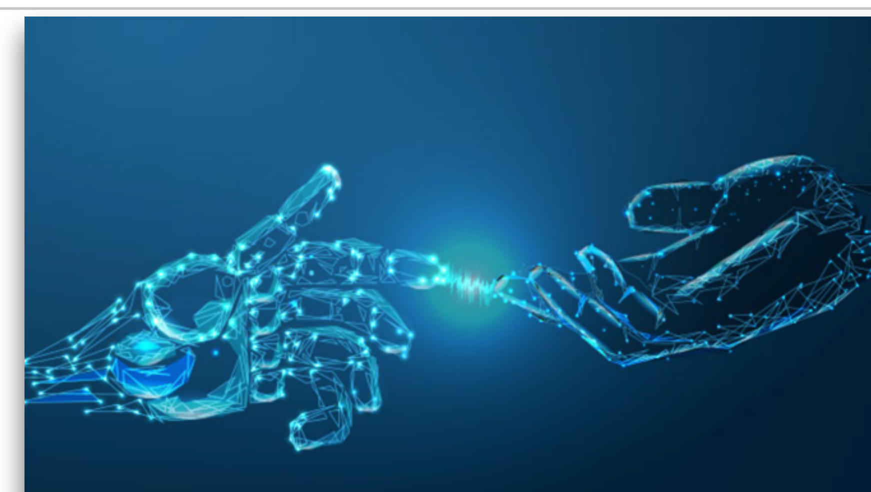
EUROPEAN ARC
ALMA Regional Centre

Bayesian Reconstruction through Adaptive Image Notion

Fabrizia Guglielmetti (ESO)



ALMA Development Workshop (ESO) - June 5, 2019

Proposal primary goals

- Enhancing CASA/tclean

  1. to **improve** emission detection for ALMA radio synthesis imaging,
  2. to **speed up** convergence procedures.

- Bayesian probability theory (BPT) is used to incorporate acquired knowledge and let the algorithm learn from the data to tackle image analysis: **BRAIN**

$$P(H_s|\mathbf{d}, I) = \frac{P(\mathbf{d}|H_s, I)P(H_s|I)}{P(\mathbf{d}|I)}$$

BRAIN phases

- Several stages are foreseen with each a prototype software

  1. automatic mask emission algorithm in tclean *(while employing one of the available deconvolution algorithms)*
  2. new standalone deconvolution algorithm *(in the minor cycle)*
  3. improvement in image analysis for mosaic of images (both in 1 and 2)
  4. improvement of source detection also in extreme sparse data
  5. introduce 3D modeling of diffuse emissions
  6. unified algorithm for the detection of both ALMA data and counting experiments
  7. introduce calibration and imaging processes in one unique algorithm

Beyond BRAIN phases

- Support CASA with a MEM reloaded task *(e.g. Strong, A., 1995, ExA, 6, 97) -see appendix*

- Improve speed in *tclean* procedure via new optimization schemes *(e.g. employing Bayesian surrogate models) -see appendix*

BRAIN in a nutshell

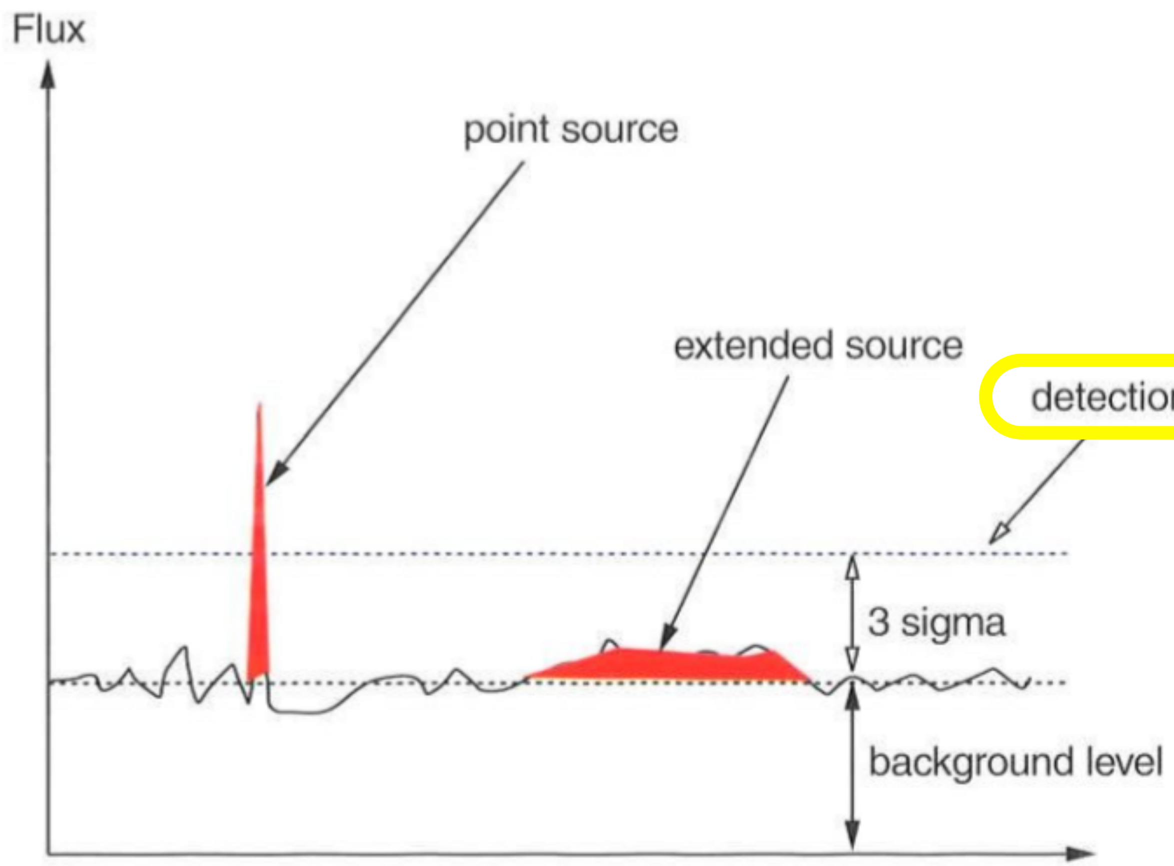


Fig. 4.1. Example of astronomical data: a point source and an extended source are shown, with noise and background. The extended object, which can be detected by eye, is undetected by a standard detection approach.

From "Astronomical Image and Data Analysis" Starck, J-L and Murtagh , F. (2006)

BRAIN in a nutshell

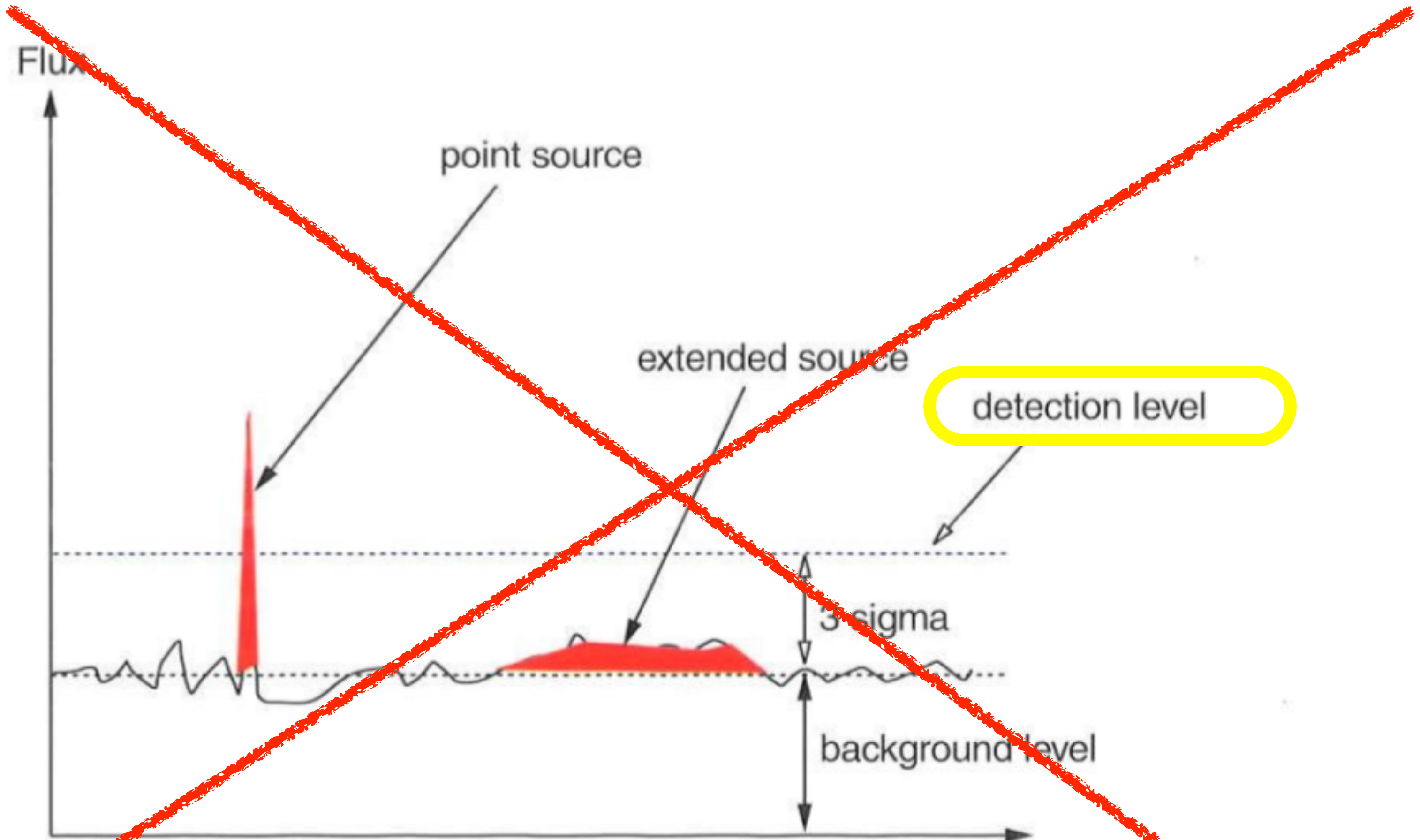


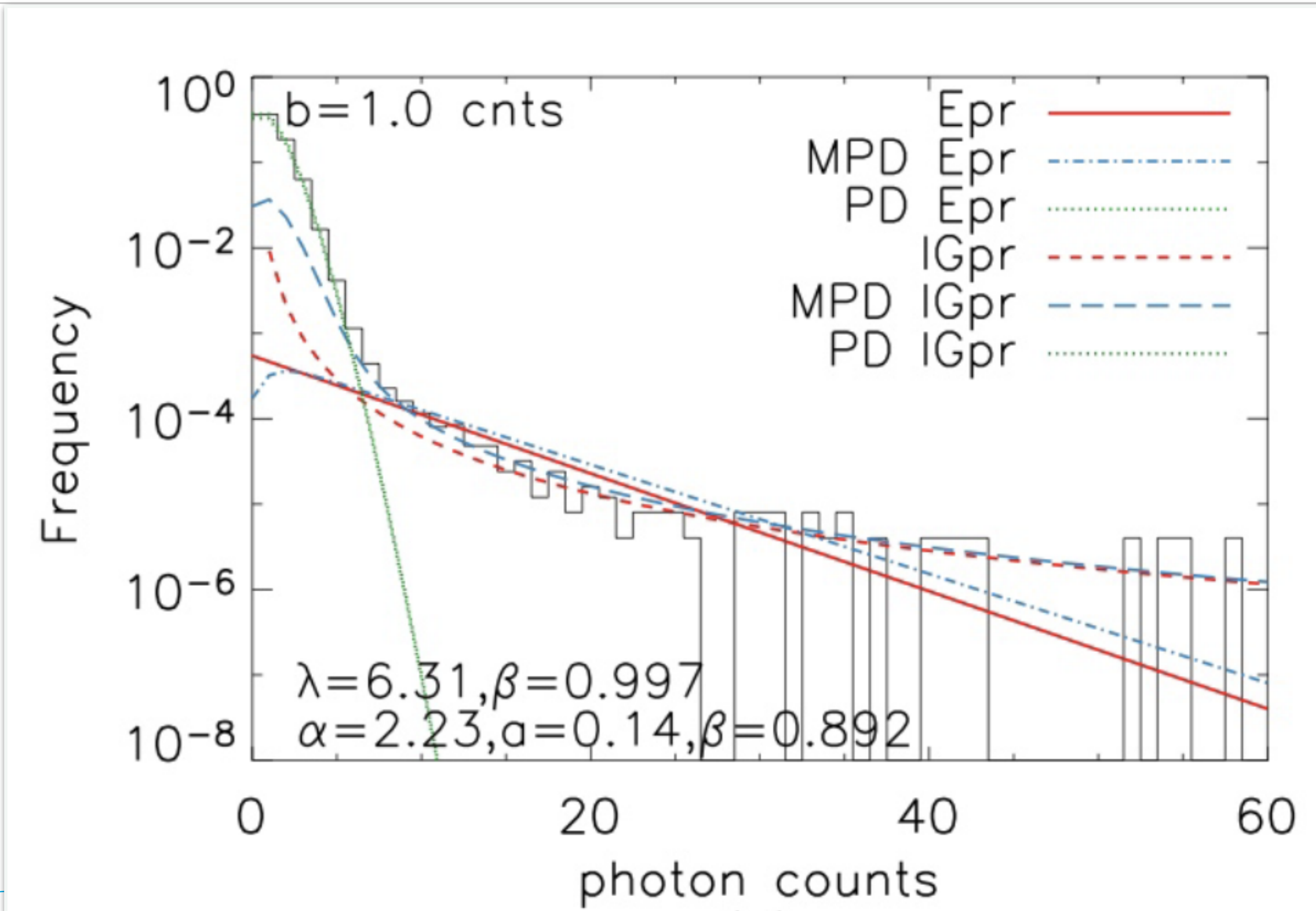
Fig. 4.1. Example of astronomical data: a point source and an extended source are shown, with noise and background. The extended object, which can be detected by eye, is undetected by a standard detection approach.

From “Astronomical Image and Data Analysis” Starck, J-L and Murtagh , F. (2006)

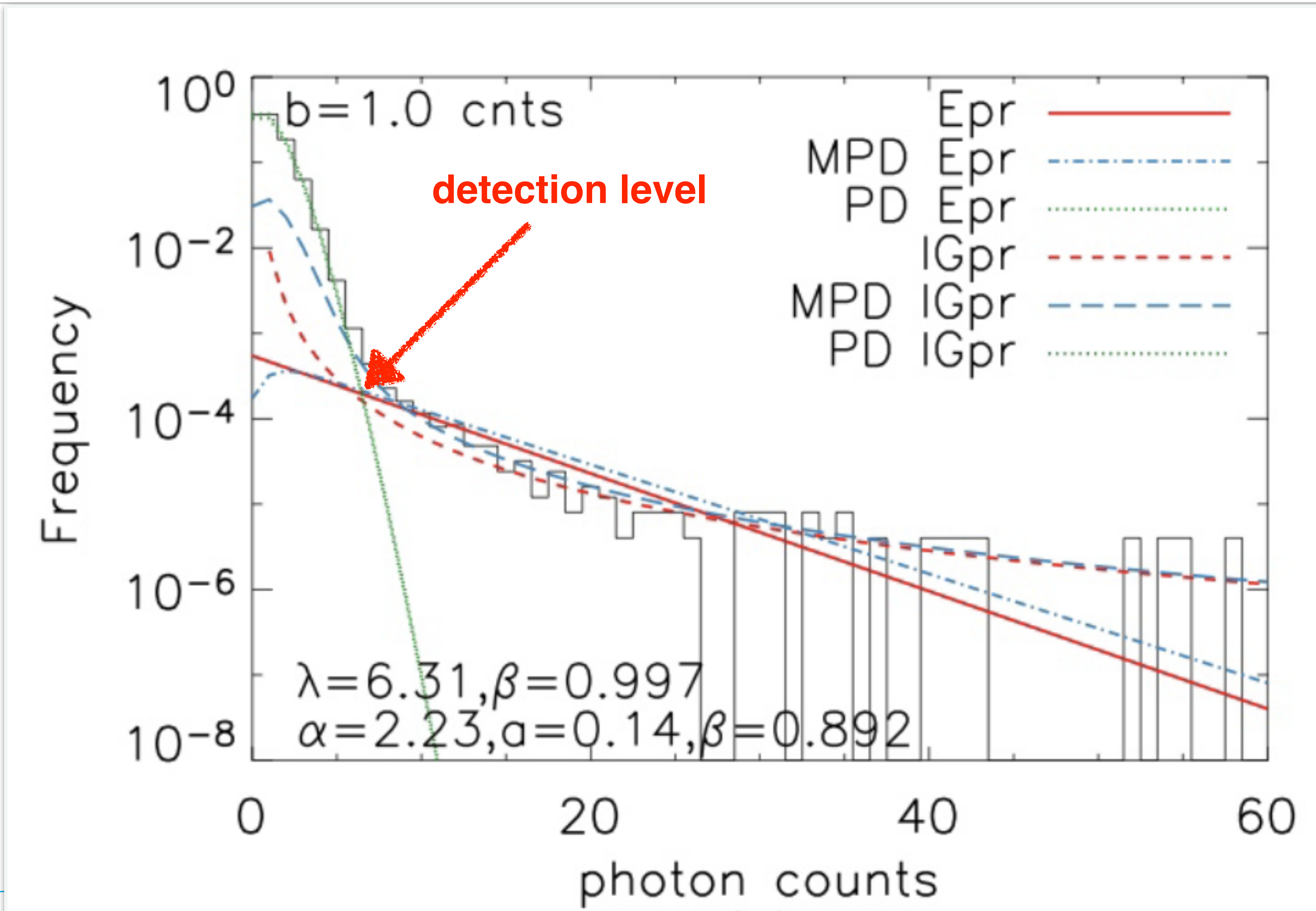
BRAIN in a nutshell

- **BPT + probabilistic mixture of models** *(Guglielmetti et al., 2019- Background-Source Separation technique)*
  - emphasis is given on the background model *(precise and accurate)*
  - automated decision for the **separation of** physically interesting sky **signals** from underlying background components (no explicit subtraction)
  - physical models and experts knowledge are incorporated in prior probability densities
  - parametric models are incorporated in the inference process
    1. parameters are estimated from the data
    2. models are compared with Bayes' factors
  - **robust uncertainty quantification** provided by the joint posterior probability densities

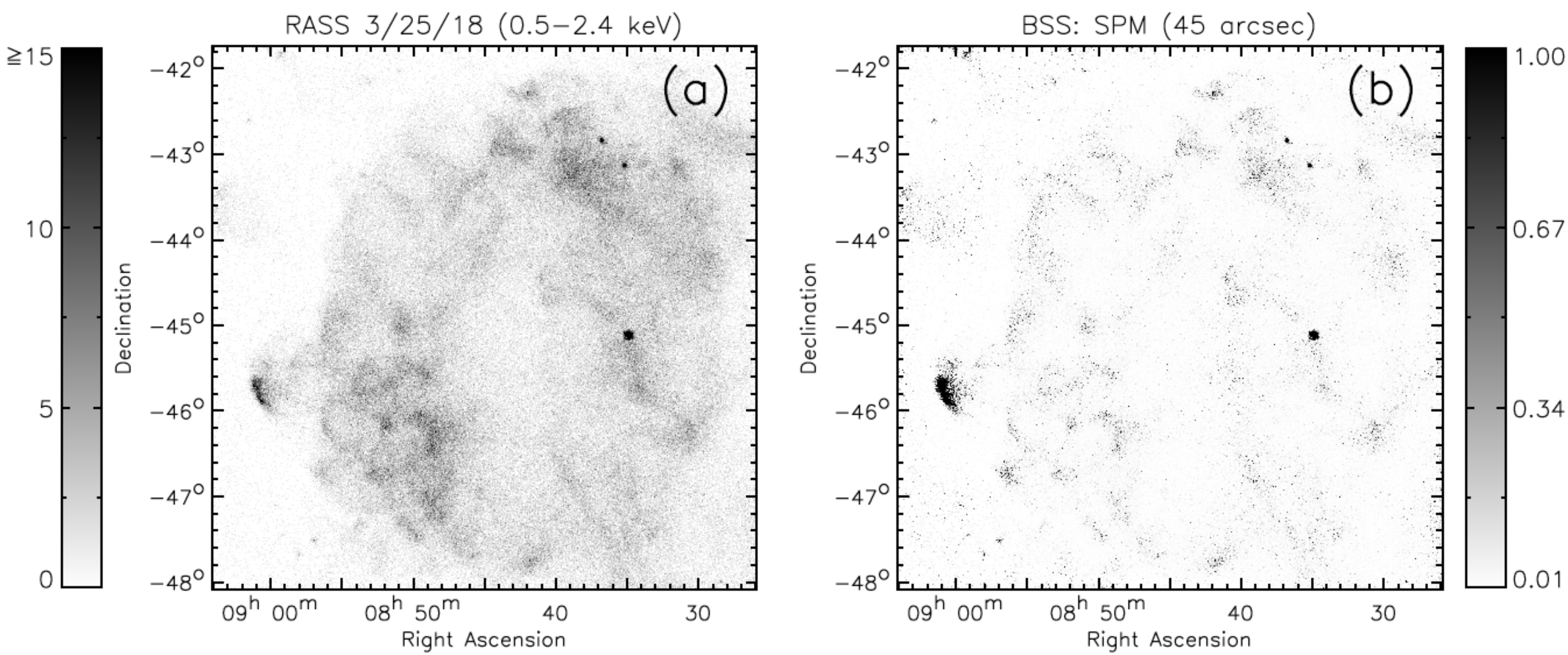
BRAIN in a nutshell



BRAIN in a nutshell



BRAIN in a nutshell

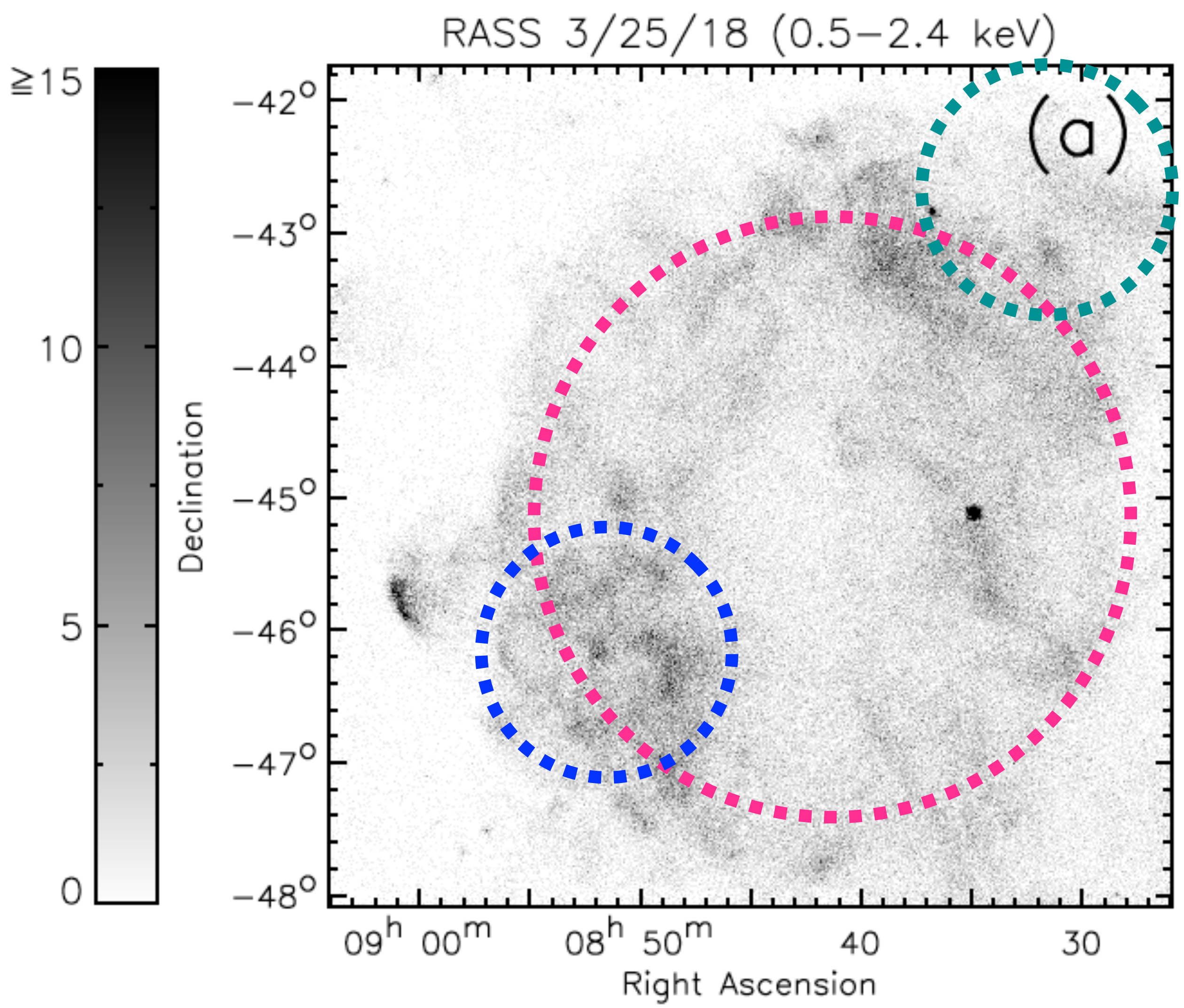


counts/pixel

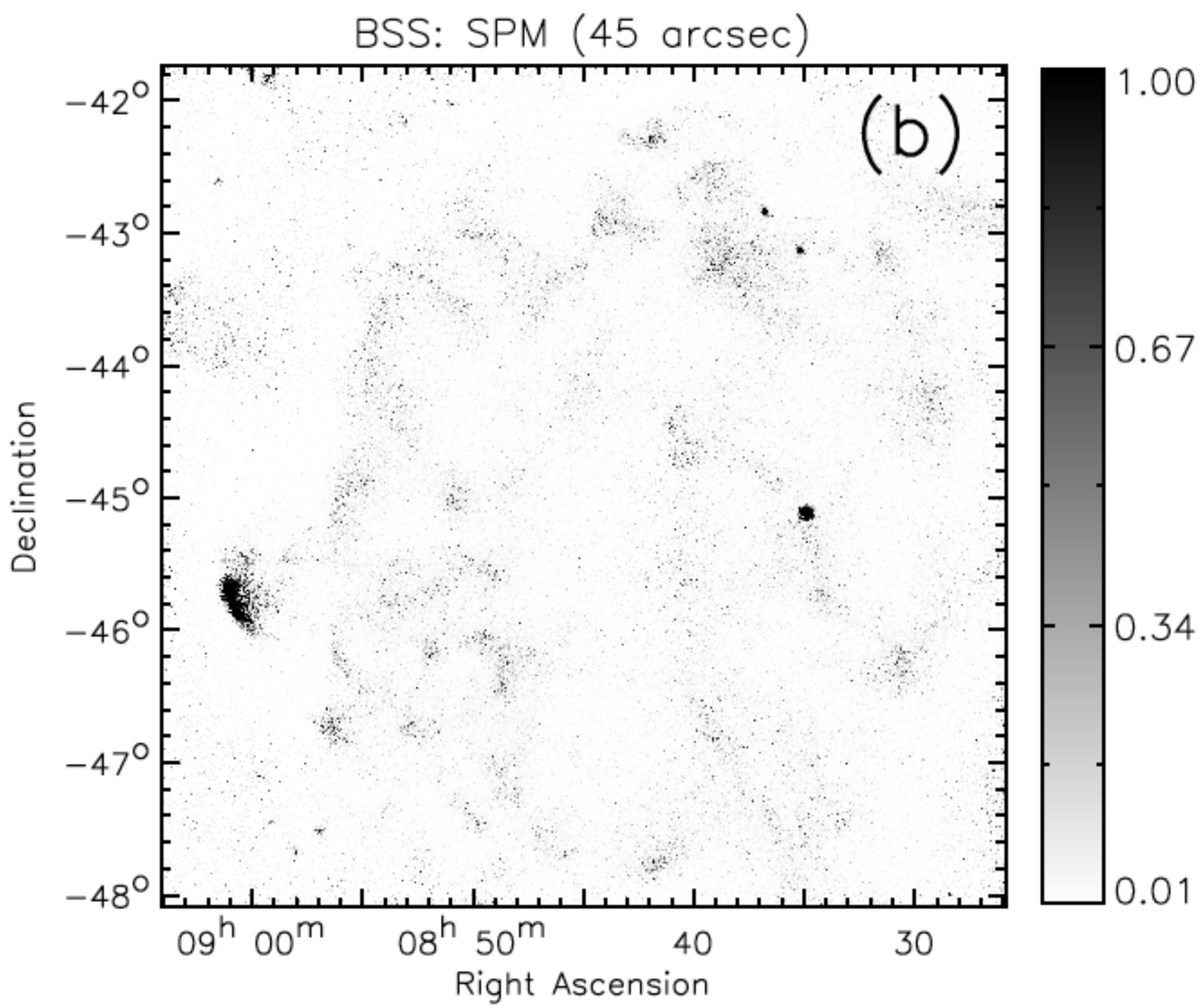
Source prob

from data space ⟷ to probability space

BRAIN in a nutshell



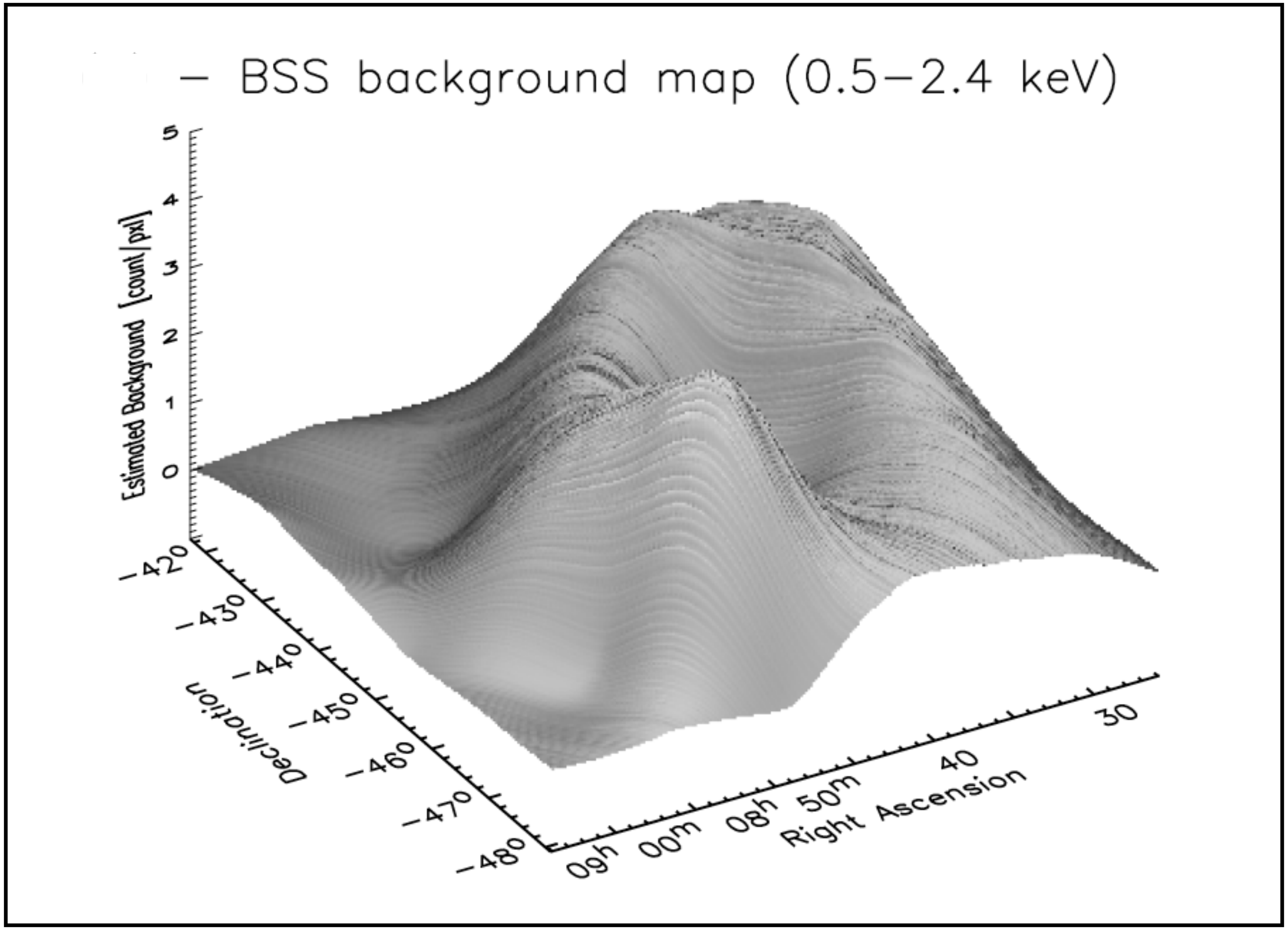
counts/pixel



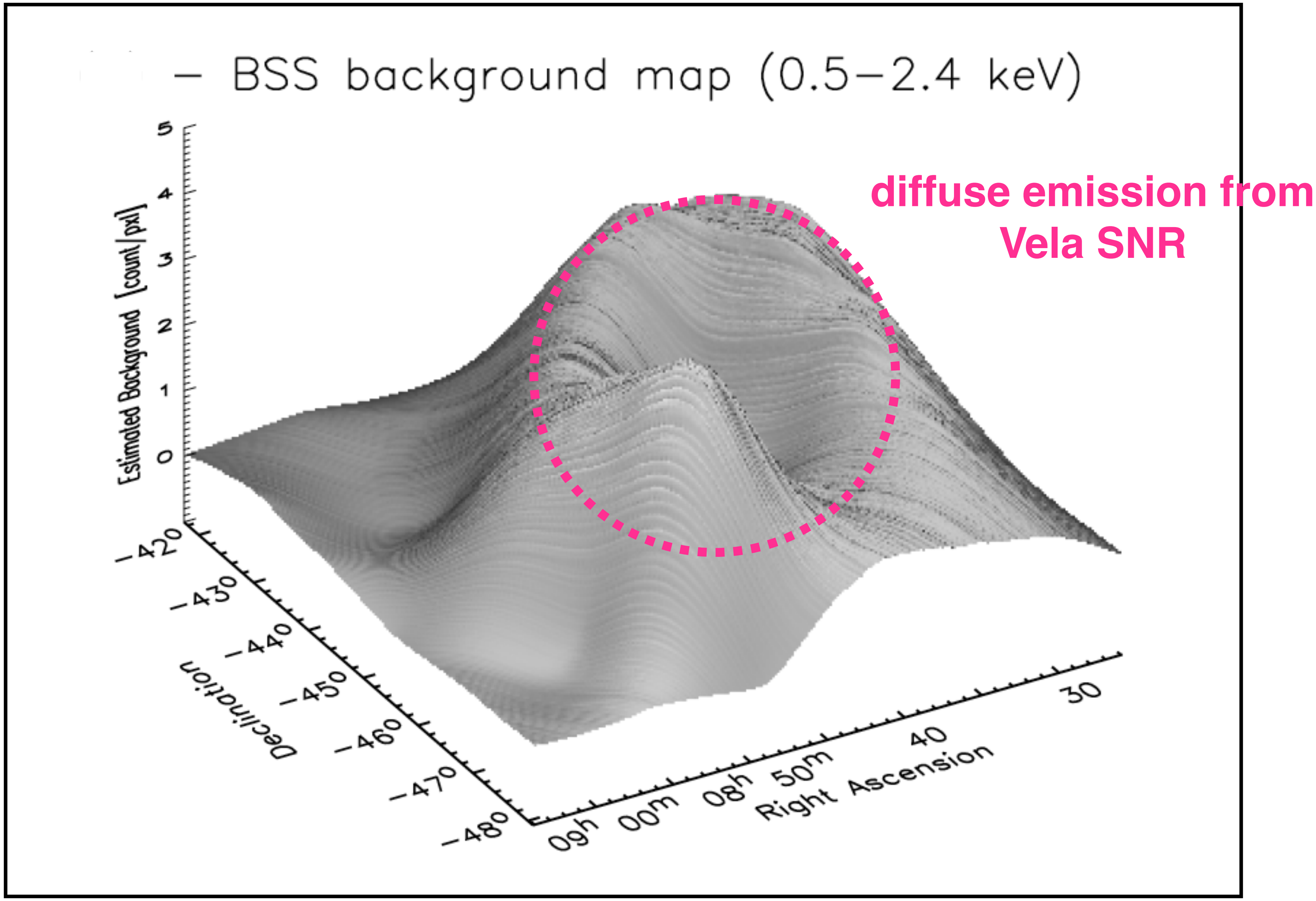
Source prob

Dataset characterised by 3 main diffuse emissions caused by 3 SNRs: Vela SNR, Vela Junior SNR, Puppis A SNR

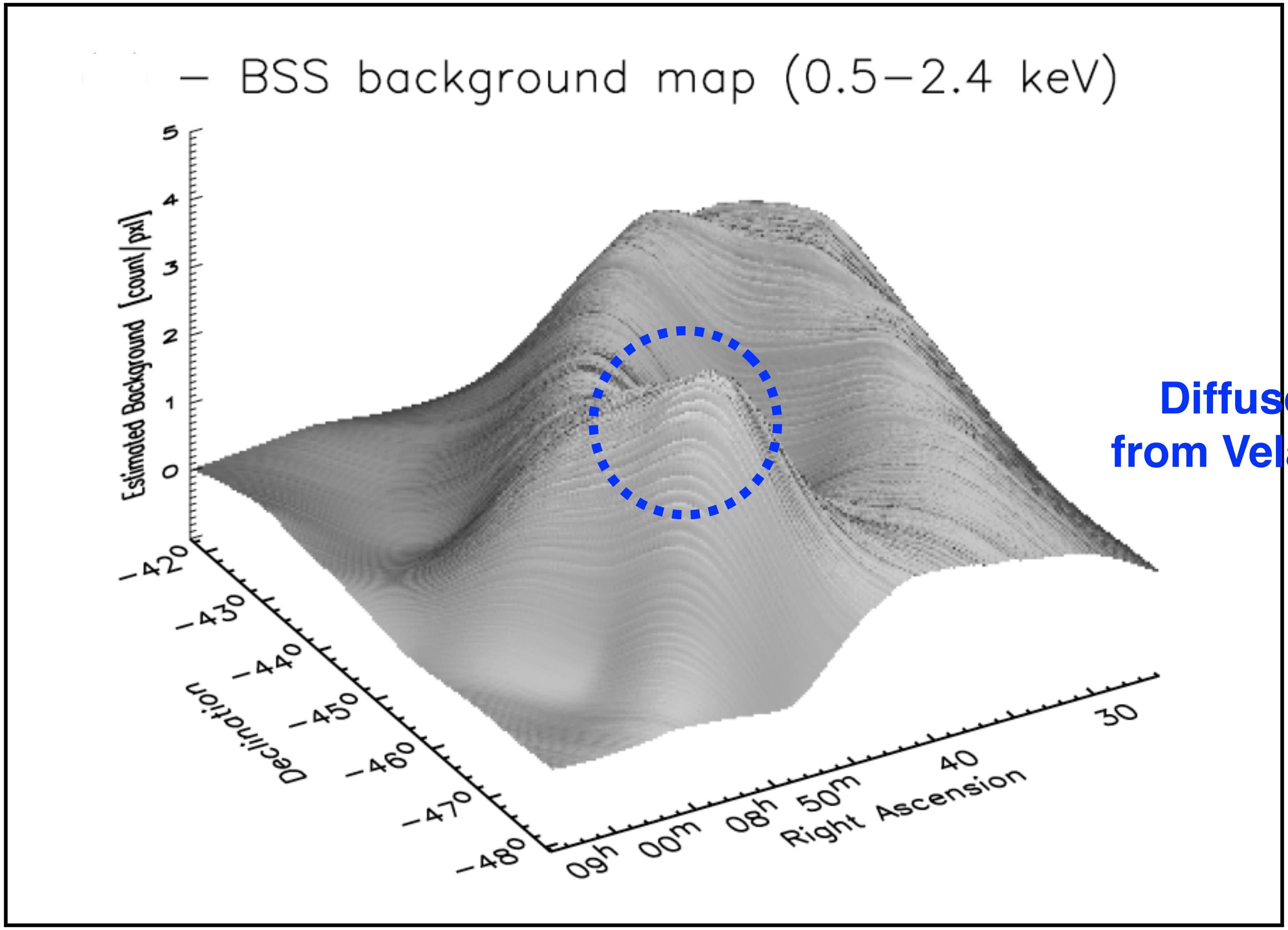
BRAIN in a nutshell



BRAIN in a nutshell

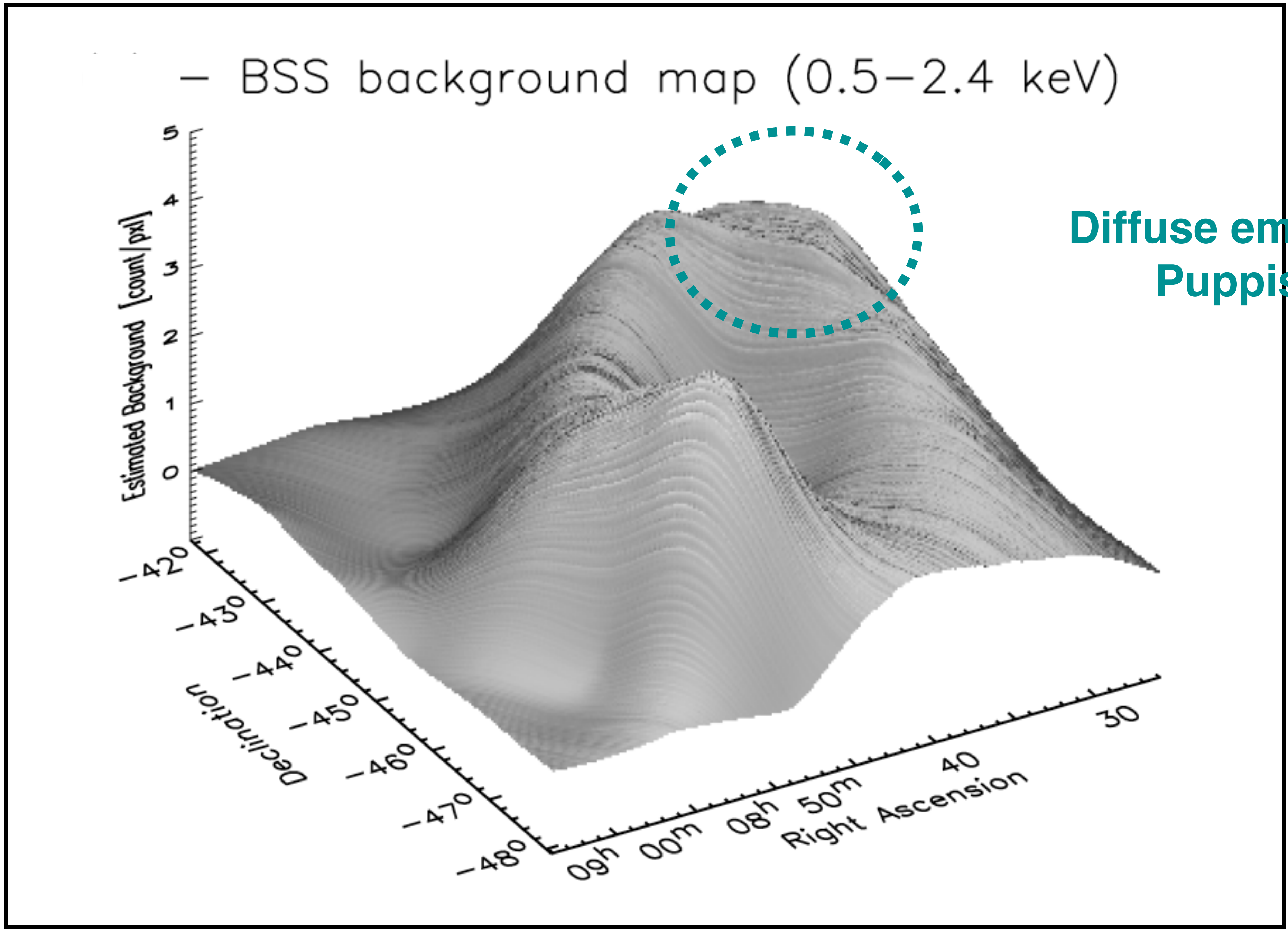


BRAIN in a nutshell



Diffuse emission from Vela Junior SNR

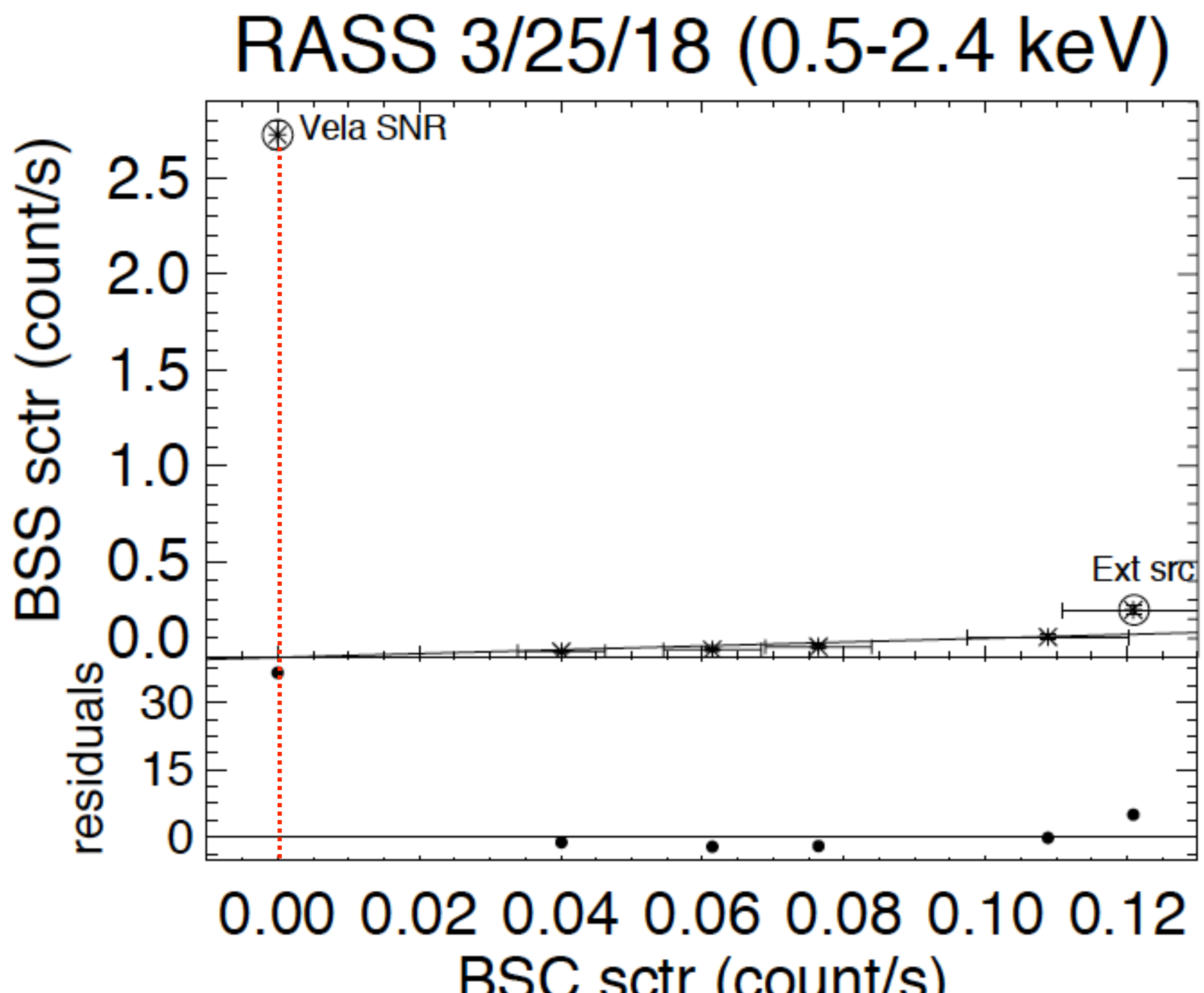
BRAIN in a nutshell



Diffuse emission from Puppis A SNR

BRAIN in a nutshell

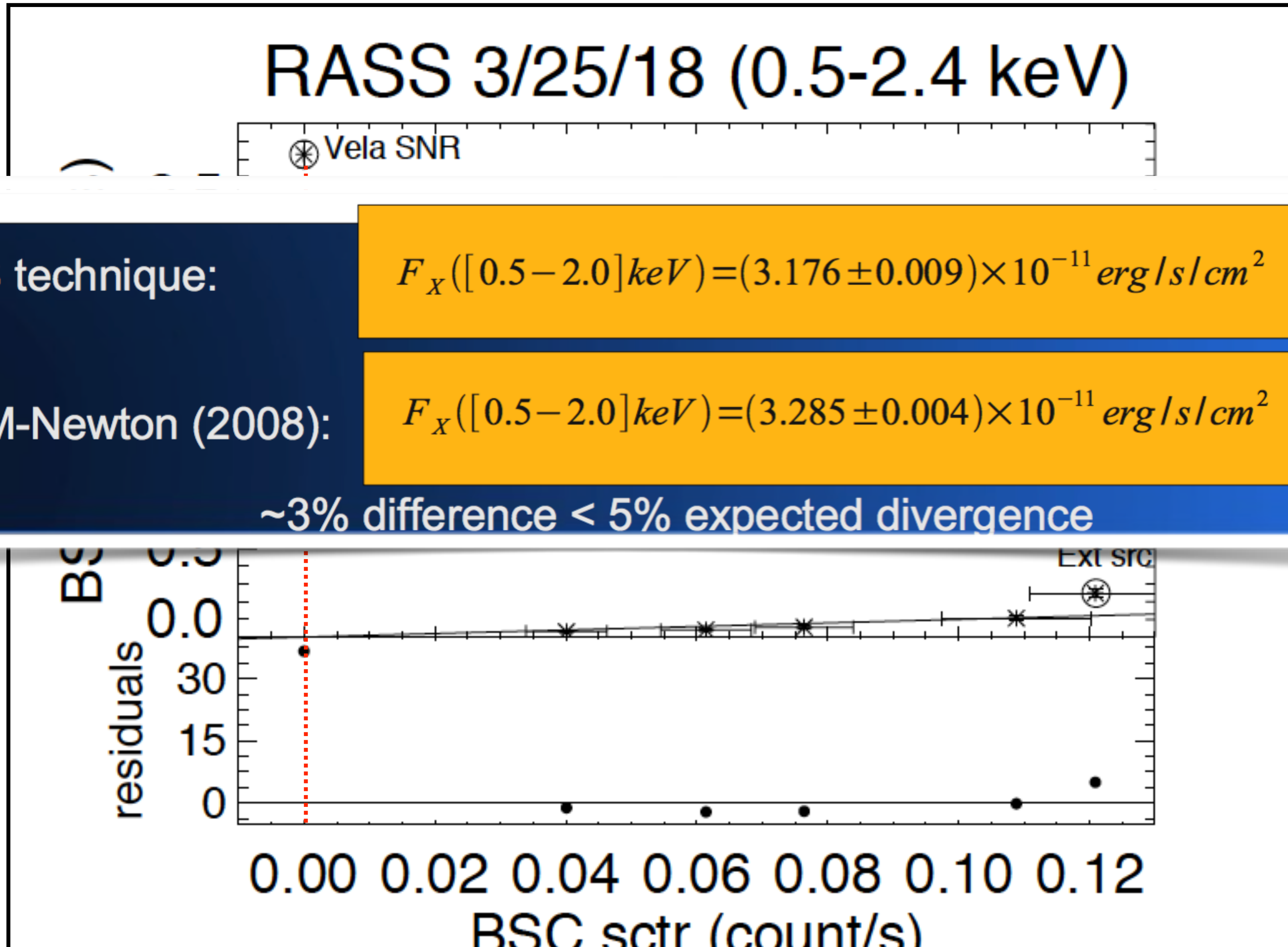
Result from background-source separation



Result from traditional technique (ML)

BRAIN in a nutshell

Result
backgr
sour
separa

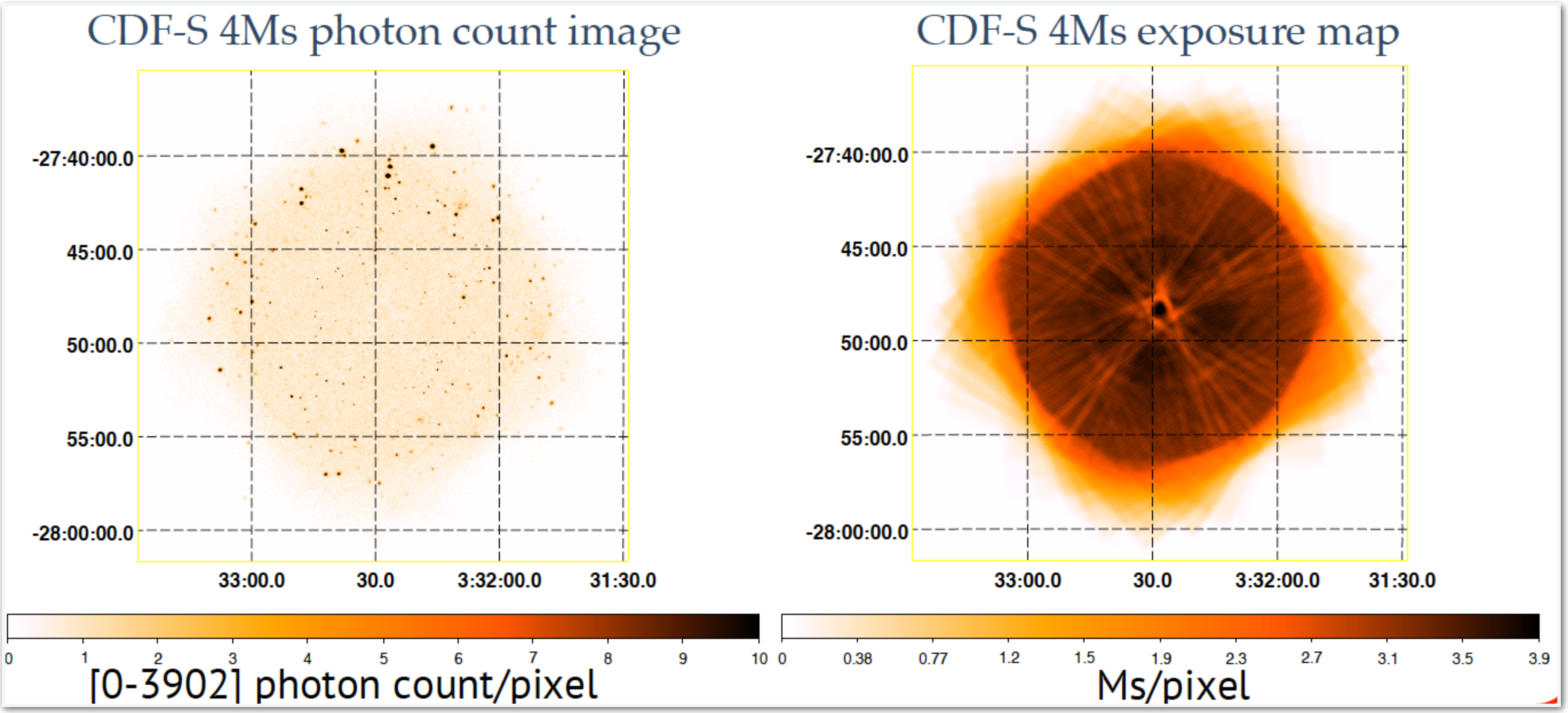


Result from traditional technique (ML)

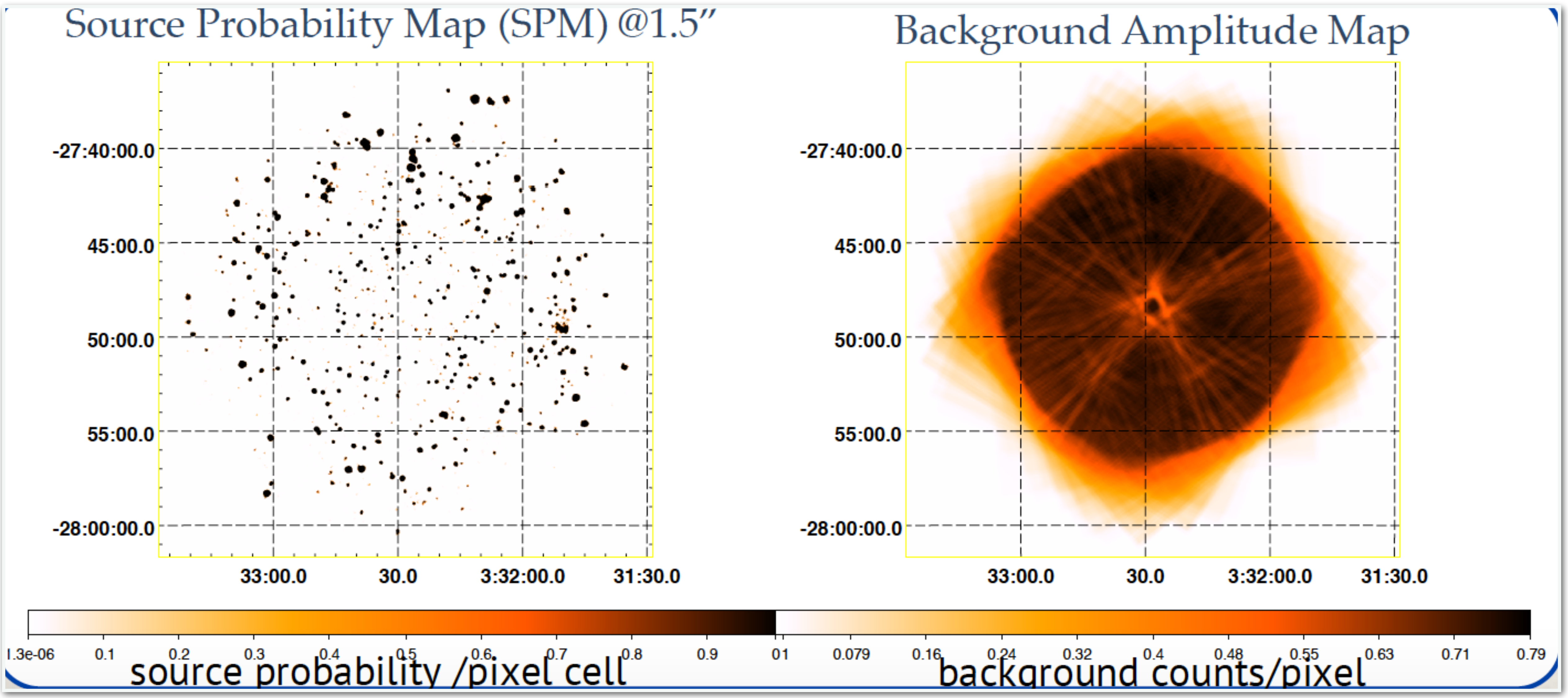
BSS technique:	$F_X([0.5-2.0]keV) = (3.176 \pm 0.009) \times 10^{-11} erg/s/cm^2$
XMM-Newton (2008):	$F_X([0.5-2.0]keV) = (3.285 \pm 0.004) \times 10^{-11} erg/s/cm^2$

~3% difference < 5% expected divergence

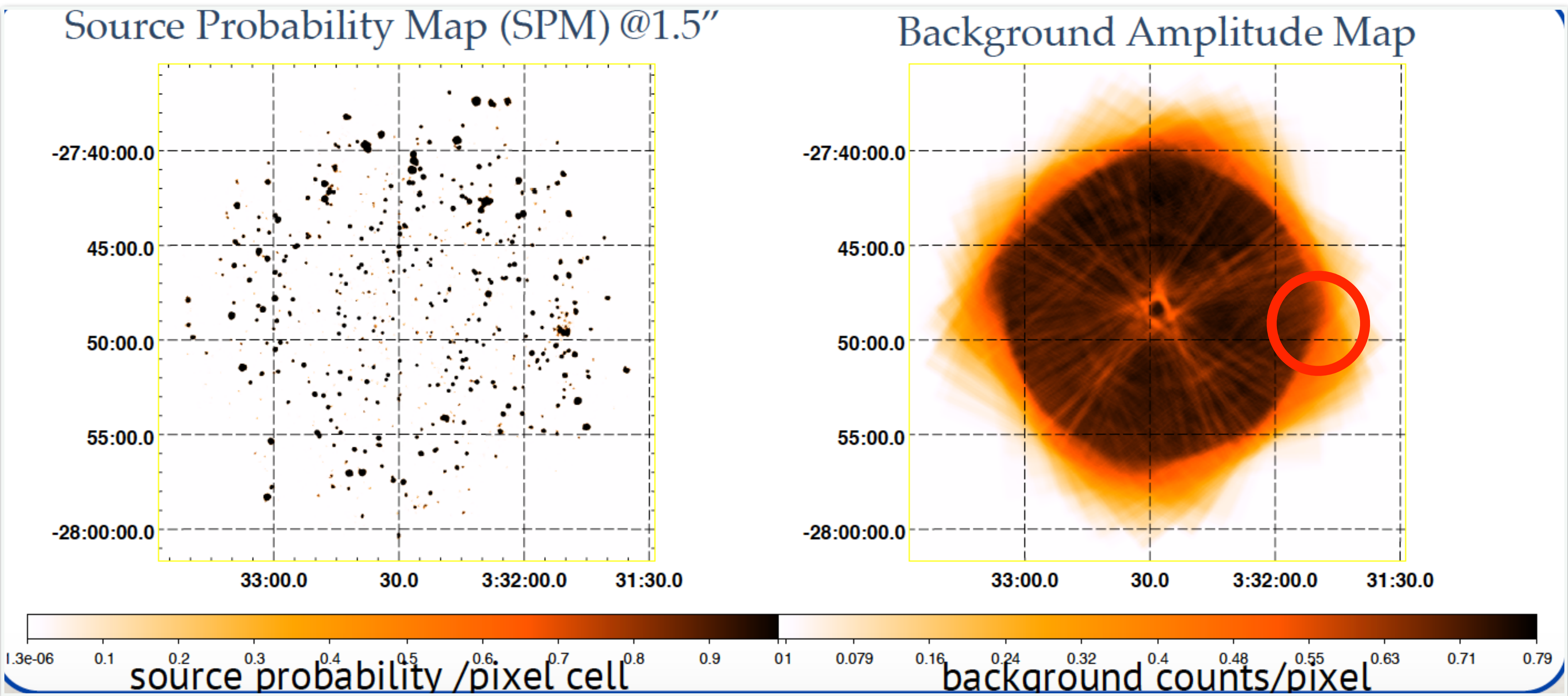
BRAIN in a nutshell



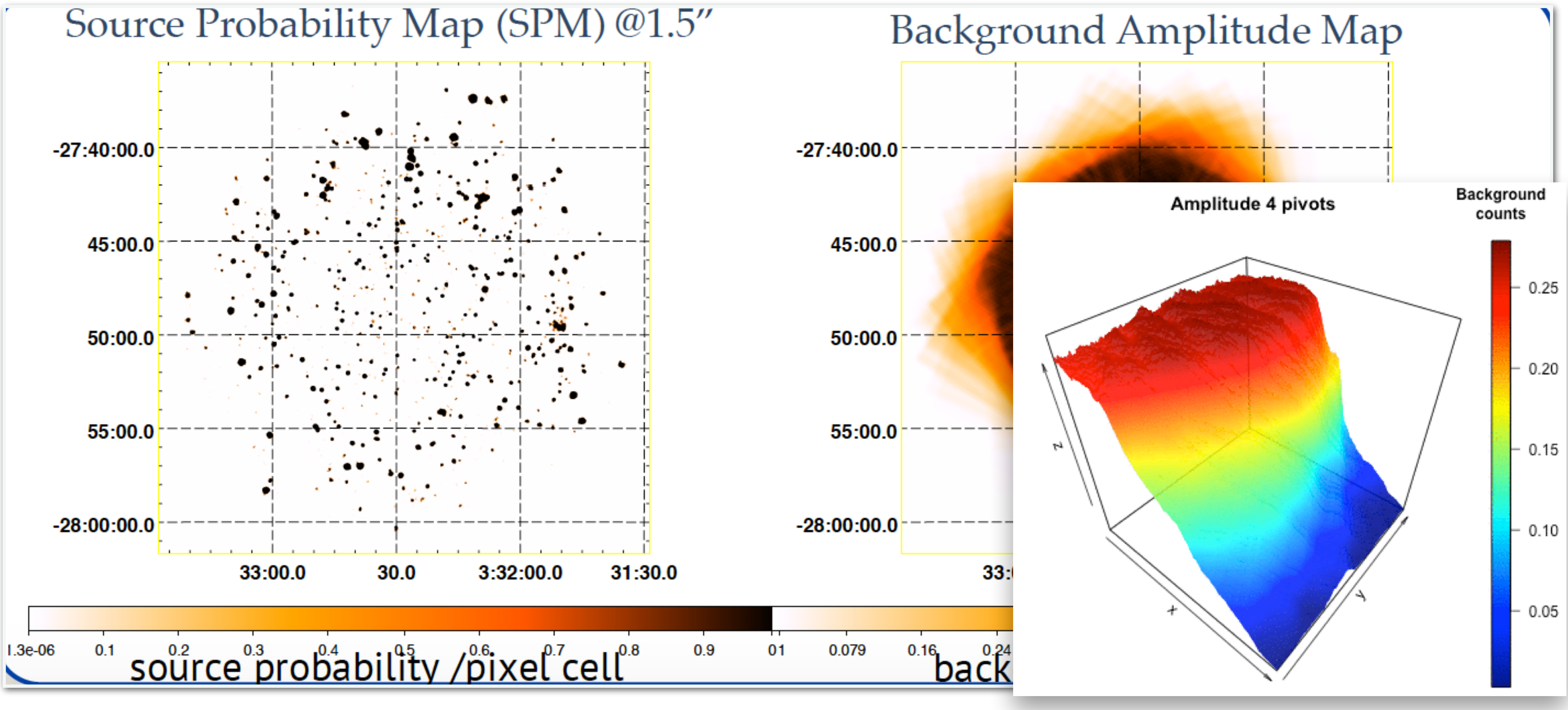
BRAIN in a nutshell



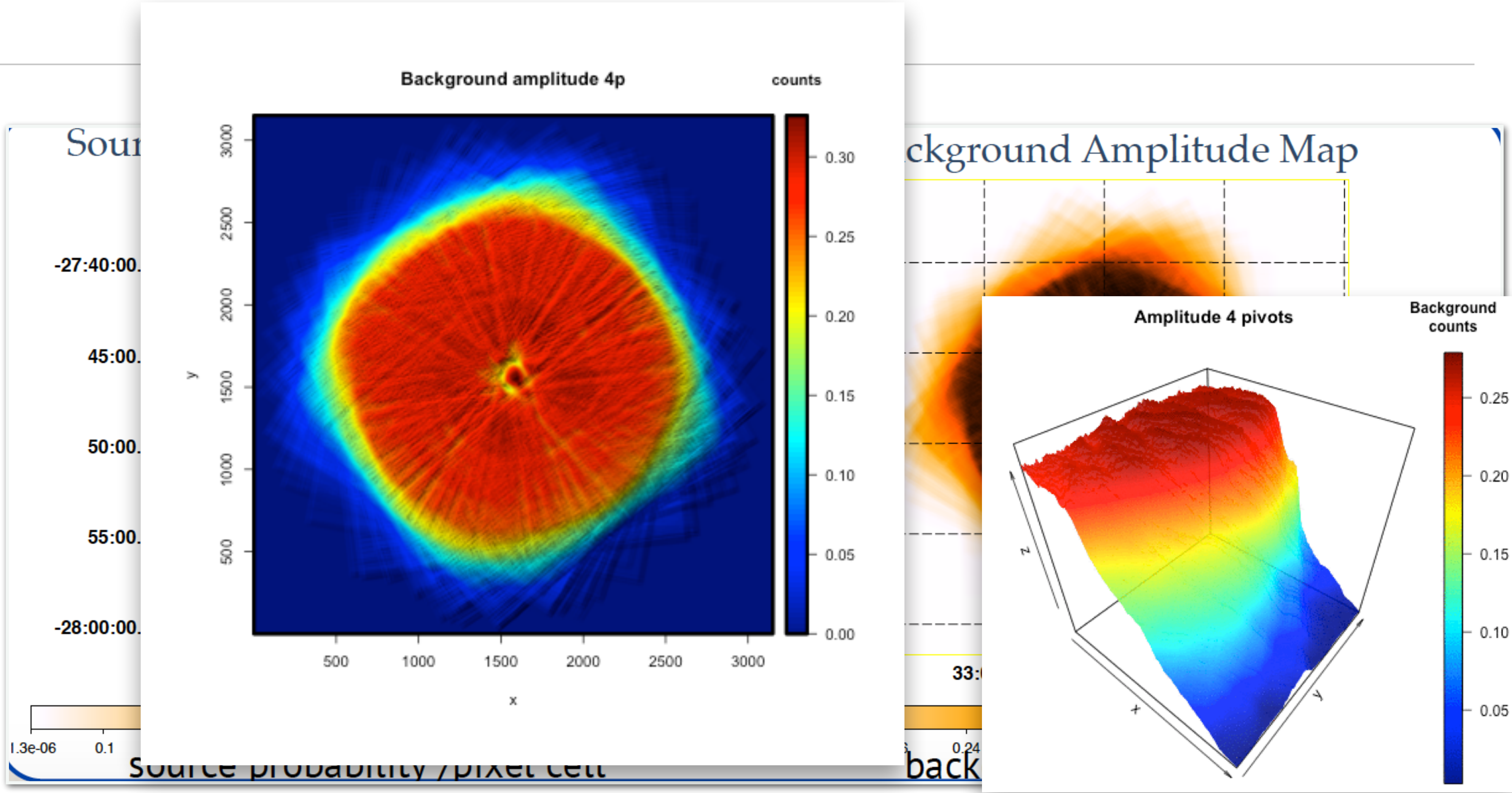
BRAIN in a nutshell



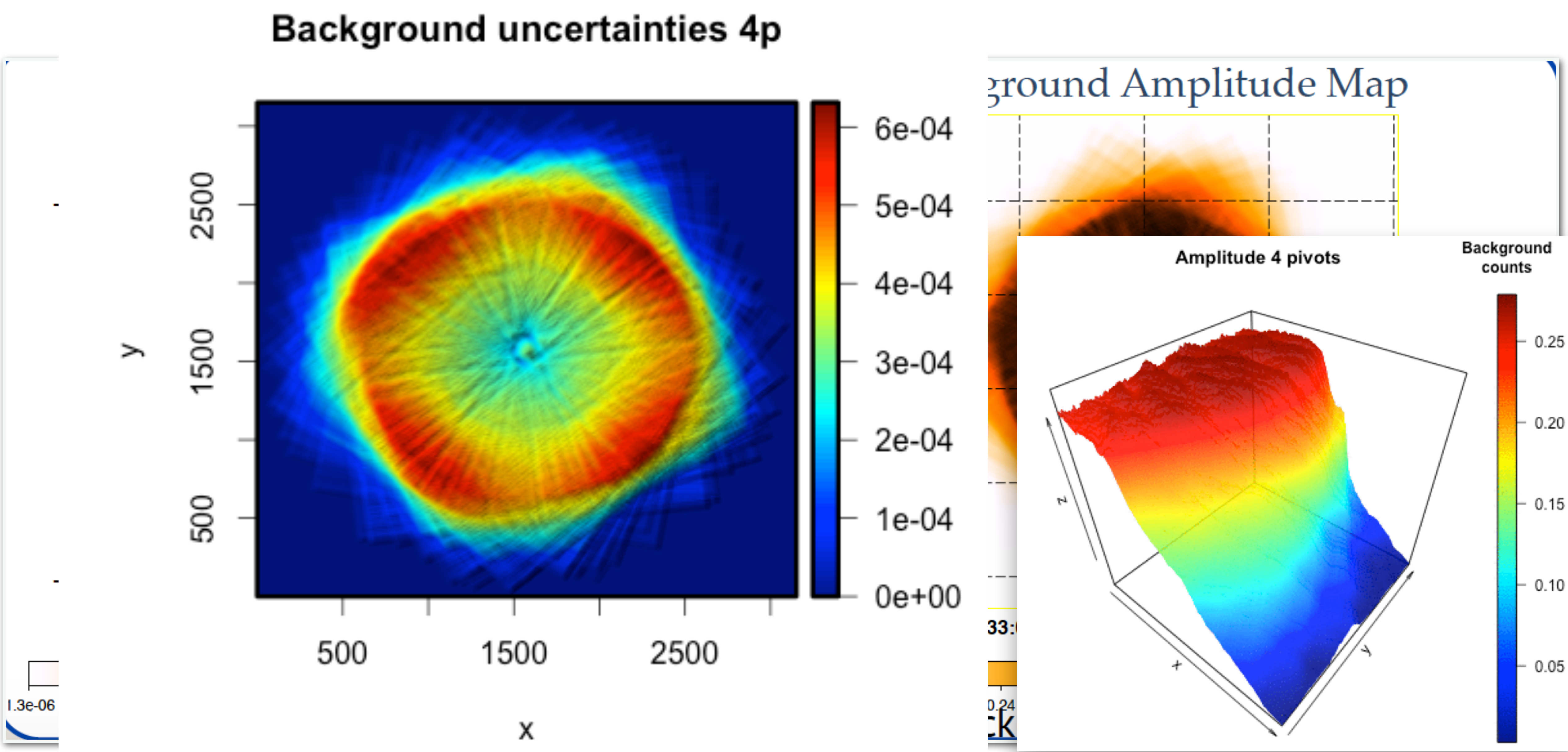
BRAIN in a nutshell



BRAIN in a nutshell

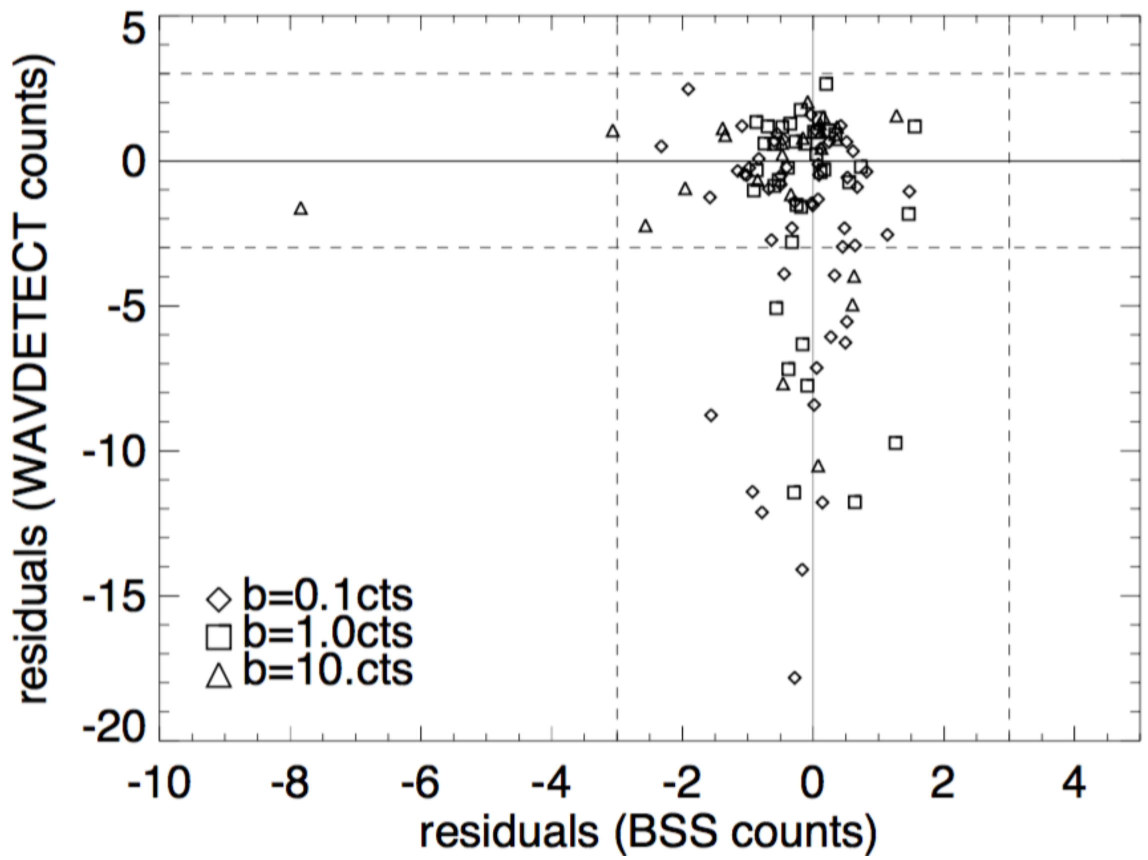


BRAIN in a nutshell



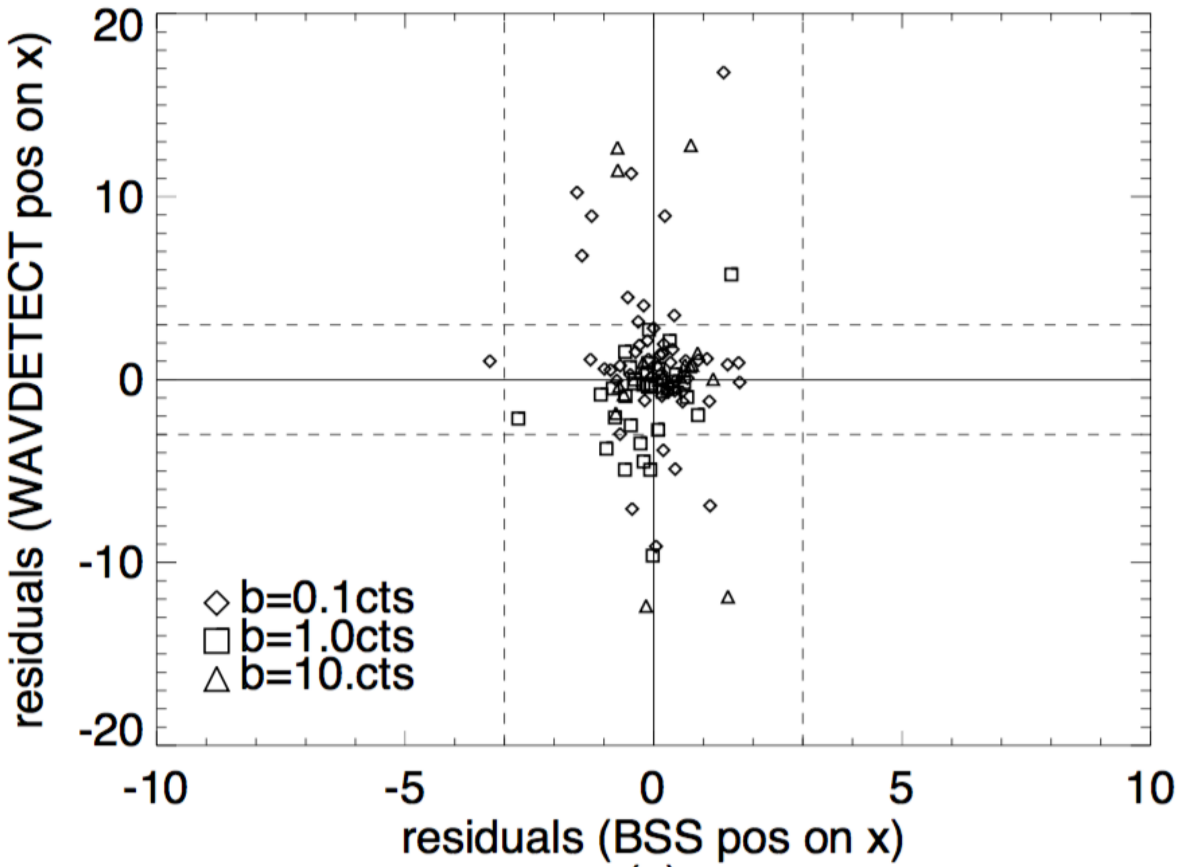
BRAIN in a nutshell

Estimated source intensities



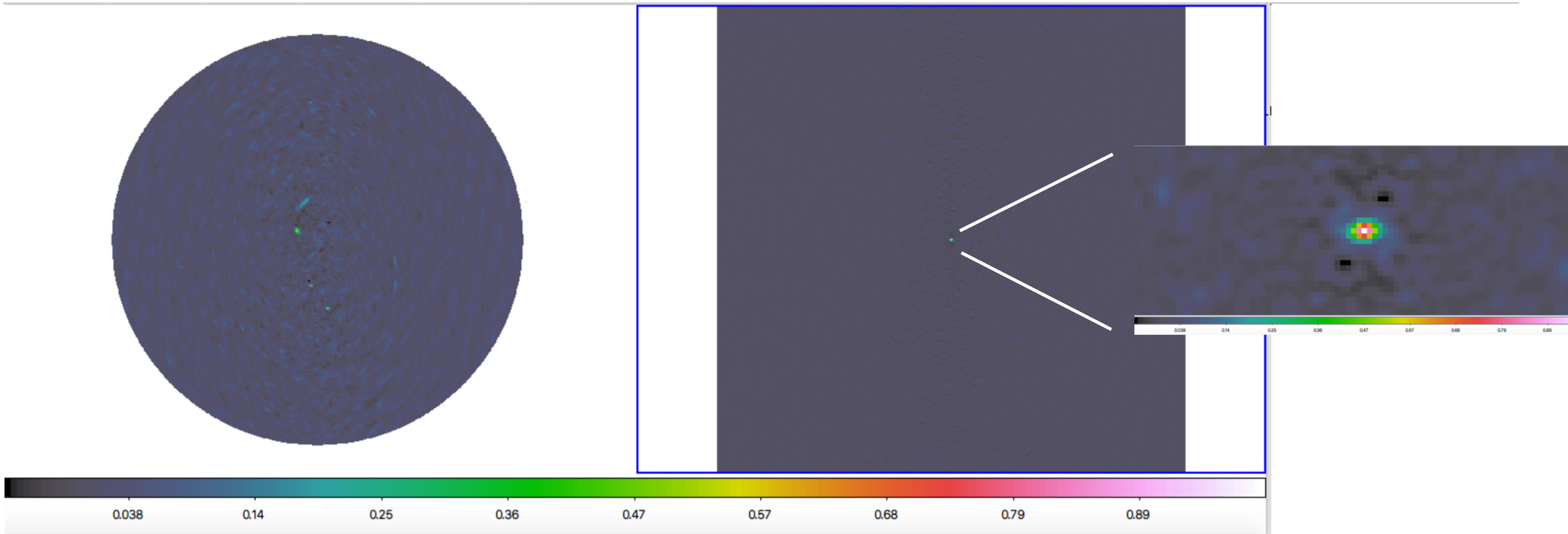
20% of the wavdetect flux sources are underestimated

Estimated source positions



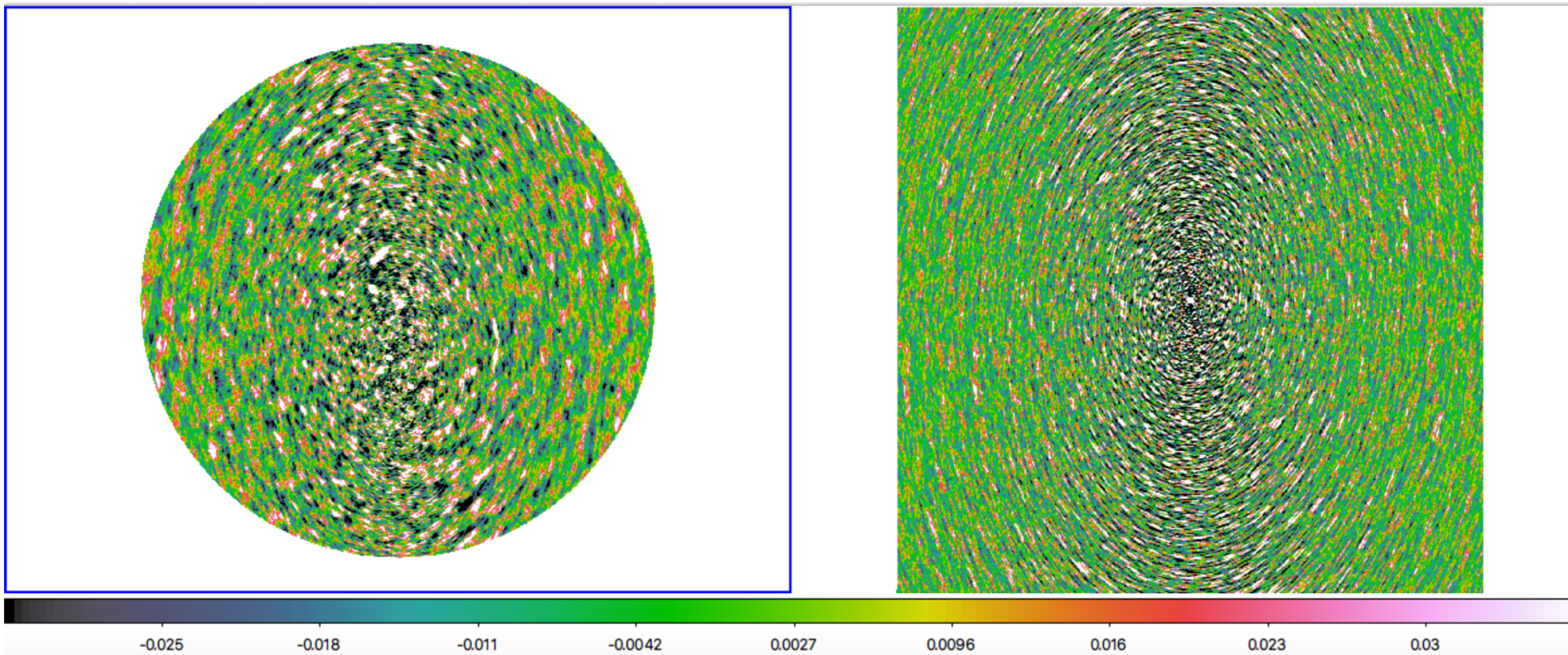
BSS residuals are a factor of 10 smaller than wavdetect ones

BRAIN further extensions



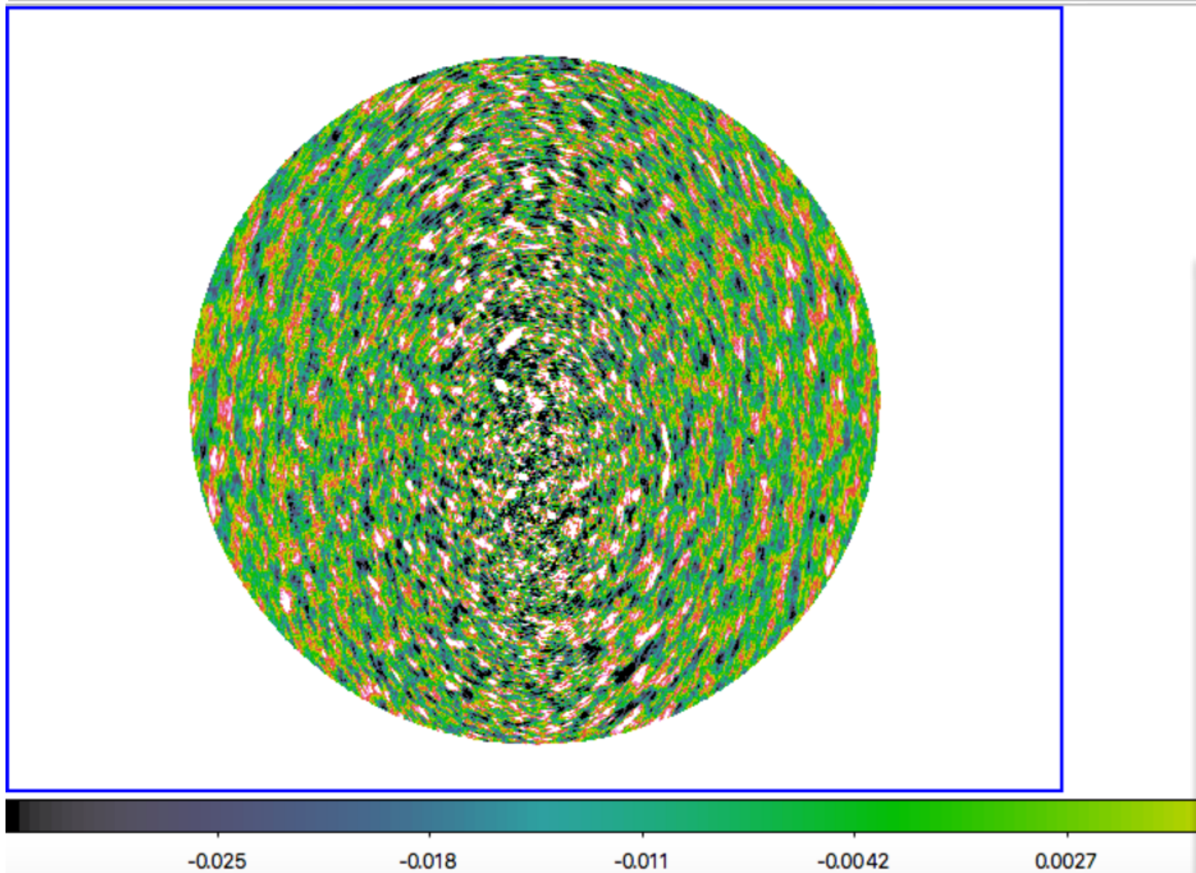
curtesy Ed Fomalont

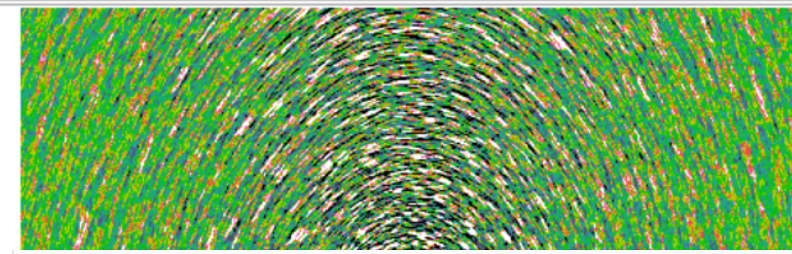
BRAIN further extensions

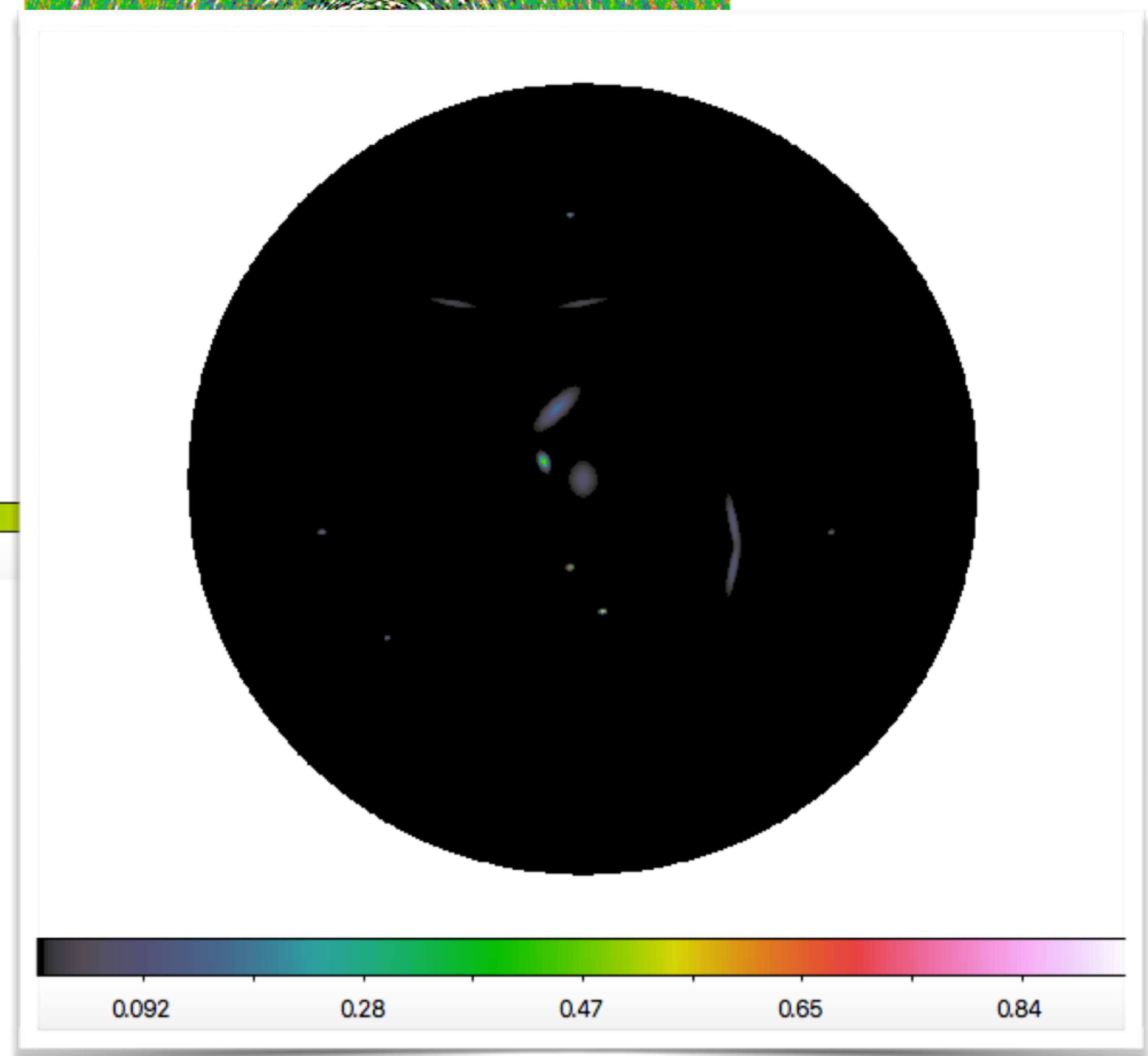


curtesy Ed Fomalont

BRAIN further extensions



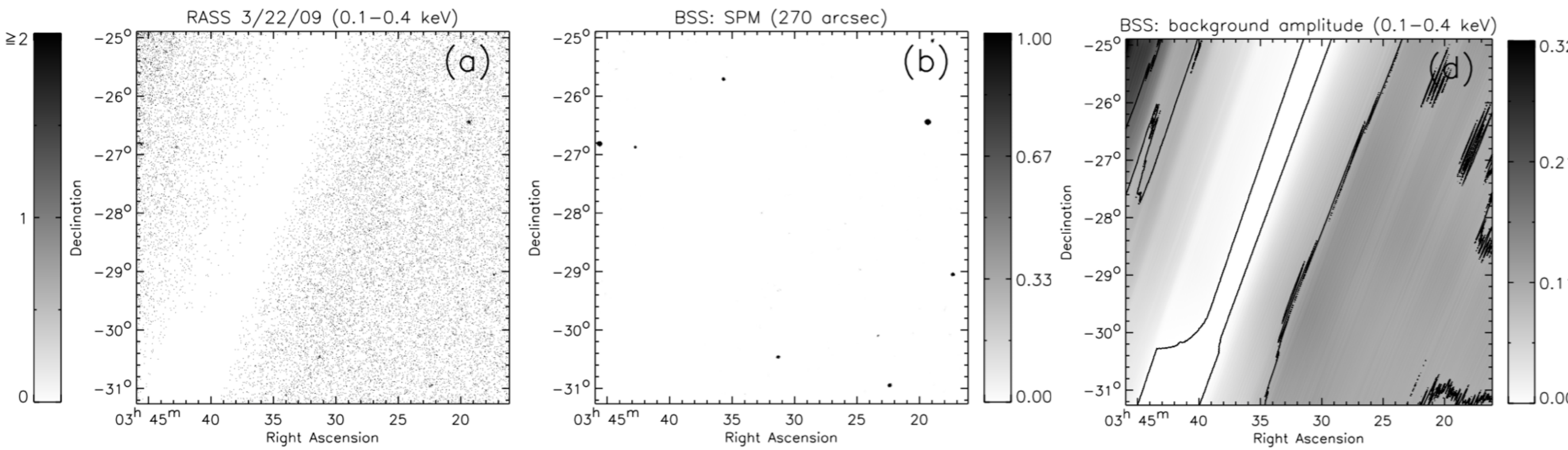




curtesy Ed Fomalont

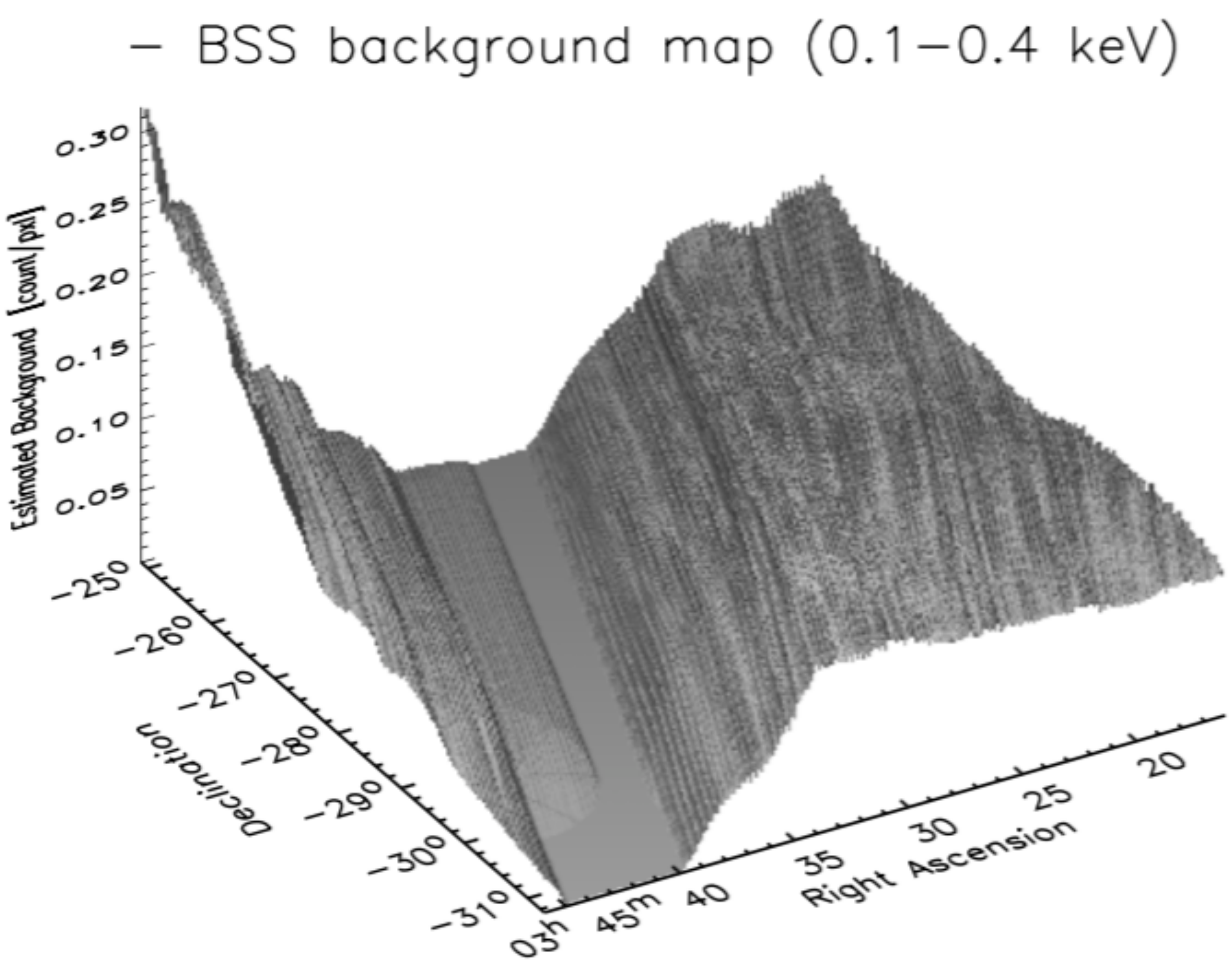
BRAIN further extensions

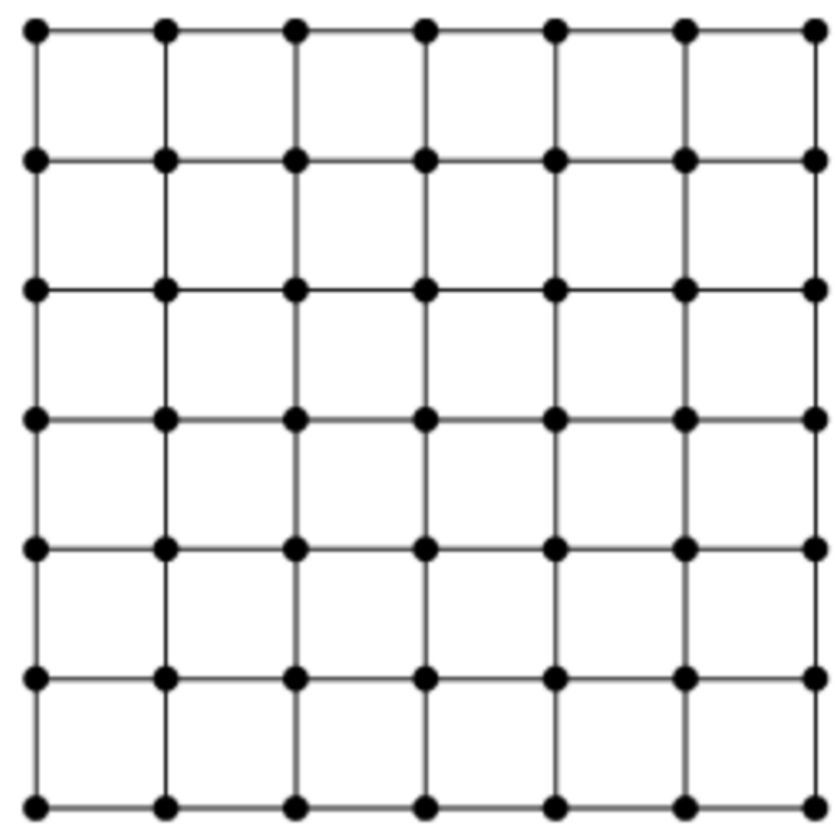
BRAIN will be equipped by a 2D adaptive kernel to improve deconvolution



BRAIN further extensions

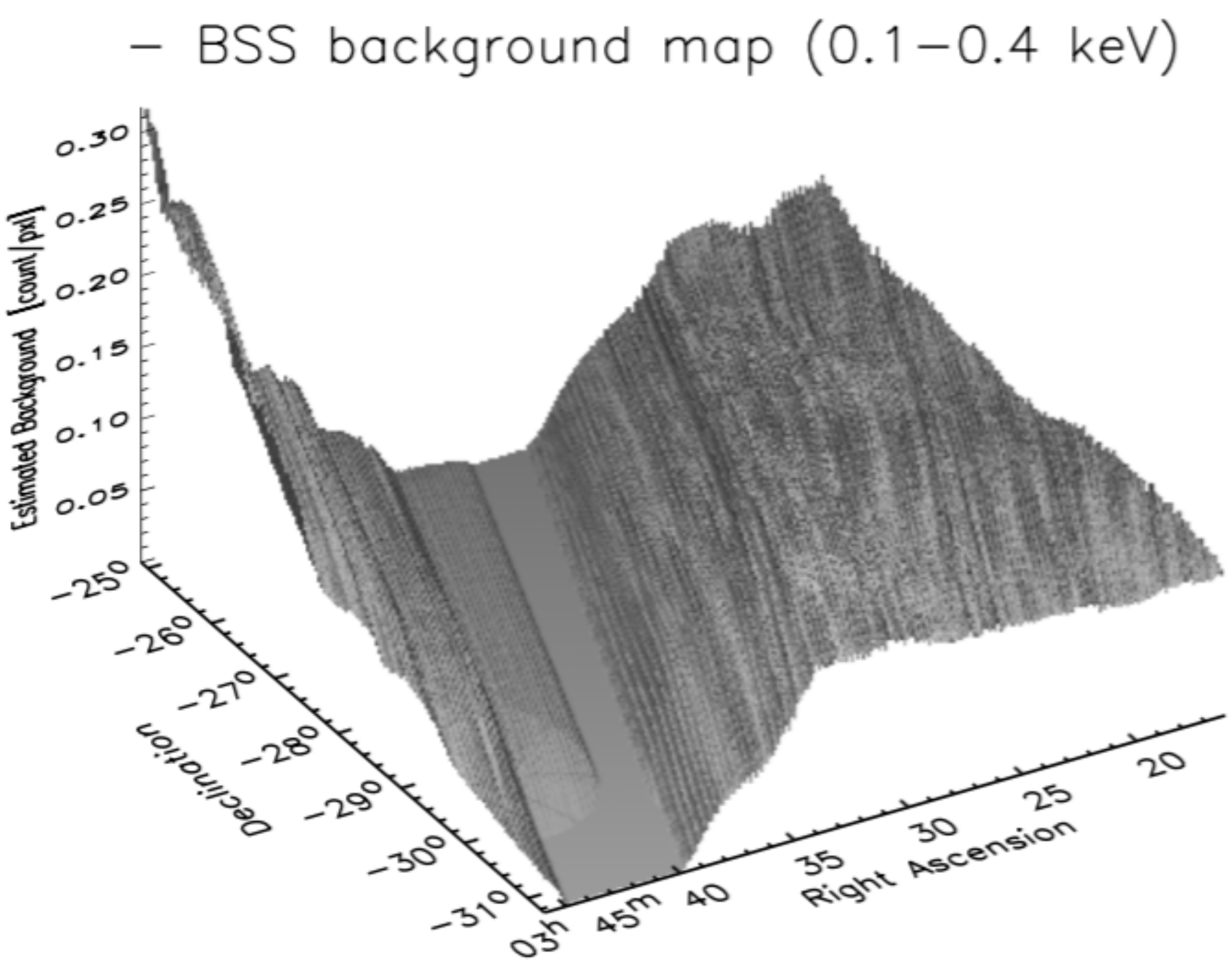
BRAIN will be equipped by a 2D adaptive kernel to improve deconvolution

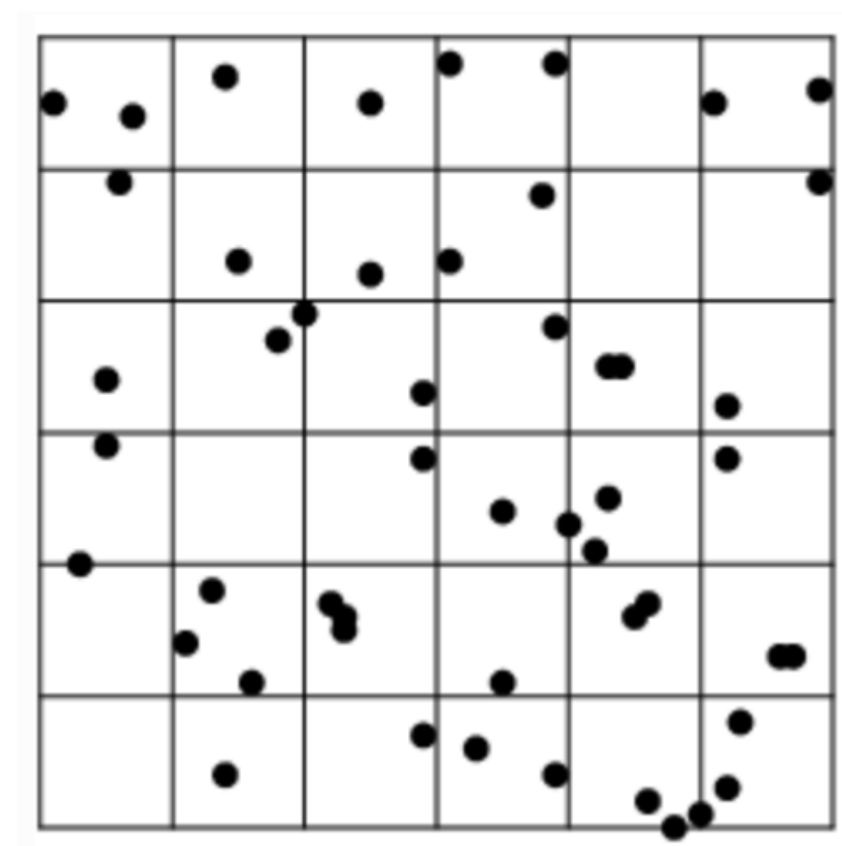




BRAIN further extensions

BRAIN will be equipped by a 2D adaptive kernel to improve deconvolution





R. Fischer et al. Phys. Rev. E61, 1152 (2000)

EUROPEAN ARC
ALMA Regional Centre

Summary

- We aim at providing new alternative tasks to CASA to support image analysis and ALMA 2030 science goals
- Bayes' theorem provides for a different interpretation of probability, wrt frequentist statistics (inductive vs deductive)

  - scientific knowledge is quantified
- In image analysis, source detection is addressed straightforwardly analysing the probability of source detection
  - model comparison and parameter estimation are at the state of the art
- Drawback of numerical complexity can be mitigated by new optimization algorithms (see appendix for more details)
- If you would like to support this study: Join the proposal!

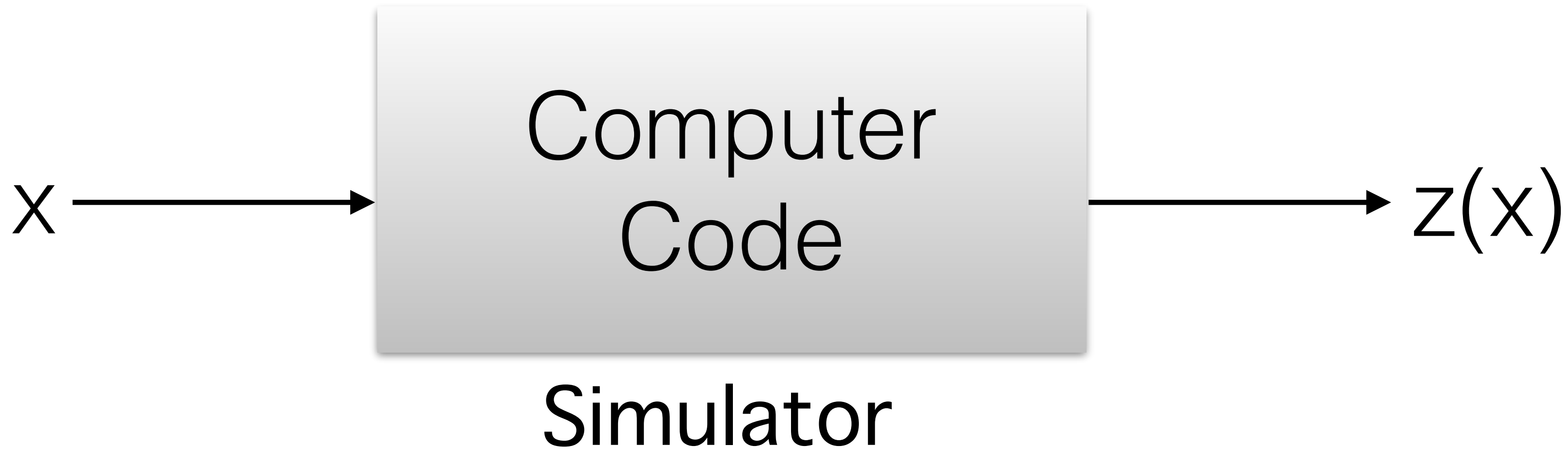
Thank you!

See below the Appendix with:

- Bayesian Surrogate models for optimization
- Application to MEM
- Resolve
- A brain teaser explained in the Bayesian framework

Appendix

Bayesian Surrogate models



Minimisation of Complex Objective function. For reasonable data size, current estimation techniques:

a. *<u>SLOW</u> if full objective function is accounted*

b. *<u>FAST</u> if do not account for full objective function*

Note: To increase computer power, performance can be enhanced via global optimization as a Bayesian decision problem where two connected statistical layers are employed

Bayesian Surrogate models

Emulator

- statistical representation of x
- expresses knowledge about z(x) at any given x
- built using prior information and training set of model runs

Kriging surrogate sampler is used to emulate the original Complex Objective function

Bayesian Surrogate models

- Data: x, y
- Model: $y = f_k(x) + \epsilon$

Gaussian likelihood: $$p(\mathbf{y}|\mathbf{x}, \mathbf{k}, M_i) \propto \prod_j \exp\left(-\frac{1}{2}(\mathrm{y_j} - \mathrm{f_k}(\mathrm{x_j}))^2/\sigma^2\right)$$

Prior over the param: $$p(\mathbf{k}|M_i)$$

Posterior param distr: $$p(\mathbf{k}|\mathbf{x}, \mathbf{y}, M_i) = \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{k}, M_i)p(\mathbf{k}|M_i)}{p(\mathbf{y}|\mathbf{x}, M_i)}$$

Make predictions:

$$p(y^*|x^*, \mathbf{x}, \mathbf{y}, M_i) = \int p(y^*|\mathbf{k}, x^*, M_i)p(\mathbf{k}|\mathbf{x}, \mathbf{y}, M_i)d\mathbf{k}$$

Bayesian Surrogate models

$$P(s|d) = \frac{P(d|s)P(s)}{P(d)} \equiv \frac{e^{-H(d,s)}}{Z_d}$$

$$d = (d_1, d_2, \ldots, d_n)^T \qquad n \in \mathbb{N}$$

$$\langle s \rangle = \mathrm{argmin}_{\langle \mathrm{s}|\mathrm{d} \rangle} \mathrm{H}(\mathrm{d}, \mathrm{s})$$

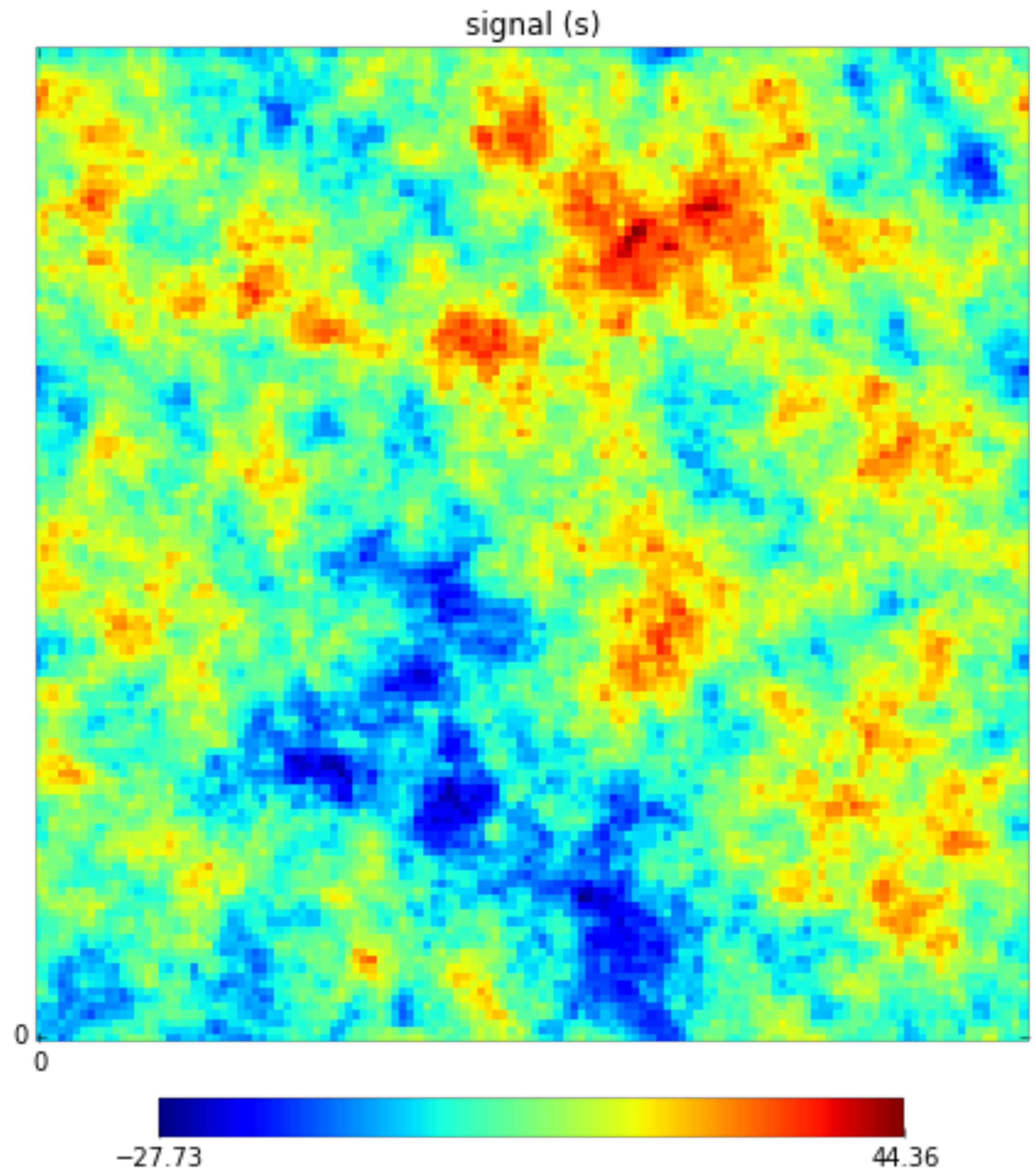
$$P(y(\mathbf{x})|\mathbf{t}_i, \mathbf{X}_i, \mathbf{I}) = \frac{P(\mathbf{t}_i|y(\mathbf{x}), \mathbf{X}_i, \mathbf{I})P(y(\mathbf{x}))}{P(\mathbf{t}_i)} \equiv \frac{e^{-H(y(\mathbf{x}),\mathbf{t}_i)}}{Z_{t_i}}$$

$$\mathbf{t}_n = \{t_i\}_{i=1}^n$$

Bayesian Surrogate models

Simulated sky signal employing NIFTY package

1.6E+4 dimensions



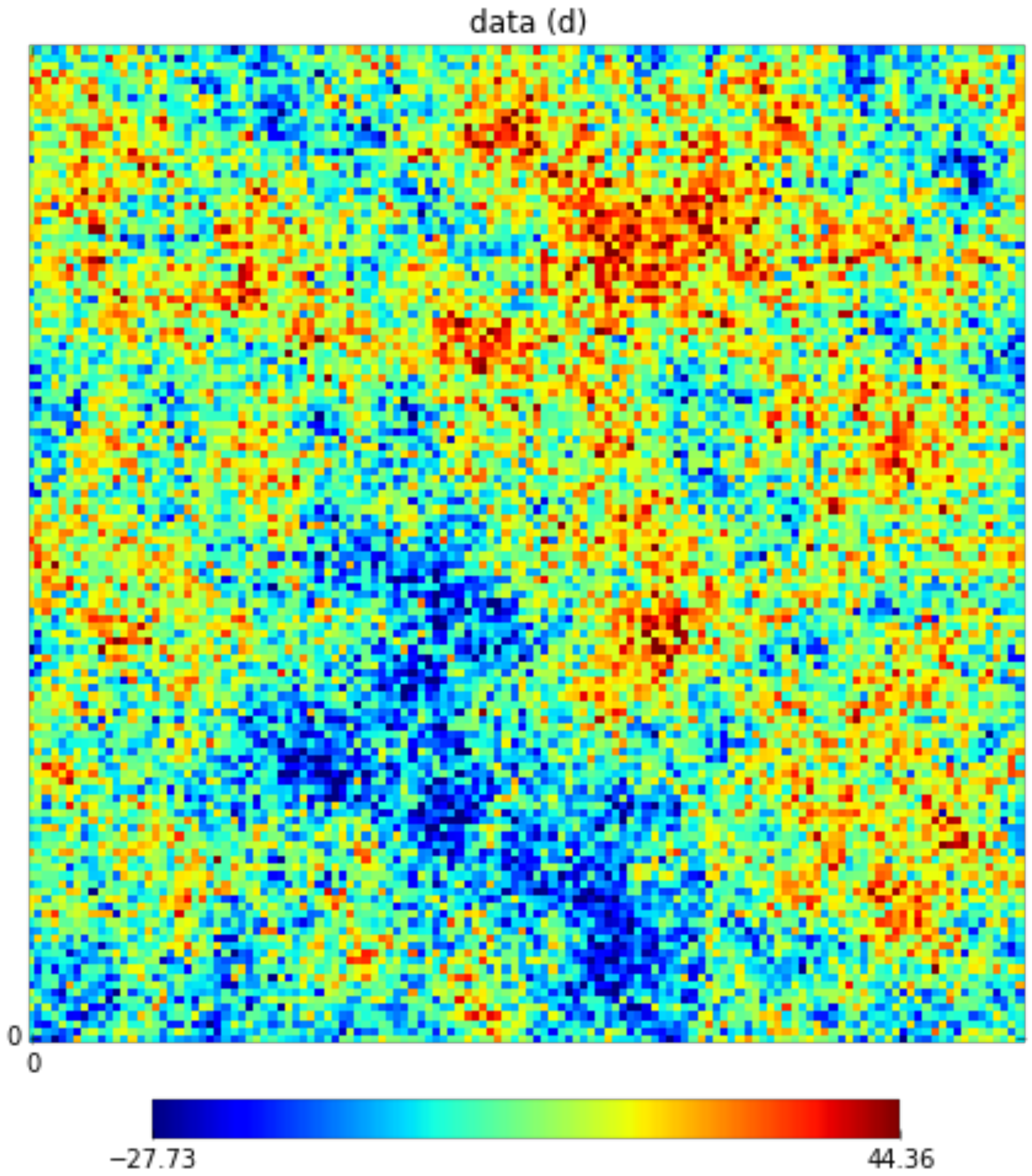
NIFTY= Numerical Information Field Theory

(M.Selig, T. Enßlin et al., 2013, A&A, 554A, 26)

Bayesian Surrogate models

Simulated dataset employing NIFTY package

$$d=R(s) + n$$



NIFTY= Numerical Information Field Theory

(M.Selig, T. Enßlin et al., 2013, A&A, 554A, 26)

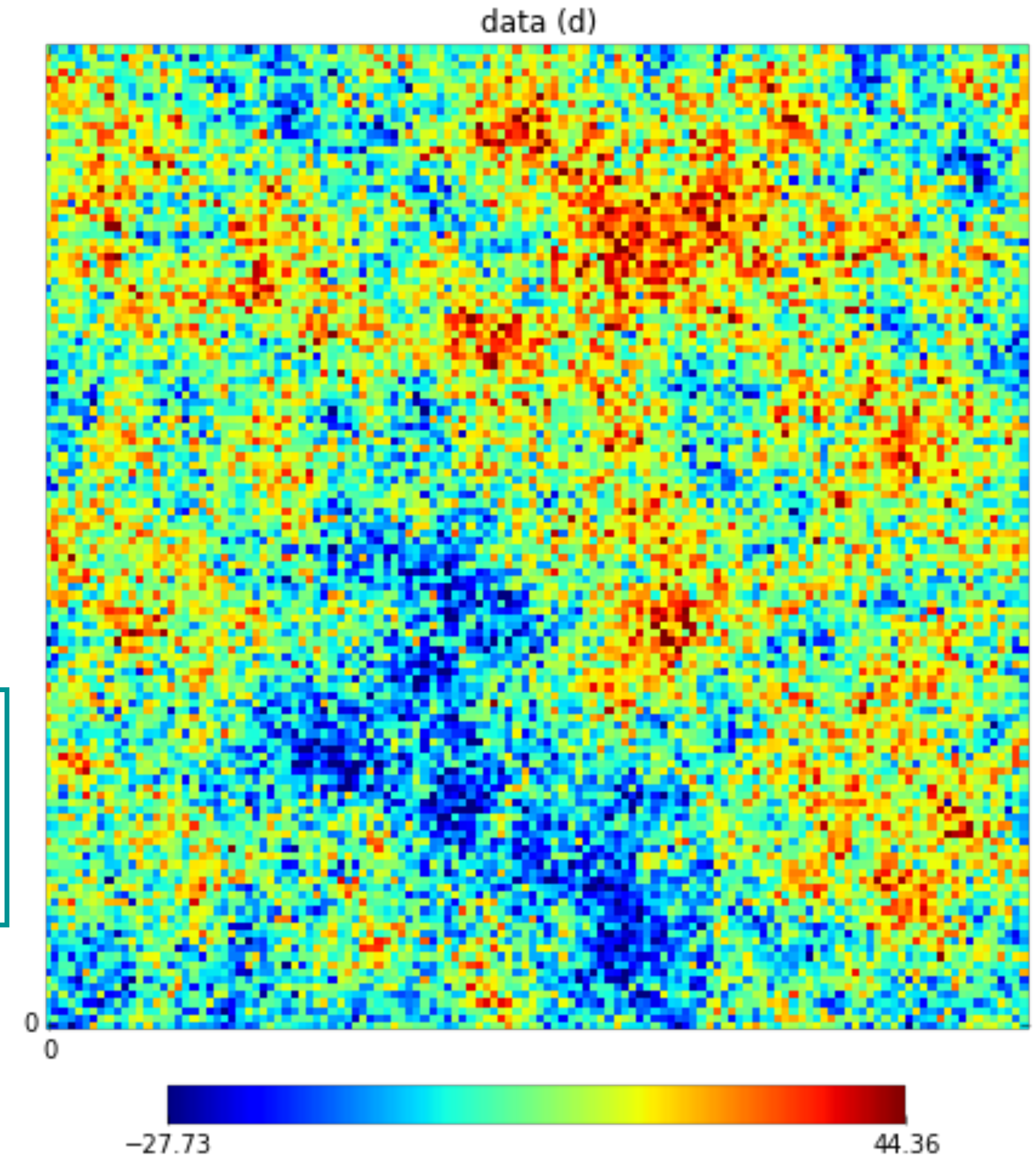
Bayesian Surrogate models

Simulated dataset employing NIFTY package

d=R(s) + n

find optimal solution for s given d

assumption: s is random field following some statistics and being constrained by the data

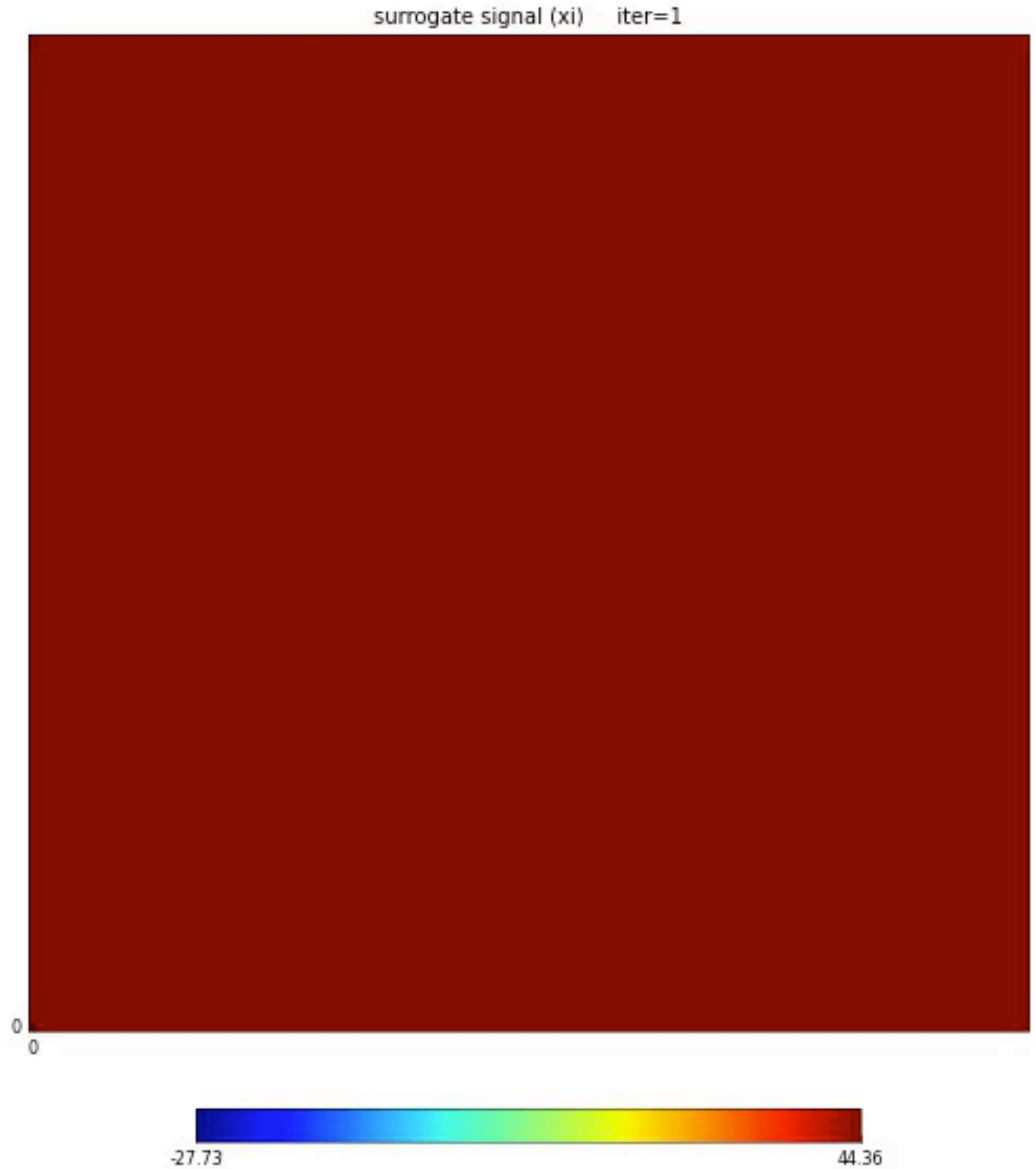


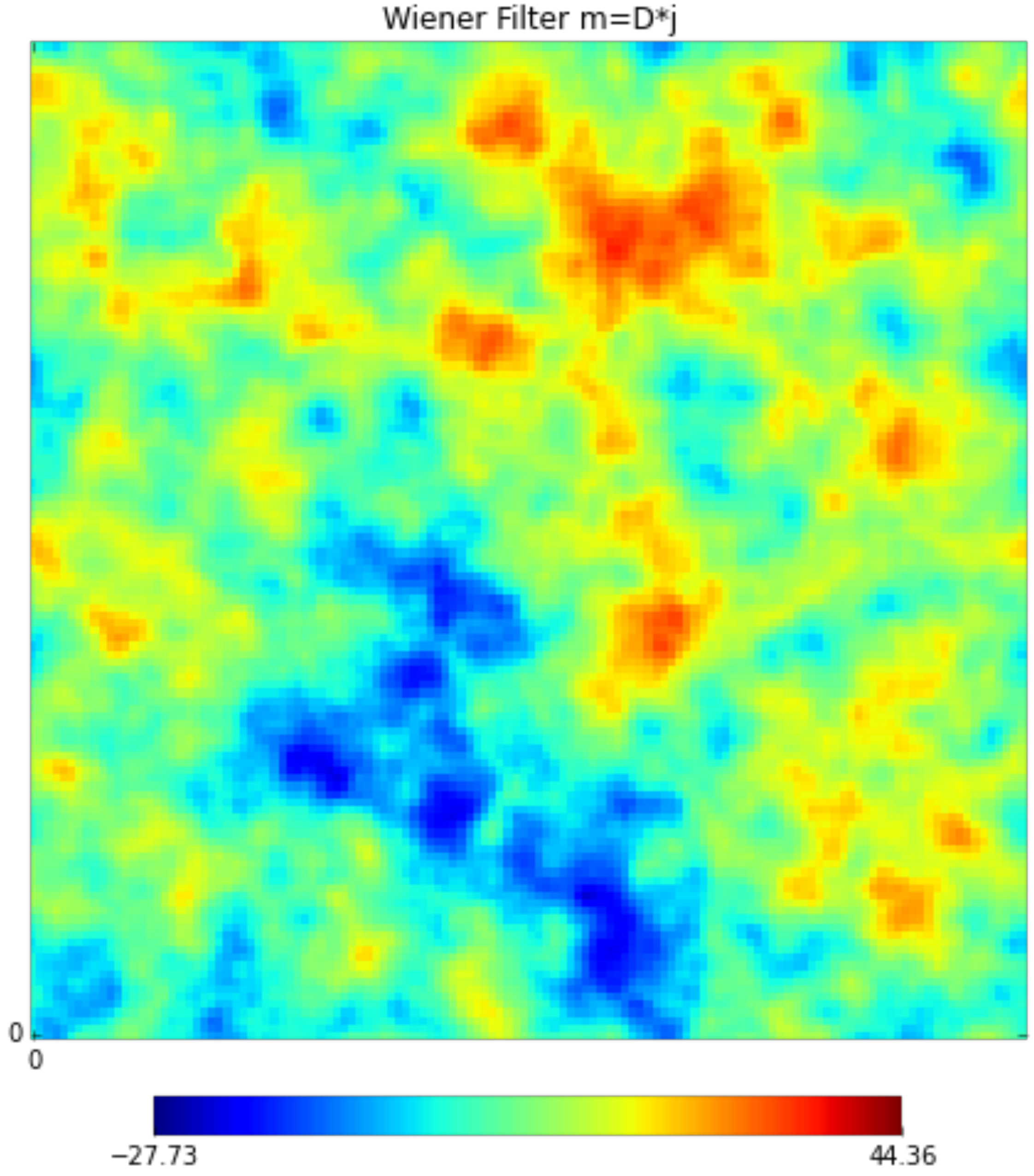
NIFTY= Numerical Information Field Theory

(M.Selig, T. Enßlin et al., 2013, A&A, 554A, 26)

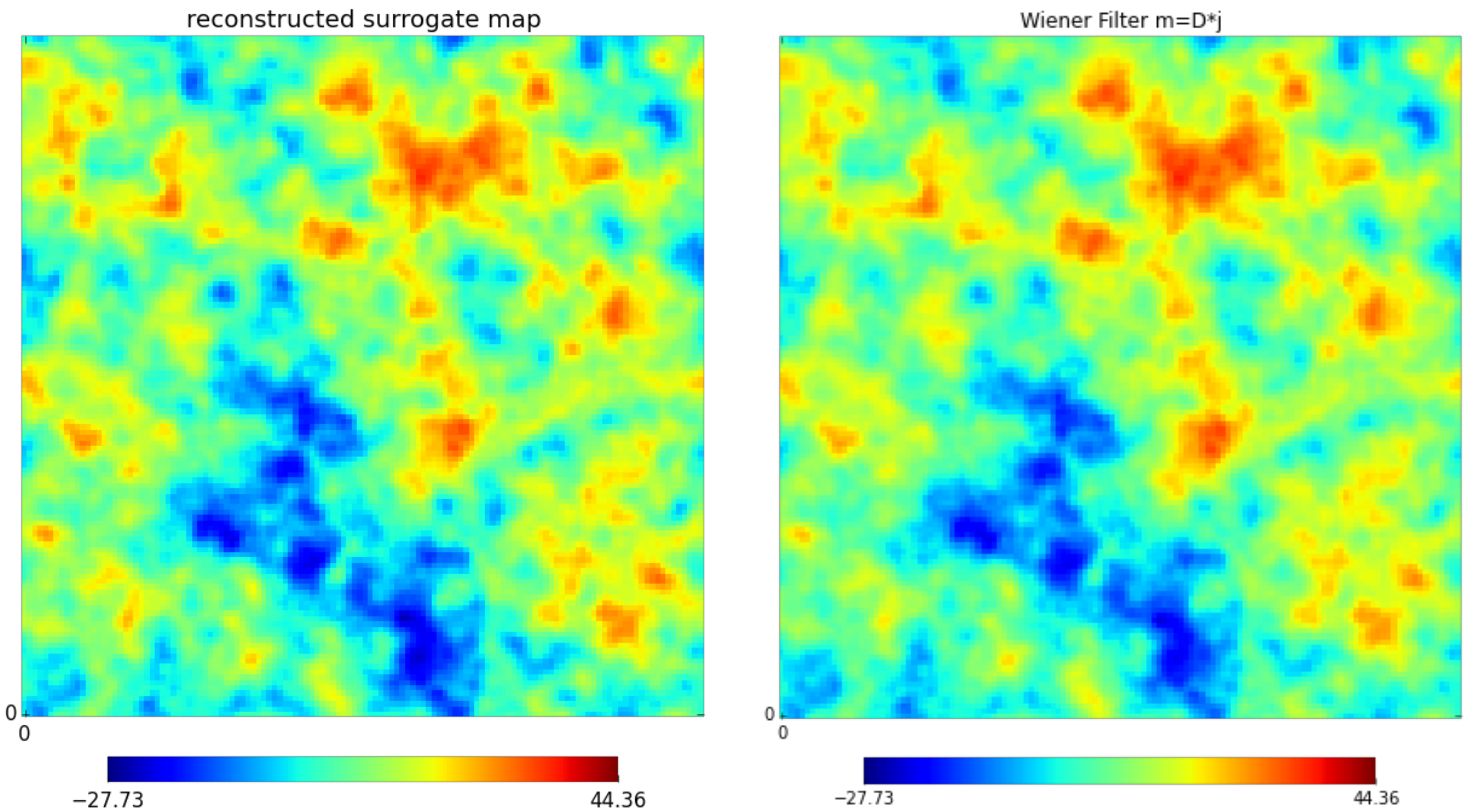
Bayesian Surrogate models

Here is a movie of the optimization scheme. If you want to see it, send me a message fguglie l@eso.org

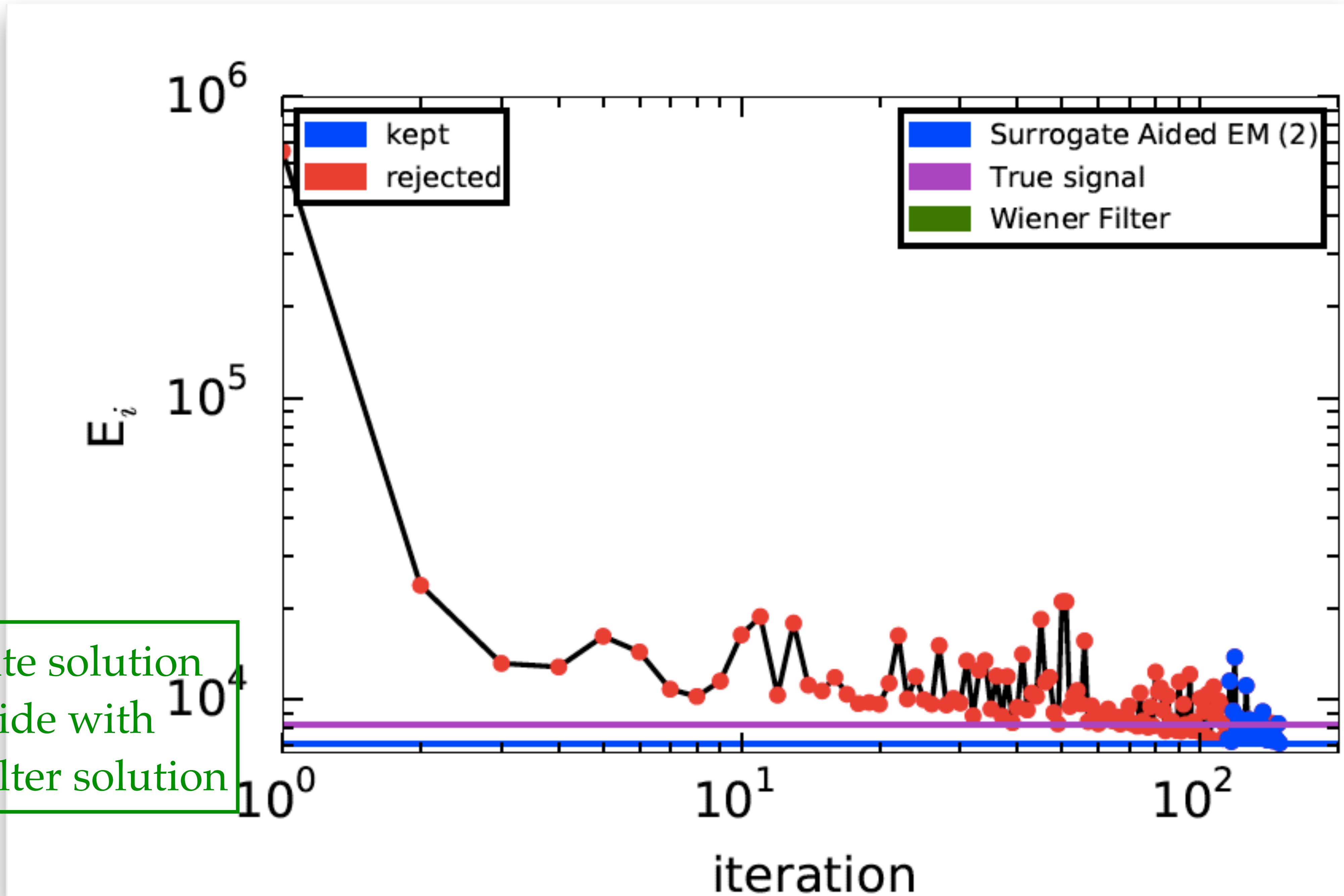




Bayesian Surrogate models

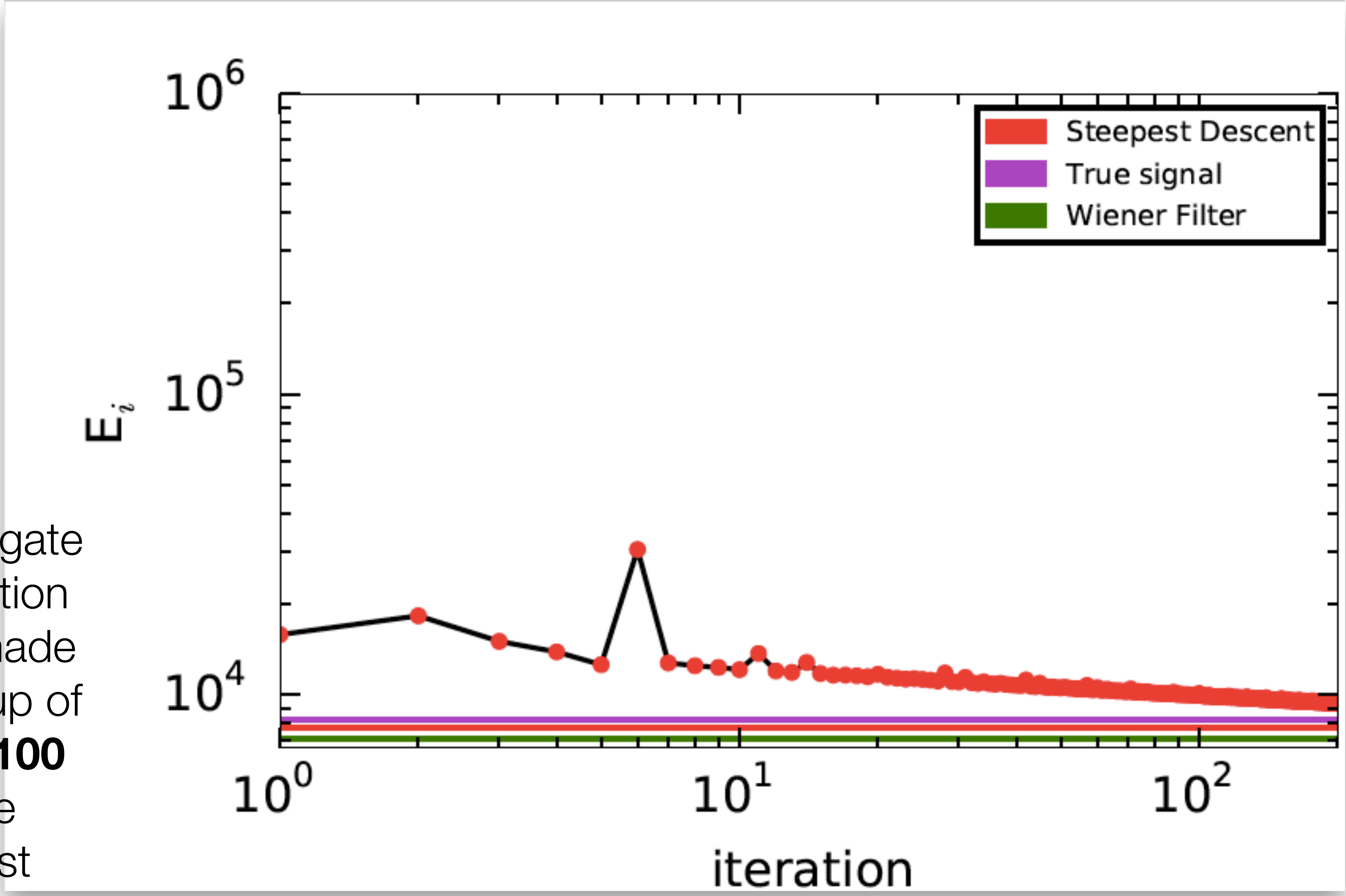


Bayesian Surrogate models



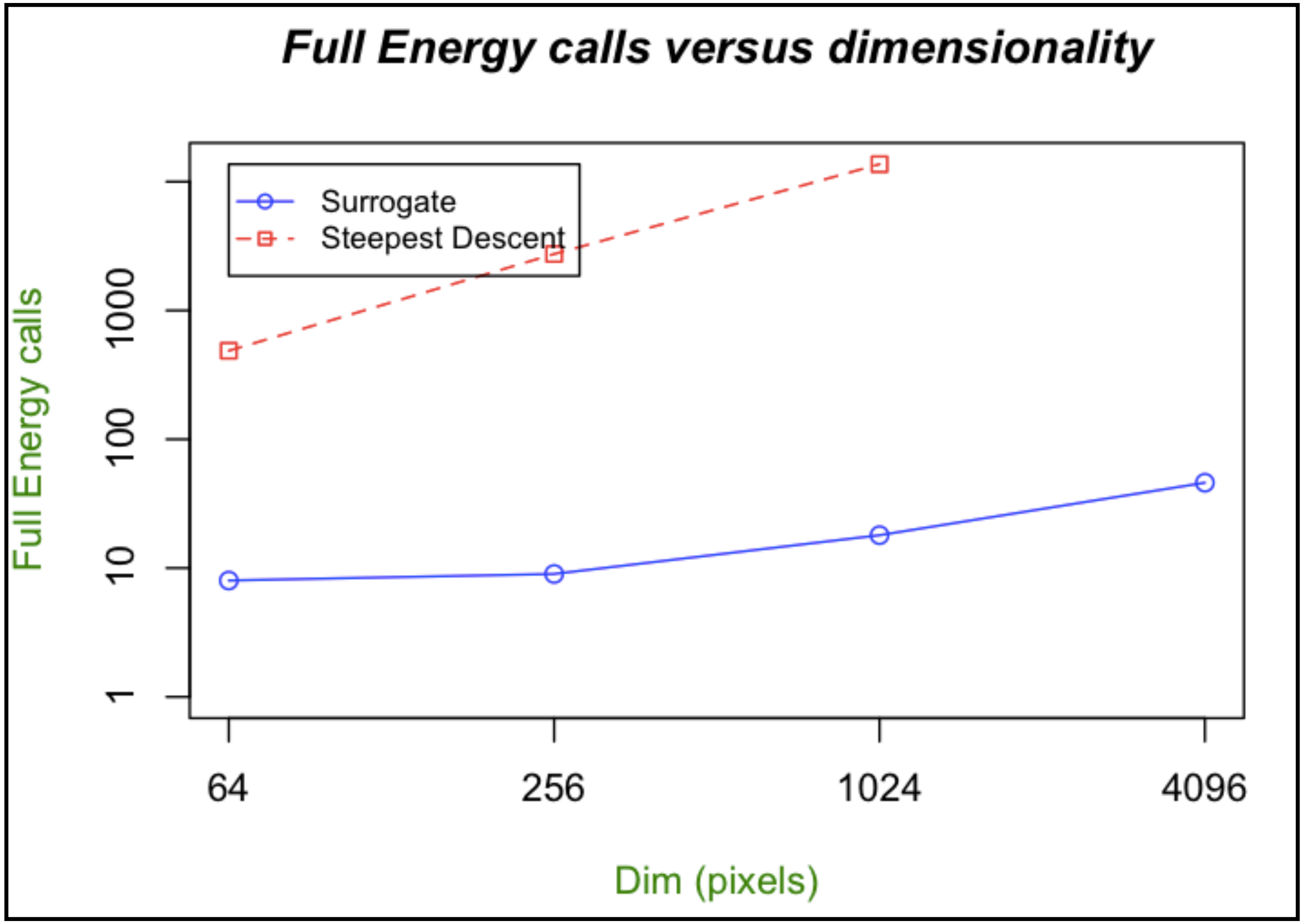
Surrogate solution coincide with Wiener filter solution

Bayesian Surrogate models



The surrogate minimisation routine made a speed up of a factor **100** wrt the steepest descent

Bayesian Surrogate models



Maximum Entropy methods

(Strong, A., 1995, ExA, 6, 97)

- Entropy: $$S = -\sum f_i log f_i$$

  (intensity in pixel i relative to model value)

- Poisson log likelihood:

$$L = logP(d|M) = \sum lne^{-x}x^n/n!$$

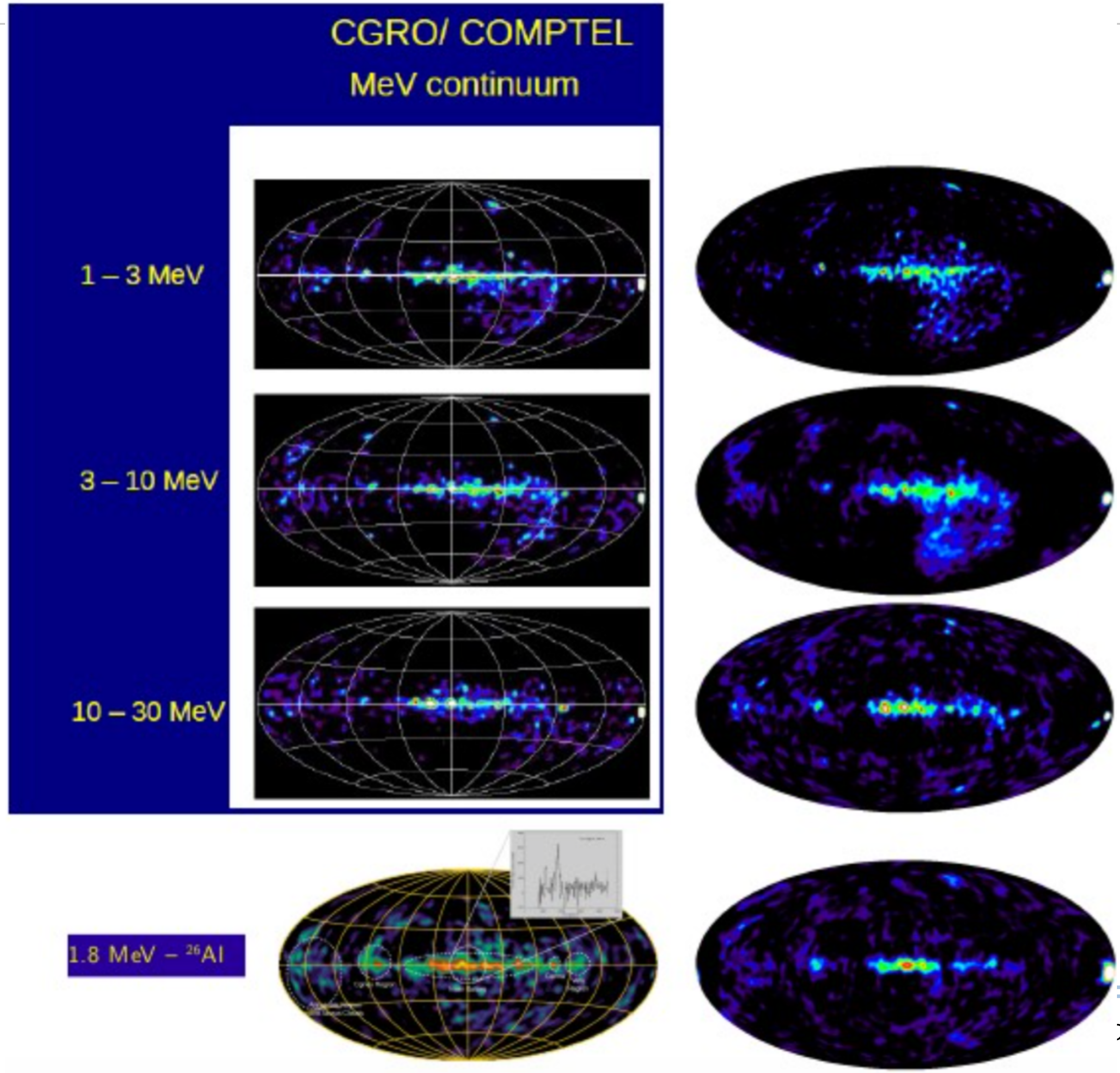
$$C = S - \lambda L$$

- S is maximised over all images consistent with the data:

  - Lagrangian multiplier =0 as start, flat map
  - C is maximised, max S for given L
  - Stop when L is consistent with the data

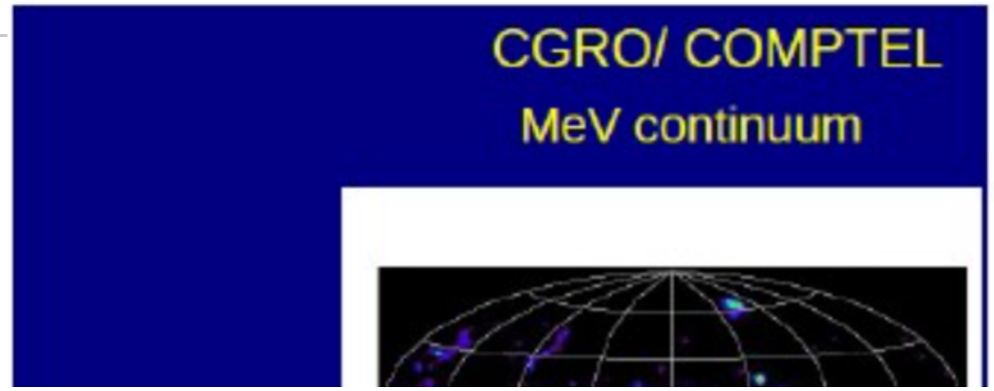
Maximum Entropy methods

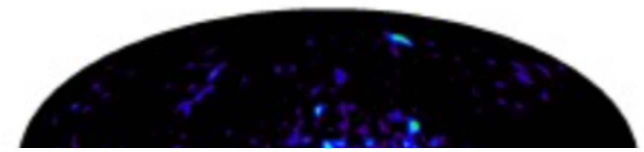
(Strong, A., 1995, ExA, 6, 97)



Maximum Entropy methods

(Strong, A., 1995, ExA, 6, 97)





COMPTEL reloaded: new initiatives in heritage MeV gamma-ray astronomy

A. W. Strong[1], W. Collmar[1], T. A. Enβlin[2], M .Reinecke[2], D. Pumpe[2], F. Guglielmetti[2]

Max-Planck-Institut für extraterrestrische Physik

1. MPE Garching 2. MPA Garching

Max-Planck-Institut für Astrophysik

ABSTRACT

The COMPTEL gamma-ray telescope on NASA's Compton Gamma Ray Observatory (CGRO) operated from 1991 to 2000. It was a double-scatter Compton instrument covering the energy range 0.75-30 MeV, both in continuum and lines. Full-sky maps and a source catalogue were the main outcome of the mission. While the Fermi-LAT instrument has now vastly enhanced our knowledge of the gamma-ray sky at higher energies, the MeV range remains devoid of new missions, so that the heritage COMPTEL data is an essential resource. Data analysis has continued at MPE Garching, with improved event processing and selections. The original skymapping method using Maximum Entropy has been adapted to current technology. A new initiative for skymapping using state-of-the-art Bayesian techniques has been started at MPA Garching; this involves Information Field Theory with the D³PO system.

INSTRUMENT

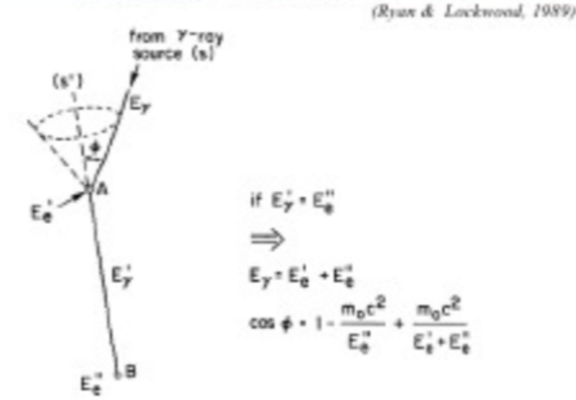
COMPTEL on CGRO (1991 - 2000)

DETECTION TECHNIQUE

The double-Compton scattering detection technique means that each photon is associated with an annulus on the sky, with known centre, radius and shape; this makes deconvolution essential for imaging. One method is Maximum Entropy (Maxent), which has been used to make all-sky images. The large instrumental background is a further challenge for any COMPTEL analysis.

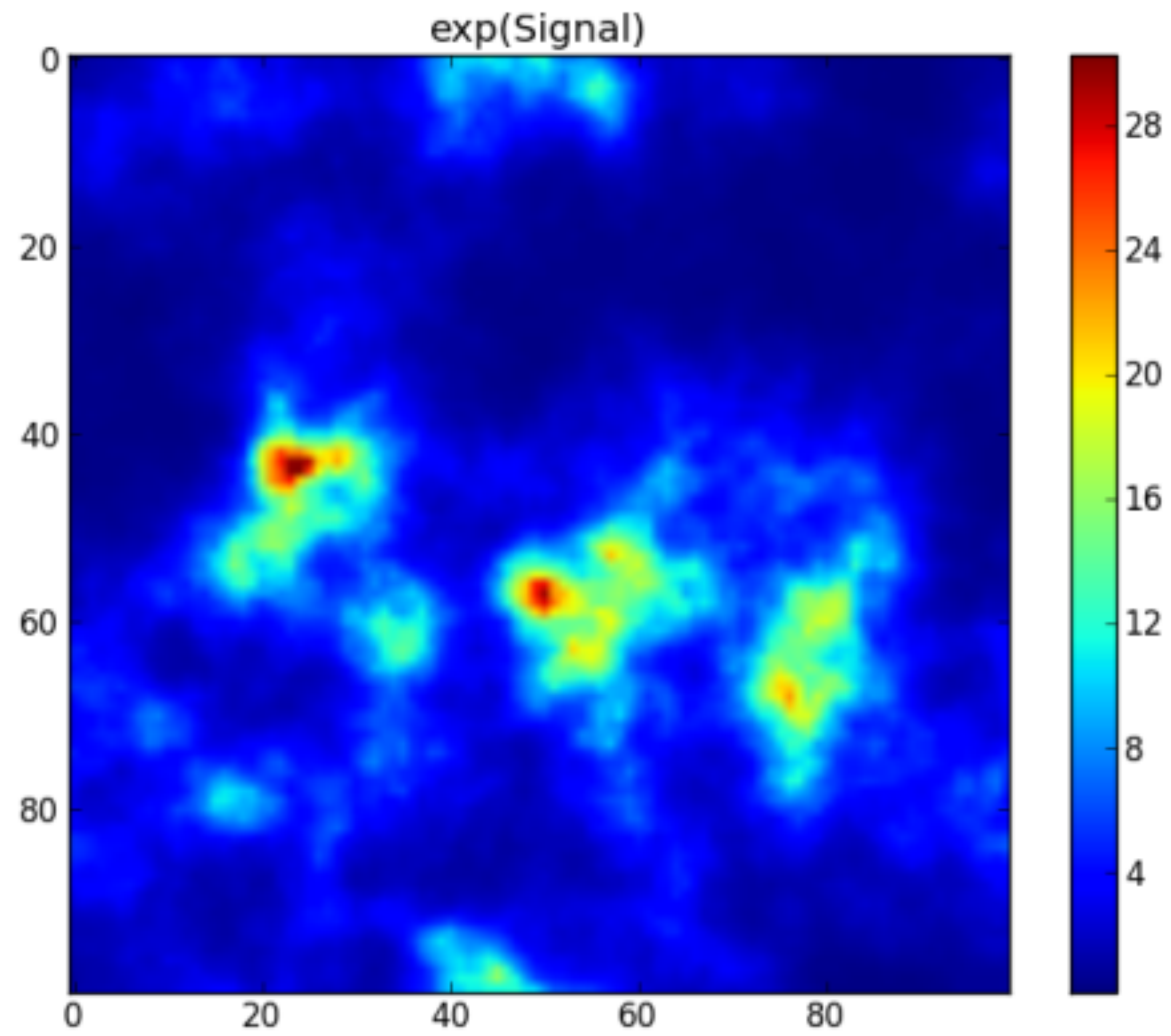
PRINCIPLE OF MEASUREMENT

(Ryan & Lockwood, 1989)



RESOLVE

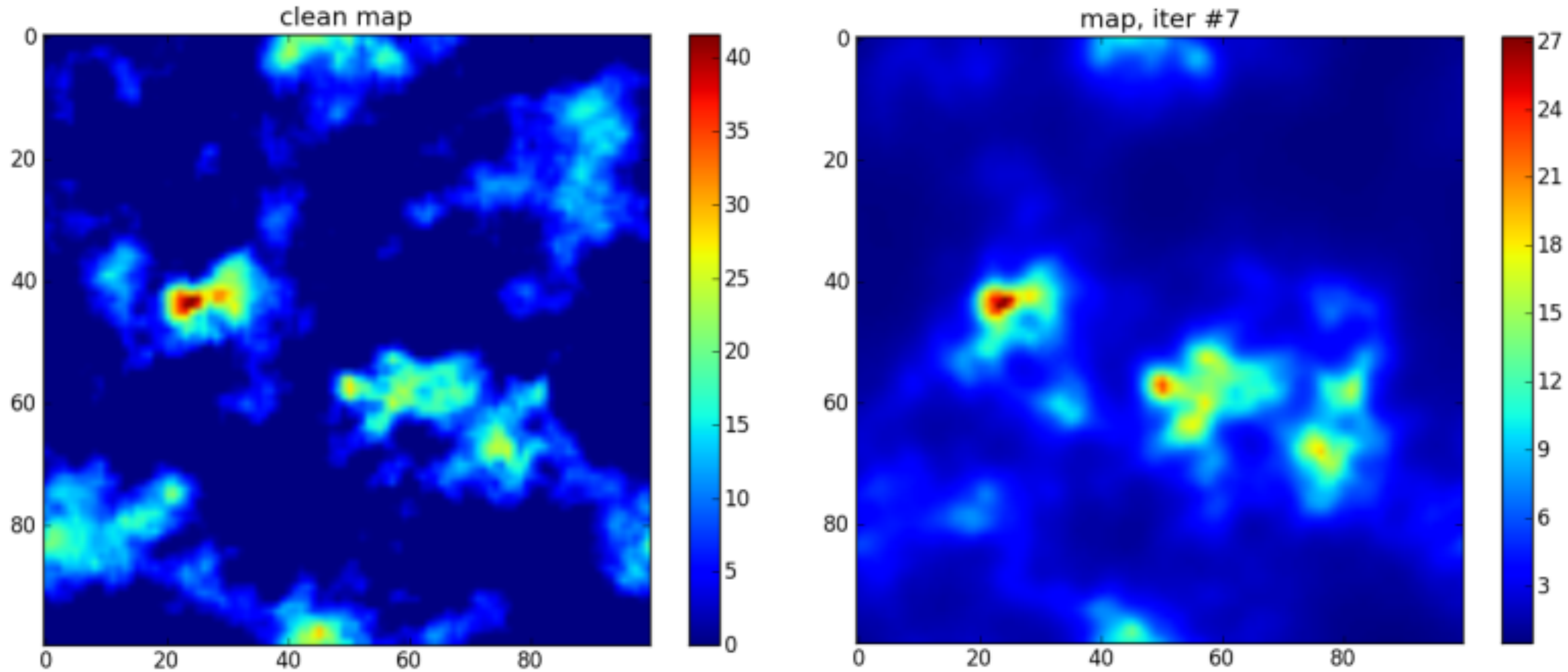
Junklewitz, H. (2016) A&A 586, 76



Simulated VLA signal (Jy / pix)

Newest version by Arras, P. (2019) arXiv:1903.11169

RESOLVE



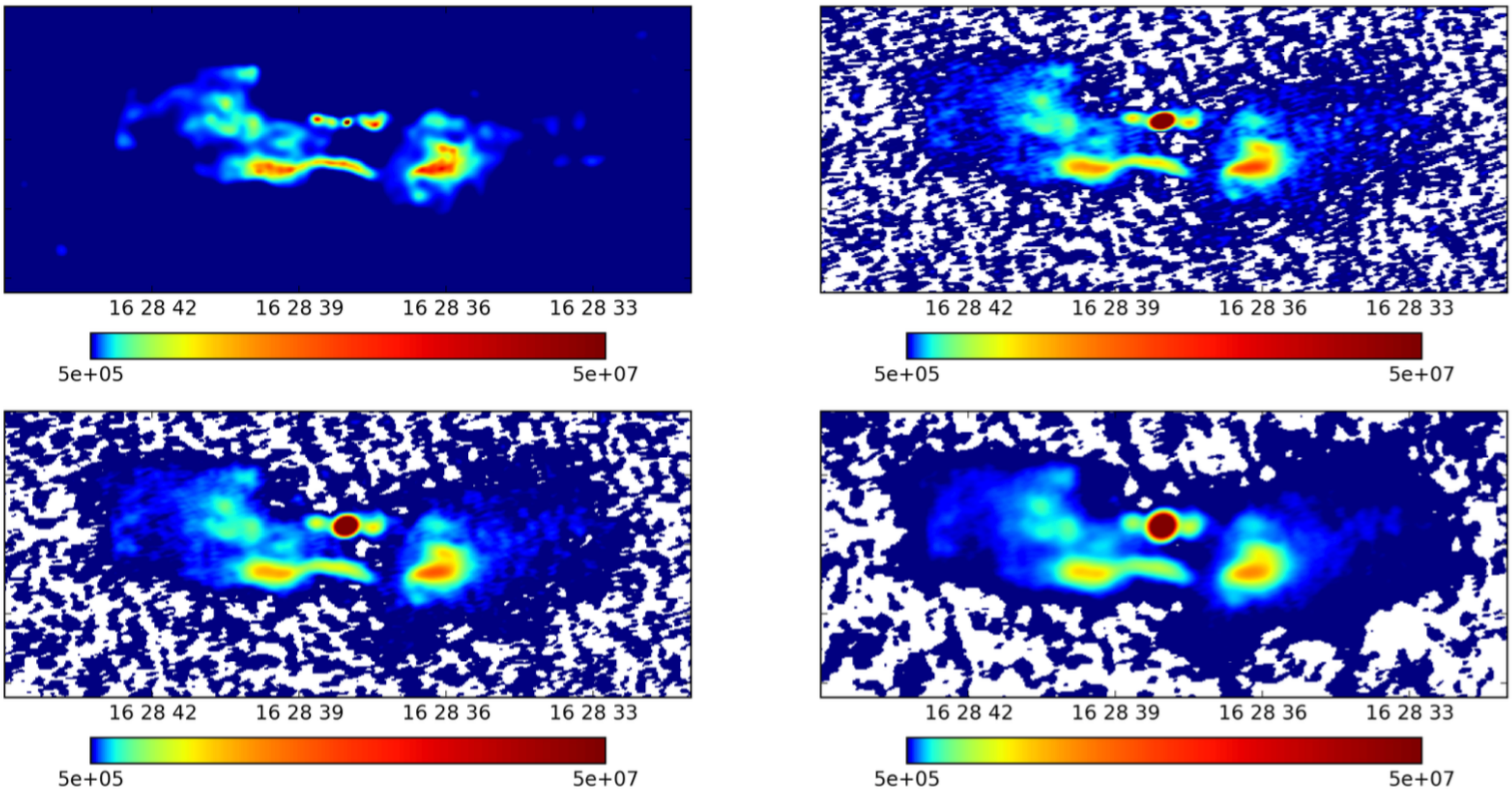
low noise, 40% uv-coverage

RESOLVE

Greiner, M. (2016) arXiv160504317

3C 338 at 8415 MHz

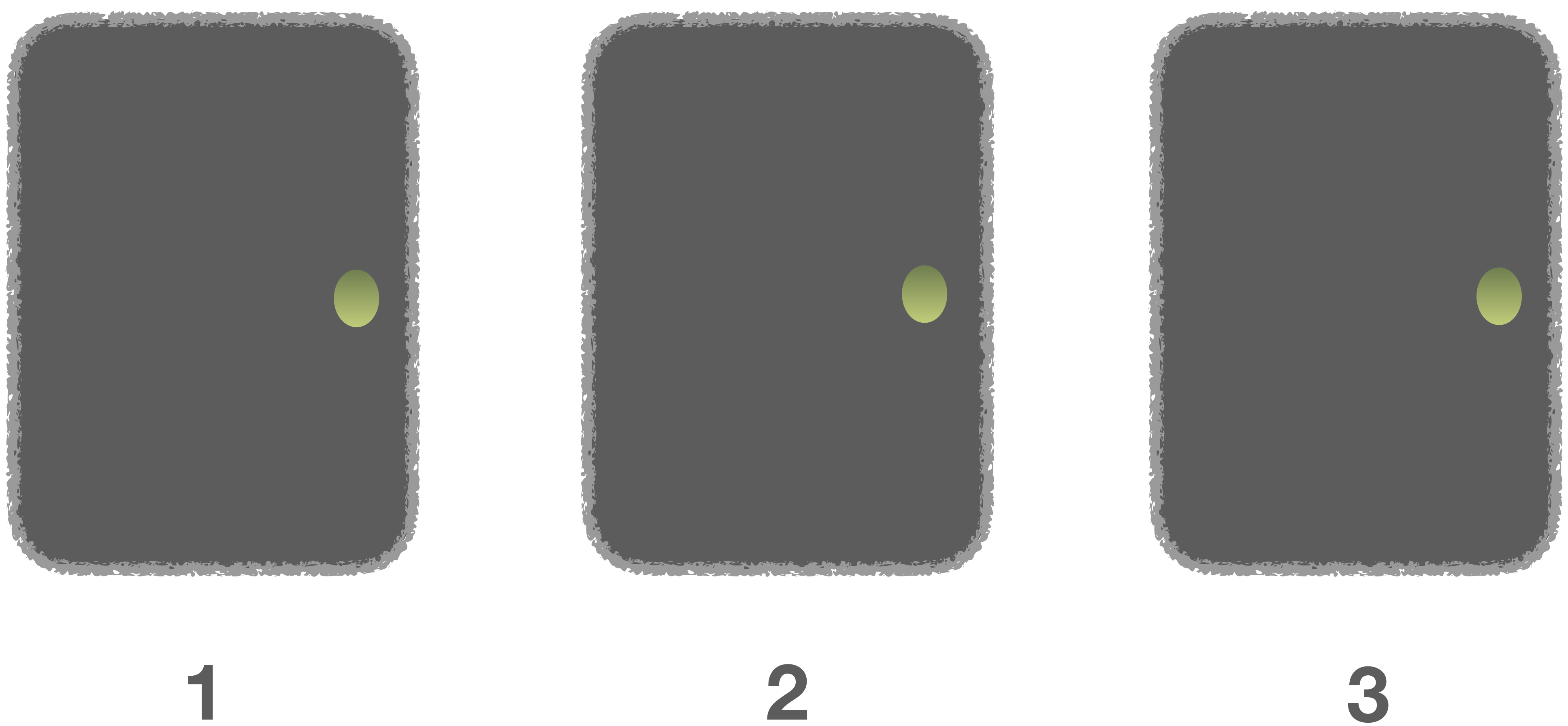
fastRESOLVE clean -uniform weight.



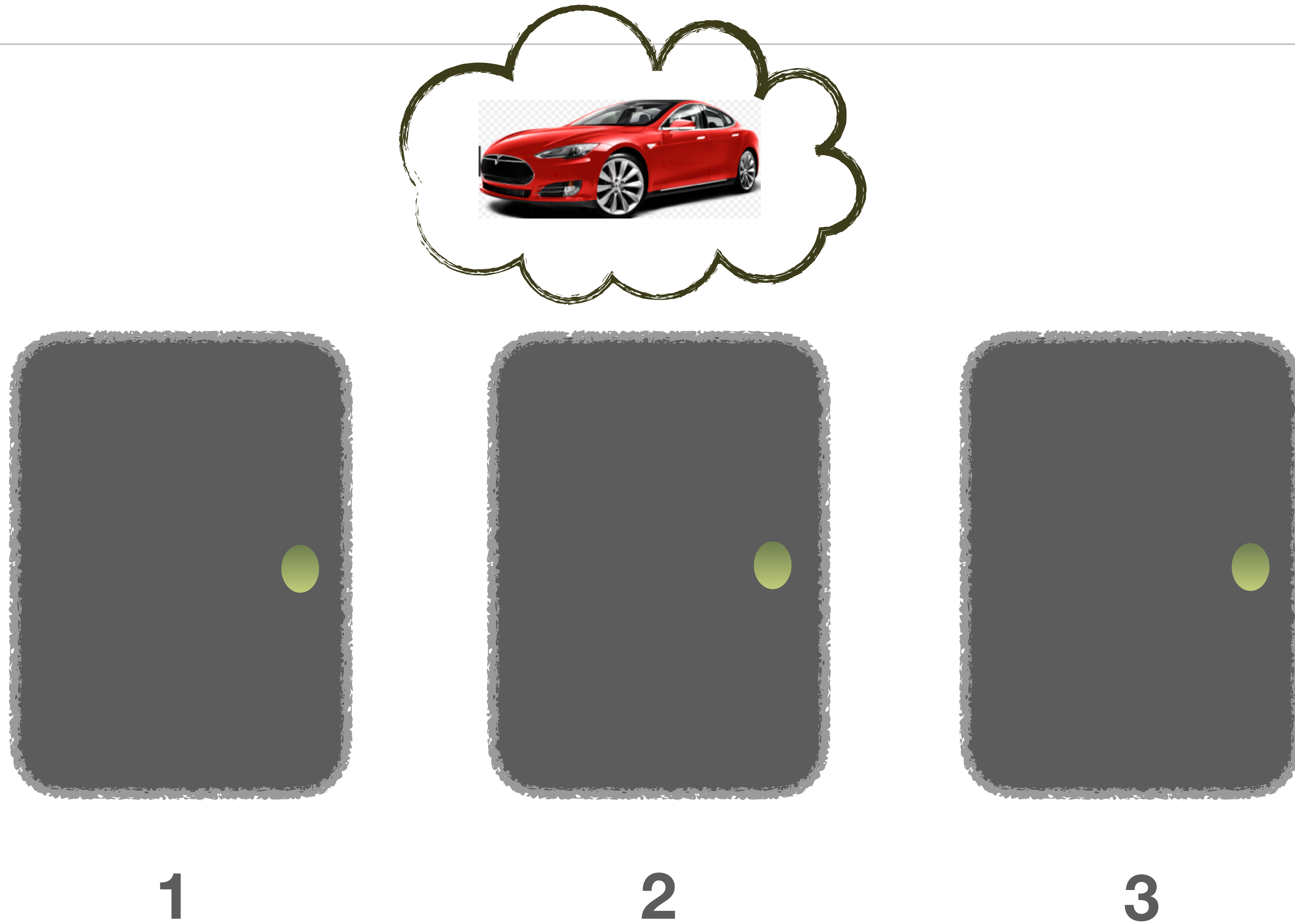
clean -intermediate weight. clean -natural weight.

Monty Hall problem

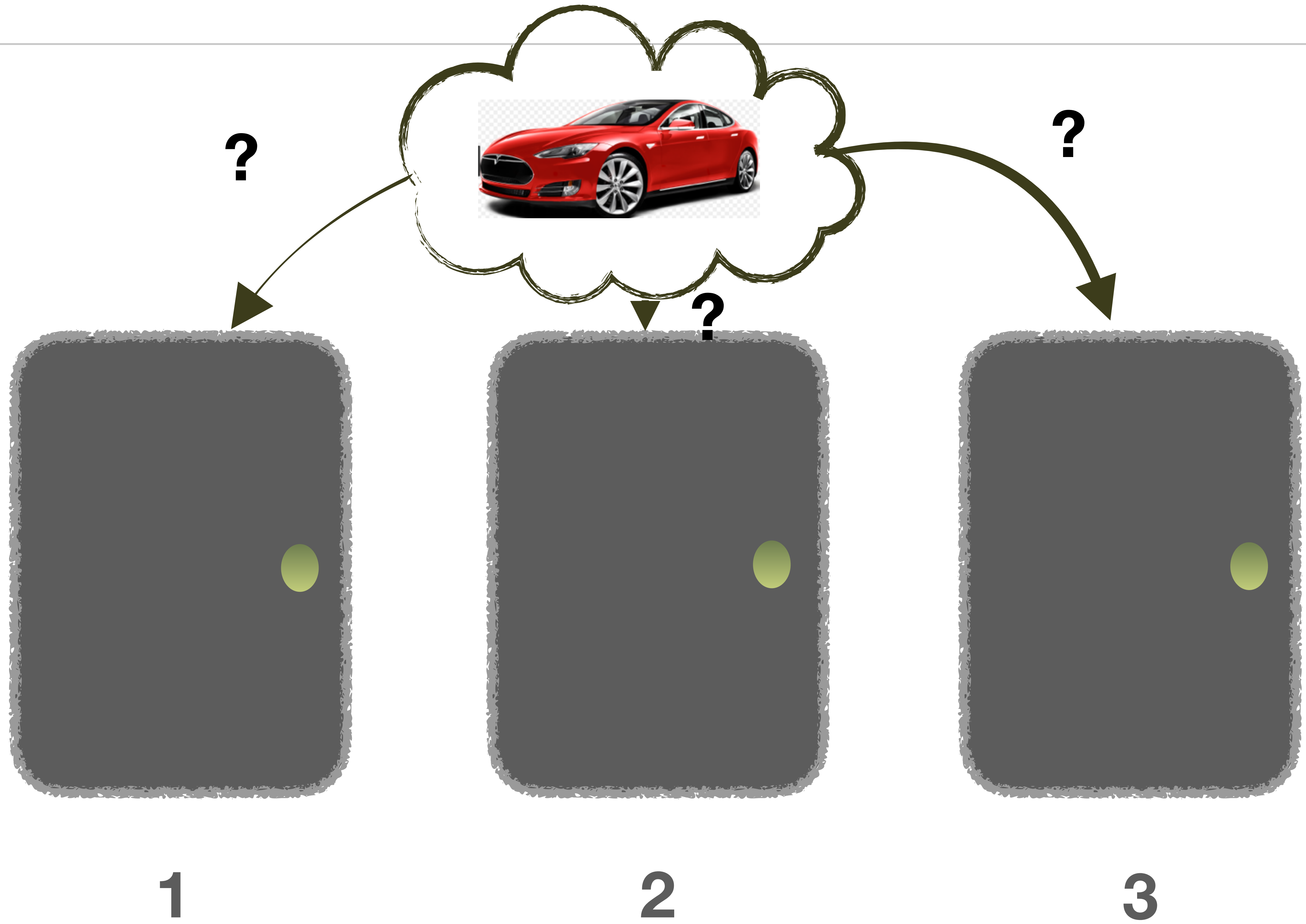
well-known brain teaser from “Let’s make a deal” TV game show (1980s).



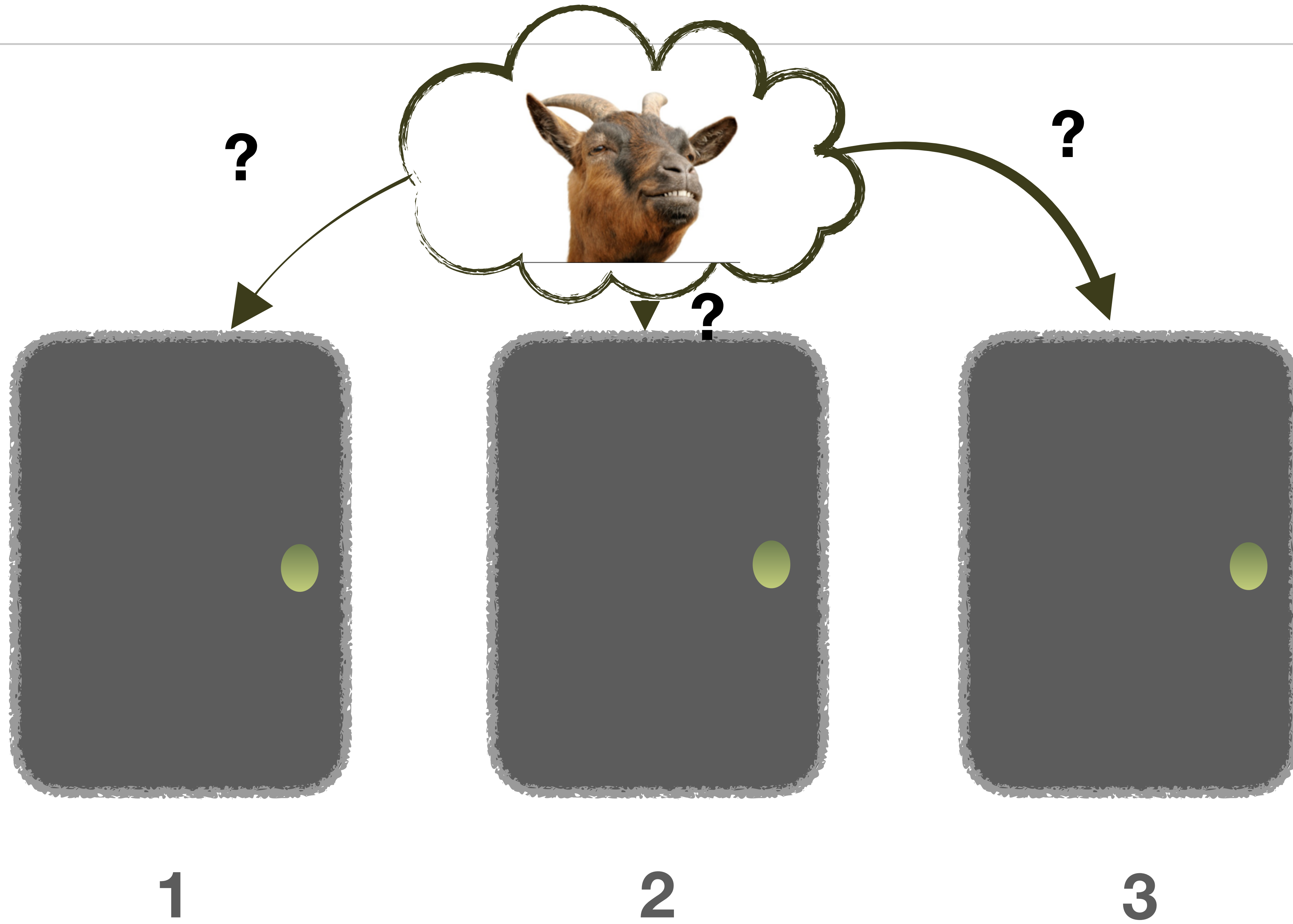
Monty Hall problem



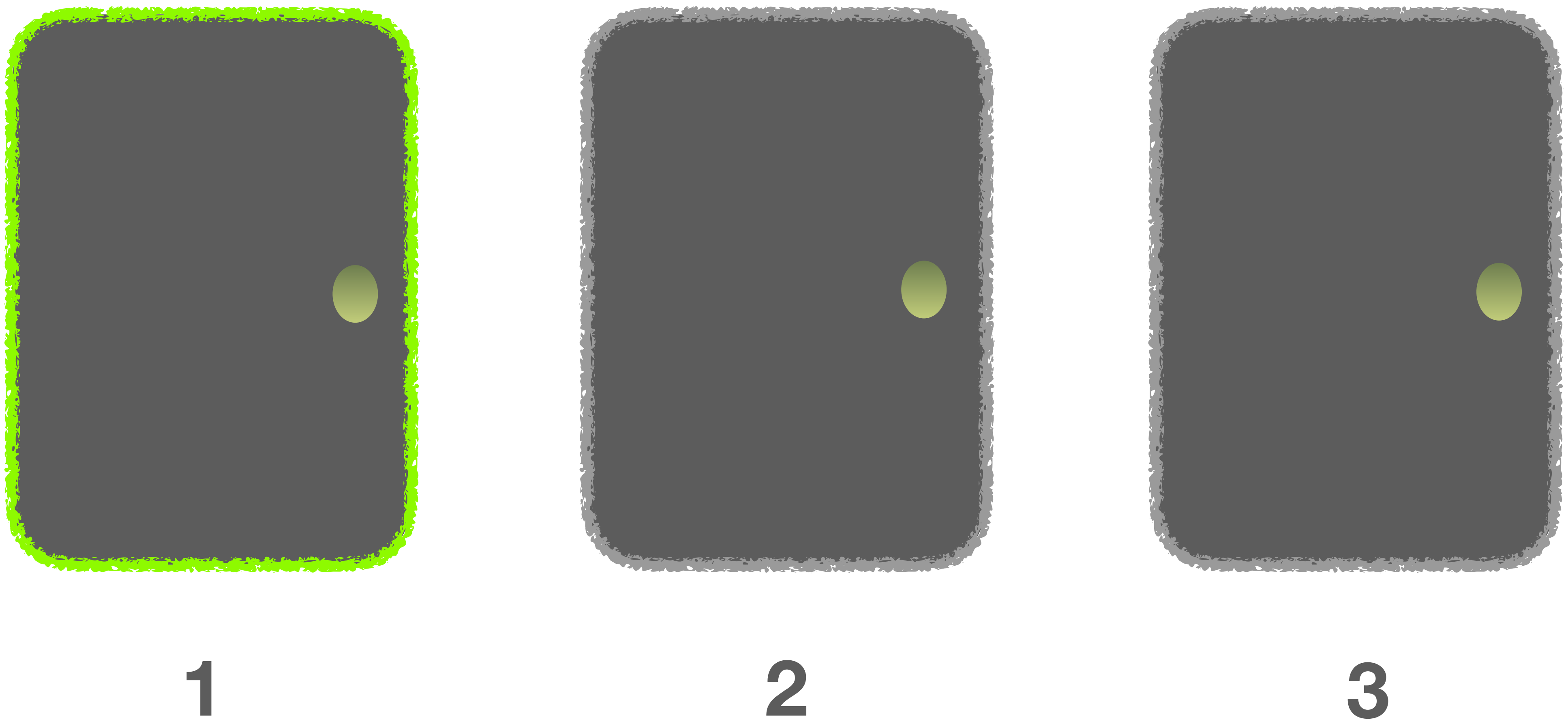
Monty Hall problem



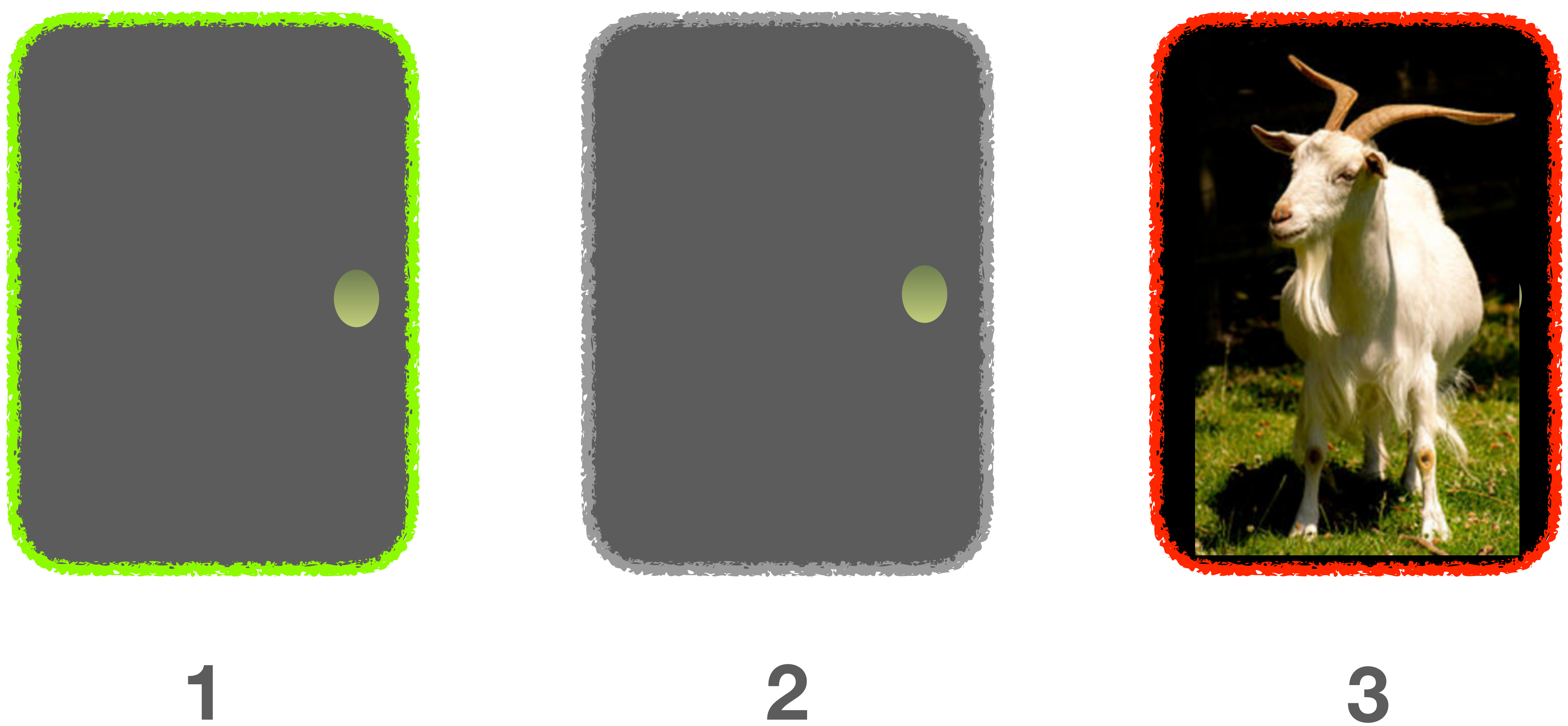
Monty Hall problem



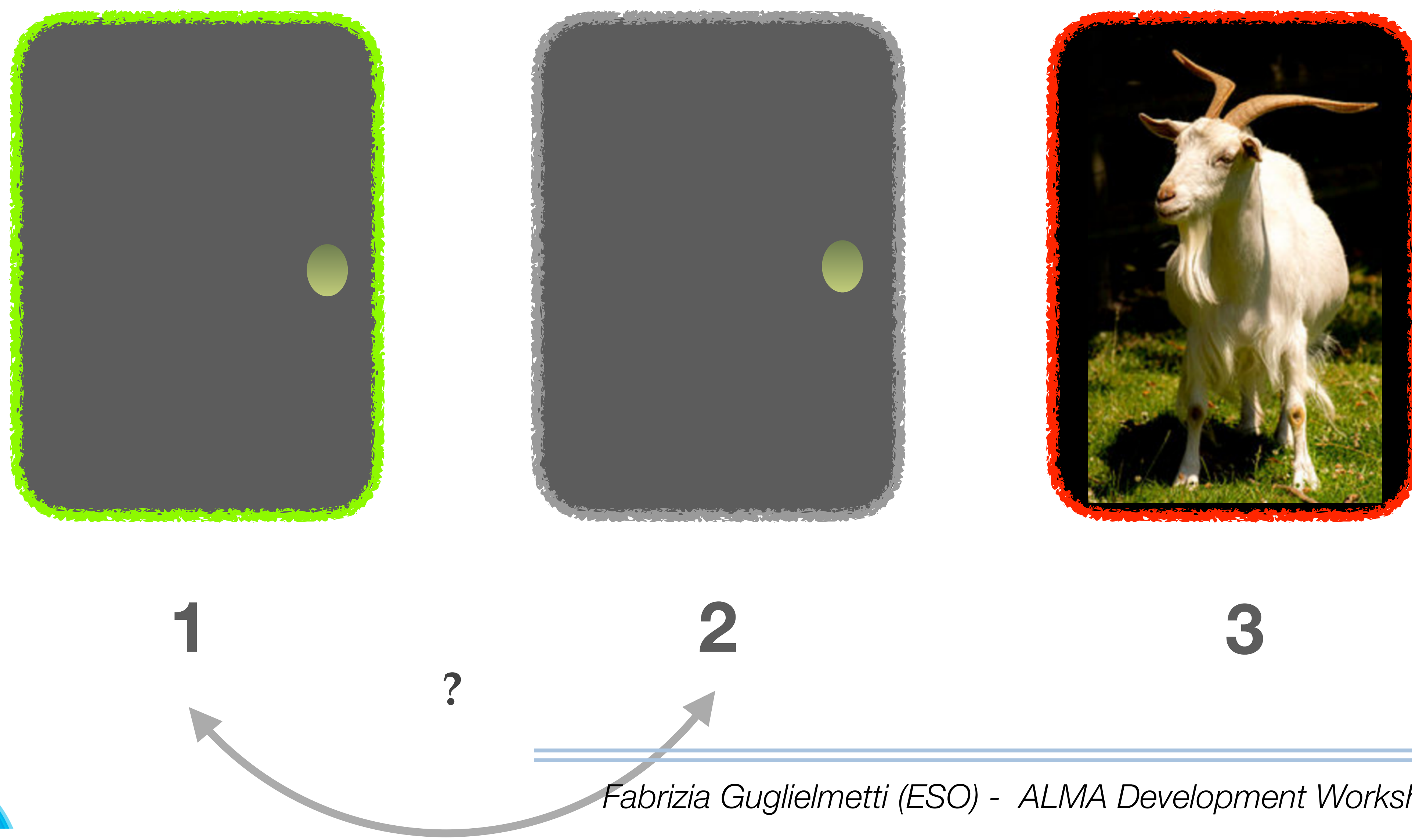
Monty Hall problem



Monty Hall problem



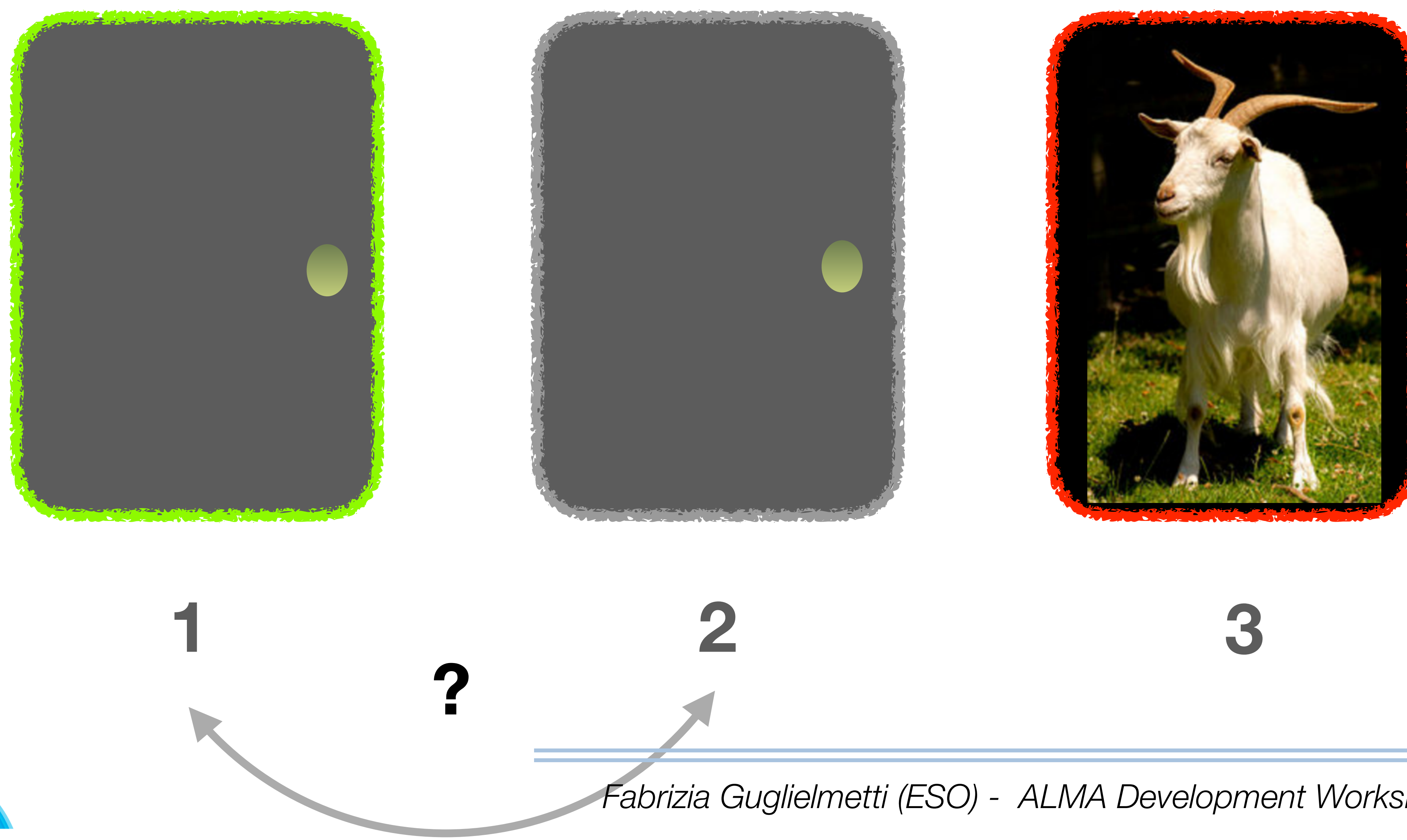
Monty Hall problem



Monty Hall problem

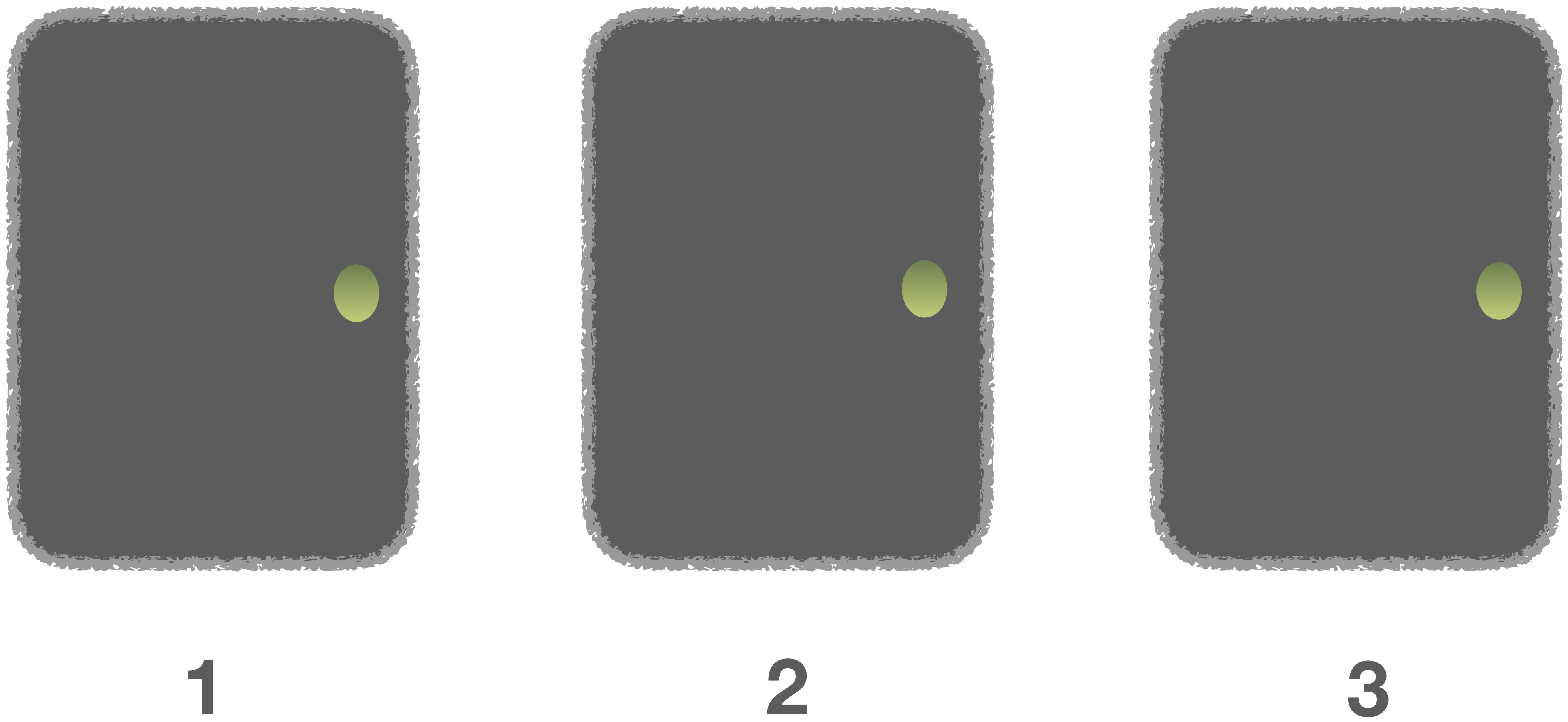
Should contestant:

- Stick with door 1?
- Switch to door 2?
- Does it make no difference?



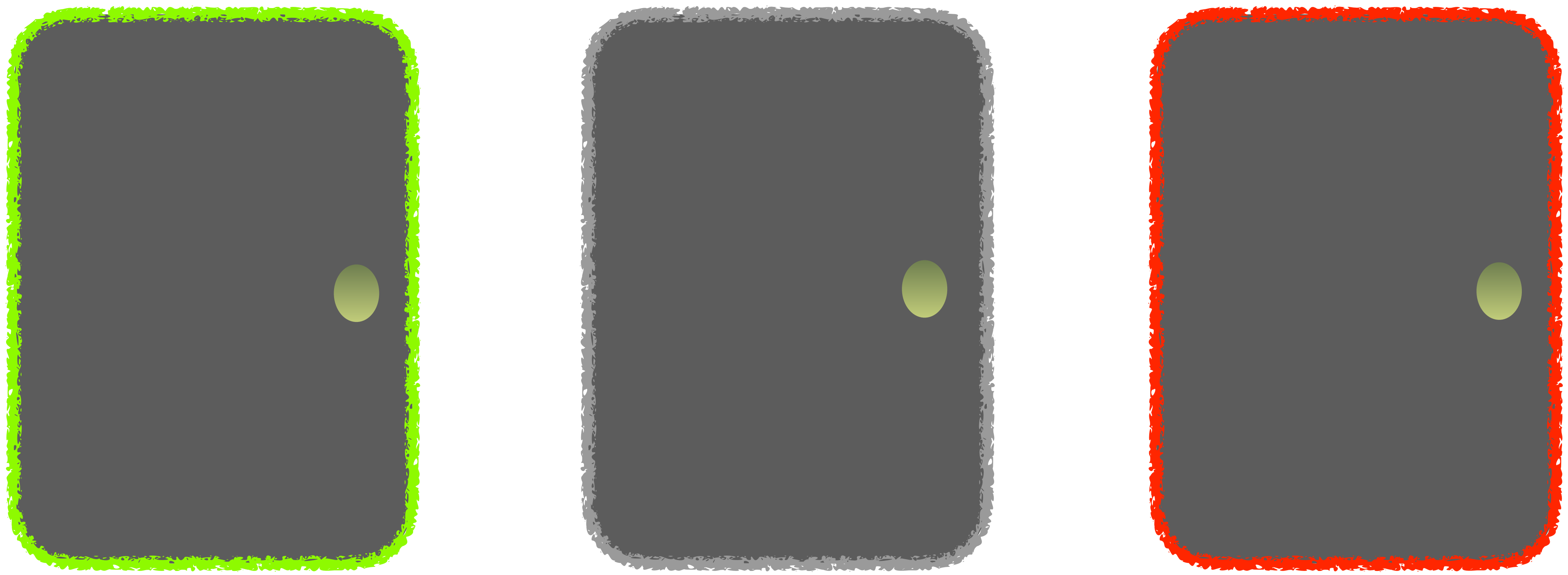
Monty Hall problem

$$P(\mathcal{H}_1) = P(\mathcal{H}_2) = P(\mathcal{H}_3) = \frac{1}{3}$$



Monty Hall problem

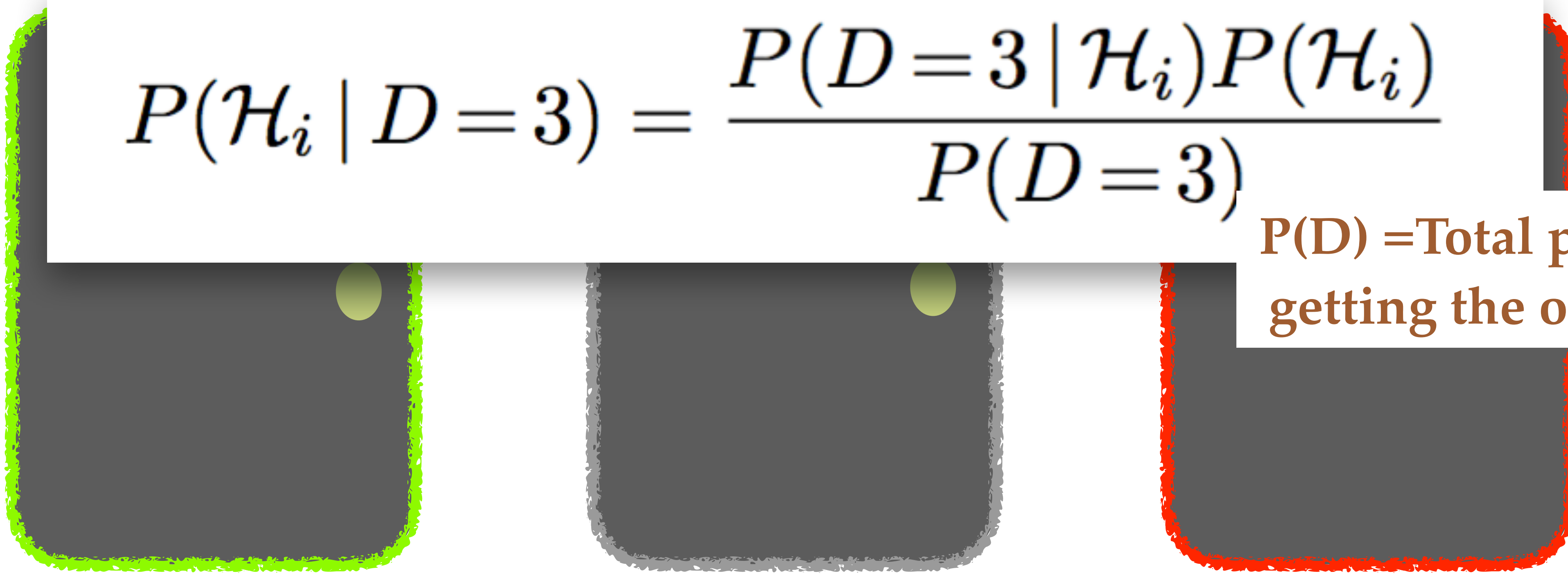
$$
\left|\begin{array}{l|l|l}
P(D=2\,|\,\mathcal{H}_1)=1/2 & P(D=2\,|\,\mathcal{H}_2)=0 & P(D=2\,|\,\mathcal{H}_3)=1 \\
P(D=3\,|\,\mathcal{H}_1)=1/2 & P(D=3\,|\,\mathcal{H}_2)=1 & P(D=3\,|\,\mathcal{H}_3)=0
\end{array}\right|
$$



1 **2** **3**

Monty Hall problem

$$
\left|\begin{array}{c} P(D=2\,|\,\mathcal{H}_1)={}^1\!/\!_2 \\ P(D=3\,|\,\mathcal{H}_1)={}^1\!/\!_2 \end{array}\right|\begin{array}{c} P(D=2\,|\,\mathcal{H}_2)=0 \\ P(D=3\,|\,\mathcal{H}_2)=1 \end{array}\left|\begin{array}{c} P(D=2\,|\,\mathcal{H}_3)=1 \\ P(D=3\,|\,\mathcal{H}_3)=0 \end{array}\right|
$$



$$
P(\mathcal{H}_i\,|\,D=3)=\frac{P(D=3\,|\,\mathcal{H}_i)P(\mathcal{H}_i)}{P(D=3)}
$$

P(D) =Total probability of getting the observed data

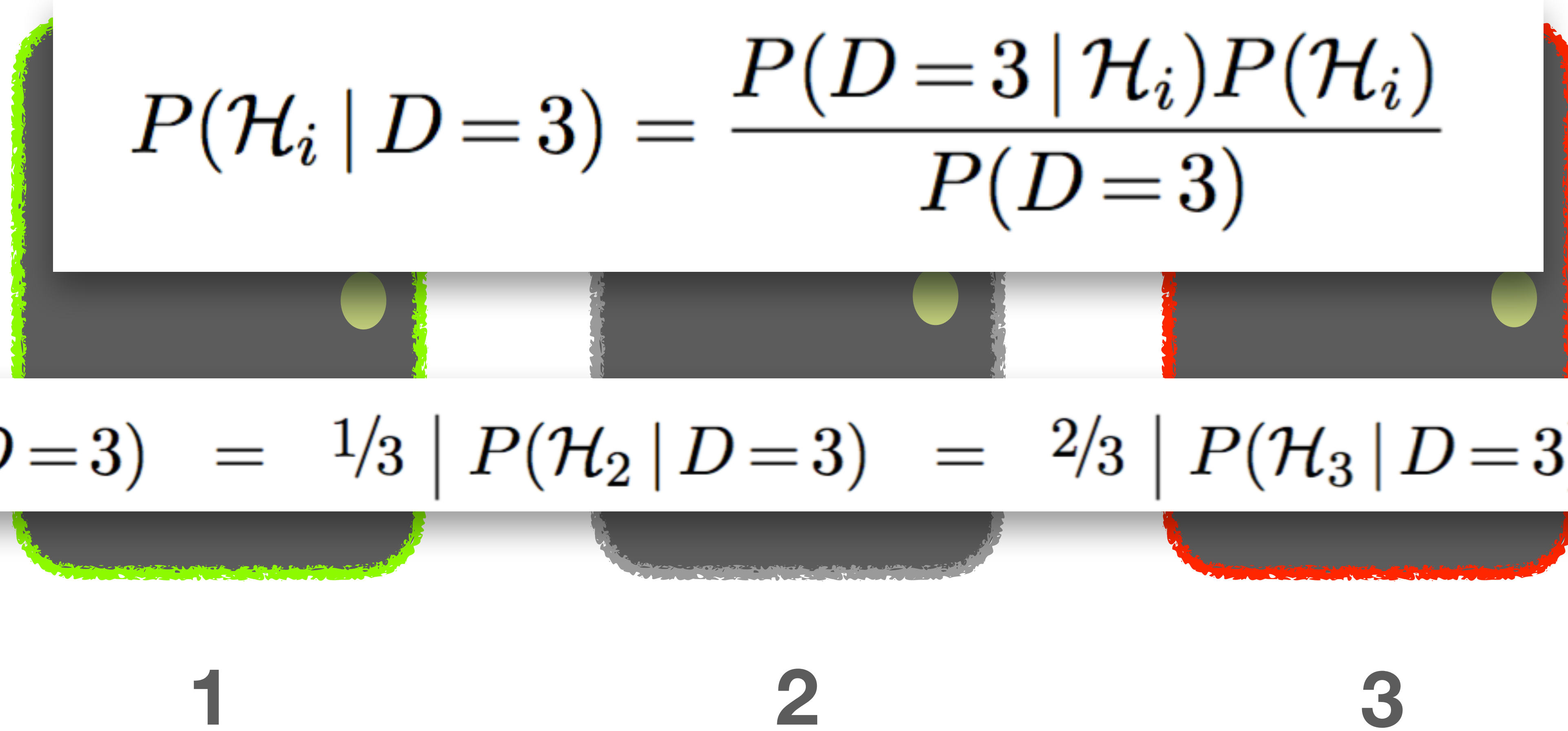
1 2 3

Monty Hall problem

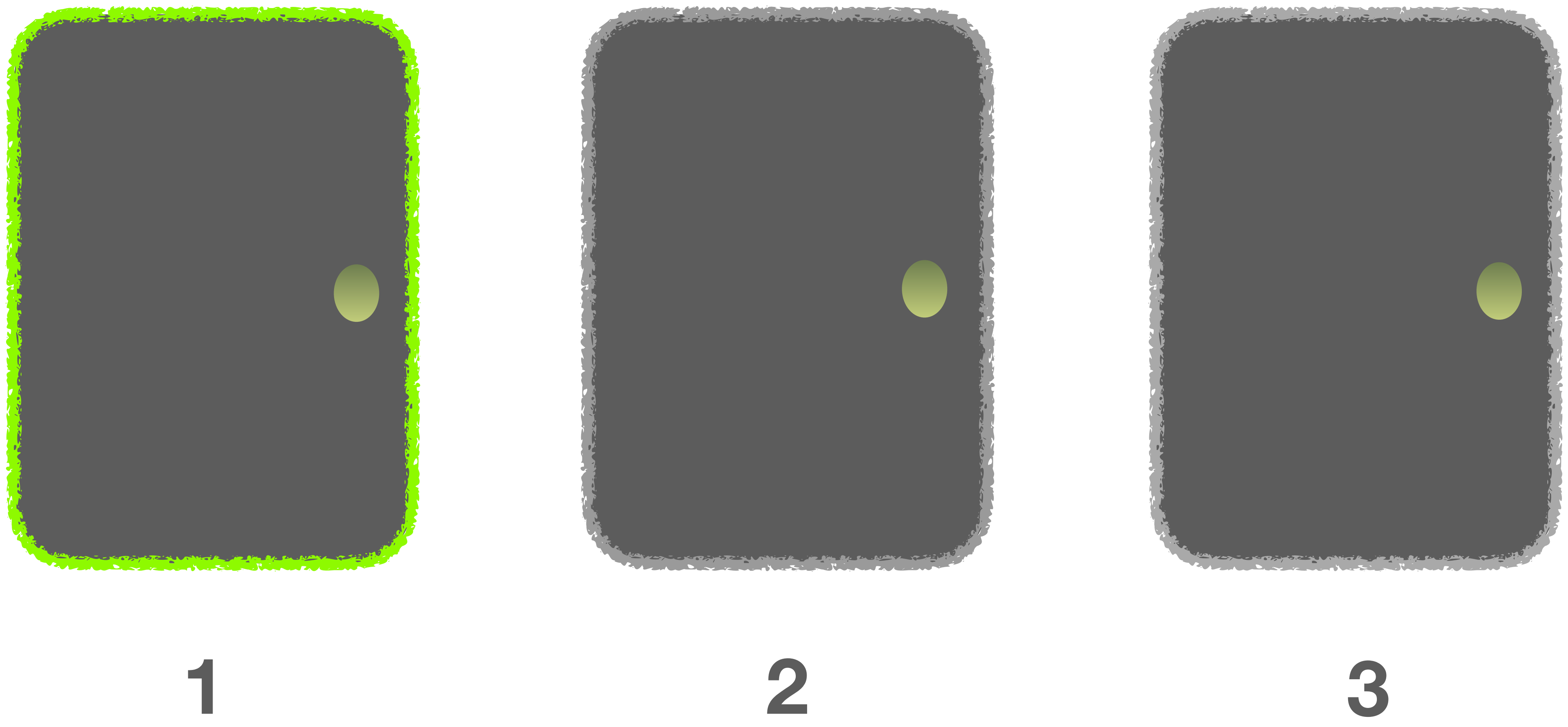
$$
\left| \begin{array}{c|c|c} P(D=2\,|\,\mathcal{H}_1) = 1/2 & P(D=2\,|\,\mathcal{H}_2) = 0 & P(D=2\,|\,\mathcal{H}_3) = 1 \\ P(D=3\,|\,\mathcal{H}_1) = 1/2 & P(D=3\,|\,\mathcal{H}_2) = 1 & P(D=3\,|\,\mathcal{H}_3) = 0 \end{array} \right|
$$

$$
P(\mathcal{H}_i\,|\,D=3) = \frac{P(D=3\,|\,\mathcal{H}_i)P(\mathcal{H}_i)}{P(D=3)}
$$

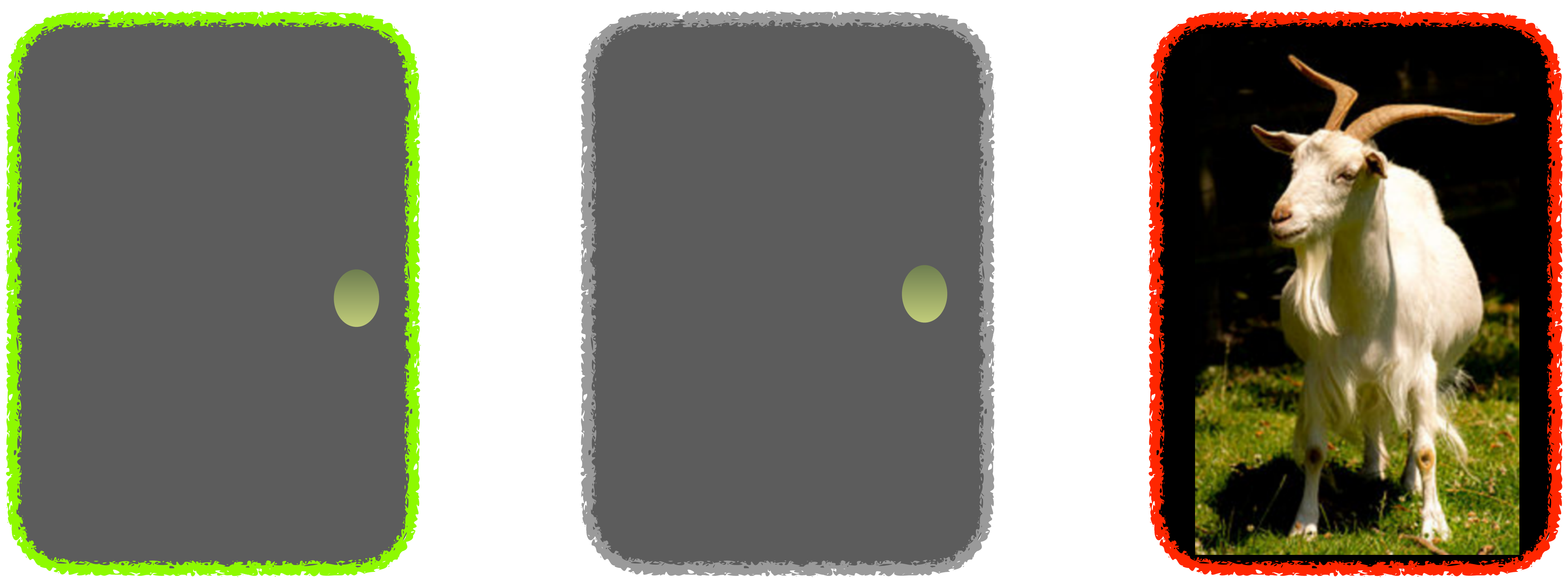
$$
\left| \begin{array}{c|c|c} P(\mathcal{H}_1\,|\,D=3) \;=\; 1/3 & P(\mathcal{H}_2\,|\,D=3) \;=\; 2/3 & P(\mathcal{H}_3\,|\,D=3) \;=\; 0. \end{array} \right|
$$



Monty Hall problem -*earthquake scenario*



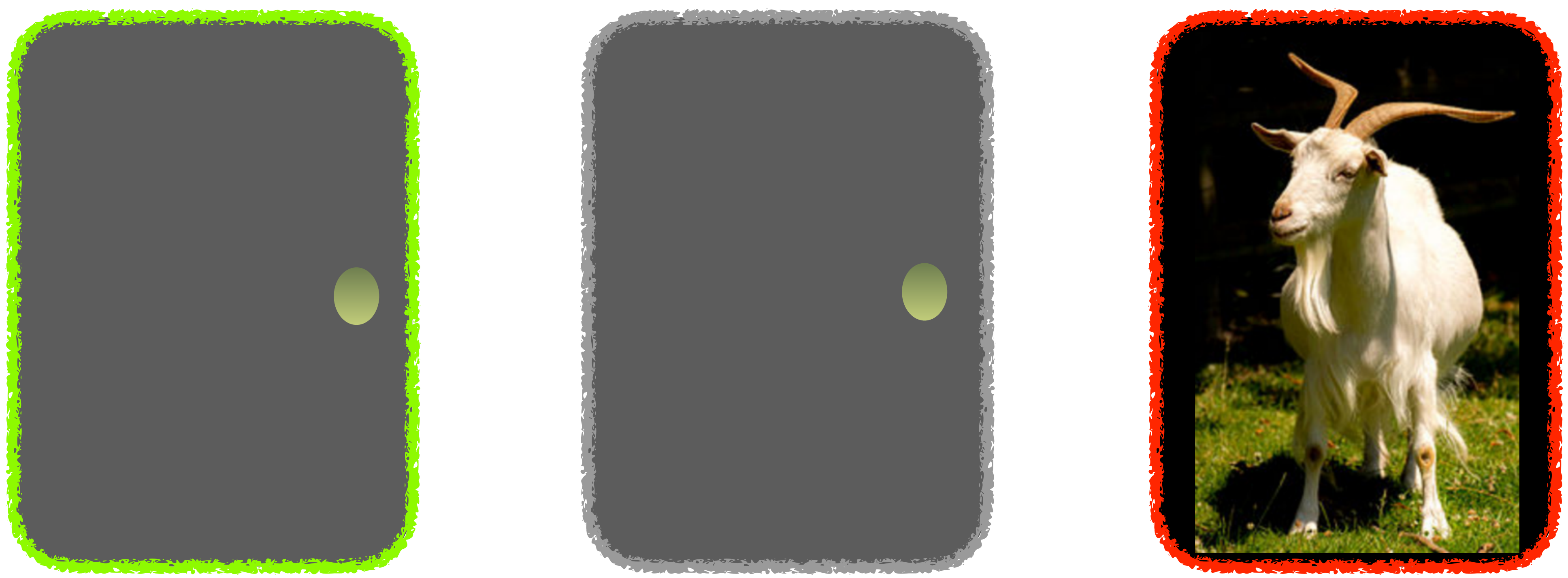
Monty Hall problem -*earthquake scenario*



1 2 3

Monty Hall problem *-earthquake scenario*

- Stick with door 1?
- Switch to door 2?
- Does it make no difference?



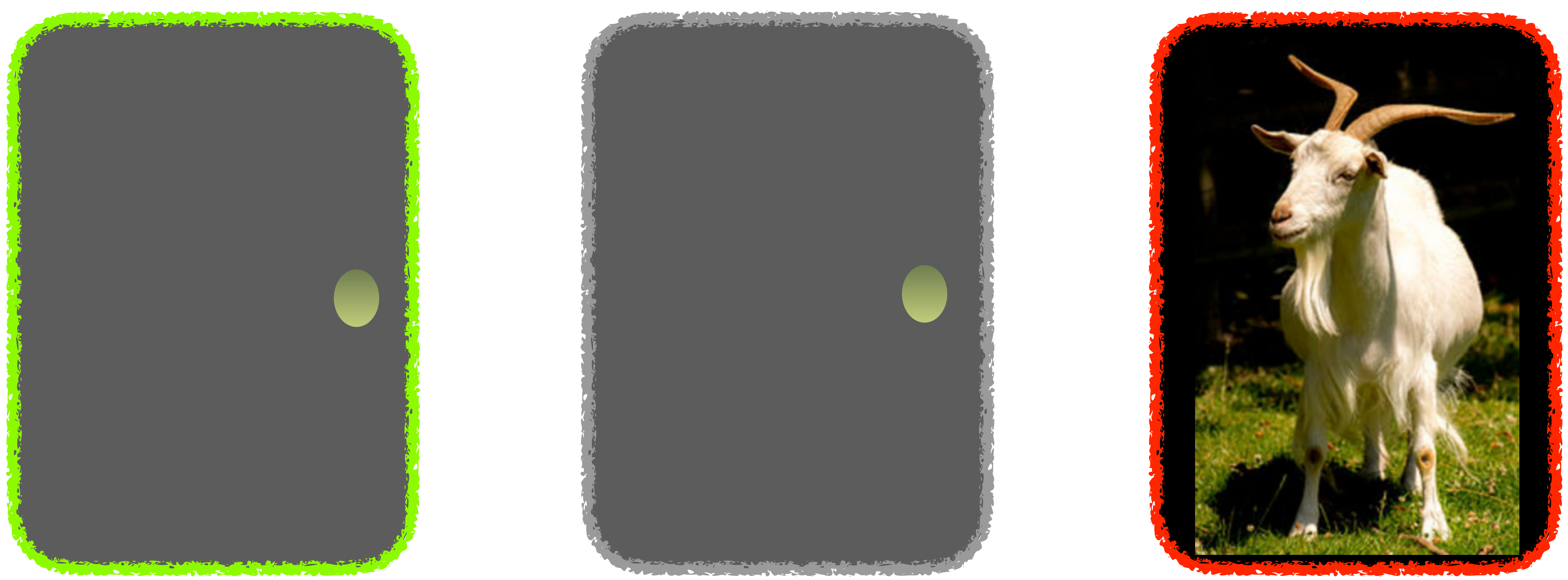
1 2 3

?

Assume the prize was placed randomly

Monty Hall problem -*earthquake scenario*

Nature of the data & probability of data differs from previous game



1 2 3

Monty Hall problem *-earthquake scenario*

Nature of the data & probability of data differs from previous game

(a) possible data outcomes are that any # of the doors might have opened

$$\mathbf{d} = (0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), \ldots, (1,1,1)$$

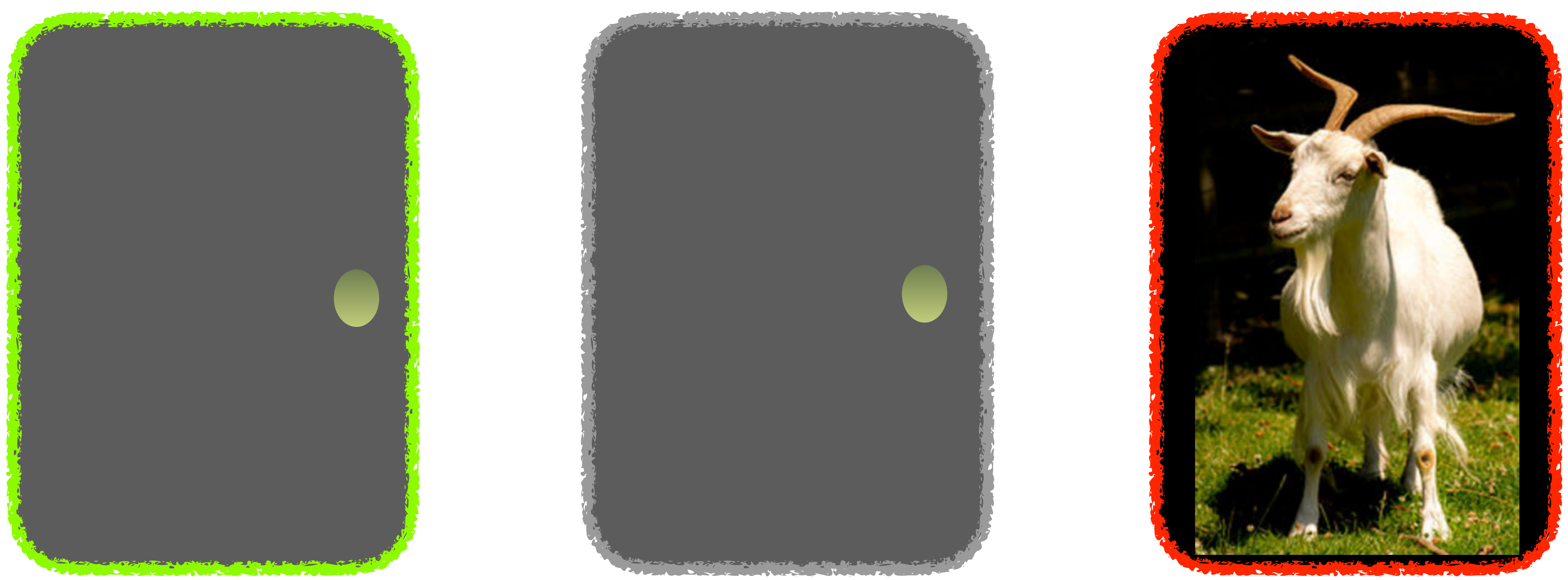
(b) Probabilities of these outcomes depend from any information about door latches and earthquake properties

1 2 3

Monty Hall problem *-earthquake scenario*

(c) We rely on the observation:

D: $\mathbf{d} = (0, 0, 1)$ + goat



1 2 3

$$P(D \mid \mathcal{H}_3) = 0$$

Monty Hall problem -*earthquake scenario*

(c) We rely on the observation:

D: $\mathbf{d} = (0, 0, 1)$ + goat

$$P(\mathcal{H}_1|D) = \frac{P(D|\mathcal{H}_1)(1/3)}{P(D)} = 1/2$$

$$P(\mathcal{H}_2|D) = \frac{P(D|\mathcal{H}_2)(1/3)}{P(D)} = 1/2$$

$$P(\mathcal{H}_3|D) = \frac{P(D|\mathcal{H}_3)(1/3)}{P(D)} = 0.$$

1 **2** **3**

$$P(D\,|\,\mathcal{H}_3) = 0$$