# Statistical inference of astrophysical point-sources

GABRIEL COLLIN

# What is a point source

Any object so small that it can be approximated as a mathematical point.

Thus, it is always located inside exactly one pixel.

However, instrumentation effects (PSF) means that it can contribute to more than one pixel. Eg:

- Diffraction spikes (seen in JWST image)
- Atmospheric diffraction
- Limited focusing (X-rays and gamma-rays)

Stars in foreground are point-sources.
Galaxies in background are not point-sources.

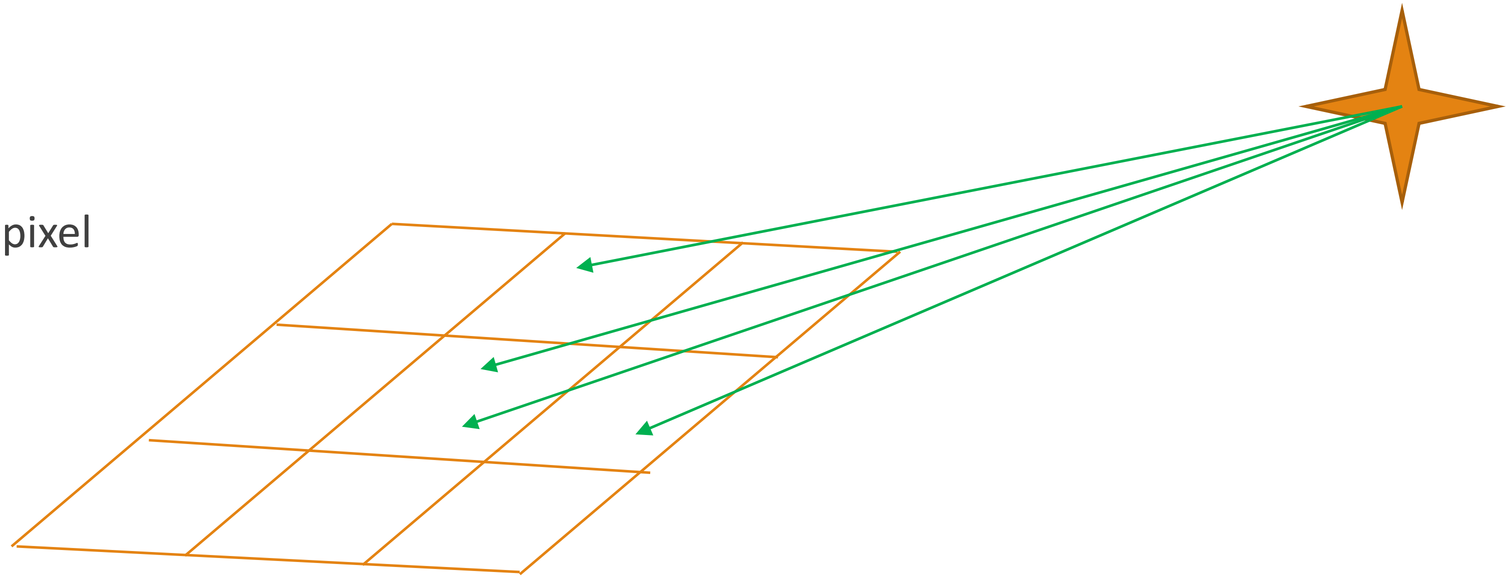
# Statistics of a point-source

A single point-source is usually treated as a Poisson distribution.

- If the flux is constant
  - (a usual approximation that we make)
- Then the arrival time of the gamma-rays is exponentially distributed.

For a pixelated detector

- There is Poisson distribution per pixel
- Weighted by
  - PSF
  - Effective area
  - Detection probability



# Point sources

Any group of point sources that have common properties is called a population.

The most familiar point source population is the stars in the night sky that are visible to the naked eye.

Here, stars are well separated.

- Images intuitively reflect the mathematical modelling of points.
- We can simply count all the stars and list their positions, in a database known as a *catalogue*.

# Crowded fields

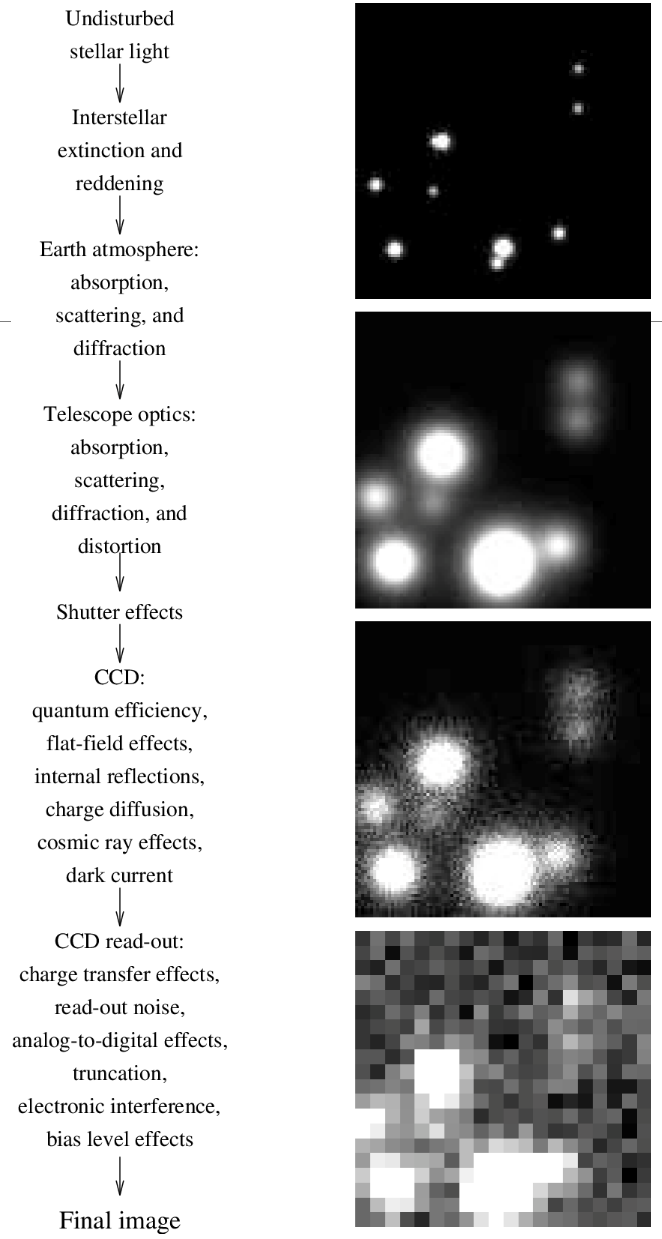
When many point sources are near each other, it can be difficult to distinguish them.

This is called a *crowded field*.

- It is entirely caused by experimental issues, such as the angular resolution of the instrument.

This kind of situation is becoming more common, as instruments push the limits.

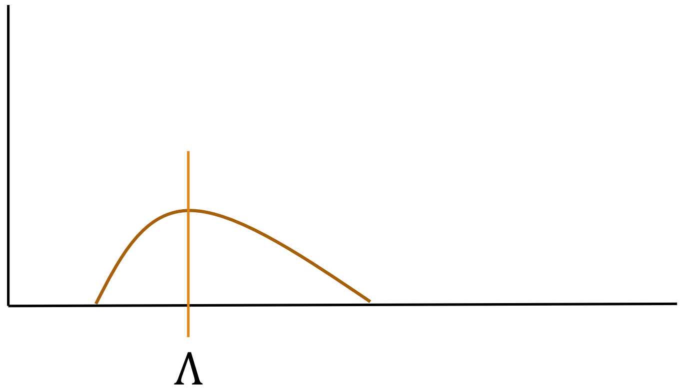
- Cheaper, smaller experiments.
- Experiments that look deeper into the universe.




Snel 1998

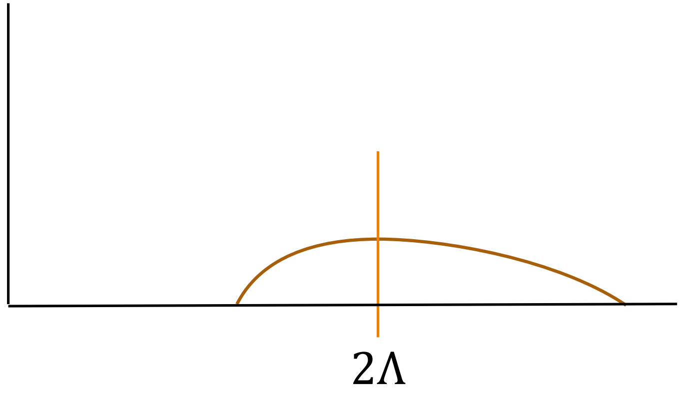
# Statistics of N sources

For one point-source, we get a Poisson distribution with mean $\Lambda$




For two point-sources, we get a Poisson distribution with mean $2\Lambda$

- (Assuming, for simplicity, the flux is the same for both)



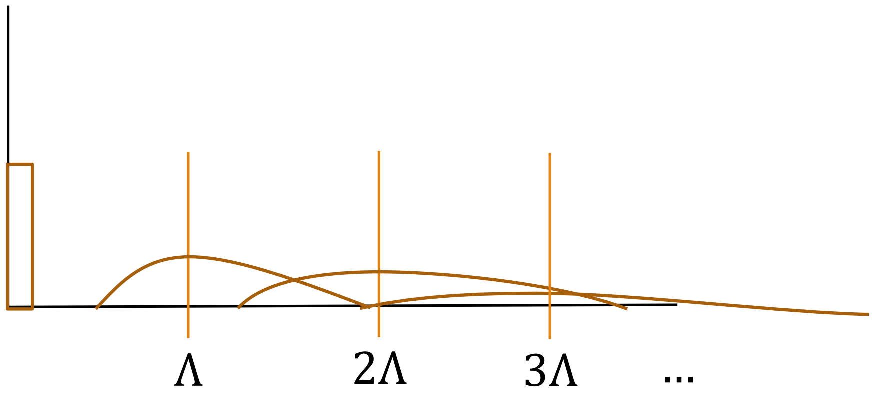
# Statistics of a population of sources

For zero point-sources, we get a Poisson distribution with a mean of zero.

- All probability mass is concentrated at $k = 0$

If we don't know the number of sources, we need to add all of these together to account for all possibilities.

- This is very different looking to a single Poisson distribution.
- There are many humps.
- Large gap between zero and the first hump.



# In numbers

Example provided by Tracy Slatyer

I expect 10 photons per pixel, in some region of the sky.

- What is my probability of finding 0 photons? 12 photons? 100 photons?

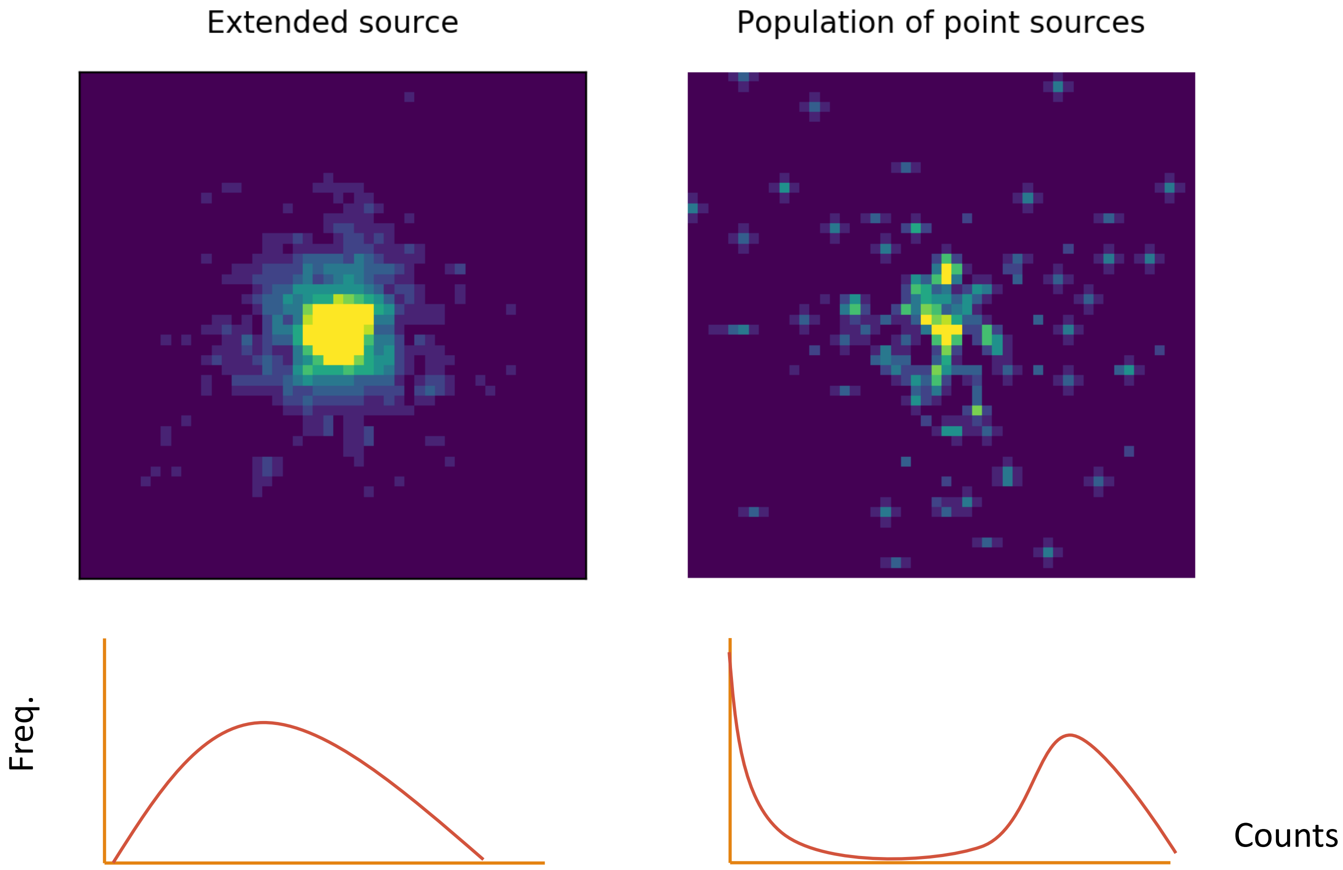
Case 1: Dark matter, Poissonian statistics

- P(12 photons) = $10^{12}\ e^{-10}/12!$ ~ 0.1
- P(0 photons) ~ $5 \times 10^{-5}$ ,
- P(100 photons) ~ $5 \times 10^{-63}$

Case 2: population of rare sources.

- Expect 100 photons/source, 0.1 sources/pixel - same expected # of photons
- P(0 photons) ~ 0.9,
- P(12 photons) ~ $0.1 \times 100^{12}\ e^{-100}/12!$ ~ $10^{-29}$ ,
- P(100 photons) ~ $4 \times 10^{-3}$
- (plus terms from multiple sources/pixel, which I am not including in this quick illustration)

# An intuitive view



# Compound Poisson Generators

Full details and alternative derivation at

arXiv > astro-ph > arXiv:2104.04529

**Astrophysics > Instrumentation and Methods for Astrophysics**

*[Submitted on 9 Apr 2021 (v1), last revised 27 Jun 2022 (this version, v2)]*

## A Compound Poisson Generator approach to Point-Source Inference in Astrophysics

Gabriel H. Collin, Nicholas L. Rodd, Tyler Erjavec, Kerstin Perez

# How to build a distribution

We need some way to combine the Poisson distributions for each choice of $N$ together.

- While accounting for the fact that $N$ is itself a random variable
  - (because we don't know how many point-sources there are).

There's an easy way to do this.

- Using **generating functions**.

The generating function for a distribution $p(k)$ is

$$G(z) = \sum_{k=0}^{\infty} p(k)\, z^k$$

$$p(k) = \frac{1}{k!}\frac{d^k}{dz^k} G(z)\bigg|_{z=0} = \sum_{n=0}^{\infty} p(n) \frac{1}{k!}\frac{d^k}{dz^k} z^n\bigg|_{z=0} = \sum_{n=0}^{\infty} p(n)\delta_{nk}$$

# The generating function

The full point-source population generating function is

$$G_{k_B}(z) = \exp\left[\underbrace{N}_{\text{Mean no. of sources}}\left(\int dF \int d\varepsilon\, e^{\varepsilon F(z-1)} \underbrace{\mu_B(\varepsilon)}_{\text{Effective effective-area distribution}}\, \underbrace{p(F)}_{\text{Flux distribution}} - 1\right)\right]$$

Mean no. of sources

Effective effective-area distribution

Flux distribution

(required to be broken power law)

All instrumental effects summarised in effective effective-area distribution

$$\mu_B(\varepsilon) = \int dx\, \underbrace{T(x)}_{\text{Spatial template}}\delta\left(\varepsilon - \underbrace{\kappa(x)}_{\text{Effective area}}\int_{\Omega_B} dy\, \underbrace{\eta(y)}_{\text{Detection prob.}}\underbrace{\phi(y|x)}_{\text{PSF}}\right)$$

Spatial template

Effective area

Detection prob.

PSF

Most often needs to be simulated

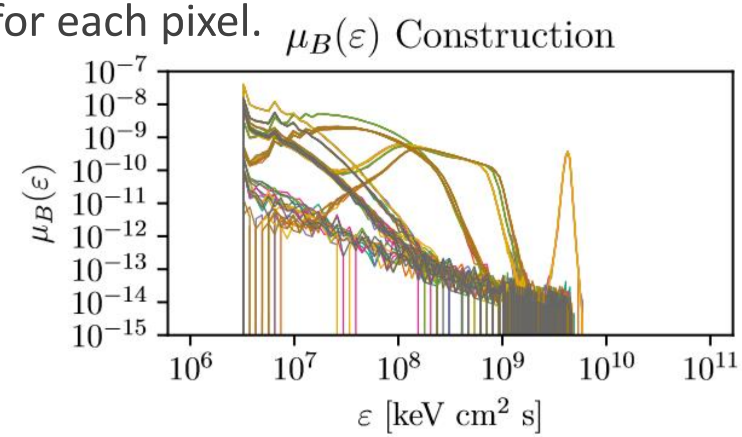
# Simulating the effective effective-area

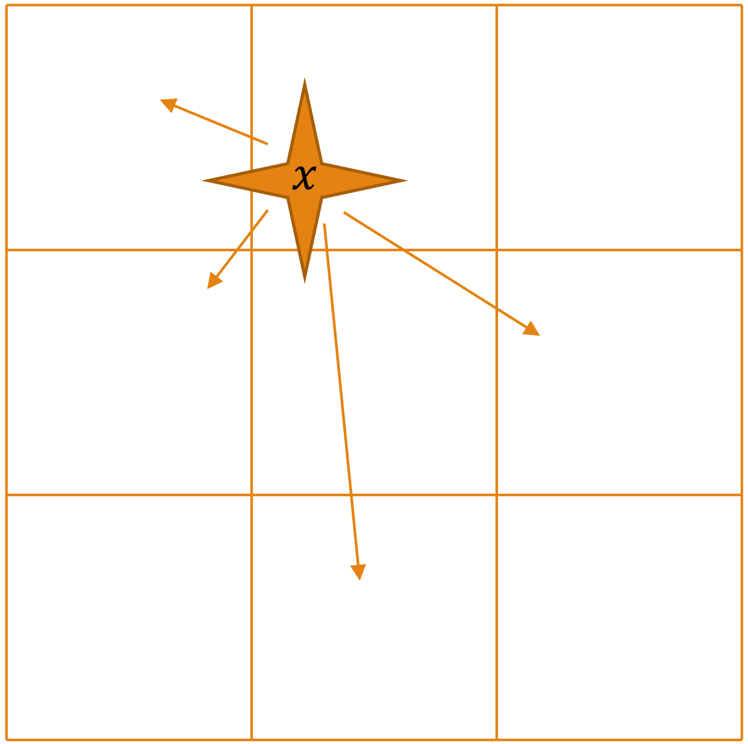
Algorithm:

- Draw a point-source location from $T(x)$
- Distribute effective area $\kappa(x)$ among all pixels
  - By drawing photons according to the PSF $\phi(y|x)$
  - weighted by the detection probability $\eta(y)$
- Repeat multiple times.

Now each pixel has a list of effective areas.

- Histogram this list for each pixel.
- This forms $\mu_B(x)$



# The total likelihood

Straightforward to get probability over pixels.

- Computationally unfeasible to account for correlations.

Instead

- Get probability for one pixel.
- Take product over all pixels.

$$p(\{n_i\}) = \prod p_i(n_i)$$

- Mean-field approximation
- Does not take into account correlations between pixels.

| | | |
|---|---|---|
| $p_0(n)$ | $p_1(n)$ | $p_2(n)$ |
| $p_3(n)$ | $p_4(n)$ | $p_5(n)$ |
| $p_6(n)$ | $p_7(n)$ | $p_8(n)$ |

# Application to galactic gamma-ray excess

Excess of gamma-rays at the galactic center

- Could be dark matter decay
- Could just be a population of previously unknown point-sources

Similar existing method called

Non-Poissonian Template Fitting (NP

Existing analysis prefers point source

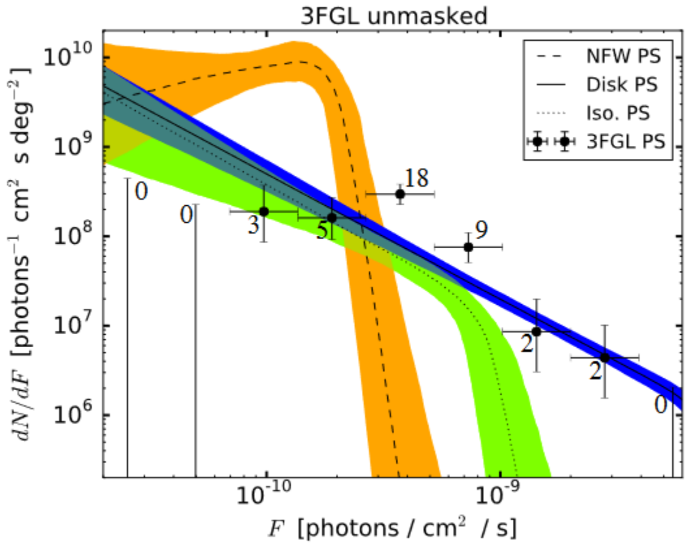
arXiv > astro-ph > arXiv:1506.05124

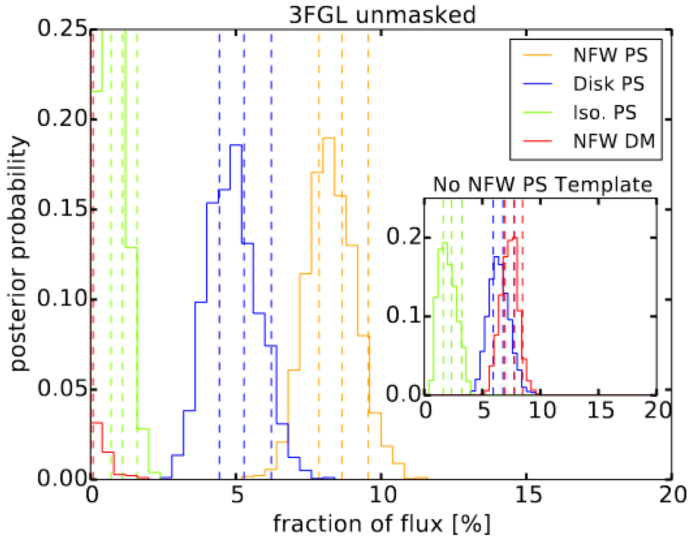
**Astrophysics > High Energy Astrophysical Phenomena**

*[Submitted on 16 Jun 2015 (v1), last revised 3 Feb 2016 (this version, v3)]*

**Evidence for Unresolved Gamma-Ray Point Sources in the Inner Galaxy**

Samuel K. Lee, Mariangela Lisanti, Benjamin R. Safdi, Tracy R. Slatyer, Wei Xue



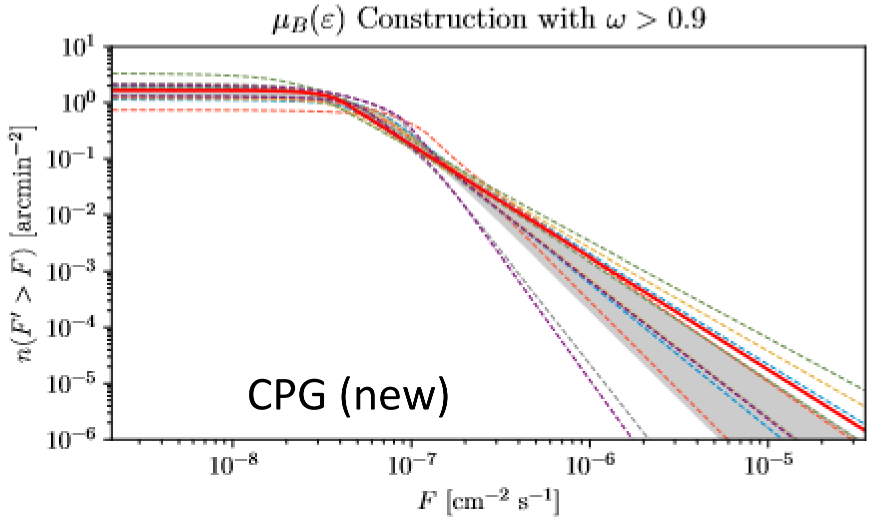



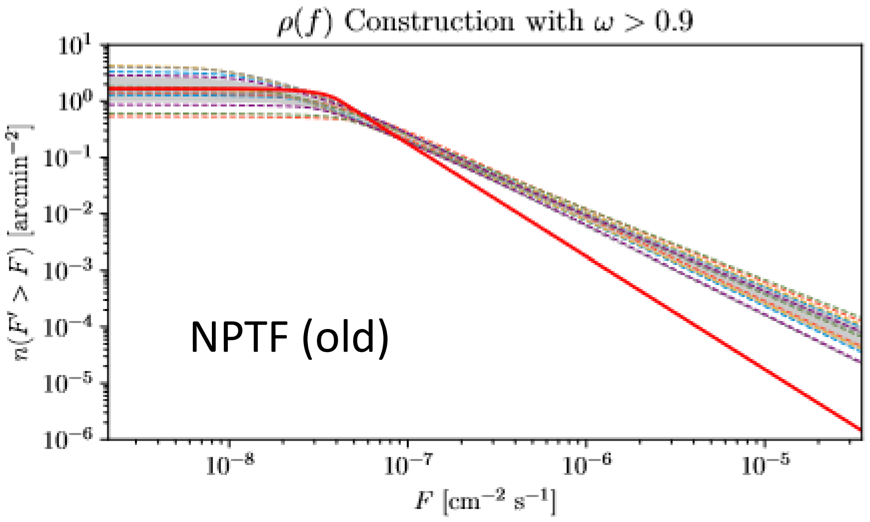
# Not the end of the story

Since then, we found that NPTF has significant problems.

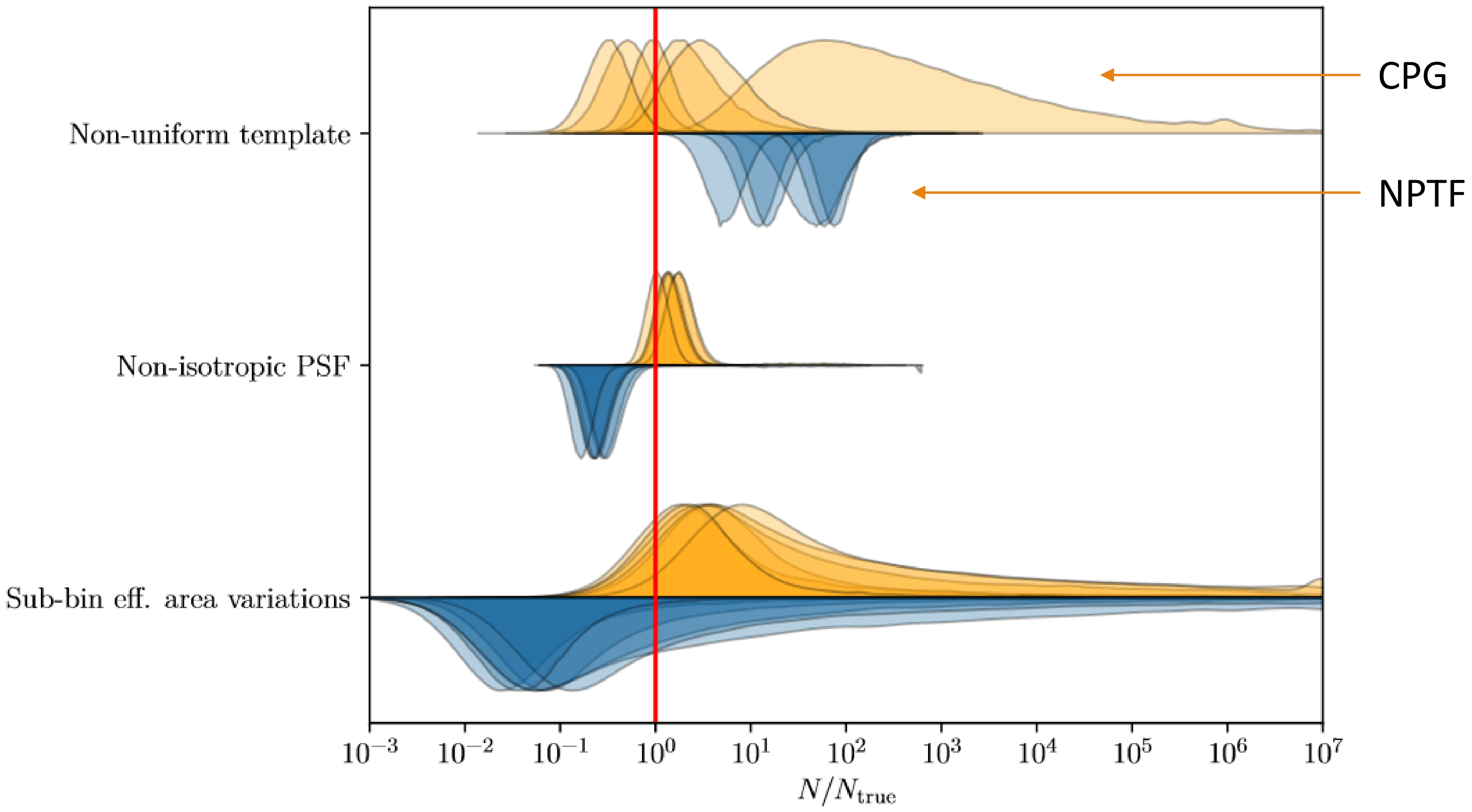
- Many instrumental effects can throw off the NPTF method.

- In X-ray astronomy these problems are most pronounced



# Instrumental effects, in detail

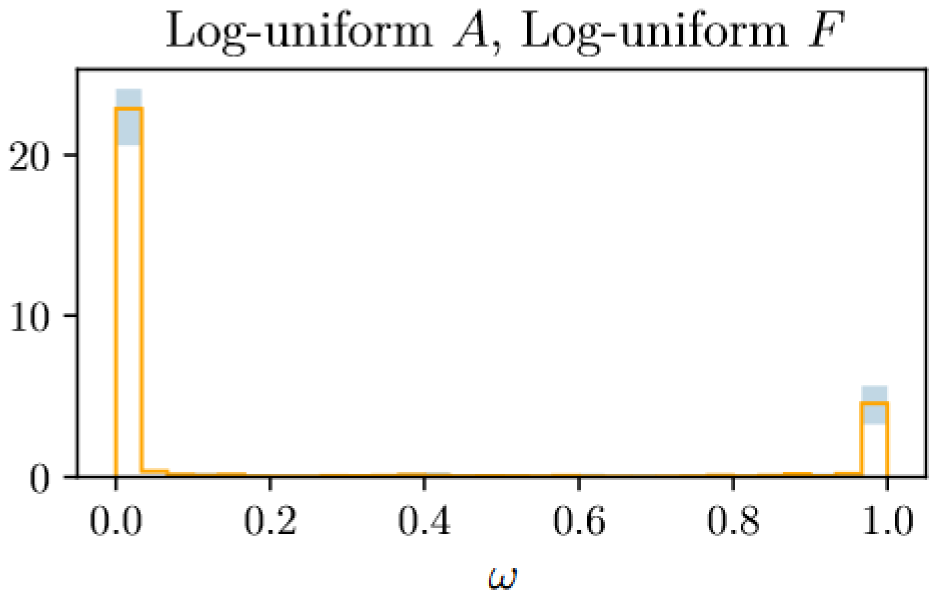


# Priors also a concern

Priors specified in terms of quantities that are not qualitatively relevant.

- Gives unexpected priors in relevant quantities!
- Always try to specify priors in quantities that you plan to plot

$$\frac{\mathrm{d}N}{\mathrm{d}F} = A \begin{cases} \left[\prod_{j=2}^{m} \left(\frac{F_{b(j+1)}}{F_{b(j)}}\right)^{-n_j}\right] \left(\frac{F}{F_{b(m)}}\right)^{-n_m} & F \in [0, F_{b(m)}] \\ \vdots & \\ \left[\prod_{j=2}^{i} \left(\frac{F_{b(j+1)}}{F_{b(j)}}\right)^{-n_j}\right] \left(\frac{F}{F_{b(i)}}\right)^{-n_i} & F \in (F_{b(i+1)}, F_{b(i)}] \\ \vdots & \\ \left(\frac{F}{F_{b(2)}}\right)^{-n_2} & F \in (F_{b(3)}, F_{b(2)}] \\ \left(\frac{F}{F_{b(2)}}\right)^{-n_1} & F > F_{b(2)} \end{cases} .$$




Fraction of flux assigned to point-sources over diffuse (dark-matter like)

# Also not the only method

Probabilistic cataloguing is a different approach

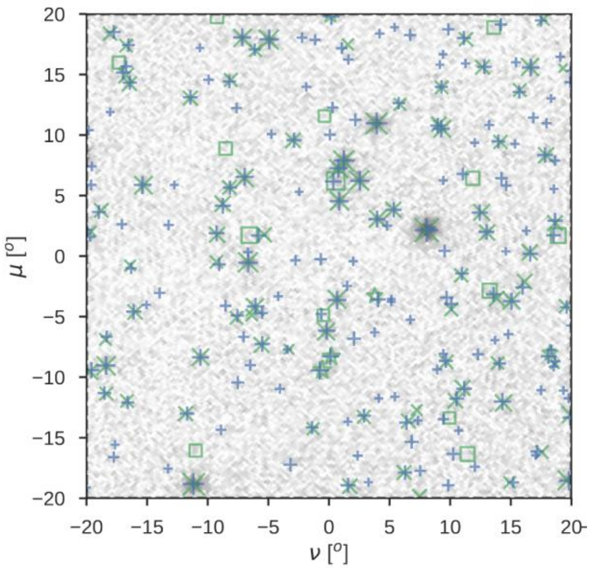
- Requires less assumptions.
- Potentially more sensitive.
- Much more computationally demanding.

Directly estimates the locations of every potential point-source.

- Means many thousands of parameters that need to be estimated!
- Requires advanced trans-dimensional statistical inference method.

**arXiv:1607.04637**

Simulation
Green marks: True point-source locations.
Blue marks: One sample from the posterior.




But, might be a better choice if expected no. of sources is low.