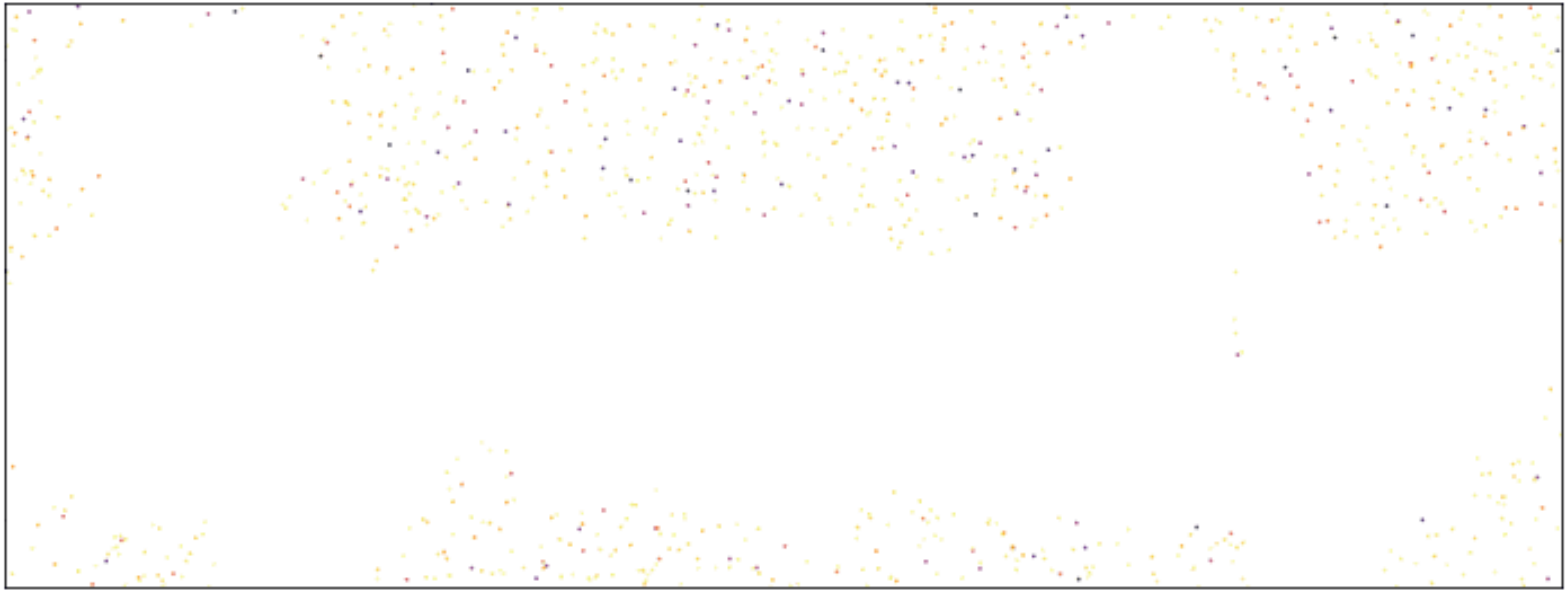
Introduction to Nested Sampling



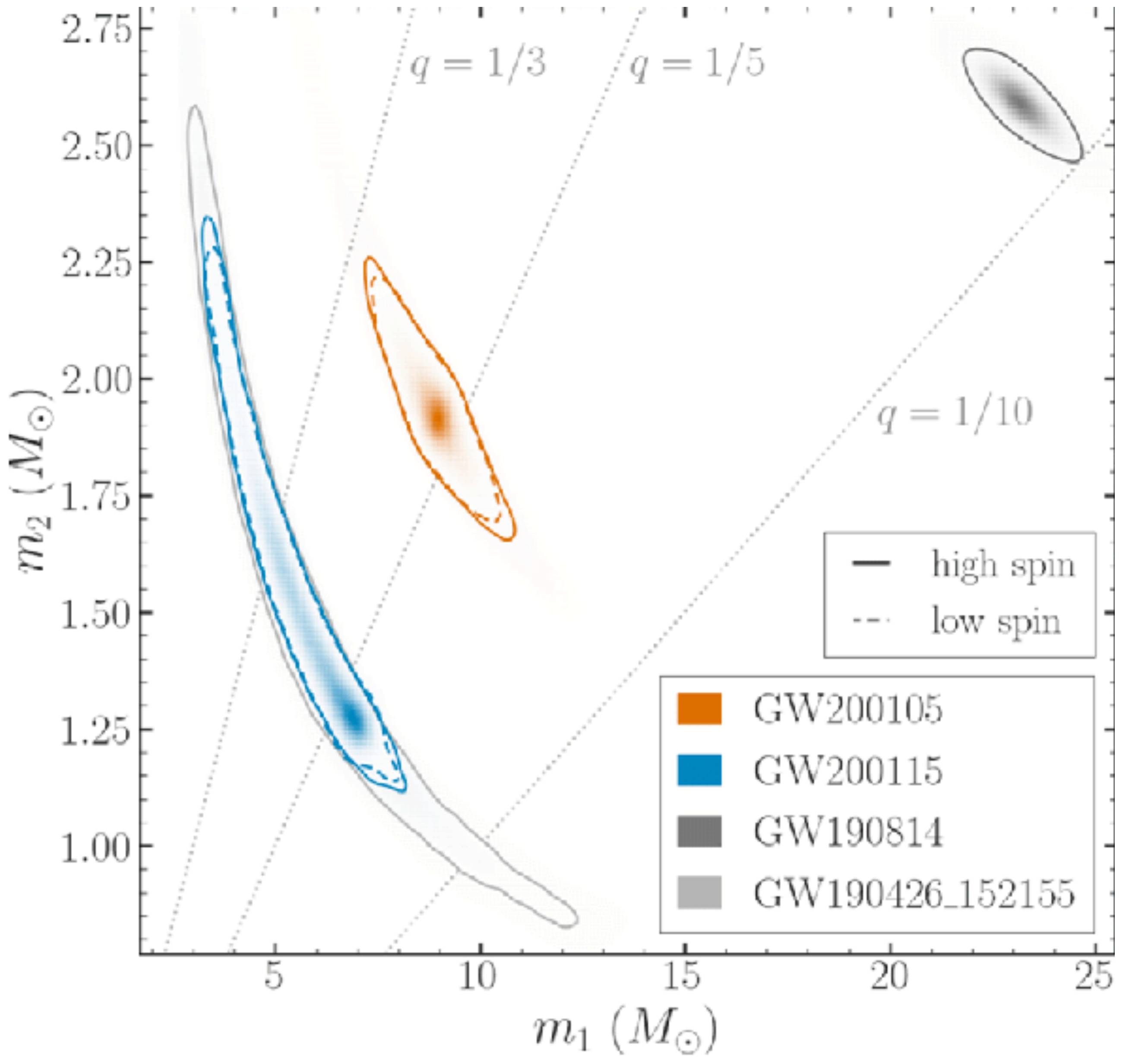
Dynesty is one of the package for Nested Sampling

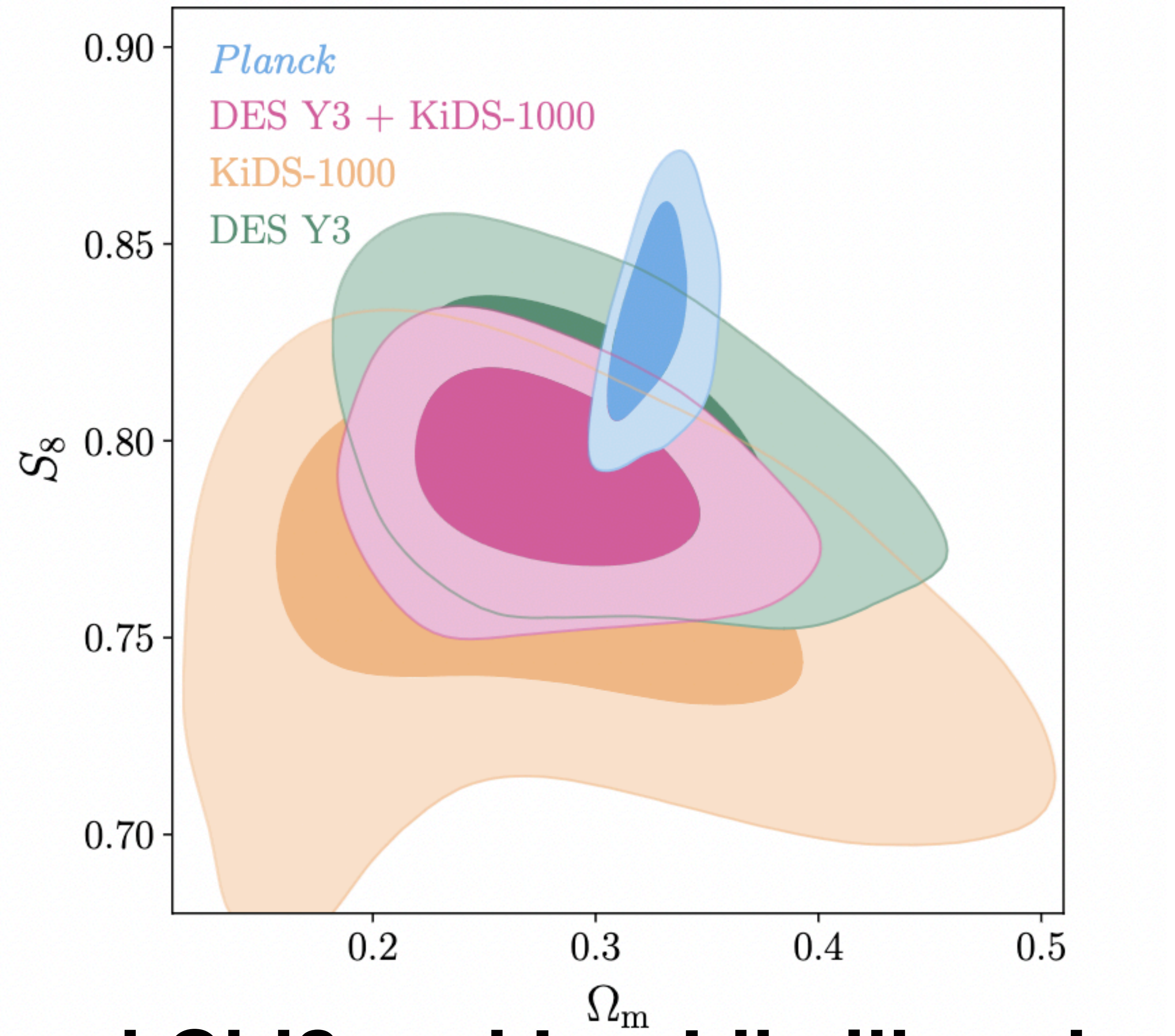
What I've understood so far ...

Fabio Acero,

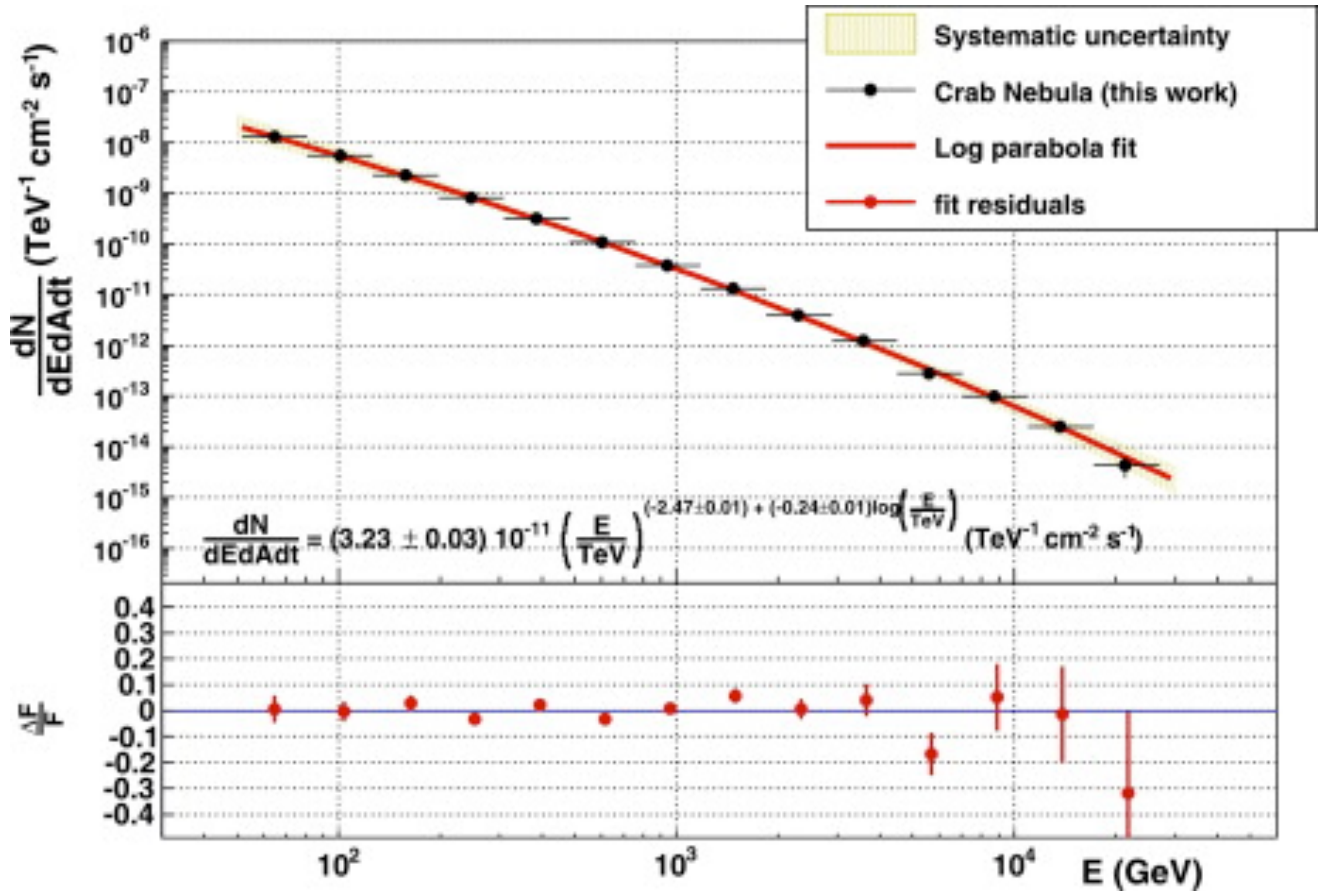
A global shift towards Bayesian frameworks

- **Bayesian analyses have been dominant in cosmology for 1 or 2 decades**
  - Complex models
  - Complex dataset.
  - Need to combine multiple datasets





- **In High-Energy astrophysics we're mostly loyal to Minuit and old fashioned Chi2 and LogLikelihood**
  - Complex sources & instruments
  - But simpler models (POW) ?
  - But moving towards Bayesian

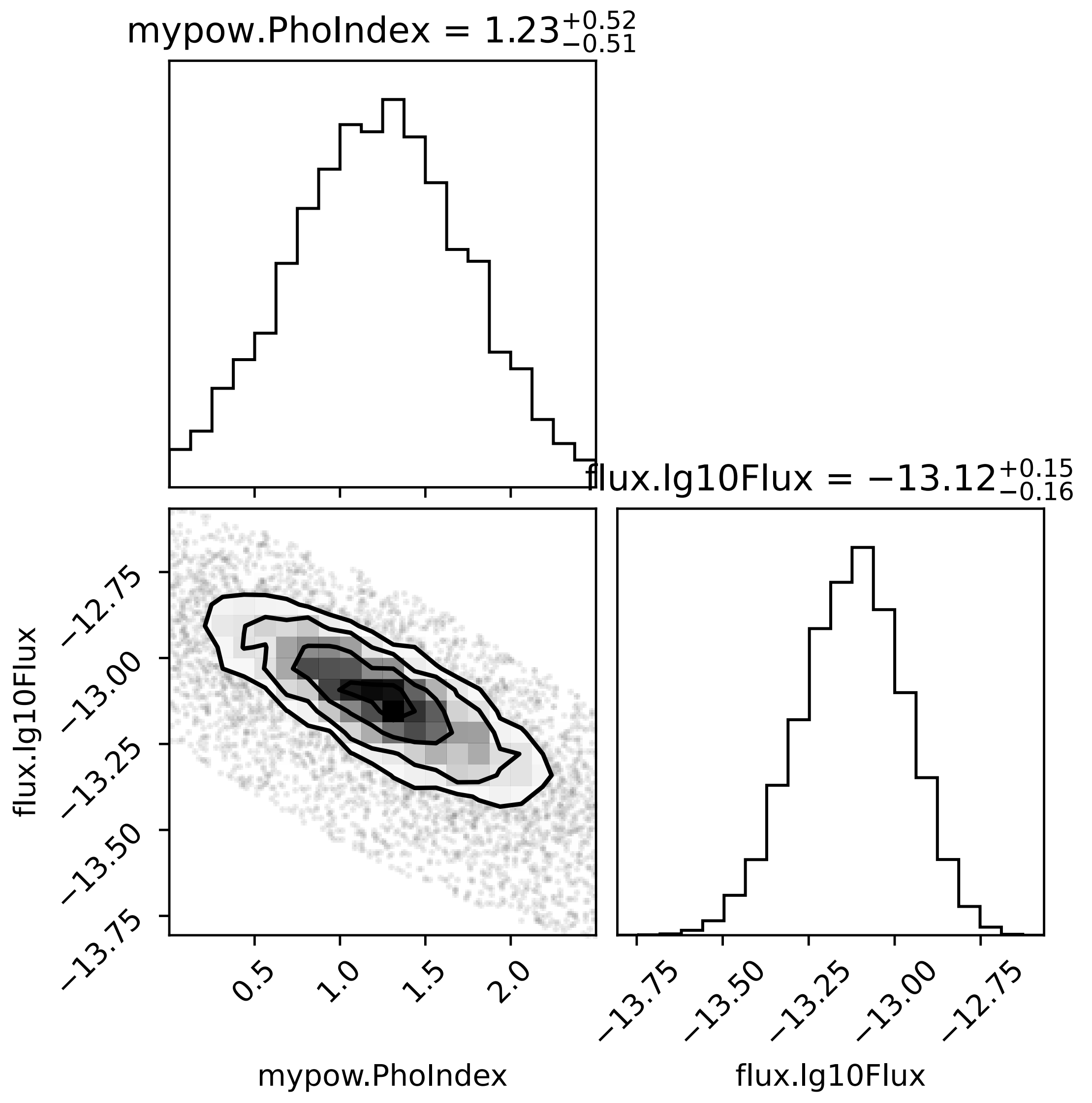


What is a Bayesian analysis (user point of view)

Priors — **Data** — **Posteriors**

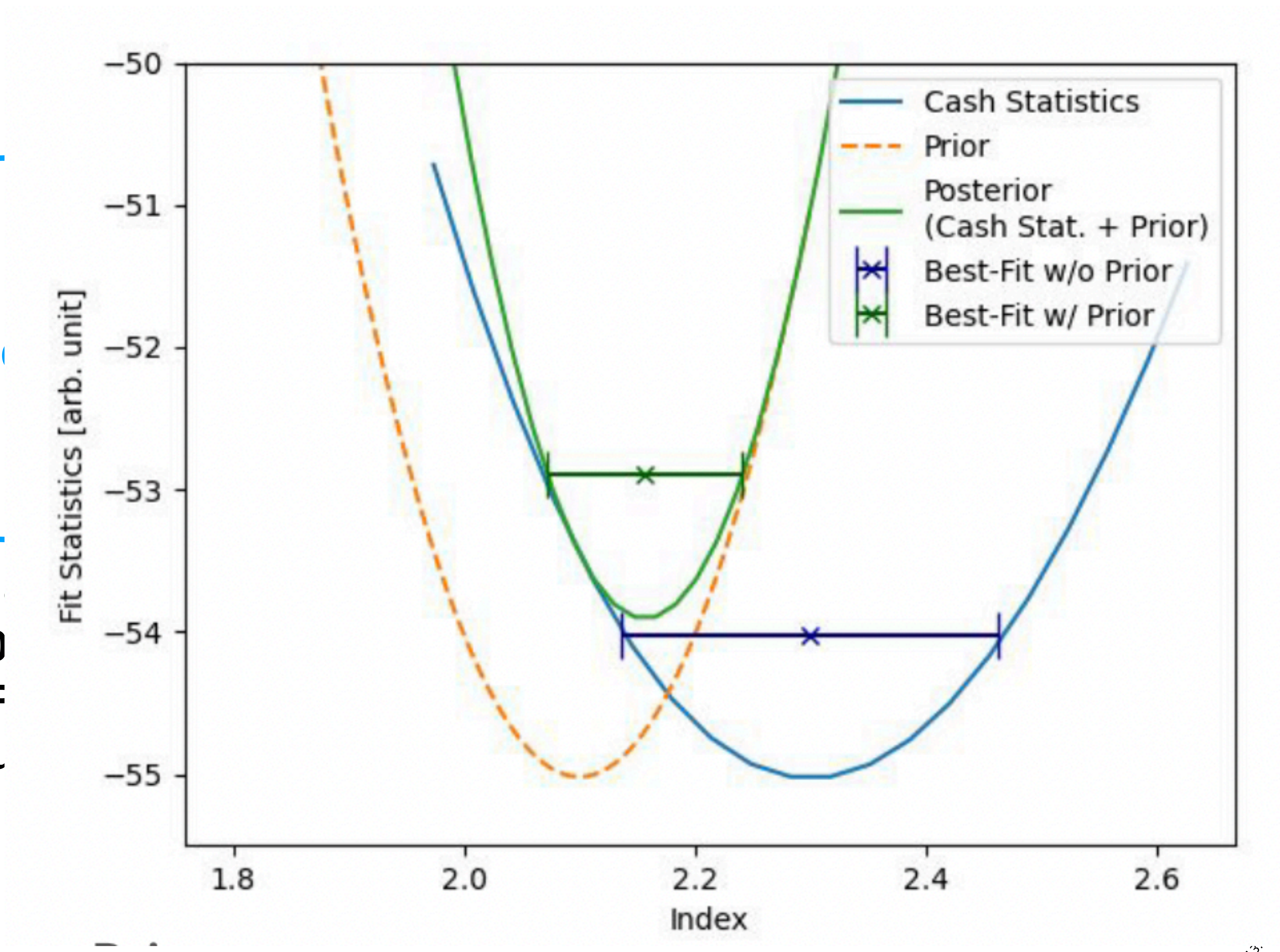
- **Prior information on parameters**
- **E.g. : Norm > 0**
- **P1 < Param < P2**
- **Norm_bkg Gaussian(1, sigma)**

- **Modified Likelihood**
- **Likelihood = data_term + priors**



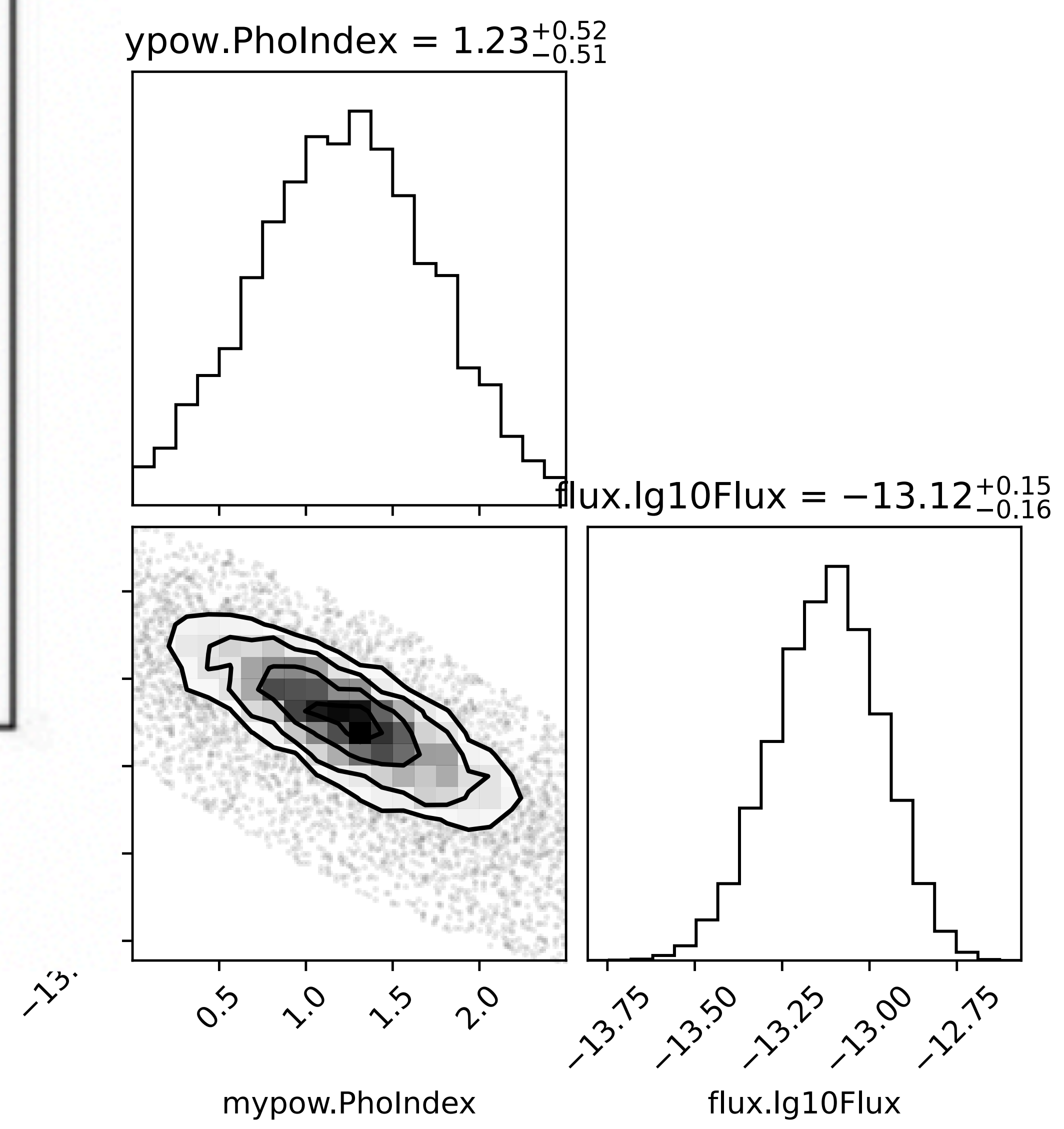
What is a Bayesian analysis (user point of view)

Pri



osteriors

- Prior informati
- E.g. : Norm > 0
- P1 < Param < P
- Norm_bkg Gau

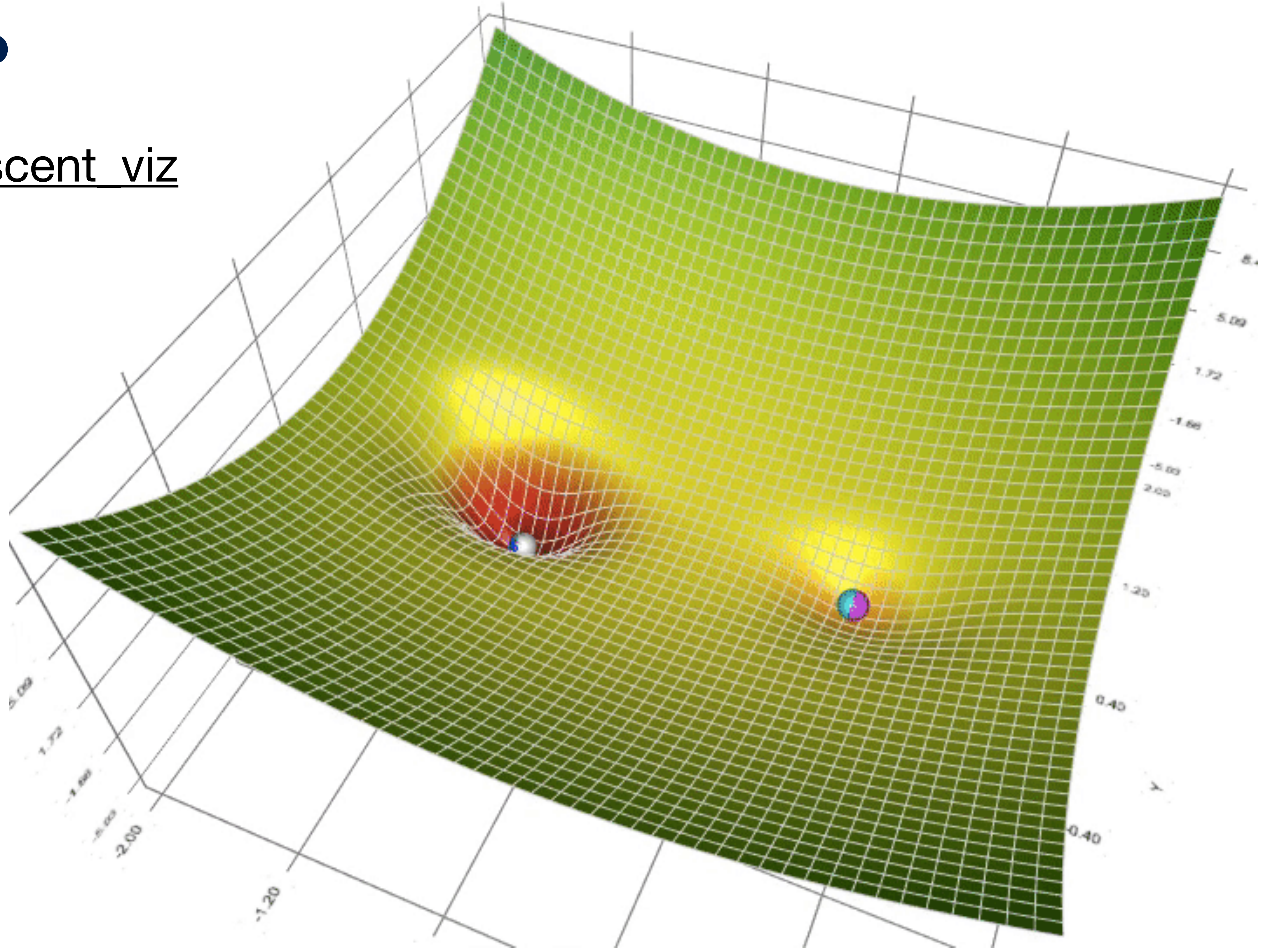


What is a fit ?

- **Need a loss function + minimizer :**
  - **Gradient Descent (e.g. Scipy minimize, iMinuit, sherpa fit, etc)**
  - **Markov chain Monte Carlo**
  - **Nested Sampling metho**

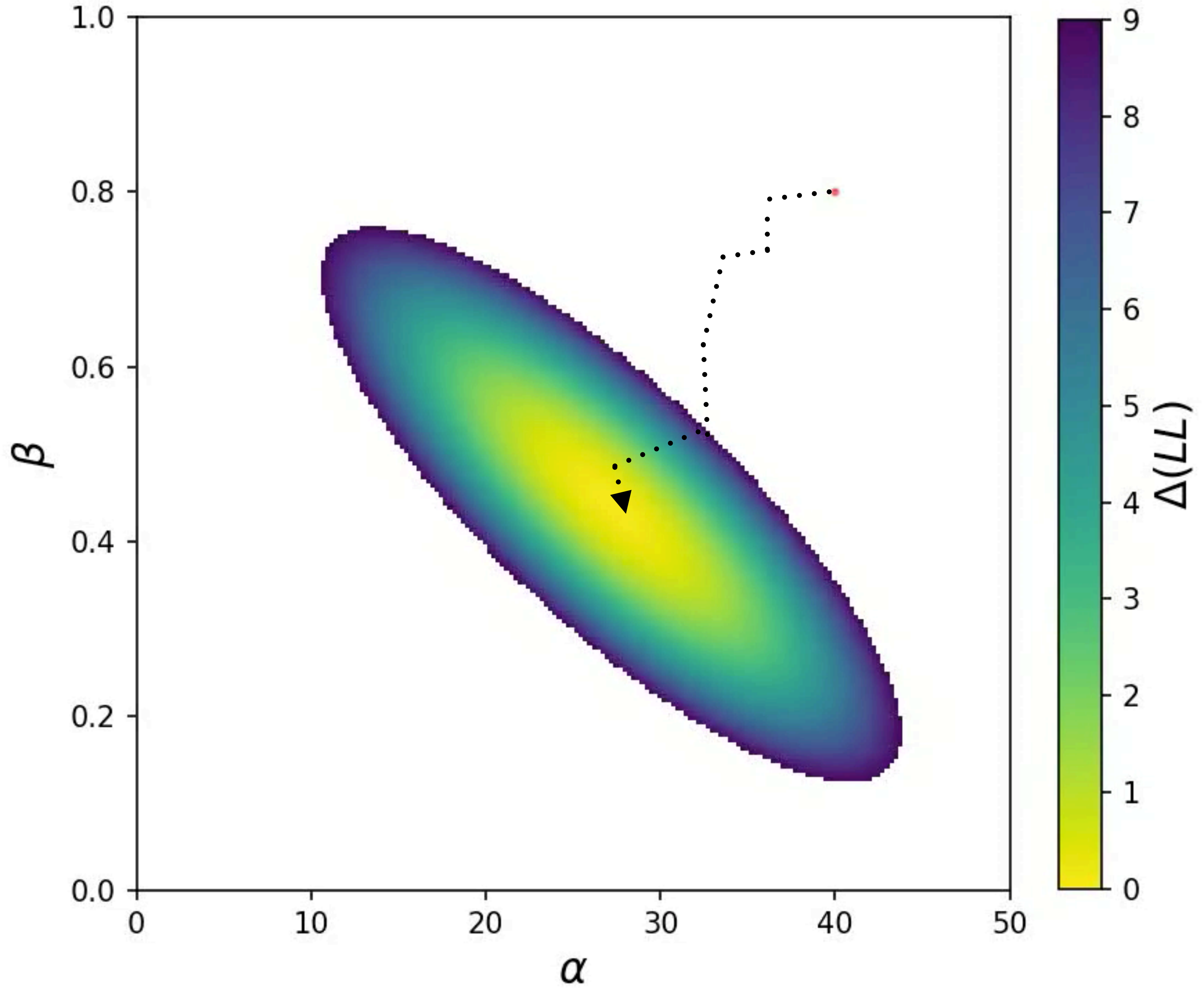
github.com/lilipads/gradient_descent_viz

Never take a best-fit for granted



Traditional fitting method

- **Gradient descent based method**
  - **Levenberg Marquardt**
  - **Migrad in Minuit for example**
  - **Gradient is estimated numerically at each step**



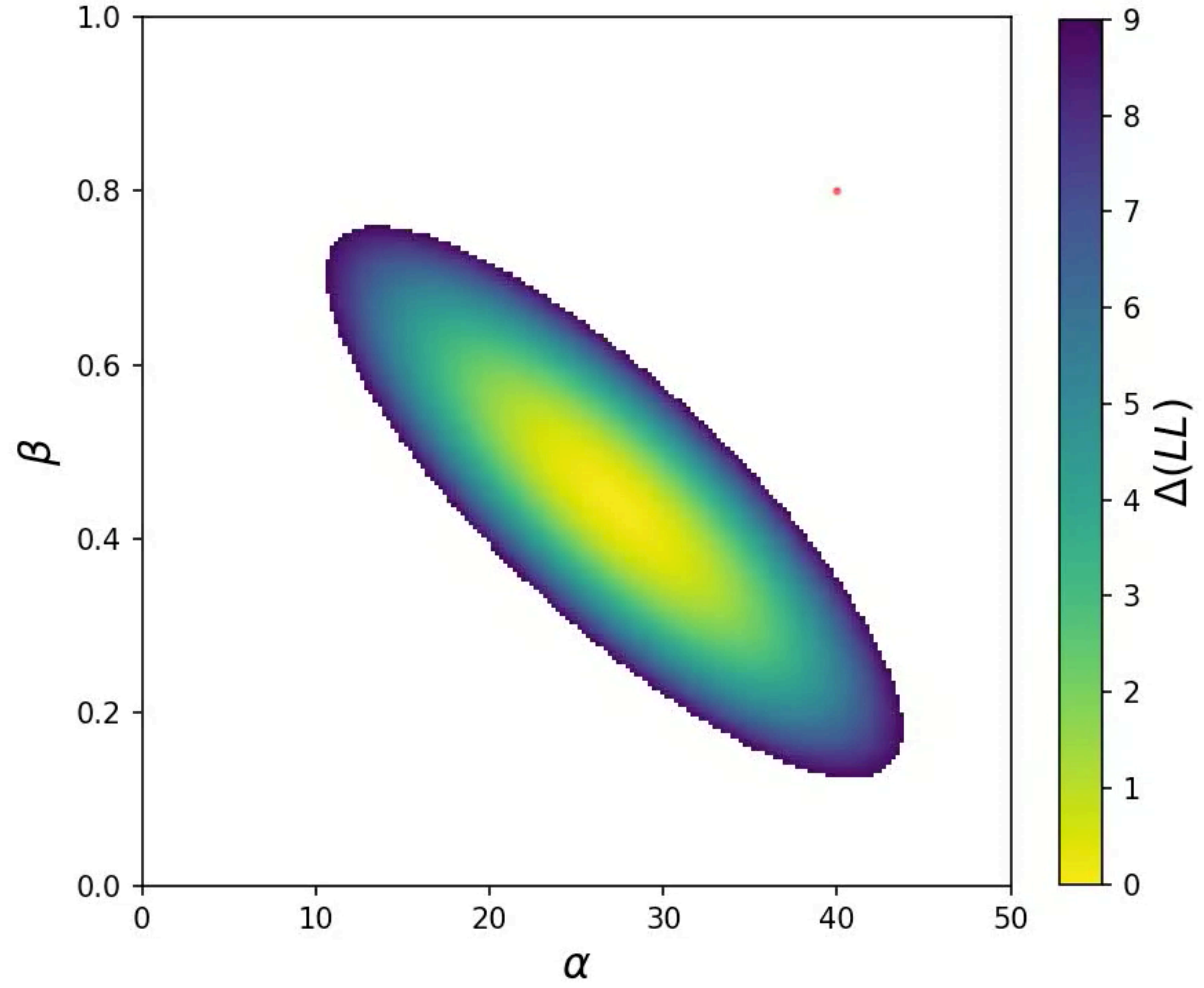
Once at best-fit stops
No information about local likelihood
Sometimes fit fails :

Migrad

FCN = -2.676e+07		Nfcn = 378		
EDM = 0.00356 (Goal: 2e-06)		time = 2.3 sec		
INVALID Minimum		No Parameters at limit		
ABOVE EDM threshold (goal x 10)		Below call limit		
Covariance	Hesse ok	APPROXIMATE	NOT pos. def.	FORCED

Markov Chain Monte Carlo (MCMC)

- **What are they:**
  - **Monte Carlo: samples are used to approximate the probability distribution**
  - **Markov Chain: semi-random walk in potential**
  - **Walkers explore the local likelihood**



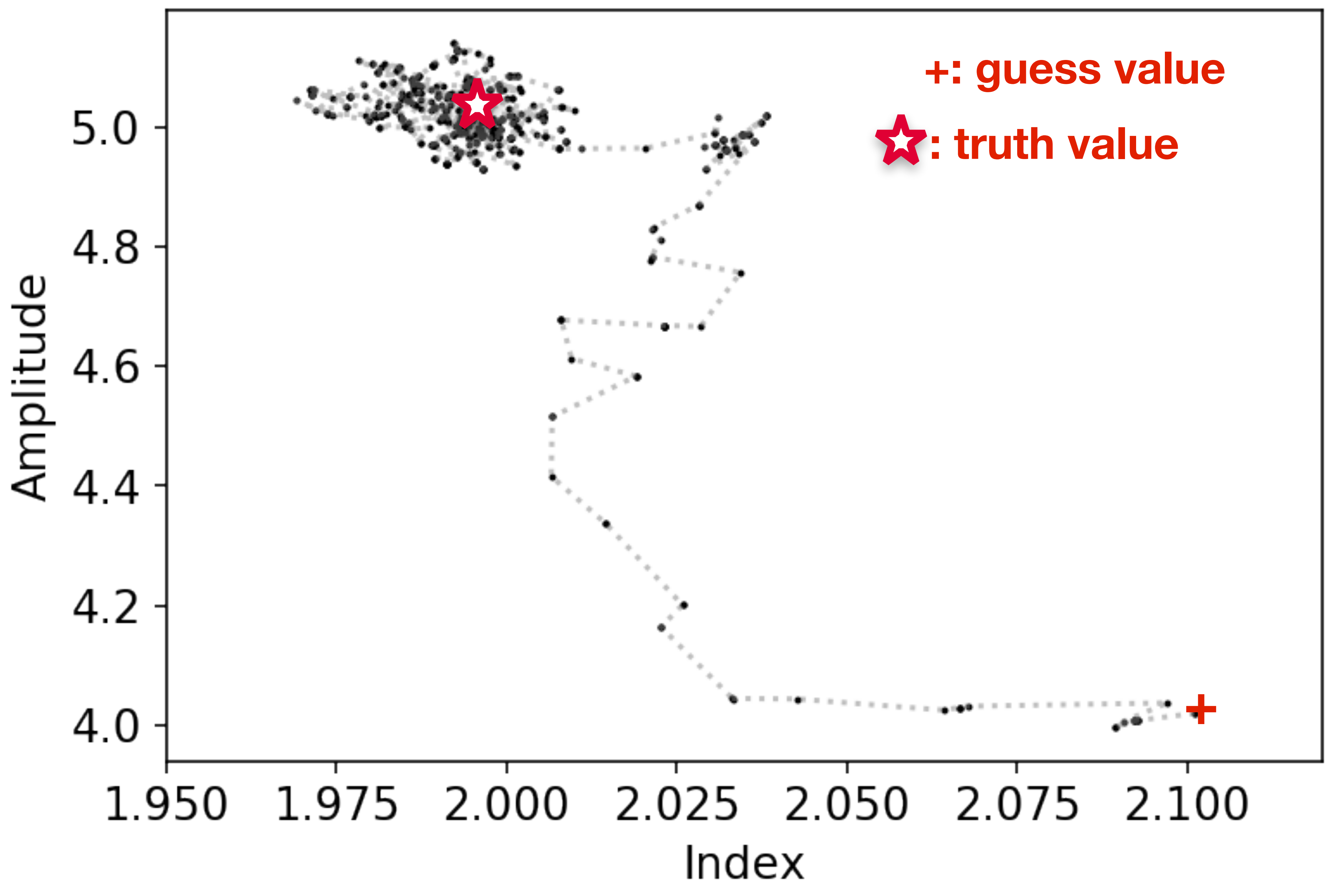
Random walk directed by potential (likelihood) spend most of their time in interesting region

Technically not a fit (no convergence) it's a phase space parameter exploration

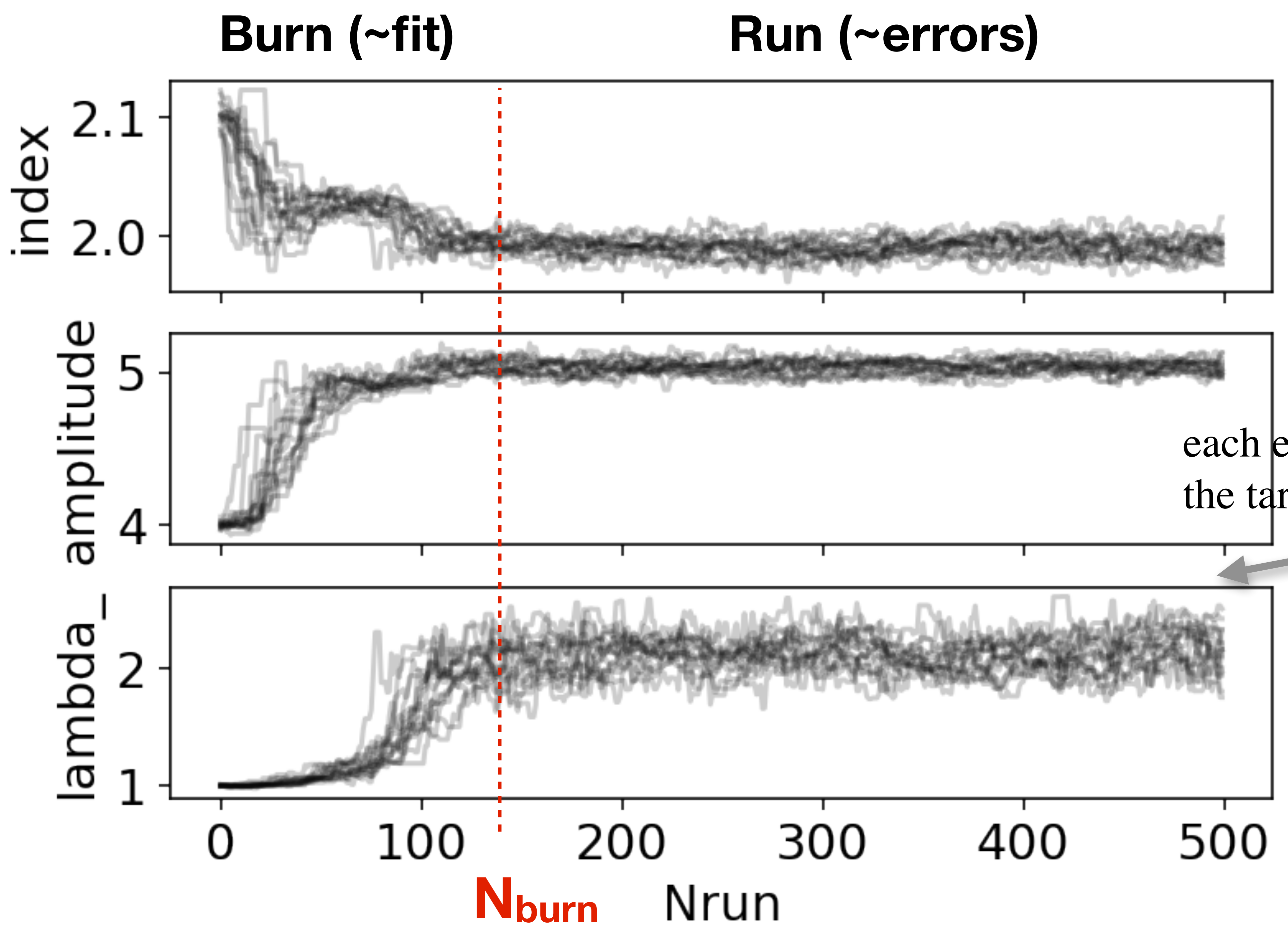
MCMC issues : inside out approach

- **Need to define an initialization point**
  - **Might stay trapped in local minima. Could miss multimodal posterior**
- **Need to define "by eye" an burn period**
  - **No clear automated criteria**
- **Need to evolve walkers for "sufficiently long" to sample target posterior**
  - **How long is good enough ?**

1 walker evolving for 500 steps



10 walkers evolving for 500 steps

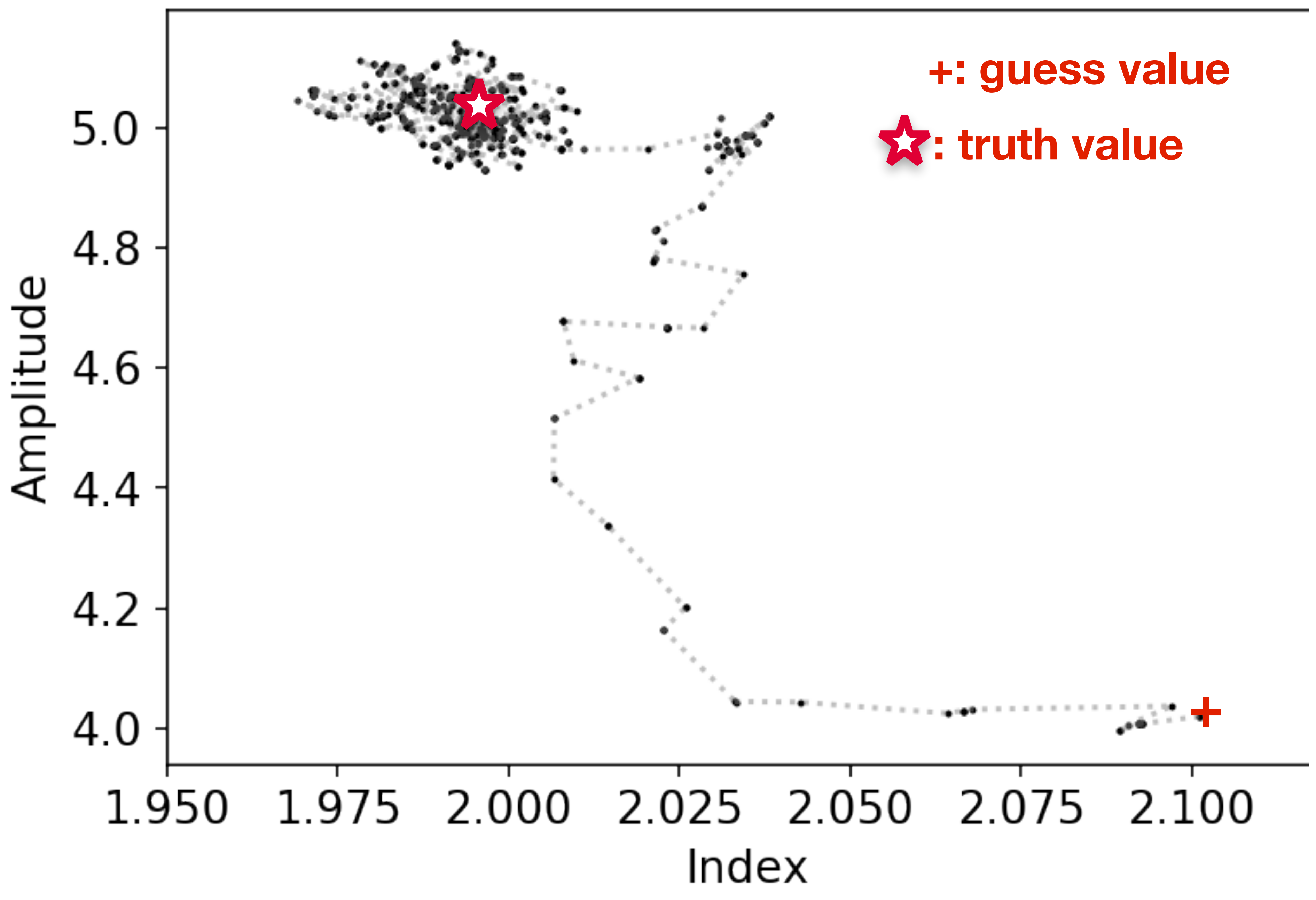


each element of the chain are samples of the target posterior distribution.

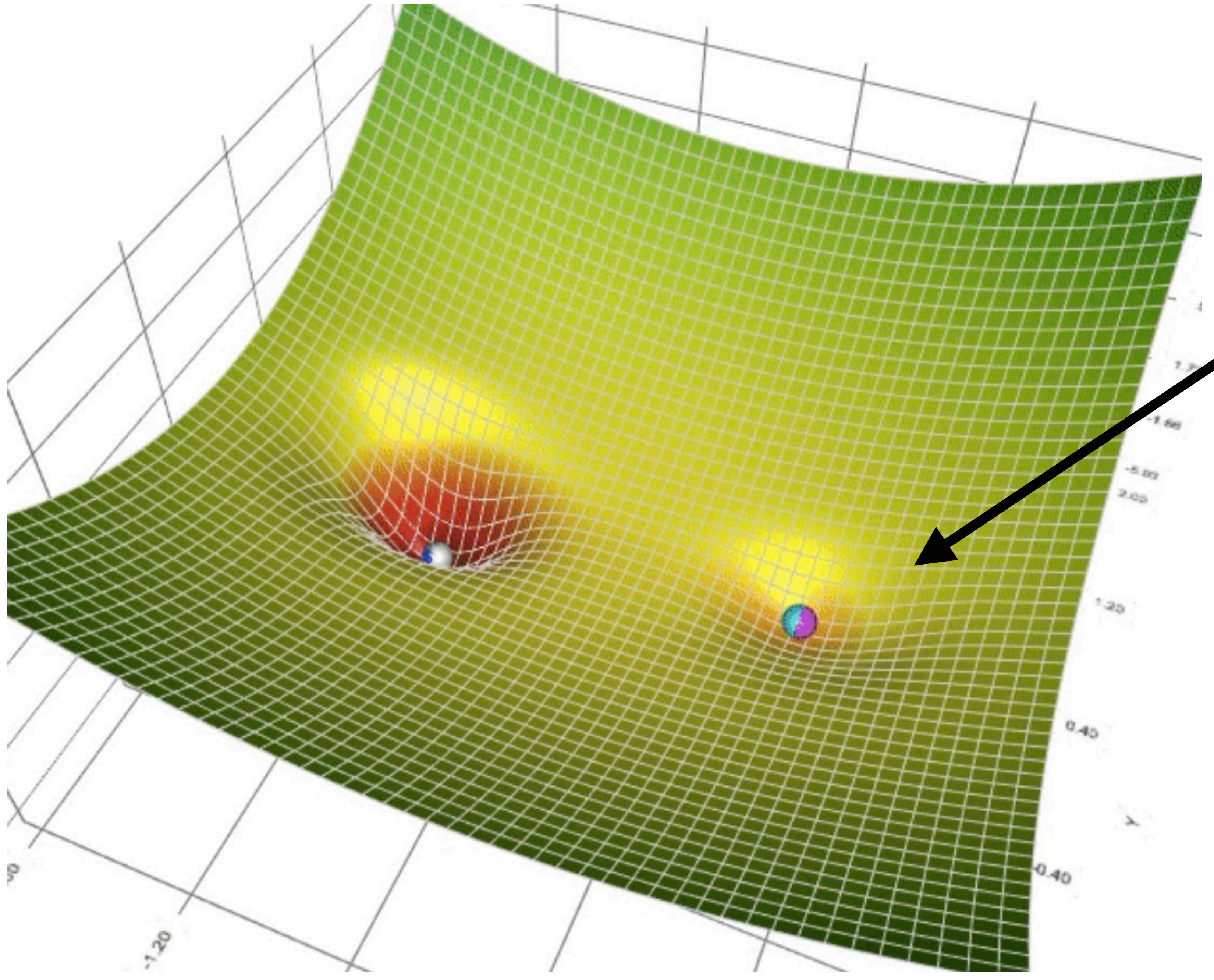
MCMC issues : inside out approach

- **Need to define an initialization point**
  - **Might stay trapped in local minima. Could miss multimodal posterior**
- **Need to define "by eye" an burn period**
  - **No clear automated criteria**
- **Need to evolve walkers for "sufficiently long" to sample target posterior**
  - **How long is good enough ?**

1 walker evolving for 500 steps



10 walkers evolving for 500 steps



But what if all your walkers end up here ?

Nested Sampling methods

- **Nested Sampling (Skilling, 2004) is a Monte Carlo algorithm for estimating an integral over a model parameter space θ**
  - **Integral :**
    - **∫ Like(Data|θ) * Prior(θ) dθ = Z = Bayesian evidence**
  - **Z can be used to compare models even if <u>not</u> nested (morphology : mwl template vs disk)**
  - **This integral is also is also what will provide the normalized posterior distributions**
- **Main idea is integration is done by switching frame from many θ to a volume variable**

Unlike MCMC methods, which attempt to estimate the posterior $\mathcal{P}(\mathbf{\Theta})$ directly, Nested Sampling instead focuses on estimating the evidence

$$\mathcal{Z} \equiv \int_{\Omega_{\mathbf{\Theta}}} \mathcal{P}(\mathbf{\Theta}) \mathrm{d}\mathbf{\Theta} = \int_{\Omega_{\mathbf{\Theta}}} \mathcal{L}(\mathbf{\Theta}) \pi(\mathbf{\Theta}) \mathrm{d}\mathbf{\Theta} \tag{6}$$

As this integral is over the entire multi-dimensional domain of $\mathbf{\Theta}$, it is traditionally very challenging to estimate.

Nested Sampling approaches this problem by re-factoring this integral as one taken over prior volume X of the enclosed parameter space

$$\mathcal{Z} = \int_{\Omega_{\mathbf{\Theta}}} \mathcal{L}(\mathbf{\Theta}) \pi(\mathbf{\Theta}) \mathrm{d}\mathbf{\Theta} = \int_0^1 \mathcal{L}(X) \mathrm{d}X \tag{7}$$

Speagle, 2020

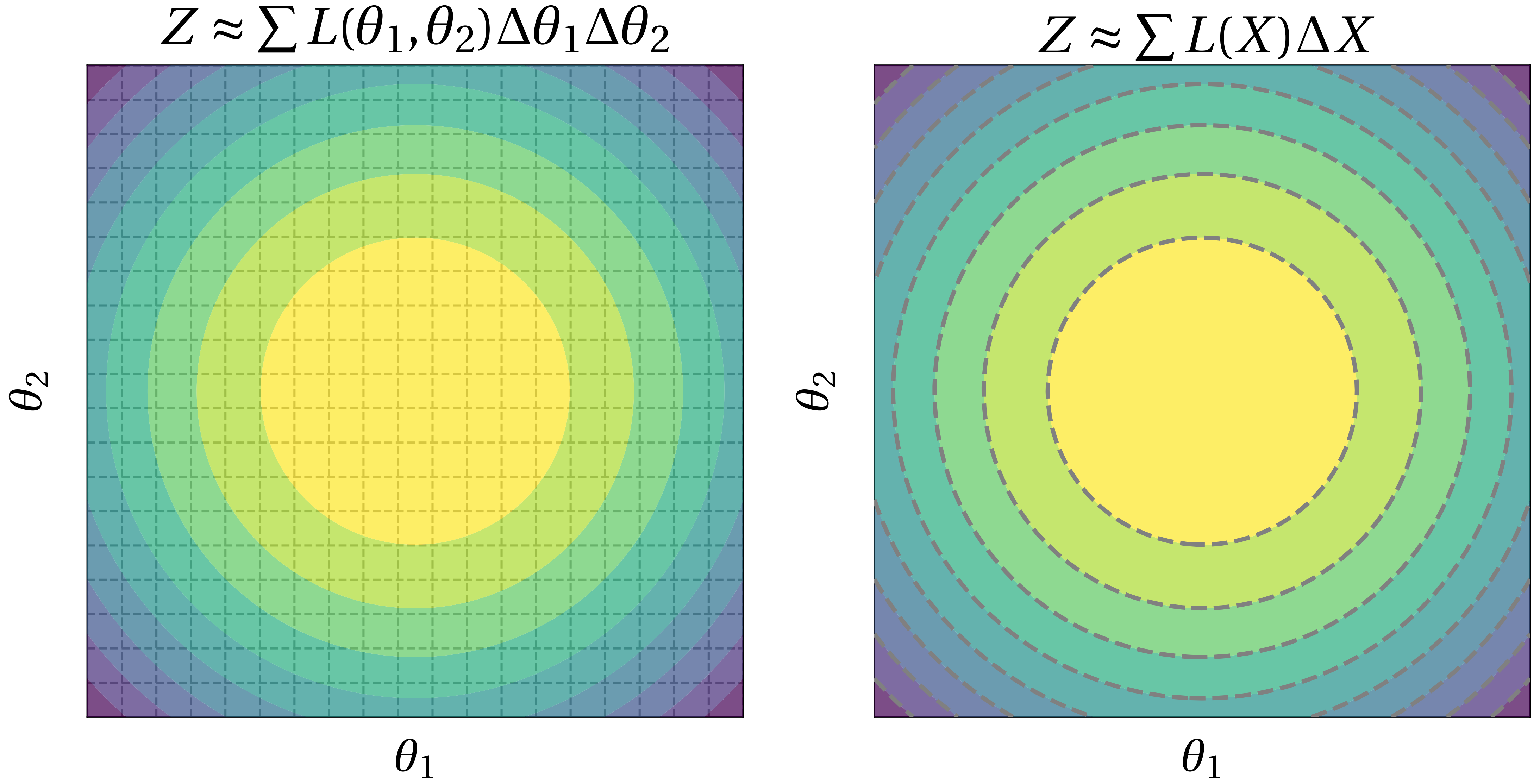
- **Analogy to spherical coordinates:**

$$\int \mathcal{P}(x, y, z) \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int \mathcal{P}(V(r)) \mathrm{d}V(r) = \int \mathcal{P}(r) 4\pi r^2 \mathrm{d}r$$

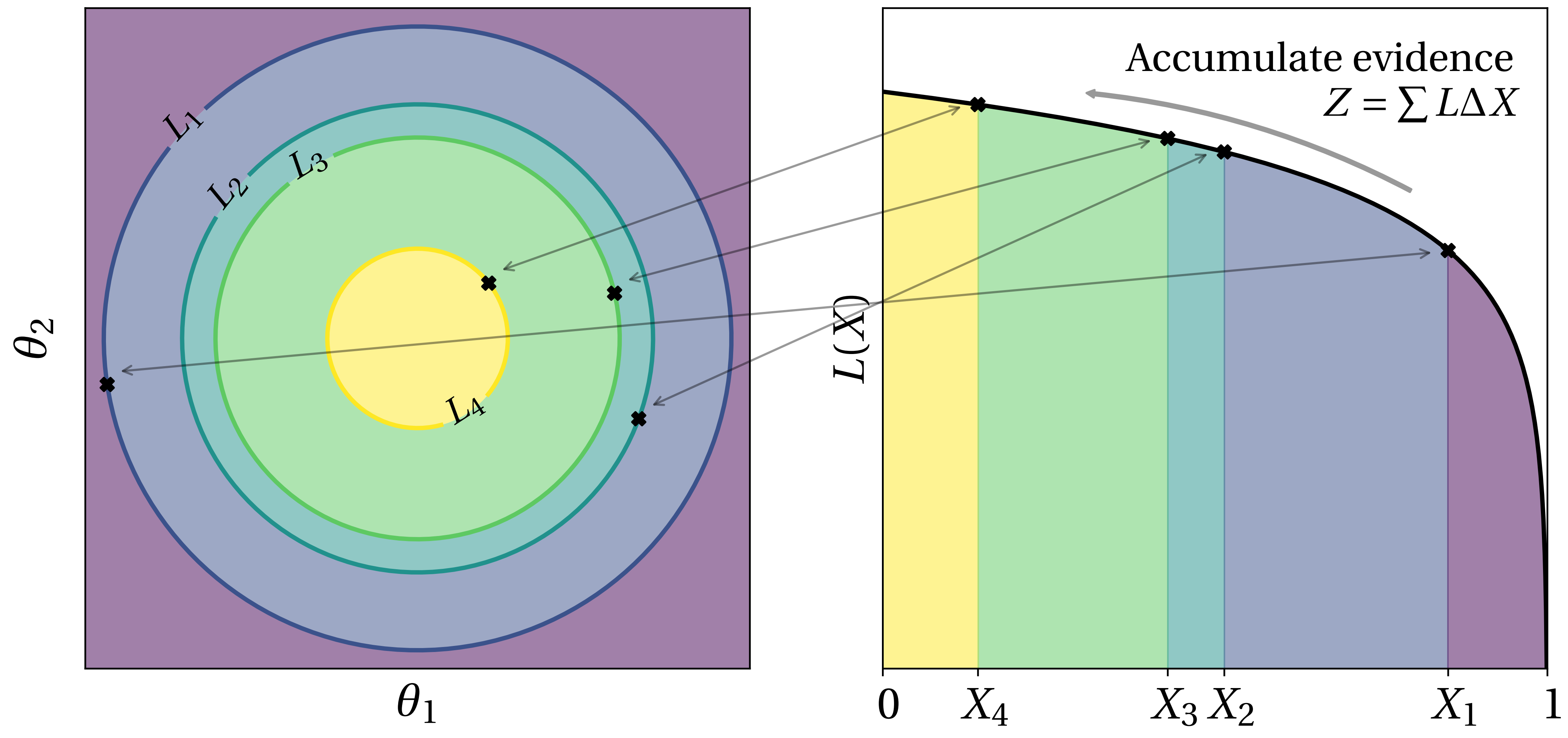
Nested means integrating in nested shells of iso-likelihood

- **Goal of the method is to integrate over shells (like an onion) to approximate the integral**

Direct space : more complex **Volume space :integration over shell of iso-LogLike**



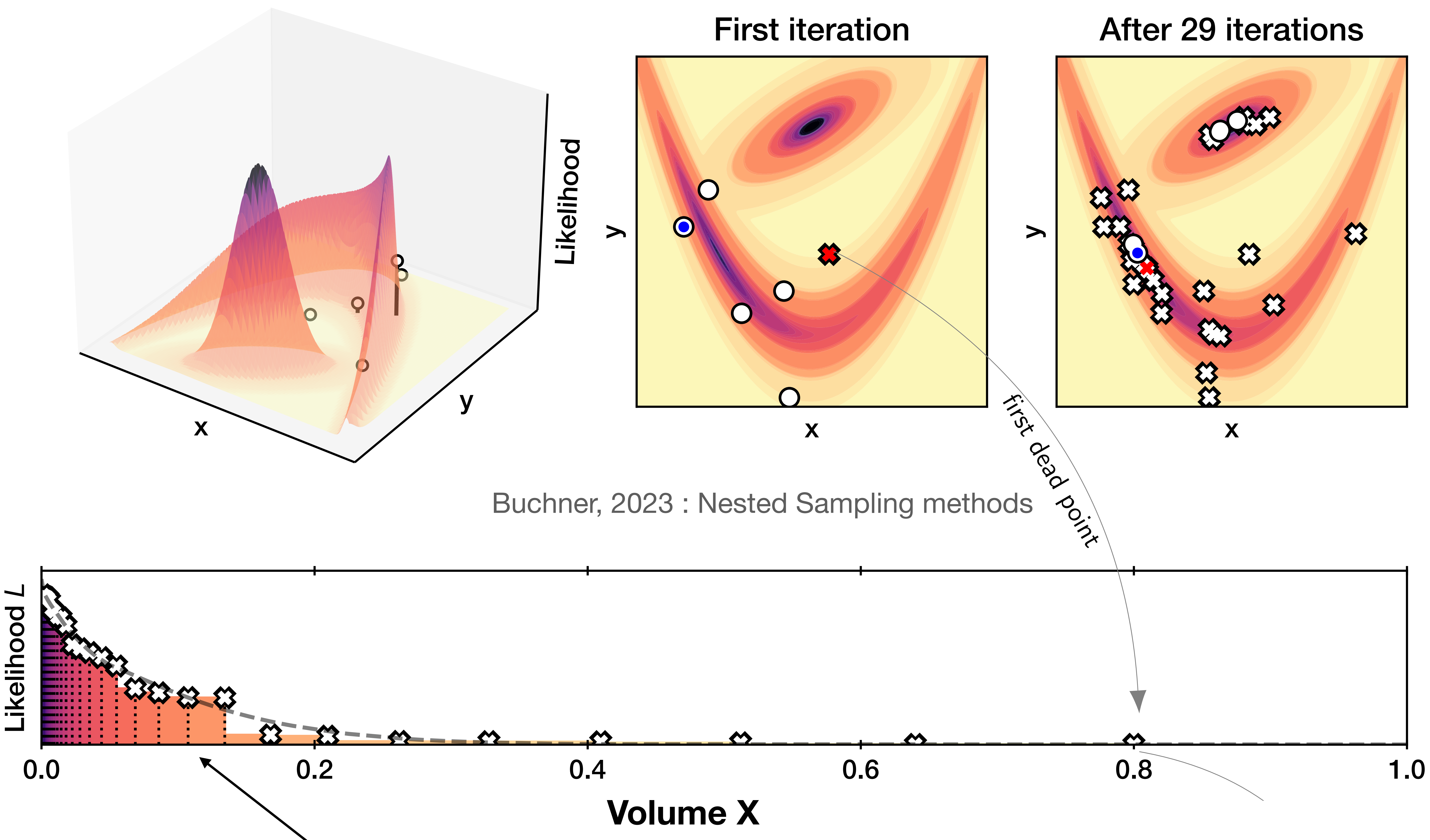
a | **The NS evidence identity.** The colours represent contours of a two-dimensional likelihood function. Rather than summing over little cubes (left), we combine cubes of similar likelihood together and sum over them (right).



b | **NS on a two dimensional problem.** We show the dead points and their iso-likelihood contours (left) and the corresponding contributions to the evidence integral (right). The volumes X_i are estimated statistically in NS.

Ashton, 2002 in Nature : Nested Sampling for physical scientists

Basic steps of Nested Sampling



Buchner, 2023 : Nested Sampling methods

$$Z = \sum L \Delta X$$ **Integration via trapezoid rule**

Z = Bayesian evidence = $\int$ Like(Data|θ) * Prior(θ) dθ

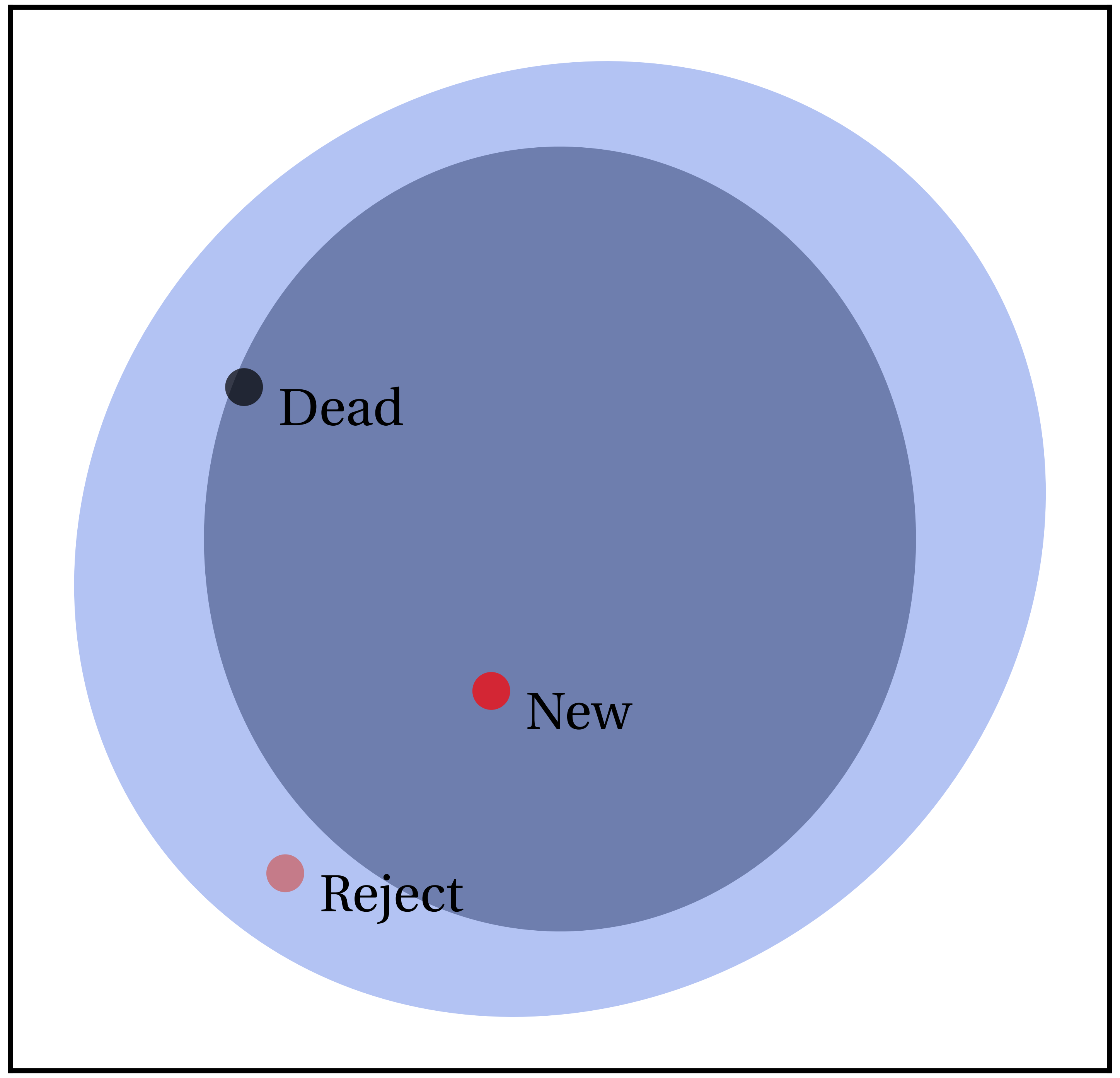
- **Define priors on params**
- **Transform param space to volume unity cube X**
- **Draw random uniform N_{live} (~400-1000)**

- **Iterate until stop criteria:**
  - **Remove worst LogLike point**
  - **Draw new point with a better LL**
    - **Likelihood-restricted prior sampling**
    - **(The tricky part)**
  - **Increment Z= Z + L*ΔX**
  - **Stop criteria : ΔZ/Z<tolerance**

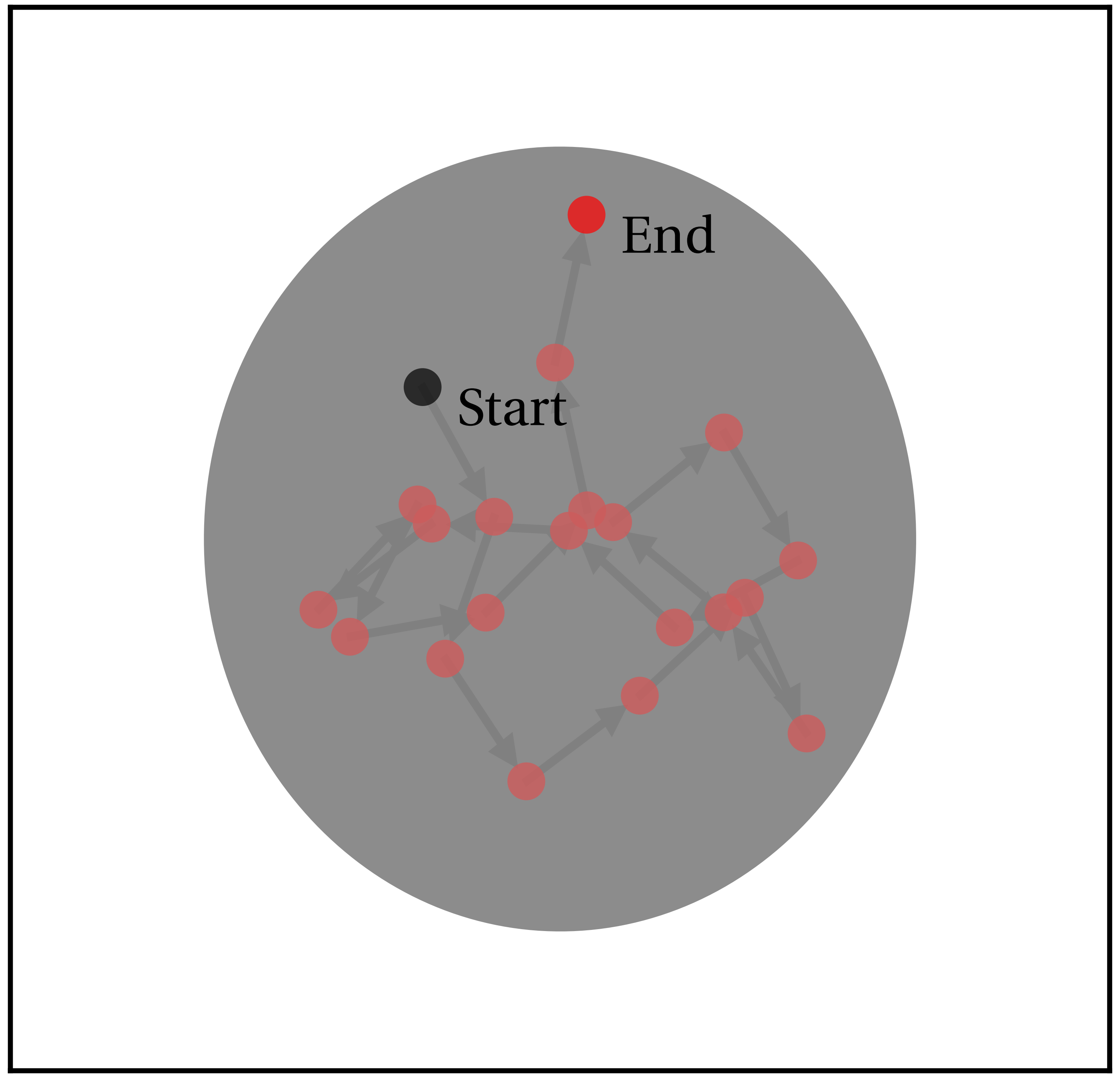
Sampling from a restricted prior ($Like_{new} > Like_{removed}$)

- The tricky part : how to sample points with a better LogLike
  - This means drawing point inside the iso LogLike contours that we don't know

- Take advantage from the fact that if NLive is large enough. Surviving points already provide trace the landscape
- So drawing an encapsulating ellipsoid and try to sample from this ellipsoid. Reject if needed
- Or if landscape too complex or high dimension use a step sampler



Ellipsoid sampling



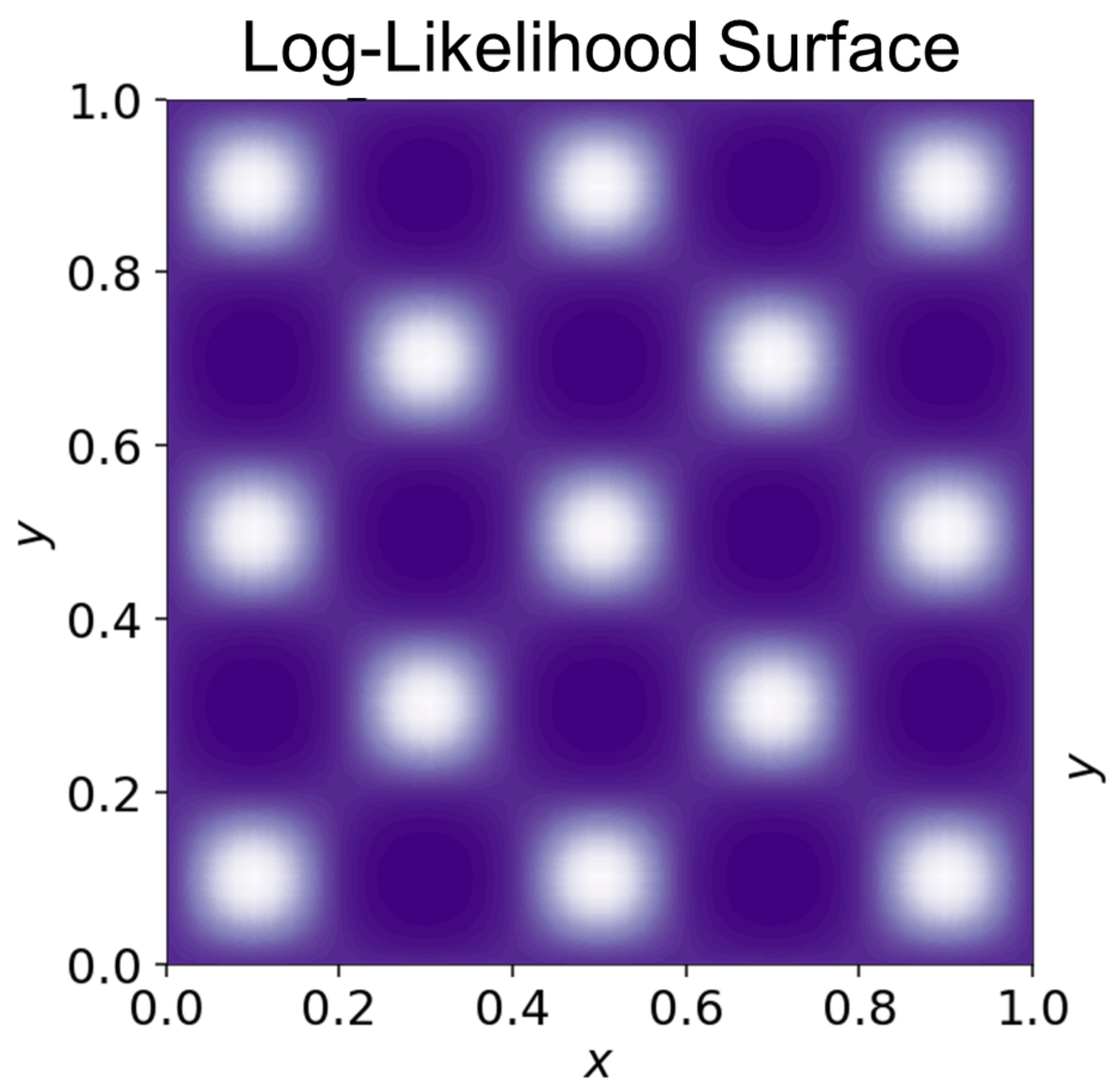
Random walk

Ashton, 2022

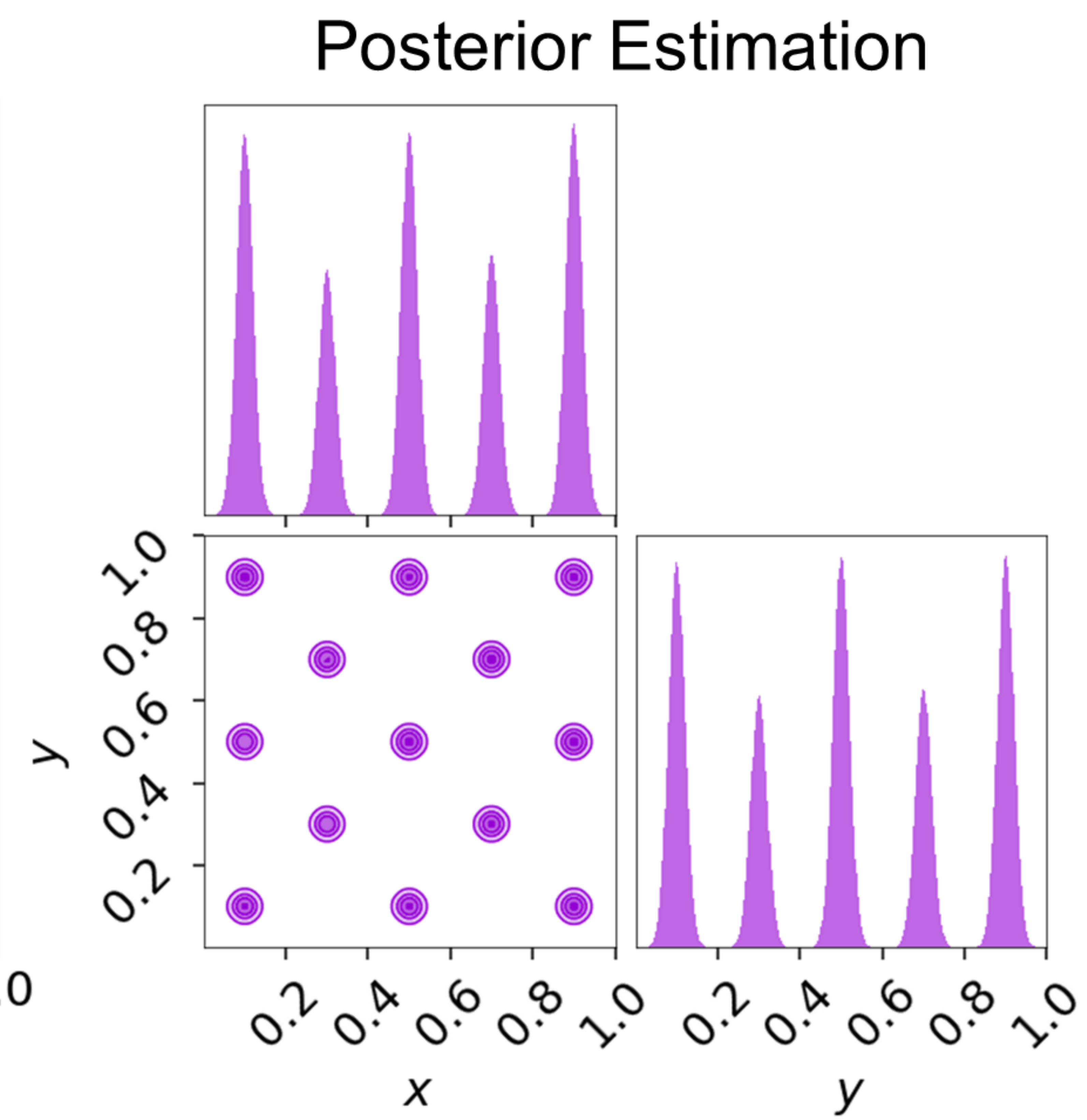
Complex landscapes

- **Works perfectly well in landscapes that would drive crazy any MCMC or gradient descent method**

Ground truth



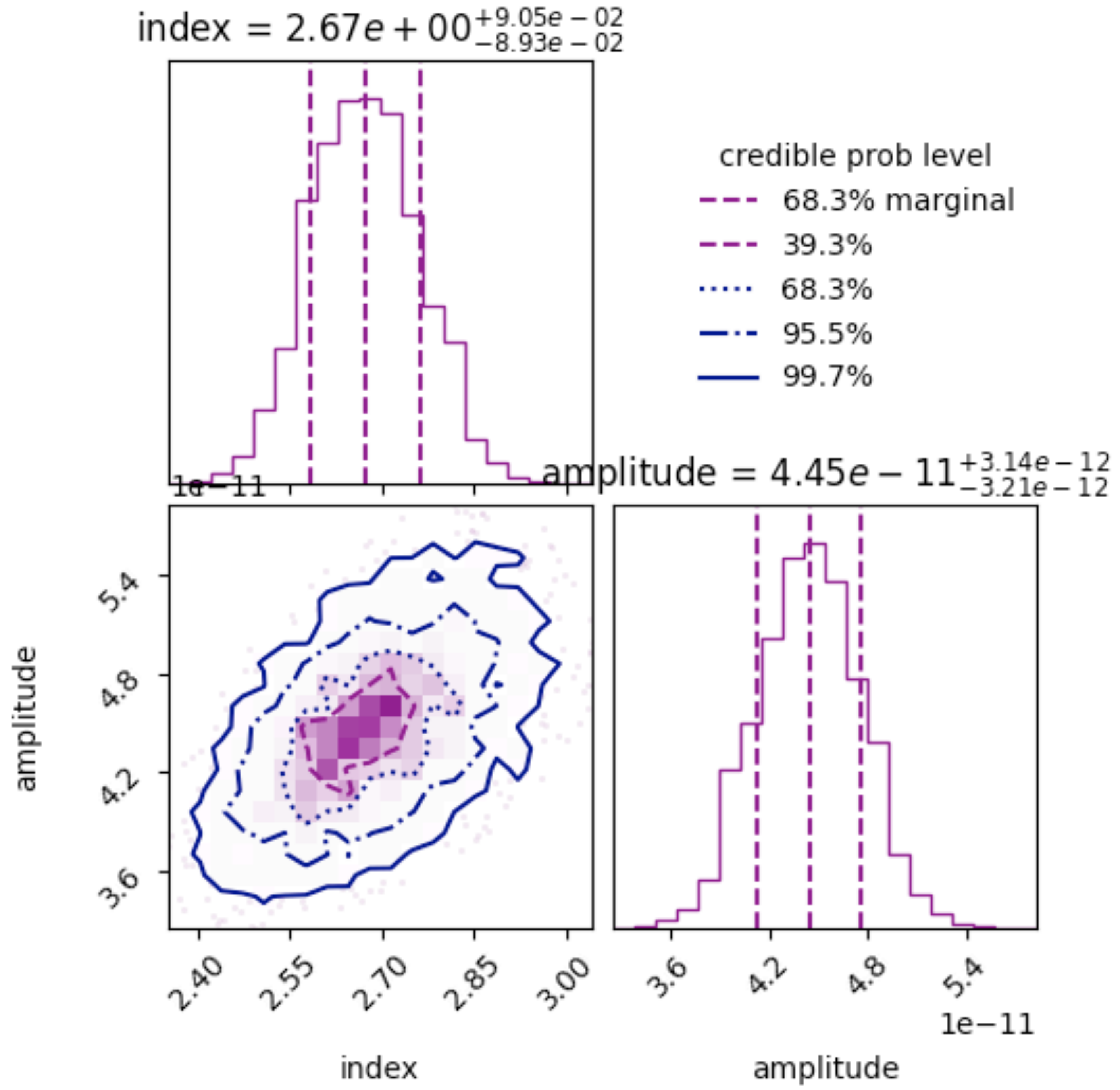
Reconstructed posterior

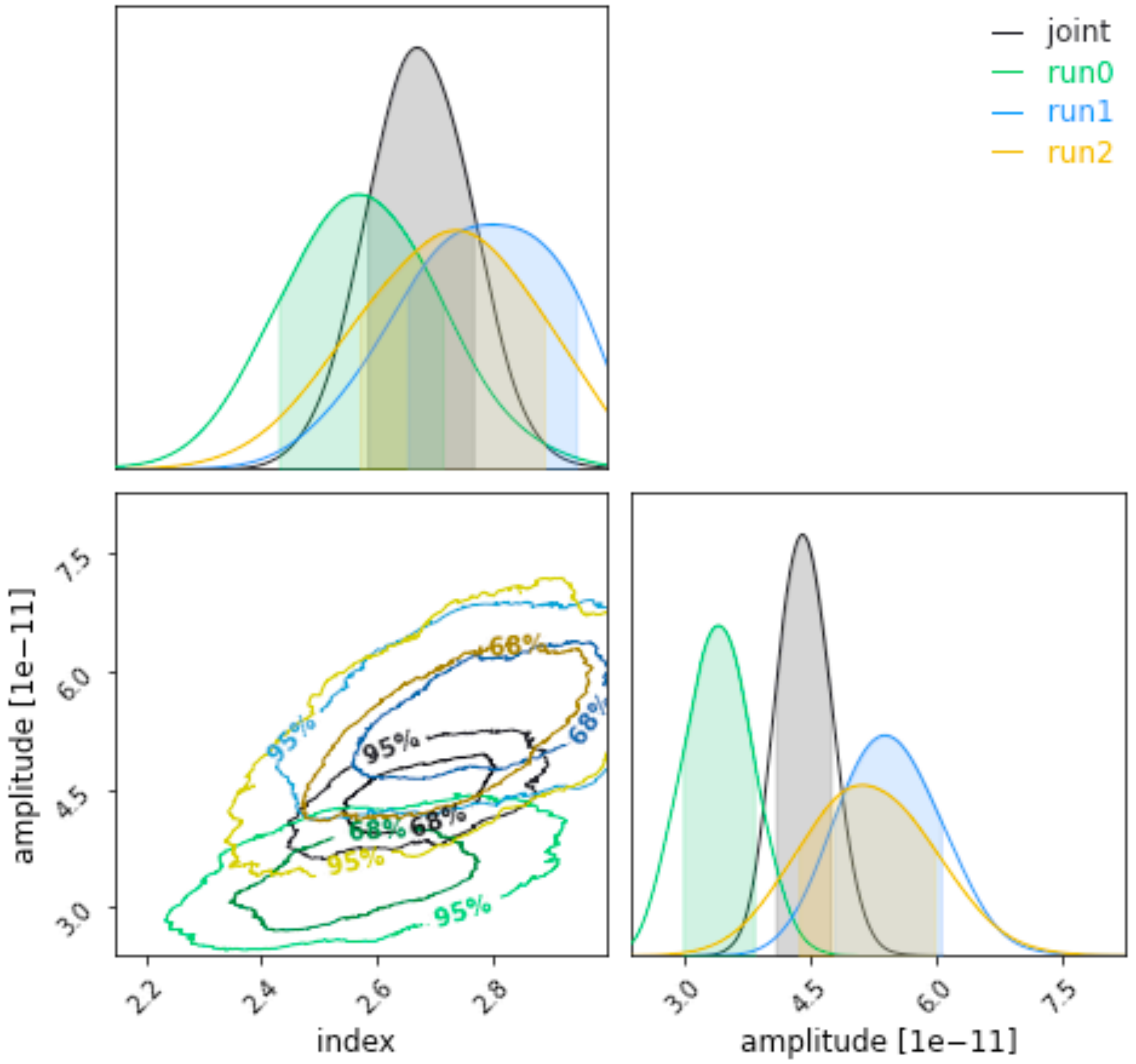


Speagle, 2020

From simple to complex analyses

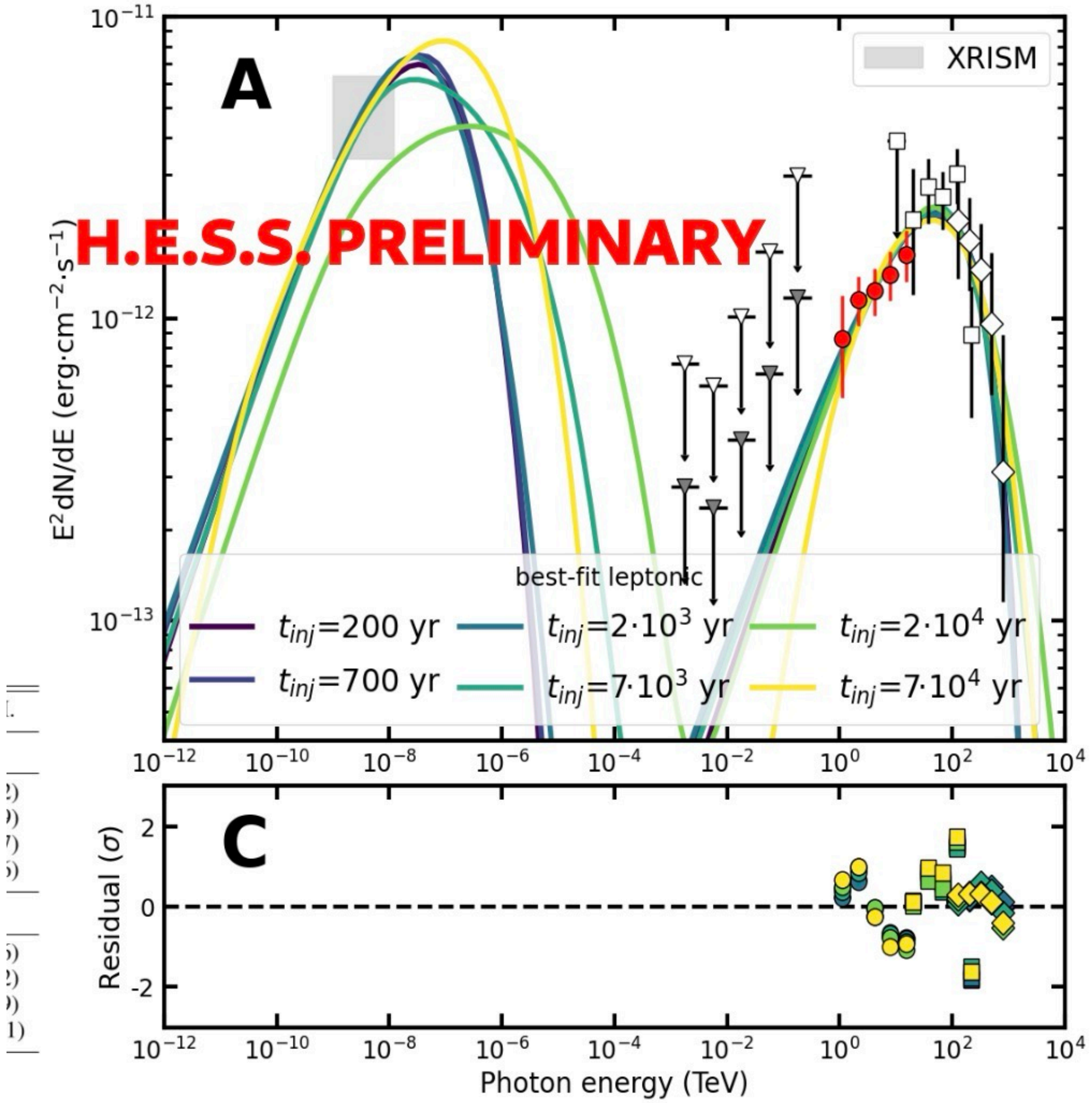
Crab spectral analysis



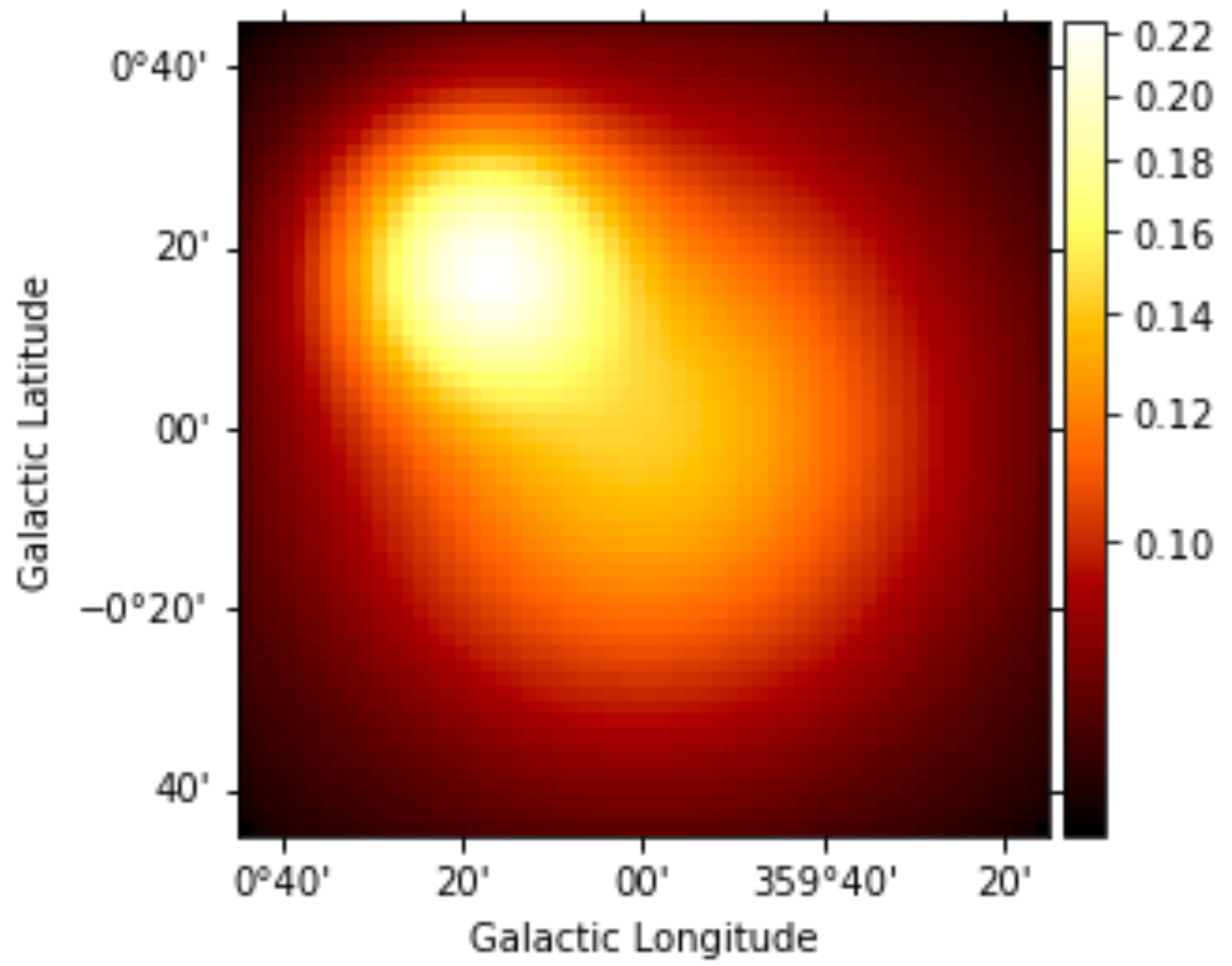


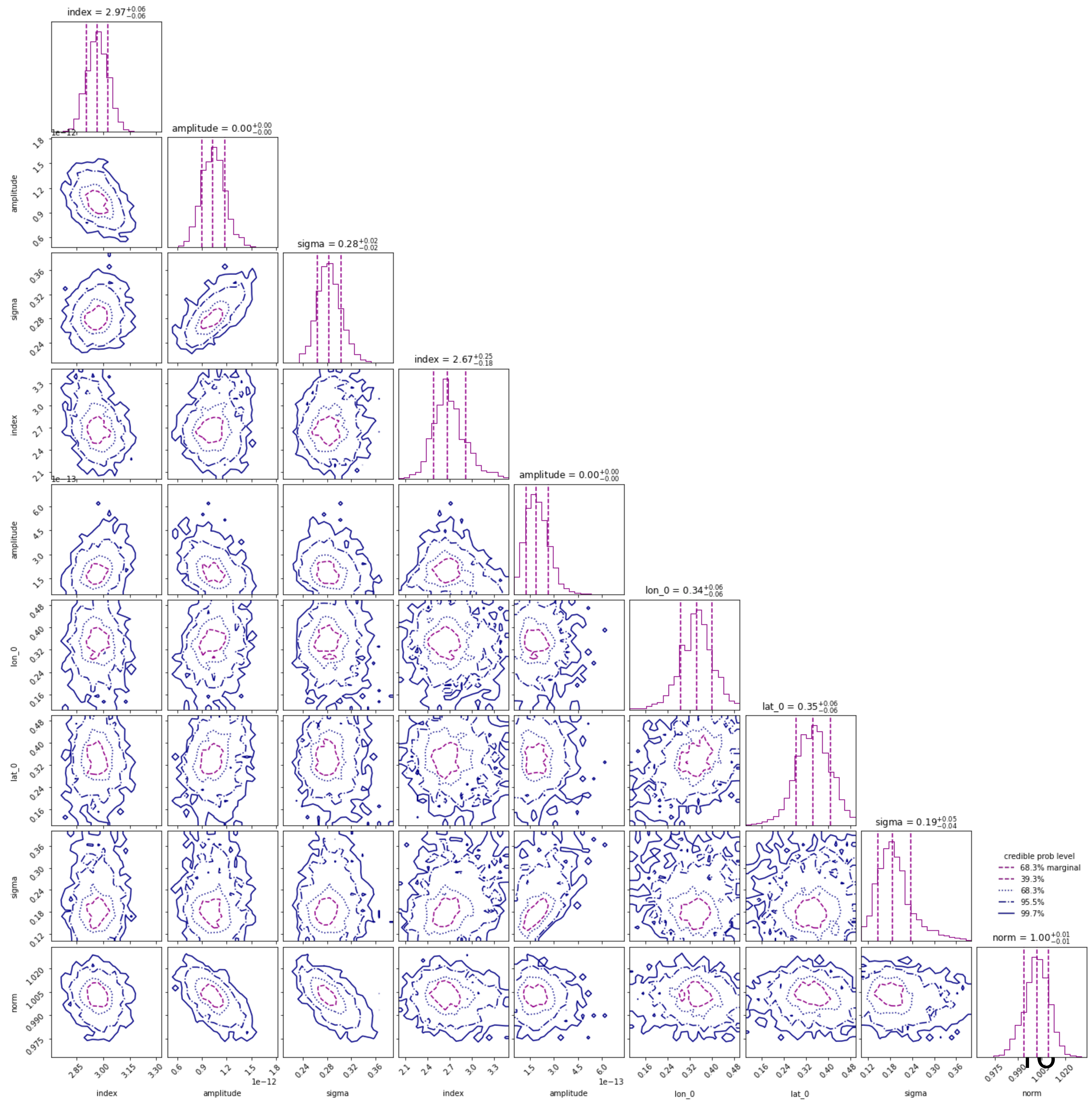
V4641 Sgr - Laura Olivera-Nieto - ICRC '25

Fit using GAMERA and bayesian sampling implemented in Gammapy 1.3

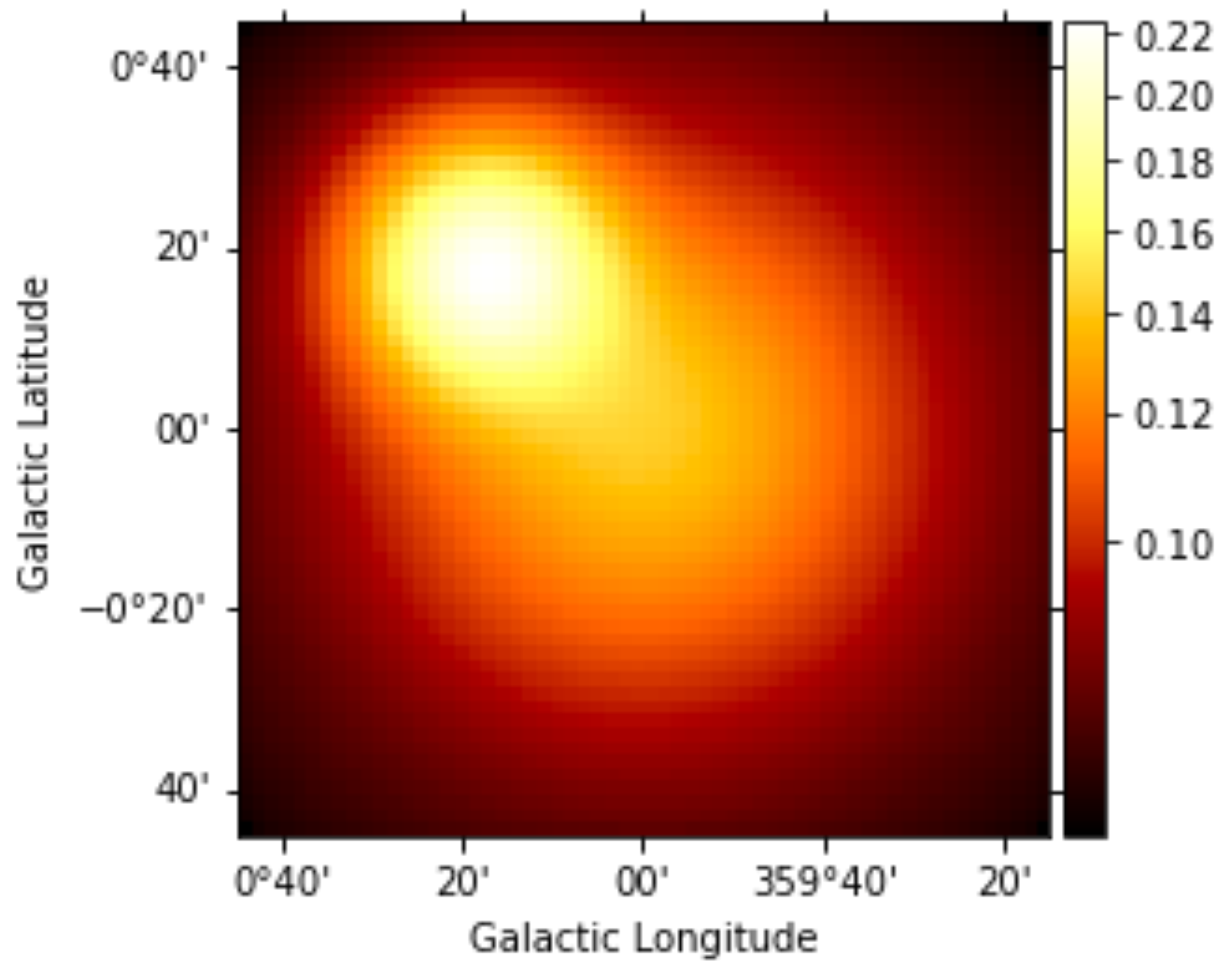


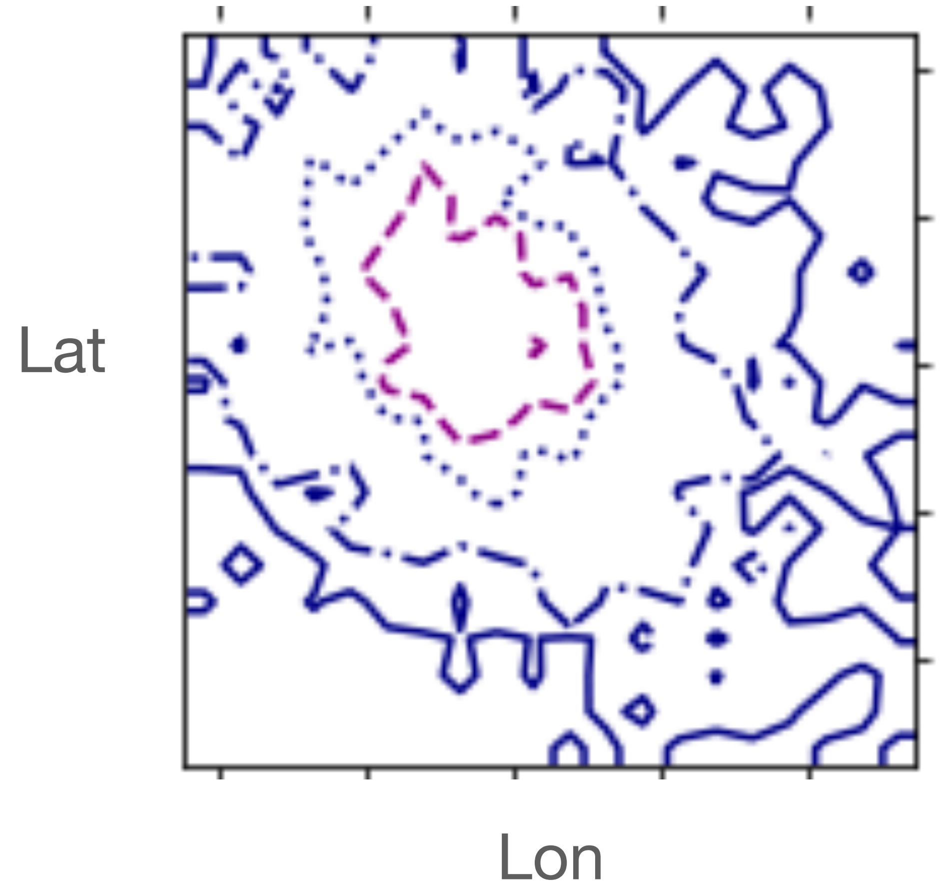
Degeneracies in complex regions

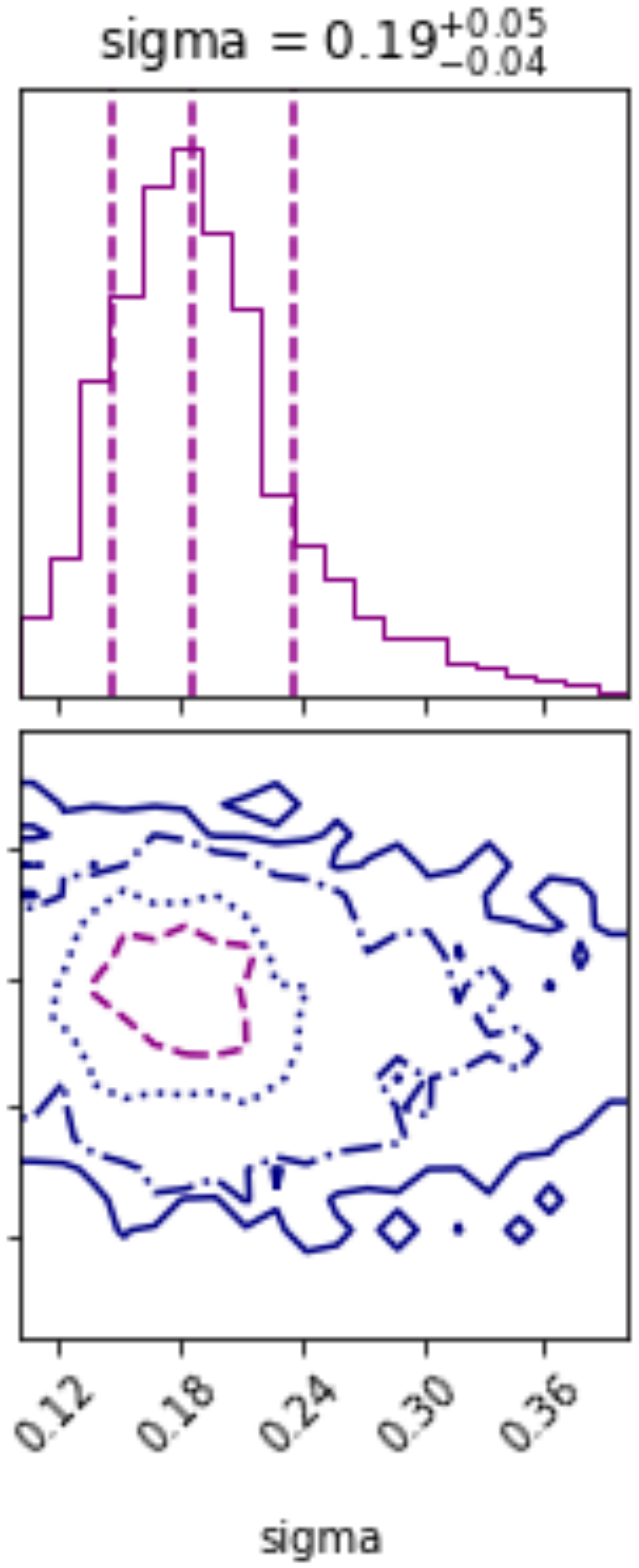




Degeneracies in complex regions







Conclusion

- **Nested Sampling Philosophy : Outside-Inside method, 500 independent live points init all over the parameter space (unlike 10-30 MCMC walkers init in small ball)**

- **Nested sample allows to lift several issues of MCMC techniques:**
  - **No need for init point. No burn-in period. No issue with multi-modal distribution.**
  - **NS has a well defined stopping criteria unlike MCMC**

- **In addition to estimating posterior distributions, calculate the Bayesian evidence Z:**
  - **This enables to compare different models that are not nested by comparing ΔZ between models**
  - **This is not possible with MCMC as only a local part of the likelihood is sampled**

- **More robust than traditional gradient descent fitting. Little human intervention needed to debug "failed fit".**
- **CONS: Need to define priors & slower but can be // easily but you get much more information than a simple fit.**

Existing Nested sampling frameworks

Ashton, 2022

Code	Methods	Dynamic	Languages	Field	Pub. Year
CosmoNest [60, 61]	ellipsoid	fixed	Fortran	Cosmology	2006
MultiNest [48, 84]	multi-ellipsoid	fixed	Fortran, C/C++, Python	Cosmology	2008
DIAMONDS [249]	multi-ellipsoid	fixed	C++	Astrophysics	2015
nestle [250]	ellipsoid, multi-ellipsoid	fixed	Python	Astrophysics	2015
nessai [90, 91]	normalising flow ellipsoid	fixed	Python	Gravitational waves	2021
(dy)PolyChord [53, 65]	slice	dynamic	Fortran, C/C++, Python	Cosmology	2015
LALInferenceNest [180]	random walk, ensemble, differential evolution	fixed	C	Gravitational waves	2015
Nested_fit [104, 257, 258]	random walk	fixed	Fortran	Atomic physics	2016
cpnest [259]	slice, differential evolution, Gauss, Hamiltonian, ensemble	fixed	Python	Gravitational waves	2017
pymatnest [44]	random walk, Galilean, symplectic Hamiltonian	fixed	Python	Materials	2017
NNest [261]	normalising flow random walk	fixed	Python	Cosmology	2019
DNest5 [55]	user-defined, random walk	diffusive	C++	Astrophysics	2020
BayesicFitting [263]	random walk, slice, Galilean, Gibbs	fixed	Python	Astronomy	2021
dynesty [52]	ellipsoid, multi-ellipsoid, MLFriends & Gauss, slice, Hamiltonian	dynamic	Python	Astrophysics	2020
UltraNest [92]	MLFriends + ellipsoid & Gauss, hit-and-run, slice	reactive	Python, Julia, R, C/C++, Fortran	Astrophysics	2020
jaxns [266]	multi-ellipsoid & slice	fixed	jax	Astronomy	2021