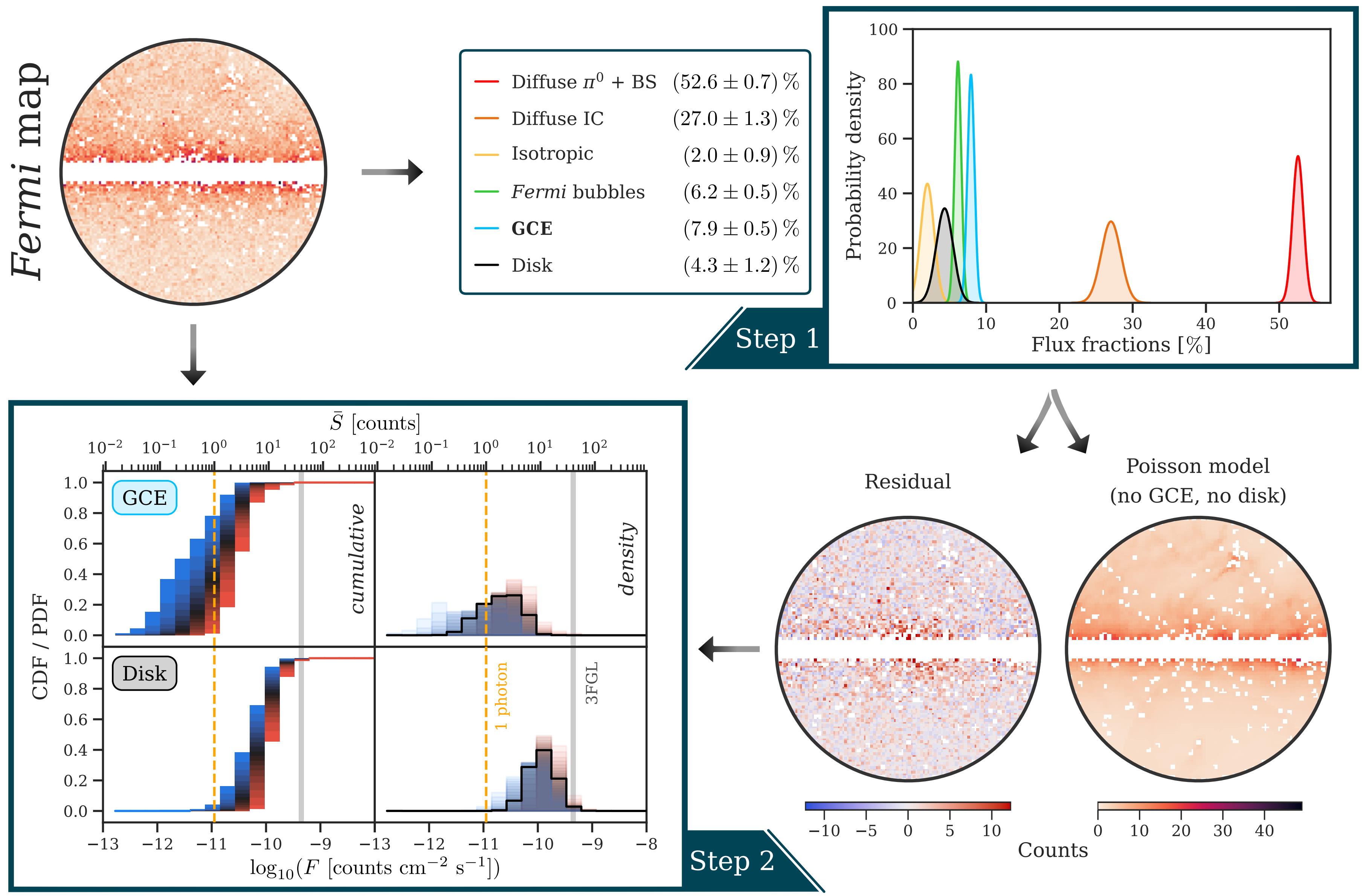
A Machine Learning Based Approach to the Galactic Center Excess

NICK RODD | PANIC 2021 | 5 SEPTEMBER 2021

2107.09070 (under review) w/ List, Lewis

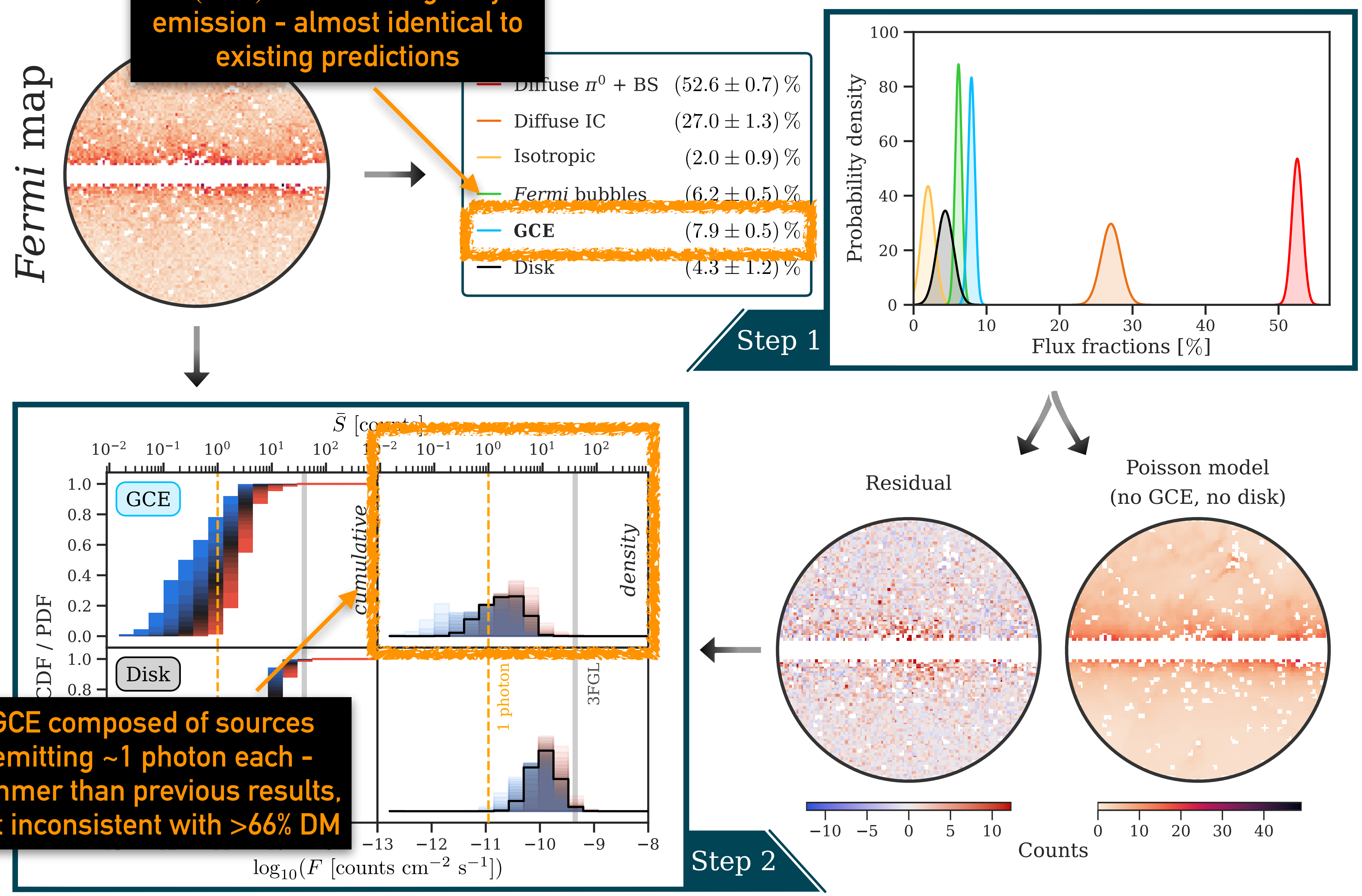
2006.12504 (PRL 2020) w/ List, Lewis, Bhat

Headline Results



[List, NLR, Lewis 2021]

Headline Results

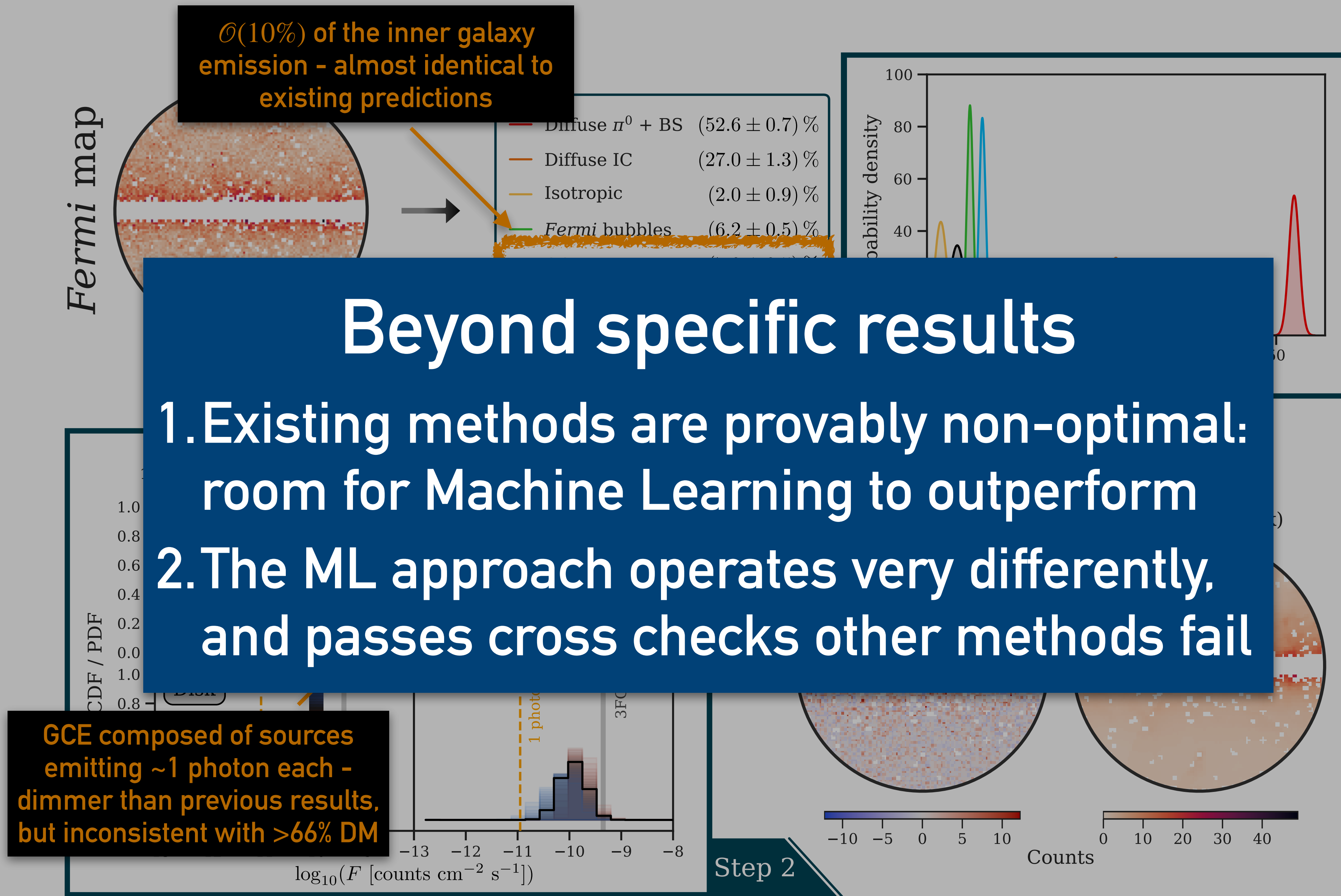


$\mathcal{O}(10\%)$ of the inner galaxy emission - almost identical to existing predictions

GCE composed of sources emitting ~1 photon each - dimmer than previous results, but inconsistent with >66% DM

[List, NLR, Lewis 2021]

Headline Results



Beyond specific results

1. Existing methods are provably non-optimal: room for Machine Learning to outperform
2. The ML approach operates very differently, and passes cross checks other methods fail

[List, NLR, Lewis 2021]

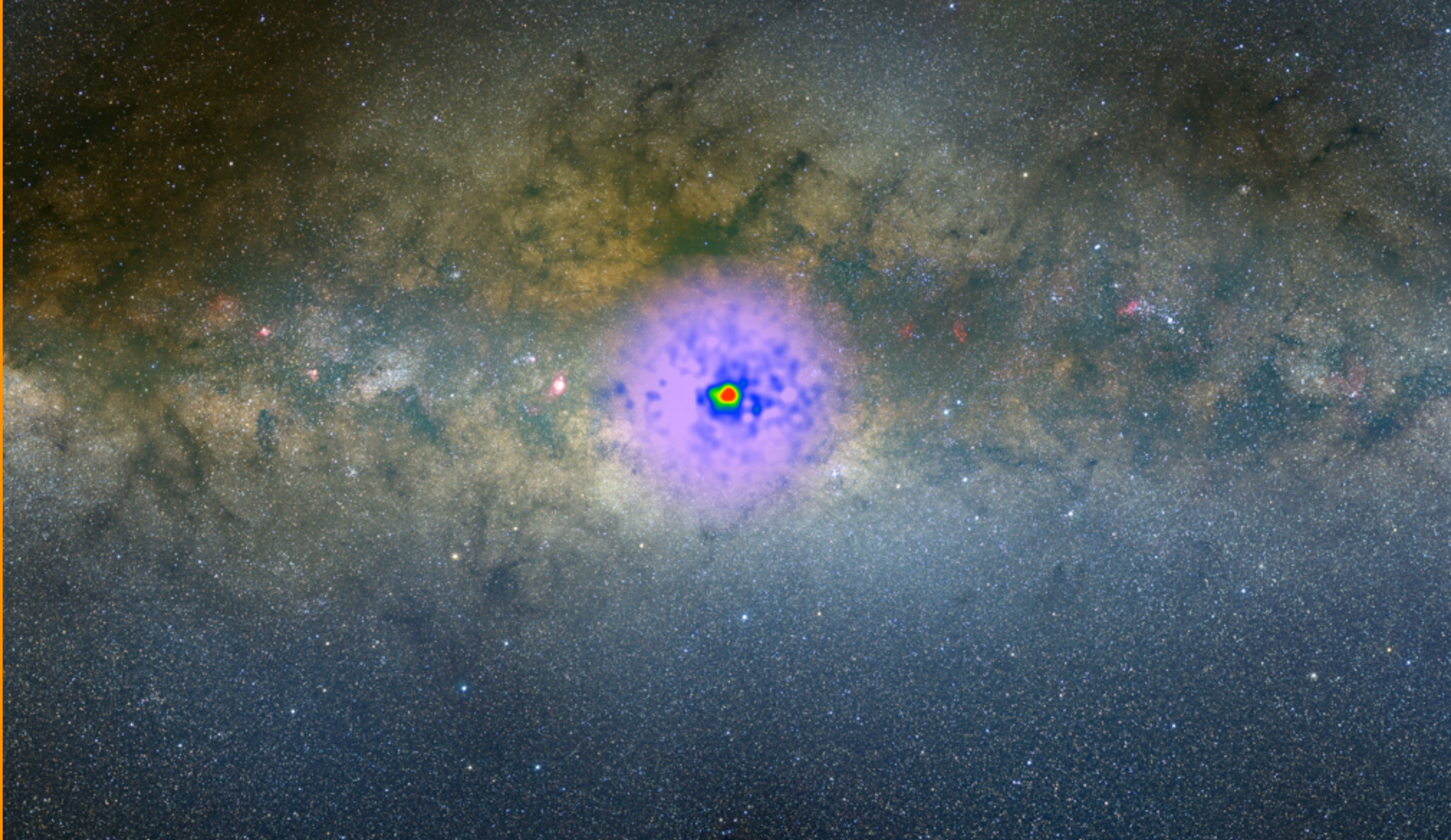
Outline

1. The GCE: dark matter or millisecond pulsars?

2. Likelihood approaches, and why ML can improve on them

3. Our convolutional neural network approach

The Galactic Center Excess

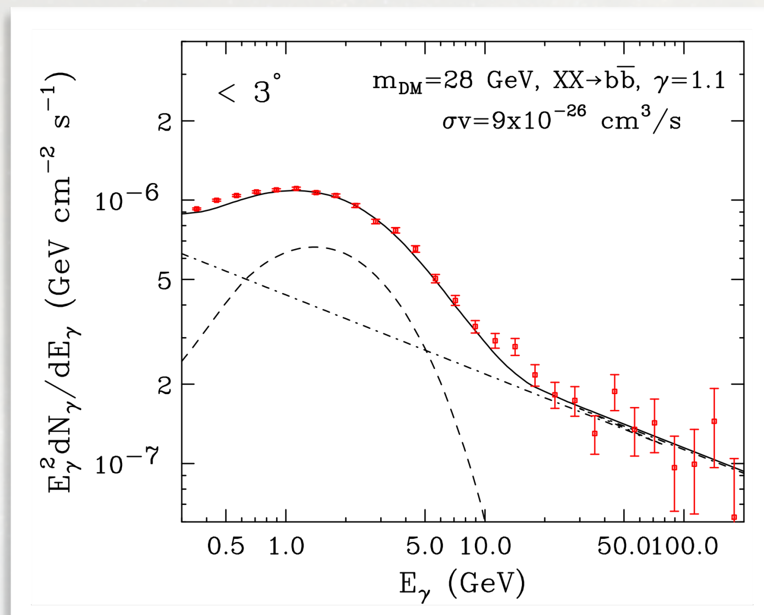


Very incomplete summary missing many important developments. For more see recent reviews in: [Murgia 2020] or [Leane 2020]

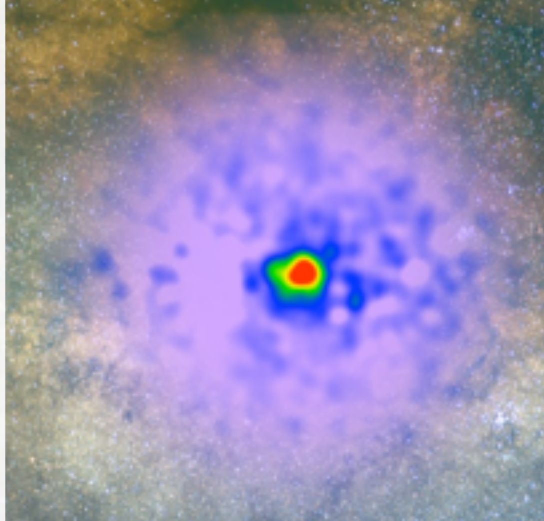
The Galactic Center Excess

Dark Matter

[Goodenough, Hooper 2009]



Exhibited many expected properties of DM, e.g. [Daylan, NLR+ 2014]

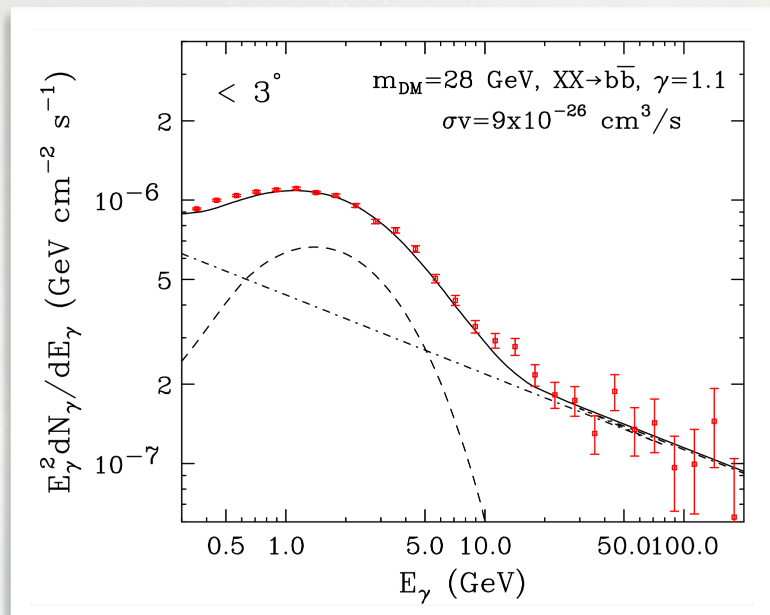


Very incomplete summary missing many important developments. For more see recent reviews in: [Murgia 2020] or [Leane 2020]

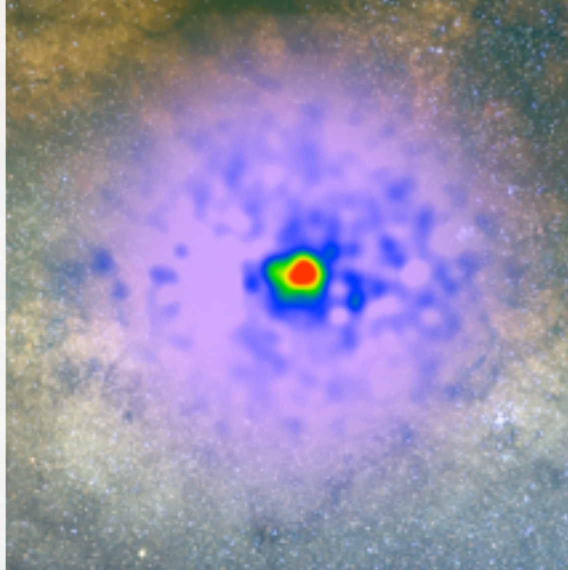
The Galactic Center Excess

Dark Matter

[Goodenough, Hooper 2009]

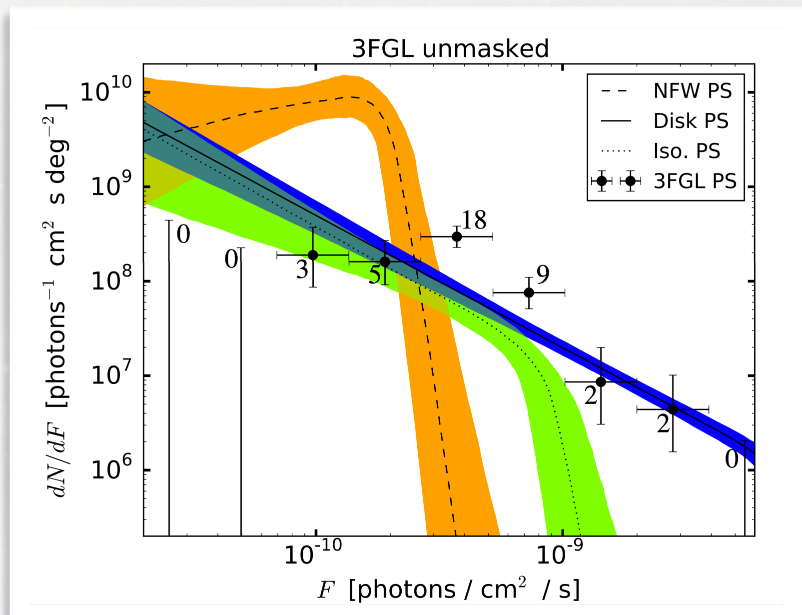


Exhibited many expected properties of DM, e.g. [Daylan, NLR+ 2014]



Millisecond Pulsars

[Lee+ 2016]



Data preferred clumpy point-source statistics rather than smoother DM

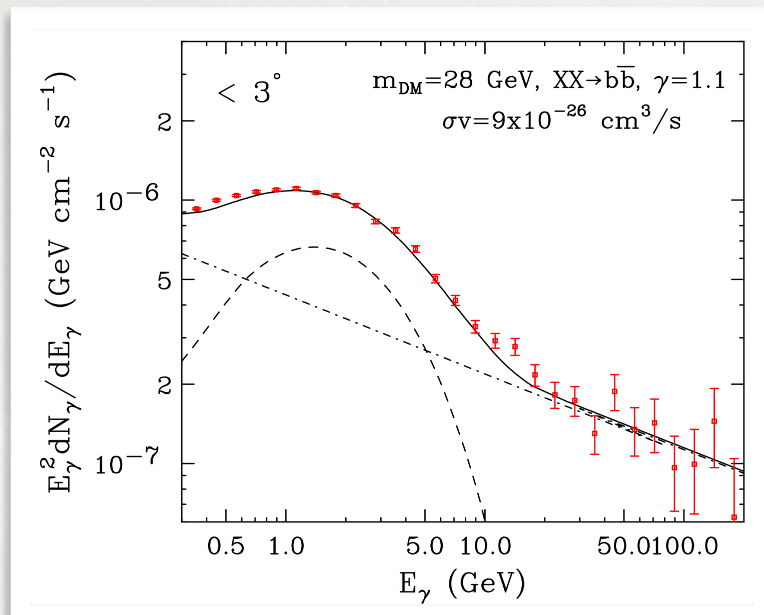
Unresolved MSPs had been suggested earlier e.g. [Wang+ 2005], [Hooper, Goodenough 2010], [Abazajian, Kaplinghat 2012]

Very incomplete summary missing many important developments. For more see recent reviews in: [Murgia 2020] or [Leane 2020]

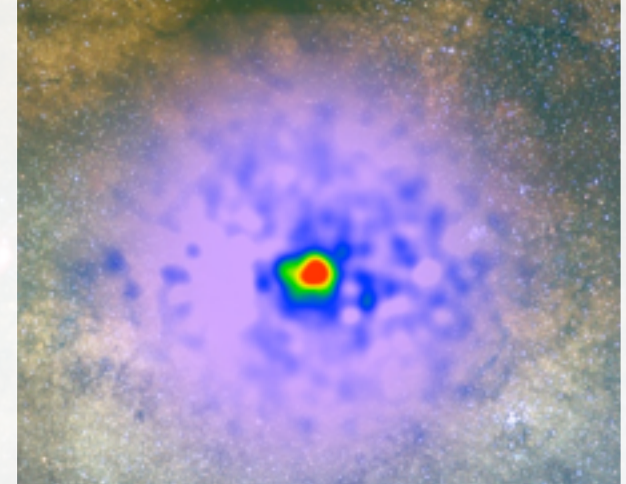
The Galactic Center Excess

Dark Matter

[Goodenough, Hooper 2009]

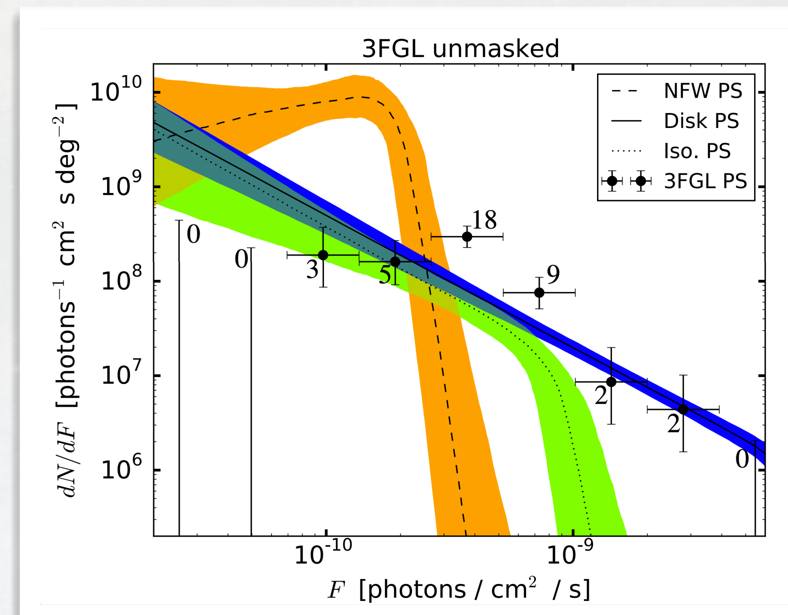


Exhibited many expected properties of DM, e.g. [Daylan, NLR+ 2014]



Millisecond Pulsars

[Lee+ 2016]



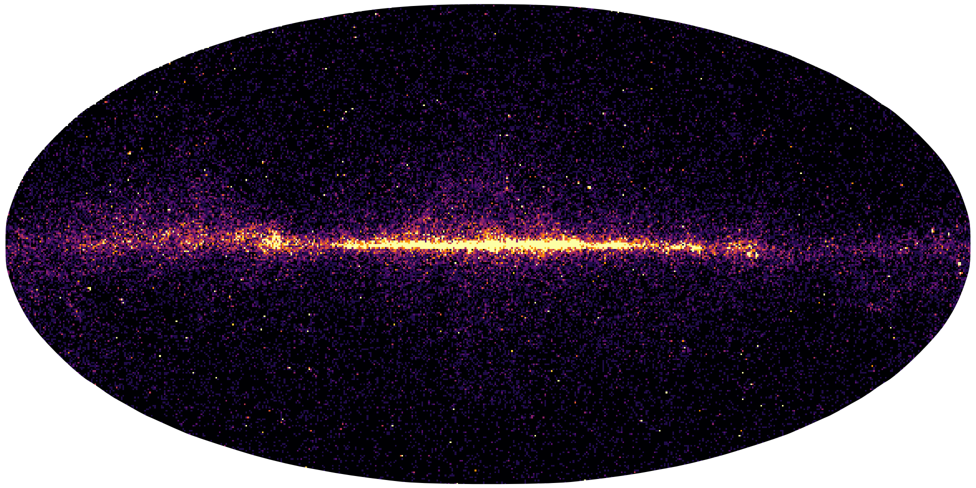
Data preferred clumpy point-source statistics rather than smoother DM

Unresolved MSPs had been suggested earlier e.g. [Wang+ 2005], [Hooper, Goodenough 2010], [Abazajian, Kaplinghat 2012]

Very incomplete summary missing many important developments. For more see recent reviews in: [Murgia 2020] or [Leane 2020]

Likelihood Methods

Fermi γ-ray data



Dataset: 11 years of PASS 8 UltracleanVeto BestPSF data (with standard quality cuts)

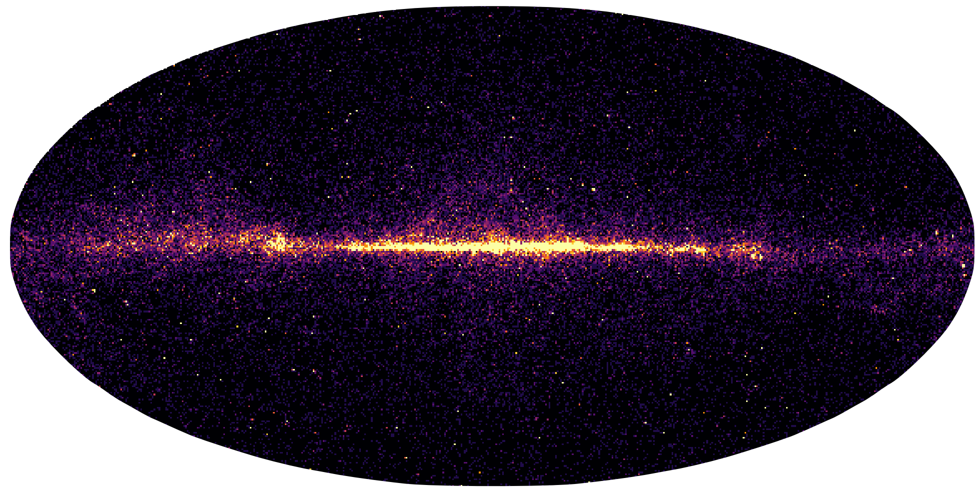
Template Fitting: model the data as a sum of spatial templates $= \sum A_t T_t$

~template fluxes

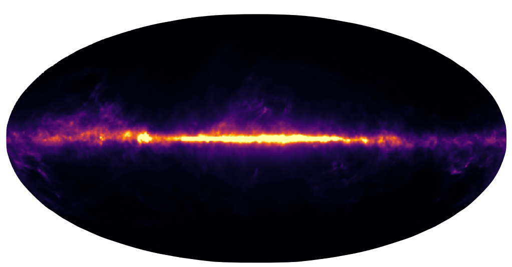
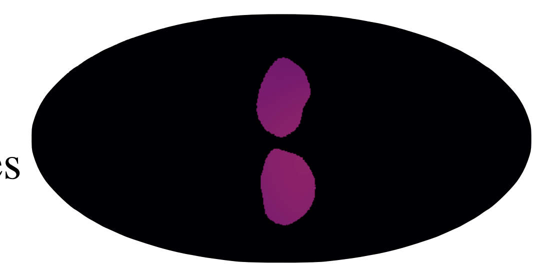
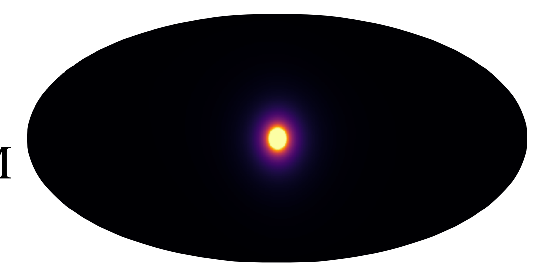
For more details, see [Mishra-Sharma, NLR, Safdi 2016]

Likelihood Methods

Fermi γ-ray data



Dataset: 11 years of PASS 8 UltracleanVeto BestPSF data (with standard quality cuts)

$= A_{\rm diffuse}$  $+A_{\rm bubbles}$  $+A_{\rm DM}$  $+\ldots$

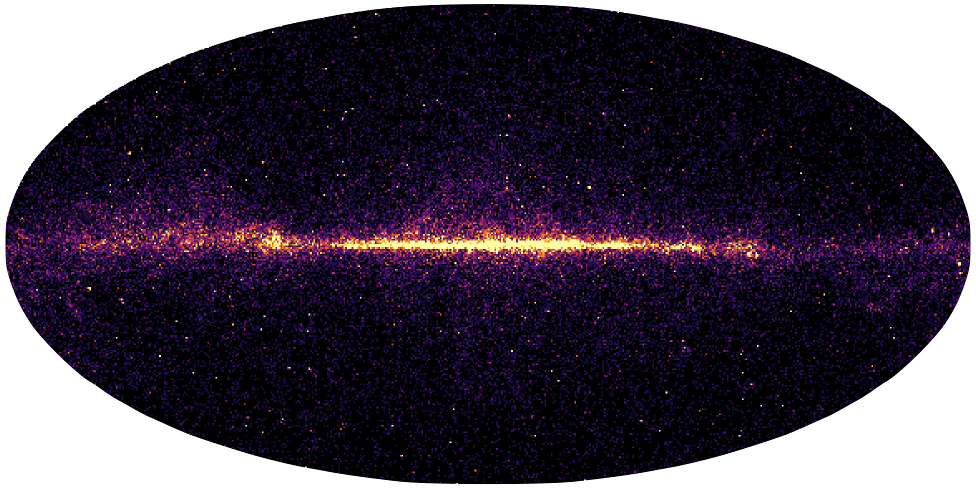
Template Fitting: model the data as a sum of spatial templates $= \sum A_t T_t$

~template fluxes

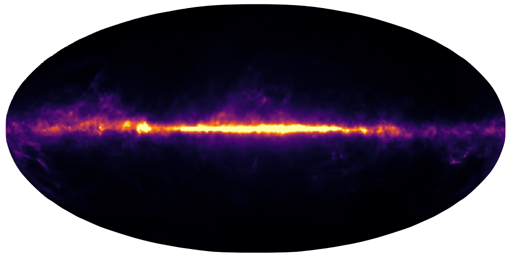
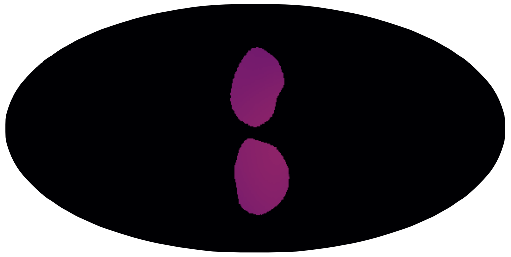
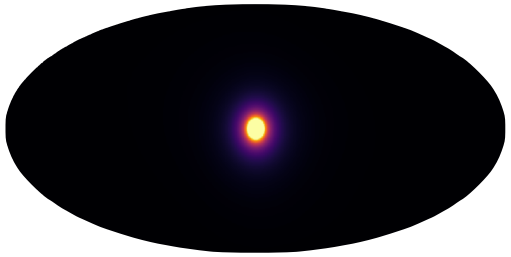
For more details, see [Mishra-Sharma, NLR, Safdi 2016]

Likelihood Methods

Fermi γ-ray data



Dataset: 11 years of PASS 8 UltracleanVeto BestPSF data (with standard quality cuts)

$= A_{\mathrm{diffuse}}$  $+A_{\mathrm{bubbles}}$  $+A_{\mathrm{DM}}$  $+\ldots$

Determine $\theta = \{A_t\}$ using Poisson likelihood

$$p(d\,|\,\theta) = \prod p^{(p)}_{\mathrm{Poisson}}$$

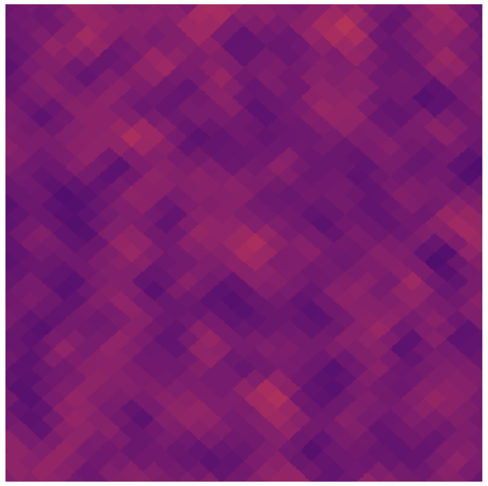
Product over all pixels p in the map

For more details, see [Mishra-Sharma, NLR, Safdi 2016]

Likelihood Methods

Non-Poissonian Template Fitting

"Smooth" Poissonian emission
(expected for dark matter)

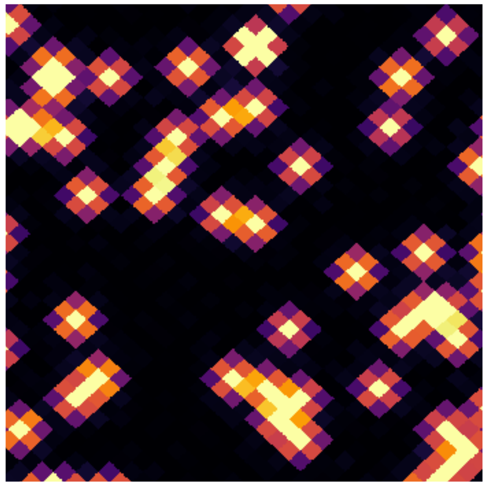


Likelihood: p_{P}

[de Moivre 1711], [Poisson 1837]

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

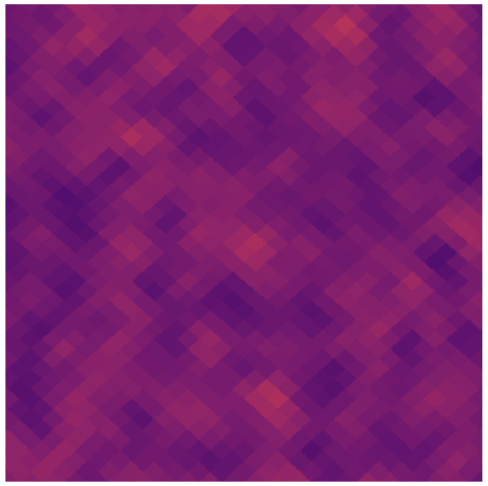
[Malyshev, Hogg 2011], [Lee+ 2016]

For more details, see [Mishra-Sharma, NLR, Safdi 2016]

Likelihood Methods

Non-Poissonian Template Fitting

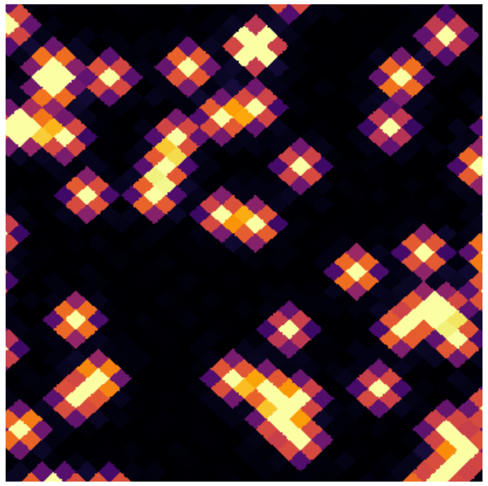
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: p_{P}

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

Template fluxes
+ PS properties

Determine θ using combined likelihood

$$p(d\,|\,\theta) = \prod p_{\mathrm{P}}^{(p)} \times p_{\mathrm{NPTF}}^{(p)}$$

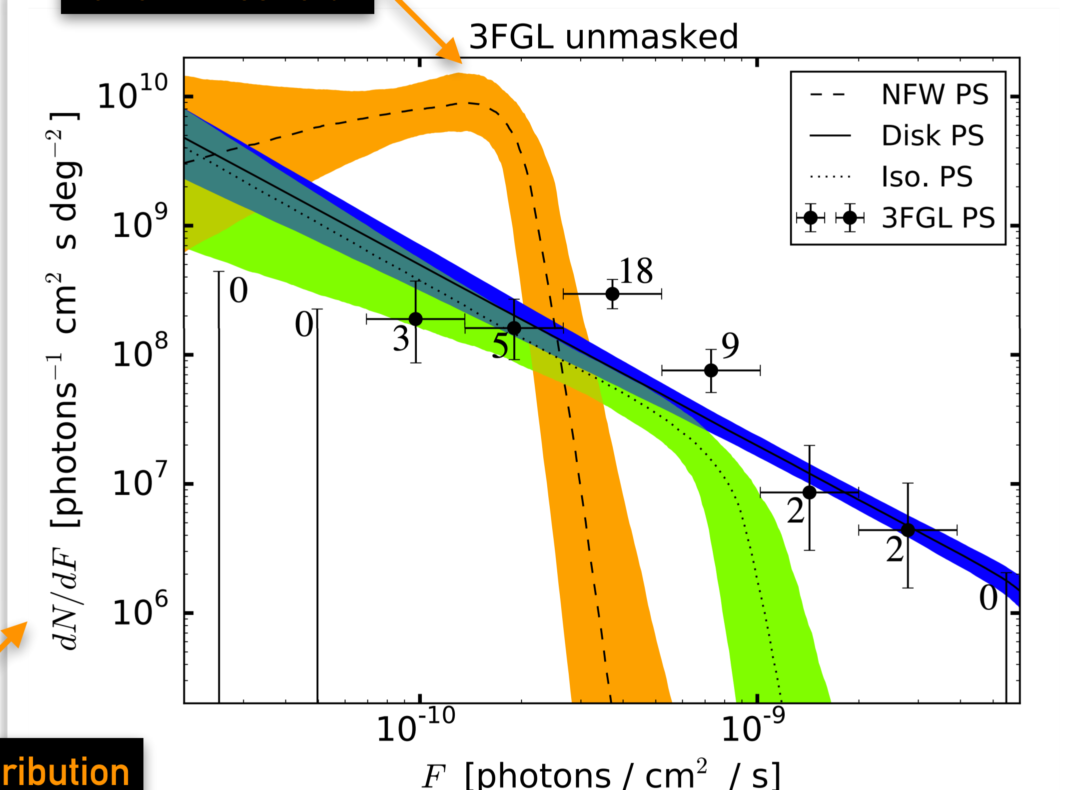
Product over all
pixels p in the map

For more details, see
[Mishra-Sharma, NLR, Safdi 2016]

Likelihood Methods

NPTF $\Rightarrow$ GCE was clumpy

GCE PSs peak just below the 3FGL threshold

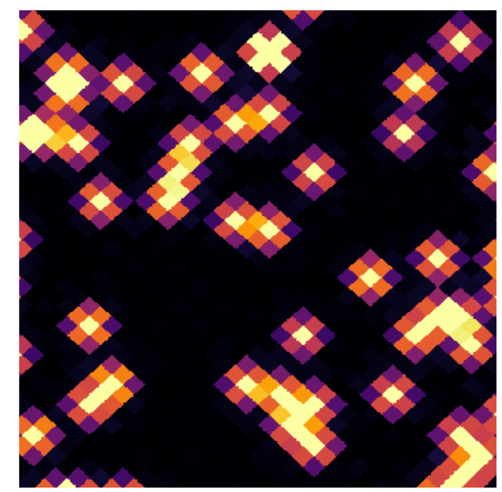


Brightness distribution of sources

[Lee, Lisanti, Safdi, Slatyer, Xue 2016]

"Clumpy" Non-Poissonian emission

(expected for millisecond pulsars)



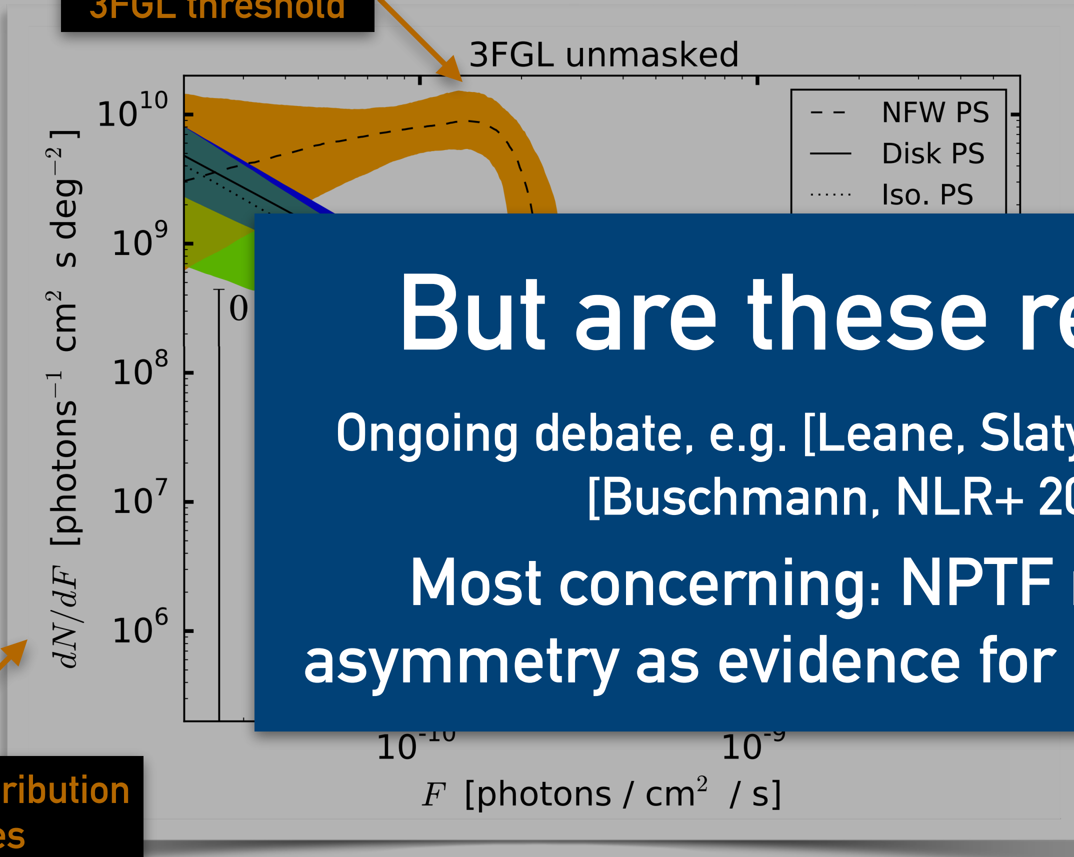
Likelihood: p_{NPTF}

Additional contemporaneous evidence
Wavelets: [Bartels, Krishnamurthy, Weniger 2016], …
Non-spherical morphology: [Macias+ 2018], …

Likelihood Methods

NPTF ⇒ GCE was clumpy

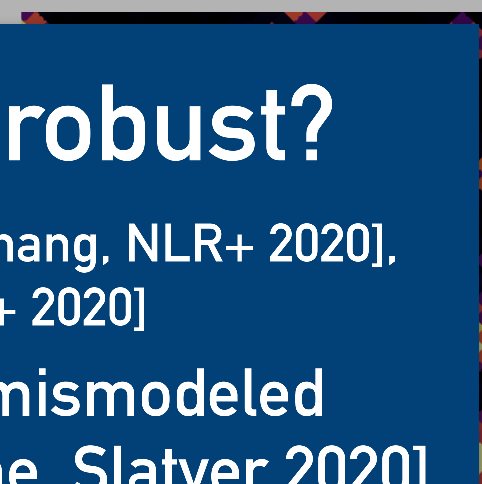
GCE PSs peak just below the 3FGL threshold



Brightness distribution of sources

[Lee, Lisanti, Safdi, Slatyer, Xue 2016]

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

But are these results robust?

Ongoing debate, e.g. [Leane, Slatyer 2019], [Chang, NLR+ 2020], [Buschmann, NLR+ 2020], [Zhong+ 2020]

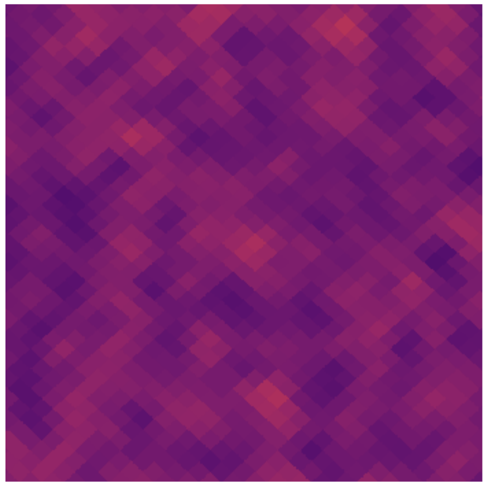
Most concerning: NPTF recovers mismodeled asymmetry as evidence for PSs [Leane, Slatyer 2020]

Additional contemporaneous evidence
Wavelets: [Bartels, Krishnamurthy, Weniger 2016], …
Non-spherical morphology: [Macias+ 2018], …

Where are the problems?

1. Poisson vs Non-Poisson divide is artificial

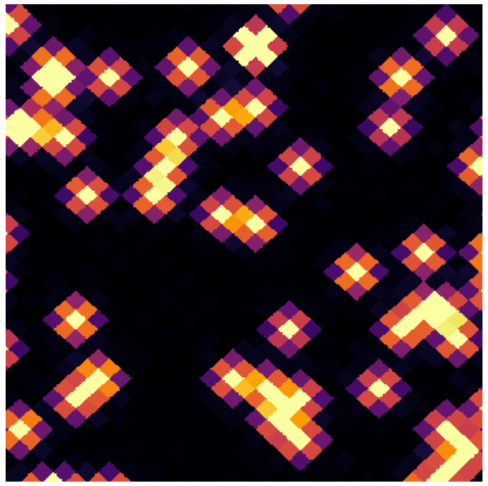
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: p_{P}

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



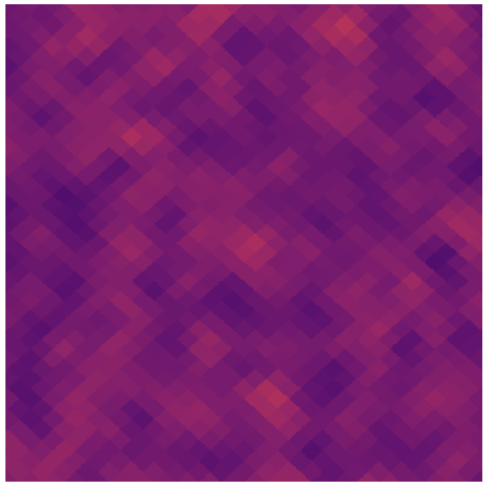
Likelihood: p_{NPTF}

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

1. Poisson vs Non-Poisson divide is artificial

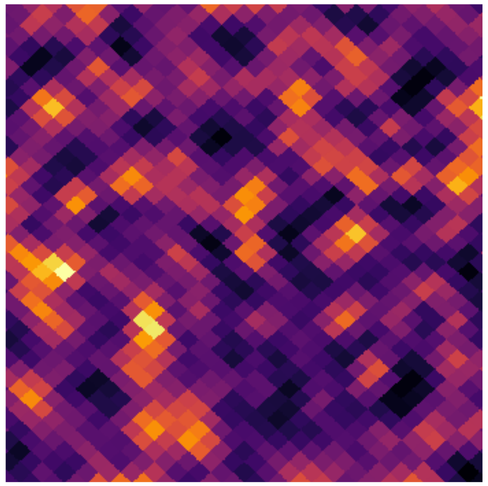
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: $p_{\rm P}$

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



increase $N_{\rm PS}$
decrease $F_{\rm PS}$
(leaving total flux unchanged)

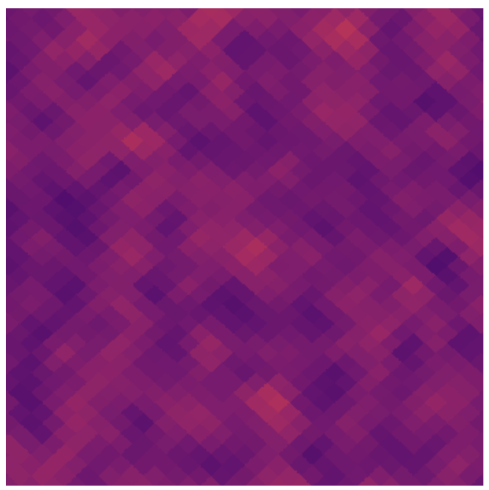
Likelihood: $p_{\rm NPTF}$

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

1. Poisson vs Non-Poisson divide is artificial

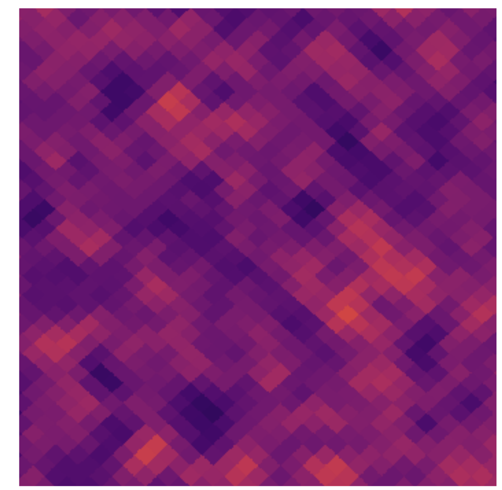
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: p_{P}

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



increase N_{PS}
decrease F_{PS}
(leaving total flux unchanged)

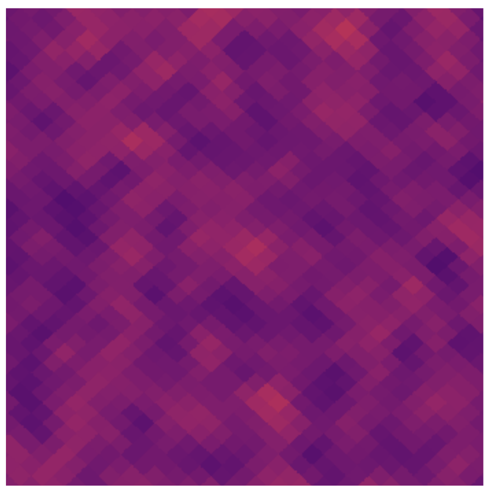
Likelihood: p_{NPTF}

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

1. Poisson vs Non-Poisson divide is artificial

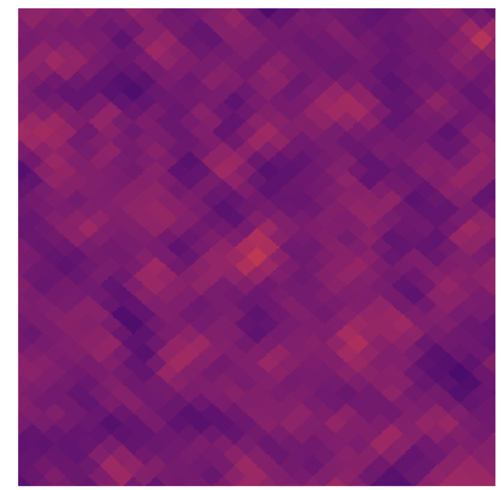
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: p_{P}

vs.

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



increase N_{PS}
decrease F_{PS}
(leaving total flux unchanged)

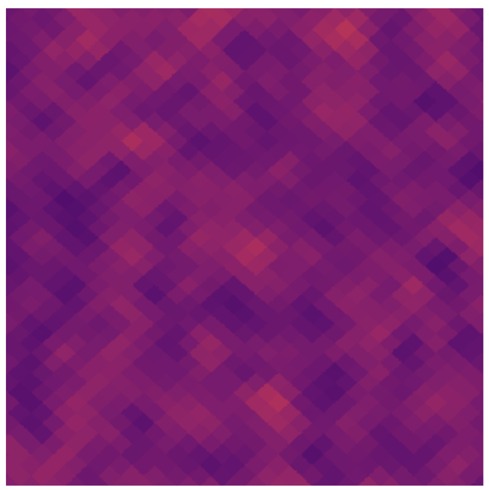
Likelihood: p_{NPTF}

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

1. Poisson vs Non-Poisson divide is artificial

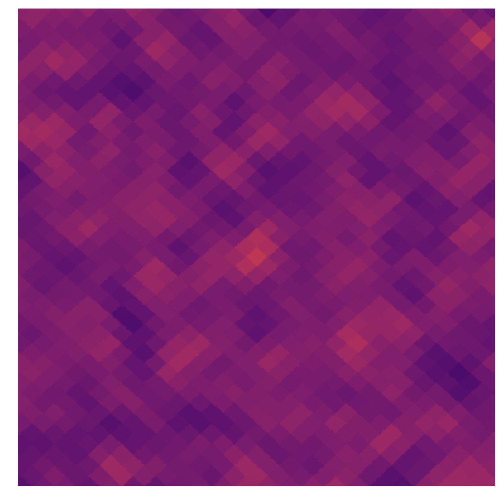
"Smooth" Poissonian emission
(expected for dark matter)



Likelihood: p_{P}

$\subset$

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

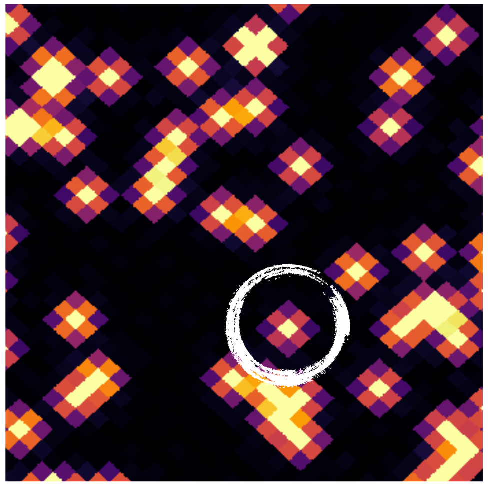
Many dim sources = Poisson emission
To address, in [List, NLR, Lewis 2021]
we only use the point-source model,
Poisson included as dim-sources

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

2. Likelihood in each pixel is not independent

"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

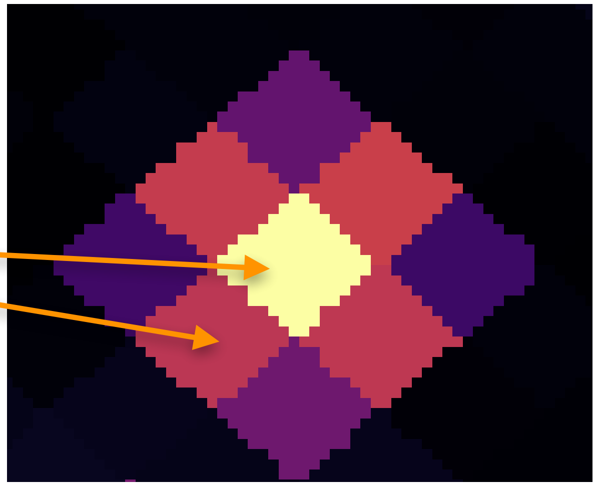
$$p(d\,|\,\theta) = \prod p_{\mathrm{P}}^{(p)} \times p_{\mathrm{NPTF}}^{(p)}$$

Product over all pixels p in the map

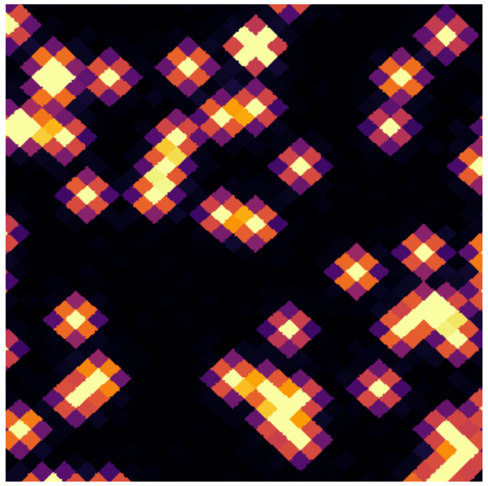
For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

Where are the problems?

2. Likelihood in each pixel is not independent



"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

$$p(d\,|\,\theta) = \prod p_{\mathrm{P}}^{(p)} \times p_{\mathrm{NPTF}}^{(p)}$$

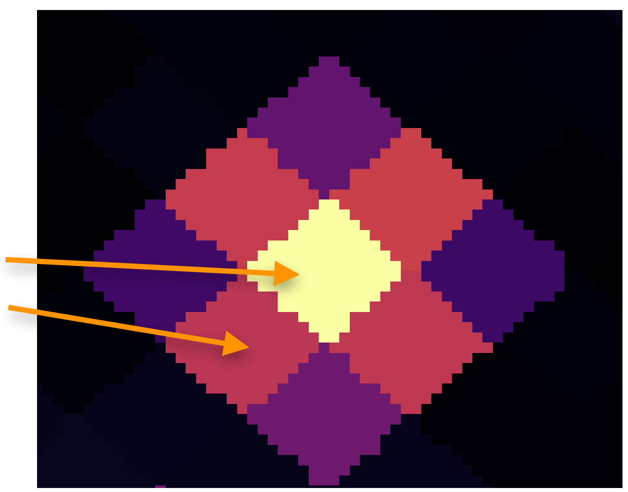
Product over all pixels p in the map

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

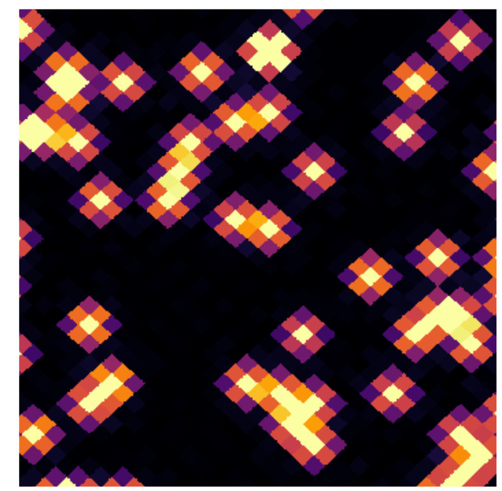
Where are the problems?

2. Likelihood in each pixel is not independent

Not independent: one point-source smeared by the instrument



"Clumpy" Non-Poissonian emission
(expected for millisecond pulsars)



Likelihood: p_{NPTF}

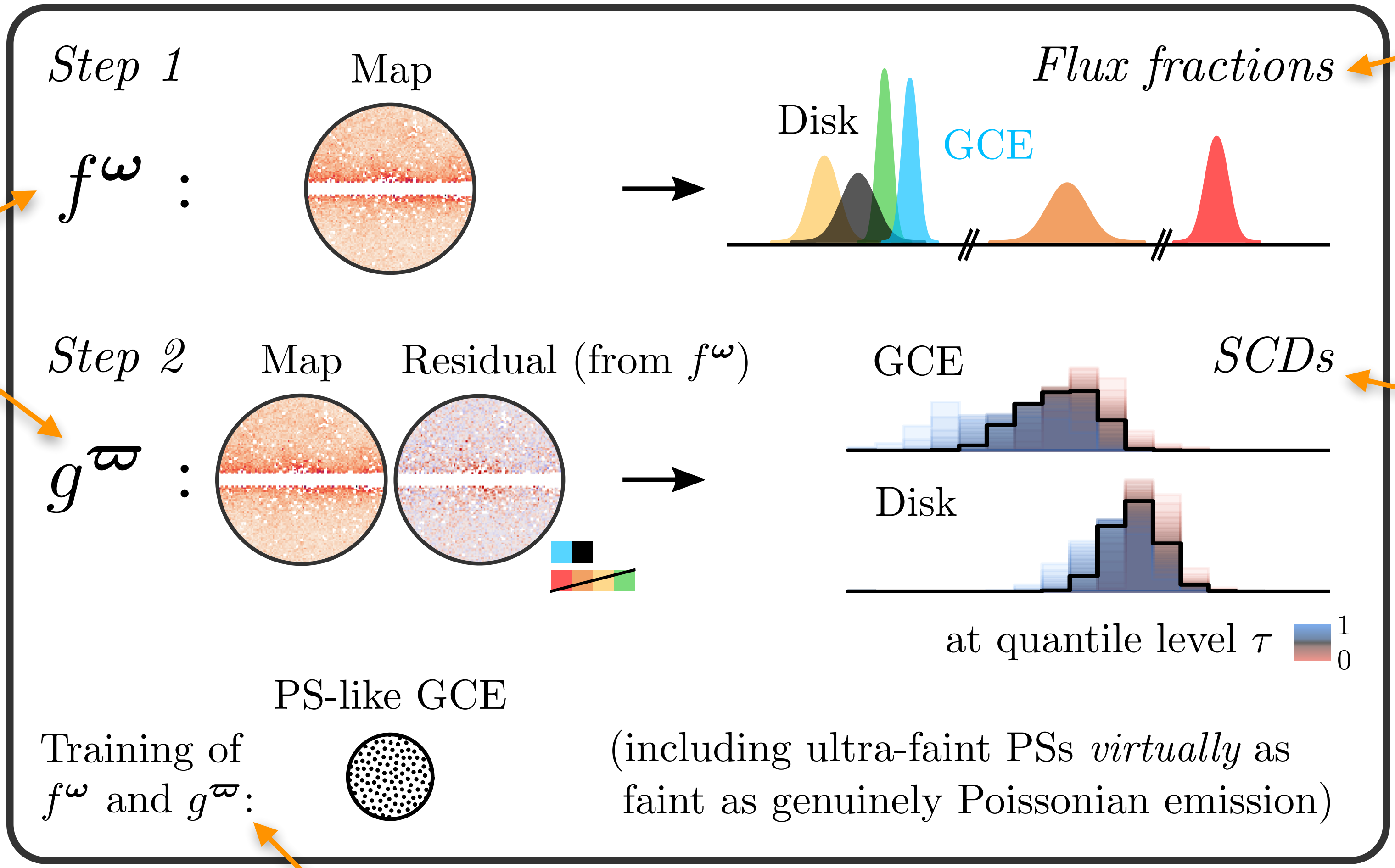
NPTF only approximates the true likelihood - unused information ML can exploit

$$p(d\,|\,\theta) \approx \prod p_{\mathrm{P}}^{(p)} \times p_{\mathrm{NPTF}}^{(p)}$$

For additional discussion, further issues with NPTF, and an improved likelihood, see [Collin, NLR, Erjavec, Perez 2021]

A Machine Learning Approach

A two step approach to the GCE



Two separate Convolutional Neural Networks (CNNs)

$= A_i$ in template fitting

$= F^2 dN/dF =$ PS brightness distribution

Fully-supervised method
Train on simulated Fermi maps

For an earlier application of CNNs to the GCE, see [Caron+ 2018]

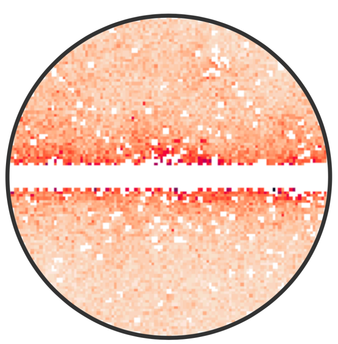
[List, NLR, Lewis 2021]

A Machine Learning Approach

Step 1: estimate template flux fractions

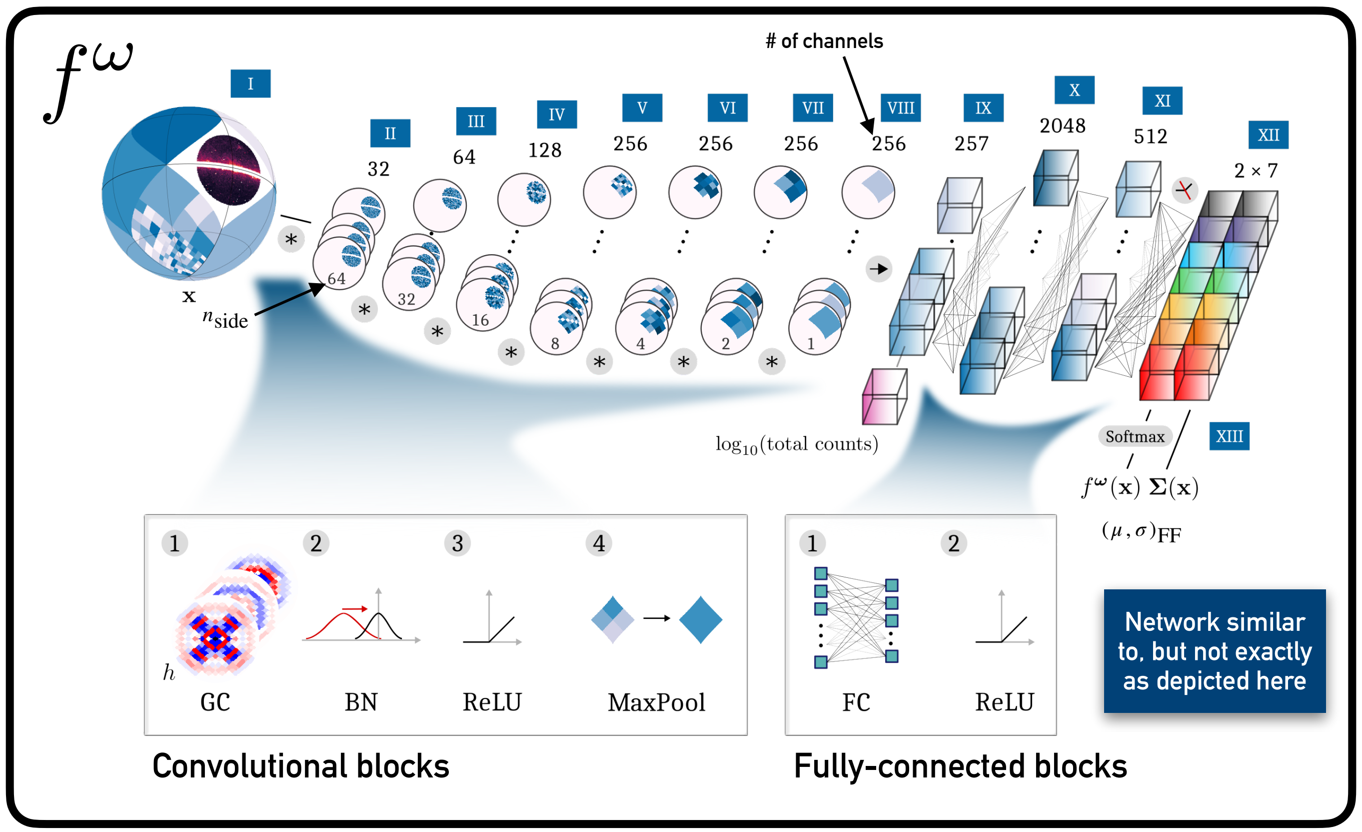
Input

Fermi map



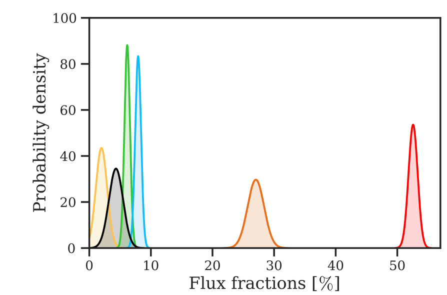
ROI: $r < 25°$; $|b| > 2°$; mask 3FGL at 95% containment

$\Rightarrow$



$\Rightarrow$

Output



Estimate FF and uncertainty with negative Gaussian LLH loss function

Image from [List, NLR, Lewis, Bhat 2020], see there for network details
We add 1 layer, as begin with $n_{\rm side} = 256$

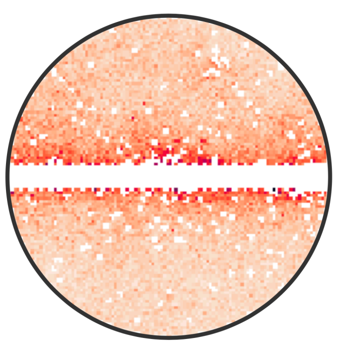
[List, NLR, Lewis 2021]

A Machine Learning Approach

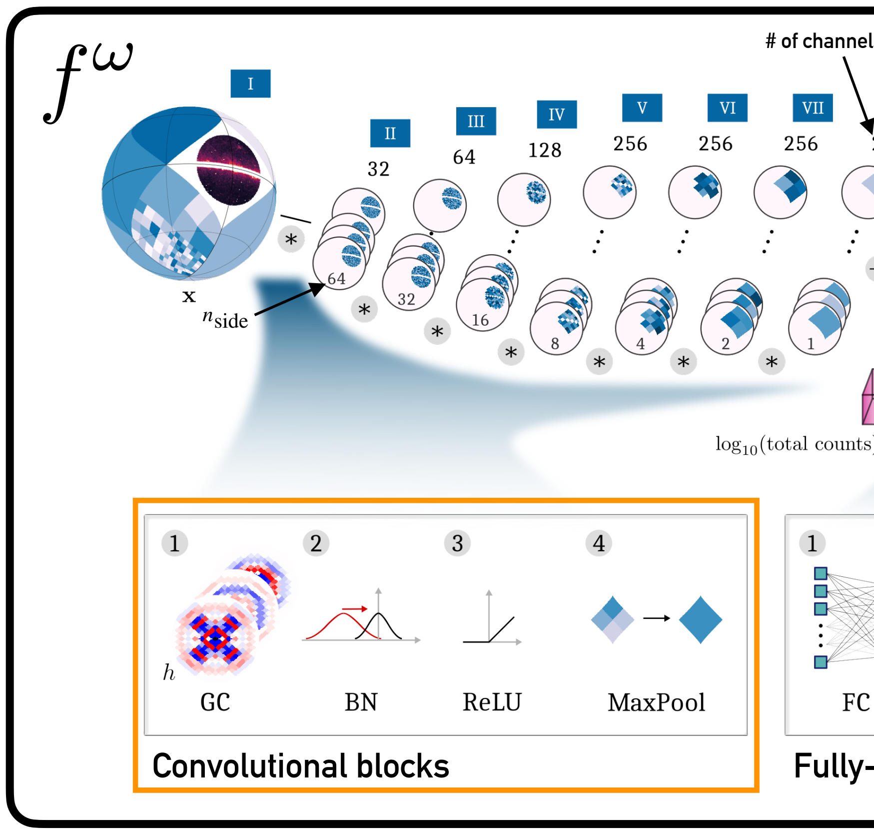
Step 1: estimate template flux fractions

Input

Fermi map



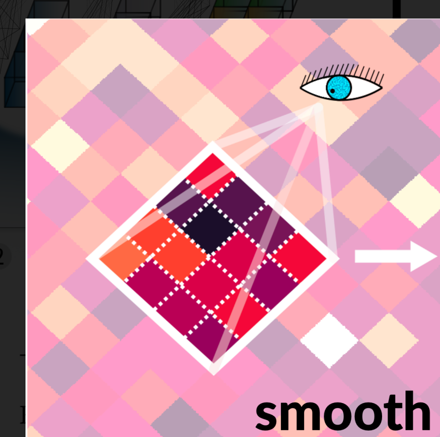
ROI: $r < 25°$; $|b| > 2°$; mask 3FGL at 95% containment



Convolutional Blocks

Assess features in regions of the map, e.g. template edges, smoothness, granularity

Use DeepSphere [Perraudin 2019], [Defferrard 2020]



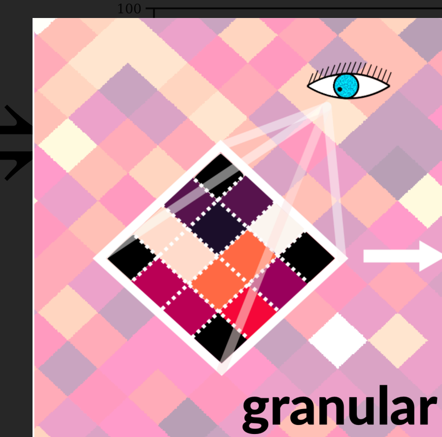


Image from [List, NLR, Lewis, Bhat 2020], see there for network details
We add 1 layer, as begin with $n_{\mathrm{side}} = 256$

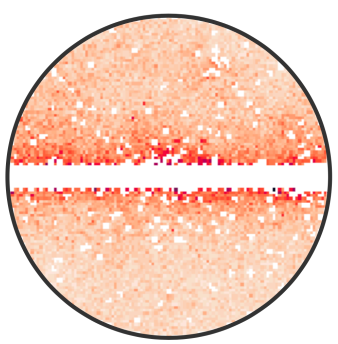
[List, NLR, Lewis 2021]

A Machine Learning Approach

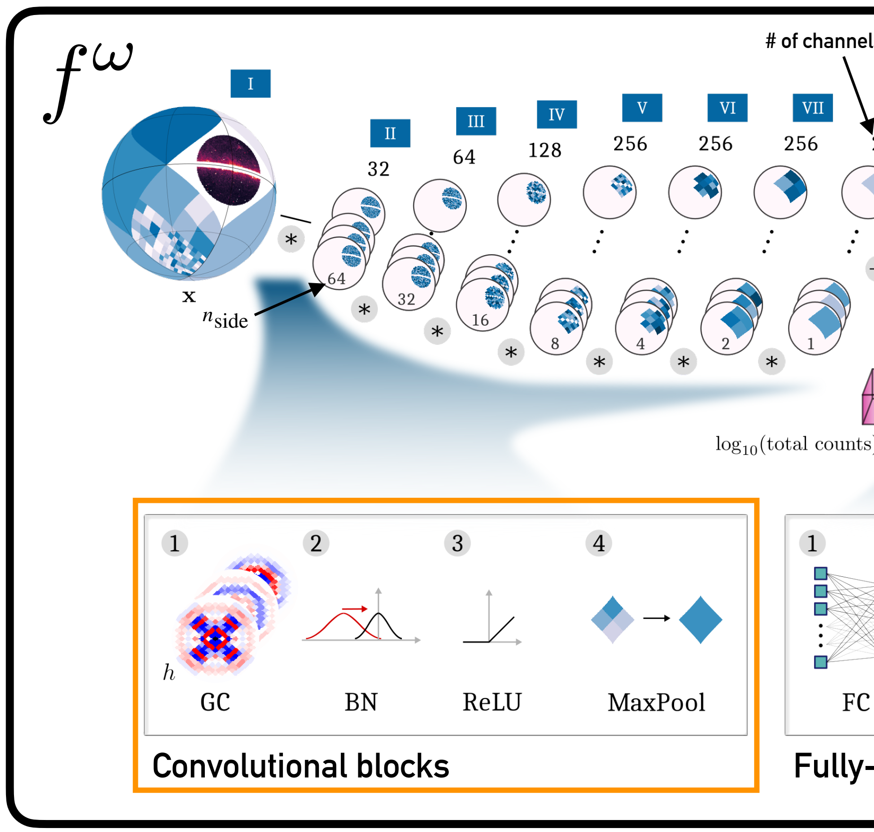
Step 1: estimate template flux fractions

Input

Fermi map



$\Rightarrow$

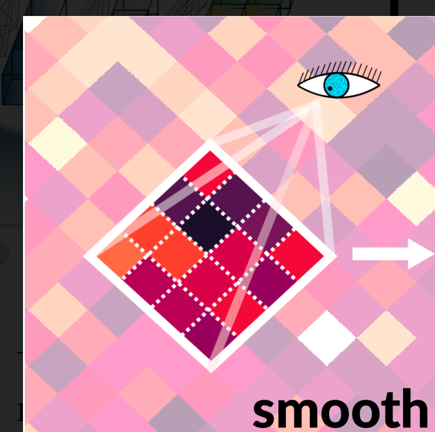


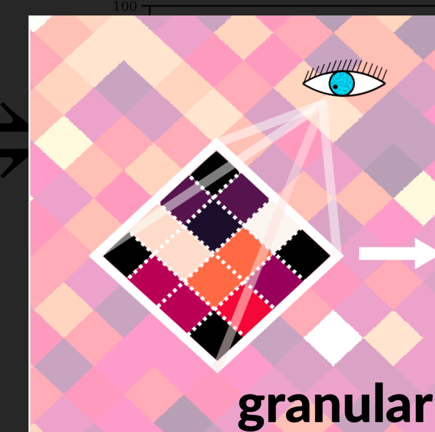
ROI: $r < 25°$; $|b| > 2°$; mask 3FGL at 95% containment

Convolutional Blocks

Assess features in regions of the map, e.g. template edges, smoothness, granularity

Use DeepSphere [Perraudin 2019], [Defferrard 2020]





Considers regions not individual pixels – very different to the likelihood approach

E.g. does not reconstruct a template asymmetry as evidence for point sources

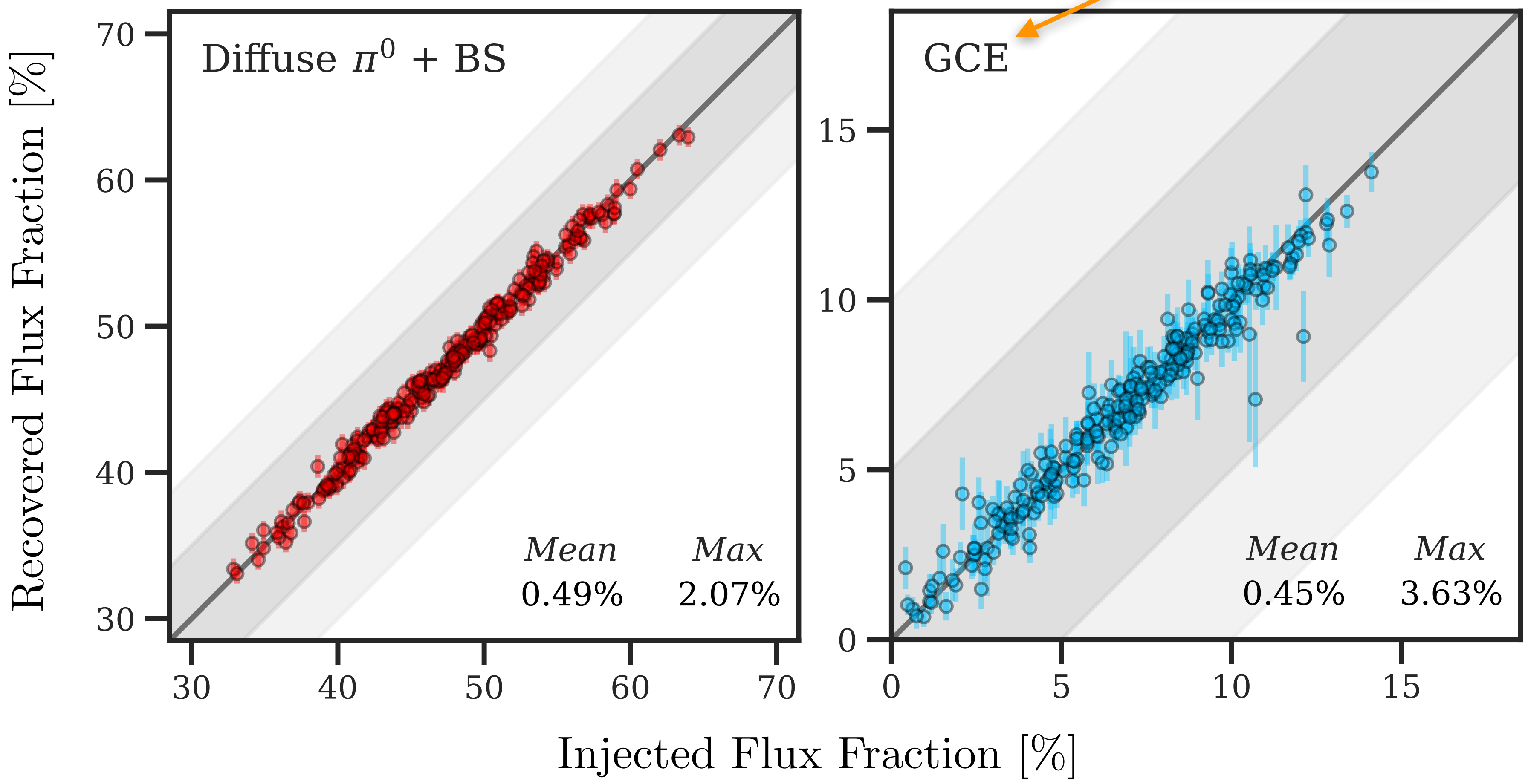
Image from [List, NLR, Lewis, Bhat 2020], see there for network details
We add 1 layer, as begin with $n_{\text{side}} = 256$

[List, NLR, Lewis 2021]

A Machine Learning Approach

Step 1: results in simulated data

Combined Poisson + PS



Not shown: estimates for diffuse IC, isotropic, Fermi bubbles, and disk

[List, NLR, Lewis 2021]

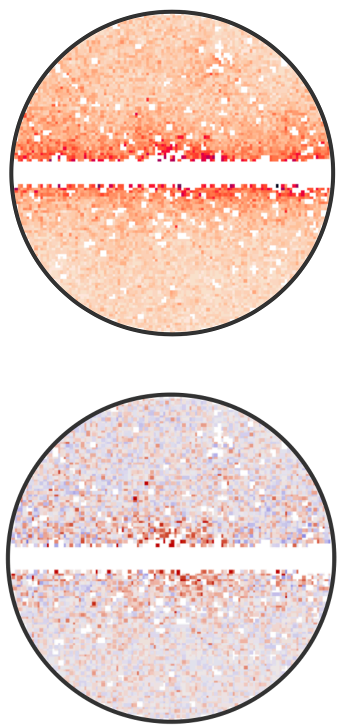
A Machine Learning Approach

Step 2: estimate GCE & disk dN/dF (SCD)

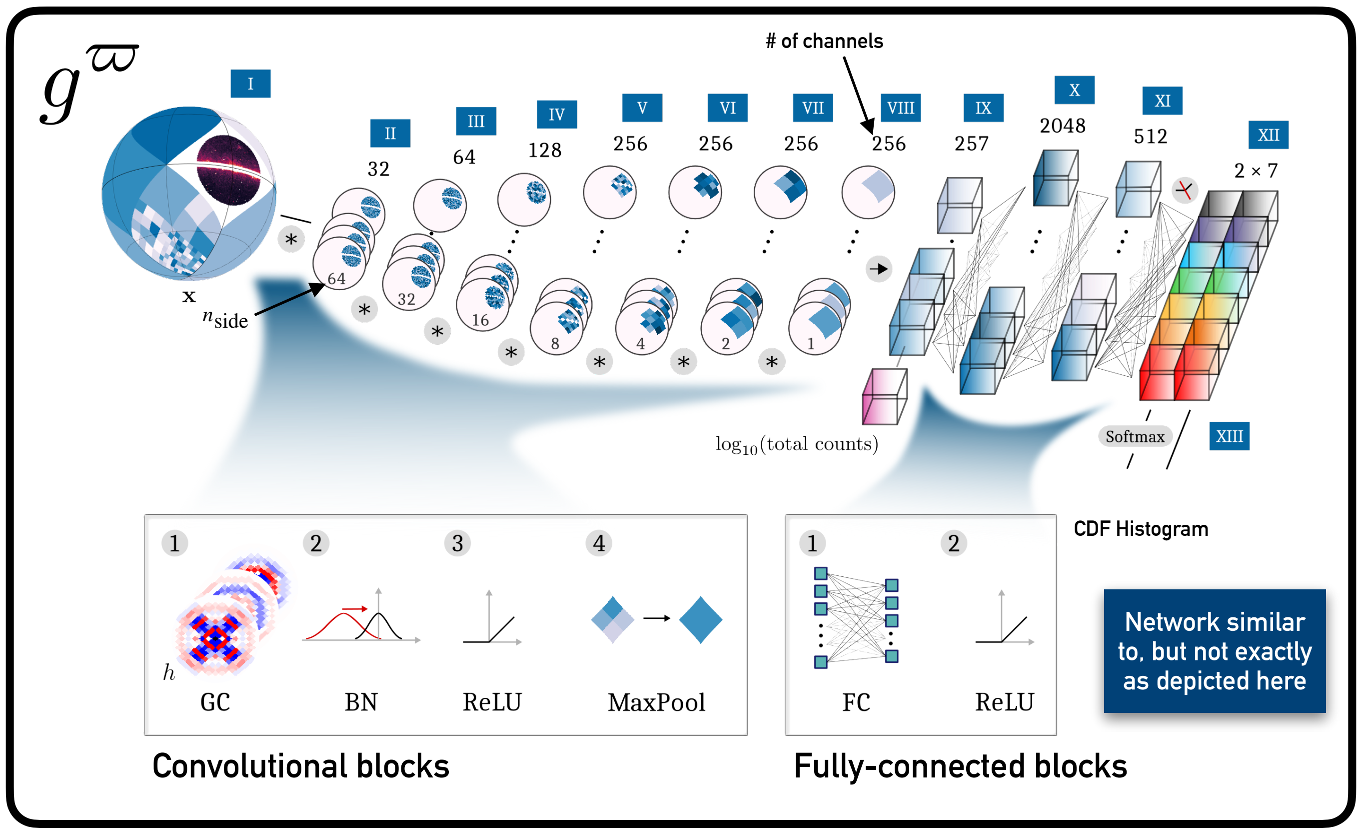
Input

Fermi map

f^{ω} residual

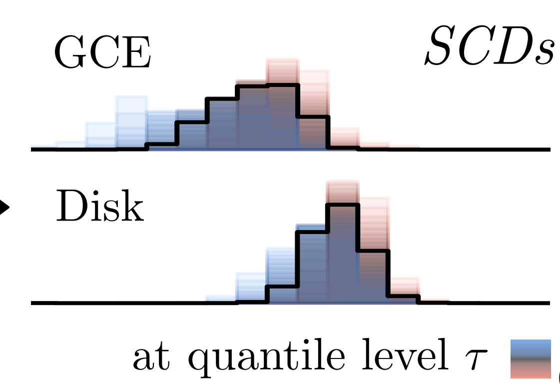


⇒



⇒

Output



Estimate quantiles of the SCD CDF histogram using the Earth Mover's Pinball Loss [List 2021]

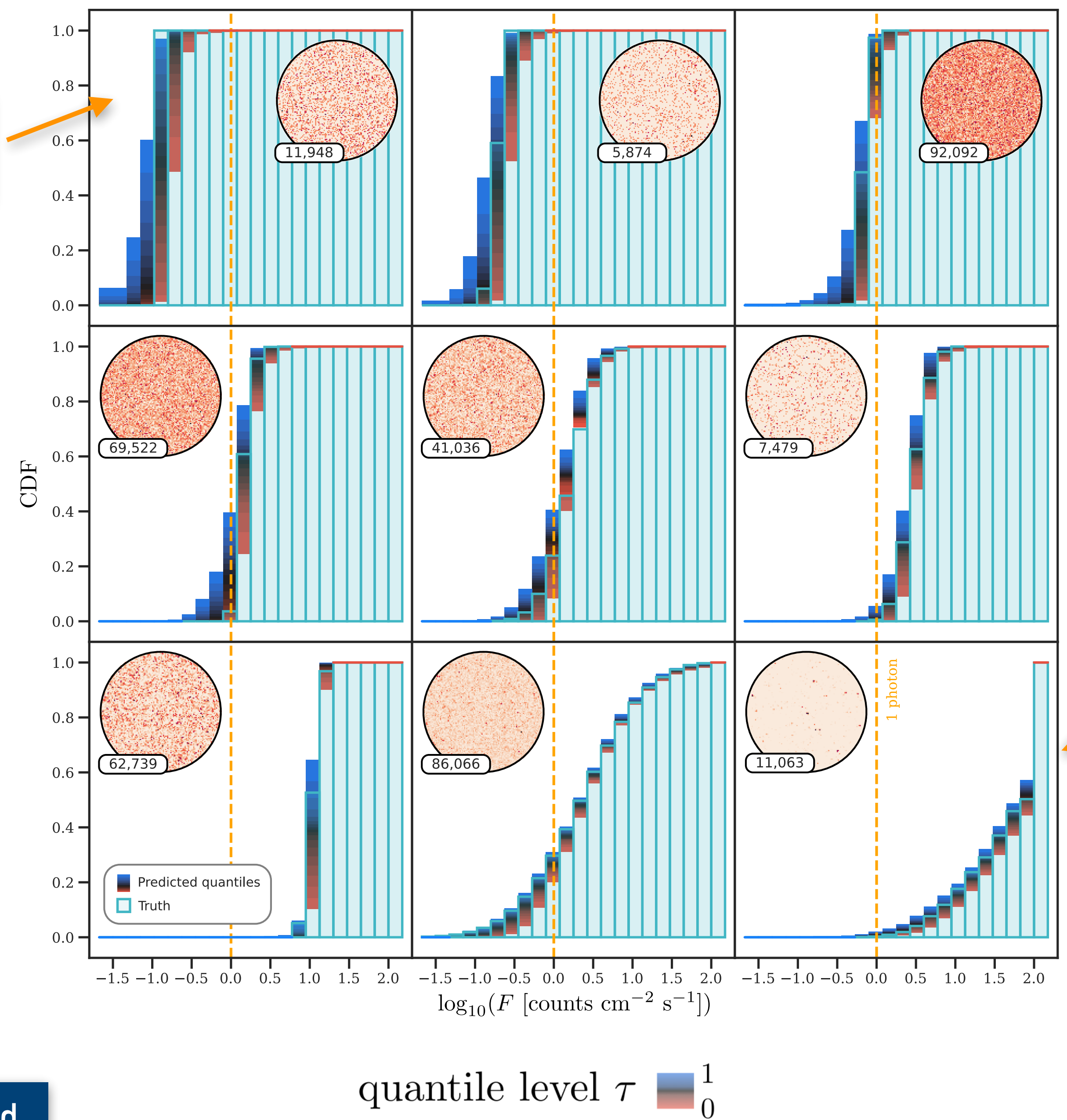
Image from [List, NLR, Lewis, Bhat 2020], see there for network details
We add 1 layer, as begin with $n_{\text{side}} = 256$

[List, NLR, Lewis 2021]

A Machine Learning Approach

Step 2: results in simulated data

Poisson-like ($F_{\rm PS} \ll 1$ photon)



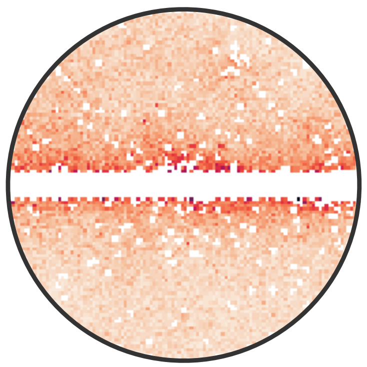
Point Source-like ($F_{\rm PS} \gg 1$ photon)

Results shown for isotropic MC - good performance also seen for disk & GCE

[List, NLR, Lewis 2021]

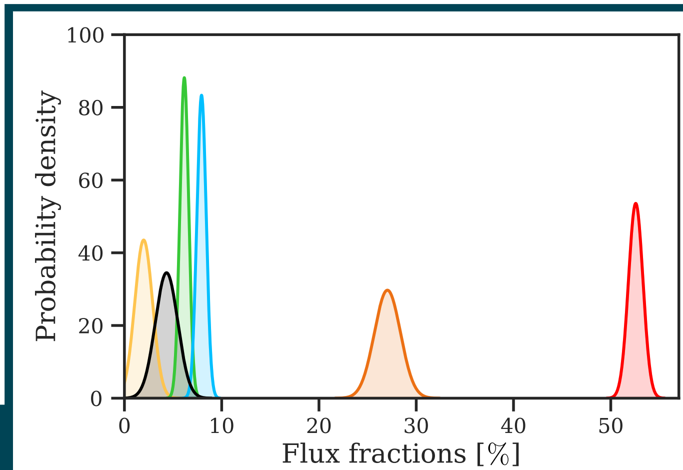
Results

Fermi map



Diffuse π^0 + BS	$(52.6 \pm 0.7)\,\%$
Diffuse IC	$(27.0 \pm 1.3)\,\%$
Isotropic	$(2.0 \pm 0.9)\,\%$
Fermi bubbles	$(6.2 \pm 0.5)\,\%$
GCE	$(7.9 \pm 0.5)\,\%$
Disk	$(4.3 \pm 1.2)\,\%$

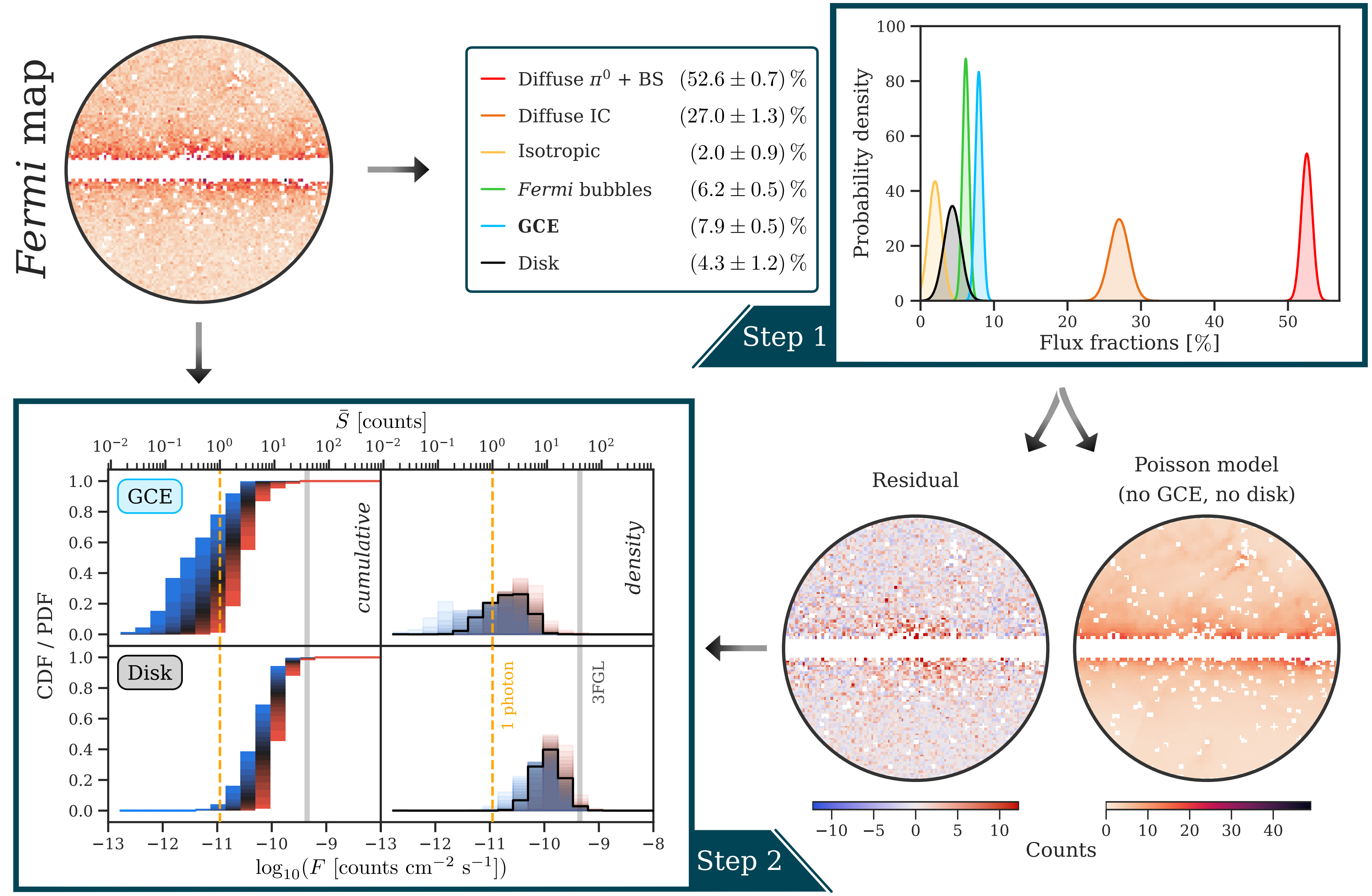
Step 1



Values & uncertainties almost identical to likelihood method

[List, NLR, Lewis 2021]

Results



We use latest diffuse models, if use with NPTF, also find a dimmer SCD

SCD $\Rightarrow N_{\rm PS} \sim 3 \times 10^4$ (100% PSs) or $\sim 6 \times 10^3$ (brightest 50% PSs)
Consistent with recent MSP population studies e.g. [Gonthier+ 2018], [Ploeg+ 2020]

Much dimmer GCE SCD than previous results

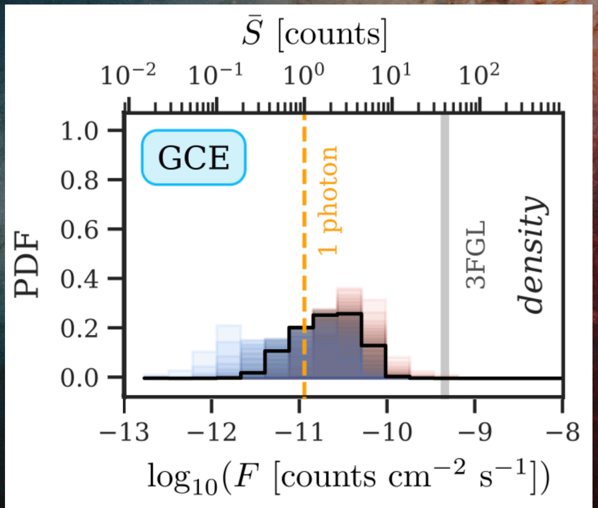
Consistent with no more than 66% Poisson emission (determined with a 3rd NN)

[List, NLR, Lewis 2021]

Conclusion

Existing GCE analyses are not optimal: room for ML

CNN finds a much dimmer source-count distribution

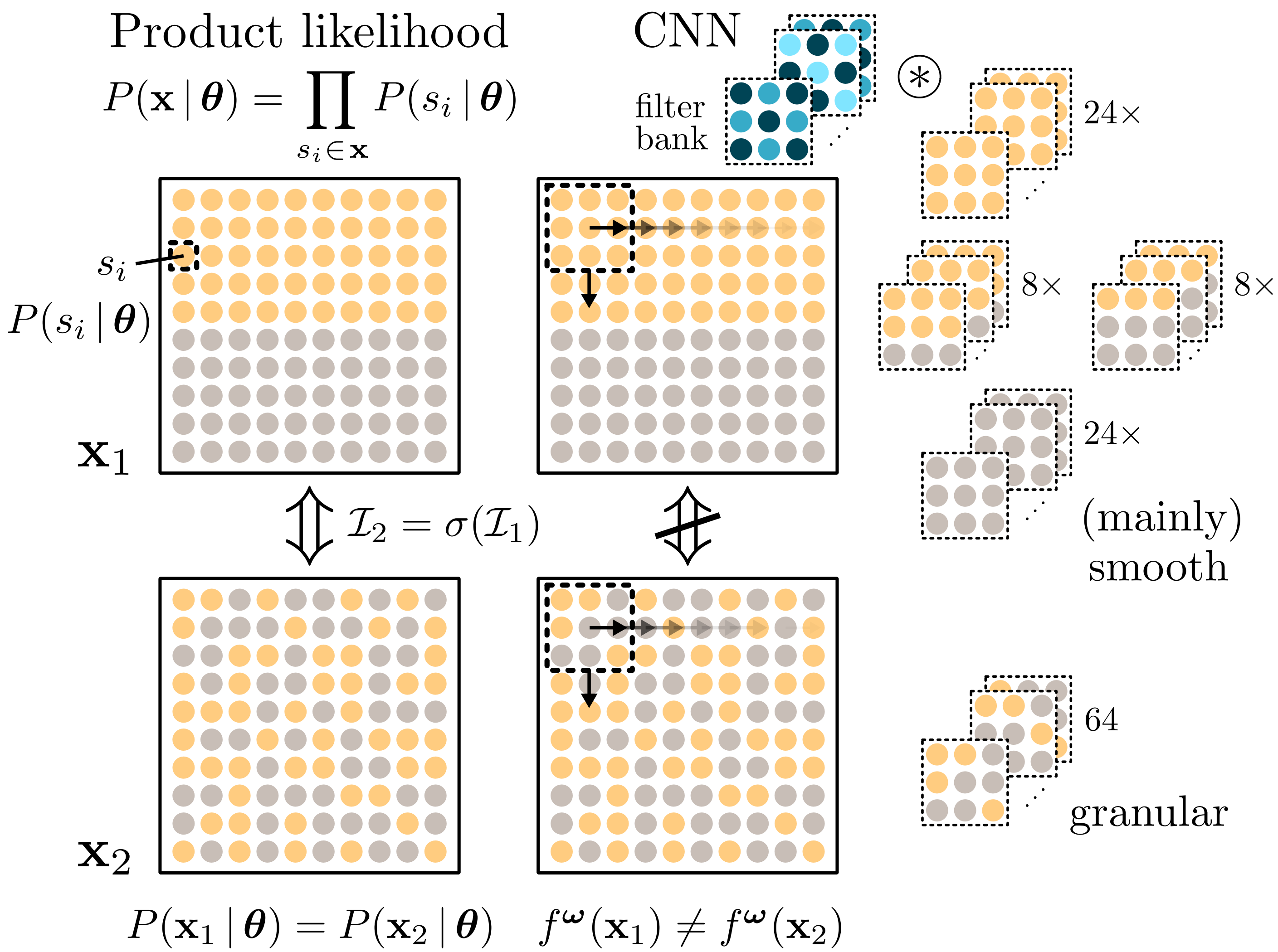


Significant scope to expand and explore ML methods

Backup Slides

Likelihood vs CNN

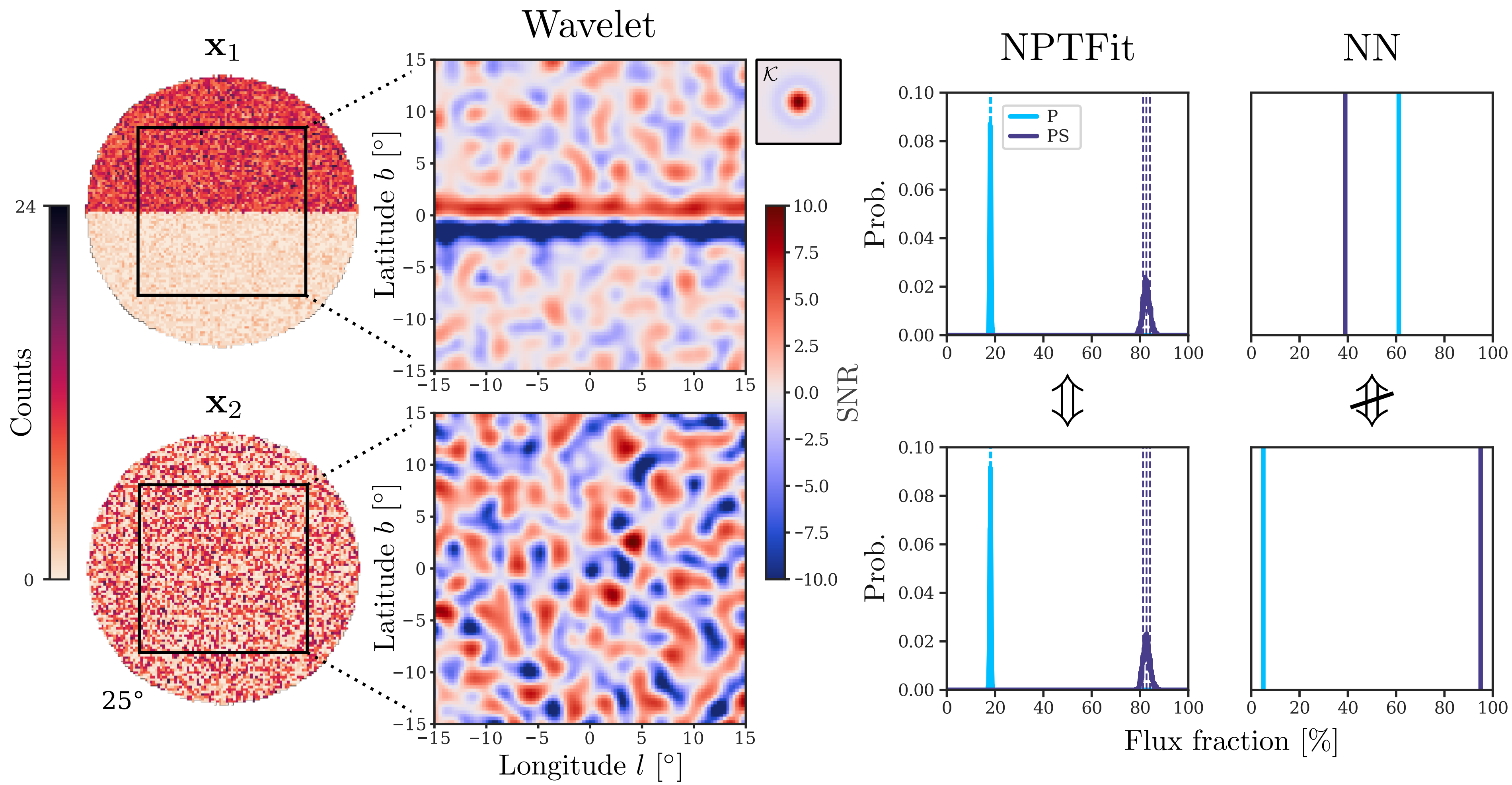
How can the NPTF reconstruct an asymmetry as PSs, but the CNN not?



[List, NLR, Lewis 2021]

Likelihood vs CNN

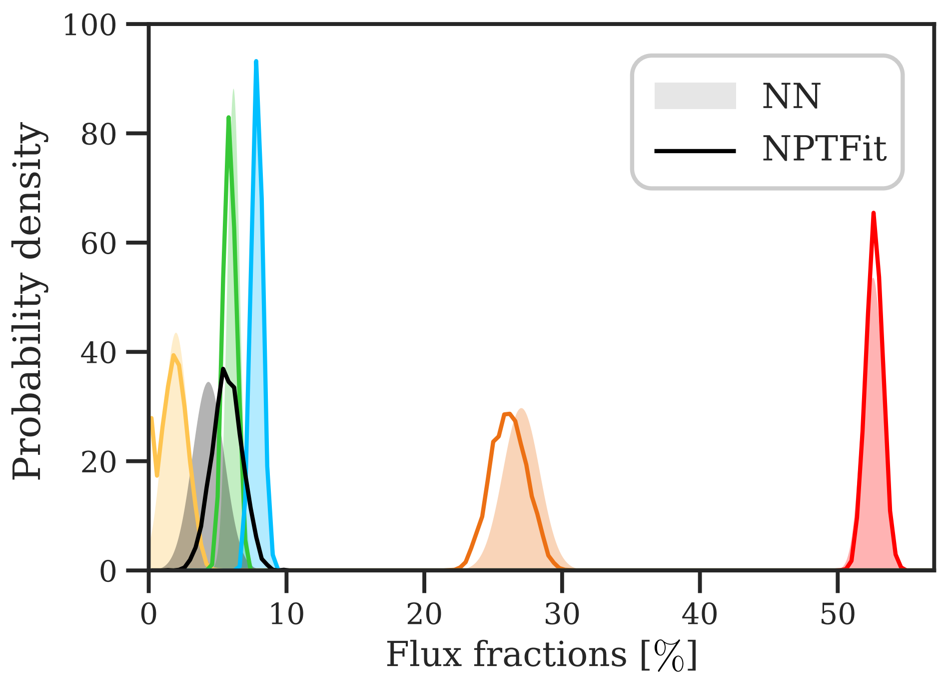
How can the NPTF reconstruct an asymmetry as PSs, but the CNN not?

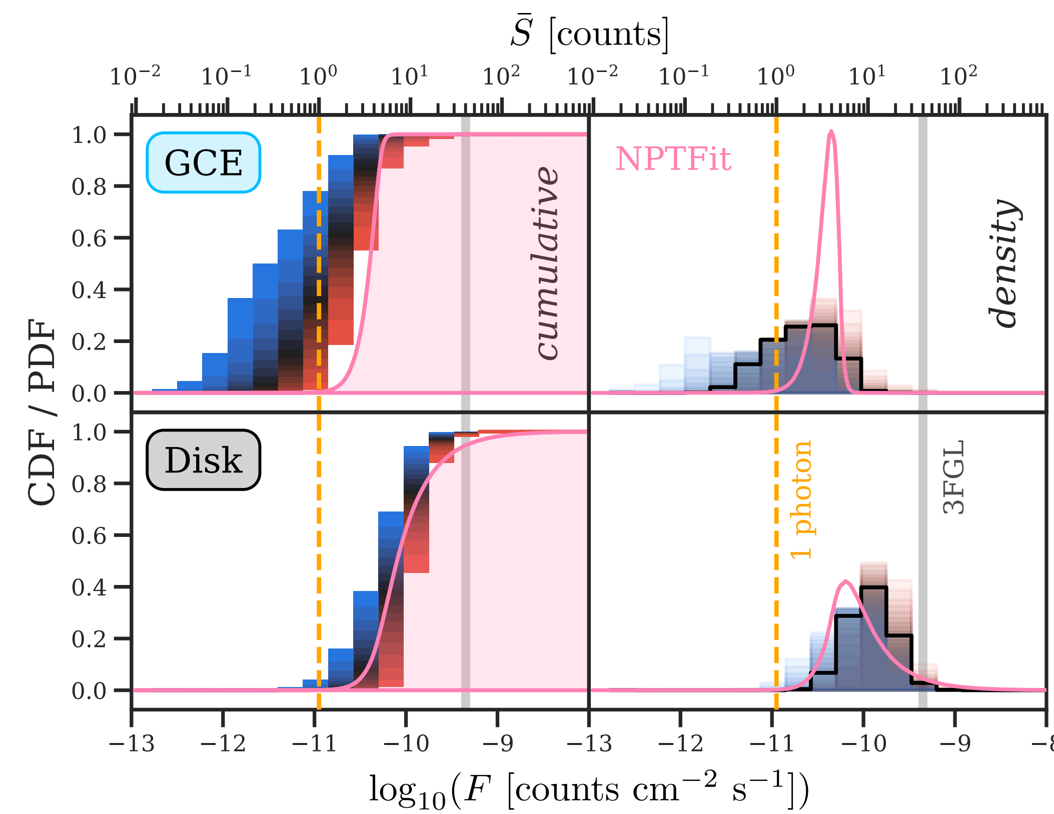


[List, NLR, Lewis 2021]

Likelihood vs CNN

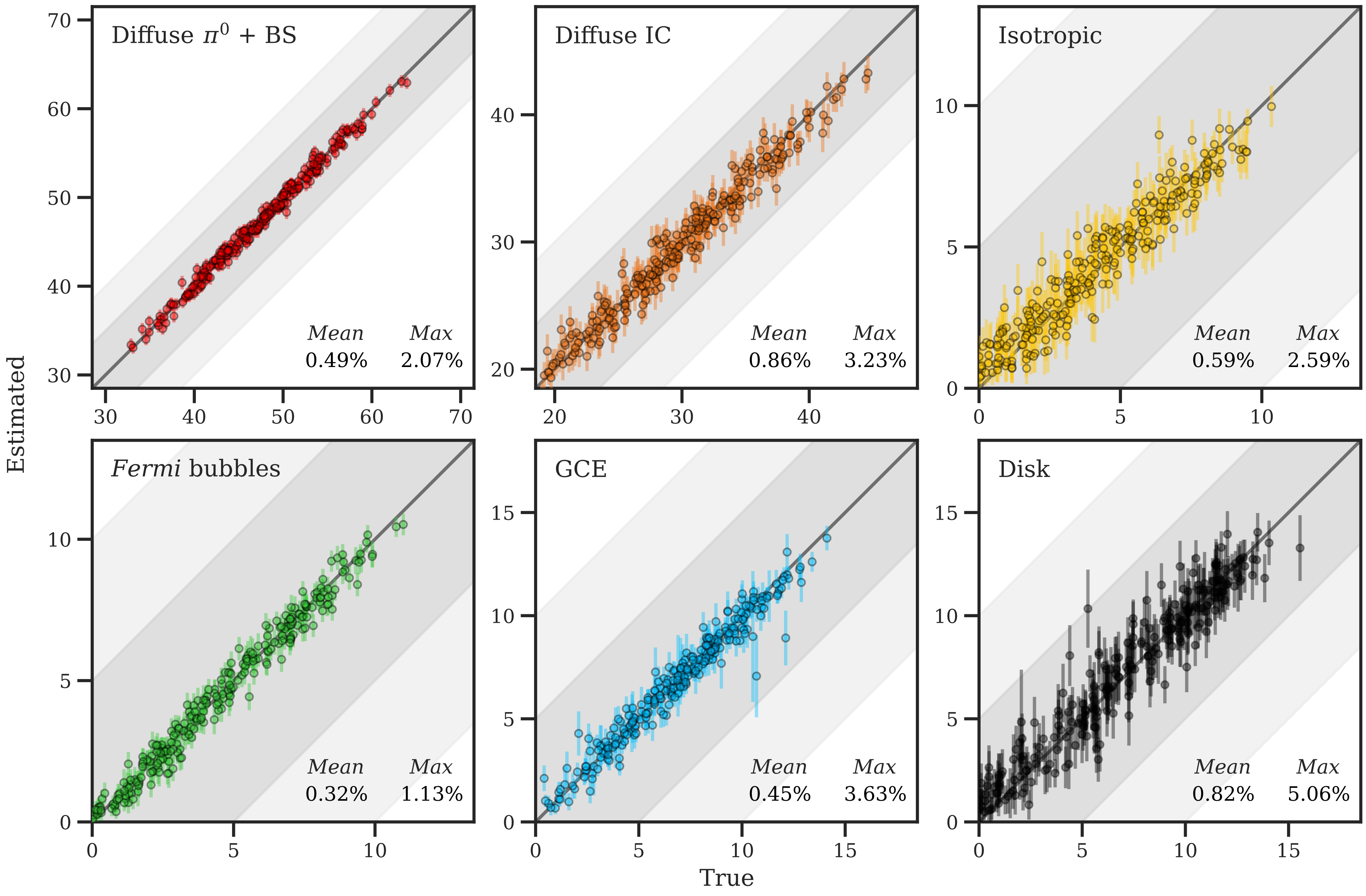
	NN	NPTFit
Diffuse π^0 + BS	$(52.6 \pm 0.7)\,\%$	$(52.6 \pm 0.6)\,\%$
Diffuse IC	$(27.0 \pm 1.3)\,\%$	$(26.1\,^{+1.4}_{-1.3})\,\%$
Isotropic	$(2.0 \pm 0.9)\,\%$	$(1.8\,^{+1.0}_{-1.1})\,\%$
Fermi bubbles	$(6.2 \pm 0.5)\,\%$	$(5.9\,^{+0.5}_{-0.4})\,\%$
GCE	$(7.9 \pm 0.5)\,\%$	$(7.9 \pm 0.4)\,\%$
Disk	$(4.3 \pm 1.2)\,\%$	$(5.7 \pm 1.1)\,\%$





[List, NLR, Lewis 2021]

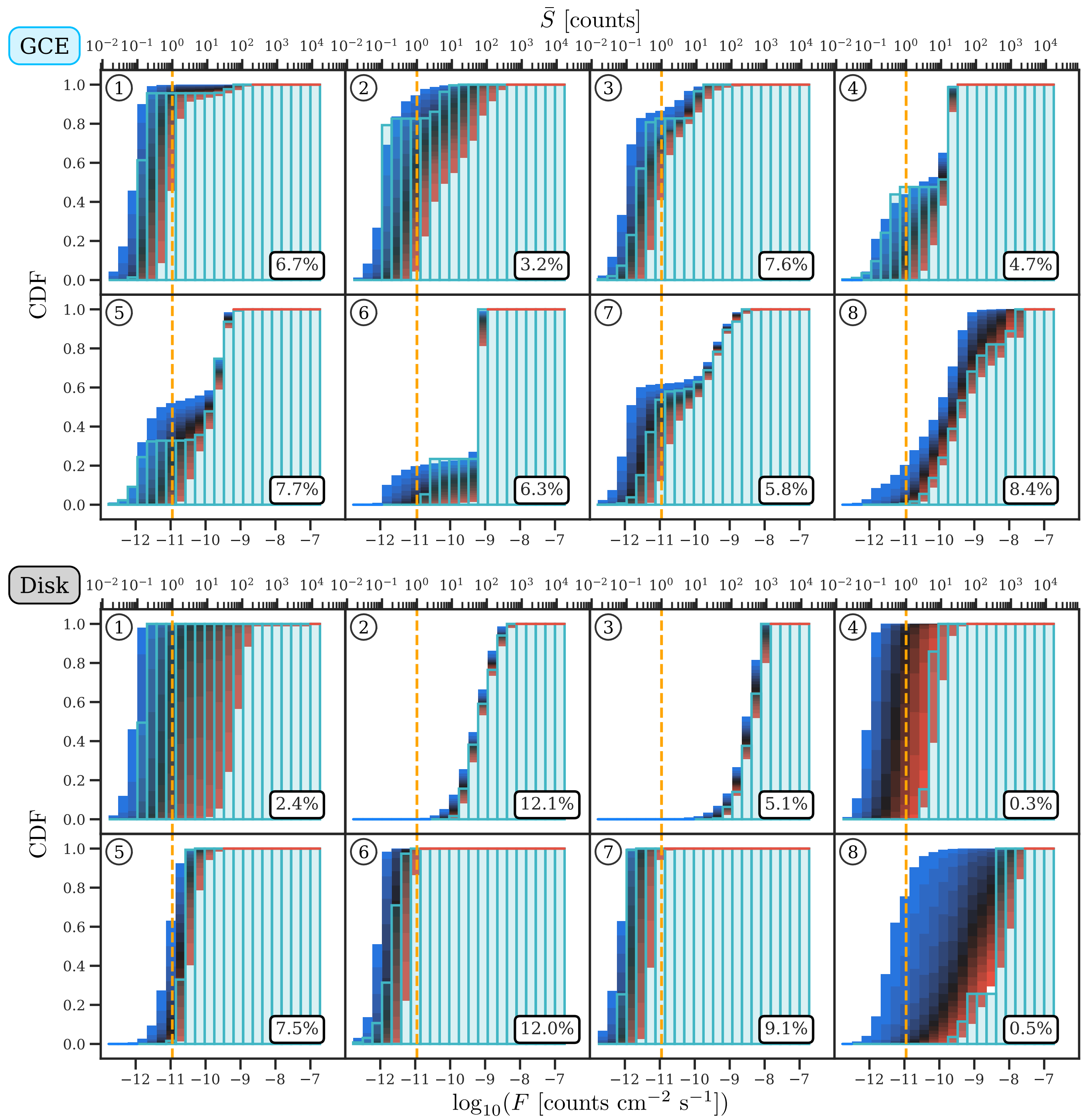
Step 1: Performance on MC



[List, NLR, Lewis 2021]

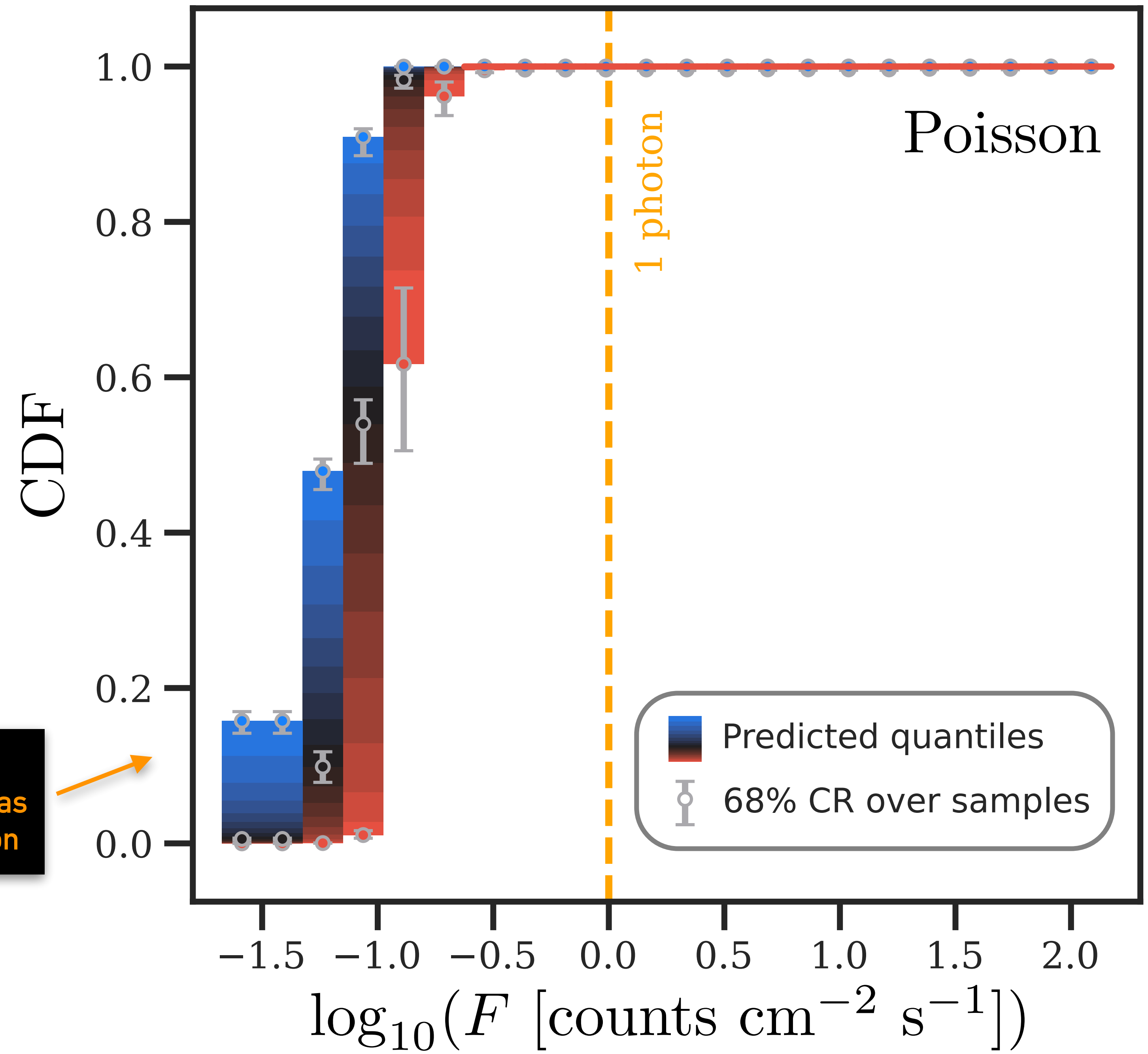
A MACHINE LEARNING APPROACH

Step 2: Performance on MC



[List, NLR, Lewis 2021]

Step 2: Poissonian MC

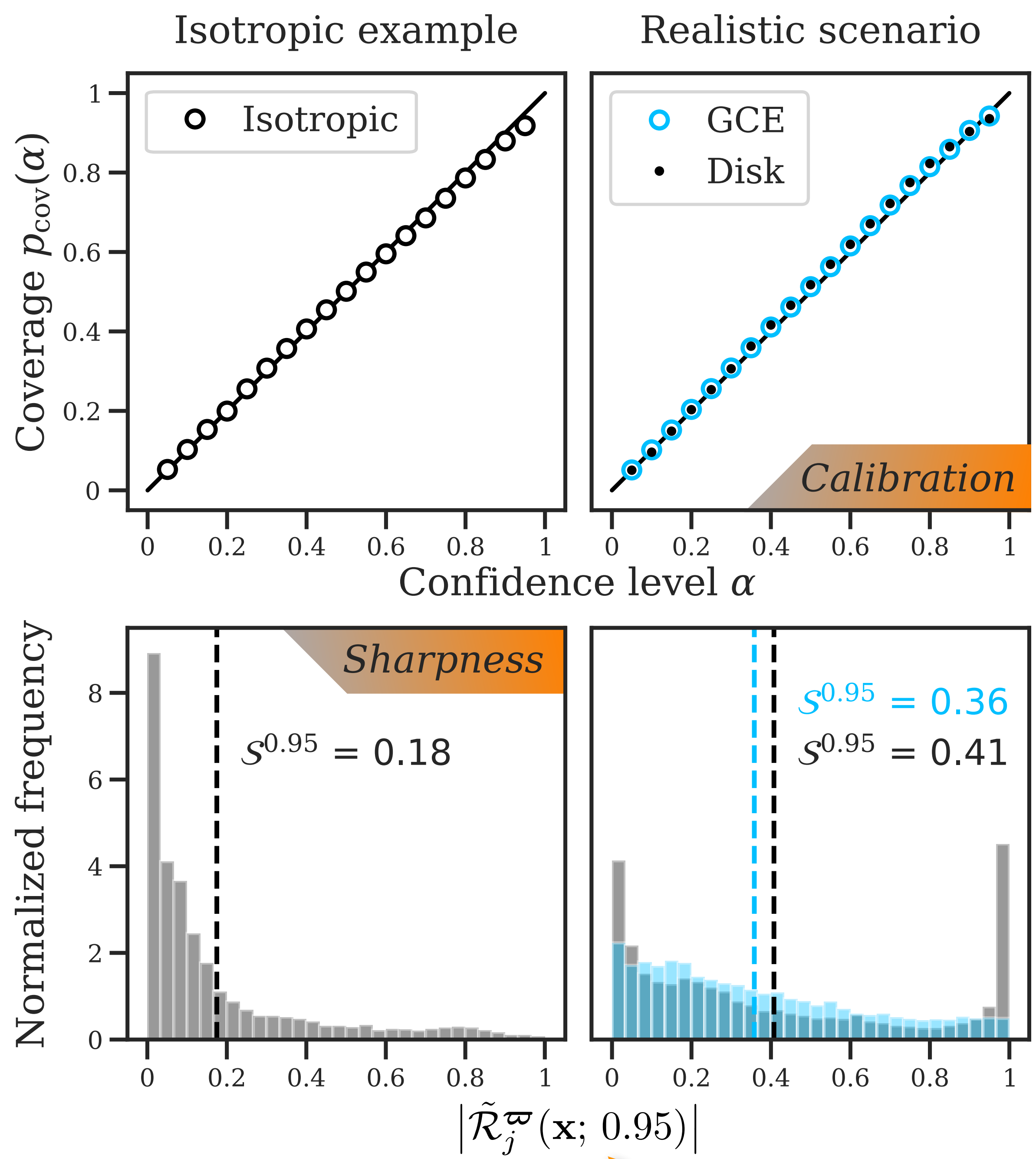


Poisson MC reconstructed as $F_{\rm PS} \ll 1$ photon

[List, NLR, Lewis 2021]

Step 2: Calibration

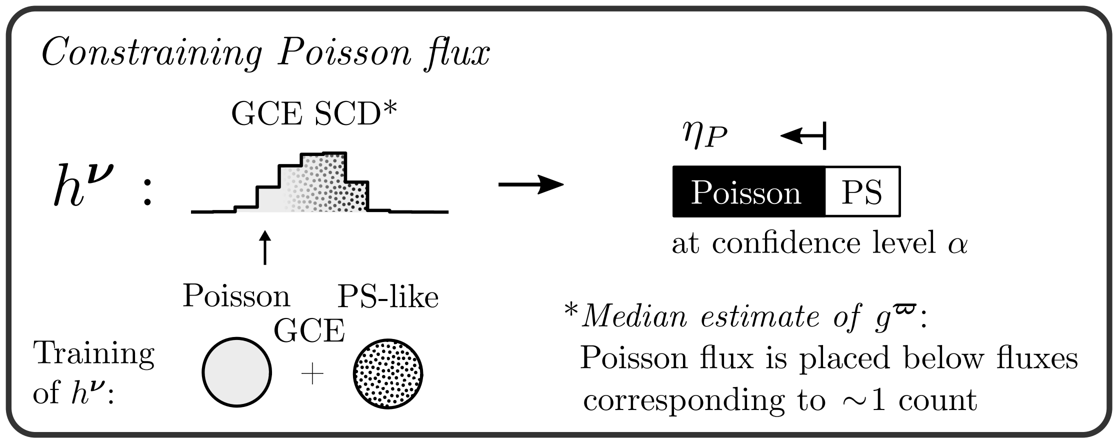
How often does the true value lie within inferred quantile range

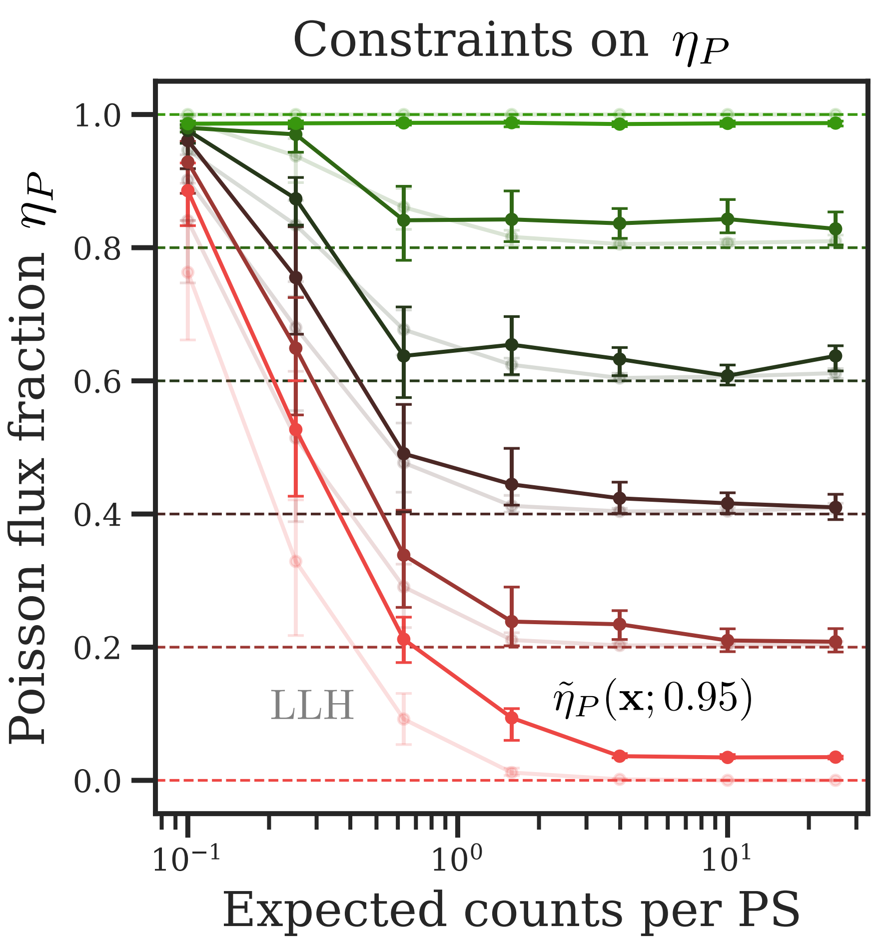


How wide is the 95%-IQR?

[List, NLR, Lewis 2021]

Step 3: Constraining η_P



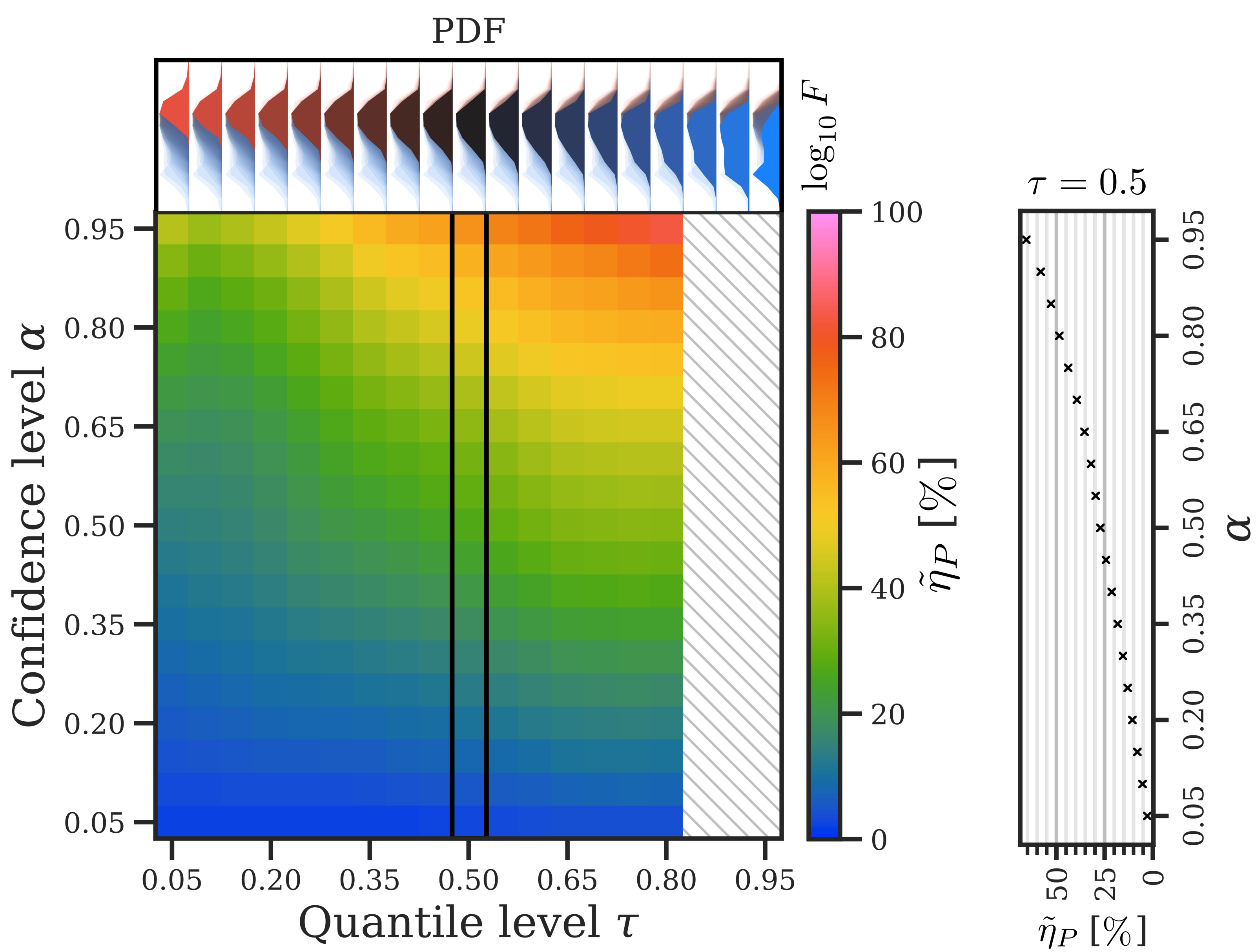


Bright PSs are inconsistent with Poisson emission
Can constrain the Poisson flux fraction η_P from the data

Performance in isotropic MC with no PSF (where LLH approach is correct)

[List, NLR, Lewis 2021]

Step 3: Constraining η_P

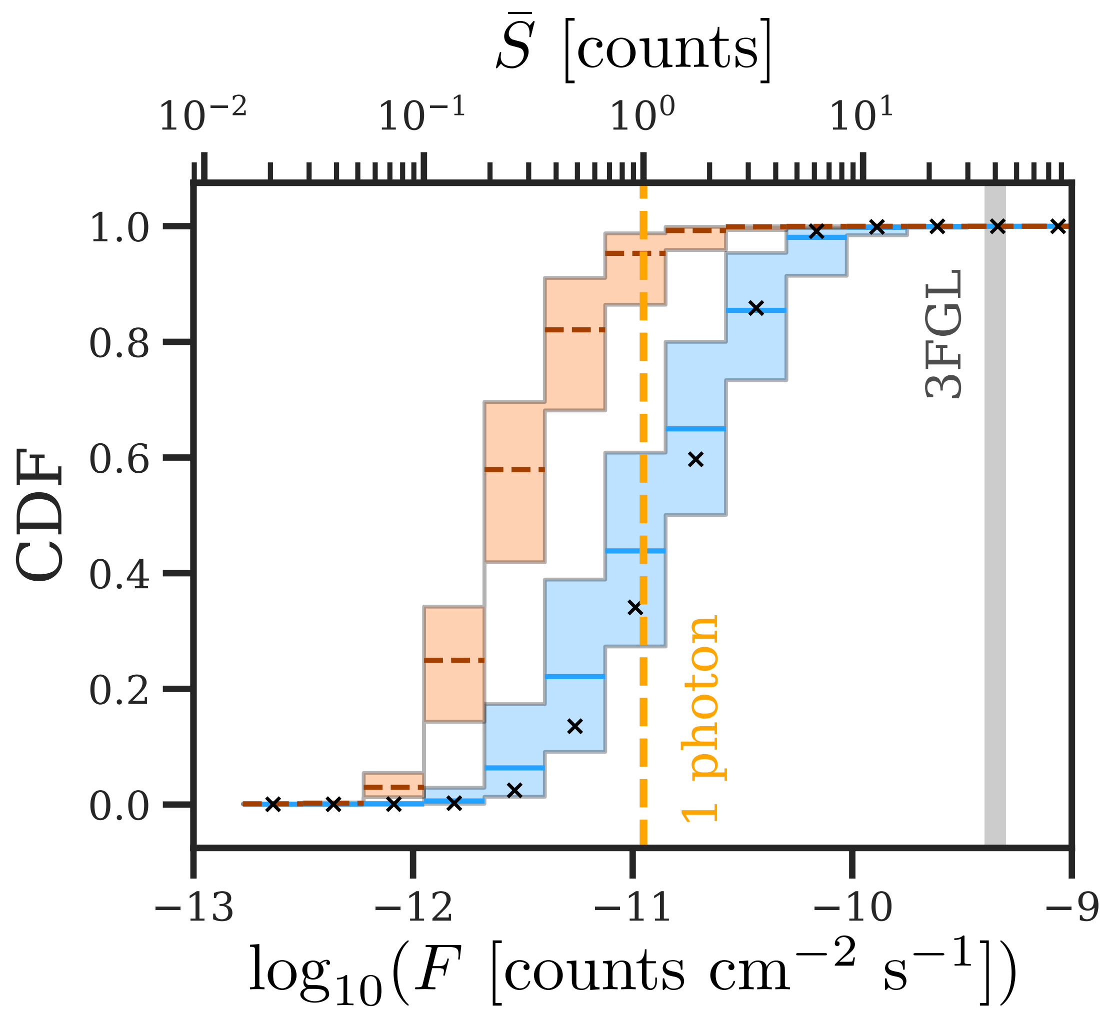


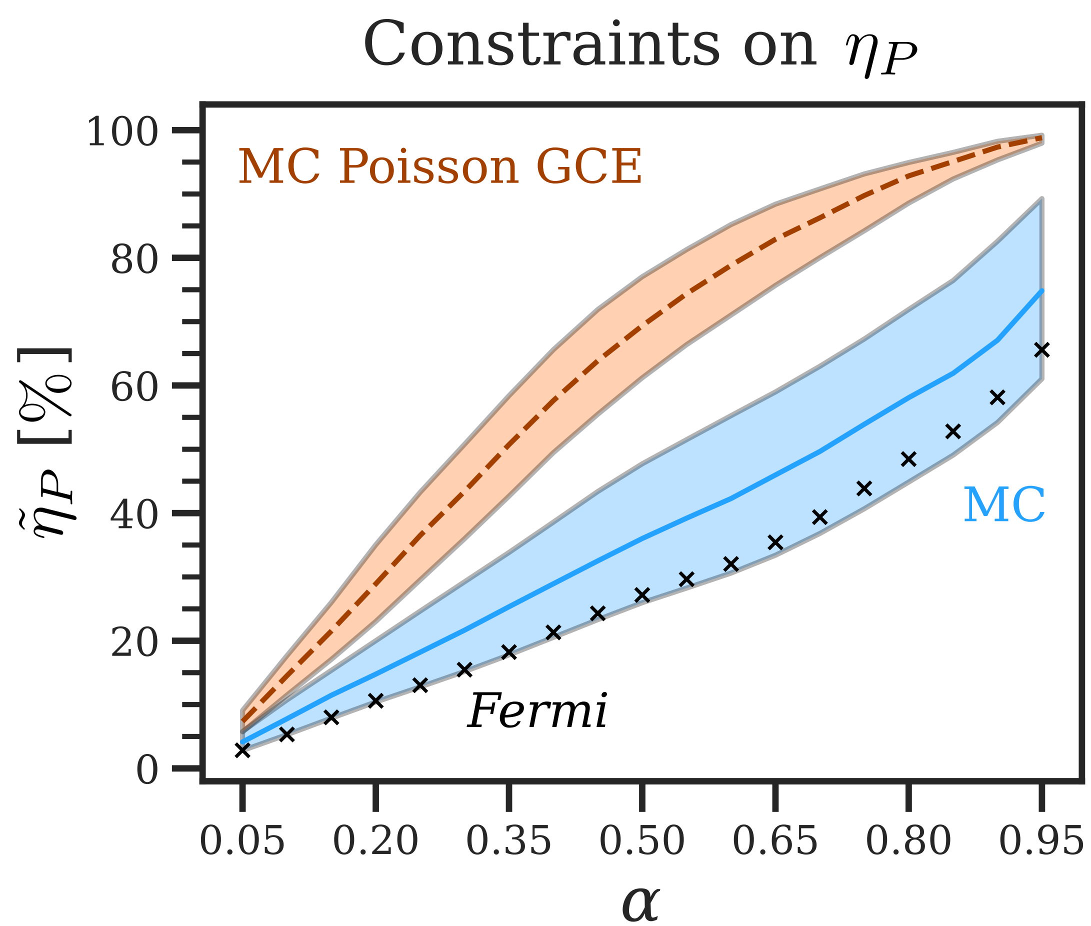
In the Fermi data: at 95% C.L. $\eta_P < 66\%$

[List, NLR, Lewis 2021]

Systematic Checks

Are the data results consistent with the equivalent MC predictions?



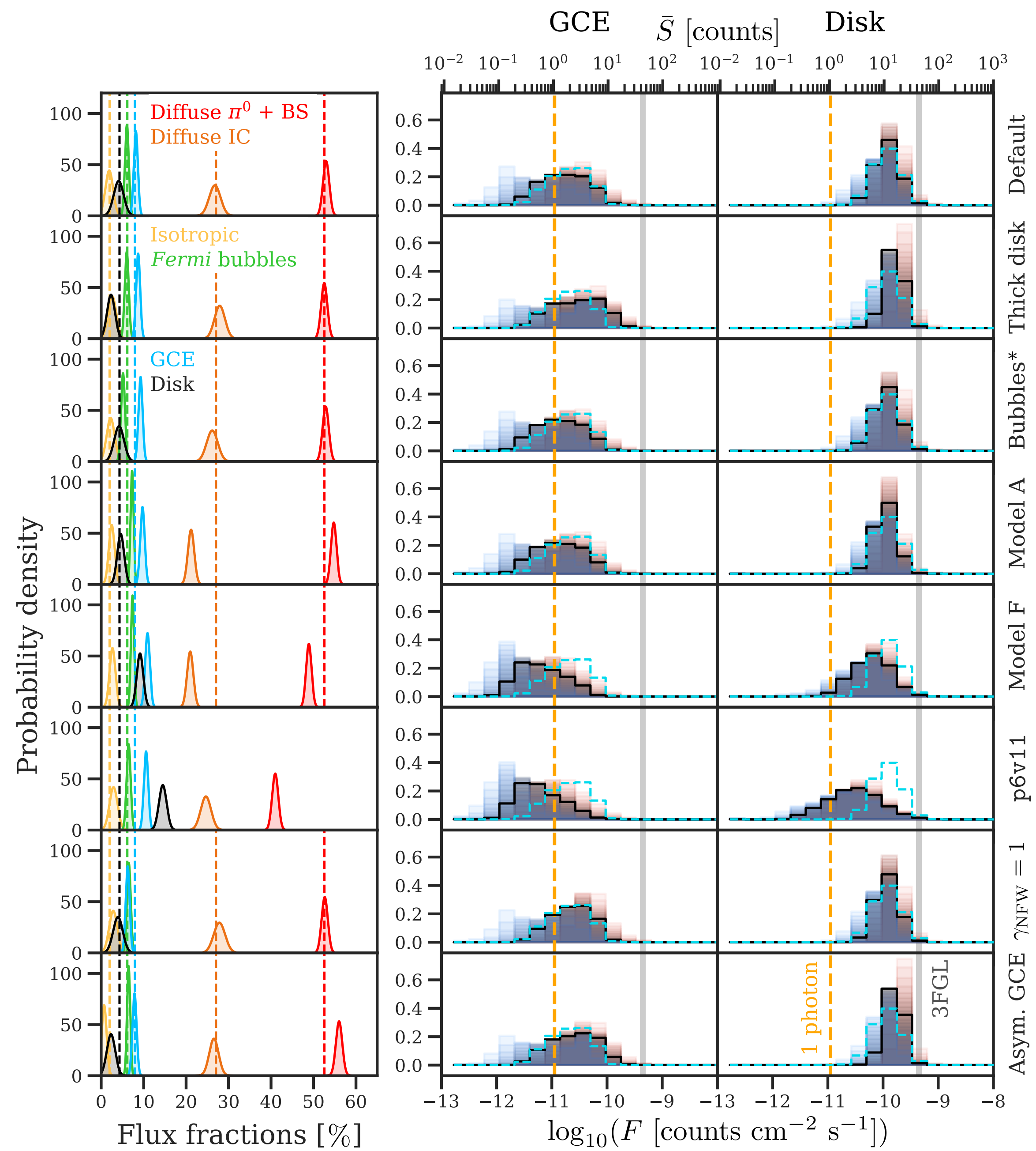


×: results from the real Fermi data

[List, NLR, Lewis 2021]

Systematic Checks

Performance with mismodeling

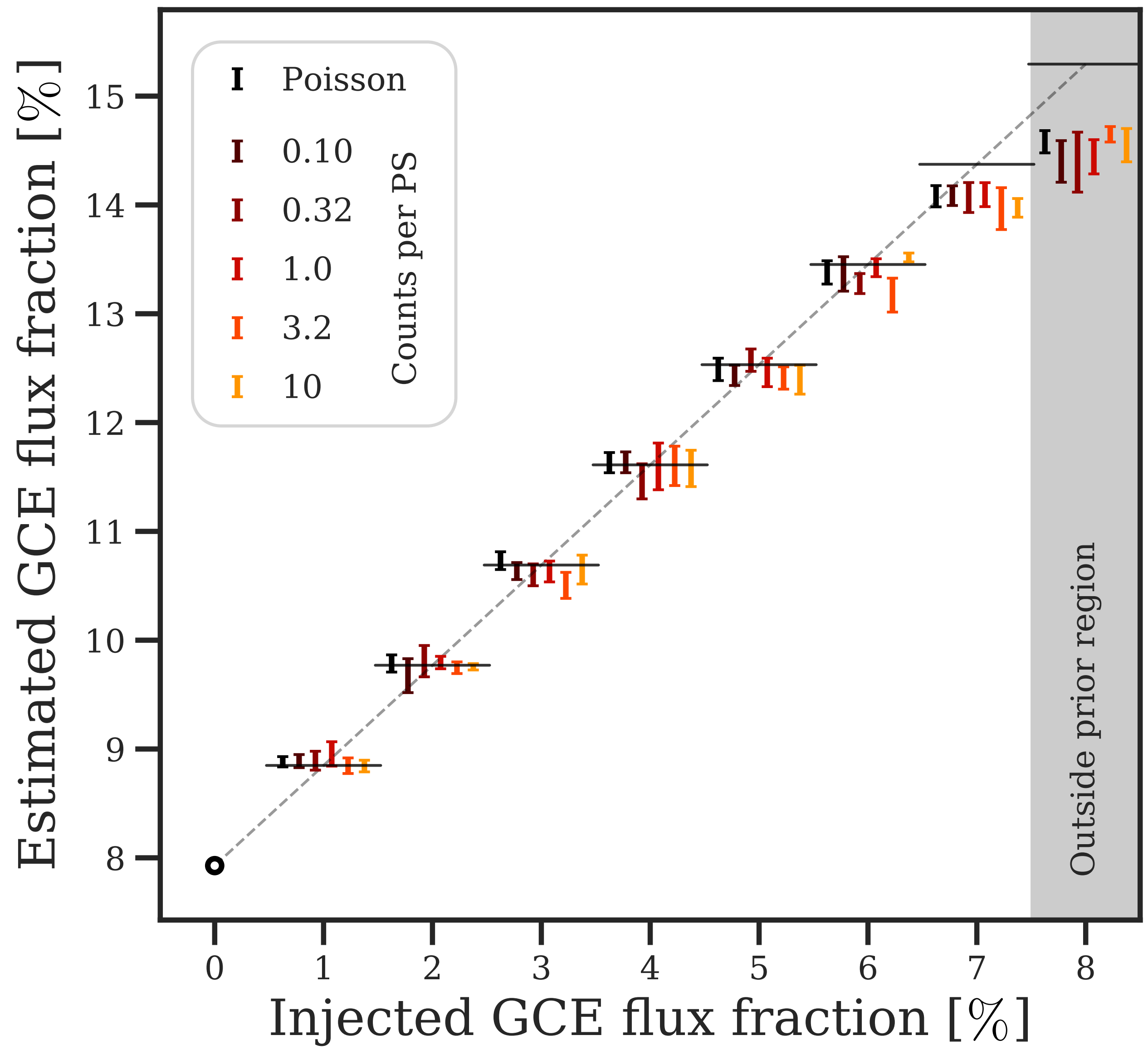


Diffuse mismodeling reconstructed as dimmer SCD - exact opposite to NPTF

[List, NLR, Lewis 2021]

Systematic Checks

Recovery of injected signal into Fermi data

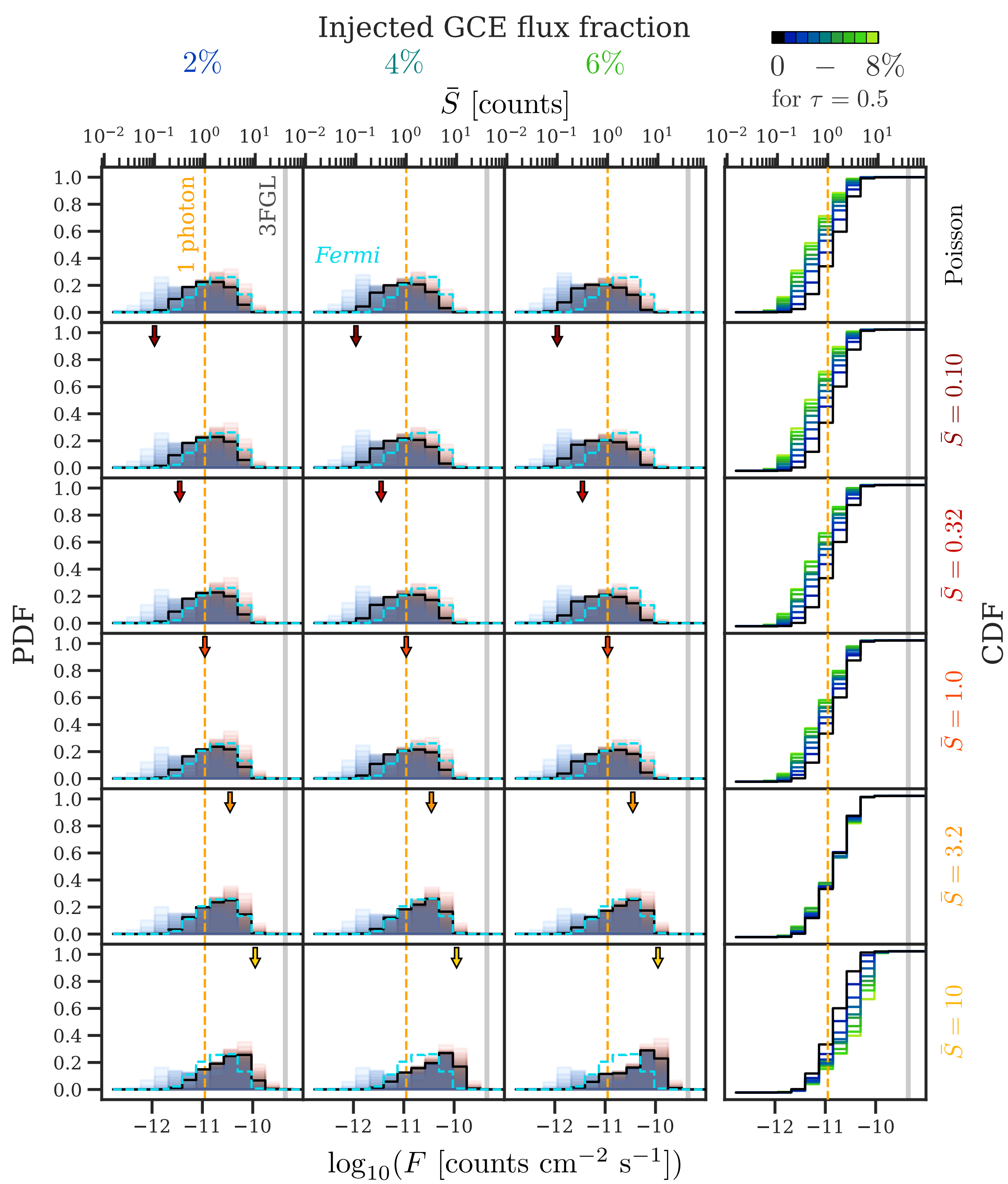


Correctly reconstructed until GCE brighter than training data

[List, NLR, Lewis 2021]

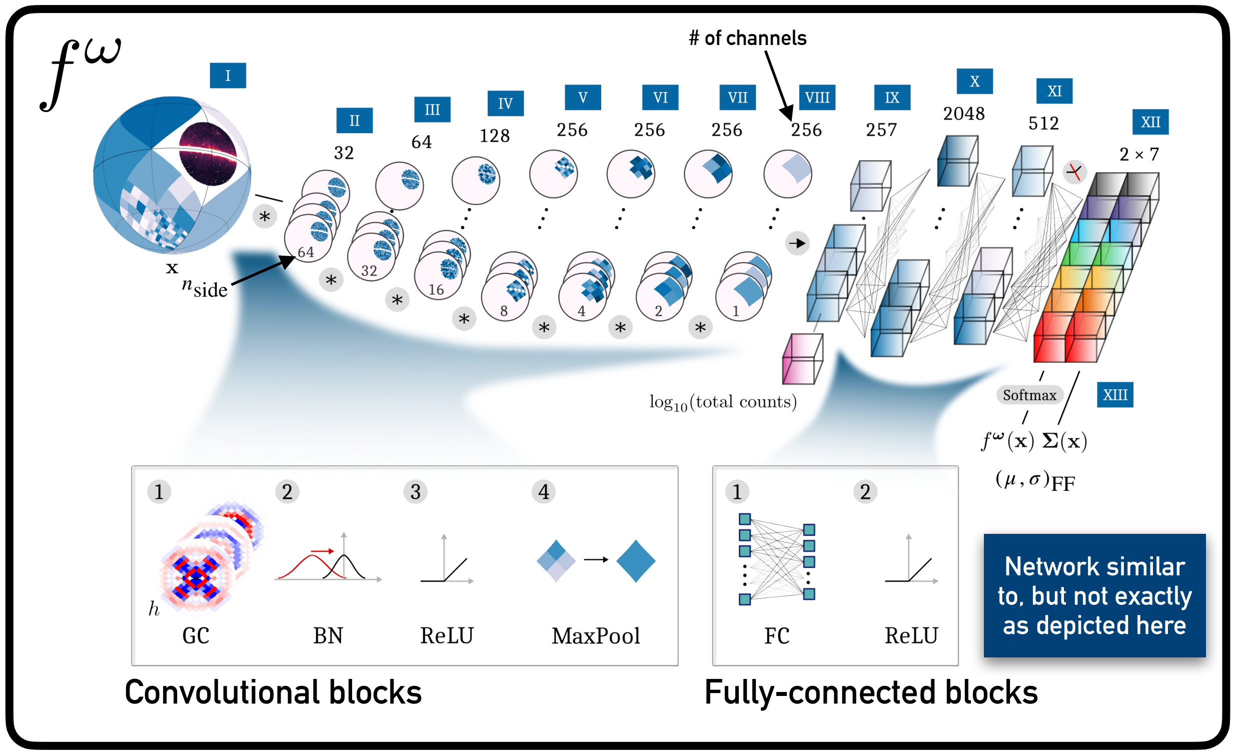
Systematic Checks

Recovery of injected signal into Fermi data



[List, NLR, Lewis 2021]

Network parameters



f^ω (map → template flux fractions):

Layer	Operations	Output shape	Output n_{side}	Trainable parameters
I	Input map (normalized)	30,805 × 1	256	–
II	ConvBlock	8,117 × 32	128	160 + 32
III	ConvBlock	2,199 × 64	64	10,240 + 64
IV	ConvBlock	598 × 128	32	40,960 + 128
V	ConvBlock	164 × 256	16	163,840 + 256
VI	ConvBlock	50 × 256	8	327,680 + 256
VII	ConvBlock	14 × 256	4	327,680 + 256
VIII	ConvBlock	4 × 256	2	327,680 + 256
IX	ConvBlock	1 × 256	1	327,680 + 256
X	Append $\log_{10}(S_{\mathrm{tot}})$	1 × 257		–
XI	ReLU ∘ FC	1 × 2,048		526,336 + 2,048
XII	ReLU ∘ FC	1 × 512		1,048,576 + 512
XIII	Reshape ∘ FC	2 × 6		6,144 + 0
XIV	Softmax (means only)	2 × 6		–
				3,111,040

g^ϖ (map → SCD histograms):

Layer	Operations	Output shape	Output n_{side}	Trainable parameters
I	Input map (normalized)	30,805 × 2	256	–
II	ConvBlock	8,117 × 32	128	320 + 32
III	ConvBlock	2,199 × 64	64	10,240 + 64
IV	ConvBlock	598 × 128	32	40,960 + 128
V	ConvBlock	164 × 256	16	163,840 + 256
VI	ConvBlock	50 × 256	8	327,680 + 256
VII	ConvBlock	14 × 256	4	327,680 + 256
VIII	ConvBlock	4 × 256	2	327,680 + 256
IX	ConvBlock	1 × 256	1	327,680 + 256
X	Append $\log_{10}(S_{\mathrm{tot}})$	1 × 257		–
XI	Append τ	1 × 258		–
XII	ReLU ∘ FC	1 × 2,048		528,384 + 2,048
XIII	ReLU ∘ FC	1 × 512		1,048,576 + 512
XIV	Reshape ∘ FC	2 × 22		22,528 + 0
XV	Normalized softplus	2 × 22		–
				3,129,632

h^ν (GCE SCD histogram → Poissonian flux fraction η_P):

Layer	Operations	Output shape	Trainable parameters
I	Input histogram	22	–
II	Append α	23	–
III	ReLU ∘ FC	256	5,888 + 256
IV	ReLU ∘ FC	256	65,536 + 256
V	Sigmoid ∘ FC	1	256 + 1
			72,193

[List, NLR, Lewis 2021]