $\frac{p(M1|D)}{p(M2|D)} = 3/1$ mean M2 is correct in a quarter of the cases?

Calibrating model decisions

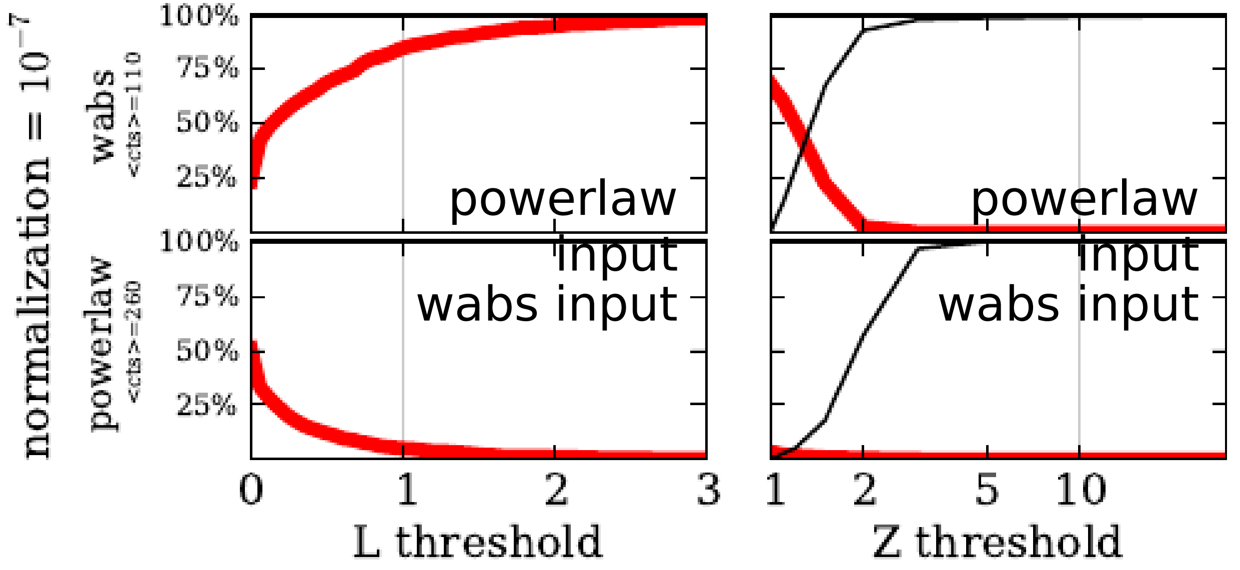
- Model probabilities → decisions
- False decision rate

  (false positives/negatives)

  - Monte Carlo simulations (parametric bootstrap)

Buchner+14

Calibrating model decisions



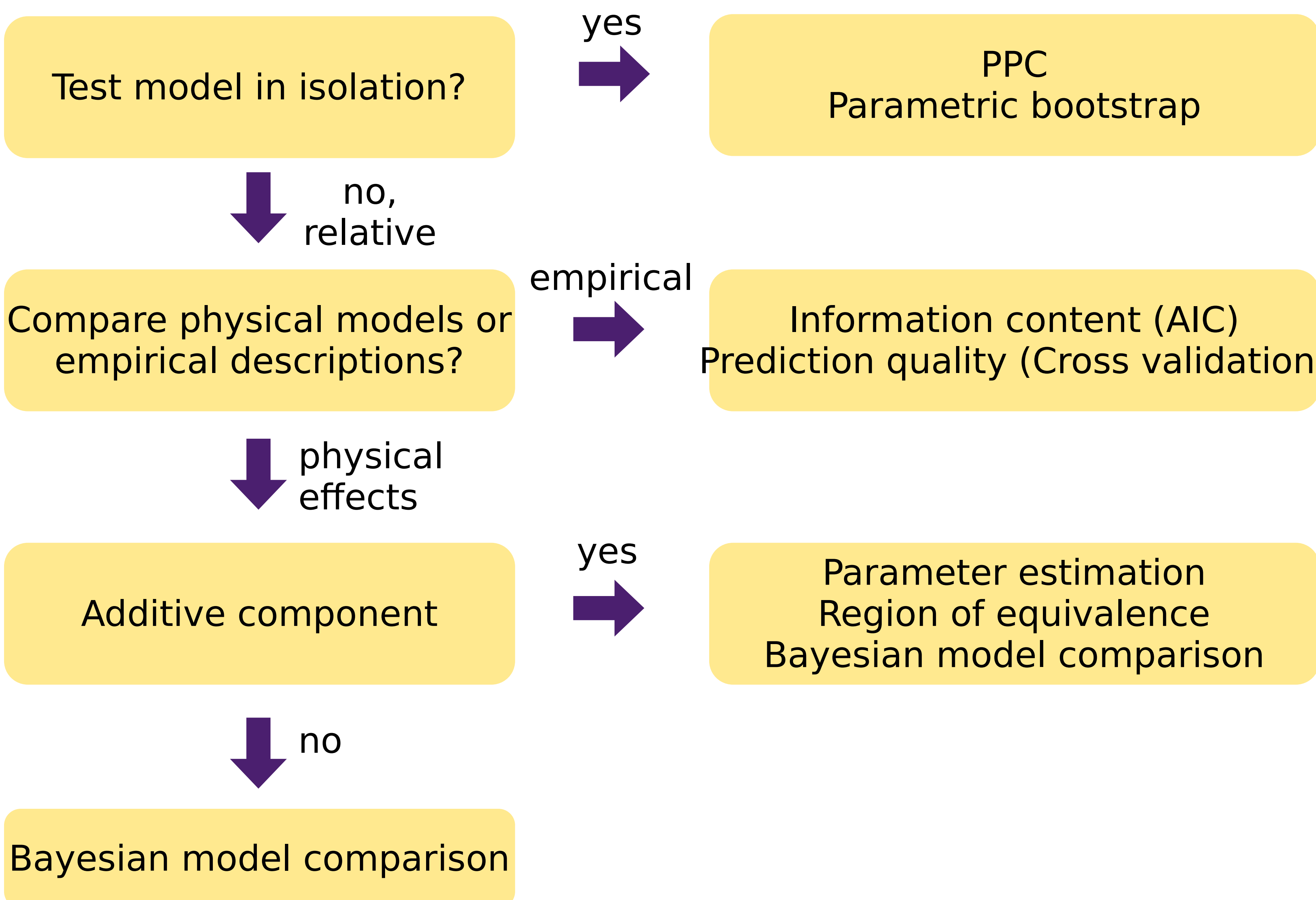
Buchner+14

False negatives
Non-decisions

Advantages:
- Get rid of parameter prior dependences
- Have frequentist properties of Bayesian method
- Completely Bayesian treatment + decisions

Disadvantages:
- Can be computationally expensive

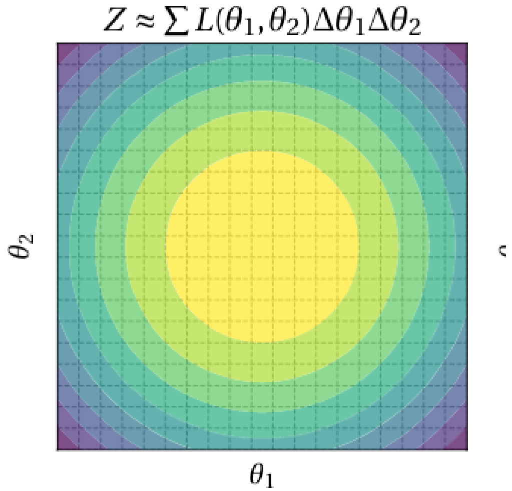
Model comparison



How to compute the Bayesian evidence Z

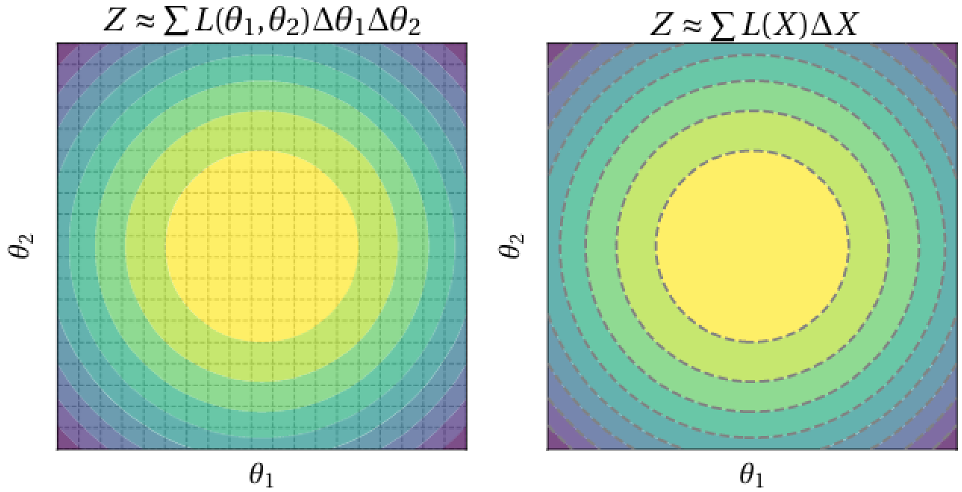
aka marginal likelihood

Nested Sampling



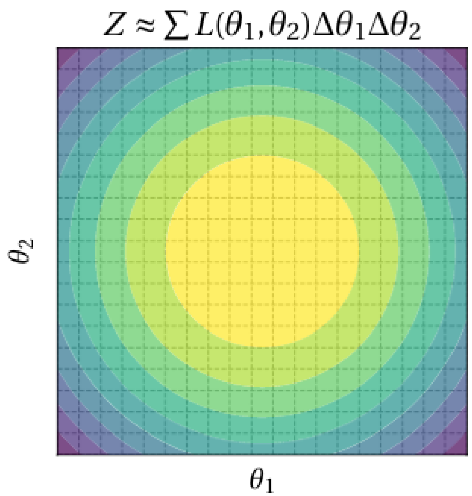
Riemann integration

Figure 1 | **Illustrations of NS algorithm.**

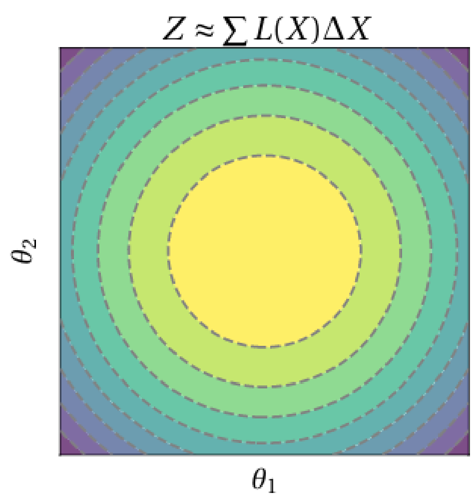


Riemann integration Lebesgue integration

Figure 1 | **Illustrations of NS algorithm.**



Riemann integration



Lebesgue integration

Uniformly distributed live points

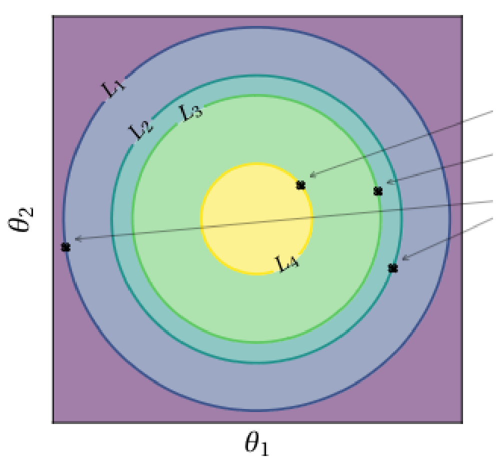
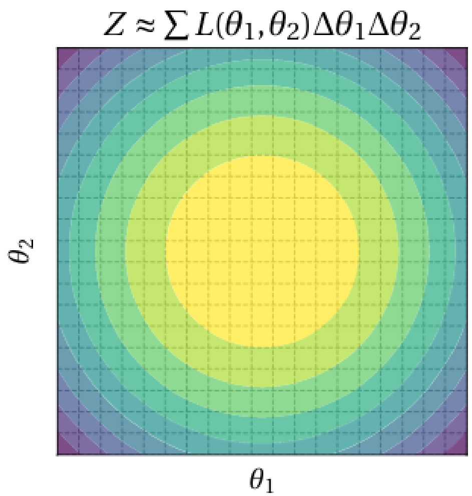
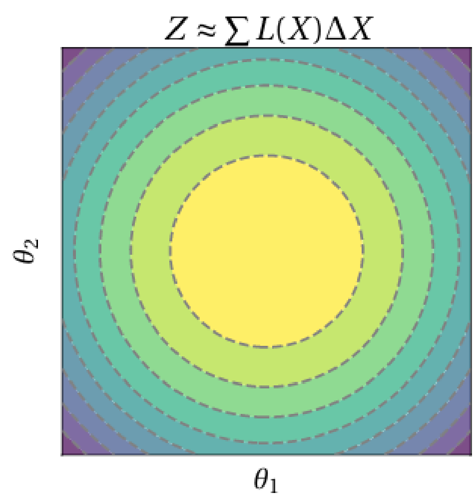


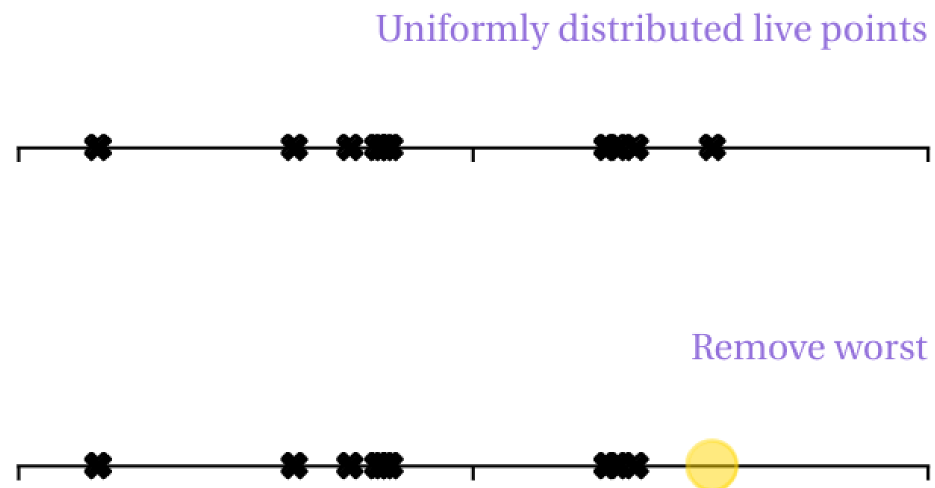
Figure 1 | **Illustrations of NS algorithm.**



Riemann integration



Lebesgue integration



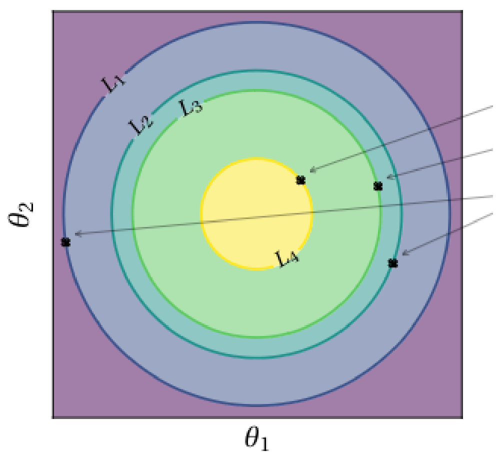
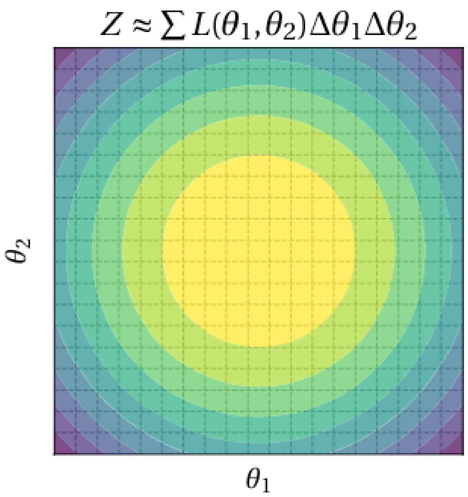
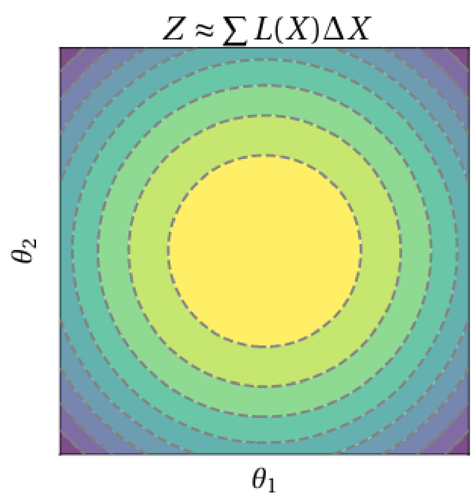


Figure 1 | **Illustrations of NS algorithm.**

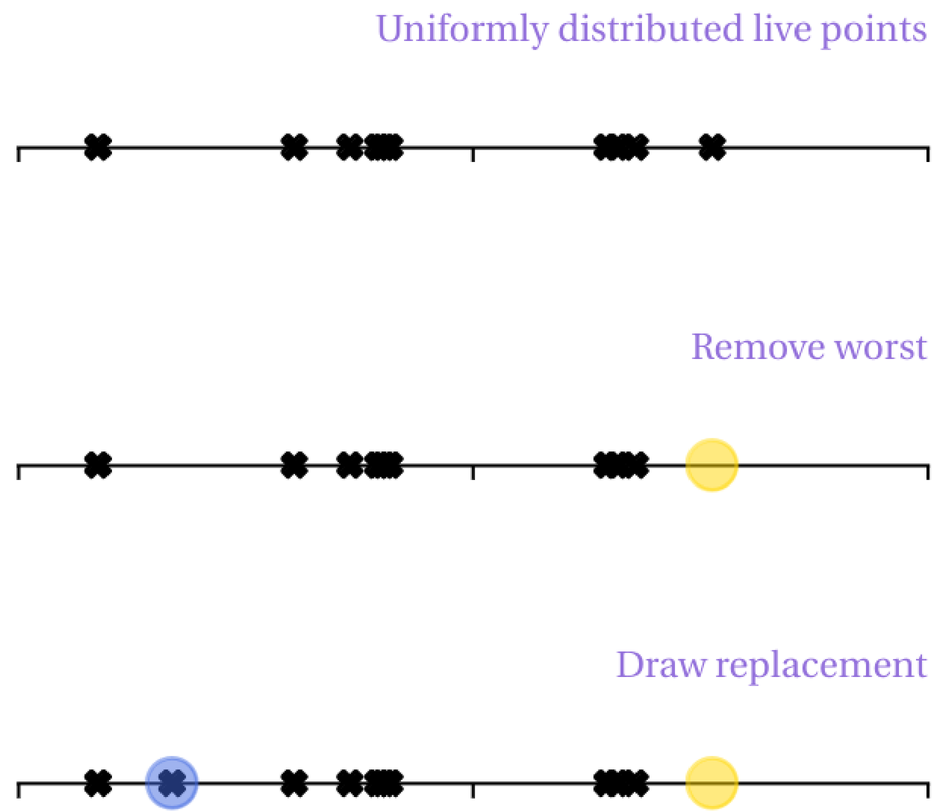
Ashton, Buchner, et al. 2022, Nature Reviews Methods Primers



Riemann integration



Lebesgue integration



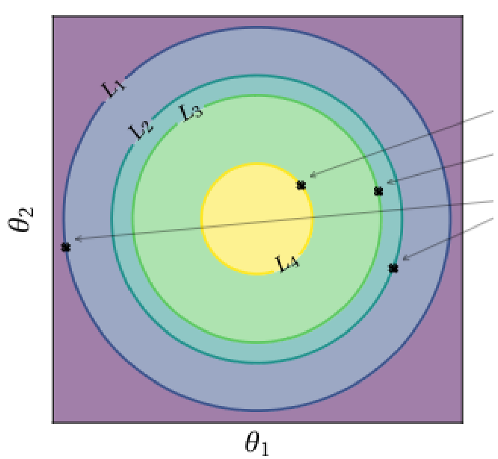
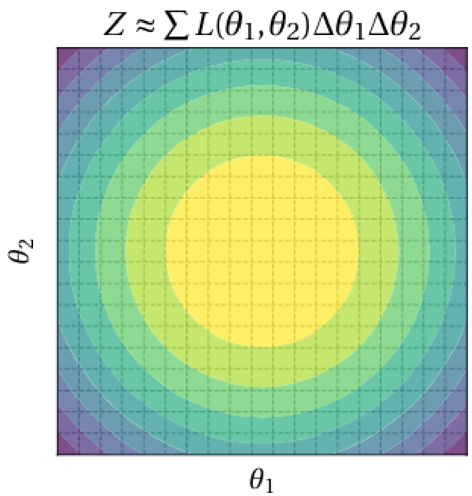
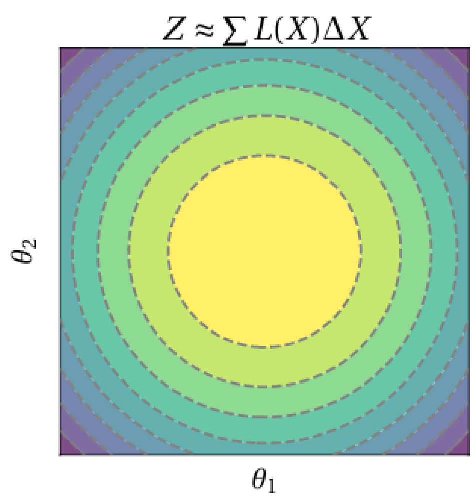


Figure 1 | **Illustrations of NS algorithm.**

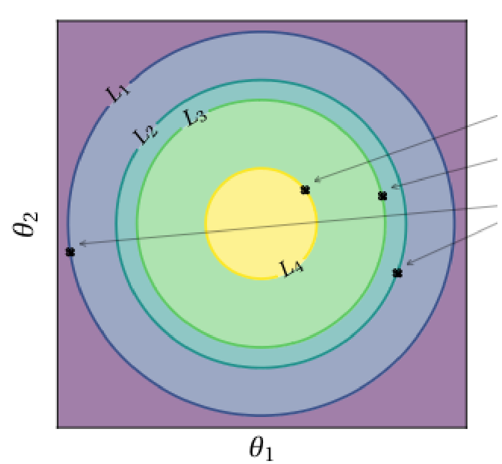
Ashton, Buchner, et al. 2022, Nature Reviews Methods Primers

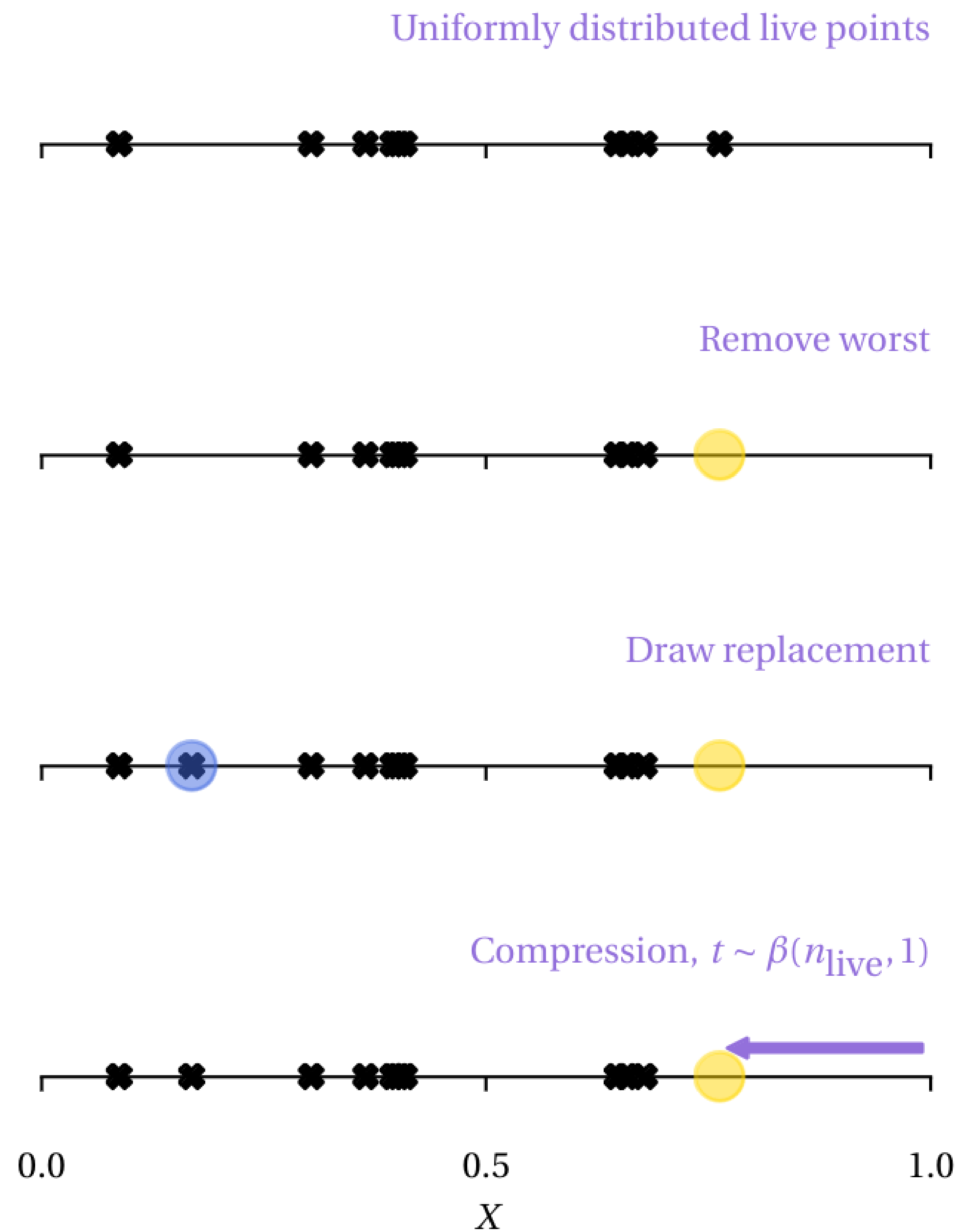


Riemann integration



Lebesgue integration

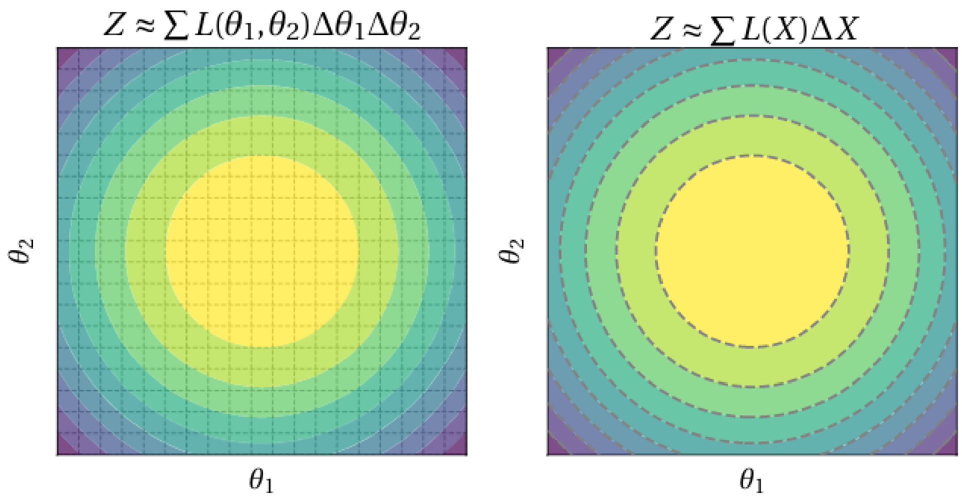




c | **Compression in one iterate of NS.**

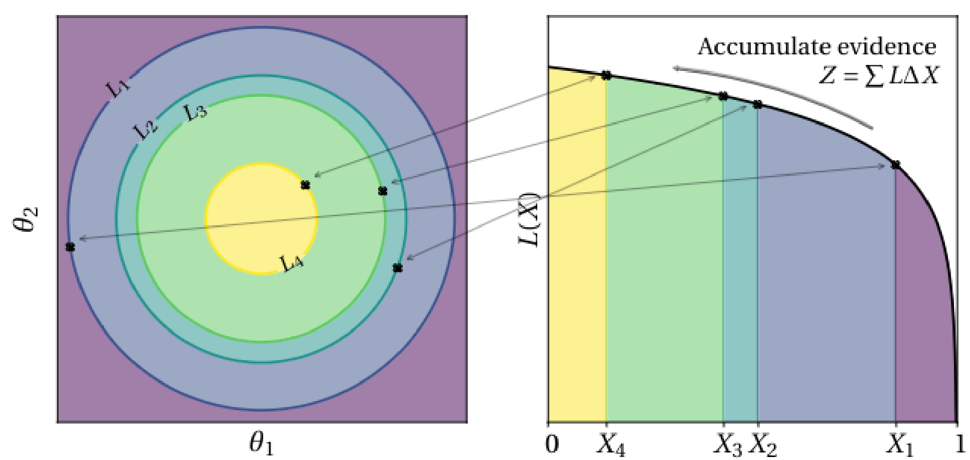
Figure 1 | **Illustrations of NS algorithm.**

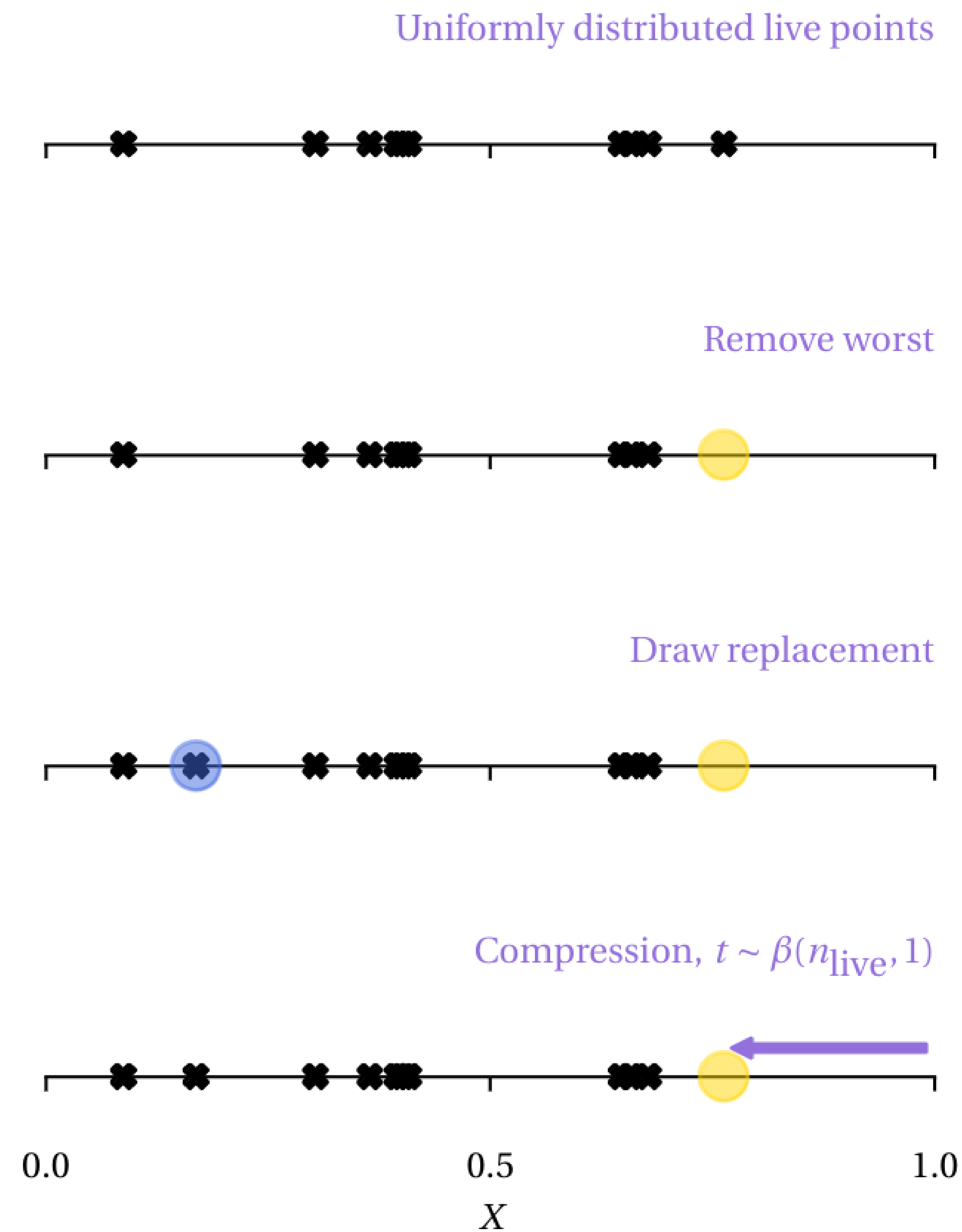
Ashton, Buchner, et al. 2022, Nature Reviews Methods Primers



Riemann integration

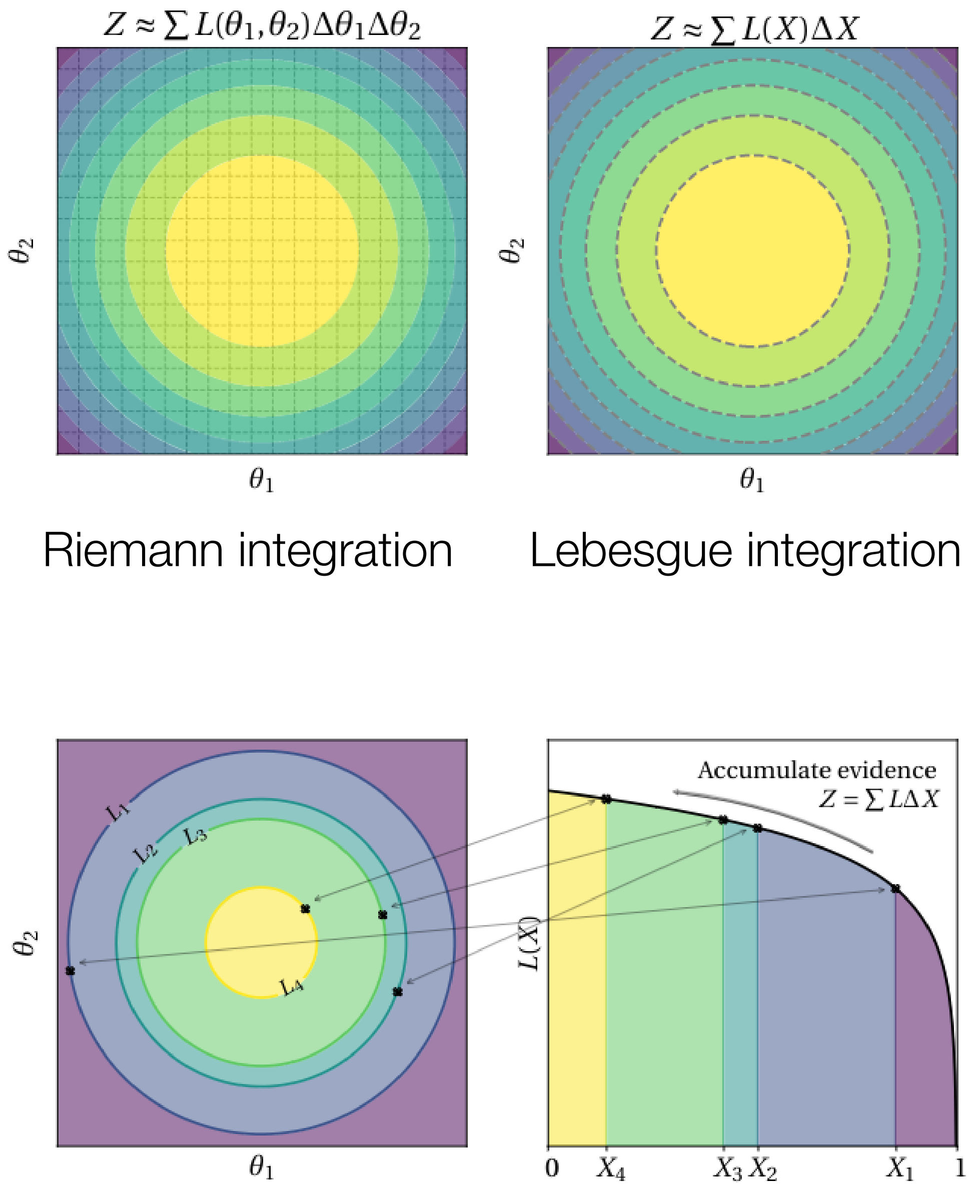
Lebesgue integration

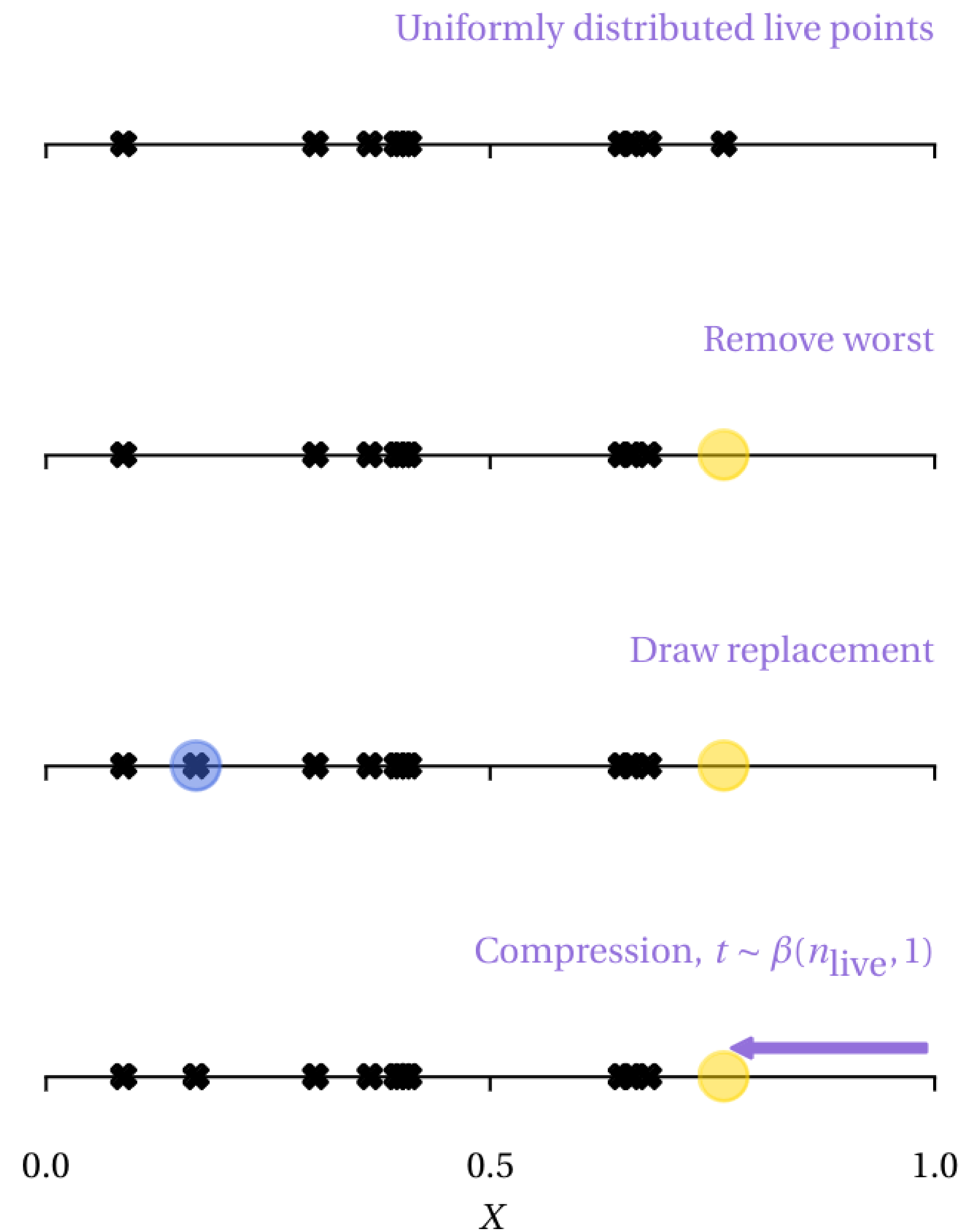




c | **Compression in one iterate of NS.**

Figure 1 | **Illustrations of NS algorithm.**

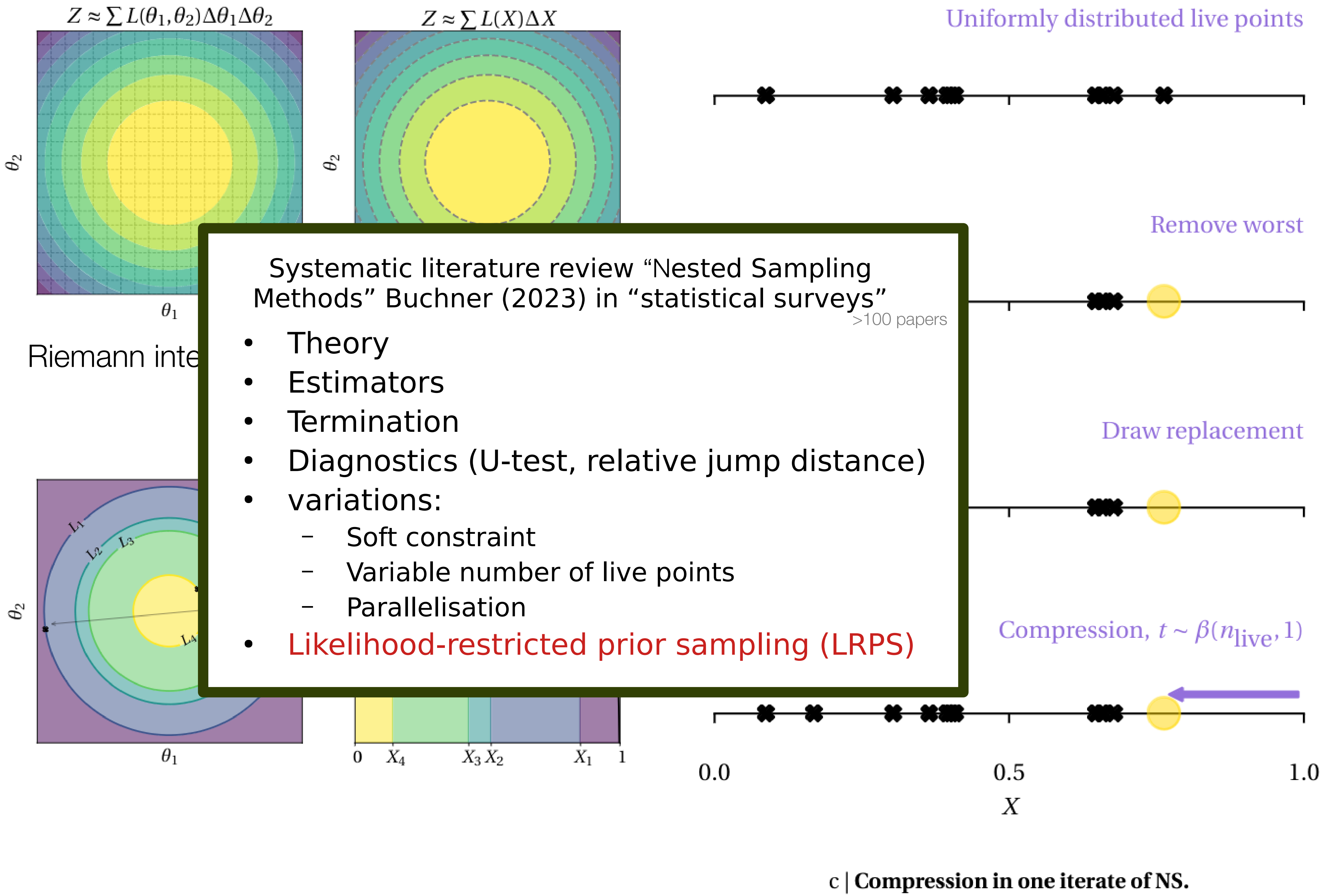




c | **Compression in one iterate of NS.**

Convergence proof of Z and posterior :
e.g. Evans (2007), Chopin&Robert (2010)

Figure 1 | **Illustrations of NS algorithm.**



Systematic literature review "Nested Sampling Methods" Buchner (2023) in "statistical surveys"

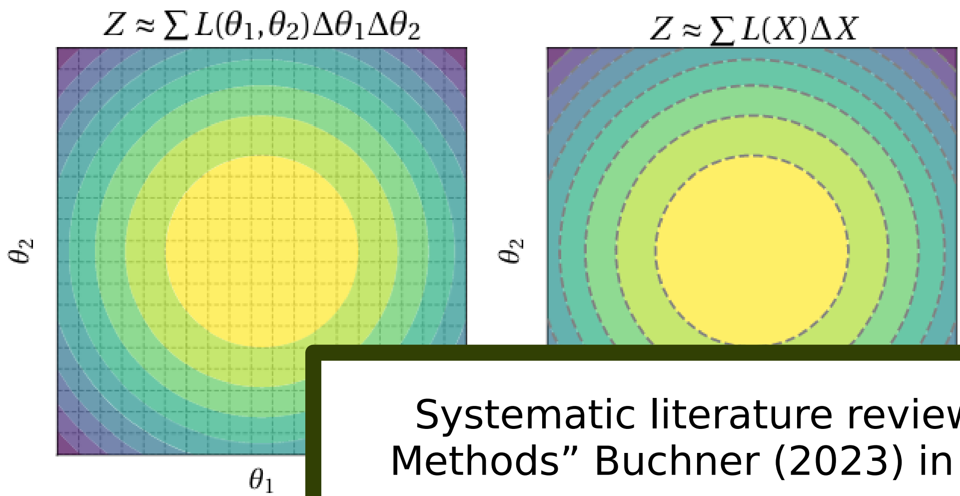
>100 papers

- Theory
- Estimators
- Termination
- Diagnostics (U-test, relative jump distance)
- variations:
  - Soft constraint
  - Variable number of live points
  - Parallelisation
- Likelihood-restricted prior sampling (LRPS)

Convergence proof of Z and posterior :
e.g. Evans (2007), Chopin&Robert (2010)

Figure 1 | **Illustrations of NS algorithm.**

Ashton, Buchner, et al. 2022, Nature Reviews Methods Primers

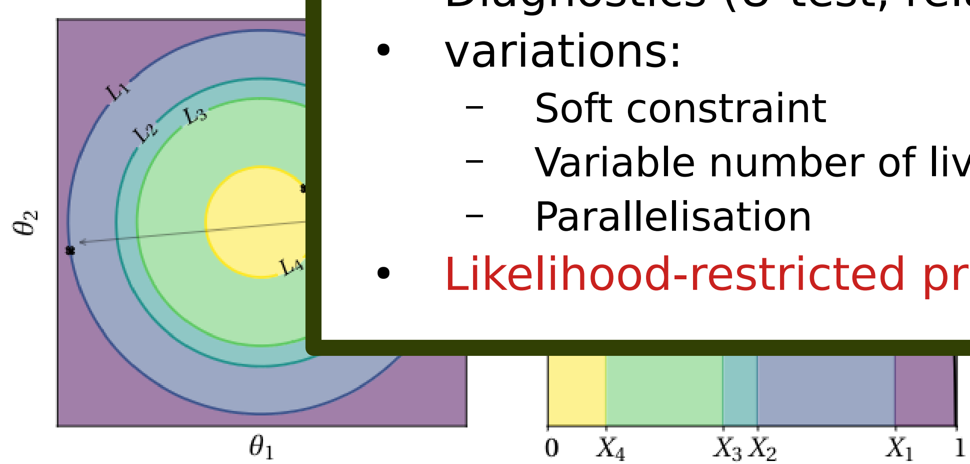


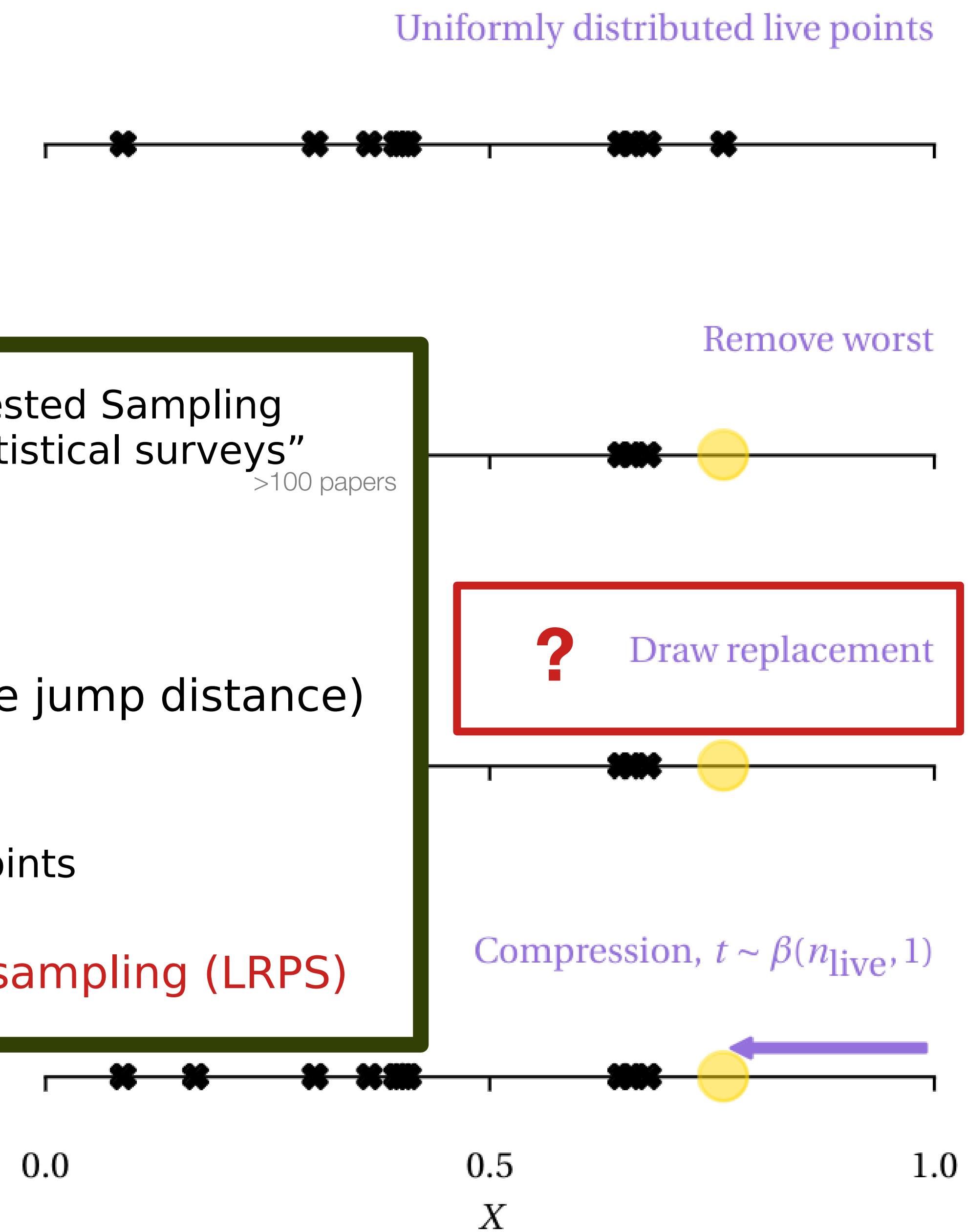
Riemann inte

Systematic literature review "Nested Sampling Methods" Buchner (2023) in "statistical surveys"

>100 papers

- Theory
- Estimators
- Termination
- Diagnostics (U-test, relative jump distance)
- variations:
  - Soft constraint
  - Variable number of live points
  - Parallelisation
- Likelihood-restricted prior sampling (LRPS)





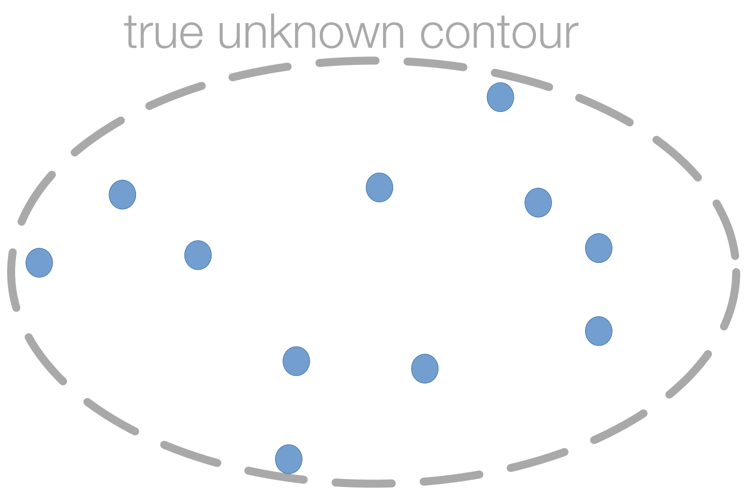
c | **Compression in one iterate of NS.**

Convergence proof of Z and posterior :
e.g. Evans (2007), Chopin&Robert (2010)

Figure 1 | **Illustrations of NS algorithm.**

- Step samplers?

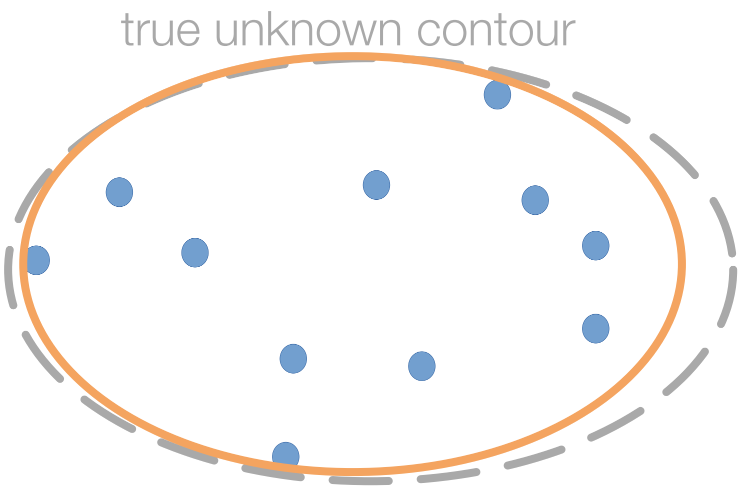
Likelihood-restricted prior sampling (LRPS)



Region-based methods:

Insight: Live points already traces out neighbourhood bounded by true unknown contour

Likelihood-restricted prior sampling (LRPS)

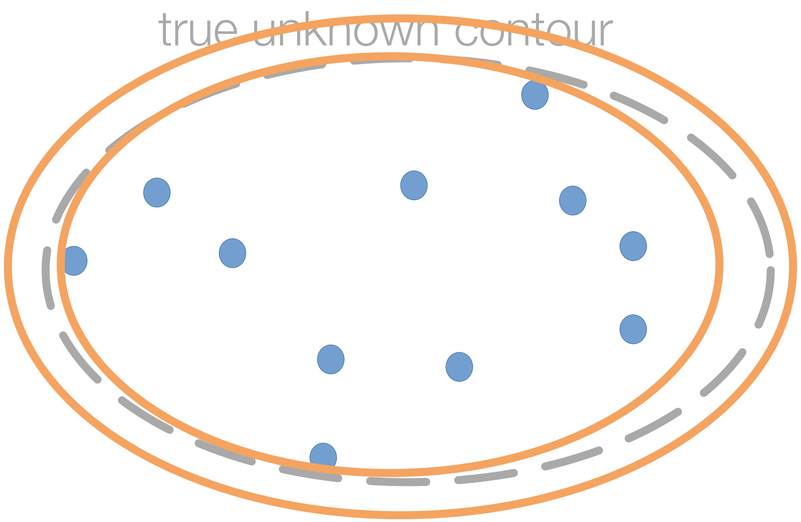


Region-based methods:

Insight: Live points already traces out neighbourhood bounded by true unknown contour

Smallest encapsulating ellipsoid (Mukherjee+06, Rollins15)

Likelihood-restricted prior sampling (LRPS)



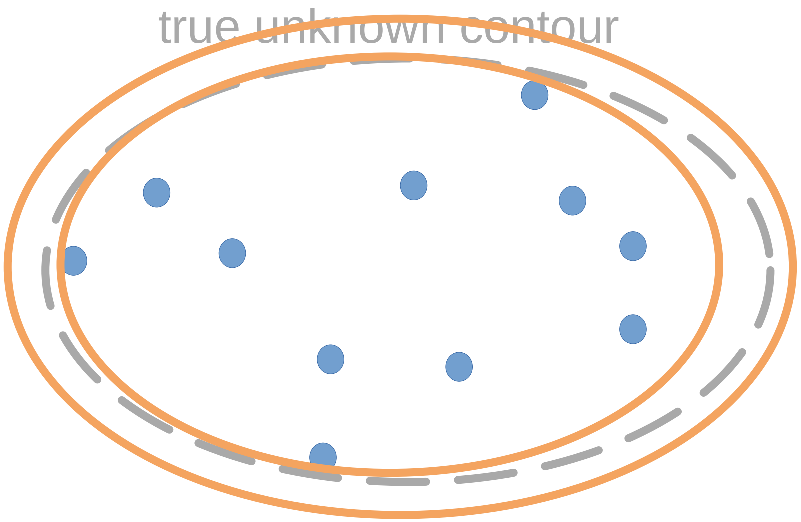
Region-based methods:

Insight: Live points already traces out neighbourhood bounded by true unknown contour

Smallest encapsulating ellipsoid (Mukherjee+06, Rollins15)
→ enlarge by a fudge factor

Sample and reject

Likelihood-restricted prior sampling (LRPS)



Region-based methods:

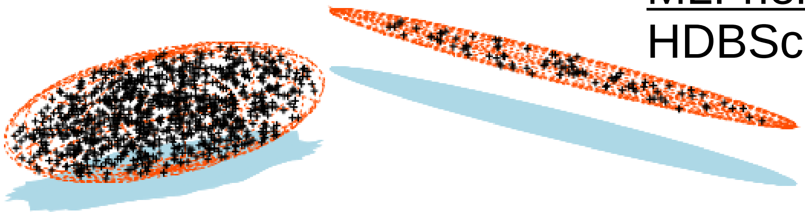
Insight: Live points already traces out neighbourhood bounded by true unknown contour

Smallest encapsulating ellipsoid (Mukherjee+06, Rollins15)
→ enlarge by a fudge factor

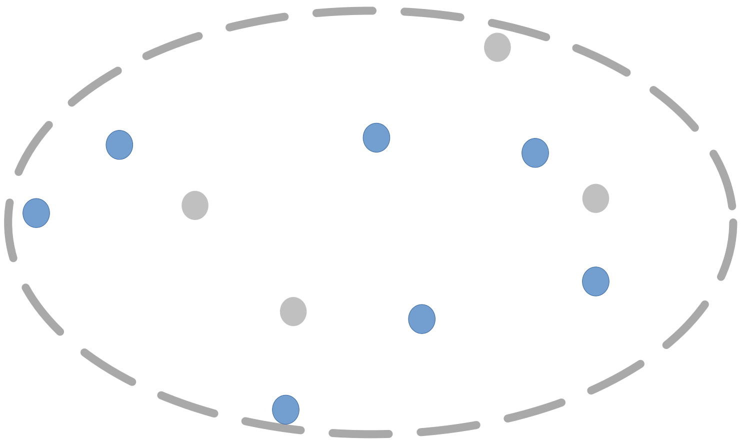
Sample and reject

Other shapes: Clustering

K-means: Shaw+07, Theisen+13
X-means: Feroz+08, Feroz+09 (MultiNest), splitting criteria variations
MLFriends: Buchner14,17 (UltraNest)
HDBScan? → unclear how to sample with fuzzy clusters

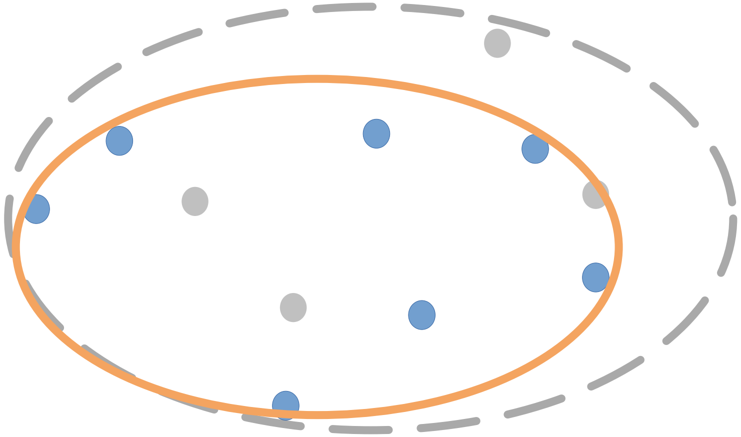


Bootstrapping: robust self-calibration



Sample with replacement
→ training sample
Left out points:
→ validation sample

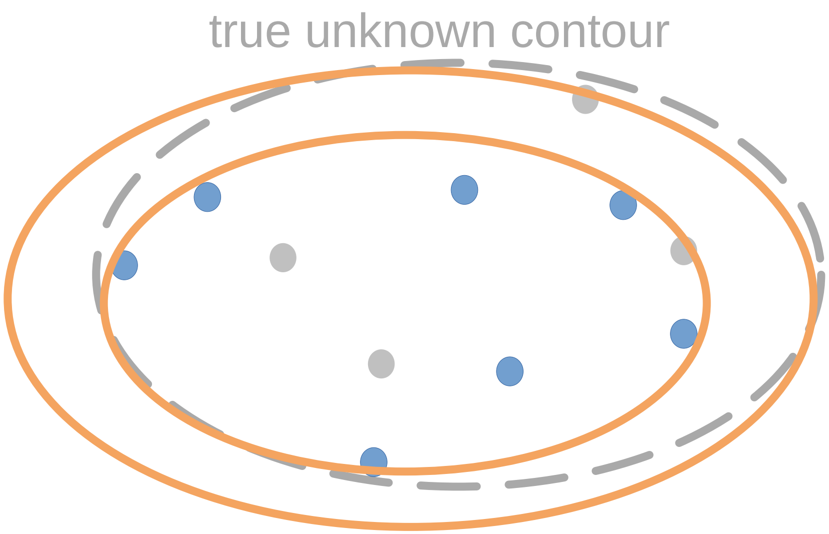
Bootstrapping: robust self-calibration



Sample with replacement
→ training sample
Left out points:
→ validation sample

Find enclosing ellipsoid

Bootstrapping: robust self-calibration

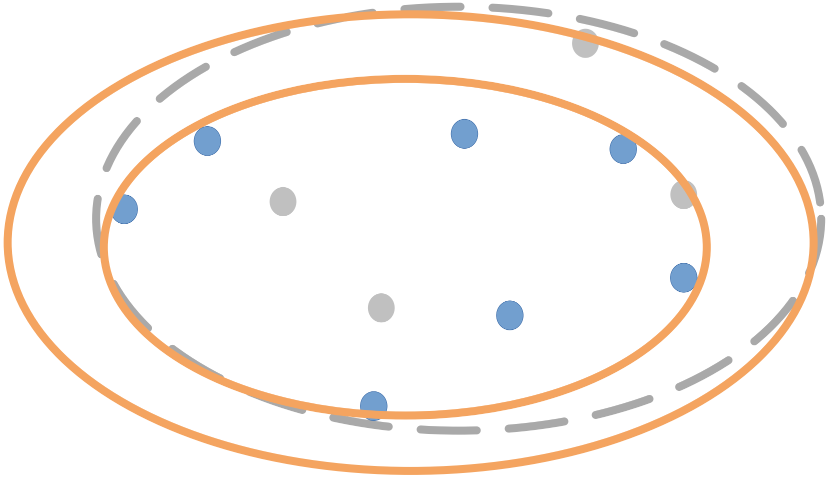


Sample with replacement
→ training sample
Left out points:
→ validation sample

Find enclosing ellipsoid

Enlarge until validation sample contained
→ enlargement factor

Bootstrapping: robust self-calibration



Sample with replacement
→ training sample
Left out points:
→ validation sample

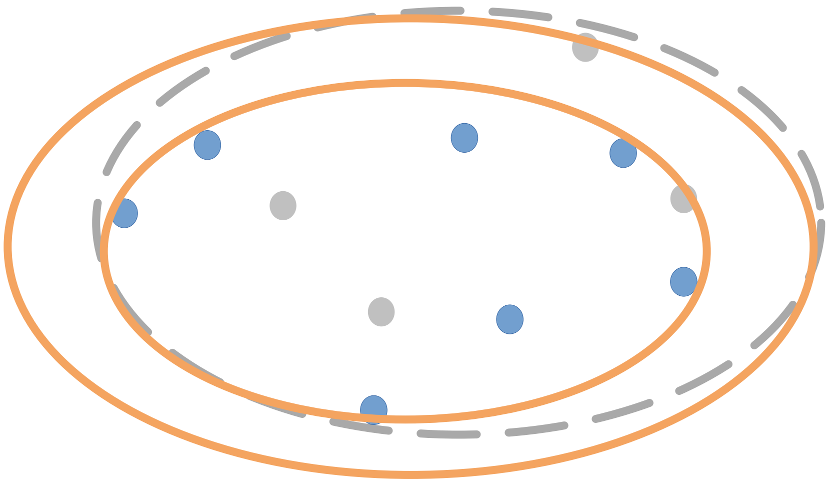
Find enclosing ellipsoid

Enlarge until validation sample contained
→ enlargement factor

Repeat a few times, retain largest enlargement factor

→ enlargement to apply to the full live point set

Bootstrapping: robust self-calibration



Sample with replacement
→ training sample
Left out points:
→ validation sample

Find enclosing ellipsoid

Enlarge until validation sample contained
→ enlargement factor

Repeat a few times, retain largest enlargement factor

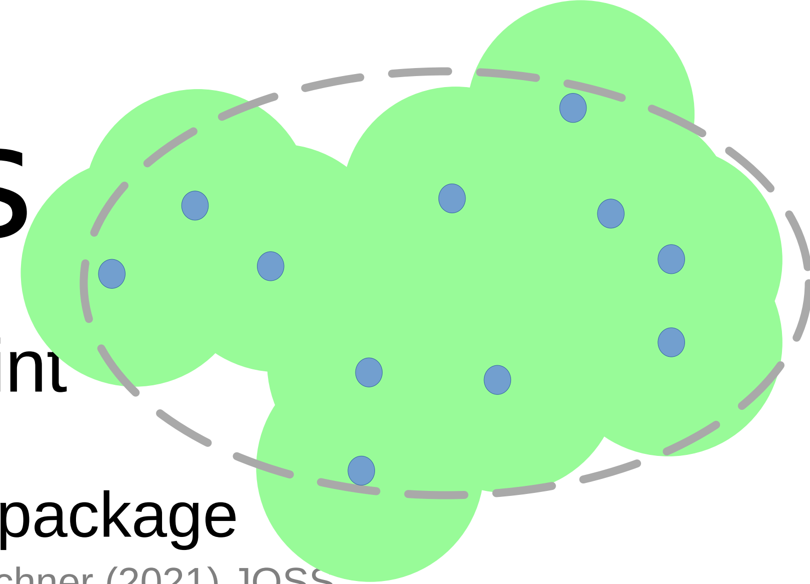
→ enlargement to apply to the full live point set

→ emulates other realisations of the nested sampling run

(Buchner14, Buchner17)

→ general, conservative approach, with safety guarantees

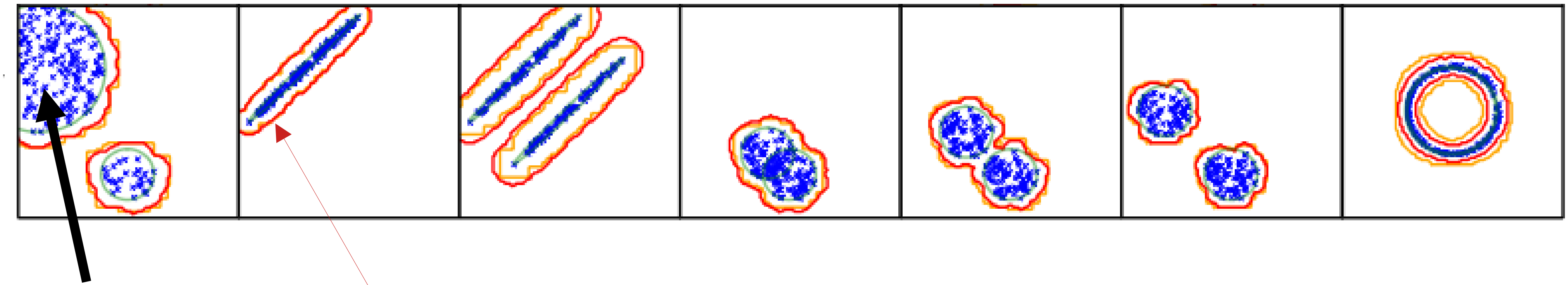
Complex shapes



MLFriends: Ellipsoid for each live point
Buchner (2014,2019)

default algorithm in the UltraNest Python package
Buchner (2021) JOSS

Adapting to complex contours



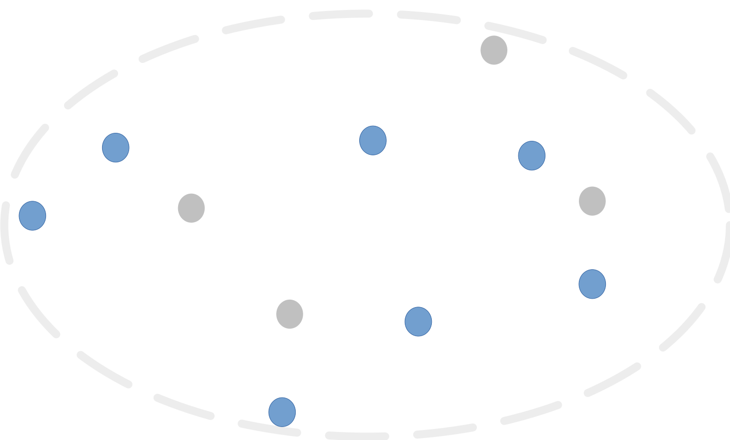
Live points

Constructed sampling region

Demo:
https://johannesbuchner.github.io/mcmc-demo/app.html#RadFriends-NS,banana

Analysing the bootstrap

sample with replacement → 2/3 in training sample, 1/3 in validation sample



(37) $$P(k) = \frac{S_2(K,k)\,K!}{K^K\,(K-k)!}$$

which combines the number of unique partitions of size k (S_2, Stirling number of the second kind) with the number of permutations (selecting k out of K). The expectation of k is

(38) $$E(k) = K\left(1-\left(1-\frac{1}{K}\right)^K\right)$$ is ~66%

with the variance:

(39) $$\mathrm{Var}(k) = n\left(1-\frac{1}{K}\right)^K + K^2\left(1-\frac{1}{K}\right)\left(1-\frac{2}{K}\right)^K - K^2\left(1-\frac{1}{K}\right)^{2K}.$$

Never in validation sample in **m** rounds:

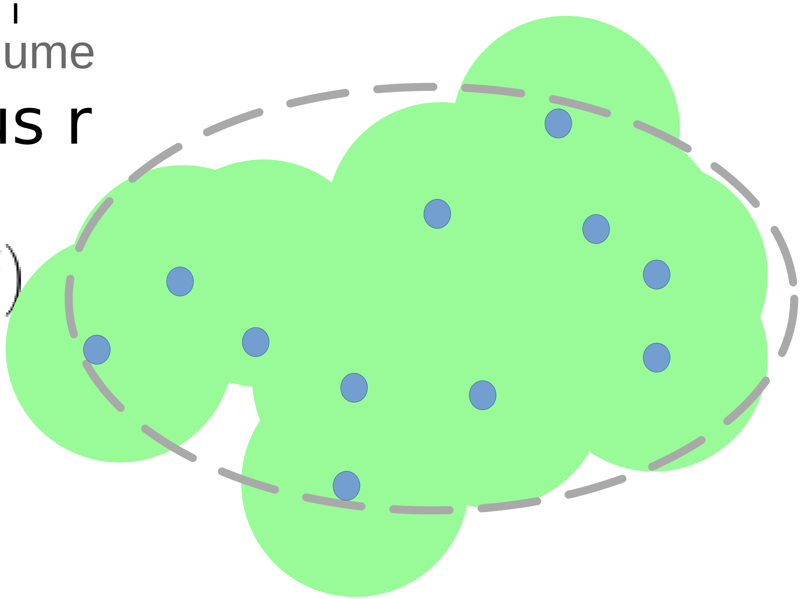
$$p_m < (1-p_1)^m \times K$$

$$p_m = 10^{-6},\ K = 1000 \rightarrow m = 45$$

Analysing region construction

- Homogeneous Poisson Point Process
- within contour “Intensity” $\lambda = K / V_i$

  Number of live points / Current prior volume
- Sphere around live point with radius r



no other live point nearby: $P(<r) = 1 - \exp\left(-\lambda V_\mathrm{d} r^d\right)$

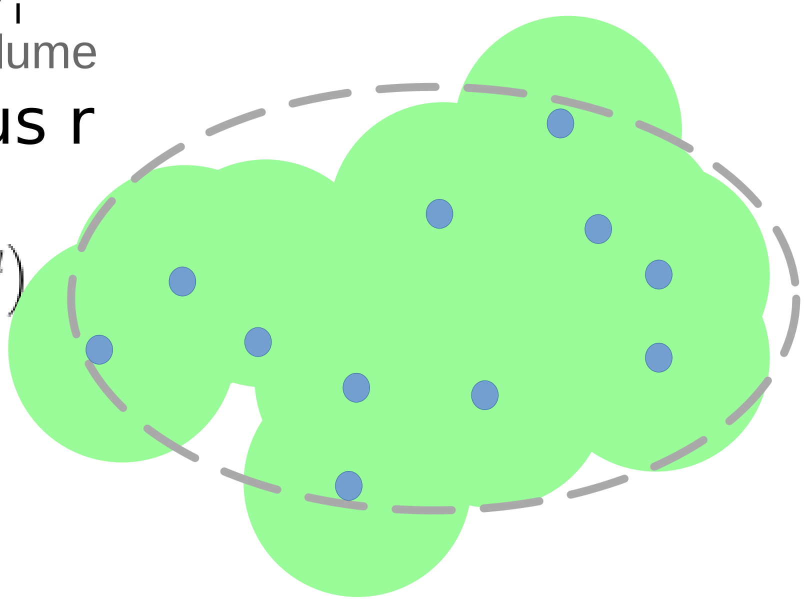
with unit n-sphere volume: $V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}$

Buchner (in prep)

Analysing region construction

- Homogeneous Poisson Point Process
- within contour "Intensity" $\lambda = K / V_i$

  Number of live points / Current prior volume
- Sphere around live point with radius r



no other live point nearby: $P(<r) = 1 - \exp\left(-\lambda V_\mathrm{d} r^d\right)$

with unit n-sphere volume: $V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}$

Radius from m bootstrap rounds with K live points $r_{\max}^d \approx \frac{\ln\left(\frac{2}{3}Km\right)}{\frac{1}{3}K} \times \frac{V_i}{V_d}$

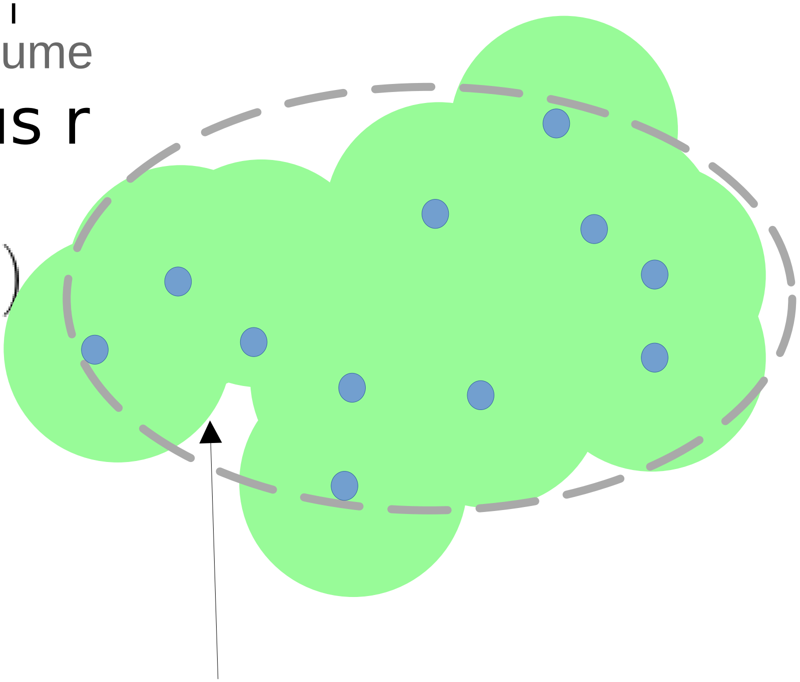
Buchner (in prep)

Analysing region construction

- Homogeneous Poisson Point Process
- within contour "Intensity" λ = K / V_i

Number of live points / Current prior volume

- Sphere around live point with radius r



no other live point nearby: $P(<r) = 1 - \exp\left(-\lambda V_d r^d\right)$

with unit n-sphere volume: $V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}+1\right)}$

Radius from m bootstrap rounds with K live points $r_{\max}^d \approx \frac{\ln\left(\frac{2}{3}Km\right)}{\frac{1}{3}K} \times \frac{V_i}{V_d}$

$$P^{\mathrm{missed}} = \exp\left(-\lambda V_d r_{\max}^d\right) = \left(\frac{2}{3}Km\right)^{-3}$$

$$p_m = 10^{-6},\ K = 1000 \rightarrow P^{\mathrm{missed}} = 4 \times 10^{-14}$$

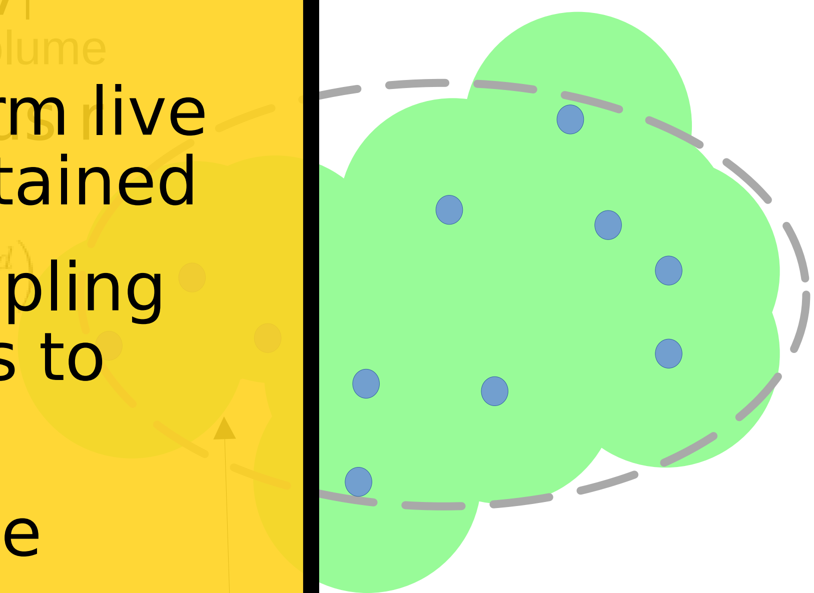
Buchner (in prep)

Analysing region construction

- Homogeneous Poisson Point Process
- within contour "Intensity" $\lambda = K / V_i$
  Number of live points / Current prior volume
- Sphere around live point with radius r

Convergence proof

- at each iteration, a uniform live point distribution is maintained
- By induction, nested sampling with MLFriends converges to posterior & evidence
- With implementable, finite compute
- Usual flag-pole caveat for all Monte Carlo algorithms (V/K resolution)



$= \left(\frac{2}{3}Km\right)^{-3}$

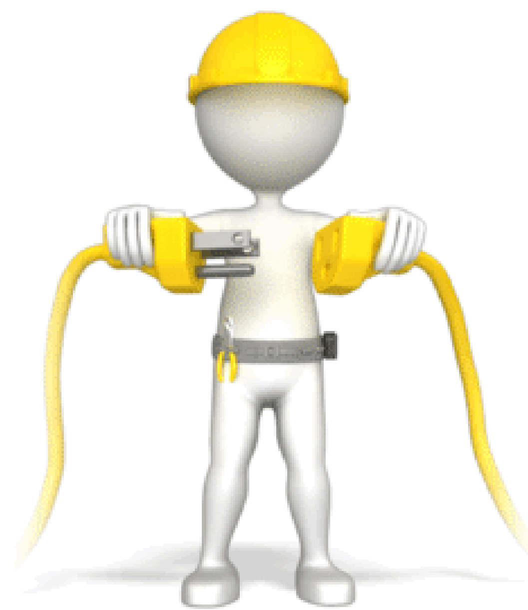
$P^{\mathrm{missed}} = 4 \times 10^{-14}$

Buchner (in prep)

So what is BXA?

Idea: make
physical parameter inference &
model comparison
easy & practical

XMM2Athena
319,565 X-ray sources
processed (Webb+23)



BXA

parallelisation,
resuming
sophisticated, robust

inference engine

based on nested sampling

MultiNest
UltraNest

community models
fully-fledged
data formats

fitting
environment

sherpa
pyxspec
(threeml)
(spex)

+ background models
+ some visualisation tools

Extra slides

- Landscape of X-ray spectral fitting
- Rules of thumb for UltraNest
- Parameter distributions
  from many data sets

An evolving software landscape

- Xspec, Sherpa, SPEX, 3ML
  - maintainance is institutional effort
  - Xspec models a community focal point
- 2014: BXA: xspec/sherpa plug-in for modern inference algorithms
- 2022: Model emulators
- 2024+: Diff PPL: e.g. jaxspec
  - Require re-implementing models!
- Missing?
  - partially diff XSF? → fastXSF
  - Spectral component emulators + BXA?

Buchner+14

Kerzendorf+22
Matzeu+22

Dupourqué+24
Barret+24

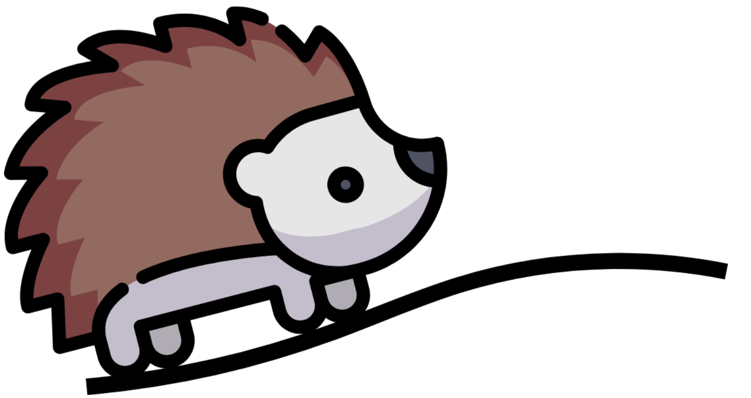
Great to see activity!

IACHEC statistics

Some nested sampling papers

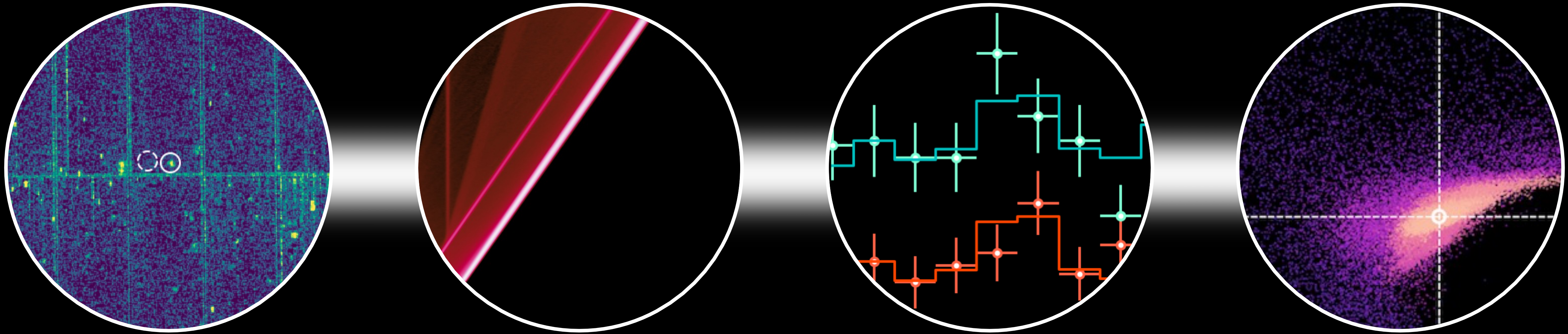
#	Bibcode	Date	Cited	Title / Author	Note
1	2016S&C....26..383B	2016/01	cited: 159	A statistical test for Nested Sampling algorithms Buchner, Johannes	MultiNest biases RadFriends algorithm
2	2019PASP..131j8005B	2019/10	cited: 209	Collaborative Nested Sampling: Big Data versus Complex Physical Models Buchner, Johannes	MLFriends algorithm (affine invariant form)
3	2021JOSS....6.3001B	2021/04	cited: 413	UltraNest - a robust, general purpose Bayesian inference engine Buchner, Johannes	UltraNest software paper
4	2021JOSS....6.3045B	2021/05	cited: 6	Bayesian X-ray Analysis (BXA) v4.0 Buchner, Johannes	BXA software paper
5	2022PSFor...5...46B	2022/12	cited: 4	Comparison of Step Samplers for Nested Sampling Buchner, Johannes	Performance comparison of step sampler proposals
6	2023StSur..17..169B	2023	cited: 107	Nested Sampling Methods Buchner, Johannes	Systematic literature review
7	2024arXiv240211936B	2024/02	cited: 3	Relative Jump Distance: a diagnostic for Nested Sampling Buchner, Johannes	stuck step samplers diagnostic

Rules of thumb for UltraNest



- Do inference correct once is faster than quick & dirty heuristics that need many verification simulations
- Read the documentation :)
- Number of live points O(1000)
- If d>20, use a step sampler
  - RJD diagnostic
- Priors do not have to be uniform – smooth edges may be useful to avoid rerunning
- Define your question well

Merci de votre attention! **Avez-vous des questions?**



Johannes Buchner

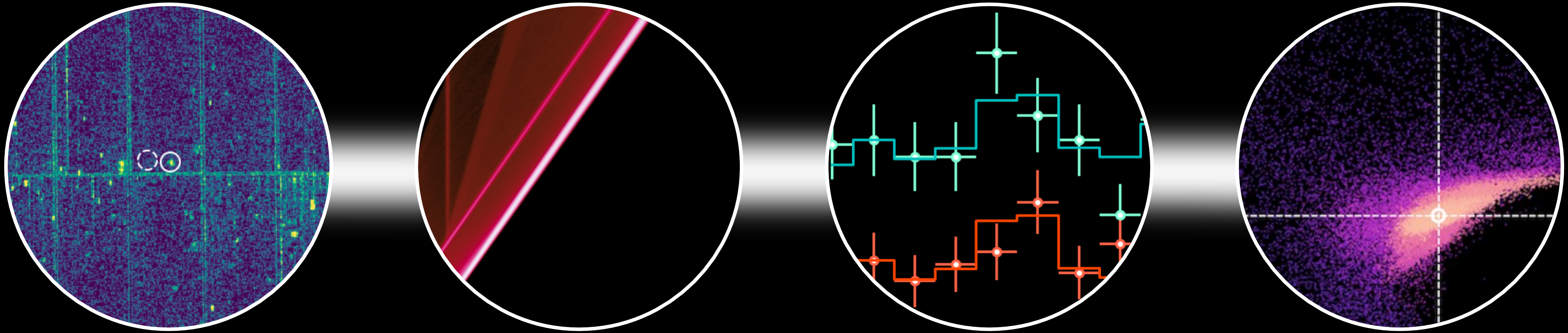
jbuchner@mpe.mpg.de

astrost.at/istics

Peter Boorman

boorman@mpe.mpg.de

peterboorman.com



peterboorman.com/tutorial_bxa