

Characterizing the expected behavior of non-Poissonian template fitting

Luis Gabriel C. Bariuan^{1,*} and Tracy R. Slatyer^{1,2,†}

¹*Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

²*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

 (Received 30 August 2022; accepted 12 April 2023; published 9 May 2023)

We have performed a systematic study of the statistical behavior of non-Poissonian template fitting (NPTF), a method designed to analyze and characterize unresolved point sources in general counts datasets. In this paper, we focus on the properties and characteristics of the Fermi-LAT gamma-ray data set. In particular, we have simulated and analyzed gamma-ray sky maps under varying conditions of exposure, angular resolution, pixel size, energy window, event selection, and source brightness. We describe how these conditions affect the sensitivity of NPTF to the presence of point sources, for inner-galaxy studies of point sources within the Galactic Center excess, and for the simplified case of isotropic emission. We do not find opportunities for major gains in sensitivity from varying these choices, within the range available with current Fermi-LAT data. We provide an analytic estimate of the NPTF sensitivity to point sources for the case of isotropic emission and perfect angular resolution, and find good agreement with our numerical results for that case.

DOI: [10.1103/PhysRevD.107.103014](https://doi.org/10.1103/PhysRevD.107.103014)

I. INTRODUCTION

Recent years have seen a number of efforts to apply photon pixel count statistics to gamma-ray data, in order to characterize populations of point sources (PSs) too faint to be individually detected at high significance (e.g. [1–11]). The general idea of these methods is to exploit the fact that an unmodeled PS population gives rise to non-Poissonian fluctuations in the number of photons per pixel, with “hot spots” corresponding to the locations of sources. Even if no individual hot spot is significant enough to be established as a PS with high probability, the distribution of fluctuations can be used to infer the properties of the population. These methods have been applied to characterize contributions to the extragalactic gamma-ray background (e.g. [5,7,8]) and to study inner Galaxy PS populations (e.g. [3,4,12]); they have also been applied to other datasets, e.g. crowded optical fields [10] and high-energy neutrinos [13].

Initially these methods focused on the case of isotropic PS populations, which is likely to be a good approximation for all-sky background radiation generated by a large ensemble of faint extragalactic sources. However, subsequent studies [3,9,11] extended this approach to the case of source populations with an arbitrary spatial distribution.

In this work we focus on one such method, *non-Poissonian template fitting* (NPTF) [2,3,14], which has been applied in a range of contexts but particularly to study

the Galactic Center excess (GCE) in public data from the *Fermi* Gamma-Ray Space Telescope (hereafter *Fermi*). The GCE is an extended and roughly spherical (not disklike) source of GeV-scale gamma rays filling the region within 1.5 kpc of the Galactic Center (GC) [15–21].

The origin of the GCE has been the subject of active controversy for the past decade, with two explanations receiving the most attention. One possibility is that the GCE originates from diffuse particle dark matter (DM) undergoing annihilation (e.g. [15,19,22]), as the flux, energy spectrum, and spatial morphology of the GCE appear broadly consistent with a DM origin. If this hypothesis were confirmed, it would be a discovery of profound importance, representing the first evidence of nongravitational interactions between DM and visible particles. However, the energy spectrum of the GCE also closely resembles that of gamma-ray pulsars observed by *Fermi*, and a number of studies have found that the spatial morphology of the GCE is a closer match to the stellar bulge than to a DM annihilation signal [23–26].¹ For these reasons, it seems plausible that the GCE represents the detection of a pulsar population in the Galactic bulge (e.g. [29–40]). If this population includes sources with brightness approaching the *Fermi* sensitivity threshold, then NPTF methods have the potential to characterize at least the

*lbariuan@mit.edu

†tslatyer@mit.edu

¹However, other recent studies [27,28] have found the opposite preference; the result appears to be sensitive to how the Galactic background emission is modeled.

bright end of this new population, and provide strong evidence against the DM hypothesis.

Previous NPTF studies have claimed evidence for a GCE source population comprised of relatively bright and rare PSs [3], but recent studies have found that those claims may have been premature due to unaccounted-for systematic errors [41–45]. Other analyses have found a preference for a significant diffuse emission component [12,46], although this does not exclude the pulsar hypothesis, since the sources might simply be too faint to be detected with current methods. At the same time, work on modeling the pulsar population in the bulge has suggested that plausible pulsar luminosity functions could generate very few *Fermi* detected sources, while yielding an appreciable number of sources in the flux range potentially detectable by NPTF methods [47–49] or related approaches using machine learning [50,51].

Given this uncertain situation, it is timely to understand how well NPTF can be expected to perform in detecting faint PS populations, and how this performance can be optimized by analysis choices. For example, many previous studies have chosen *Fermi* event selections to optimize angular resolution, at the cost of exposure. While several studies have explored the effect on their results of varying the event selection (e.g. [5,44,45]), this has not yet been done in a systematic way.

In this work, we systematically explore the ability of the public NPTFit algorithm (as described in Ref. [14]) to reconstruct faint sources in simulated data, as a function of the instrument capabilities and analysis choices. We focus primarily on the analysis of the inner Milky Way, as relevant for the GCE, but also provide results for the simpler case where signal and background are both isotropic.

We begin in Sec. II by discussing how we expect the likelihood ratio in favor of a point-source population to behave, in a simplified approximate context that can be treated analytically, by approximating some or all of the relevant Poisson distributions as Gaussian. This approximation is not expected to hold in detail in the cases of greatest interest to us, but it is helpful for building intuition.

In Sec. III, we then move on to our numerical study, starting by discussing the procedure by which we perform fits to the real *Fermi* data to derive reasonable baseline estimates for the properties of the background model and PSs. We use these results to generate simulated data that is similar to the true gamma-ray sky as observed by *Fermi*, using the public code NPTFit-Sim, a package designed to simulate populations of unresolved PSs.² In this section we also discuss our methodology for fitting to simulated data, and the test statistic we will use to describe the sensitivity of NPTF methods to faint sources.

In Sec. IV we lay out the parameters we will vary in our simulations: exposure, angular resolution, energy window,

pixel size, and source brightness. We describe the procedure for varying each of these parameters using NPTFit and NPTFit-Sim, including any associated modifications to the prior ranges.

In Sec. V we perform an initial analysis and comparison between simulated data and our analytic approximations, in the simplified scenario where the PS and smooth contributions to the gamma-ray sky are both isotropic.

We then move on to a full realistic inner Galaxy analysis; conduct variations of the various analysis parameters, singly and in combination; and present the (numerical) results in Sec. VI. In particular, we explore the individual effects of varying the exposure level and the point spread function (PSF), and map out the tradeoff when exposure level is increased (reduced) with the effect of worsening (improving) angular resolution, using the specific examples of *Fermi* event classes sorted by angular resolution. Modifying the energy window varies the effective exposure, the PSF, and also (in real data) the relative amplitude of the various background and signal components; we explore these effects independently. We then demonstrate the effect of varying the brightness of the PSs while keeping the total flux of the population constant (as appropriate for hypothetical source populations that explain the bulk of the GCE). Finally, we examine the question of the optimal pixel size for NPTF analyses, exploring both the sensitivity to faint PSs and accuracy of the parameter reconstruction.

In Sec. VII we summarize our results and discuss some implications for NPTF analyses of *Fermi* gamma-ray data in the inner Galaxy.

In Appendix A we present further details of our simulation parameters and fitting methodology; in Appendix B we discuss the degree to which our source count functions model a single-brightness PS population; in Appendix D we show additional results for the simpler case where both signal and background are isotropic; and in Appendix E we show the results of using an alternative Galactic diffuse model as the basis for our simulations.

II. ANALYTIC APPROXIMATIONS FOR NON-POISSONIAN TEMPLATE FITTING

Let us begin by building some intuition for how the detectability of PSs is likely to scale in a NPTF-like setup. We will initially follow the approach of Ref. [45], essentially replacing the Poisson distributions with Gaussians; this will be a good approximation when the number of sources per pixel and number of counts/source are both large, and can more generally provide qualitative insights into how various inputs affect the PS sensitivity.

Here we will compute likelihoods and likelihood ratios as a measure of sensitivity, whereas in the numerical analysis of later sections we will perform a Bayesian analysis and evaluate Bayes factors. The Bayesian evidence is an integral of the likelihood weighted by the priors, and

²<https://github.com/nickrodd/NPTFit-Sim>.

TABLE I. Summary of $\langle \Delta \ln \mathcal{L} \rangle$ and $\text{Var}(\Delta \ln \mathcal{L})$ for a single pixel (as calculated in Sec. II), where the log likelihood difference is evaluated between two Gaussian models for $P(N)$ with parameters (X, σ^2) and (Y, τ^2) . In the upper part of the table, the true model is assumed to be the Gaussian with parameters (X, σ^2) , and we consider two regimes characterized by the detectability parameter $\delta \equiv \sigma^2/\tau^2$. In the lower part of the table, we rewrite σ, τ, X and Y in terms of parameters describing the PS population (see text), where k denotes the proportion of the total emission that is attributed to the PS population, s denotes the expected number of photons per source, and n_0 is the number of sources per pixel. In this part of the table we assume $\delta = ks \gg 1$. In the final line, we furthermore employ a more accurate approximation for the true model for $P(N)$ rather than assuming it to be Gaussian (see Sec. IIF).

	$\langle \Delta \ln \mathcal{L} \rangle$	$\text{Var}(\Delta \ln \mathcal{L})$
<i>General Gaussian Model Comparison</i>		
High Detectability $\delta \gg 1$	$\frac{1}{2}[\delta - \ln(1 + \delta)]$	$2\langle \Delta \ln \mathcal{L} \rangle^2$
Low Detectability $\delta \ll 1$	$\frac{\delta^2}{4}$	$2\langle \Delta \ln \mathcal{L} \rangle$
<i>PS Signal + Diffuse Background, $ks \gg 1$</i>		
Gaussian probability distribution	$\frac{1}{2}[ks - \ln(1 + ks)]$	$2\langle \Delta \ln \mathcal{L} \rangle^2$
More accurate probability distribution: accounting for rare sources	$\frac{1}{2}[ks - \ln(1 + ks)]$	$\langle \Delta \ln \mathcal{L} \rangle^2 (2 + \frac{1}{n_0})$

so loosely speaking we expect them to have qualitatively similar properties under variations of the source brightness, exposure, etc. However, our expressions for the likelihood ratios should not *a priori* be expected to accurately approximate the Bayes factors, since Bayes factors incorporate information from the priors (including the number of free parameters in the model), and the likelihood ratios do not.³ We summarize the key results within this section in Table I.

A. Pixel likelihood to observe N photons

Let us first review some relevant results from Ref. [45]. Consider a simplified scenario where our PS population model predicts n_0 sources per pixel, and all sources are identical, with an expected number of photons per source of s . For the moment, we will ignore leakage out of the pixel due to the nontrivial angular resolution, but as a first approximation the effect of such leakage would be to reduce s . We are interested in calculating the probability to observe N photons in a pixel.

If we fix the number of observed sources (in a given pixel) to be n , then the total number of photons in the pixel will follow a Poisson distribution with mean ns . For $ns \gg 1$, we can approximate this distribution as a Gaussian with a mean and variance of ns , via the central limit theorem. Then the probability to observe N photons is given approximately by:

$$P(N|\{n, s\}) = \frac{1}{\sqrt{2\pi ns}} e^{-(N-ns)^2/(2ns)}. \quad (1)$$

³However, in practice, we will find that for our default choice of priors, the differences between the likelihood ratios and Bayes factors are small compared with other differences between the analytic and numerical results.

As a note, this expression can be thought of as a continuous probability density function (PDF), but also as a measure of the finite probability to observe N photons by integrating the PDF over a bin of width $dN = 1$ (i.e. the difference between adjacent values of N). Provided the PDF does not vary rapidly over the bin, this integral can simply be approximated by the value of the PDF at the center of the bin. We will use both interpretations of $P(N|\{n, s\})$ and similar quantities in the following calculations.

This distribution function is convolved with $P(n|n_0)$, a distribution that describes the probability of drawing n sources given that the expected number of sources is n_0 . The resulting function, which we denote $P(N|n_0, s)$, describes the likelihood of obtaining N photons given that the number of sources is described by a Poisson distribution with an expectation value of n_0 . If the number of sources is large, we can also approximate the distribution that describes the number of sources with a Gaussian with mean and variance n_0 . Furthermore, the integrand is dominated by the region where $ns \approx N$, so we can set $ns \approx N$ except where $N - ns$ appears in an exponent.

These approximations yield the following equation for the probability to observe N photons given n_0 and s :

$$\begin{aligned} P(N|\{n_0, s\}) &= \int dn P(N|\{n, s\}) P(n|n_0) \\ &\approx \int dn \frac{1}{\sqrt{2\pi N}} e^{-(ns-N)^2/(2N)} \\ &\quad \times \frac{1}{\sqrt{2\pi n_0}} e^{-(n_0-n)^2/(2n_0)}. \end{aligned} \quad (2)$$

The integral over n can be performed analytically and takes a simple form, if we assume the peak in the integrand is sufficiently far away from the limits of integration that we can take those limits to $\pm\infty$ without affecting the result. Furthermore, around the peak of the probability distribution

we have $N \approx n_0 s$, so we can approximate $N \approx n_0 s$ except when $N - n_0 s$ appears in an exponent. These approximations yield:

$$\begin{aligned} P(N|\{n_0, s\}) &\approx \frac{1}{\sqrt{2\pi(N + n_0 s^2)}} e^{\frac{-(N - n_0 s)^2}{2(N + n_0 s^2)}} \\ &\approx \frac{1}{\sqrt{2\pi n_0 s(1 + s)}} e^{\frac{-(N - n_0 s)^2}{2n_0 s(1 + s)}}. \end{aligned} \quad (3)$$

That is, under these approximations the probability of observing N photons takes a Gaussian form (at least near the peak of the distribution), but with an inflated variance of $n_0 s(1 + s)$, a factor of $(1 + s)$ greater than the expectation value of $n_0 s$.

If $s \ll 1$, our model corresponds to a very faint source population that should be indistinguishable from diffuse emission. In this case, we recover the standard Gaussian approximation to the Poisson distribution, with equal mean and variance of $n_0 s$,

$$P(N|\{n_0, s\}) \approx \frac{1}{\sqrt{2\pi n_0 s}} e^{\frac{-(N - n_0 s)^2}{2n_0 s}}. \quad (4)$$

Thus the characteristic feature of a PS population (within these approximations) is an enhanced variance, by a factor of $1 + s$.

In the event that the number of counts per source satisfies $s \gg 1$ but the number of sources per pixel $n_0 \lesssim 1$, a more refined approximation for the distribution may be useful (going beyond the results of Ref. [45]). If i sources are drawn in a given pixel (i being an integer), the number of counts from those sources will be Poisson-distributed with expectation si ; for $s \gg 1$ and $i > 0$, we can approximate each of these individual distributions as a Gaussian with mean and variance si , so overall we have:

$$P(N|\{n_0, s\}) \approx p_0(n_0)\delta_{N,0} + \sum_{i=1}^{\infty} p_i(n_0) \frac{1}{\sqrt{2\pi si}} e^{\frac{-(N - si)^2}{2si}}, \quad (5)$$

where $p_i(n_0)$ is the Poisson probability of drawing i sources when n_0 are expected.

B. Likelihood ratio between models (Gaussian approximation)

Now suppose the true underlying model for a given pixel yields a Gaussian distribution for N with mean X and variance σ^2 . We wish to evaluate the expected log likelihood ratio between the correct model and an alternative model that predicts mean Y and variance τ^2 . We will denote these models respectively as (X, σ^2) and (Y, τ^2) . This result has been computed previously in Ref. [45]; we review it here.

For context, the correct model might represent a linear combination of a PS population and a diffuse signal, while the alternative model allows only for a diffuse signal; the expected log likelihood ratio in this case then gives a measure of how well we will be able to exclude the all-diffuse model and thus detect the PS population. We will work out the case for general (X, Y, σ^2, τ^2) first, under the approximation where all the relevant probability distributions are Gaussian, and then apply this general result to several scenarios in which we might wish to detect PS populations.

For a single pixel, the probability of finding N photons predicted by the model (Y, τ^2) is

$$\mathcal{L} = P(N|\{Y, \tau^2\}) = e^{[-(N - Y)^2]/(2\tau^2)} / \sqrt{2\pi\tau^2} \quad (6)$$

corresponding to a log likelihood of $\ln \mathcal{L} = -\frac{(N - Y)^2}{2\tau^2} - \frac{1}{2} \ln 2\pi\tau^2$. To get the expected value of the log likelihood with respect to the true model, we can integrate against the true distribution of N , i.e. $P(N|\{X, \sigma^2\})$. This yields [45]:

$$\langle \ln \mathcal{L}(Y, \tau^2) \rangle = -\frac{[(X - Y)^2 + \sigma^2]}{2\tau^2} - \frac{1}{2} \ln(2\pi\tau^2). \quad (7)$$

Note we use the $\langle \rangle$ notation generally to denote expected values with respect to the true model.

Now we diverge from Ref. [45], which focused on determining the best-fit choice for τ^2 given a discrepancy between X and Y . Let us instead simply examine the expected $\Delta \ln \mathcal{L}$ between the fitted model (Y, τ^2) and the best-fit model (X, σ^2) , which is given by:

$$\begin{aligned} \langle \Delta \ln \mathcal{L} \rangle &\equiv \langle \ln \mathcal{L}(X, \sigma^2) - \ln \mathcal{L}(Y, \tau^2) \rangle \\ &= \frac{[(X - Y)^2 + \sigma^2]}{2\tau^2} + \frac{1}{2} \ln(2\pi\tau^2) - \frac{1}{2} - \frac{1}{2} \ln(2\pi\sigma^2) \\ &= \frac{[(X - Y)^2 + \sigma^2]}{2\tau^2} - \frac{1}{2} \left[1 + \ln\left(\frac{\sigma^2}{\tau^2}\right) \right]. \end{aligned} \quad (8)$$

If both models produce a very similar expected number of photons, i.e. $X \approx Y$, and differ only in their variances, then this result can be simplified to:

$$\langle \Delta \ln \mathcal{L} \rangle = \frac{\sigma^2}{2\tau^2} - \frac{1}{2} \left[1 + \ln\left(\frac{\sigma^2}{\tau^2}\right) \right]. \quad (9)$$

Note however that if Y and τ^2 are allowed to vary within certain limits or while satisfying certain conditions, then it is not guaranteed that the best-fit point lies at $Y = X$; if the global likelihood maximum (at $Y = X, \tau^2 = \sigma^2$) cannot be attained, then the best-fit value of Y will depend on the value of τ^2 (and vice versa). Most simply, this can occur

when the model is Poissonian, in which case τ^2 is fixed to Y , but the data has non-Poissonian components and so σ^2 differs from X in the true underlying model. A related scenario, studied in Refs. [44,45], occurs when the model requires the same value of Y in multiple pixels but the true underlying model varies across those pixels; this leads to a best-fit model variance τ^2 that differs from the true underlying variance σ^2 (possibly leading to misattribution of the enhanced variance to a PS population).

C. Variance between realizations (Gaussian approximation)

In addition to working out the expected log likelihood ratio as a measure of sensitivity to incorrect modeling (such as attempting to describe PSs with a Poissonian template), it is helpful to understand the expected variability in this ratio between different realizations. In the limit where the number of pixels is large, the total $\Delta \ln \mathcal{L}$ for the image is the sum of many independent random variables ($\Delta \ln \mathcal{L}$ for each pixel), and so is expected to follow a Gaussian probability distribution by the central limit theorem (even if the probability distribution for $\Delta \ln \mathcal{L}$ in a single pixel is highly non-Gaussian). Consequently, in this limit, we expect the distribution of the total $\Delta \ln \mathcal{L}$ (summed over pixels) to be well characterized by its expectation value and variance.

As in the previous subsection, we will work out the result initially for general choices of the PDF parameters for the true and alternative hypotheses, (X, Y, σ^2, τ^2) . We will then apply these results to specific scenarios, in particular where the true model [described by (X, σ^2)] includes a PS component but the alternative model [described by (Y, τ^2)] does not.

We can estimate the variance of $\Delta \ln \mathcal{L}$ by evaluating $\text{Var}(\Delta \ln \mathcal{L}) \equiv \langle (\Delta \ln \mathcal{L})^2 \rangle - \langle \Delta \ln \mathcal{L} \rangle^2$. Let us first focus on the case where $\sigma \gg \tau$ and $\langle \Delta \ln \mathcal{L} \rangle \gg 1$ and so the first term dominates in Eq. (9). This can occur, for example, where there is a bright PS population inducing a large variance $\sigma^2 \gg X$, which cannot be replicated by an alternative model based solely on diffuse emission with Poissonian statistics; in that sense this is a high-detectability limit.

Then using the estimates above and again taking $X \approx Y$, we find that:

$$\begin{aligned} \langle (\Delta \ln \mathcal{L})^2 \rangle &\approx \int dN \left[-\frac{(N-X)^2}{2\tau^2} + \frac{(N-X)^2}{2\sigma^2} \right]^2 \\ &\quad \times P(N|\{X, \sigma^2\}) \\ &\approx \frac{3}{4} \left(\frac{\sigma}{\tau} \right)^4, \end{aligned} \quad (10)$$

and thus:

$$\begin{aligned} \text{Var}(\Delta \ln \mathcal{L}) &\approx \frac{3}{4} \left(\frac{\sigma}{\tau} \right)^4 - \frac{1}{4} \left(\frac{\sigma}{\tau} \right)^2 \\ &= \frac{1}{2} \left(\frac{\sigma}{\tau} \right)^4. \end{aligned} \quad (11)$$

Thus we expect the standard deviation in this regime to be

$$\begin{aligned} \text{std}(\Delta \ln \mathcal{L}) &\approx \frac{1}{\sqrt{2}} \frac{\sigma^2}{\tau^2} \\ &\approx \sqrt{2} \langle \Delta \ln \mathcal{L} \rangle. \end{aligned} \quad (12)$$

We see that we generically expect the scatter in $\Delta \ln \mathcal{L}$ (from a single pixel) to be of the same order as its expected value. When combining n_{pix} pixels, the expectation value and variance are both enhanced by a factor of n_{pix} , so the standard deviation should be suppressed relative to the expectation value by a factor of $1/\sqrt{n_{\text{pix}}}$.

In this high-detectability, purely Gaussian case, there is actually a simple analytic expression for the full PDF of $\Delta \ln \mathcal{L}$, which we derive in detail in Appendix F:

$$P(\Delta \ln \mathcal{L} = x) = \frac{1}{\sqrt{\pi x \delta}} e^{-x/\delta}, \quad x \geq 0, \quad (13)$$

where $\delta \equiv (\sigma^2/\tau^2) - 1$. It can be readily checked that this distribution reproduces the expectation value and variance given above for $\delta \gg 1$. Note that this distribution is not at all Gaussian; however, as discussed above, combining a large number of pixels and summing their $\Delta \ln \mathcal{L}$ contributions is expected to give an approximately Gaussian PDF by the central limit theorem.

If we instead consider the low-detectability case where $\tau^2 \approx \sigma^2$, i.e. $\delta = (\sigma^2/\tau^2) - 1 \ll 1$, then we instead obtain:

$$\begin{aligned} \langle (\Delta \ln \mathcal{L})^2 \rangle &\approx \int dN \left[-\frac{(N-X)^2}{2\tau^2} + \frac{(N-X)^2}{2\sigma^2} \right. \\ &\quad \left. - \frac{1}{2} \ln \frac{\tau^2}{\sigma^2} \right]^2 P(N|\{X, \sigma^2\}) \\ &\approx \delta^2 \int dN \left[-\frac{(N-X)^2}{2\sigma^2} + \frac{1}{2} \right]^2 P(N|\{X, \sigma^2\}) \\ &\approx \delta^2/2, \end{aligned} \quad (14)$$

where we have used the approximation $\ln(1+\delta) \approx \delta$ and taken the limits of integration to $\pm\infty$. In the same limit,

$$\langle \Delta \ln \mathcal{L} \rangle \approx \delta^2/4. \quad (15)$$

Thus for $\delta \ll 1$, the first term dominates the variance and we have:

$$\text{Var}(\Delta \ln \mathcal{L}) \approx \delta^2/2 \approx 2 \langle \Delta \ln \mathcal{L} \rangle. \quad (16)$$

Thus in this case the square root of the variance is parametrically enhanced (by a factor of $1/\delta$) relative to the expectation value. The variance and expectation value are parametrically similar and will both be enhanced by a factor of n_{pix} when multiple pixels are combined, and so in this regime the standard deviation (square root of the variance) should be of the same order as the square root of the expectation value.

Now we will apply these results to estimate the expected log likelihood ratio between a model containing PSs and one that omits them, when a real population of PSs is present in the data. This $\Delta \ln \mathcal{L}$ will tell us the confidence level with which we expect to be able to exclude the model with no PSs, and hence the confidence level for PS detection. It is similar to the metric we will use for sensitivity to a PS population in our numerical studies.

D. Single component (100% PS emission)

Let us begin by assuming that the data is completely described by a PS population (of identical sources, as described above) without any contribution from a smooth background source. The PS emission has a mean and variance approximated by $(X, \sigma^2) = (N, N(1+s))$, where s is the number of photons per source and N the total number of photons.

Let us consider the expected $\Delta \ln \mathcal{L}$ between the correct PS-based model, and a model that includes only smooth emission, but which correctly predicts the expected number of photons N . Such a smooth model must have equal mean and variance, so we must have $(Y, \tau^2) = (N, N)$.

Using Eq. (9), we plug in these parameters and obtain:

$$\langle \Delta \ln \mathcal{L} \rangle \approx \frac{1}{2} [s - \ln(1+s)]. \quad (17)$$

Note that *all* dependence on the total number of photons N has canceled out; only the number of photons per source is relevant. In particular, this property ensures the likelihood ratio will go to 1 when $s \ll 1$ as required (since this corresponds to the limit of many very faint sources, at which point the smooth model is perfectly adequate), even if the number of sources is very large. Specifically, at small s we have $\langle \Delta \ln \mathcal{L} \rangle \approx (s/2)^2$.

However, this behavior also has the perhaps-surprising implication that having *more* sources (and hence more photons) of fixed brightness in a single pixel neither increases nor decreases the PS sensitivity based on the pixel likelihood, at least once the numbers are large enough that the relevant likelihoods can be approximated as Gaussian.

The leading order behavior of this function at large s is $\langle \Delta \ln \mathcal{L} \rangle \approx s/2$, i.e. the log likelihood in favor of PSs grows linearly with the brightness of the sources. Since the number of photons seen from a given source is directly

proportional to the exposure (i.e. time viewing the source multiplied by the effective area of the instrument), we expect that (at least in this background-free case) $\langle \Delta \ln \mathcal{L} \rangle$ will also grow linearly with exposure. The normalization factor here is also familiar. $2\Delta \ln \mathcal{L} \approx s$ is often used as a test statistic, whose square root translates to the significance in sigma; thus roughly speaking, we expect the detection significance of the PS population (measured in sigma) from a given pixel to approach \sqrt{s} for large s .

If we do not impose the condition that the expected number of photons is N , we can maximize the likelihood for this model under the condition $Y = \tau^2$, obtaining:

$$Y_{\text{optimal}} = \frac{1}{2} \left(\sqrt{4X^2 + 4\sigma^2 + 1} - 1 \right). \quad (18)$$

For the case at hand, this yields $Y_{\text{optimal}} = \frac{1}{2} (\sqrt{4N^2 + 4N(1+s)} + 1 - 1)$. If $N \gg s$, $N \gg 1$ (i.e. the number of both sources and photons is large, consistent with our Gaussian approximations), then to a good approximation $Y_{\text{optimal}} \approx N$ and the estimates above should be reasonable. It is also true in practice, in NPTF analyses of the GCE, that the total photon flux associated with the best-fit GCE model is typically very similar when comparing the fits with and without a model for GCE PSs (e.g. [3]).

E. Generalization to arbitrary ratio of PS and smooth emission

Now let us consider the scenario in which a fraction k of the emission is associated with PSs and the remainder with smooth emission. We seek to evaluate $\langle \Delta \ln \mathcal{L} \rangle$ between the best-fit model (corresponding to the truth) and the model with only smooth emission.

The total number of predicted photons is the sum of the predicted photons associated with each component. The sum of two Gaussian-distributed random variables is also Gaussian-distributed, with mean (variance) given by the sum of the means (variances) for the individual distributions. Thus within our approximations, the best-fit model (matching the truth) has a Gaussian probability distribution for the number of photons N with parameters $(X, \sigma^2) = (kN + (1-k)N, kN(1+s) + (1-k)N)$.

The model with only smooth components that matches the total number of photons has (as in our previous example) $(Y, \tau^2) = (N, N)$.

Using Eq. (9), we obtain:

$$\langle \Delta \ln \mathcal{L} \rangle \approx \frac{1}{2} [ks - \ln(1+ks)]. \quad (19)$$

Thus the effect of a nonzero background fraction on the sensitivity is equivalent to rescaling the photon flux of

individual sources. In this case, the change in scaling behavior from $\langle \Delta \ln \mathcal{L} \rangle \propto s^2$ to $\langle \Delta \ln \mathcal{L} \rangle \propto s$ will occur parametrically around $ks \sim 1$. It is worth noting that if there are a large number of pixels, a significant detection may be consistent with $ks \ll 1$ from every individual pixel, and in this case we should expect a faster-than-linear scaling of the log likelihood ratio with increasing s (or k).

F. A more accurate probability distribution: Accounting for rare sources

We can also compute the expected value of the likelihood ratio and its variance, between two Gaussian models, if the underlying “true” probability distribution is given by Eq. (5), for the case $n_0 \lesssim 1$ where the Gaussian approximations break down. For a Gaussian distribution that describes the expected counts with mean X and variance σ^2 , we find:

$$\begin{aligned} \langle \ln \mathcal{L} \rangle &= \int dN \left[-\frac{(N-X)^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right] P(N|n_0, s) \\ &= p_0(n_0) \left[-\frac{X^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right] \\ &\quad + \sum_{i=1}^{\infty} p_i(n_0) \int dN \left[-\frac{(N-X)^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) \right] \\ &\quad \times \frac{1}{\sqrt{2\pi is}} e^{-(N-is)^2/2is} \\ &= -\frac{1}{2} \sum_{i=0}^{\infty} p_i(n_0) \left[\frac{(is-X)^2 + is}{\sigma^2} + \ln(2\pi\sigma^2) \right], \quad (20) \end{aligned}$$

where again we have made the approximation of taking the limits of integration to $\pm\infty$, relying on $is \gg 1$ for $i \geq 1$ (so that the Gaussians are centered well away from the limits of integration). Now the infinite sums over i can be computed by using the fact that the Poisson probabilities $p_i(n_0) = n_0^i e^{-n_0}/i!$ satisfy $\sum_{i=0}^{\infty} p_i(n_0) = 1$. In particular, by relabeling dummy indices the following identities can easily be proved:

$$\begin{aligned} \sum_{j=0}^{\infty} j p_j(n_0) &= n_0, \\ \sum_{j=0}^{\infty} j^2 p_j(n_0) &= n_0(n_0 + 1), \\ \sum_{j=0}^{\infty} j^3 p_j(n_0) &= n_0^3 + 3n_0^2 + n_0, \\ \sum_{j=0}^{\infty} j^4 p_j(n_0) &= n_0^4 + 6n_0^3 + 7n_0^2 + n_0. \quad (21) \end{aligned}$$

Applying these results we find:

$$\langle \ln \mathcal{L} \rangle = -\frac{1}{2} \left[\ln(2\pi\sigma^2) + \frac{X^2 + s^2 n_0(n_0 + 1) + s n_0(1 - 2X)}{\sigma^2} \right]. \quad (22)$$

In particular, if we hold the variance constant then the likelihood is maximized for $X = n_0 s$, and if we set $X = n_0 s$ (i.e. the model matches the expected total number of photons), then we obtain:

$$\langle \ln \mathcal{L} \rangle = -\frac{1}{2} \left[\ln(2\pi\sigma^2) + \frac{n_0 s(1+s)}{\sigma^2} \right]. \quad (23)$$

The likelihood is then maximized for $\sigma^2 = n_0 s(1+s)$, which is the same variance we found when we directly approximated the probability distribution for the point-source population as Gaussian.

If we examine the expected $\Delta \ln \mathcal{L}$ between this best-fit Gaussian model and a Gaussian model with $\sigma^2 = X = n_0 s$ (representing a purely diffuse signal), we find:

$$\begin{aligned} \langle \Delta \ln \mathcal{L} \rangle &= -\frac{1}{2} \left[\ln(1+s) + \frac{n_0 s(1+s)}{n_0 s(1+s)} - \frac{n_0 s(1+s)}{n_0 s} \right] \\ &= \frac{1}{2} [s - \ln(1+s)]. \quad (24) \end{aligned}$$

Remarkably, this is exactly the same result we found under the Gaussian approximation for the underlying probability distribution [Eq. (17)], suggesting that this result is quite robust even when the assumptions needed to justify the Gaussian approximation break down. We will see in future sections that this result works fairly well to explain scaling relationships for the (numerically computed) sensitivity as we vary the properties of the sources and diffuse background.

As previously, we can generalize to the case where PSs constitute a fraction k of the total emission, so the total expected photon count is $n_0 s/k$ with an expected number of $(1-k)n_0 s/k$ photons originating from diffuse emission. In this case the probability distribution for N given in Eq. (5) must be updated accordingly. As previously, we approximate the probability distribution for the number of photons from diffuse emission as a Gaussian with mean and variance $(1-k)n_0 s/k$; for each choice i for the number of sources drawn, the emission from sources (mean and variance si) can be added to that from diffuse emission by the usual prescription for the sum of normally distributed random variables (i.e. the means and variances add). Thus the overall distribution becomes:

$$P(N|n_0, s, k) \approx \sum_{i=0}^{\infty} p_i(n_0) \frac{e^{-\{N-s[i+n_0(1/k-1)]\}^2/2s[i+n_0(1/k-1)]}}{\sqrt{2\pi s[i+n_0(1/k-1)]}}, \quad (25)$$

where as previously $p_i(n_0)$ is the Poisson probability of drawing i sources when n_0 are expected.

Under this distribution, if we compute the expected likelihood of a Gaussian model with mean X and variance σ^2 , we find (by the same methods as previously):

$$\langle \ln \mathcal{L} \rangle = -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(X^2 - \frac{2Xn_0s}{k} + \frac{n_0s(k + s(k^2 + n_0))}{k^2} \right) + \ln(2\pi\sigma^2) \right]. \quad (26)$$

For fixed σ^2 , this is maximized for $X = n_0s/k$ (as expected, when the model matches the total number of counts); if we fix $X = n_0s/k$, then we obtain:

$$\langle \ln \mathcal{L} \rangle = -\frac{1}{2} \left[\frac{1}{\sigma^2} \left(\frac{n_0s(ks + 1)}{k} \right) + \ln(2\pi\sigma^2) \right]. \quad (27)$$

The expected log likelihood difference between the purely diffuse Gaussian model with $\sigma^2 = X = n_0s$ and the Gaussian model with $\sigma^2 = (n_0s/k)(ks + 1)$ (matching our previous prescription in the case of mixed PS and smooth emission) is then given by,

$$\langle \Delta \ln \mathcal{L} \rangle = \frac{1}{2} \left[ks + \ln \frac{1}{1 + ks} \right], \quad (28)$$

exactly as previously.

However, while the expected log likelihood is unchanged by shifting to this modified probability distribution, the variance differs. Working in the limit where the log terms in the delta log likelihood can be ignored, let us examine the variance of the delta log likelihood between the Gaussian models with $\sigma^2 = n_0s$ and $\sigma^2 = (n_0s/k)(ks + 1)$. In both cases we take $X = n_0s/k$. Then using the identities in Eq. (21), we obtain:

$$\begin{aligned} \text{Var}(\Delta \ln \mathcal{L}) &= \langle (\Delta \ln \mathcal{L})^2 \rangle - \langle \Delta \ln \mathcal{L} \rangle^2 \\ &\approx k^2 \frac{s(s+6) + 3}{4n_0} \left(\frac{ks}{ks+1} \right)^2 + \frac{1}{2} (ks)^2 \\ &\rightarrow \langle \Delta \ln \mathcal{L} \rangle^2 \left(\frac{1}{n_0} + 2 \right), \quad ks \gg 1. \end{aligned} \quad (29)$$

In particular, we observe that there is now an additional term in the variance which scales as $1/n_0$. Consistently with our previous calculation, this term will be negligible when $n_0 \gg 1$ and our original Gaussian approximation holds, but it can lead to a significant enhancement to the variance when $n_0 \ll 1$. In particular, for $ks \gg 1$ and $n_0 \ll 1$, we

expect that the variance over the full dataset can be approximated as:

$$\begin{aligned} \text{Var}(\Delta \ln \mathcal{L})_{\text{overall}} &\approx n_{\text{pix}} \langle \Delta \ln \mathcal{L} \rangle_{\text{per pixel}}^2 / n_0 \\ &\approx \langle \Delta \ln \mathcal{L} \rangle_{\text{overall}}^2 / n_0 n_{\text{pix}}, \\ \Rightarrow \text{std}(\Delta \ln \mathcal{L})_{\text{overall}} &\approx \frac{\langle \Delta \ln \mathcal{L} \rangle_{\text{overall}}}{\sqrt{n_0 n_{\text{pix}}}}. \end{aligned} \quad (30)$$

Thus we see that in this case, rather than the suppression of $1/\sqrt{n_{\text{pix}}}$ that we found earlier (for the high- ks case), instead the suppression is only $1/\sqrt{n_{\text{tot}}}$, where $n_{\text{tot}} = n_0 n_{\text{pix}}$ is the total number of PSs in the image.

Broadly speaking, the standard deviation in $\Delta \ln \mathcal{L}$ is always related to $\langle \Delta \ln \mathcal{L} \rangle$ by a factor of $1/\sqrt{A}$, but A can be either the number of pixels, the number of PSs (when the number of sources per pixel is small), or the test statistic $\langle \Delta \ln \mathcal{L} \rangle$ itself (when the contribution to the test statistic per pixel is small). In the examples we have checked, it is always the smallest of these three parameters that dominates the variance, which is intuitively sensible.

Note that in particular this means the variance can be much larger than one might naively estimate from the square root of the test statistic; if the number of sources is only $\mathcal{O}(100)$, then the variance in the test statistic will be consistently at the $\mathcal{O}(10\%)$ level even if the sources are bright and the significance of detection is very high. Furthermore, we have so far neglected contributions to the variance from the width of the source count function (SCF) (which will modify the effective s entering these calculations from realization to realization, and hence increase the variance), the presence of a nonzero point spread function (likewise), and cross-talk and degeneracies with other background components.

G. Implications for analysis choices

If the overall number of photon counts increases, due to increased exposure (i.e. increased observation time or effective area), the signal fraction k remains constant while s varies linearly. Consequently, we expect $\langle \Delta \ln \mathcal{L} \rangle$ to depend linearly on exposure for sufficiently large s , with the transition from quadratic to linear scaling beginning around $s \sim 1/k$.

Suppose a nonzero angular resolution for the instrument causes the expected number of photons from a single source in a pixel to be reduced, due to leakage into neighboring pixels. Then $\langle \Delta \ln \mathcal{L} \rangle$ will be reduced by the same factor, in the regime where the delta log likelihood scales linearly with s . The signal fraction k should not be affected by this leakage unless the overall distribution of either the signal or background varies rapidly relative to the angular resolution scale; if there is such a rapid variation, there may also be a correction corresponding to the change in k .

The fact that the sensitivity depends only on ks suggests that it is generally more important to have a low background fraction (high k) than a high density of sources (since the latter has no effect on the expected sensitivity in the regimes of validity of our analytic approximations). This suggests that the sensitivity is likely to be dominated by pixels where the expected PS signal is brightest as a fraction of all diffuse backgrounds (which may not be the pixels with the largest number of sources).

We have derived these results for the contribution to $\langle \Delta \ln \mathcal{L} \rangle$ from a single pixel, but the overall log likelihood is simply the sum of the results for the individual pixels. We can thus apply these results even to the realistic case where the background and signal models can have quite different spatial distributions: we simply calculate the appropriate k -value in each pixel and estimate the contribution to $\langle \Delta \ln \mathcal{L} \rangle$ accordingly. Also note that the smooth model can be arbitrarily complicated; the only information we have used is that it has Poissonian statistics.

III. INPUTS AND METHODOLOGY FOR NUMERICAL CALCULATIONS

A. Data selection

To calibrate our simulations to the real gamma-ray sky, we employ eleven years of the Pass 8 public *Fermi* data. The data were collected over 573 weeks from August 4, 2008 to June 19, 2019. To employ the most stringent cosmic-ray rejection criteria, we restrict our selection to the ULTRACLEANVETO event class. For most tests, except those that varied energy range, we limited the energy range to 2–20 GeV, following the default in NPTFit and previous NPTF analyses. We restrict ourselves to an analysis of the top three quartiles of the data graded by angular resolution, as this provides enough range in angular resolution to explore the tradeoff with exposure, and the angular resolution degrades significantly in the bottom quartile.

B. NPTFit scan setup

We employ v.0.2 of NPTFit, together with MultiNest, a Bayesian inference tool that implements a nested sampling algorithm [14,52]. For fits of the simulated data, the number of live points is described within the individual procedure sections; when not otherwise specified we used $n_{\text{live}} = 100$. At each experiment, we checked the recovered evidences at different n_{live} values to determine if the scans were reasonably converged. We found that changes in $\langle \ln \text{BF} \rangle$ at n_{live} values beyond 100 were consistently very small, and thus the scans are well-converged.

When fitting to simulated data, our region of interest (ROI) is centered on the GC and has a radius of 15° . We exclude the band with galactic latitude $|b| < 2^\circ$. This ROI is chosen for computational efficiency and to minimize contamination from background emissions, while preserving sensitivity to the GCE, motivated by a recent study

finding that sensitivity to GCE PSs plateaus for ROIs with radii between 15° and 20° [43]. Our expectation is that shifting to a modestly different ROI would not significantly affect the scaling with exposure, angular resolution, etc, that we study in this work, although the overall sensitivity would change and so should not be compared directly between analyses with different ROIs.

During NPTFit scans, the non-Poissonian components of the sky maps must be exposure corrected at each pixel (a computationally costly process) since the exposure map (a map that reflects the duration of observation and the stringency of data selection) is nonuniform [14,52]. In order to optimize computational efficiency (and consistent with the recommendations in NPTFit), we set $n_{\text{exp}} = 5$ to divide the ROI into 5 distinct regions within which the exposure is treated as uniform.

C. Modeling the gamma-ray sky

We conducted a Bayesian NPTFit analysis of the (real data) *Fermi* sky map at $n_{\text{live}} = 500$, modeling the sky as a linear combination of spatial templates characterized by parameters and prior distributions that are described in Appendix A. For this analysis, and subsequent fits to real data (which were used only to choose parameters for the subsequent simulations), we extended the radius of the ROI from 15° to 30° , but retained the mask of the Galactic plane (consistent with defaults in NPTFit). Smooth/diffuse templates were included for the *Fermi* Bubbles (“Bub”), smooth isotropic emission (“Iso”), smooth GCE (“GCE”), and Galactic diffuse emission (“Dif”). Templates were included for PS populations associated with the GCE (“GCE PS”), isotropic/extragalactic sources (“Iso PS”), and the Galactic disk (“Disk PS”). Smooth/diffuse templates each have one associated parameter, A_{smooth} , controlling their overall normalization in the model; PS population templates (hereafter “PS templates”) are associated with an overall normalization parameter A_{PS} which controls the number of sources, and with a SCF which describes the number of sources as a function of their flux. We use a singly-broken power law model for the SCF, as is the default in NPTFit:

$$\frac{dN}{dS} = A_{\text{PS}} T_{\text{PS}} \begin{cases} \left(\frac{S}{S_b}\right)^{-n_1} & S \geq S_b \\ \left(\frac{S}{S_b}\right)^{-n_2} & S \leq S_b \end{cases}, \quad (31)$$

where A_{PS} is an overall normalization factor, and T_{PS} is the position-dependent template with the fixed normalization given in the NPTFit code (see Appendix A for details). Note the parameter S_b controls the expected number of photon counts per source at the position of the break in the power law. The expected number of sources in a given pixel is then set by:

$$N_{\text{tot}} = A_{\text{PS}} T_{\text{PS}} S_b \left(\frac{1}{n_1 - 1} + \frac{1}{1 - n_2} \right), \quad (32)$$

whereas the expected number of photons is set by:

$$S_{\text{tot}} = A_{\text{PS}} T_{\text{PS}} S_b^2 \left(\frac{1}{n_1 - 2} + \frac{1}{2 - n_2} \right). \quad (33)$$

Note that in the main text of the paper we employ the default Galactic diffuse emission model from NPTFit, constructed from the *Fermi* Collaboration’s p6v11 diffuse model. This model is known to have features that can bias the results [43] when it is used directly to reconstruct PS populations from the real data; however, it should provide a reasonable description of the data when we are only interested in constructing and analyzing simulations (where the model is correct by construction). To check this assertion, in Appendix E we recalculate our results with a different Galactic diffuse emission model and comment on the differences. Either of these Galactic diffuse emission models reconstruct the GCE as being 100% PSs, with the smooth GCE component being negligible: consequently, our simulations will generally explore the sensitivity of NPTF methods to a PS population bright enough to explain the full GCE. (However, note that there are other models of the Galactic foregrounds where the flux attributed to the smooth GCE component is not negligible [43,44].)

After performing this fit, we extracted the posterior median parameters associated with each template, which were then used as the baseline inputs to simulations for the rest of the paper. These simulation parameters are displayed in Table V in Appendix A. Note in particular that (consistent with previous NPTF studies) the inferred shape of the SCF for the GCE PS is quite sharply peaked around S_b , so we will generally be simulating GCE PS populations where the PSs all have roughly the same brightness as observed at Earth (fixed by S_b). This is likely *not* a realistic luminosity function, but serves as a convenient basis for understanding the sensitivity of NPTF methods. We discuss the sharpness of the SCF peak further in Appendix B.

D. Producing simulated sky maps

For each template, we generated realizations based on the posterior median parameters from the real data. For the smooth/diffuse emission components, we performed a Poisson draw from the associated template (with normalization given by the simulation parameters taken from the fit to real data). To obtain realizations of PS populations, we employed NPTFit-Sim.

To obtain the full skymaps, the individual components were summed. All skymaps were binned using HEALPix, a package designed to allow equal-area pixelization of the sky [53]. The *nside* value controls the pixel size, with the sky having a total of $12 \times \text{nside}^2$ equal-area pixels. By default we set *nside* to 128, which is also the NPTFit default,

and corresponds to roughly a 0.5° mean spacing between individual pixel centers in the region toward the GC. For *nside* = 128, there are 2808 pixels within our ROI.

E. Sensitivity figure of merit

A NPTFit analysis returns posterior probability distributions for each of the parameters, and an estimate of the overall Bayesian evidence for the model. Comparing two NPTFit analyses, with different template choices, allows us to evaluate the Bayes factor (BF) between the two scenarios, as the ratio of their evidences. In particular, we can define the sensitivity to a GCE PS population in terms of the BF in favor of a model that contains the complete set of templates (Dif, Bub, Iso, GCE, GCE PS, Disk PS, Iso PS) compared with a model that excludes the GCE PS template. A high value of this BF corresponds to a high-significance detection of the GCE PS template, over and above the smooth GCE template. For convenience, we will generally work with $\ln \text{BF}$ rather than the BF itself. Where $\text{BF} \lesssim 1$ and so $\ln \text{BF}$ is negative, there is no detection of GCE PSs.

The BF directly gives the ratio of Bayesian probabilities that the model with the GCE PS template is correct, compared to the model without that contribution. For those more accustomed to frequentist statistics, it may be helpful to think of the BF as comparable to a likelihood ratio $\mathcal{L}_1/\mathcal{L}_2$, with additional terms that penalize models with more degrees of freedom. In this sense $2 \ln \text{BF}$ is broadly analogous to the commonly used test statistic $2\Delta \ln \mathcal{L}$, which for a likelihood that is Gaussian near its maximum ($\mathcal{L}(x) \propto e^{-x^2/(2\sigma^2)}$) can be written as $2\Delta \ln \mathcal{L} \approx (x/\sigma)^2$, and thus can be thought of as the “number of sigma” squared associated with the deviation from the best-fit point.

The $\ln \text{BF}$ in favor of a GCE PS population can vary widely between realizations. For our main figure of merit for sensitivity, we will use the expected value of $\ln \text{BF}$ obtained by taking the average across realizations, $\langle \ln \text{BF} \rangle$, although we will also show the scatter between realizations.

IV. PROCEDURES FOR PARAMETER VARIATION

Within each subsection below, we describe the general procedure for varying different inputs: exposure, angular resolution, source brightness, and pixel size. We describe the method for adjusting parameters in the simulation of skymaps as well as how to account for these variations through the priors when analyzing the skymaps using NPTFit. If the test involves combinations of these variations, then the priors must be modified by simultaneously implementing the adjustment factors to the priors for each alteration performed.

A. Exposure

Although *Fermi* is a space-based telescope, it does not observe every part of the sky simultaneously. As a result, an exposure map is needed to keep track of how long *Fermi*

observed a particular region of the sky and with what effective area. The exposure map provided by the Fermi-LAT Collaboration and implemented in NPTFit has units of cm^2s [14].

Increasing the amplitude of the exposure map could describe longer observations with *Fermi* or less stringent cuts on photons as part of the data selection. We define an exposure rescaling factor χ , which allows us to vary the intensity of the exposure map through a scalar multiplicative factor, hence rescaling the expected number of photons present in simulated data. In our baseline case, $\chi = 1$. In general, the exposure rescaling could be position-dependent (e.g. corresponding to longer observations of only part of the ROI). However, we expect such position-dependent variations to be modest for the inner Galaxy region, as the size of our ROI is smaller than the field of view of *Fermi*.

We implemented the variation of exposure in the simulated data by modifying the template parameters as follows. For smooth/diffuse templates, the template normalization A_{smooth} is multiplied by the rescaling factor χ , since A_{smooth} determines the mean photon counts within each pixel. For non-Poissonian templates, when the other parameters are held fixed, A_{PS} determines how *many* sources are present, which is not a function of exposure. Therefore, we instead multiply the counts break S_b by χ , as S_b controls the expected number of photons emitted by a source lying at the break in the SCF; as discussed previously, for the fits we perform, S_b corresponds to the typical number of photons per source. To ensure that the total number of sources does not change, we divide A_{PS} by χ following Eq. (33).

When performing NPTFit analyses on these modified skymaps, the input exposure map must be multiplied by χ . Furthermore, we adjust the range of priors governing A_{smooth} for Poissonian sources and A_{PS} and S_b for non-Poissonian sources, so that they correspond to the same underlying physical emission parameters as in the original $\chi = 1$ analysis. For example, the boundaries of a uniform prior on $\log A_{\text{smooth}}$ are shifted by $+\log \chi$; the boundaries of a log prior on $\log A_{\text{PS}}$ are shifted by $-\log \chi$; and the boundaries of the linear prior on S_b are multiplied by χ . (The original values of all priors are displayed in Table VI in Appendix A.) For simulated data, the number of live points we utilized in the scans is $n_{\text{live}} = 300$. We checked that the relative changes in the recovered evidences (under variations to n_{live}) are negligible for all individual realizations.

B. Angular resolution

Angular resolution, characterized by the point spread function (PSF), represents how well a telescope such as *Fermi* is able to reconstruct the original direction of a detected photon. A nondelta-function PSF represents an uncertainty in the direction of a photon's origin. As a result, the image produced of a photon source is "smeared" across one or more pixels. Since the NPTFit implementation does not account for correlations between neighboring pixels

(see e.g. [11] for a discussion), this smearing has the potential to bias the recovered SCF.

Modifications to the photon direction reconstruction, or construction of future gamma-ray telescopes, may allow for better angular resolution (equivalently, a narrower PSF) than *Fermi* can currently achieve. However, even within the *Fermi* dataset photons can be separated by the quality of their directional reconstruction, allowing us to improve angular resolution at the cost of exposure. Specifically, *Fermi* photons are divided into four quartiles ranked by angular resolution, and separate PSF estimates are provided for each of the quartiles. Furthermore, lower-energy photons have intrinsically worse angular resolution, so a cut on photon energy has the effect (among others) of modifying the effective PSF.

Fermi's PSF is modeled by a pair of King functions [defined in Eq. (35)] and is characterized by a set of several parameters. The PSF is approximately Gaussian near the core, with larger non-Gaussian tails. Equation (34) displays the full functional form of *Fermi*'s PSF.⁴

$$P(x, \vec{\alpha}_p) = f_{\text{core}} K(x, \sigma_{\text{core}}, \gamma_{\text{core}}) + (1 - f_{\text{core}}) K(x, \sigma_{\text{tail}}, \gamma_{\text{tail}}) \quad (34)$$

$$K(x, \sigma, \gamma) = \frac{1}{2\pi\sigma^2} \left(1 - \frac{1}{\gamma}\right) \left[1 + \frac{x^2}{2\gamma\sigma^2}\right]^{-\gamma}. \quad (35)$$

Here x is a rescaled distance from the center of the source, with an energy-dependent scale factor $S_p(E)$:

$$x = \frac{\delta p}{S_p(E)} \\ \delta p = 2\sin^{-1}\left(\frac{|\hat{p}' - \hat{p}|}{2}\right), \quad (36)$$

where \hat{p} and \hat{p}' are the unit vectors corresponding respectively to the true and reconstructed directions of the photon. The parameters that define the PSF [$S_p(E)$, f_{core} , and γ and σ for the two King functions in Eq. (34)] are provided with the *Fermi* dataset as functions of energy and event selection.

Note that because of the rather complex form of the *Fermi* PSF, different event selections may have PSFs that are not related simply by an overall shift in scale [e.g. by a modification to $S_p(E)$], but are different in shape. To maximize the practical applicability of our work, rather than simply rescaling the PSF, we test the effect of using the true PSFs for different quartiles of the *Fermi* data ranked by PSF, and for different energy ranges. However, we will show that within the range of event selections we study, the effect on sensitivity of changing the PSF can be quite well

⁴https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_PSF.html.

described by the variation of a summary parameter such as the 68% containment angle, suggesting that the detailed form of the *Fermi* PSF is not a crucial ingredient.

We divide the dataset into 40 log-spaced energy bins spanning the range from 0.2 GeV to 2000 GeV (i.e. 10 bins per decade). For each quartile and energy bin, we re-simulate the data with the same underlying model parameters but different PSF parameters. We stack these simulations together where appropriate (e.g. when testing multiple quartiles simultaneously, or when considering a broad energy range). When we analyze the simulated data, we use the worst PSF for any subset of the simulated data, which is consistent with what has been done in previous studies on the real data [3,45]. For example, if the simulated data involved photons from the top three PSF quartiles and a range of energies from E_{\min} to E_{\max} , the PSF parameters used will correspond to the third-best PSF quartile and the energy E_{\min} (since the angular resolution of *Fermi* improves monotonically with increasing energy).

C. Energy range

Varying the energy range of the data selection has multiple effects. Including a wider range of energies effectively increases the exposure; including lower-energy photons worsens the angular resolution. As mentioned above, we use the real PSF of *Fermi* for different energy ranges as one way to probe the effects of varying angular resolution. However, changing the energy range has additional effects that are not reducible to changes to angular resolution and exposure: low-energy photons are more abundant than high-energy ones in general, but also the spectra of the various emission components are different. Consequently, changing the energy range will modify the flux fraction associated with the GCE (and all other components). The choice of energy range thus needs to be optimized depending on the signal of interest.

To address the specific question of the optimal energy range for GCE studies, we change our energy cut on the real data and then repeat the analysis described in Sec. III. That is, we refit the templates to the real data with the new energy range, simulate the data based on these new template parameters, and determine the sensitivity to PSs as a function of energy range.

D. Pixel size

Another somewhat *ad-hoc* choice in the standard NPTF analysis is the choice of spatial binning for the photons, i.e. the size of the equal-area HEALPIX pixels (or equivalently, their number). Following previous studies such as [3,14,41], we utilized $n_{\text{side}} = 128$ for the majority of our analysis. The reason for this choice is the similarity between the $n_{\text{side}} = 128$ pixel radius and *Fermi*'s angular resolution in the energy range of interest.

If the pixel size chosen is substantially smaller than the angular resolution, PSs will always occupy multiple pixels,

and the fact that NPTFit does not model correlations between neighboring pixels could lead to a loss of sensitivity. In the extreme limit of small pixel size, where all pixels contain either 0 or 1 photons, all sensitivity to PS populations would be lost. On the other hand, pixels much larger than the angular resolution increase the background from diffuse signals in any given pixel, and again this might be expected to reduce the sensitivity to PSs.

To examine these effects in detail, we use the same template parameters derived from the data for $n_{\text{side}} = 128$ (without any adjustment of parameters or priors) but simulate skymaps at $n_{\text{side}} = 512$ using the procedure described in Sec. III. After generating simulated skymaps at this higher resolution, we can increase the pixel size for each realization as desired, using a HEALPIX package that combines “children” pixels to create a superpixel, while accounting for proper normalization of photon counts [53]. This preserves the distribution of sources in the individual realizations as we explore different pixel sizes.

For variations in pixel size, priors do not need to be adjusted. Instead, the templates must be properly normalized within the ROI to ensure accurate scaling during parameter retrieval (see Appendix A).

E. Source brightness

The previously discussed parameters describe instrumental properties and analysis choices. In this section, we discuss how the template model parameters are adjusted to describe a genuinely different source population. In analyses of sensitivity as a function of source brightness, our goal is to understand the potential of the NPTFit algorithm to detect faint subthreshold sources.

The brightness of PSs can be varied by adjusting the S_b and A_{PS} parameters of the SCF [Eq. (31)]. We tested the effect of varying the brightness of individual PSs while keeping the total flux in the PS population fixed (as appropriate for a PS population making up most or all of the GCE). Using Eq. (33), this requirement can be satisfied by simultaneously varying S_b and A_{PS} as follows:

$$S_b \rightarrow nS_b, \quad A_{\text{PS}} \rightarrow \frac{1}{n^2}A_{\text{PS}}. \quad (37)$$

For example, if $n = 1/2$, the number of photons each source emits is reduced by half, however, the number of sources increases by a factor of 2 to compensate. This equates to a factor of 4 increase in the template normalization factor A_{PS} , since the number of sources scales as $A_{\text{PS}}S_b$.

As for the exposure tests, when we rescale the simulated parameters we also rescale the priors in the fit to simulated data. For example, if $S_b \rightarrow nS_b$ and there was initially a linear prior on S_b in the range [0.05, 80], the new prior is linear with range [0.05 n , 80 n]. If the prior on $\log_{10} A_{\text{PS}}$ is initially [-6, 1], it is adjusted to [-6 - 2 $\log_{10} n$, 1 - 2 $\log_{10} n$]. We also checked the effect of keeping the

priors fixed; except in situations where the true parameters approached the edge of the prior (or fell outside it), the effect was minimal.

For all simulations involved in the variation of source brightness, the number of live points we used to scan the data was set to $n_{\text{live}} = 100$ for computational efficiency.

V. A SIMPLIFIED ISOTROPIC SCENARIO

Before analyzing the results with all templates, we perform a simplified analysis including only isotropic components in our simulations, and approximating the exposure map as uniform. This analysis serves as a test of our analytic predictions. We describe some additional studies under this simplified scenario in Appendix D.

In this case, our normalization convention for the emission templates requires that the templates T and T_{PS} are both 1 in all pixels within the ROI. The normalization of the simulated signals was determined by matching the parameters for the PS component (A_{PS} , n_1 , n_2 , and S_b) to the isotropic PS component extracted from the real *Fermi* data. For our baseline analyses, the smooth component normalization was chosen such that the total flux contributions of the smooth isotropic and PS components are equal. Explicitly, given our normalization convention for the templates T and T_{PS} , this means that $A(\theta)_{\text{smooth}}$ is given by:

$$A(\theta)_{\text{smooth}} = A(\theta)_{\text{PS}} S_b^2 \left[\frac{(n_1 - n_2)}{(n_1 - 2)(2 - n_2)} \right], \quad (38)$$

where (as previously) $A(\theta)_{\text{PS}}$ is the template normalization for the emission associated with the isotropic PS population, S_b is the break of the source-count function, and n_1, n_2 are

the slope of the source count functions defined by a singly-broken power law.

As previously, we use NPTFit-Sim to simulate PSs and a Poisson draw to simulate the smooth component. The priors on the various parameters are set as discussed in Sec. IV and Appendix A. (Note that if the simulated value of A_{smooth} was outside the prior range on A_{iso} in the main analysis, we would need to adopt different priors for this isotropic study, but in fact it lies well within the prior range so this is not a problem.)

A. Variation of exposure (narrow PSF)

The analytic approximations we derived in Sec. II assumed that the PSs were not smeared by the PSF. Thus as a cross-check of the scaling behavior estimated from our analytic results, we performed an initial set of simulations where the PSF was taken to be extremely narrow (i.e. the angular reconstruction was effectively perfect), covering a range of exposure levels χ . Specifically, we sampled exposure rescaling factors χ between 10^{-2} and 10 (recall $\chi = 1$ corresponds to the baseline exposure), and for each case generated and scanned skymaps that employed a Gaussian PSF with a tiny variance 10^{-20} . For each choice of χ we ran the analysis for 10 simulated realizations, and for each realization evaluated the ln BF between models with and without isotropically distributed PSs (both models allow for an isotropically distributed smooth component).

Figure 1 plots the Bayes factor preference in favor of PSs as a function of exposure, together with the analytic solution for the likelihood ratio in the case where PSs and background are equally bright [Eq. (19) with $k = 1/2$ and $s = S_b$]. We also show the result of including only the linear $ks/2$ term in the analytic estimate of Eq. (19).

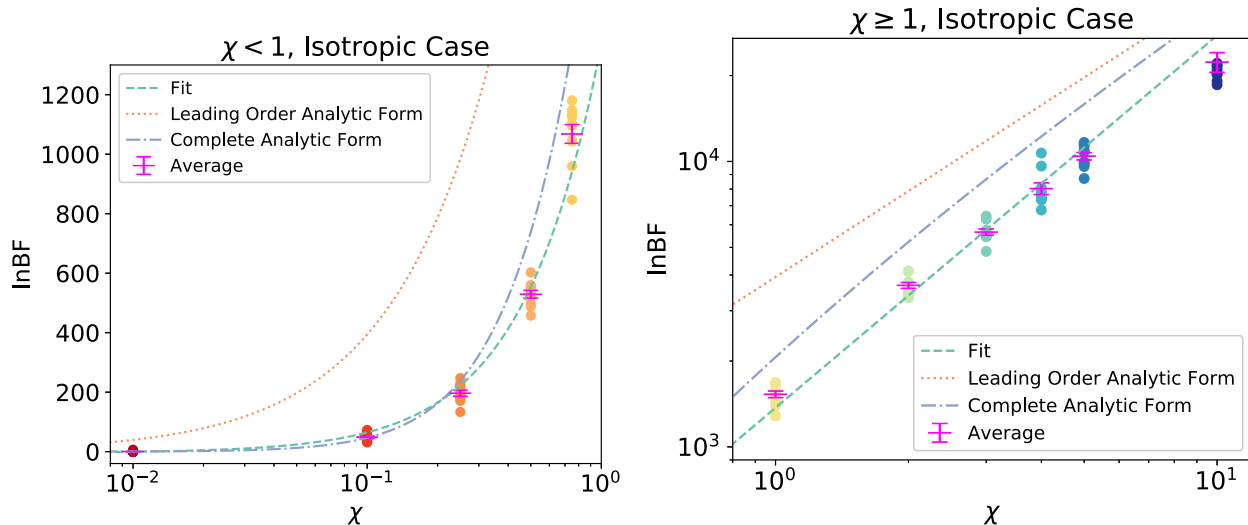


FIG. 1. ln BF across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), for varying values of χ /exposure. Signal and background are isotropic and both simulations and scans employ a narrow Gaussian PSF. The green dashed line denotes the modified power law fit defined in Eq. (39). The orange dotted line denotes the first (linear) term of the analytic form Eq. (19), while the blue dash-dot line denotes the full analytic solution described in Eq. (19).

For each choice of exposure, we evaluate $\langle \ln \text{BF} \rangle(\chi)$ by taking the average of $\ln \text{BF}$ across realizations at each exposure level; we also compute the standard error of the mean across the realizations for this quantity at each χ value (indicated by magenta vertical bars in Fig. 1).

We work by default (here and in the remainder of this work) with the log of the Bayes factor between models, rather than the likelihood ratio; however, in this specific example we also evaluated the log likelihood ratio and found that it was generically quite close to the log Bayes factor (and in particular the difference between the two was not responsible for the difference between the numerical results and the analytic approximation for $\langle \Delta \ln \mathcal{L} \rangle$). Thus we treat our analytic approximation as a rough estimate for $\langle \ln \text{BF} \rangle$.

For these parameter choices, we observe that the analytic form mildly overestimates the sensitivity, by a factor of roughly 20%–30% in $\langle \ln \text{BF} \rangle$ at high exposure, but accurately captures the fall-off of the detection sensitivity at low exposure, and the scaling at high exposure. The remaining discrepancy is likely due to the approximations we have made in deriving our analytic results (e.g. relating to the shape of the probability distribution, and assuming we can treat all integrals as having limits $\pm\infty$, as well as approximating the SCF as a delta-function).

We observe a consistent scatter at the $\mathcal{O}(10\%)$ level in $\langle \ln \text{BF} \rangle$ between different realizations, which does not obviously decrease at large χ . (Note that here we are discussing the standard deviation across realizations, not the standard error of the mean; the latter is smaller by a factor of $1/\sqrt{n_{\text{realizations}}}$.) This can be understood in terms of our variance calculations in Sec. II. The parameters we have simulated correspond to 5.61χ photons/source, 2808 pixels,

and an average of 0.11 sources/pixel; thus we expect a total number of sources in the ROI around 280, and a standard deviation in the log likelihood ratio that is of order $\langle \ln \text{BF} \rangle / \sqrt{280} \sim 0.06 \langle \ln \text{BF} \rangle$ or $\sqrt{\langle \ln \text{BF} \rangle}$, whichever is larger. This is consistent with the $\mathcal{O}(10\%)$ scatter we observe at high exposure.

In addition to the comparison to the analytic prediction, we can parametrize the scaling of the sensitivity with χ as a power law and fit for the parameters (although power-law behavior should be expected to break down at sufficiently small χ , where $\ln \text{BF}$ can attain negative values). The fitting function we use is

$$\langle \ln \text{BF} \rangle(\chi) = \alpha \chi^\beta + \gamma, \quad (39)$$

where the offset parameter γ serves to correct the behavior at small χ where there is not enough data to detect a significant signal. For each value of χ we took the central value of $\langle \ln \text{BF} \rangle(\chi)$ to be the average over realizations, with an error bar determined by the standard error of the mean, and performed a least-squares fit. The resulting best-fit model is also plotted in Fig. 1.

B. Variation of exposure (realistic PSF)

In a realistic scenario, we will always have to deal with a PSF that is not arbitrarily narrow. We repeat the simulation and analysis described above using the full PSF appropriate to the real *Fermi* dataset (for the top PSF quartile), and show results in Fig. 2. We compare these results to the same analytic solution (i.e. with no allowance for the PSF) as in the previous analysis, and again perform a least-squares fit to a simple power law fitting function.

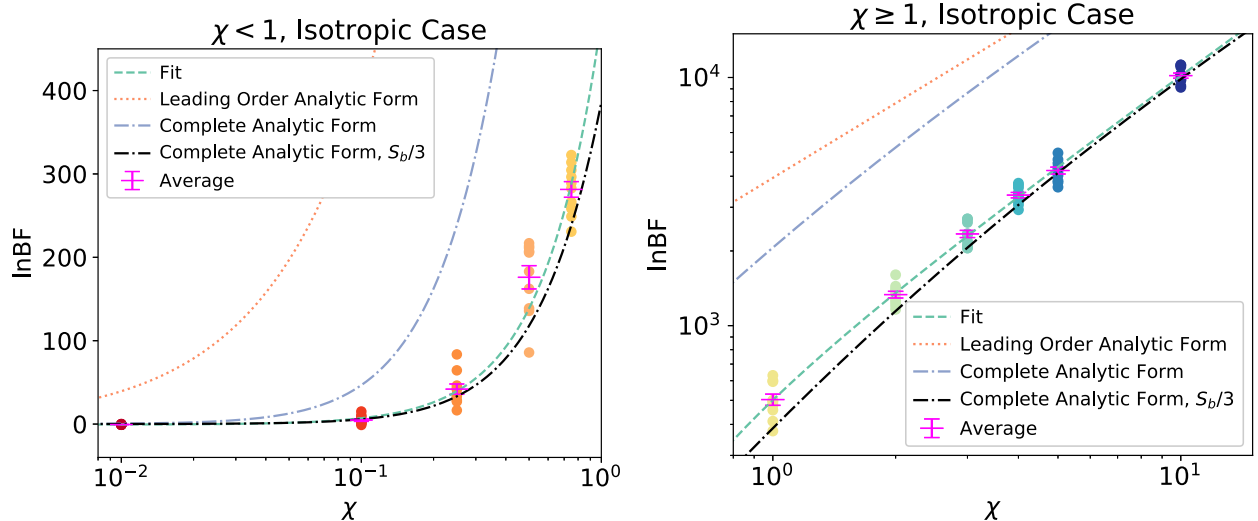


FIG. 2. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), for varying values of χ /exposure. Signal and background are isotropic and both simulations and scans employ a realistic PSF model. The green dashed line denotes the modified power law fit defined in Eq. (39). The orange dotted line denotes the first (linear) term of the analytic form in Eq. (19), while the blue dash-dot line denotes the full analytic approximation from the same equation with $s \rightarrow S_b$. The black dash-dot line shows the analytic approximation with $s \rightarrow S_b/3$.

We observe that the analytic solution still describes the shape of $\langle \ln \text{BF} \rangle$ as a function of χ quite well, but now the discrepancy in our sensitivity metric is more pronounced (a factor of a few at high χ). At least qualitatively, this discrepancy can be largely absorbed by taking s/S_b to be a constant other than unity; Fig. 2 shows the effect of using the analytic approximation with s replaced by $S_b/3$. The variance remains $\mathcal{O}(10\%)$ at high χ , which can be understood as discussed above.

In this more realistic case, we thus recommend using the analytic estimate only to understand scaling behavior rather than as a quantitative estimate of the expected sensitivity, although a reasonably good description can be obtained by fitting a constant rescaling factor to be applied to s .

C. Variation of relative flux contributions

We can also test the effects of varying the relative flux contributions of the smooth and PS components while allowing the total flux to remain constant. Figure 3 demonstrates how the sensitivity changes as the PS flux fraction is varied, for both the narrow PSF and realistic PSF (top quartile) cases. For comparison, we also overlay the predictions given by our analytic approximations, Eq. (19), with $s \rightarrow S_b$ and $s \rightarrow S_b/3$. We find that as the relative contribution of the PS component increases, the sensitivity of NPTFit to PSs naturally increases with a shape consistent with the analytic prediction provided by Eq. (19), and in the narrow-PSF case the $s \rightarrow S_b$ substitution provides quantitatively accurate results. Although the case where the map is simply produced with a smooth component is not shown

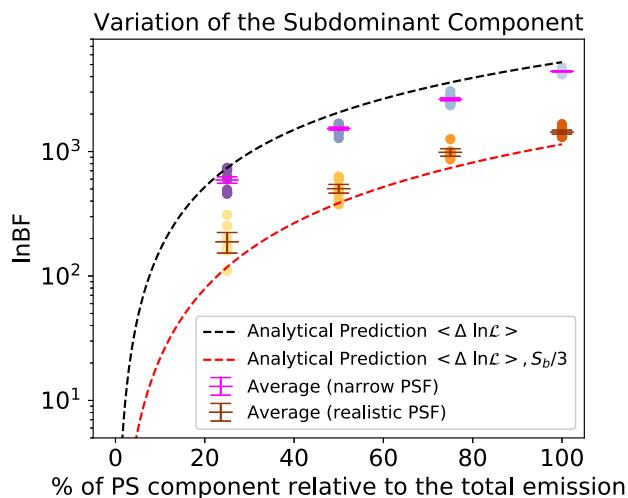


FIG. 3. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta, brown), as the relative contributions between a smooth and PS component within an isotropic map are varied, for realistic and narrow PSF prescriptions. Dashed lines indicate the analytic prescription of Eq. (19) with the replacements $s \rightarrow S_b$ (black) and $s \rightarrow S_b/3$ (red).

in the figure due to the log-scaling, the result averages to -0.54 ± 0.067 in the realistic-PSF case and -0.56 ± 0.14 in the narrow-PSF case, both of which are small, as expected. In the regime where there is no significant preference for PSs, we expect the Bayes factor in favor of the model without PSs to be highly dependent on the choice of priors (as also discussed in e.g. [11,42]). We explore this point further in Appendix C.

VI. RESULTS OF SIMULATED PARAMETER VARIATIONS IN THE FULL INNER GALAXY ANALYSIS

In this section we now proceed to a numerical analysis using simulated *Fermi* data for the inner Galaxy, employing the complete set of templates discussed in Sec. III. The results of this section can thus be used directly to optimize NPTFit-based approaches to studies of the inner Galaxy and GCE.

A. Varying the exposure

As in the isotropic case, we sampled exposure rescaling factors χ between 10^{-2} and 10. For each choice of χ we ran the analysis for 20 simulated realizations, and for each realization evaluated the $\ln \text{BF}$ between models with and without the GCE PS template.

Figure 4 shows the resulting values of $\ln \text{BF}$ for each exposure level. We evaluate $\langle \ln \text{BF} \rangle(\chi)$ by taking the average of $\ln \text{BF}$ across realizations at each exposure level (indicated by magenta vertical bars in the figure along with error bars that denote the $\sigma/N_{\text{realizations}}$ (where $N_{\text{realizations}}$ is the number of realizations in a given sample) standard error of the mean across all the realizations within a particular case).

As discussed in Sec. V, we fit a power-law function [Eq. (39)] to the data for $\langle \ln \text{BF} \rangle(\chi)$. The resulting best-fit parameters are given in Table II, and the best-fit model is plotted in Fig. 4 (solid blue line). We find approximately that $\langle \ln \text{BF} \rangle \propto \chi^{0.8}$ at large BF. This is broadly consistent with our expectation from the analytic estimate in Sec. II that $\langle \ln \text{BF} \rangle$ should scale \sim linearly in χ .

B. Varying the angular resolution

1. PSF models for different quartiles

We begin by examining how the sensitivity of NPTFit varies when the top three quartiles of data by PSF are analyzed separately, keeping the energy range fixed at its default value of 2–20 GeV. Quartiles are labeled in order of decreasing angular resolution (so e.g. “PSF 01” represents the best quartile).

Within each quartile, we simulated and analyzed 20 realizations. All simulations are generated at a rescaling factor of $\chi = 1$. Figure 5 shows a striking decline in sensitivity in the quartiles with worse angular resolution.

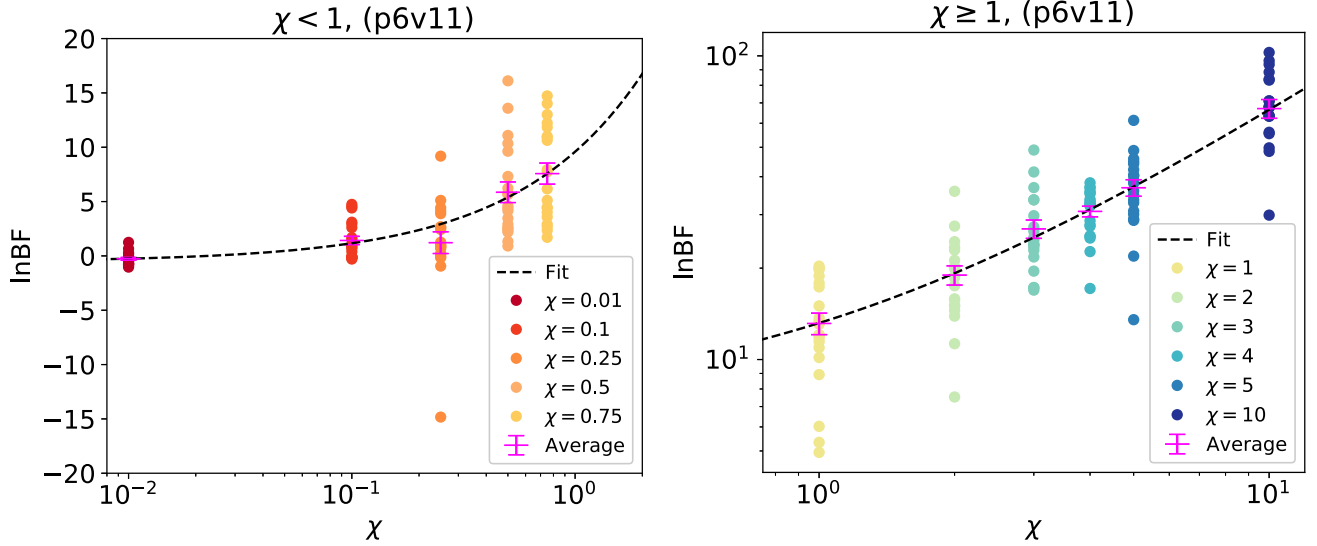


FIG. 4. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), for varying values of χ /exposure. Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. The best-fit line is a standard power law with an additive shift defined in Eq. (39).

As previously, we display both the scatter between realizations and the average $\langle \ln \text{BF} \rangle$ across realizations for a given quartile.

2. PSF models for different energies

Another practical way to vary the angular resolution in *Fermi* data is to modify the energy window. We first examined the (theoretical) case where the angular resolution is varied while keeping all other parameters constant. We resimulated the data with the original exposure map ($\chi = 1$) but with PSF corresponding to the appropriate *Fermi* PSF for energies between 0.6 GeV and 3.2 GeV (recall that the baseline analysis uses the *Fermi* PSF at 2 GeV), in the top PSF quartile. In each case we performed 20 realizations.

In Fig. 6 we plot the resulting values of $\ln \text{BF}$, against the value of the 68% containment angle associated with each PSF model (in degrees), which we denote η . We also include the results for the 2 GeV PSF in all three quartiles (discussed above).

To summarize the results, we fit the data with a power law, $\langle \ln \text{BF} \rangle(\eta) = a\eta^b$, using the same least-squares analysis as described above for the case of $\langle \ln \text{BF} \rangle(\chi)$. Table III

TABLE II. Best-fit parameters obtained using least-squares regression method along with the 1σ error for the power law fit of $\langle \ln \text{BF} \rangle$ to χ , as defined in Eq. (39).

Recovered parameters	p6v11
α (coefficient)	11.30 ± 0.69
β (power)	0.76 ± 0.04
γ (shift)	-0.62 ± 0.20

displays the resulting best-fit parameters. We find that $\langle \ln \text{BF} \rangle(\eta) \propto \eta^{-1}$, i.e. the sensitivity appears to scale approximately inversely with the containment radius, at least while holding the pixel size constant at $\text{aside} = 128$.

Thus as a rule of thumb, we expect an increase in the exposure by a factor of n to be approximately compensated by an increase in the containment radius (not the containment area) by a factor of n ; if the exposure can be more than doubled while worsening the containment angle by less

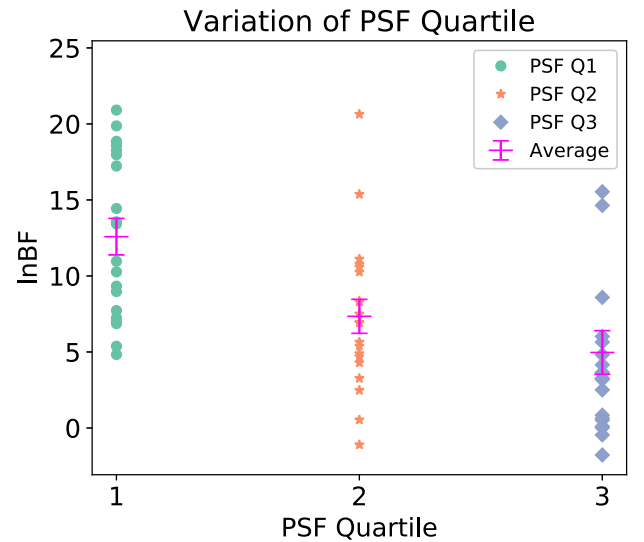


FIG. 5. $\ln \text{BF}$ across 20 realizations (circle/star/diamond markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), for the top three PSF quartiles.

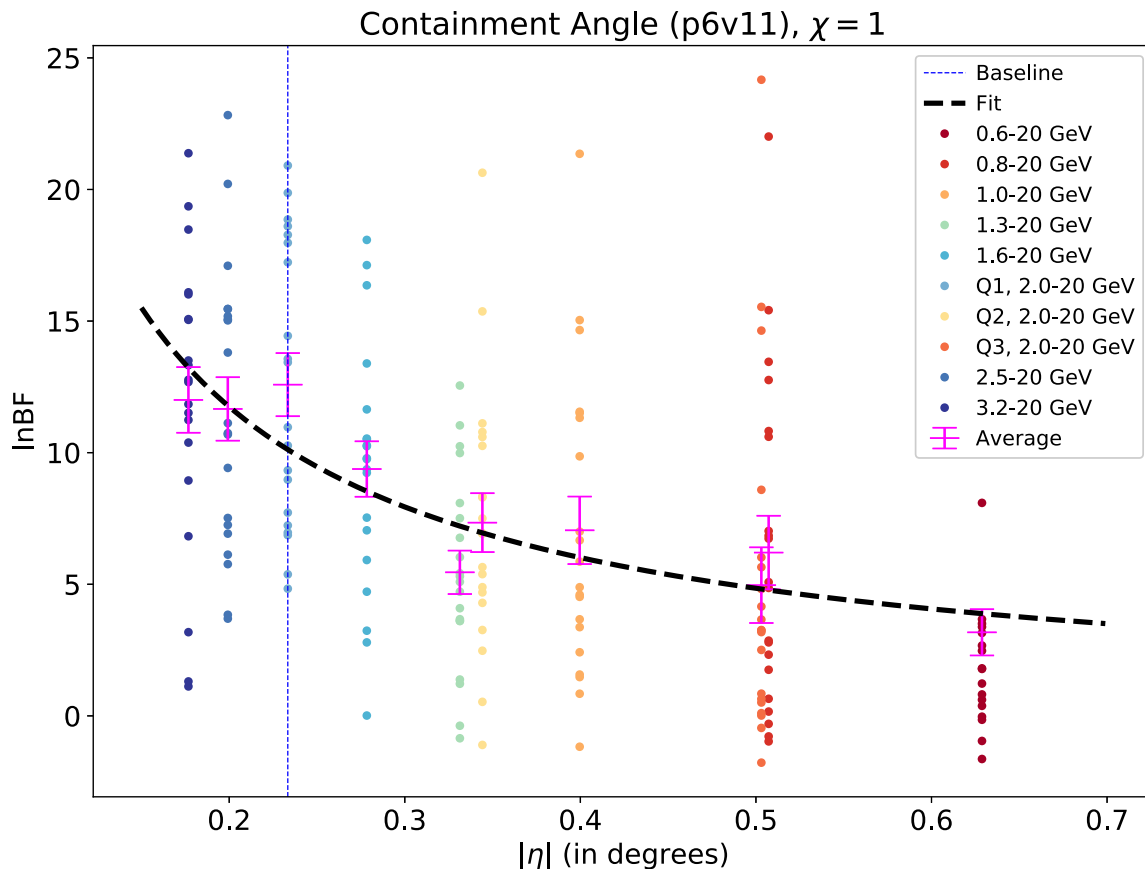


FIG. 6. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), as a function of the 68% containment angle of the PSF. Realizations are obtained by considering the top PSF quartile for different energy ranges; for the baseline 2–20 GeV range, we test each of the top three quartiles individually. The best-fit line (black) is the power-law fit to the $\langle \ln \text{BF} \rangle$ obtained using least-squares. The baseline case (blue vertical line) denotes the baseline case 2–20 GeV for quartile 1, typically used in previous *Fermi* analyses.

than a factor of two, this will generally be a beneficial tradeoff.

Note that our choice of $n_{\text{side}} = 128$ corresponds to a mean pixel spacing (0.46°) that exceeds or is comparable to the 68% containment angle for all but the widest energy range (0.6–20 GeV) that we consider. We will show in Sec. VI C that decreasing the pixel size below the PSF does not appear to have large effects on the expected sensitivity (although it can increase the variance), but studies focusing on a broader energy range might still wish to test smaller n_{side} values to reduce leakage of PSs into neighboring pixels.

TABLE III. Recovered parameters and the 1σ error for the power-law fit to $\langle \ln \text{BF} \rangle$ as a function of 68% containment angle.

Parameter	p6v11
a (coefficient)	2.5 ± 0.6
b (power)	-1.0 ± 0.2

3. Simultaneous variation of exposure and angular resolution

To check the stability of the scaling rules we have found so far and the validity of this simple estimate, we now explicitly test the effect of simultaneously varying the angular resolution and the exposure. In many realistic situations, and in particular for *Fermi* data, relaxing cuts on photon quality will simultaneously increase the effective exposure and worsen angular resolution.

We repeat the analysis described in Sec. VI A for simulated data using the appropriate PSF model for PSF quartiles 2 and 3, with 20 realizations for each combination of χ and quartile. We scanned the realizations at $n_{\text{live}} = 300$. Our results for $\langle \ln \text{BF} \rangle(\chi)$ for each quartile are summarized in Fig. 7. As in Eq. (39), we fit the data for each quartile with a power law with a constant offset, and provide the best-fit parameters in Table IV.

In general we observe that the slope appears to become steeper (more rapid increase in sensitivity with exposure) in quartiles with worse angular resolution. This reflects that significant detection of PSs requires a higher χ value when

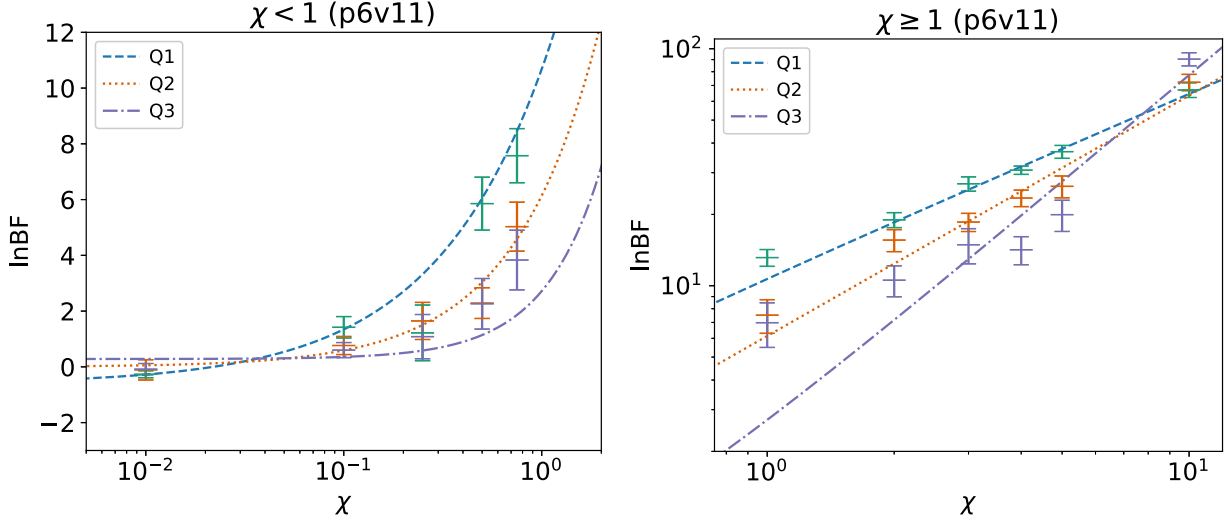


FIG. 7. $\langle \ln \text{BF} \rangle$ and the 1σ standard error of the mean across 20 realizations for the top three quartiles graded in angular resolution sampling different values of χ . Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. The best-fit lines show the Eq. (39) fit to the data.

the angular resolution is worse, but for sufficiently large χ factors, the significance becomes almost independent of angular resolution. This may be related to the pixels surrounding a PS becoming bright enough to be individually detected as significant PSs.

We can also test the effect of stacking together the simulated data corresponding to the different quartiles, which has the effect of increasing the effective exposure $\chi_{\text{eff}} > \chi$ relative to the one-quartile case. The sum of the first and second quartile has $\chi_{\text{eff}} = 2$, and the combined top three quartiles have $\chi_{\text{eff}} = 3$.

TABLE IV. Best-fit parameters obtained using least-squares regression method for the power law fit to the $\ln \text{BF}$ values (based on 20 realizations for each point) as we varied the source brightness of the GCE PS component in p6v11. Parameter uncertainties obtained from the standard error of the mean across 20 realizations are also displayed.

	$n = 1/4$	$n = 1/2$	$n = 1$
<i>Quartile 1</i>			
α (coefficient)	3.18 ± 0.70	7.35 ± 0.49	11.30 ± 0.69
β (power)	1.43 ± 0.13	0.97 ± 0.05	0.76 ± 0.04
γ (shift)	0.0003 ± 0.31	-0.40 ± 0.14	-0.62 ± 0.20
<i>Quartile 2</i>			
α (coefficient)	0.48 ± 0.18	3.97 ± 0.48	6.14 ± 0.78
β (power)	2.24 ± 0.20	1.22 ± 0.07	1.01 ± 0.08
γ (shift)	0.11 ± 0.26	-0.34 ± 0.18	-0.002 ± 0.37
<i>Quartile 3</i>			
α (coefficient)	0.02 ± 0.02	0.88 ± 0.39	2.44 ± 0.96
β (power)	3.67 ± 0.40	2.01 ± 0.22	1.50 ± 0.21
γ (shift)	0.18 ± 0.30	-0.23 ± 0.19	0.30 ± 0.37

Figure 8 shows the sensitivity based on 20 realizations for each of these three cases scanned at $n_{\text{live}} = 300$. We find that to quite a good approximation, the increased number of photons simply cancels out the effects of worsening the angular resolution on average. As a simple estimate we calculated the combined effects of the predictions of varying

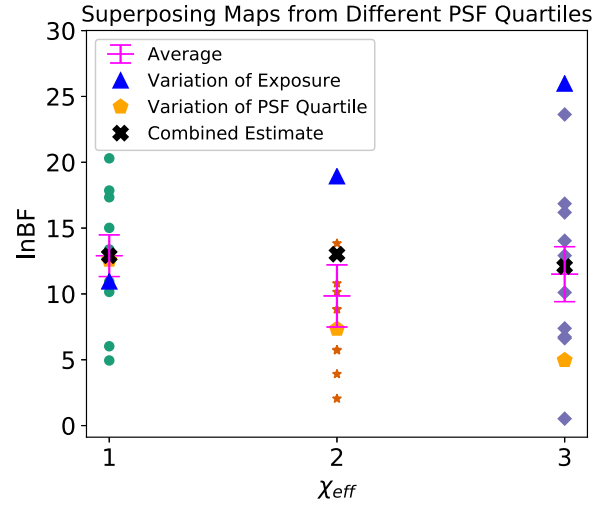


FIG. 8. $\ln \text{BF}$ across 10 realizations (circle/star/diamond markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), stacking increasing numbers of skymaps generated with different angular resolutions (corresponding to the PSF quartiles). All scans assumed the angular resolution of the worst included quartile. The blue triangles indicate the increased sensitivity as predicted by varying the exposure, while the orange pentagons indicate the expected worsening of sensitivity due to angular resolution degradation. The black filled “X” symbols display an estimate for the overall sensitivity change from combining the two (see text for details).

the exposure and angular resolution. To do so, we define a rescaling factor $r = r_{\text{exposure}}(\chi)r_{\text{PSF}}(\text{Quartile})$, where $r_{\text{exposure}}(\chi)$ is the ratio of the expected log BF at exposure χ , denoted $\langle \ln \text{BF} \rangle(\chi)$, to the baseline expected log BF $\langle \ln \text{BF} \rangle(\chi = 1)$, as obtained from Eq. (39) and Table II. Thus $r_{\text{exposure}}(\chi)$ characterizes the increase in sensitivity with enhanced exposure. $r_{\text{PSF}}(\text{Quartile})$ is the ratio of the expected log BF for a specified single quartile, denoted $\langle \ln \text{BF} \rangle(\text{Quartile})$, to the baseline expected log BF $\langle \ln \text{BF} \rangle(\text{Quartile} = 1)$, obtained from Fig. 5. Thus $r_{\text{PSF}}(\text{Quartile})$ characterizes the decline of sensitivity with worsening angular resolution. $\langle \ln \text{BF} \rangle(\chi)$ and $\langle \ln \text{BF} \rangle(\text{Quartile})$ are denoted as blue stars and orange pentagons on Fig. 8, respectively. To obtain the combined estimate denoted by the black filled “X,” we multiplied the calculated r factor with the baseline value $\langle \ln \text{BF} \rangle(\text{Quartile} = 1)$ obtained from the realizations for the baseline case in Fig. 8. We find that this estimate agrees with the simulation results that on average adding quartiles with worse angular resolutions to gain exposure does not yield large increases (or decreases) in the average sensitivity to PSs.

Quantitatively, Q3 has a containment angle slightly more than twice that of Q1 (see Fig. 6), while including Q2 and Q3 triples the exposure. The scaling of the sensitivity with exposure is slightly sublinear whereas for containment angle it is linear to a good approximation, and in practice we find that these two effects almost completely cancel out. Thus the overall sensitivity is (perhaps surprisingly, and somewhat coincidentally) insensitive to the inclusion of additional quartiles.

We might wonder if by approximating the PSF in the stacked dataset as the PSF of the worst quartile, we introduce biases in the recovered parameters. We checked this explicitly over our sample of 20 realizations. On average, we find that the median parameter deviations were rather small, mostly at $< 1\sigma$ level, with some exceptions for components such as the isotropic emission. However, we also checked cases where we stacked different realizations of the same quartile, so that the PSF was the same between different subpopulations and was thus modeled in the same way for the simulated data and the fit. We found, on average, that the biases were similar in these cases; they were not obviously worsened by stacking maps with different PSFs, while modeling with the worst PSF. Thus, the misreconstruction cannot be attributed to mismodeling of the PSF in a subset of the data.

4. Varying the energy range

While we have previously explored the effect of changing the PSF to one appropriate for other energy ranges, we now explore the effect of changing the energy range itself. We kept the upper limit of the energy range fixed at 20 GeV, since high-energy photons are rare and their inclusion/exclusion is unlikely to qualitatively change the results. We varied the low-energy limit of the energy range between

0.6–3.2 GeV, spanning the peak of the GCE, by including or excluding low-energy bins. As discussed previously, the bin boundaries are log-spaced in energy, with 10 bins per decade, starting at 0.2 GeV.

As a first test, we sought to understand how the sensitivity could be expected to vary just as a result of the modified angular resolution combined with the larger number of photons in low-energy bins. To explore this question, we held the underlying physical model fixed, and treated the enhanced number of photons as an effective exposure factor χ_{eff} , while using the appropriate PSF for the lowest-energy photons in the analysis. Specifically, we took χ_{eff} to be the ratio of the total number of photons in the real data (over the whole sky) in the modified energy range, to the total number of photons in the original energy range.

Figure 9 shows the results of this test. The results indicate that due to the worse angular resolution obtained by including data from lower energy ranges, we expect at best a mild increase in the (expected) sensitivity, compared with a substantial increase in the case where only the exposure is varied. As a first-order comparison to our simulated results, we analyzed the combination of the effects of varying the exposure and PSF as discussed in Secs. IVA and VIB 1. Similar to Sec. VIB 4, we define a rescaling factor $r = r_{\text{exposure}}(\chi)r_{\text{PSF}}(\eta)$, where $r_{\text{exposure}}(\chi)$ is the ratio of $\langle \ln \text{BF} \rangle(\chi)$ to $\langle \ln \text{BF} \rangle(\chi = 1)$, obtained from Eq. (39) and Table II. $r_{\text{PSF}}(\eta)$ is the ratio of $\langle \ln \text{BF} \rangle(\eta)$ to

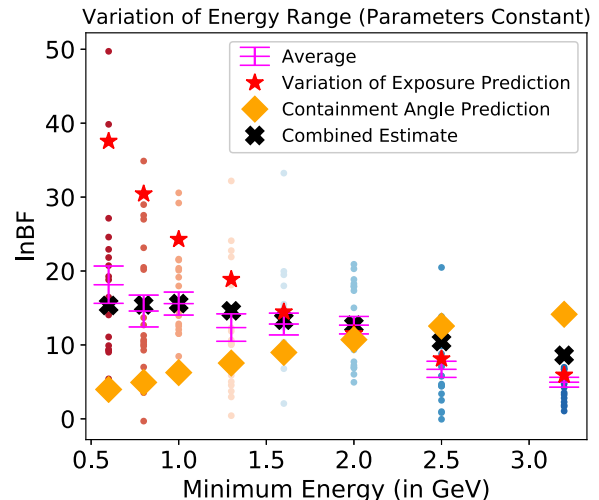


FIG. 9. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), varying minimum energy while holding the parameters constant. All scans employ the angular resolution of the lowest-energy photons in the range. The increase in photon counts is captured through an effective exposure factor χ_{eff} . The red stars display the predicted effect of this exposure variation, while the orange diamonds predict the effect of changing the containment angle. The black filled “X” symbols display an estimate for the overall sensitivity change from combining the two (see text for details).

$\langle \ln \text{BF} \rangle(\eta[2-20 \text{ GeV}])$ (baseline case), obtained from Fig. 6. $\langle \ln \text{BF} \rangle(\chi)$ and $\langle \ln \text{BF} \rangle(\eta)$ are denoted as red stars and orange diamonds on Fig. 9, respectively. To obtain the combined estimate denoted by the black filled “X,” we multiplied the calculated r to the $\langle \ln \text{BF} \rangle[2-20 \text{ GeV}]$ obtained from the simulations in Fig. 9. The estimates indicate a fairly flat scaling behavior of sensitivity across different energy ranges. This suggests that the beneficial effects of increasing sensitivity are canceled out by the worsening of angular resolution at lower energy ranges.

As an example, consider varying the minimum energy of the event selection from 2.0 to 1.0 GeV. The 68% containment angle of the PSF increases from the baseline 0.23° to 0.40° . As shown in Fig. 6, this change in PSF induces a decrease in $\langle \ln \text{BF} \rangle$ by a factor of 0.58. However, the larger number of photon counts with a minimum energy of 1.0 GeV corresponds to an effective rescaling factor of $\chi_{\text{eff}} = 2.75$ relative to the case with minimum energy 2.0 GeV (ignoring differences in the spectrum between the different components). Therefore, based on Table II, the value of $\ln \text{BF}$ should increase by a factor of ~ 2.10 , if this exposure change were the only factor. The combined effect would correspond to only a $\sim 22\%$ increase in $\ln \text{BF}$. Thus in this case, we would expect the increase in sensitivity from additional photons to come close to offsetting the loss of angular resolution, leading to very little net change in sensitivity (with perhaps a slight advantage for a 1.0 GeV minimum energy). This resembles the roughly flat behavior with energy we actually observe in Fig. 9.

This is how the sensitivity would behave if the signal and backgrounds had identical spectra, but of course this is not the case for the GCE. For a specific signal, such as the GCE in this case, we need to either input a theoretical spectrum for each component, or re-fit the model parameters from the real data in each energy band. We take the latter approach here, and then repeat the sensitivity analysis on data simulated using these updated, energy-dependent parameters.

Figure 10 shows the result of these simulations and analyses. If only photon number and angular resolution were relevant, there would be a strong argument for extending the energy range for the analysis all the way down to 0.6 MeV (or lower), but for the actual GCE spectrum we observe that the highest expected sensitivity is obtained for a minimum energy of 1.0 or 1.6 GeV. This energy scale roughly coincides with the peak of the GCE distribution.

One might ask if features in Fig. 10 simply reflect fluctuations in the total GCE flux inferred from the real data in different energy ranges (used to fix the simulation parameters). We checked this explicitly and found no evidence of such an association; the parameters controlling the simulated GCE PS flux vary smoothly over the relevant range of threshold energies, and the fluctuations in Fig. 10 are thus likely to be statistical.

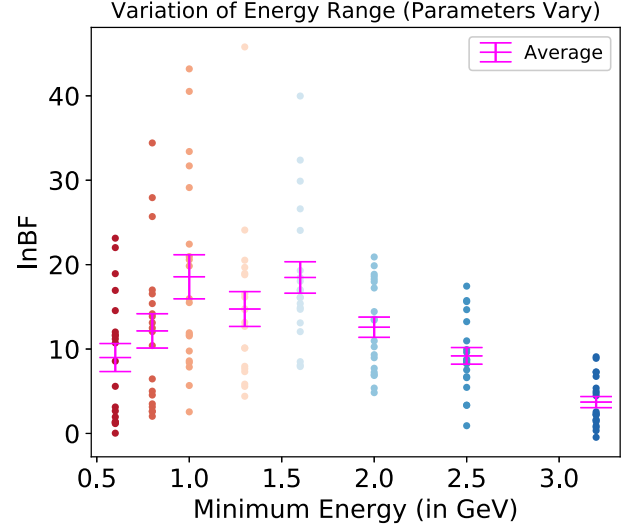


FIG. 10. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), varying the minimum energy and updating the template parameters to match the posterior median parameters from fits to the real data in the same energy range. All scans employ the angular resolution of the lowest-energy photons in the range.

C. Pixel size variation

We examined a wide range of pixel size to determine an optimal value for analysis. We started at an n_{side} value of 512 and downgraded to n_{side} values of 256, 128, 64, 32. At each pixel level, we computed $\ln \text{BF}$ across 20 realizations.

Figure 11 shows the recovered sensitivity as a function of n_{side} . We find that there is a slight increase in the

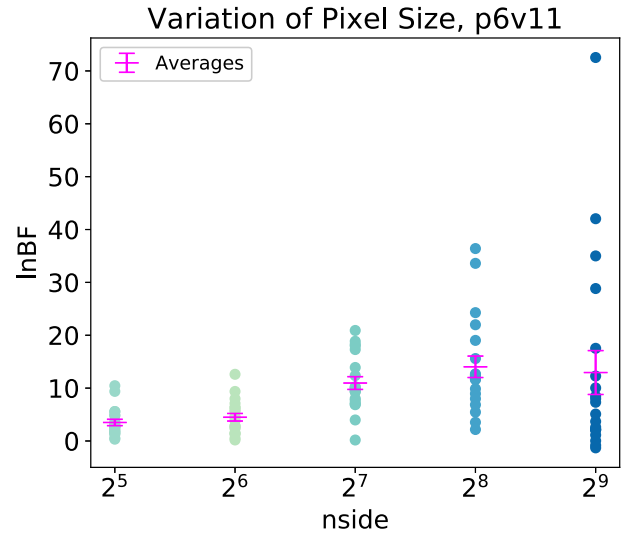


FIG. 11. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), varying the n_{side} parameter. The pixel area varies inversely with n_{side} .

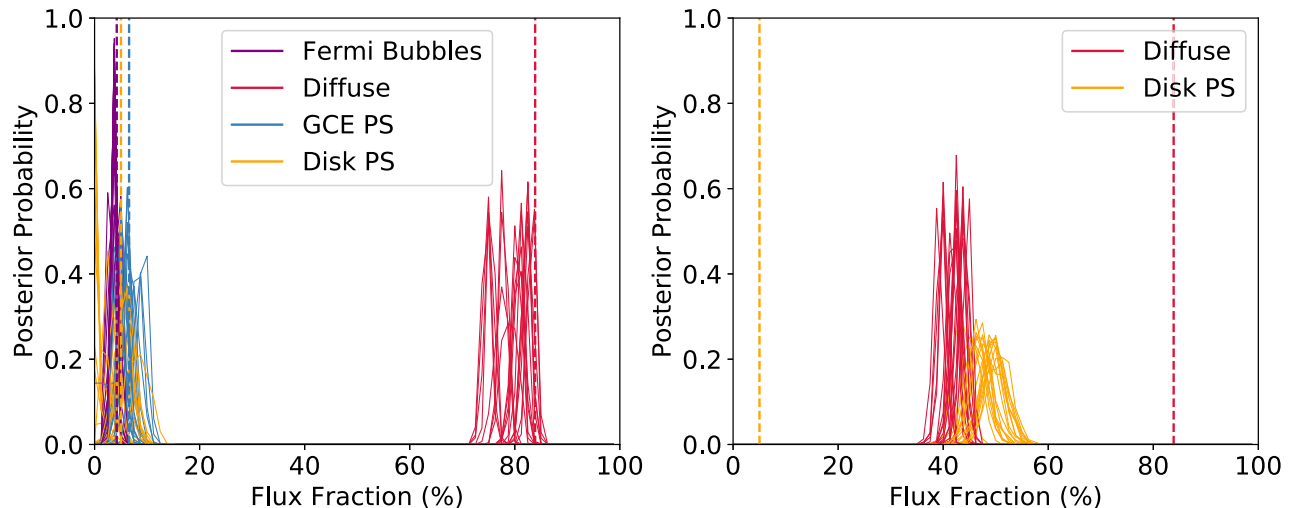


FIG. 12. Left: Flux fraction plot at the common pixel size employed in previous analyses ($n_{\text{side}} = 128$) of certain emissions: *Fermi* bubbles (purple), diffuse emission (red), GCE PS (blue), disk PS (orange). Right: flux fraction plot demonstrating misreconstruction of the diffuse (red) and disk PS (orange) components at a low n_{side} value ($n_{\text{side}} = 32$). The recovered flux fraction are obtained from 20 simulations. The vertical dashed lines denote the flux fraction injected into the simulations.

sensitivity to a population of PSs as resolution is improved. However, the scatter between individual realization is also increased. It is plausible that this occurs because with small pixel sizes there is a greater risk that a relatively-bright source happens to land near a pixel boundary and consequently loses significance. In realizations where this behavior happens to be rare, the significance is naturally higher than for smaller n_{side} (as the likelihood contributions from a larger number of pixels are summed), but in other realizations this effective dimming of the sources will markedly decrease the inferred significance of the population. (An alternative way to think about this is that pixels much smaller than the PSF are not independent data points, and so by treating them as independent we may artificially enhance the apparent significance of the result [11],⁵ but may also miss correlations that could reveal a signal.) In the regime where the pixel size is significantly larger than the PSF, we expect the significance to be reduced because we have reduced the number of independent data points (pixels), discarding information in the process.

It appears that $n_{\text{side}} 128$ and 256 are likely the optimal values: $n_{\text{side}} 256$ has a slightly higher average sensitivity but with considerably more scatter between realizations. The relative insensitivity of NPTF methods to pixel size, in a simpler context, was previously studied in Ref. [1].

On the other end of the spectrum, low n_{side} values cause severe discrepancies in the recovered flux fraction for Galactic diffuse emission and PS populations associated with the Galactic disk (Disk PS). At an n_{side} value of 32 , for example, a significant portion of the injected Galactic diffuse emission is absorbed into the Disk PS template. In

contrast, at the standard choices of pixel size (such as $n_{\text{side}} = 128$), the recovered flux fractions are fairly consistent with the injection values. The left panel of Fig. 12 shows the recovered flux fractions and the injected values (dashed lines) for $n_{\text{side}} = 128$ for the *Fermi* bubbles, diffuse, GCE PS, and disk PS contributions. The right panel of Fig. 12 displays the misreconstruction of the diffuse and disk PS emissions due to a large pixel size ($n_{\text{side}} = 32$). The recovered flux fractions are systematically biased across all realizations, even in this case where the model is correct by construction. Thus low- n_{side} analyses should be treated with considerable caution, in addition to their lack of sensitivity.

D. Source brightness

Finally, we would like to understand how sensitivity scales with the brightness of the sources. This is important both for understanding the prospects for future detection of a bulge PS population with *Fermi*, and in understanding the likely sensitivity of future telescopes with different exposure and angular resolution.

We perform the same test as described in Sec. VI B 3, where the default source brightness parameter for the GCE PS component $S_b = 17.11$ is multiplied by a factor of $n = 1/2$ or $1/4$ (and the number of sources is modified to keep the total flux constant). Figure 13 shows the results, while Table IV shows the corresponding best-fit parameters and uncertainties obtained from the power-law fit in Eq. (39), using the least-squares method.

Our results indicate that the parameter (of the power-law fit) α increases at higher values of n . Since α corresponds to the sensitivity at $\chi = 1$, it is expected to observe an increase as the brightness of the sources increase. On the other hand,

⁵We thank Nicholas Rodd for pointing out this effect.

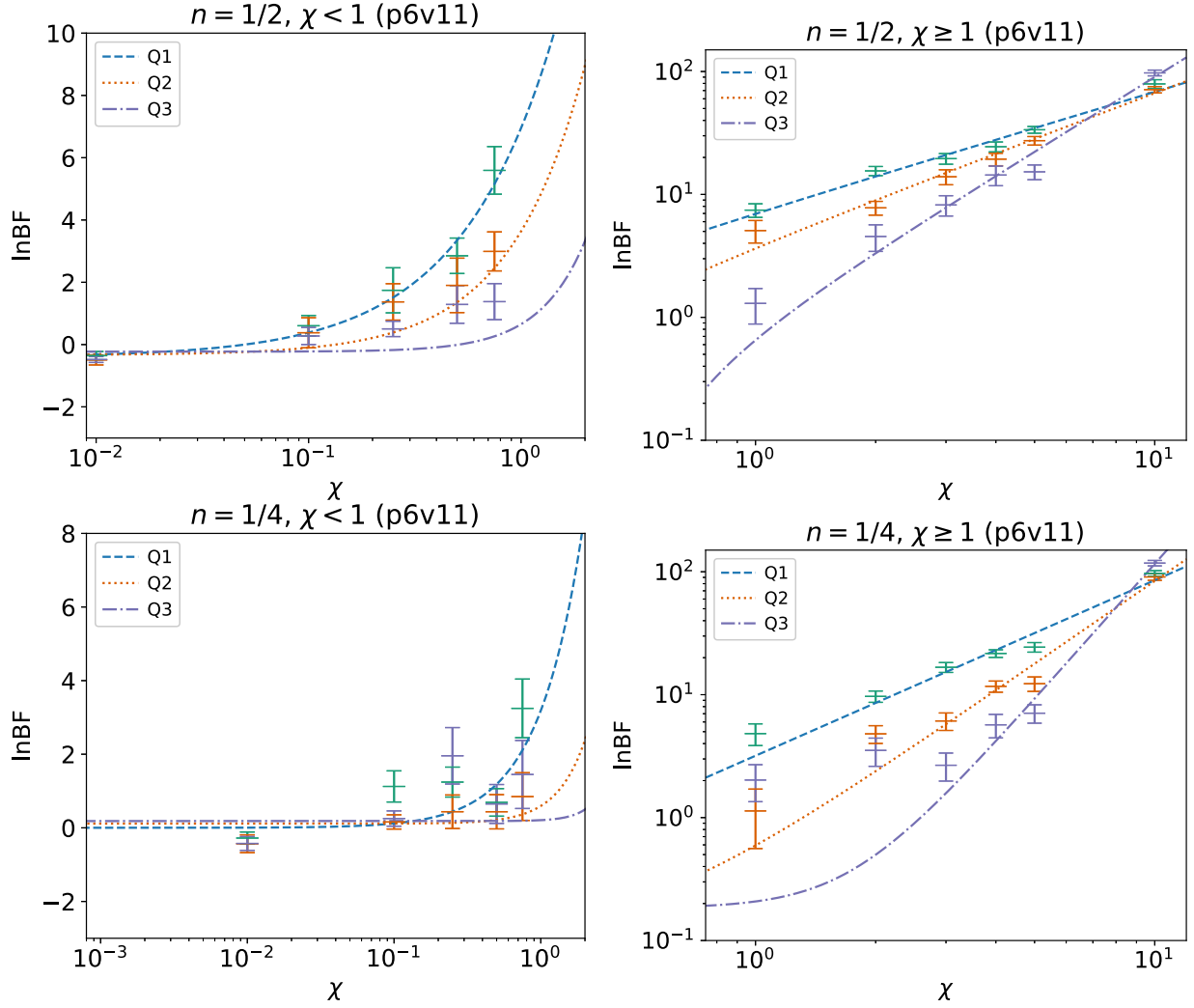


FIG. 13. $\langle \ln \text{BF} \rangle$ and the $\sigma/\sqrt{20}$ standard error of the mean across 20 realizations at different χ values for the top three quartiles graded by angular resolution. Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. In the upper panels we simulated the sources at half of the baseline brightness ($S_b = 17.11$ from Table V), but doubled the number of sources, in accordance with Eq. (33); in the lower panels we simulated the sources at 1/4 of the baseline brightness but quadrupled the number of sources. The dashed, dotted, and dash-dotted lines correspond to the shifted power law [Eq. (39) for quartiles 1, 2, and 3, respectively].

β decreases modestly as n increases. Assuming that $\langle \ln \text{BF} \rangle$ has the same functional behavior as $\langle \Delta \ln \mathcal{L} \rangle$, our expectation from the analytic results is $\beta \approx 1$, at least if the fit is dominated by the region where the expected number of counts/source and hence $\langle \ln \text{BF} \rangle$ is large. However, at low brightness levels, as predicted by the analytic equations, we expect a quadratic scaling, $\beta \approx 2$. Thus it is reasonable to see a stronger scaling for smaller values of n , where the quadratic behavior is relevant for a larger range of χ . In other words, increased exposure is more important for fainter sources.

To clearly demonstrate the effect of varying the source brightness on the overall sensitivity, we plotted $\langle \ln \text{BF} \rangle$ ($\chi = 1$) across 20 realizations as a function of S_b , for the top three PSF quartiles, in Fig. 14. Within the S_b range we tested, we find that, on average, sensitivity increases as the

PS population brightens. The decrease in sensitivity qualitatively matches expectations based on the theoretical results presented in Sec. II, where the expected $\Delta \ln \mathcal{L}$ monotonically decreases as s is lowered, although we observe significant scatter across the different quartiles.

We eventually expect to lose all sensitivity to point sources as their brightness becomes sufficiently small; as $S_b \rightarrow 0$, it becomes impossible to distinguish the PS flux and the smooth emission, and we expect the effects of priors to dominate the results (see Appendix C for a study of the prior dependence). Roughly speaking, we expect the onset of this regime to occur around $S_b \sim 1$ (as for sources with counts $s \ll 1$ it will be rare to observe multiple photons from a single source); more quantitatively, we observe from the analytic approximations [Eq. (19)] that the log likelihood difference per pixel becomes rapidly

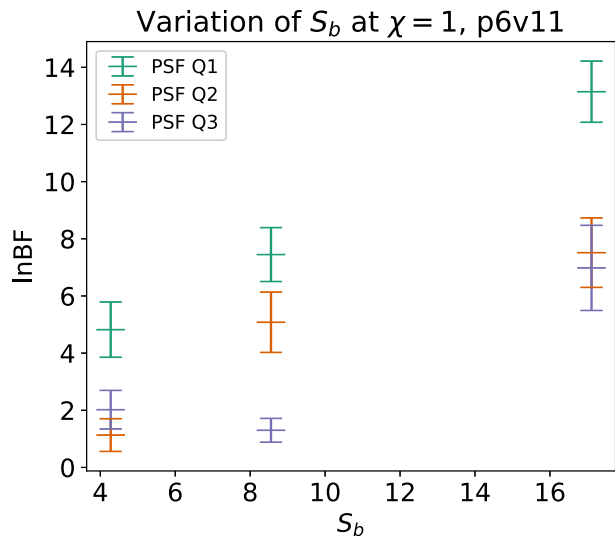


FIG. 14. $\langle \ln \text{BF} \rangle$ and the $\sigma/\sqrt{20}$ standard error of the mean (obtained from 20 realizations at each point) across the top three quartiles at varying brightness level for the full *Fermi* case, with the baseline exposure. The best-fit S_b in the real data is 17.11; we test the effect of reducing S_b by a factor of 2 or 4. The baseline case was scanned using $n_{\text{live}} = 300$, while the other cases were performed using $n_{\text{live}} = 100$ for computational efficiency.

smaller once ks drops below 1. (Nonetheless, given a sufficiently large number of pixels, it may still be possible to probe the properties of sources with $s \lesssim 1$, as discussed in e.g. [50].)

Approximating the sensitivity-exposure relation by the power law with an additive shift as in Eq. (39), we can estimate the χ value that corresponds to $\ln \text{BF} \geq 1$ for Quartile 1, as a function of the brightness of the source population (indicating some sensitivity to the sources; we could increase this threshold to require a more significant detection). Using the best-fit parameters in Table IV, we find that for $n = 1$, this threshold corresponds to $\chi \geq 0.08$; for $n = 1/2$, to $\chi \geq 0.18$; and for $n = 1/4$, to $\chi \geq 0.45$. As expected, a higher level of exposure is required to detect fainter populations of PSs.

VII. CONCLUSIONS

We have investigated the statistical behavior of non-Poissonian template fitting, as implemented in the NPTFit public code, when characterizing unresolved point sources. In particular we have explored the sensitivity to point sources both analytically and numerically, in a simplified isotropic-emission scenario and a realistic scenario relevant to gamma rays from the inner Galaxy. We define the sensitivity to point sources (or detectability of point sources) as the ratio of the maximum likelihood (analytic case) or Bayesian evidence (numerical case) between the true underlying model versus a model that excludes point sources associated with the signal

component. We first derived analytic estimates of the sensitivity for point sources with a delta-function SCF, where all sources have the same expected number of photons per source s . We found that the expected contribution to the log likelihood ratio from a given pixel is a function only of ks , where k is the fraction of the emission that is attributed to point sources; the scaling of the log likelihood with ks is linear for $ks \gg 1$ and quadratic for $ks \ll 1$. We also examined the variance in this sensitivity, reflecting the expected scatter between realizations. By exploring a range of scenarios, we found that the standard deviation of our sensitivity metric was generically smaller than the expectation value by a factor of \sqrt{A} , where A can be the total number of sources in the ROI, the total number of pixels in the ROI, or the log likelihood ratio itself; in general, whichever parameter is smallest dominates the variance. This behavior can lead to a relatively large scatter between the sensitivity inferred from different simulations, which we indeed observe in the numerical data. We tested our analytic predictions using numerical simulations in a simplified case where both point sources and smooth emission are isotropic (detailed in Sec. V and Appendix D), and found that the analytic results were quantitatively quite accurate in the case where the PSF is very narrow, and provide a good description of various scaling relations even with a more realistic PSF. The analytic and isotropic results may be relevant to other analyses employing non-Poissonian template fitting, e.g. the analysis of the neutrino background presented in [54].

We then numerically investigated the role of several key parameters in the NPTFit analysis of a population of point sources associated with the *Fermi* gamma-ray skymap. The parameters we tested included exposure, angular resolution, source brightness and pixel size. This analysis was performed within the full *Fermi* scenario using both p6v11 templates and Model A templates for the Galactic diffuse emission (detailed in Appendix E). The results we quote below are based on simulations with the default p6v11 template from the public NPTFit code, but we found consistent results using an alternative background model denoted Model A.

For the cases we tested, we found the following general relationships between exposure, angular resolution, and sensitivity:

- (i) Gaining exposure alone induces an increase in sensitivity that is roughly linear for the exposure range and analysis choices we focused on. The analytic approximation predicts a scaling between linear and quadratic; our results are broadly consistent with this expectation, but in complex/realistic sky models involving multiple templates with different morphologies, we have observed both slightly sub-linear scaling and stronger-than-quadratic scaling, the latter in quartiles with poor angular resolution.

- (ii) Worse angular resolution results in lower sensitivity (as expected) such that the third-highest-quality angular resolution quartile has average sensitivity approximately 60% lower than of the highest quality quartile (although the degree of degradation can vary depending on the exposure level and the brightness of the sources, with the loss of sensitivity being more pronounced for fainter sources and lower exposure).
- (iii) As a simplified parametrization of the angular resolution (which varies according to quartile selection and included energy range), we examined how the sensitivity varies according to the 68% containment angle radius η , and found that sensitivity has an approximate inverse proportionality relation to η .
- (iv) With these results, we tested the effects of increasing exposure at the expense of angular resolution in practical situations relevant to *Fermi* data, by varying the quartile selection and energy range. We find, on average, that the increase in sensitivity from a larger exposure is offset by the degraded angular resolution to a good approximation.
- (v) Hence, this tradeoff only produces a small net change in sensitivity; there is a broad range of possible analysis choices that yield similar expected sensitivity. The large scatter in sensitivity between realizations means that some caution is needed when interpreting changes in the Bayes factor in real data upon addition of extra quartiles, energy bins, ROI, etc; it appears possible for changes in the analysis choice that modify the dataset to have a large apparent effect on the Bayes factor purely due to this scatter.

Note that the properties (for the exposure vs angular resolution tradeoff) identified above are specific to the current implementation of NPTFit, which assumes that all photons share the same angular resolution (fixed by the worst angular resolution in the dataset). It is possible that a more sophisticated approach that tracks angular resolution separately for different categories of photons could lead to a greater benefit from including lower-angular-resolution quartiles or lower-energy photons.

We also examined the role of source brightness to the sensitivity of NPTFit to point sources to understand how the sensitivity falls off as the source brightness declines. For sufficiently faint point sources, as expected, NPTFit is unable to distinguish point sources from background smooth emission. More specifically, in the top graded quartile for angular resolution, the minimum exposure (in relation to the baseline exposure) required to achieve any hint of detection (defined somewhat arbitrarily as an average Bayes factor $\ln \text{BF} \geq 1$) scales as $O(S_b^{-1.28})$, where S_b is the peak number of photons per Galactic Center excess point source.

We explored simulations with different pixel sizes, and found that pixel size was not a crucial factor in determining the sensitivity of NPTFit, except in the case of extremely large or small pixel sizes. Very large pixel sizes induced inaccurate recovery of the various physical emission components; small pixel sizes led to higher variance across realizations.

These results serve as a systematic demonstration of the behavior of NPTFit under a range of conditions relevant to analyses of *Fermi* data from the inner Galaxy (and potentially more broadly).

ACKNOWLEDGMENTS

The authors thank Siddharth Mishra-Sharma, Nicholas Rodd, and the anonymous referee for helpful feedback and suggestions. The work of T. R. S. was supported by the US Department of Energy, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under grant Contract No. DE-SC0012567 through the Center for Theoretical Physics at MIT; the National Science Foundation under Cooperative Agreement No. PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, [55]); and a grant from the Simons Foundation (Grant No. 929255, T.R.S.). The work of L. G. C. B. was funded by the Massachusetts Institute of Technology Undergraduate Research Opportunities Program (UROP) office.

APPENDIX A: DETAILED METHODOLOGY

1. Analysis templates

In this appendix we detail the spatial templates we used in our analyses, and the physical processes they attempt to capture. For our primary (p6v11) analysis we largely used the templates that were publicly released with the NPTFit code package, as we summarize below. In Appendix E we test the effects of using an alternative pair of templates to describe the Galactic diffuse emission (vs a single template in the p6v11 analysis), collectively denoted Model A.

(i) *Fermi* bubbles

The *Fermi* bubbles are large lobes of extended hard gamma-ray emission that appear to be emanating from the vicinity of the GC [56]. The exact origin of this large structure is currently unknown. The default template in NPTFit is based on Ref. [56] and has a uniform intensity within the bubbles region (prior to exposure effects). This template has a single associated degree of freedom, its normalization parameter A_{bub} .

(ii) Isotropic background

An appreciable fraction of the overall gamma-ray emission is expected to originate from extragalactic sources and possess an approximately isotropic distribution [57–60]. We thus include a template for emission that is isotropic across the full sky

(prior to applying exposure effects). This emission includes photons from both relatively bright sources and much fainter sources, so we include two isotropic components, one representing smooth emission and one representing PSs.

The smooth isotropic template has a single degree of freedom: its template normalization parameter A_{iso} . The non-Poissonian template for isotropic PSs has one degree of freedom for its normalization ($A_{\text{ISO}}^{\text{PS}}$) and three degrees of freedom associated with its SCF, as described by Eq. (31), $\{n_1^{\text{ISO-PS}}, n_2^{\text{ISO-PS}}, S_b^{\text{ISO-PS}}\}$. Note that we label PS templates with capital letters and Poissonian templates with lower-case letters.

(iii) Galactic Center excess (NFW profile)

The template for the GCE is constructed from the line-of-sight projection of the square of a generalized Navarro-Frenk-White (NFW) profile [9,61,62]. The explicit form of the density profile of the generalized NFW for the Milky Way is

$$\rho(r) \propto \frac{(r/r_s)^{-\gamma}}{(1+r/r_s)^{3-\gamma}}, \quad (\text{A1})$$

where $r_s = 20$ kpc, $\gamma = 1.25$. The line of sight projection to determine the 2D spatial flux distribution is

$$J(\psi) = \int_0^\infty \rho^2(r) ds, \quad (\text{A2})$$

where ψ is the angle away from the GC, s is the line-of-sight distance, and $r^2 = R^2 + s^2 - 2Rs \cos \psi$ where R is the distance between the Earth and the GC.

As for the isotropic emission, we use this spatial morphology for two different emission components, one smooth and one representing a PS population. The smooth isotropic template has a single degree of freedom: its template normalization parameter A_{gce} . The non-Poissonian template for GCE-distributed PSs has one degree of freedom for its normalization ($A_{\text{GCE}}^{\text{PS}}$) and three degrees of freedom associated with its SCF, as described by Eq. (31), $\{n_1^{\text{GCE-PS}}, n_2^{\text{GCE-PS}}, S_b^{\text{GCE-PS}}\}$.

(iv) Galactic diffuse emission

The Galactic diffuse emission is the dominant contributor to the total gamma-flux in the energy range relevant to *Fermi*. There are three main contributors to this emission: (1) proton collisions with the gas, which produce pions that decay to photons, $pp \rightarrow X + \pi^0 \rightarrow X + \gamma\gamma$, (2) inverse

Compton scattering (ICS) driven by cosmic ray electrons upscattering abundant low-energy photons in the interstellar radiation field, and (3) cosmic ray electrons scattering on the ambient gas, which produces photons via bremsstrahlung. The first two processes generally dominate the total emission.

In our default analysis, the diffuse emission is modeled as a single smooth template similar to the `p6v11` diffuse model packaged with the public version of `NPTFit`, but accounting for 573 weeks of data. This template has a single degree of freedom: its template normalization parameter $A_{\text{dif-p6v11}}$. This template is quite old and is known to have significant deficiencies for modeling the real Galactic sky (e.g. [43]), so in Appendix E we tested the effect of replacing this template with two separate templates based on the `Model A` of Ref. [20]. One template corresponds to the π^0 and bremsstrahlung emission (which traces the gas), the other to the ICS component. Each template has a single degree of freedom corresponding to its normalization parameter; these parameters are denoted A_{pibrem} and A_{ics} respectively.

(v) Disk-correlated sources

The galactic disk of the Milky Way contains many gamma-ray PSs, which we model as having a doubly exponential thick-disk source distribution [3]. The 3D number density of sources is approximated as:

$$n(z, R) \propto \exp\left[\frac{-R}{5 \text{ kpc}}\right] \exp\left[\frac{-|z|}{1 \text{ kpc}}\right], \quad (\text{A3})$$

where R and z are respectively the radial distance from the GC and the height above the Galactic disk. This number density function is then integrated along the line of sight as described for the GCE template.

(vi) Template normalization

We normalize each of these templates according to the standard `NPTFit` convention. For smooth/Poissonian components, we store counts templates denoted by T , which include a factor of the exposure map (i.e. they are in counts not flux), are smoothed by the PSF, and are normalized to have an average of one count/pixel in the region within 30° of the GC where $|b| > 2^\circ$. The maps of expected counts per pixel for a smooth/Poissonian component are thus given by $A_{\text{smooth}}T$ for each component, and A_{smooth} describes the average number of counts/pixel in the normalization region. For PS components, the templates T_{PS} denote the number of sources per pixel rather than the number of photon counts per pixel, and so do not include the instrument response functions (i.e. PSF smoothing and multiplication

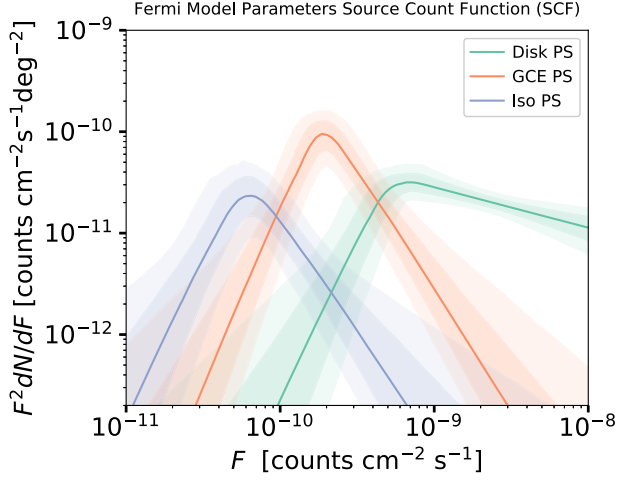


FIG. 15. The singly-broken power law form (median and the 68% and 95% containment bands) that characterizes the source count functions for the disk PS, GCE PS, and isotropic PSs components obtained from *Fermi* data collected until 2019. These functions were also used to describe and generate simulated data.

by the exposure map). The T_{PS} templates are normalized so that when T_{PS} is multiplied by the exposure map, the mean of the resulting map is equal to the mean of the exposure map (in the same normalization region as discussed above). The map of expected sources per pixel is proportional to this normalized T_{PS} map, with the normalization controlled by A_{PS} and the properties of the SCF, as in Eq. (31).

2. Analysis parameters

The baseline analysis parameters we used to create the sky map simulations were obtained by fitting the real *Fermi* data (with the data selection described in Sec. III A) using NPTfit. We perform the fit at $n_{\text{live}} = 500$ for greater accuracy, and we choose the ROI as described in Sec. III B. The priors used in this fit, which are also used to analyze simulation data, are given in Table VI. In Fig. 15, we plot the best fit source count functions (SCF) we obtained, which describe PSs we injected into our skymap simulations. Note, however, that these source count functions are quite peaked in flux, and the GCE SCF in particular can be treated approximately as a delta function (as we discuss in Appendix B).

The resulting posterior median parameters, which are used to generate simulations, are given in Table V. As we will discuss shortly, we repeated the analysis for two different choices of Galactic diffuse emission model, labeled p6v11 and Model A, which include different numbers of templates; consequently, some templates are only relevant for one of the two diffuse models.

TABLE V. Posterior median parameters extracted from data collected until 2019 by *Fermi*, and used to generate simulated data. The two columns correspond to (left) use of a single diffuse Galactic emission template, labeled “dif-p6v11,” and (right) use of two diffuse Galactic emission templates, labeled “pibrem” and “ics.”

Parameters	p6v11	Model A
$\log_{10} A_{\text{bub}}$	0.03	-0.03
$\log_{10} A_{\text{iso}}$	-2.16	-1.17
$\log_{10} A_{\text{gce}}$	-2.31	-2.09
$\log_{10} A_{\text{dif-p6v11}}$	1.22	...
$\log_{10} A_{\text{pibrem}}$...	0.97
$\log_{10} A_{\text{ics}}$...	0.84
$\log_{10} A_{\text{GCE}}^{\text{PS}}$	-2.47	-1.32
$n_1^{\text{GCE-PS}}$	4.39	4.48
$n_2^{\text{GCE-PS}}$	-1.57	-1.37
$S_b^{\text{GCE-PS}}$	17.11	5.56
$\log_{10} A_{\text{DSK}}^{\text{PS}}$	-3.73	-2.53
$n_1^{\text{DSK-PS}}$	2.40	2.28
$n_2^{\text{DSK-PS}}$	-1.23	0.08
$S_b^{\text{DSK-PS}}$	48.64	12.86
$\log_{10} A_{\text{ISO}}^{\text{PS}}$	-1.60	-4.92
$n_1^{\text{ISO-PS}}$	4.25	3.69
$n_2^{\text{ISO-PS}}$	-1.24	-0.57
$S_b^{\text{ISO-PS}}$	5.61	27.41

3. Simulation procedure

We generated exposure maps for our event selection using the *Fermi* Science Tools, and then simulated

TABLE VI. Prior ranges employed for analyses of both real and simulated data, unless specified otherwise in the text.

Parameter	p6v11	Model A
$\log_{10} A_{\text{bub}}$	[-3, 1]	[-3, 1]
$\log_{10} A_{\text{iso}}$	[-3, 1]	[-3, 1]
$\log_{10} A_{\text{gce}}$	[-3, 1]	[-3, 1]
$\log_{10} A_{\text{dif-p6v11}}$	[-3, 1]	...
$\log_{10} A_{\text{pibrem}}$...	[-2, 2]
$\log_{10} A_{\text{ics}}$...	[-2, 2]
$\log_{10} A_{\text{GCE}}^{\text{PS}}$	[-6, 1]	[-6, 1]
$n_1^{\text{GCE-PS}}$	[2.05, 5]	[2.05, 5]
$n_2^{\text{GCE-PS}}$	[-3, 1.95]	[-3, 1.95]
$S_b^{\text{GCE-PS}}$	[0.05, 80]	[0.05, 80]
$\log_{10} A_{\text{DSK}}^{\text{PS}}$	[-6, 1]	[-6, 1]
$n_1^{\text{DSK-PS}}$	[2.05, 5]	[2.05, 5]
$n_2^{\text{DSK-PS}}$	[-3, 1.95]	[-3, 1.95]
$S_b^{\text{DSK-PS}}$	[0.05, 80]	[0.05, 80]
$\log_{10} A_{\text{ISO}}^{\text{PS}}$	[-6, 1]	[-6, 1]
$n_1^{\text{ISO-PS}}$	[2.05, 5]	[2.05, 5]
$n_2^{\text{ISO-PS}}$	[-3, 1.95]	[-3, 1.95]
$S_b^{\text{ISO-PS}}$	[0.05, 80]	[0.05, 80]

realizations of the gamma-ray sky using the parameters of Table V, modified by any variations that we wanted to test. When we performed variations that shifted the simulated values of individual parameters, we shifted the relevant priors to ensure they were still appropriate; see the sections relevant to individual analyses for details.

We worked in Python 3.6.4. For templates representing smooth emission, we performed a standard random Poisson draw to obtain the counts per pixel; for non-Poissonian templates, we simulated sky realizations using the public code package NPTFit-Sim.

All simulations were done using HEALPix with $n_{\text{side}} = 128$, with the exception of tests described in Sec. VIC and analogous tests in the appendices). We simulated and saved realizations of each individual template, and then summed them to obtain mock realizations of the *Fermi* dataset.

APPENDIX B: DEGREE TO WHICH THE SCF APPROXIMATES A DELTA-FUNCTION

To demonstrate the sharpness of the peak of the SCF, we studied how the sensitivity of our NPTFit-based pipeline to GCE PSs varies as the SCF width of the GCE PS component narrows. In the SCF described by Eq. (31), we varied the value of n_2 while constraining n_1 to its posterior median value (4.39), which corresponds to a steep drop at high flux. We adjusted A_{PS} to preserve the total number of photons across each scenario.

We began by testing very negative values of $n_2 = -30, -20, -10$, corresponding to near-delta-function forms for the SCF. The left panel of Fig. 16 displays the resulting $\ln \text{BF}$ values (with 10 realizations at each n_2 value and $n_{\text{live}} = 300$). The sensitivity results were very stable as we varied n_2 in this range, indicating that our baseline choice

of SCF is not meaningfully different from a delta function for the purpose of sensitivity calculations. We then continued increasing n_2 to larger values (up to 0.8), as shown in the right panel of Fig. 16; the results remained relatively stable for $n_2 < 0$, but the sensitivity dropped off after that point, by approximately a factor of 2.

Our ability to test even broader SCFs is limited by our functional form for the SCF, which only has one break in the power law; increasing n_2 above 1 would mean the sources are predominantly low-luminosity, but in the absence of a low-end break in the SCF, this would yield an infinite number of sources. We attribute the lack of a precipitous sensitivity decline to the fact that we are restricted to $n_2 < 1$, where consequently both the number of sources and the overall photon flux are dominated by relatively bright sources, with fluxes in the neighborhood of S_b . For example, at $n_2 = 0$, 45% of flux comes from sources above S_b , and even at the highest n_2 value we tested $n_2 = 0.8$, 33% of flux is drawn from sources above S_b . To see a very large decrease in sensitivity from increasing n_2 (rather than lowering S_b), we would likely need to allow for more flexible SCF prescriptions with a large n_2 but a low-luminosity cutoff.

APPENDIX C: EFFECTS OF PRIORS ON $\Delta \ln \mathcal{L}$ AND $\ln \text{BF}$

Since we work in a Bayesian framework, we expect our results for the sensitivity to depend on the priors to some degree. Especially where the PSs are faint enough that the data cannot effectively discriminate between PSs and diffuse emission, the flux attribution to PSs vs Poissonian components may also be significantly influenced by the choice of priors (see also discussions in [11,42]). In this appendix we discuss the effect of changing selected priors from log to

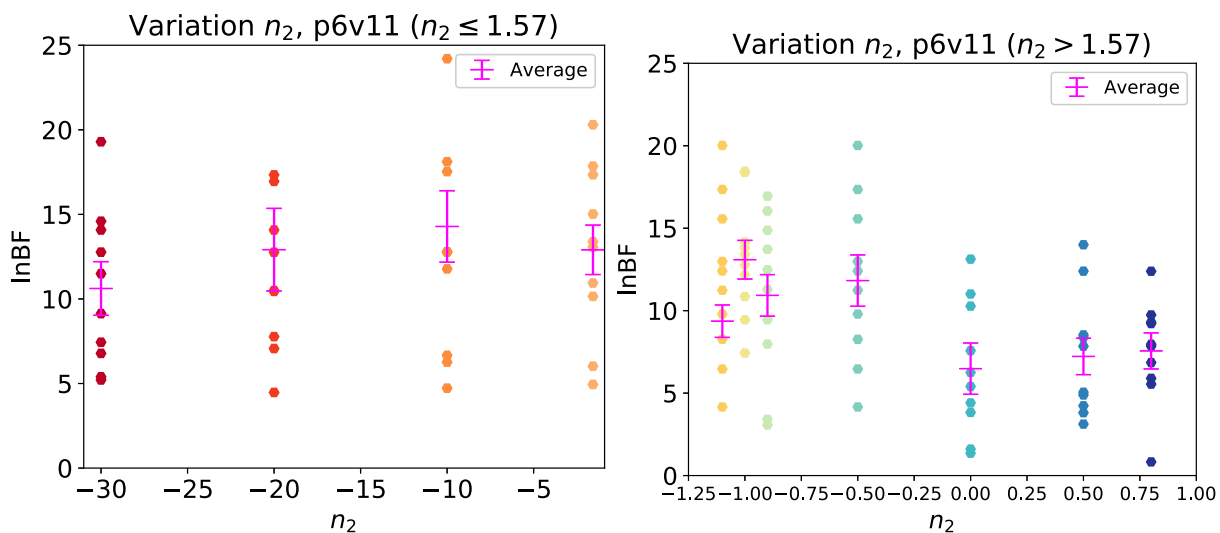


FIG. 16. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), varying n_2 of the SCF for the GCE PS component in the case with all *Fermi* templates.

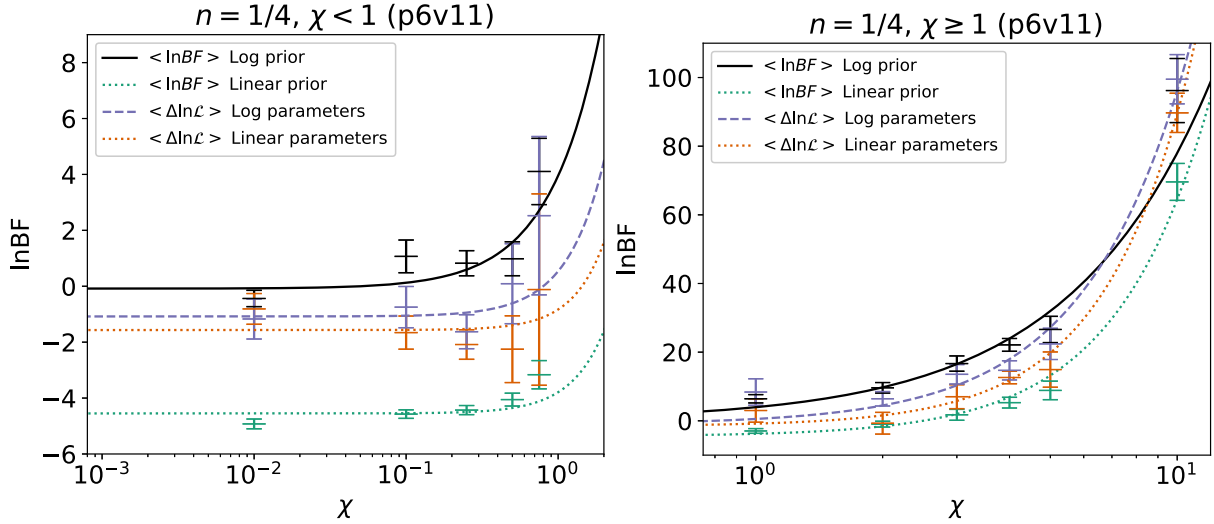


FIG. 17. $\langle \ln BF \rangle$ and $\langle \Delta \ln \mathcal{L} \rangle$, with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean across 10 realizations, varying the χ /exposure factor. We show results for log uniform and linearly uniform priors on the normalization of the GCE components, and likelihoods are evaluated at the posterior median parameters.

linear: specifically, we modify the priors that correspond to the normalization factors of the smooth GCE component A_{gce} and the GCE PS population $A_{\text{GCE}}^{\text{PS}}$. We keep the prior boundaries fixed.

We compared the $\ln BF$ in favor of PSs for a set of simulations of the full *Fermi* skymap, across exposure factors $\chi = 0.01$ to $\chi = 10$, for both the case of the best-fit S_b from the real data, and with S_b reduced to 1/4 of its baseline value (corresponding to fainter point sources). Figure 17 shows the effect of changing the priors for the faint (1/4 baseline brightness) sources, where the impact is most pronounced. In this figure we also plot the results of evaluating $\Delta \ln \mathcal{L}$ between the posterior median parameters obtained from the runs with/without GCE PSs; the likelihood itself does not depend on the priors, but since the posterior distribution is prior-dependent, we observe a small residual difference at the level of the likelihoods as well. (Note that the posterior median parameters do not generally correspond to the maximum likelihood point; evaluating the likelihood ratio between the maximum likelihood points should of course give a prior-independent result.)

We observe that in general using linearly uniform priors on A leads to a smaller sensitivity figure of merit compared to the use of log uniform priors (this also holds for point sources with the baseline brightness level); the linearly uniform priors also give rise to a modestly steeper power-law slope in $\langle \ln BF \rangle$ with increasing χ , as the difference between log and linear priors becomes less pronounced at high exposure. For sufficiently faint sources and/or low exposures, the use of linear priors can lead to an apparent preference for the model with no GCE PSs even when they are present in the simulation (this is as expected; in the absence of data favoring PSs, the Bayesian approach should prefer the simpler model).

APPENDIX D: ADDITIONAL ANALYSES FOR THE SIMPLIFIED ISOTROPIC SCENARIO

In this appendix we discuss several additional analyses of the simplified isotropic case discussed in Sec. V, to test the degree to which results from the full inner Galaxy analysis (Sec. VI) carry over to this simpler scenario.

1. Tradeoffs between PSF and exposure

We apply the same procedure as described in the main text (Sec. VI) to test the response of NPTFit to a simultaneous variation of angular resolution and exposure level, in the presence of a population of PSs. Note that for these simulations, the number of live points we used for the scans is set to $n_{\text{live}} = 500$.

Figure 18 shows the results of the simulations and analysis. We fitted the exposure dependence for the three separate quartiles using a power law with a constant additive shift; results are given in Table VII (the $n = 1$ column). We find similar patterns as those described in Sec. VIB 3, although the overall sensitivity is higher (presumably because the signal-to-background ratio is much higher) and the sensitivity increases modestly faster than linearly with exposure in this case, with a slope of ~ 1.4 – 1.5 . We observe that moving from Quartile 1 to Quartile 2, and Quartile 2 to Quartile 3, produces a fairly exposure-independent degradation in sensitivity of roughly a factor of two.

2. Detection limit for faint sources

Similarly to the inner Galaxy analysis in the main text, we also tested the faintest isotropic PSs NPTFit is able to

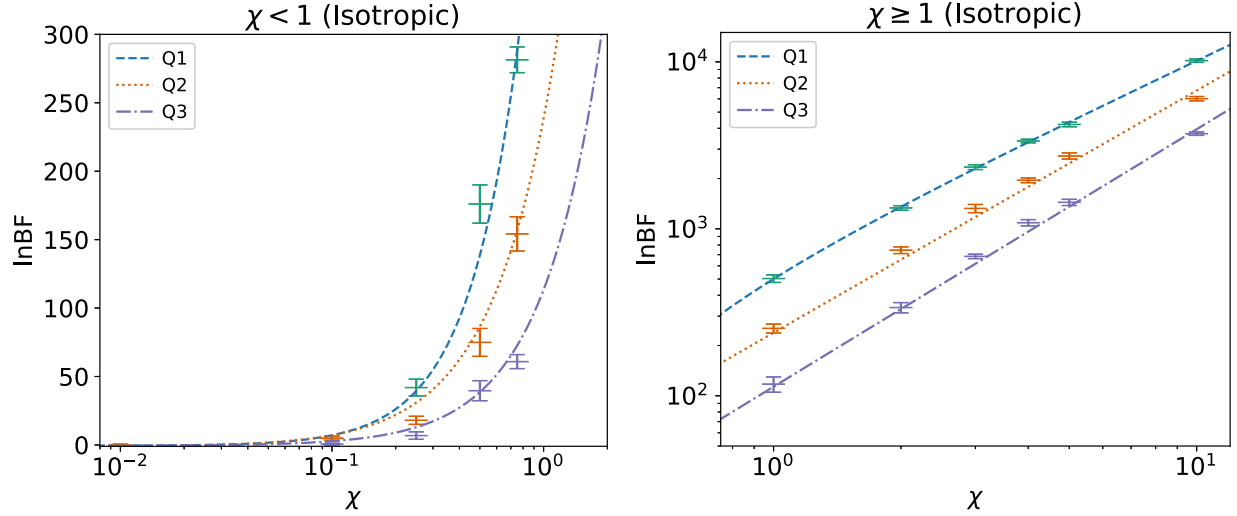


FIG. 18. $\langle \ln \text{BF} \rangle$ and the $\sigma/\sqrt{10}$ standard error of the mean across 10 realizations sampled at each quartile across varying levels of χ . Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. The parameters that describe the best-fit lines (dashed, dotted, and dot-dashed) are located in Table VII.

detect. The following graphs show the results of this analysis. In the panels of Fig. 19, n is used in the same way as in Eq. (37). Aside from testing fainter PSs, we also tested brighter PSs. The flux remained constant by reducing the number of sources present. Similar to the approach described in the main text, we model the dependence of $\langle \ln \text{BF} \rangle$ on the exposure as a power law with a constant additive shift, separately for n varying from $1/16$ to 2 .

Note that at lower values of χ , especially for faint sources, the data is not well-described by the best-fit curve. As in the main text, this is most likely just a signal that the power-law fit is inappropriate when the sensitivity to PSs is low.

We also plotted $\langle \ln \text{BF} \rangle(\chi = 1)$ across 10 realizations for varying levels of S_b for the isotropic scenario as shown in Fig. 20. The baseline case corresponds to $S_b = 5.61$. Similar to the full *Fermi* case the sensitivity increases as the population of PSs brightens. In the isotropic case, the increase in sensitivity is more stable and consistent, likely due to a lack of cross-talk with other templates that can occur (and impact the sensitivity especially for faint signals) in the baseline case. For comparison, we also plotted the analytic solution of Eq. (19) and we find that the power law scaling behavior of the simulated data is well captured by the estimate.

Comparing the full baseline case with the isotropic case, we find qualitatively similar patterns within the coefficients.

TABLE VII. Best-fit parameters and the 1σ errors obtained using least-squares regression method for the power law fit to the $\ln \text{BF}$ values of the isotropic scenario where we varied the source brightness while keeping the overall flux and the flux ratio between the PS and smooth component constant.

	$n = 1/16$	$n = 1/8$	$n = 1/4$	$n = 1/2$	$n = 1$	$n = 2$
<i>Quartile 1</i>						
α (coefficient)	3.77 ± 0.69	18.21 ± 2.47	71.90 ± 7.07	189.36 ± 14.38	438.79 ± 30.71	941.70 ± 47.79
β (power)	2.04 ± 0.11	1.78 ± 0.07	1.60 ± 0.06	1.48 ± 0.04	1.41 ± 0.04	1.34 ± 0.40
γ (shift)	-0.37 ± 0.05	-0.39 ± 0.07	-0.56 ± 0.28	-1.50 ± 1.22	-1.23 ± 0.44	-2.49 ± 1.21
<i>Quartile 2</i>						
α (coefficient)	1.01 ± 0.34	6.16 ± 0.96	25.63 ± 2.54	97.05 ± 7.25	240.09 ± 19.31	554.90 ± 43.34
β (power)	2.16 ± 0.16	1.93 ± 0.09	1.73 ± 0.05	1.53 ± 0.06	1.45 ± 0.05	1.36 ± 0.05
γ (shift)	-0.37 ± 0.11	-0.35 ± 0.12	-0.50 ± 0.06	-0.58 ± 0.23	-1.40 ± 1.75	-1.67 ± 1.87
<i>Quartile 3</i>						
α (coefficient)	0.23 ± 0.08	1.70 ± 0.64	9.83 ± 0.87	32.16 ± 5.70	113.91 ± 9.10	287.81 ± 19.51
β (power)	2.44 ± 0.17	2.17 ± 0.19	1.86 ± 0.05	1.73 ± 0.09	1.45 ± 0.04	1.48 ± 0.04
γ (shift)	-0.28 ± 0.04	-0.49 ± 0.10	-0.35 ± 0.05	-0.70 ± 0.26	-0.58 ± 0.29	-1.02 ± 0.66

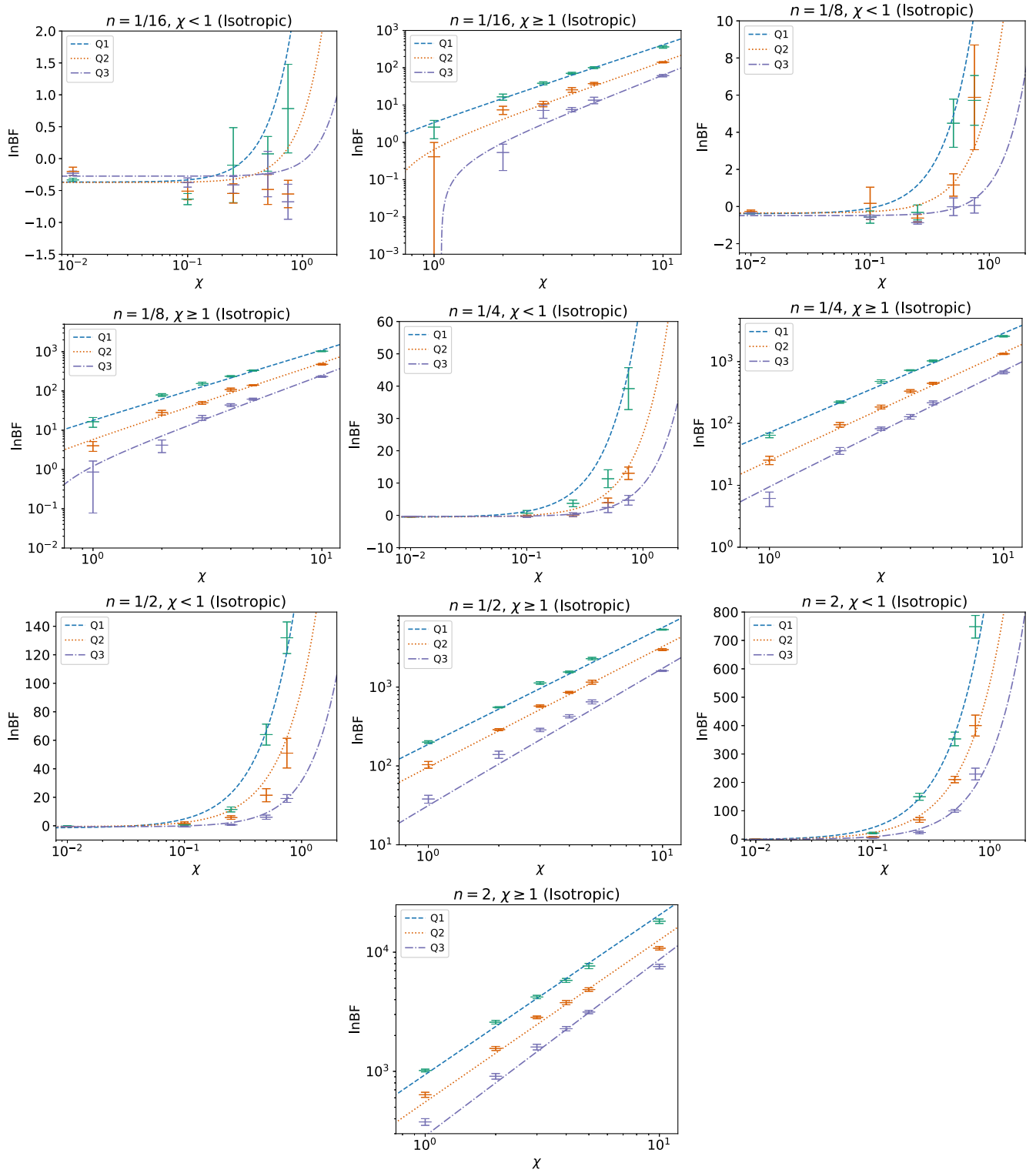


FIG. 19. $\langle \ln \text{BF} \rangle$ and the $\sigma/\sqrt{10}$ standard error of the mean across 10 realizations sampled across different χ values for each of the top three quartiles graded by angular resolution, in the simplified isotropic analysis. Each plot is labeled with the χ (to denote which exposure rescaling factor range is covered) and n values, where n corresponds to the source brightness as described in Eq. (37).

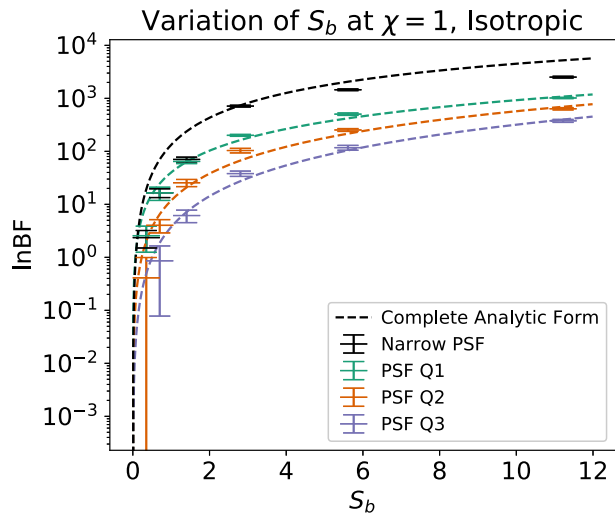


FIG. 20. $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean across 10 realizations at each of the top three quartiles and narrow PSF at varying brightness level for the isotropic case, with the baseline exposure. The best-fit S_b in the real data is 5.61; we test the effect of reducing and increasing S_b . All cases were scanned at $n_{\text{live}} = 500$. The dashed lines represent the results of a power-law fit to $\langle \ln \text{BF} \rangle$ as a function of S_b (holding exposure fixed at $\chi = 1$). The explicit form of the equation is Eq. (D1). For comparison, we also plot the analytic description of Eq. (19) as a function of s .

At higher source brightness (increased n), the parameter α (corresponding approximately to the sensitivity at $\chi = 1$) increases. More quantitatively, we fitted:

$$\langle \ln \text{BF} \rangle|_{\chi=1} = a S_b^p, \quad (\text{D1})$$

where a , p are numerical constants. We find that $\langle \ln \text{BF} \rangle$ scales as $O(S_b^{1.4})$, $O(S_b^{1.7})$, $O(S_b^{1.9})$, for quartiles 1, 2, and 3 respectively. We plotted the results of this fit in Fig. 20 and

find that the scaling behavior is approximately the same as the analytic form Eq. (19), if we allowed s , the number of photons per source to vary. The β parameter, describing the rate at which sensitivity increases with respect to exposure, is enhanced at lower values of n for the range of exposure values we test, rising to ~ 2 (compared to ~ 1.3 for the brightest sources). This is consistent with our analytic estimates, where we found that we expected the significance to scale with the square of the exposure in the faint-source limit.

Similar to the analysis performed in Sec. VID, if we assume that Eq. (39) is a reasonable description of the underlying relation of sensitivity to exposure, we can use this fit to determine χ values at each brightness level (for PSFQ1) that correspond to $\ln \text{BF} \gtrsim 1$, indicating the first hint of evidence for a detection. Using the best-fit parameters in Table VII, we find that for $n = 1/16$, $n = 1/8$, $n = 1/4$, $n = 1/2$, $n = 1$ and $n = 2$ these “exposure thresholds” correspond respectively to $\chi \geq 0.61, 0.24, 0.091, 0.054, 0.024, 0.015$, scaling roughly inversely with n .

3. Pixel size variation

As in the main text, we tested the effect of varying the pixel size. In this simplified (isotropic) scenario, we also considered the interplay between choosing smaller or larger pixel sizes and having fainter or brighter PSs—for example, we might ask whether different pixel sizes are more or less optimal depending on the source brightness. Figure 21 shows the results of our analysis; similarly to our previous results, we found that the ideal pixel size for optimizing sensitivity to PSs lies in the intermediate range around $n_{\text{side}} = 128$. The relative enhancement at the intermediate range, however, decreases with fainter sources.

An interesting feature present across the plots in Fig. 21 is the sudden spike in sensitivity at large pixel sizes

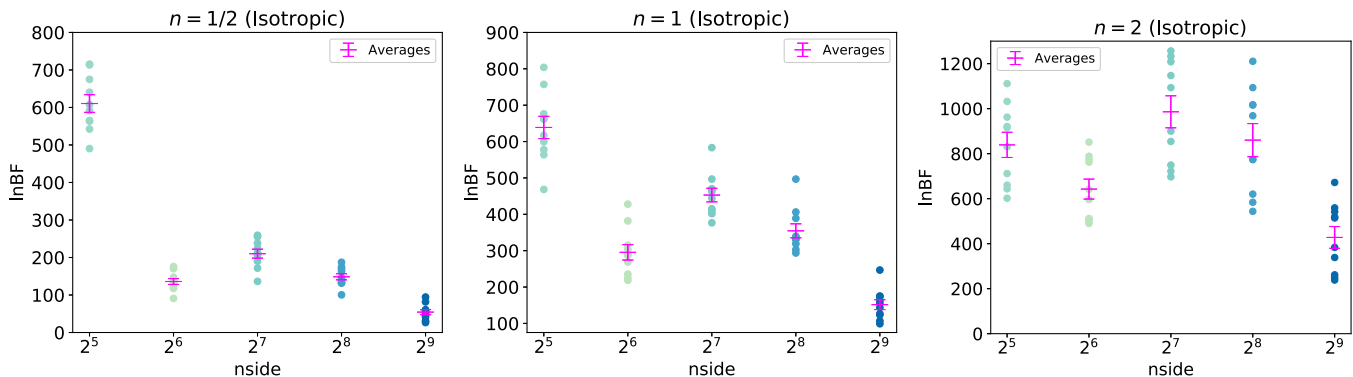


FIG. 21. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), sampling different pixel sizes (n_{side}) at various brightness levels of an isotropic PS population. Note that the anomalously high significance values attached to the largest pixel size ($n_{\text{side}} = 32$) are associated with a failure to correctly reconstruct the input parameters.

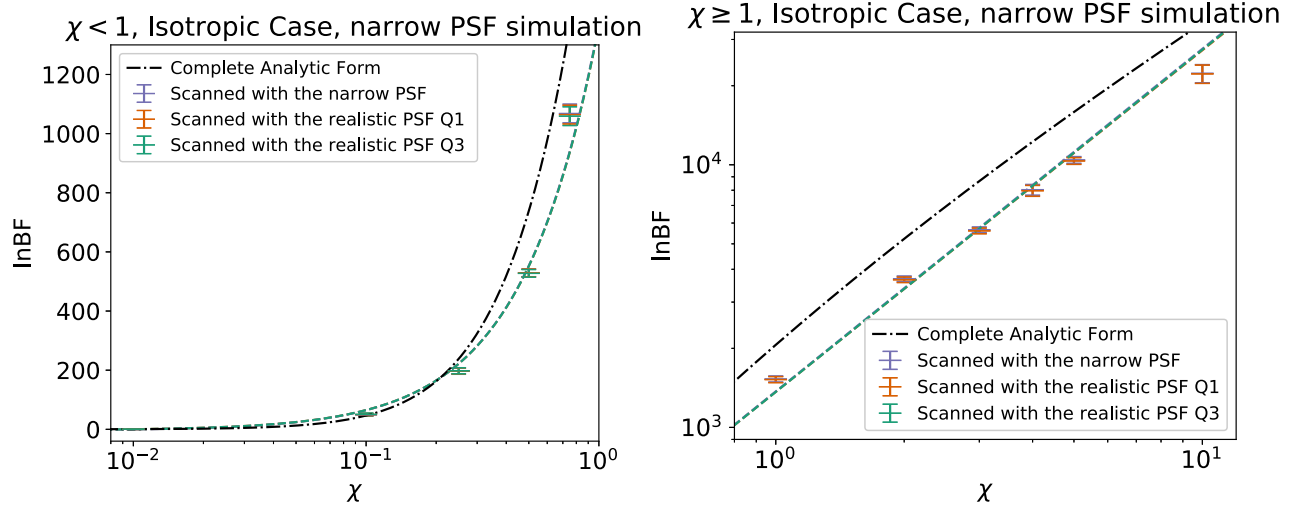


FIG. 22. $\langle \ln \text{BF} \rangle$ with errors obtained from the $\sigma/\sqrt{10}$ standard error of the mean across 10 realizations across different levels of exposure χ . The data was simulated using a narrow PSF and scanned in three separate ways: (1) using the narrow PSF that matches the simulation, (2) using the realistic *Fermi* PSF at Q1, and (3) using the realistic *Fermi* PSF at Q3.

(low nside). Upon closer inspection of the returned flux fraction, `NPTFit` incorrectly recovers the flux fraction for both smooth and PS isotropic emission (similar to the misattribution at low nside in the main text). Specifically, the fit attributes most of the emission to the isotropic PS population instead of distributing the flux equally to both sources. Examining the flux fraction plots for smaller pixel sizes, on the other hand, the correct flux fraction ratio is recovered.

4. Tests with an overly conservative PSF

The convention in the literature (which we have generally followed in this work, e.g. in Fig. 8) has been to use the PSF associated with the worst-angular-resolution photons in the dataset, for the purpose of `NPTFit` scans. In this section we test the impact of using an overly conservative (i.e. too large) PSF on data with a better angular resolution, in the simplified isotropic scenario.

We simulated narrow-PSF isotropic data as previously discussed. We then performed three sets of `NPTFit` scans of the same data set: (1) using the narrow PSF, (2) using the realistic *Fermi* PSF appropriate to the top PSF quartile, Q1, and (3) using the realistic *Fermi* PSF appropriate to the third PSF quartile, Q3.

Figure 22 displays the results of this test. We find that the three cases are generally consistent with each other. Thus, utilizing an overly conservative PSF within the `NPTFit` scan does not significantly impact the resultant sensitivity to PSs, at least in the cases we have tested. However, it is possible that assuming an isotropic background and signal reduces the effect of PSF mismodeling; the sensitivity to PSF mismodeling may also depend on the pixel size.

APPENDIX E: TESTS WITH AN ALTERNATIVE DIFFUSE MODEL

In this appendix we repeat selected analyses with a different model for the Galactic diffuse emission. `Model A` [20] splits the diffuse emission among two different components: π^0 decay from proton-proton collisions plus bremsstrahlung (“Pibrem”) and inverse Compton scattering (“ICS”). Separating the diffuse emission into two different components grants an extra degree of freedom during the fitting process, and allows for a better fit. We note that we scanned all realizations for this subsection at `nlive` = 100 for computational efficiency.

1. Variation of exposure

We began by repeating the exposure-variation analysis of Sec. VI A. Figure 23 shows the sensitivity as a function of the exposure rescaling factor χ . As previously, we fit this curve with a power law with a constant additive offset, and plot the resulting best-fit parameters in Table VIII.

The behavior of $\langle \ln \text{BF} \rangle$ as a function of the rescaling factor χ is quite similar between `p6v11` and `Model A` analyses, but the power-law slope in the current analysis is 1.0, corresponding to a slightly steeper scaling than the slope of 0.76 found in the main text. Both are qualitatively similar to the expectation of linear scaling from the analytic estimates for the behavior of the delta log likelihood.

2. Variation of PSF quartile

We performed the same PSF variation test described in VIB 1 using the alternative (`Model A`) diffuse templates. The results are shown in Fig. 24; we find the sensitivity is

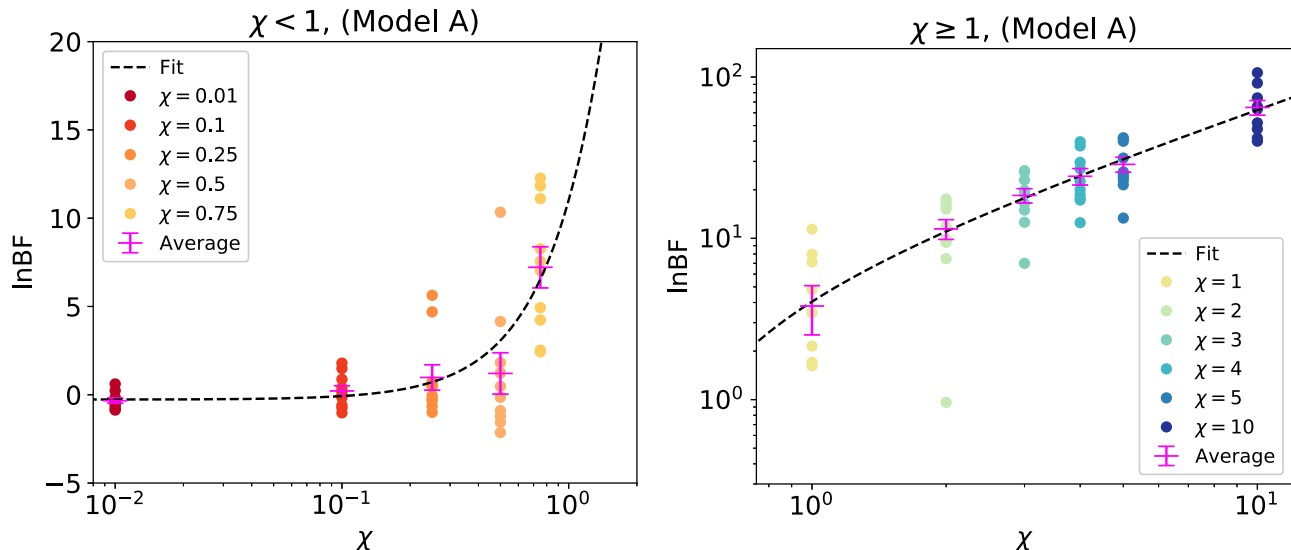


FIG. 23. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), sampling across various levels of exposure χ with Model A as the diffuse template. Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. The best-fit line is a standard power law with an additive shift, defined in Eq. (39).

somewhat degraded in the other PSF quartiles relative to the best quartile. For Model A, even in the top quartile the sensitivity is rather modest, and so an inverse scaling with the containment angle would be expected to lead to low sensitivity in the other quartiles, which is observed (but with substantial uncertainties).

3. Trade-offs in sensitivity between exposure and angular resolution

We repeated the exposure variation test for each of the PSF quartiles, to test the tradeoff between increased exposure and improved angular resolution. Figure 25 shows the results. Best-fit parameters for the power-law fits to each set of results are given in Table IX. The power-law slope with respect to exposure steepens slightly in the second quartile relative to the first. In Q3, the apparent best-fit slope is substantially steeper than in the other two quartiles, but the quality of the power-law fit is quite poor; the steep slope is driven by a high sensitivity at $\chi = 10$

TABLE VIII. Best-fit parameters obtained using least-squares regression method for the power law fit (Model A). Parameter uncertainties obtained from the standard error of the mean across 10 realizations are also displayed.

Parameter	Model A
α (coefficient)	6.0 ± 0.7
β (power)	1.0 ± 0.1
γ (shift)	-0.4 ± 0.2

combined with a relatively low and stable sensitivity at $\chi = 2-5$.

We examine the effect of “stacking” skymaps generated at different PSF quartiles to increase the effective exposure. We perform the same comparison described in Sec. VIB 3 to obtain the combined estimate of varying exposure and angular resolution. Similar to the results obtained from p6v11, the $\langle \ln \text{BF} \rangle$ across the three levels of χ_{eff} are consistent with each other to within 1σ . Hence, we find that the increase in sensitivity from a larger dataset is offset by the worsening of angular resolution. Figure 26 displays the results along with the combined estimate.

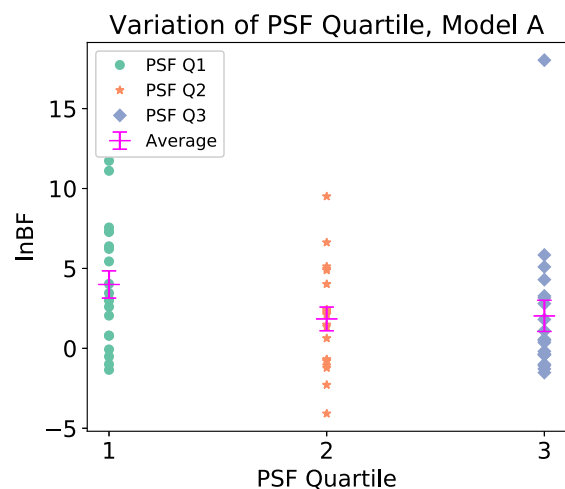


FIG. 24. $\ln \text{BF}$ across 20 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{20}$ standard error of the mean (magenta), at each PSF quartile, using Model A.

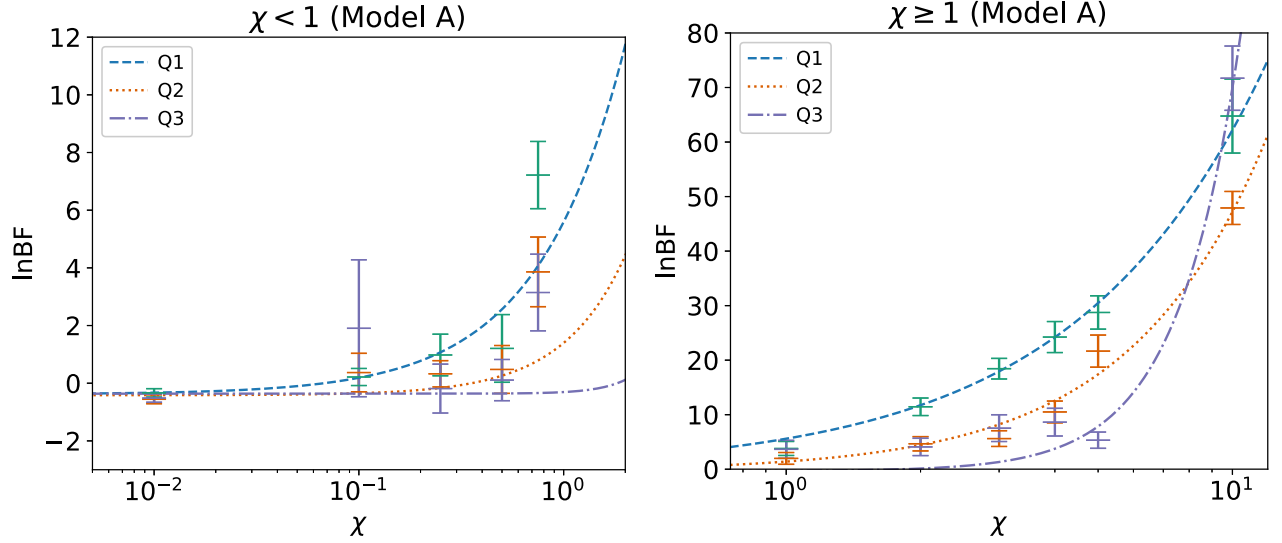


FIG. 25. $\langle \ln \text{BF} \rangle$ and the 1σ standard error of the mean across 10 realizations for top three quartiles graded in angular resolution sampling different values of χ using an alternative diffuse template (Model A). Left: realizations with $\chi < 1$. Right: realizations with $\chi \geq 1$. The best-fit lines show the Eq. (39) fit to the data.

TABLE IX. Model A power law parameters across the top three quartiles.

Parameter	Q1	Q2	Q3
α (coefficient)	6.0 ± 0.7	1.8 ± 0.5	0.1 ± 0.1
β (power)	1.0 ± 0.1	1.4 ± 0.1	3.1 ± 0.5
γ (shift)	-0.4 ± 0.2	-0.4 ± 0.2	-0.4 ± 0.3

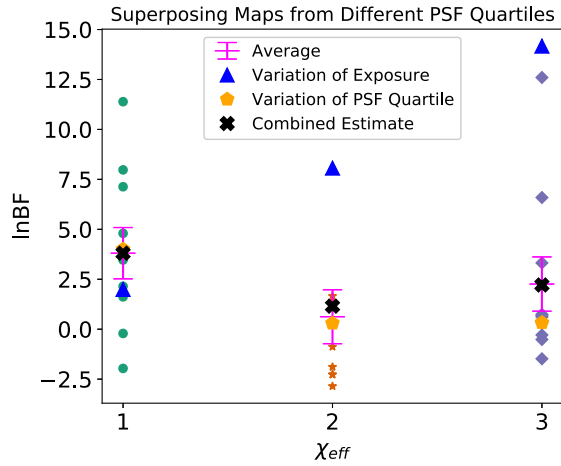


FIG. 26. $\ln \text{BF}$ and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta) across 10 realizations that stacked skymaps generated with different angular resolutions (PSF quartiles) for skymaps simulated using the Model A diffuse template. The scans assumed the worst angular resolution. The blue triangles indicate the increased sensitivity as predicted by varying the exposure, while the orange pentagons indicate the worsening of sensitivity due to angular resolution degradation. The black filled “X” display the combined estimate from varying the two parameters.

4. Pixel size variation

Finally, we tested the effects of varying the pixel size when using Model A for the Galactic diffuse emission. We applied the same procedure as described in Sec. VI C. Figure 27 shows the sensitivity level as a function of pixel size.

A first look at this figure seems to indicate that the optimal pixel size is the smallest value tested, $n_{\text{side}} = 32$. However, this high apparent sensitivity turns out to be

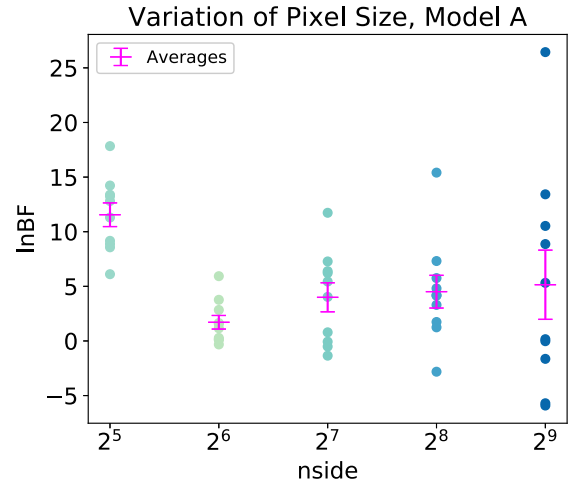


FIG. 27. $\ln \text{BF}$ across 10 realizations (circle markers), and $\langle \ln \text{BF} \rangle$ with error bars obtained from the $\sigma/\sqrt{10}$ standard error of the mean (magenta), using an alternative diffuse model (Model A) at five different pixel sizes. The pixel size decreases as n_{side} increases.

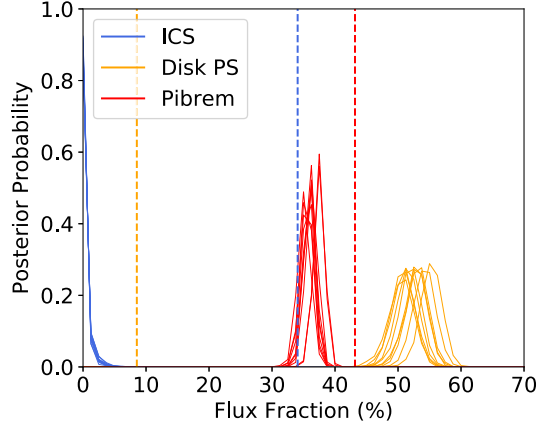


FIG. 28. Flux fraction plot demonstrating the misreconstruction due to a low n_{side} value ($n_{\text{side}} = 32$). The vertical dashed lines denote the injected values obtained from the parameter posterior medians from the real *Fermi* data. The flux fraction posterior probability distributions across 10 realizations (fainter lines) are displayed for the components of Model A diffuse emission and emission from Disk PSs.

accompanied by a failure to correctly reconstruct the various template fluxes, similar to the effect discussed for isotropic emission in Appendix D. In particular, the diffuse component of Model A which is comprised of the inverse Compton scattering (ICS) and the pion plus bremsstrahlung (Pibrem) component appears to be degenerate with the disk PS component, and the flux associated with these components is frequently misreconstructed, as shown in Fig. 28. Taking this into account, it appears that the ideal pixel size remains at intermediate values, $n_{\text{side}} = 128\text{--}256$.

APPENDIX F: DERIVATION OF THE PDF FOR THE TEST STATISTIC IN THE GAUSSIAN AND HIGH-DETECTABILITY LIMIT

To derive Eq. (13), we first evaluate the log likelihood between two Gaussian models, parametrized by (X, σ^2) and (Y, τ^2) , as a function of the observed number of counts N :

$$\begin{aligned} \Delta \ln \mathcal{L} &= P(N|\{X, \sigma^2\}) - P(N|\{Y, \tau^2\}) \\ &= -\frac{(N-X)^2}{2\sigma^2} - \frac{1}{2} \ln(2\pi\sigma^2) + \frac{(N-Y)^2}{2\tau^2} + \frac{1}{2} \ln(2\pi\tau^2) \\ &= \frac{-(N-X)^2}{2\sigma^2} - \frac{-(N-Y)^2}{2\tau^2} - \frac{1}{2} \ln(\sigma^2/\tau^2). \end{aligned} \quad (\text{F1})$$

In the high-detectability limit, we can make the approximation that the log term is small and can be ignored. Furthermore, we are interested in the comparison of two distributions that have the same expectation value but different variances (parametrizing the degree to which the distribution is non-Poissonian), so we can set $X = Y$. Writing $\delta = (\sigma^2/\tau^2) - 1$ as in the main text, we then have $\tau^2 = \sigma^2/(1 + \delta)$, and so:

$$\begin{aligned} \Delta \ln \mathcal{L} &\approx -\frac{(N-X)^2}{2} \left(\frac{1}{\sigma^2} - \frac{1}{\tau^2} \right) \\ &= \delta \frac{(N-X)^2}{2\sigma^2}. \end{aligned} \quad (\text{F2})$$

Note that under these approximations, $\Delta \ln \mathcal{L}$ is always non-negative for $\delta > 0$. Now we can evaluate the probability that this expression takes some value $x > 0$ under the true distribution of N , which we take to be given by $P(N|\{X, \sigma^2\})$.

$$\begin{aligned} P(\Delta \ln \mathcal{L} = x) &= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(N-X)^2/2\sigma^2} \delta\left(x - \delta \frac{(N-X)^2}{2\sigma^2}\right) dN \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x/\delta} \int \delta\left(x - \delta \frac{(N-X)^2}{2\sigma^2}\right) dN \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x/\delta} \sqrt{\frac{\sigma^2}{2x\delta}} \times \int dN \left[\delta\left(N - X + \sqrt{2\sigma^2 x/\delta}\right) + \delta\left(N - X - \sqrt{2\sigma^2 x/\delta}\right) \right] \\ &= \frac{1}{\sqrt{\pi x \delta}} e^{-x/\delta}, \end{aligned} \quad (\text{F3})$$

where in the second-last line we have evaluated:

$$\begin{aligned} \left| \frac{d}{dN} (x - \delta(N-X)^2/(2\sigma^2)) \right| &= |(N-X)|\delta/\sigma^2 \\ &= \sqrt{\frac{2x\sigma^2}{\delta}} \delta/\sigma^2 \\ &= \sqrt{2x\delta/\sigma^2} \end{aligned} \quad (\text{F4})$$

on the support of the delta function. Equation (F3) matches Eq. (13) and is the desired result.

- [1] Dmitry Malyshev and David W. Hogg, Statistics of gamma-ray point sources below the Fermi detection limit, *Astrophys. J.* **738**, 181 (2011).
- [2] Samuel K. Lee, Mariangela Lisanti, and Benjamin R. Safdi, Distinguishing dark matter from unresolved point sources in the inner galaxy with photon statistics, *J. Cosmol. Astropart. Phys.* **05** (2015) 056.
- [3] Samuel K. Lee, Mariangela Lisanti, Benjamin R. Safdi, Tracy R. Slatyer, and Wei Xue, Evidence for Unresolved γ -Ray Point Sources in the Inner Galaxy, *Phys. Rev. Lett.* **116**, 051103 (2016).
- [4] Tim Linden, Nicholas L. Rodd, Benjamin R. Safdi, and Tracy R. Slatyer, High-energy tail of the Galactic Center gamma-ray excess, *Phys. Rev. D* **94**, 103013 (2016).
- [5] Mariangela Lisanti, Siddharth Mishra-Sharma, Lina Necib, and Benjamin R. Safdi, Deciphering contributions to the extragalactic gamma-ray background from 2 GeV to 2 TeV, *Astrophys. J.* **832**, 117 (2016).
- [6] Hannes-S. Zechlin, Alessandro Cuoco, Fiorenza Donato, Nicolao Fornengo, and Andrea Vittino, Unveiling the gamma-ray source count distribution below the Fermi detection limit with photon statistics, *Astrophys. J. Suppl. Ser.* **225**, 18 (2016).
- [7] H. S. Zechlin, A. Cuoco, F. Donato, N. Fornengo, and M. Regis, Statistical measurement of the gamma-ray source-count distribution as a function of energy, *AIP Conf. Proc.* **1792**, 070020 (2017).
- [8] Hannes-S. Zechlin, Silvia Manconi, and Fiorenza Donato, Constraining galactic dark matter with gamma-ray pixel counts statistics, *Phys. Rev. D* **98**, 083022 (2018).
- [9] Tansu Daylan, Stephen K.N. Portillo, and Douglas P. Finkbeiner, Inference of unresolved point sources at high galactic latitudes using probabilistic catalogs, *Astrophys. J.* **839**, 4 (2017).
- [10] Stephen K. N. Portillo, Benjamin C. G. Lee, Tansu Daylan, and Douglas P. Finkbeiner, Improved point-source detection in crowded fields using probabilistic cataloging, *Astron. J.* **154**, 132 (2017).
- [11] Gabriel H. Collin, Nicholas L. Rodd, Tyler Erjavec, and Kerstin Perez, A compound Poisson generator approach to point-source inference in astrophysics, *Astrophys. J. Suppl. Ser.* **260**, 29 (2022).
- [12] F. Calore, F. Donato, and S. Manconi, Dissecting the Inner Galaxy with γ -Ray Pixel Count Statistics, *Phys. Rev. Lett.* **127**, 161102 (2021).
- [13] M. G. Aartsen *et al.* (IceCube Collaboration), A search for neutrino point-source populations in 7 yr of IceCube data with neutrino-count statistics, *Astrophys. J.* **893**, 102 (2020).
- [14] Siddharth Mishra-Sharma, Nicholas L. Rodd, and Benjamin R. Safdi, *NPTFit*: A code package for non-Poissonian template fitting, *Astron. J.* **153**, 253 (2017).
- [15] Lisa Goodenough and Dan Hooper, Possible evidence for dark matter annihilation in the inner milky way from the Fermi gamma ray space telescope, [arXiv:0910.2998](https://arxiv.org/abs/0910.2998).
- [16] Dan Hooper and Lisa Goodenough, Dark matter annihilation in the galactic center as seen by the Fermi gamma ray space telescope, *Phys. Lett. B* **697**, 412 (2011).
- [17] Dan Hooper and Tim Linden, On the origin of the gamma rays from the galactic center, *Phys. Rev. D* **84**, 123005 (2011).
- [18] Dan Hooper and Tracy R. Slatyer, Two emission mechanisms in the fermi bubbles: A possible signal of annihilating dark matter, *Phys. Dark Universe* **2**, 118 (2013).
- [19] Tansu Daylan, Douglas P. Finkbeiner, Dan Hooper, Tim Linden, Stephen K.N. Portillo, Nicholas L. Rodd, and Tracy R. Slatyer, The characterization of the gamma-ray signal from the central milky way: A case for annihilating dark matter, *Phys. Dark Universe* **12**, 1 (2016).
- [20] Francesca Calore, Ilias Cholis, and Christoph Weniger, Background model systematics for the Fermi GeV excess, *J. Cosmol. Astropart. Phys.* **03** (2015) 038.
- [21] M. Ajello *et al.* (Fermi-LAT Collaboration), Fermi-LAT observations of high-energy γ -ray emission toward the galactic center, *Astrophys. J.* **819**, 44 (2016).
- [22] Christopher Karwin, Simona Murgia, Tim M. P. Tait, Troy A. Porter, and Philip Tanedo, Dark matter interpretation of the Fermi-LAT observation toward the galactic center, *Phys. Rev. D* **95**, 103005 (2017).
- [23] Oscar Macias, Chris Gordon, Roland M. Crocker, Brendan Coleman, Dylan Paterson, Shunsaku Horiuchi, and Martin Pohl, Galactic bulge preferred over dark matter for the Galactic centre gamma-ray excess, *Nat. Astron.* **2**, 387 (2018).
- [24] Richard Bartels, Emma Storm, Christoph Weniger, and Francesca Calore, The Fermi-LAT GeV excess as a tracer of stellar mass in the Galactic bulge, *Nat. Astron.* **2**, 819 (2018).
- [25] Oscar Macias, Shunsaku Horiuchi, Manoj Kaplinghat, Chris Gordon, Roland M. Crocker, and David M. Nataf, Strong evidence that the Galactic bulge is shining in gamma rays, *J. Cosmol. Astropart. Phys.* **09** (2019) 042.
- [26] Martin Pohl, Oscar Macias, Phaedra Coleman, and Chris Gordon, Assessing the impact of hydrogen absorption on the characteristics of the Galactic Center excess, *Astrophys. J.* **929**, 136 (2022).
- [27] Mattia Di Mauro, Characteristics of the galactic center excess measured with 11 years of *Fermi*-LAT data, *Phys. Rev. D* **103**, 063029 (2021).
- [28] Ilias Cholis, Yi-Ming Zhong, Samuel D. McDermott, and Joseph P. Surdutovich, The return of the templates: Revisiting the Galactic Center excess with multi-messenger observations, *Phys. Rev. D* **105**, 103023 (2022).
- [29] Kevork N. Abazajian and Manoj Kaplinghat, Detection of a gamma-ray source in the galactic center consistent with extended emission from dark matter annihilation and concentrated astrophysical emission, *Phys. Rev. D* **86**, 083511 (2012); **87**, 129902(E) (2013).
- [30] Kevork N. Abazajian, Nicolas Canac, Shunsaku Horiuchi, and Manoj Kaplinghat, Astrophysical and dark matter interpretations of extended gamma-ray emission from the Galactic Center, *Phys. Rev. D* **90**, 023526 (2014).
- [31] Dan Hooper, Ilias Cholis, Tim Linden, Jennifer Siegal-Gaskins, and Tracy Slatyer, Pulsars cannot account for the inner galaxy's GeV excess, *Phys. Rev. D* **88**, 083009 (2013).
- [32] N. Mirabal, Dark matter vs. Pulsars: Catching the impostor, *Mon. Not. R. Astron. Soc.* **436**, 2461 (2013).

- [33] F. Calore, M. Di Mauro, F. Donato, and F. Donato, Diffuse gamma-ray emission from galactic pulsars, *Astrophys. J.* **796**, 14 (2014).
- [34] Ilias Cholis, Dan Hooper, and Tim Linden, Challenges in explaining the Galactic Center gamma-ray excess with millisecond pulsars, *J. Cosmol. Astropart. Phys.* **06** (2015) 043.
- [35] Qiang Yuan and Kunihito Ioka, Testing the millisecond pulsar scenario of the Galactic Center gamma-ray excess with very high energy gamma-rays, *Astrophys. J.* **802**, 124 (2015).
- [36] Ryan M. O’Leary, Matthew D. Kistler, Matthew Kerr, and Jason Dexter, Young pulsars and the Galactic Center GeV gamma-ray excess, [arXiv:1504.02477](https://arxiv.org/abs/1504.02477).
- [37] Harrison Ploeg, Chris Gordon, Roland Crocker, and Oscar Macias, Consistency between the luminosity function of resolved millisecond pulsars and the Galactic Center excess, *J. Cosmol. Astropart. Phys.* **08** (2017) 015.
- [38] Dan Hooper and Tim Linden, Millisecond pulsars, TeV halos, and implications for the Galactic Center gamma-ray excess, *Phys. Rev. D* **98**, 043005 (2018).
- [39] R. T. Bartels, T. D. P. Edwards, and C. Weniger, Bayesian model comparison and analysis of the galactic disk population of gamma-ray millisecond pulsars, *Mon. Not. R. Astron. Soc.* **481**, 3966 (2018).
- [40] Richard Bartels, Francesca Calore, Emma Storm, and Christoph Weniger, Galactic binaries can explain the Fermi Galactic Center excess and 511 keV emission, *Mon. Not. R. Astron. Soc.* **480**, 3826 (2018).
- [41] Rebecca K. Leane and Tracy R. Slatyer, Revival of the Dark Matter Hypothesis for the Galactic Center Gamma-Ray Excess, *Phys. Rev. Lett.* **123**, 241101 (2019).
- [42] Laura J. Chang, Siddharth Mishra-Sharma, Mariangela Lisanti, Malte Buschmann, Nicholas L. Rodd, and Benjamin R. Safdi, Characterizing the nature of the unresolved point sources in the Galactic Center, *Phys. Rev. D* **101**, 023014 (2020).
- [43] Malte Buschmann, Nicholas L. Rodd, Benjamin R. Safdi, Laura J. Chang, Siddharth Mishra-Sharma, Mariangela Lisanti, and Oscar Macias, Foreground mismodeling and the point source explanation of the Fermi galactic center excess, *Phys. Rev. D* **102**, 023023 (2020).
- [44] Rebecca K. Leane and Tracy R. Slatyer, Spurious Point Source Signals in the Galactic Center Excess, *Phys. Rev. Lett.* **125**, 121105 (2020).
- [45] Rebecca K. Leane and Tracy R. Slatyer, The enigmatic galactic center excess: Spurious point sources and signal mismodeling, *Phys. Rev. D* **102**, 063019 (2020).
- [46] Florian List, Nicholas L. Rodd, Geraint F. Lewis, and Ishaan Bhat, The GCE in a New Light: Disentangling the γ -Ray Sky with Bayesian Graph Convolutional Neural Networks, *Phys. Rev. Lett.* **125**, 241102 (2020).
- [47] Harrison Ploeg, Chris Gordon, Roland Crocker, and Oscar Macias, Comparing the galactic bulge and galactic disk millisecond pulsars, *J. Cosmol. Astropart. Phys.* **12** (2020) 035.
- [48] Anuj Gautam, Roland M. Crocker, Lilia Ferrario, Ashley J. Ruiter, Harrison Ploeg, Chris Gordon, and Oscar Macias, Millisecond pulsars from accretion induced collapse naturally explain the Galactic Center gamma-ray excess, *Nat. Astron.* **6**, 703 (2022).
- [49] Jack T. Dinsmore and Tracy R. Slatyer, Luminosity functions consistent with a pulsar-dominated galactic center excess, *J. Cosmol. Astropart. Phys.* **06** (2022) 025.
- [50] Florian List, Nicholas L. Rodd, and Geraint F. Lewis, Extracting the galactic center excess’ source-count distribution with neural nets, *Phys. Rev. D* **104**, 123022 (2021).
- [51] Siddharth Mishra-Sharma and Kyle Cranmer, A neural simulation-based inference approach for characterizing the Galactic Center γ -ray excess, *Phys. Rev. D* **105**, 063017 (2022).
- [52] F. Feroz, M. P. Hobson, and M. Bridges, MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics, *Mon. Not. R. Astron. Soc.* **398**, 1601 (2009).
- [53] K. M. Gorski, Eric Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, HEALPix—A Framework for high resolution discretization, and fast analysis of data distributed on the sphere, *Astrophys. J.* **622**, 759 (2005).
- [54] M. G. Aartsen *et al.* (IceCube Collaboration), A search for neutrino point-source populations in 7 yr of IceCube data with neutrino-count statistics, *Astrophys. J.* **893**, 102 (2020).
- [55] <http://iaifi.org/>.
- [56] Meng Su, Tracy R. Slatyer, and Douglas P. Finkbeiner, Giant gamma-ray bubbles from Fermi-LAT: Active galactic nucleus activity or bipolar galactic wind?, *Astrophys. J.* **724**, 1044 (2010).
- [57] W. L. Kraushaar, G. W. Clark, G. P. Garmire, R. Borke, P. Higbie, V. Leong, and T. Thorsos, High-energy cosmic gamma-ray observations from the OSO-3 satellite, *Astrophys. J.* **177**, 341 (1972).
- [58] C. E. Fichtel, R. C. Hartman, D. A. Kniffen, D. J. Thompson, G. F. Bignami, H. Ögelman, M. E. Özel, and T. Tümer, High-energy gamma-ray results from the second small astronomy satellite, *Astrophys. J.* **198**, 163 (1975).
- [59] P. Sreekumar *et al.* (EGRET Collaboration), EGRET observations of the extragalactic gamma-ray emission, *Astrophys. J.* **494**, 523 (1998).
- [60] Mattia Di Mauro and Fiorenza Donato, Composition of the Fermi-LAT isotropic gamma-ray background intensity: Emission from extragalactic point sources and dark matter annihilations, *Phys. Rev. D* **91**, 123001 (2015).
- [61] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White, A universal density profile from hierarchical clustering, *Astrophys. J.* **490**, 493 (1997).
- [62] Francesca Calore, Mattia Di Mauro, Fiorenza Donato, Jason W. T. Hessels, and Christoph Weniger, Radio detection prospects for a bulge population of millisecond pulsars as suggested by Fermi LAT observations of the inner galaxy, *Astrophys. J.* **827**, 143 (2016).