The Astrophysical Journal Supplement Series, 260:29 (35pp), 2022 June https://doi.org/10.3847/1538-4365/ac5cb7

A Compound Poisson Generator Approach to Point-source Inference in Astrophysics

Gabriel H. Collin[1,2,6], Nicholas L. Rodd[3,4], Tyler Erjavec[1], and Kerstin Perez[1,5]
[1] Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; gabrielc@mit.edu
[2] Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA
[4] Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5] The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, USA
Received 2021 May 4; revised 2022 March 8; accepted 2022 March 9; published 2022 June 17

Abstract

The identification and description of point sources is one of the oldest problems in astronomy, yet even today the correct statistical treatment for point sources remains one of the field's hardest problems. For dim or crowded sources, likelihood-based inference methods are required to estimate the uncertainty on the characteristics of the source population. In this work, a new parametric likelihood is constructed for this problem using compound Poisson generator (CPG) functionals that incorporate instrumental effects from first principles. We demonstrate that the CPG approach exhibits a number of advantages over non-Poissonian template fitting (NPTF)—an existing method—in a series of test scenarios in the context of X-ray astronomy. These demonstrations show that the effect of the point-spread function, effective area, and choice of point-source spatial distribution cannot, generally, be factorized as they are in NPTF, while the new CPG construction is validated in these scenarios. Separately, an examination of the diffuse-flux emission limit is used to show that most simple choices of priors on the standard parameterization of the population model can result in unexpected biases: when a model comprising both a point-source population and diffuse component is applied to this limit, nearly all observed flux will be assigned to either the population or to the diffuse component. A new parameterization is presented for these priors that properly estimates the uncertainties in this limit. In this choice of priors, CPG correctly identifies that the fraction of flux assigned to the population model cannot be constrained by the data.

Unified Astronomy Thesaurus concepts: Astrostatistics strategies (1885); Prior distribution (1927); Parametric hypothesis tests (1904); Bayesian statistics (1900); X-ray telescopes (1825); Spatial point processes (1915); X-ray point sources (1270)

1. Introduction

A practical and unbiased method of point-source inference is central to the task of recovering the physical characteristics of point-source populations. When point sources are bright and well-separated in the sky, their identification and cataloging is straightforward and requires little statistics beyond quantification of uncertainties on location and brightness. The primary difficulty arises when point sources are dim, where one must distinguish a putative source from a background fluctuation and account for the possibility of multiple closely overlapping sources—called a crowded field. The calculation of the uncertainty on the number of sources is particularly demanding, as the number of sources is a discrete parameter. Parametric point-source inference methods, such as the non-Poissonian template fitting (NPTF) method (Lee et al. 2016), sidestep this issue by specifying a likelihood conditioned on the characteristics of a population. In particular, the likelihood can be expressed in terms of the mean number of sources—a continuous parameter for which uncertainty calculation is considerably easier.

NPTF is currently the most widely applied parametric point-source inference method in gamma-ray and neutrino astronomy, especially in the analysis of gamma-rays from the Galactic Center. However, there is growing concern that results obtained with NPTF must be interpreted cautiously to avoid potential biases; see the discussion in Leane & Slatyer (2019, 2020a, 2020b), Chang et al. (2020), and Buschmann et al. (2020). In the present work, we will identify several clear biases of the NPTF framework that become most obvious in the following three specific test scenarios motivated by X-ray astronomy: (i) a spatially nonuniform point-source population model; (ii) a nonisotropic instrumental point-spread function (PSF); and (iii) variation in the instrumental detector response, namely the effective area, on scales smaller than the choice of binning—that is typically, in turn, on the scale of the PSF. In all cases, we observe a bias in the recovered number of point sources as summarized in Figure 1.

To resolve these biases, we develop a new parametric point-source inference likelihood called the compound Poisson generator (CPG) likelihood. This new likelihood addresses the biases in the NPTF apparent in the X-ray domain, and further provides a rigorous treatment of several approximations made in the conventional NPTF approach. In addition, we demonstrate the importance of a carefully chosen prior parameterization in Bayesian analyses. Priors chosen directly on the standard parameterization of the differential source-count function can, in the limit where these models are formally indistinguishable, lead to posteriors where the observed flux is assigned to either the point-source population model or an associated diffuse emission model. We define a new coordinate system for the specification of priors that removes this unwanted effect as summarized in Figure 2.

This work does not address biases that may result from incorrect modeling of the spatial distributions of populations (see, in particular, Leane & Slatyer 2020a, 2020b), effective

[6] Corresponding author.

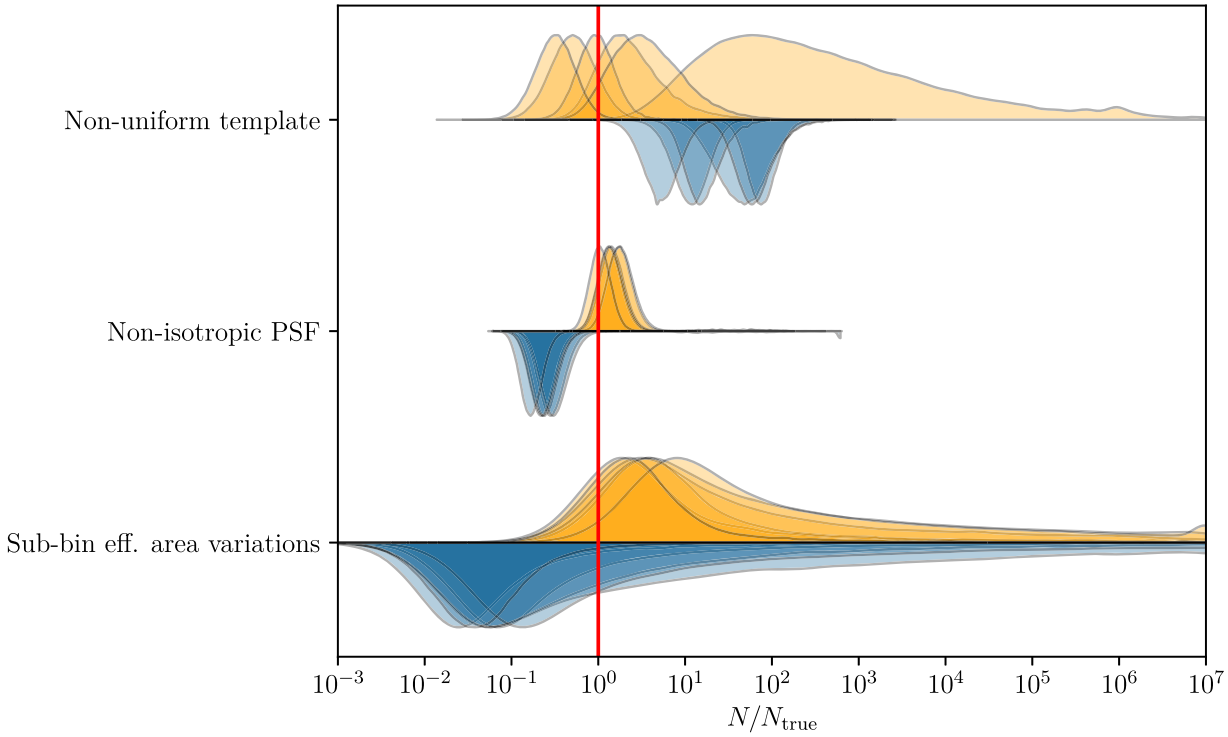

Figure 1. The recovered posterior in the number of point sources in the population, N, for both the novel CPG construction and the existing NPTF approach, for the three test scenarios considered in this work. Upright distributions are the posteriors for individual trials from the CPG construction. Inverted distributions are the posteriors recovered by NPTF. A clear bias away from the true number of sources, $N_{\rm true}$ (red line), is observed from NPTF. We will describe the results shown here in detail in Section 6. For example, the extension of many posteriors to $N \gg N_{\rm true}$ is explained in Section 6.3.

area, exposure area, or the PSF. Such biases are not limitations of the point-source inference method; instead, they are concerns for individual applications of point-source inference to specific analyses—although they are no less important for obtaining correct results. In this work, we assume that these effects are modeled correctly, and investigate the statistical methodology of point-source inference.

We organize the remainder of the discussion as follows. Section 2 reviews the current state of point-source inference, with a more detailed discussion on the concerns with NPTF, where the CPG enters, and how it resolves some of these issues. Section 3 outlines the point-source population model and the various instrumental effects introduced by standard X-ray instruments such as the NuSTAR satellite telescope, which we use as an example. Having outlined the background and challenges, in Section 4 the CPG likelihood is constructed from first principles. Section 5 introduces the priors on the population model parameters and demonstrates how the new coordinate system correctly handles the combination of point source and diffuse models. The NPTF method is introduced in more detail in Section 6, where we consider several scenarios motivated by instrumental effects, results from CPG and NPTF on these cases are directly compared, and the general efficacy of the new likelihood is demonstrated. An implementation of the CPG is made publicly available here.[7]

2. Review of Point-source Inference

To begin with, we will review the history and current state of point-source inference as it applies to parametric inference in the regime of dim sources and crowded fields. In this regime, automated source extraction is most commonly employed. This involves such methods as thresholding, peak searches, and wavelet decomposition, among others (Masias et al. 2012). Source-extraction algorithms can be broadly categorized as point-estimation methods—they produce a single best-fit or most-significant solution, and this solution can be represented as a point in the parameter space of point sources. For example, in thresholding and peak searches, a pixel is considered part of a source if the measured pixel intensity exceeds some defined threshold, which may be manually selected or defined by a statistical significance estimated from the image. The set of pixels that are classified as belonging to a source constitute a single point in the parameter space of all such sets of pixels, and may be further transformed into point-source locations by taking the pixels in each set, and for example, computing the centroid of those pixels. The resultant point in parameter space is the algorithm's best guess as to the true parameters. However, this guess is just that: an estimate of the true locations of the point sources in an image. As such, the calculation of uncertainties is required in order to quantify the quality of this single best guess.

In this problem, there are two classes of parameters: the individual point-source parameters, $\boldsymbol{\vartheta}$, such as location and brightness, which are continuous quantities; and the number of sources in the image, $\tilde{N}$, which is a discrete quantity. The measured data, such as the observed image, will be distributed according to the likelihood $\mathcal{L}(\boldsymbol{d}|\tilde{N}, \{\boldsymbol{\vartheta}_i\})$, where $\{\boldsymbol{\vartheta}_i\}$ is a set of size $\tilde{N}$, one for each of the $\tilde{N}$ sources. The pair $(\tilde{N}, \{\boldsymbol{\vartheta}_i\})$ form a point in the parameter space of the likelihood, and this point is often referred to as a catalog. When it comes to the calculation of uncertainties, discrete and continuous parameters must be handled in different ways.

[7] https://github.com/ghcollin/cpg_likelihood

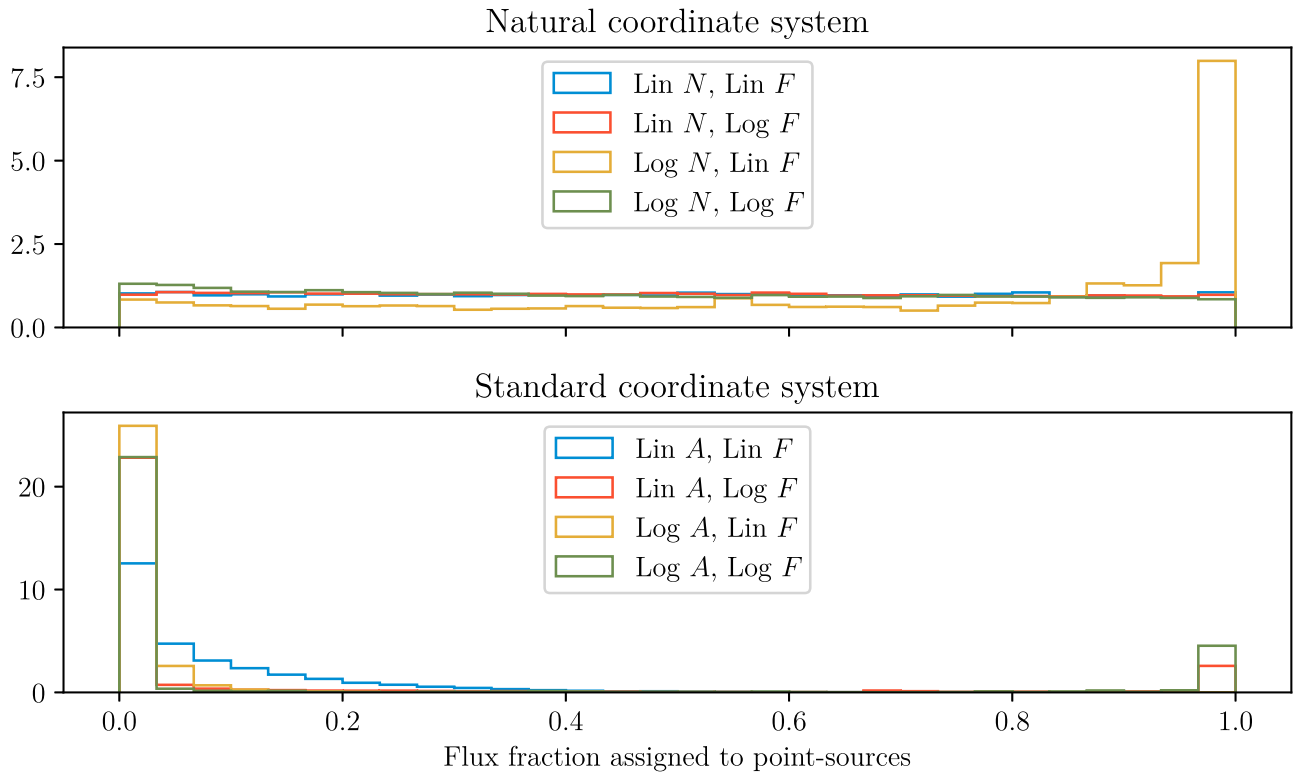

Figure 2. The median recovered posterior for the novel natural coordinate system (top) and standard coordinate system (bottom), for multiple choices of priors on the number of sources (N, natural only), flux (F), and differential source-count function normalization (A, standard only). The posterior is shown for the fraction of flux assigned to the population model, as compared to a diffuse background model. In this scenario, only diffuse flux is present, and so these posteriors should be uniform, as diffuse emission is formally indistinguishable from a population of dim point sources. This is observed only for the natural coordinate system, with one exception discussed further in Section 5.

Deriving uncertainties on the continuous individual source parameters is the easier of the two methods. Assuming the likelihood distribution is available, uncertainties can be derived using variational methods, or confidence regions can be constructed should Wilks' theorem apply. Deriving uncertainties on the discrete number of sources, $\tilde{N}$, is more fraught. The calculation of uncertainty on a discrete parameter is equivalent to the problem of model selection. Each choice for the number of sources should be considered as a separate model. Two choices for the number of sources may then be compared using a likelihood ratio between the likelihoods for each of the associated models. This likelihood ratio is a test statistic, and the probability distribution for the ratio must be known if it is to be converted to a p-value. The p-values may then be used to construct a confidence interval on the number of sources. Most often, Wilks' theorem is used to assume that the likelihood ratio is χ^2 distributed; however, this is only guaranteed in the asymptotic limit and any deviation of the true distribution from this assumption will result in incorrect confidence intervals.

An alternative approach to model selection is the calculation of marginal likelihoods:[8]

$$\mathcal{L}(\boldsymbol{d}|\tilde{N}) = \int \mathcal{L}(\boldsymbol{d}|\tilde{N}, \{\vartheta_i\})[\prod_{i=1}^{\tilde{N}} p(\vartheta_i) d\vartheta_i], \quad (1)$$

where $p(\vartheta_i)$ is a prior on the individual source parameters. From this expression, the posterior $p(\tilde{N}|\boldsymbol{d})$ can be found through Bayes' theorem, and so a credible region can be constructed on the number of sources—avoiding the need for a test statistic as in the likelihood ratio case. Variational methods provide only a lower bound on the marginal likelihood of the model, and thus provide no more than an estimate for the posterior distribution on the number of sources.

A more common approach is to sample the posterior $p(\{\vartheta_i\}|\boldsymbol{d}, \tilde{N})$ directly using nested sampling (Skilling 2004) or a Markov Chain Monte Carlo (MCMC). Nested sampling provides an estimate of the marginal likelihood directly, while various techniques exist for deriving the marginal likelihood from MCMC—for details, consult the review by Gelman & Meng (1998). Although this provides the most principled approach to the problem for small numbers of sources, in practice it exhibits a critical flaw. When the number of sources is potentially large, a marginal likelihood must be computed for each possible choice of $\tilde{N}$—a flaw also shared by the variational and likelihood ratio methods. For hundreds of sources, this can rapidly become computationally infeasible.

Ultimately, this flaw can be understood to be a manifestation of a one-dimensional grid search over $\tilde{N}$. In fact, MCMC is designed to avoid grid searches by concentrating sampling on regions of parameter space that are most likely; however, the MCMC algorithm is designed exclusively for continuous parameters, while $\tilde{N}$ is a discrete parameter that, when varied, also modifies the number of ϑ parameters. The Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm is an extension of MCMC to discrete parameters that control the

[8] Point-source inference can be approached in a frequentist or Bayesian framework, and we will use the language of both throughout. Nevertheless, when adopting a Bayesian approach—as has commonly been done for the NPTF—priors must be chosen carefully, as we discuss in Section 5.

number of parameters in the distribution (Green 1995). Probabilistic cataloging (Daylan et al. 2017; Portillo et al. 2017) is the application of RJMCMC to the problem of point-source inference. While the point-estimation methods produce a single best-guess catalog, probabilistic cataloging produces posterior samples in the space of all possible catalogs. From this catalog posterior, a posterior on the number of sources can be calculated. A catalog posterior provides—essentially by definition—the most general solution to the problem of point-source inference. This generality comes at a high cost: the transdimensional parameter transformations required by RJMCMC has to be carefully selected, as they must be tuned to the specific application to ensure that transdimensional moves are efficient. In addition, the dimensionality of the likelihood is at least two or three times larger than the number of sources—depending on the number of parameters per source. For thousands of sources, it can be unrealistic to assume that the RJMCMC will properly equilibrate for even a given number of sources, let alone across the number of sources; the application of RJMCMC will require careful and extensive diagnostics, none of which are foolproof.

The solution to this problem requires a revisitation to the stated inference goal. If the goal truly necessitates the location and intensity of each individual source, then probabilistic cataloging is necessary. But in this problem area, where the number of sources is large, they also likely form a crowded field. In this case, locations of individual sources will be poorly defined, as they form a near-uniform density within which sources are easily interchangeable. Instead of attempting to track each individual source, a model can be defined for the entire population of sources. The most common choice of model is a differential source-count function: $dN/d\boldsymbol{\vartheta}(\boldsymbol{\theta})$. This function then has its own parameters, $\boldsymbol{\theta}$, which are characteristics of the population as a whole, such as the spatial distribution of sources, or the average flux of the population. The differential source-count function gives the density of sources as a function of $\boldsymbol{\vartheta}$ given $\boldsymbol{\theta}$, and thus implicitly defines a probability distribution:

$$\frac{dN}{d\boldsymbol{\vartheta}}(\boldsymbol{\theta}) = Np(\boldsymbol{\vartheta}|\boldsymbol{\theta}), \tag{2}$$

where N is the mean number of sources in the population. This allows marginalization of the likelihood over both the set of $\boldsymbol{\theta}$ and further the possible number of sources,

$$\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta}) = \sum_{\tilde{N}} \int \mathcal{L}(\boldsymbol{d}|\tilde{N}, \{\boldsymbol{\vartheta}_i\})[\prod_{i=1}^{n} p(\boldsymbol{\vartheta}_i|\boldsymbol{\theta})d\boldsymbol{\vartheta}_i]p(\tilde{N}|N), \tag{3}$$

where $p(\tilde{N}|N)$ is a prior on the number of sources given the mean number of sources—usually assumed to be a Poisson distribution. Thus, sampling directly from the posterior of $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ reduces the high-dimensional nonparametric sampling problem in the space of all possible catalogs to a low-dimensional parametric sampling problem in $\boldsymbol{\theta}$. However, to achieve this we have marginalized out the $\boldsymbol{\vartheta}$ parameters, and so identification of the location or brightness of individual sources is no longer possible—the inference problem is now on the population itself.

To achieve this vast simplification, the likelihood $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ must be constructed directly so that it may be evaluated without computing the multidimensional integrals in Equation (3). An early attempt at this construction, called the $P(D)$ method (Scheuer 1957; Miyaji & Griffiths 2002; Barcons 1992), used instrument simulations to numerically estimate the likelihood as a histogram over the number of detected photons. This procedure estimates $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ by sampling $\boldsymbol{d}$ from the right-hand side of Equation (3). As this is effectively equivalent to an importance-weighted integration of those multidimensional integrals, this method ultimately carries the same computational burden of probabilistic cataloging.

In Malyshev & Hogg (2011), the authors noted that the number of detected photons is a sum of the number of detected photons produced by each source—itself distributed by a Poisson distribution. This allows the likelihood to be constructed using probability generating functions, and provides an analytic formula for $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\theta})$ assuming a perfect instrument. To incorporate instrumental effects, a heuristic argument is used to justify a semi-analytic expression for the likelihood in terms of a detector effect correction function, $\rho(f)$, that is generated through Monte Carlo simulation.

This method was further extended by Lee et al. (2016) to images of multiple pixels by parameterizing the source count function in terms of a spatial template. Known as NPTF, this is currently a leading method for parametric point-source inference in gamma-ray astronomy, and the method has also been applied to the search for astrophysical neutrino point sources in data from the IceCube telescope (Aartsen et al. 2020). The NPTF was primarily developed to analyze the excess of Galactic Center (GCE) gamma-rays observed by the Fermi telescope (Goodenough & Hooper 2009; Hooper & Goodenough 2011; Hooper & Linden 2011; Abazajian & Kaplinghat 2012; Hooper & Slatyer 2013; Gordon & Macias 2013; Abazajian et al. 2014, 2015; Daylan et al. 2016; Calore et al. 2015; Ajello et al. 2016; Linden et al. 2016; Macias et al. 2018; Clark et al. 2018), and concluded that the observed excess was better described by a population of point sources, in comparison to a dark matter annihilation origin (Lee et al. 2015, 2016; see also Bartels et al. 2016).[9] An approach similar to the NPTF that has also been widely used is the one-point fluctuation analysis or one-point PDF method; see, for example, Lee et al. (2009), Feyereisen et al. (2015, 2017), Zechlin et al. (2016a, 2016b, 2018), Manconi et al. (2020), and Calore et al. (2021). In the present work, we will focus on comparisons between the CPG and the NPTF, but many of our conclusions would be similar if we compared to these alternative approaches.

Recent investigations have suggested that the NPTF results must be interpreted carefully, raising the possibility that the nature of the GCE has not yet been conclusively resolved. First, Leane & Slatyer (2019) demonstrated that the NPTF could attribute an injected dark matter signal to the point-source model, although it appears this concern can be addressed. In particular, an improved treatment of the background models largely resolves the issue (Buschmann et al. 2020), and further, as was emphasized in Chang et al. (2020), a degree of confusion is unavoidable given the inherent degeneracy between purely Poisson emission and a population of point sources that produce at most one photon each in the data set. Nevertheless, when performing a Bayesian analysis, confusion between point-source and pure Poisson emission can be exacerbated by a poor choice of priors—a point we will return

[9] The NPTF has also been applied to the problem of determining the point-source contribution to the extragalactic gamma-ray background (Lisanti et al. 2016).

to in Section 5. An additional concern regarding the NPTF that has not yet been addressed is that it appears the presence of an unmodeled asymmetry in the data can significantly bias the method, in that the asymmetry leads the NPTF to return strong evidence for a point-source population, even when there is none (Leane & Slatyer 2020a, 2020b). Taken together, the above results emphasize that the output of the NPTF must be interpreted cautiously—indeed, whether the GCE contains the first hints of the particle nature of dark matter remains an open problem that even more recent wavelet (Zhong et al. 2020) and machine-learning (List et al. 2020) approaches have not conclusively resolved.

Nevertheless, as yet, the modeling of the detector effect correction function of NPTF, $\rho(f)$, has not been questioned, despite the heuristic justification for its original inclusion in the method. In this work, we will show that there are instances where the NPTF construction explicitly breaks down as a result of the mathematical construction of $\rho(f)$, and that this represents an obstacle to extending the method to X-ray data sets. In detail, we present a first-principles construction of the parametric likelihood in Equation (3), which incorporates the correction of detector effects in a statistically justified manner. The result, which we call the Compound Poisson Generator, includes a new detector effect correction function that is an analytic expression of the basic components of instrumental effects: the PSF, effective area, and photon detection probability. The new construction demonstrates that the spatial distribution of the point-source population cannot be disentangled from the detector effect correction function, nor can the PSF be disentangled from the effective area. This shows that the NPTF—which factorizes the spatial distribution, PSF, and effective area—cannot describe point-source statistics in general. To show this lack of generality, the NuSTAR X-ray telescope is used as a test case for both the CPG and NPTF. The substantial differences between the detector response in NuSTAR and Fermi will reveal the stated deficiencies in NPTF.

As for the effect of unmodeled asymmetries, these occur when the spatial distribution for an emitter is specified incorrectly. As NPTF and CPG have additional explanatory power above that of a simple Poisson model, they will produce a better fit to the data even if the emission is truly diffuse. This same effect would be observed when using probabilistic cataloging, as the location of the sources will migrate to explain the deviation from the specified spatial distribution. The CPG likelihood does not address this issue, as it is, in fact, not an issue with the point-source model at all. To see this, consider the following analogy with statistical mechanics. We would be surprised to observe a box where all gas molecules are located in one corner, as this situation is a very unlikely macrostate for a gas. On the other hand, from the perspective of the microstate description of the gas, the gas is in as likely a configuration as any other configuration including those where the system is more thoroughly mixed. Probabilistic cataloging, NPTF, and CPG all specify the likelihood of the data, which is a microstate description of the population. Thus, a configuration where all sources are on one side of the image is as likely as any other, and these methods make no distinction.

The problem here lies in the diffuse model that these population models are compared to. The introduction of an unmodeled asymmetry will have only a small effect on the population likelihood; in comparison, the likelihood for the diffuse model will suffer greatly due to this mismodeling. When the two models are compared, it appears that the population model is erroneously describing sources, but it is the diffuse model that is erroneously rejecting the diffuse hypothesis. To resolve the issue, the spatial distribution must be given additional degrees of freedom so that the diffuse model can account for deviations from the expected spatial distribution. Much like in statistical mechanics, care can be taken by examining the macrostate of the fitted population. The quantities analogous to macrostates in point-source inference are the population parameters. The differential source-count function can be adjusted to include, for example, an $N_{\rm top}$ and $N_{\rm bottom}$ for the top and bottom of the image. The ratio of these means can then be computed, and an unlikely value for this ratio will form a diagnostic signal for a mismodeling issue.

3. Point-source Population and Instrument Models

This section describes the point-source population model that will be used in this investigation, as well as the instrument model that is necessary to generate simulated observations. The instrument model is also needed in Section 4 to incorporate a detector correction into the likelihood.

3.1. Population Model

In the parametric approach, an assumption must be made to select a model that describes the population of sources. Here, we describe the population using a differential source-count function: dN/dF. This function describes the number of sources, N, as a differential over the individual point-source flux, F.

The point-source flux F is defined as a normalization factor in a power-law flux energy spectrum:

$$\frac{d\Phi}{dE} = F\left(\frac{E}{E_0}\right)^{-\gamma}, \tag{4}$$

where Φ is the number of photons per unit area and time, E is the photon energy, E_0 is the power-law scale, and γ is the power-law index (also known as the photon index). As a result, the dimensions of F are photons per unit area, time, and energy. Power-law spectra are common for the high-energy tails of X-ray sources (Fleishman & Bietenholz 2007; Mukai 2017; Hong et al. 2016), due to various processes such as synchrotron emission and inverse Compton scattering, and thus make a natural choice for this investigation. A power-law index of $\gamma = 1.5$ is used for all scenarios, as spectra with this index are common near the Galactic Center (Hong et al. 2016), and the power-law scale is set to $E_0 = 1$ keV.

The contribution of each source to the image is determined by converting the flux of the individual source to an expected number of photons based on the response of the detector. For a telescope, this response is commonly specified in terms of the exposure time, $t_{\rm exp}$, which is the time the instrument spent collecting the flux of interest, and further, the effective area, $\mathcal{A}_{\rm eff}$, which is the collecting area of an equivalent idealized telescope that detects all incident photons, such that the effective area is strictly less than the actual size of the real instrument. The mean number of detected photons (also called

the mean number of counts), S, is then

$$S = t_{\rm exp} \int_{E_i}^{E_f} dE\ \mathcal{A}_{\rm eff}(E) \frac{d\Phi}{dE}, \tag{5}$$

where the number of counts is defined within an energy band of interest, $E \in [E_i, E_f]$. As the only model parameter in this equation is F—through $d\Phi/dE$—this expression can be simplified to $S = F\kappa$, where κ is called the detector response. Generally speaking, the exposure time and effective area are position-dependent, so this is a function of position in the image, $\boldsymbol{x}$:

$$\kappa(\boldsymbol{x}) = t_{\rm exp}(\boldsymbol{x}) \int_{E_i}^{E_f} dE\ \mathcal{A}_{\rm eff}(E, \boldsymbol{x}) \left(\frac{E}{E_0}\right)^{-\gamma}. \tag{6}$$

Equations (4) and (5) are not necessary choices for the methods employed in this investigations. All that is required is for the number of counts, $S_{\boldsymbol{x}}$, received at location $\boldsymbol{x}$ to be able to be specified in terms of the individual source model parameter F and another quantity $\kappa(\boldsymbol{x})$. The value of $\kappa(\boldsymbol{x})$ need not even be based on an assumed energy spectrum, it must only be known and calculable for any location $\boldsymbol{x}$ and satisfy the relation $S_{\boldsymbol{x}} = \kappa(\boldsymbol{x})F$.

As reviewed above, for a single source, the key parameter dictating how many photons we expect to observe is the flux, F. When studying a population of sources, these fluxes can vary between the individual sources. The distribution of fluxes is described with a differential source-count function, dN/dF, which encodes the number of sources with flux between F and $F + dF$. For the investigations in this article, the differential source-count function is assumed to be a broken power law. A singly broken power law has the following form:

$$\frac{dN}{dF} = A \begin{cases} \left(\dfrac{F}{F_{b(2)}}\right)^{-n_2} & F \leqslant F_{b(2)} \\ \left(\dfrac{F}{F_{b(2)}}\right)^{-n_1} & F_{b(2)} < F, \end{cases} \tag{7}$$

where A is a normalization factor, $F_{b(2)}$ is the location of the break in flux, n_1 is the power index after the break, and n_2 is the index before the break. The generalization of this distribution to multiple breaks is discussed in Section 5. The ultimate goal is then to infer the dN/dF model parameters, denoted in aggregate by the vector $\boldsymbol{\theta}$, from the statistics of the number of counted photons received by the telescope or detector. To be explicit, for the parameterization given in Equation (7), $\boldsymbol{\theta} = \{A, F_{b(2)}, n_1, n_2\}$.

Generally, the differential source-count function can be posed in terms of a distribution, $p(F)$, which gives the probability density of an individual source having flux, F, through the relation

$$\frac{dN}{dF} = Np(F), \tag{8}$$

where N is the mean number of sources. This representation is more useful when describing the statistical process underlying point sources.

Another common representation is the cumulative source-count function (also called simply the source-count function):

$$N(F' > F) = \int_F^\infty \frac{dN}{dF'} dF', \tag{9}$$

which specifies the number of sources in the population with a flux greater than F. If the population is large in spatial extent, this may be written as the areal source-density function, $n(F' > F)$, often in numbers of sources per steradian or arcminutes squared.

Power-law-type source-count functions are common for astrophysical populations generally, and for X-ray emitter populations specifically (Mukai 2017; Hong et al. 2016). This is not unexpected, as the inverse square reduction in apparent brightness with distance will give most populations a power-law-like distribution in the observed brightness. More generally, power-law distributions are common in nature (see, for example, the Pareto distribution) and are the maximum entropy distributions for a logarithmic parameter with a specified mean.

3.2. X-Ray Instrumentation

Although the $P(D)$ method has been used extensively to study X-ray sources (Miyaji & Griffiths 2002), the current leading parametric inference method, NPTF, has yet to be applied to this regime. X-ray astronomy poses a number of novel complications—not present in gamma-ray and neutrino astronomy—to the application of parametric point-source inference. Overcoming these challenges is one of the central aims of this work.

The NuSTAR telescope (Harrison et al. 2013; Wik et al. 2014; Madsen et al. 2015, 2017), in particular, possesses most of these complications; for this reason, NuSTAR is used as detector model to explore the effect that these complications have on parametric point-source inference. As compared to gamma-ray and neutrino astronomy, the NuSTAR X-ray data set presents the following unique challenges:

1. NuSTAR, like most X-ray telescopes, has a far narrower field of view (FOV) of 12′ per side, with hard edges due to the use of a detector plane with focusing optics. Fermi and IceCube have FOVs of approximately 40° and the full sky, respectively;
2. Although NuSTAR has a much narrower angular resolution, the NuSTAR PSF is a larger fraction of the FOV compared to Fermi and IceCube;
3. The NuSTAR PSF varies significantly as a function of position on the detector plane within a given observation; and
4. To compensate for the narrow FOV, multiple observations are often compiled together into a mosaic. This creates a complex and discontinuous detector response, with multiple overlapping PSFs for each observation. While the instrument response of both Fermi and particularly IceCube do vary across the sky, the variation is significantly smoother than common for X-ray data sets.

In order to study all of these effects, we have developed a detector simulation suite for NuSTAR. This simulation injects point sources into an binned map that forms a image. In this investigation, the bin sizes are much larger than the pixelization[10] of the NuSTAR X-ray detectors, and so the effect of this intrinsic pixelization on the image binning is not considered

[10] For the remainder of this article, "pixel" will be used to refer to the physical detector pixelization of the NuSTAR detector, while "bin" will refer the aggregation of pixels according to a scheme chosen by the analyzer. For the investigations presented here, the bins are 20 or more times larger than the physical pixelization.

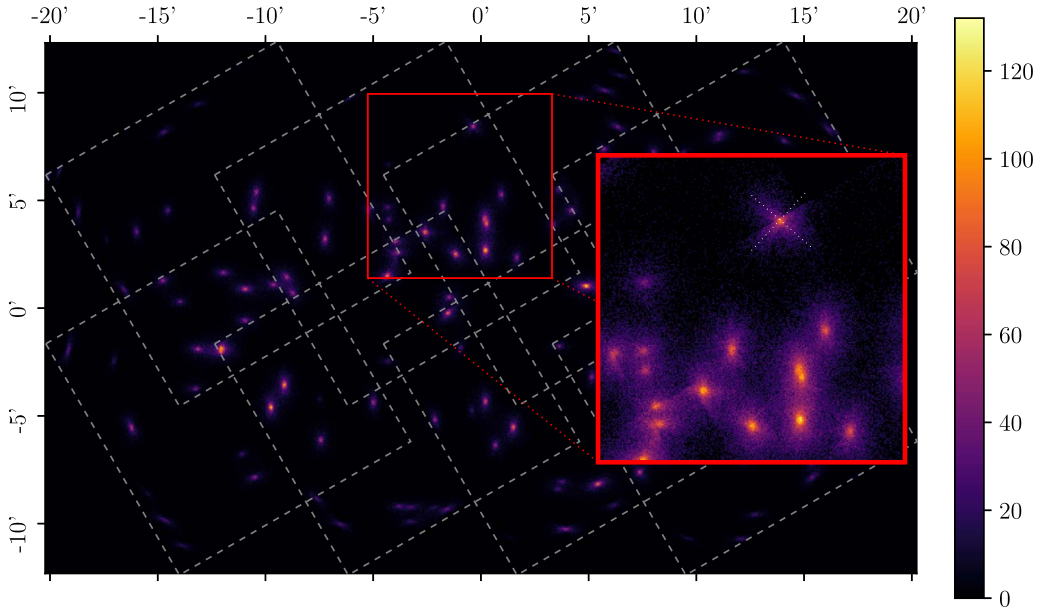

Figure 3. Multiple simulated observations are formed into a mosaic for the final image. The color scale shows the total number of detected photons per bin. The gray dashed lines show the individual observations that comprise the mosaic. The axes are centered around an arbitrary sky location in units of arcminutes. The red inset shows the details of a few sources in this image. The color scale of the inset has been scaled by a square root to improve contrast. For one particular source within the inset, the white dotted lines are each perpendicular to the optical axes of the nearby observations, and show the direction along which the PSF lies for that observation. The simulation details are the same as Section 6.5, except that the density of sources is five times less, the average flux per source 600 times greater, and the bin size is five times smaller than the NuSTAR detector pixelization.

here. The NuSTAR effective area—including vignetting effects—and PSF are incorporated into the simulation, and can be individually altered or simplified to assess the effect of each on the performance of the point-source methods investigated here. The simulation also requires a spatial distribution for the point-source population, as well a dN/dF function as defined by Equation (8). Further details on NuSTAR and the simulation procedure are given in Appendix A.

An example image generated by the simulation is shown in Figure 3. This image demonstrates many of the complications arising from the NuSTAR telescope. Multiple observations are stacked into a mosaic, and the boundaries between the observations, shown by gray dashed lines, cause discontinuities in the images of nearby point sources. In addition, the anisotropic PSF of NuSTAR causes complicated PSF shapes for sources near these boundaries, as shown in the red inset. The white dotted lines within the inset highlight one particular source that lies on the boundary of an observation. Two observations contribute to the image of this source, and so two PSF shapes (along the dotted lines) combine to create a highly irregular PSF shape.

The X-ray data will contain emission from contributions other than point sources. In particular, we expect both detector backgrounds to populate the collected data, as well as smooth astrophysical emission associated with, for instance, the cosmic X-ray background. We collectively refer to these non-point-source flux contributions as diffuse emission. The total diffuse emission will have an associated spatial map that specifies the mean expected flux at each location, which we can then combine with the detector response—as we did for the point-source flux F—to determine the mean predicted diffuse counts at each location. Simulation of these contributions is then performed by generating a map that is a draw from a Poisson distribution that is associated with the mean value of the diffuse map.

4. Derivation of the CPG Likelihood

We now turn toward the central goal of this paper, the construction of the CPG likelihood. The construction of the likelihood will heavily involve the use of probability-generating functions and functionals, as we will exploit their useful properties for constructing compound distributions. Generating functions are an alternative representation of a discrete probability distribution over non-negative integers. Suppose we have a distribution $P(k)$, which determines the probability of observing an integer k. The generating function is then defined by the Z-transform of the distribution:

$$G(z) = \mathbb{E}[z^k] = \sum_{k=0}^{\infty} P(k) z^k. \tag{10}$$

For example, the generating function for the Poisson distribution with mean λ is

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} z^k = e^{\lambda(z-1)}. \tag{11}$$

The generating function approach will be convenient for a number of reasons; however, a central property that we will exploit is as follows. Consider forming a sum of M independent and identically distributed random variates, each of which has a generating function $G_1(z)$, however with M itself an independent random variate, with its own generating function G_2. Then the generating function for the sum is given by $G_2(G_1(z))$, which follows directly from the expectation value definition of the generating function and the law of total expectation. For the problem at hand, this will arise in the context of counting the total number of counts detected from a population of point sources, for which the number of sources is itself a random variable.

We will build up the full CPG likelihood over the course of several steps, where each part describes the distributions and

generating functions for a major concept in the construction. Specifically, we divide the discussion as follows.

1. 4.1: Generating function for a single point source.
2. 4.2: Generating function for multiple sources.
3. 4.3: Definition of the detector effect correction.
4. 4.4: Calculation of the single-bin likelihood from the full generating function.
5. 4.5: Extending the single-bin likelihood to an image.
6. 4.6: Incorporating multiple population and diffuse emission models.

The goal of this section is to build an intuition for the moving parts of this construction. A more direct, but abstract, construction is provided in Appendix B, along with further discussion of the properties of generating functions and functionals. In that appendix, we also show the unbinned likelihood, as well as the likelihood that accounts for the correlations point sources induce between neighboring bins, as well as a discussion on why these correlations must be ignored for computational reasons in these demonstrations.

4.1. Single-source Generating Function

To begin with, consider the emission from a single source located at a position $\boldsymbol{x}$. The number of detected photons from a time-integrated observation of that source follows a Poisson distribution. This is a result of the exponential nature of the interdetection time distribution, and the statistical independence of multiple detections conditioned on the mean number of detected photons, S_B. Here, the index B denotes the spatial bin in which the counts are detected. Continuing, let $P_P(s_B|S_B)$ be the probability mass function for the Poisson distribution describing s_B detected photons with mean S_B. Then, given the true spatial location of the source, $\boldsymbol{x}$, the probability mass function for s_B at this location is

$$P(s_B|\boldsymbol{x}) = \int dS_B P_P(s_B|S_B) p(S_B|\boldsymbol{x}). \tag{12}$$

Here, we introduced $p(S_B|\boldsymbol{x})$, which is the probability distribution for the mean, S_B, for a given source location of $\boldsymbol{x}$, i.e., given a point source at $\boldsymbol{x}$, it is the probability that it produces a mean number of counts S_B in the bin B. As the exact value of S_B is unknown, we marginalize over it in the above expression. The exact configuration is shown in Figure 4(a), where a source at location $\boldsymbol{x}$ contributes s_B detected photons via the distribution $P(s_B|\boldsymbol{x})$ to bin B, which has a spatial size Ω_B.

We can move from the expected number of counts from a source, S_B, to the physical flux F, by combining a conditional distribution $p(S_B|F, \boldsymbol{x})$ and a distribution over source flux, $p(F)$, as follows:

$$p(S_B|\boldsymbol{x}) = \int dF p(S_B|F, \boldsymbol{x}) p(F). \tag{13}$$

This construction explicitly assumes that the flux distribution is isotropic, such that $p(F|\boldsymbol{x}) = p(F)$, an assumption that holds when the sources are themselves identically distributed—a key property that will be needed in the next section, and which we will assume throughout. The flux distribution itself has already been defined: it is given by the differential source-count function, dN/dF, in Equation (8). The conditional distribution $p(S_B|F, \boldsymbol{x})$, however, will depend centrally on the detector effect correction—as discussed already, the conversion from flux to counts intimately depends on the detector response—and it will determine the amount of flux F that contributes to the bin B. For the moment, we will leave it unspecified, postponing a definition until Section 4.3.

Combining Equations (12) and (13), we have

$$P(s_B|\boldsymbol{x}) = \int dS_B P_P(s_B|S_B) \times \int dF p(S_B|F, \boldsymbol{x}) p(F), \tag{14}$$

which fully specifies the probability of observing s_B counts from a single point source in terms of the differential source-count function, in $p(F)$, the instrumental response, in $p(S_B|F, \boldsymbol{x})$, and of course, the Poisson distribution, $P_P(s_B|S_B)$. From this, we can immediately determine the generating function for a single source located at a given position $\boldsymbol{x}$, $G_{s_B|\boldsymbol{x}}(z)$,

$$\begin{aligned} G_{s_B|\boldsymbol{x}}(z) &= \mathbb{E}_F[\mathbb{E}_{S_B|F,\boldsymbol{x}}[\mathbb{E}_{s_B|S_B}[z^n]]] \\ &= \int dS_B e^{S_B(z-1)} \int dF p(S_B|F, \boldsymbol{x}) p(F), \end{aligned} \tag{15}$$

which follows as the generating function for the Poisson distribution $P_P(s_B|S_B)$, that is, $\mathbb{E}_{s_B|S_B}[z^n] = e^{S_B(z-1)}$.

4.2. Multiple-source Generating Function

We next extend the discussion to account for the emission detected from a population of sources. Statistically, the locations of these point sources follows a spatial Poisson point process. This follows from the observation that:

1. The number of sources can be any non-negative integer.[11]
2. The occurrence of each source is independent from other sources. The presence of a point source does not have an effect on the probability of a new source forming.
3. Two sources never occupy the same spatial location. An infinitesimal area on the sky has only zero or one sources in it.

Spatial clustering of sources may violate the last two assumptions. If the clustering is known in advance (e.g., it is known that the sources cluster around the galactic center), then the construction presented here will render the source locations independent by conditioning on the spatial distribution of the sources. Some clustering processes cannot be made conditionally independent in this way; e.g., if the presence of one source increases the probability of further sources forming nearby. In this case, the construction presented here does not apply for counting sources. Instead, the method can count entire clusters as single sources, and the flux distribution, $p(F)$, would describe the flux of clusters. The size of this effect will depend on the degree to which the spatial distribution assumption is violated by the clustering. As an explicit example, the presence of binary emitters would violate these assumptions, and the CPG will count binary systems, rather than sources.[12]

To proceed, let $T(\boldsymbol{x})$ denote the spatial distribution for the location of a source (or system) in the population. In terms of $T(\boldsymbol{x})$, often referred to as the spatial template of the point

[11] In reality, there will be a physical upper bound to the number of sources one can find in most populations. Nevertheless, the number typically is sufficiently large that an unbounded process is an excellent approximation.

[12] For this caveat to be relevant, both objects must be emitters that can be detected by the instrument. If only one object in a binary is an emitter, the presence of the other is irrelevant to these assumptions.

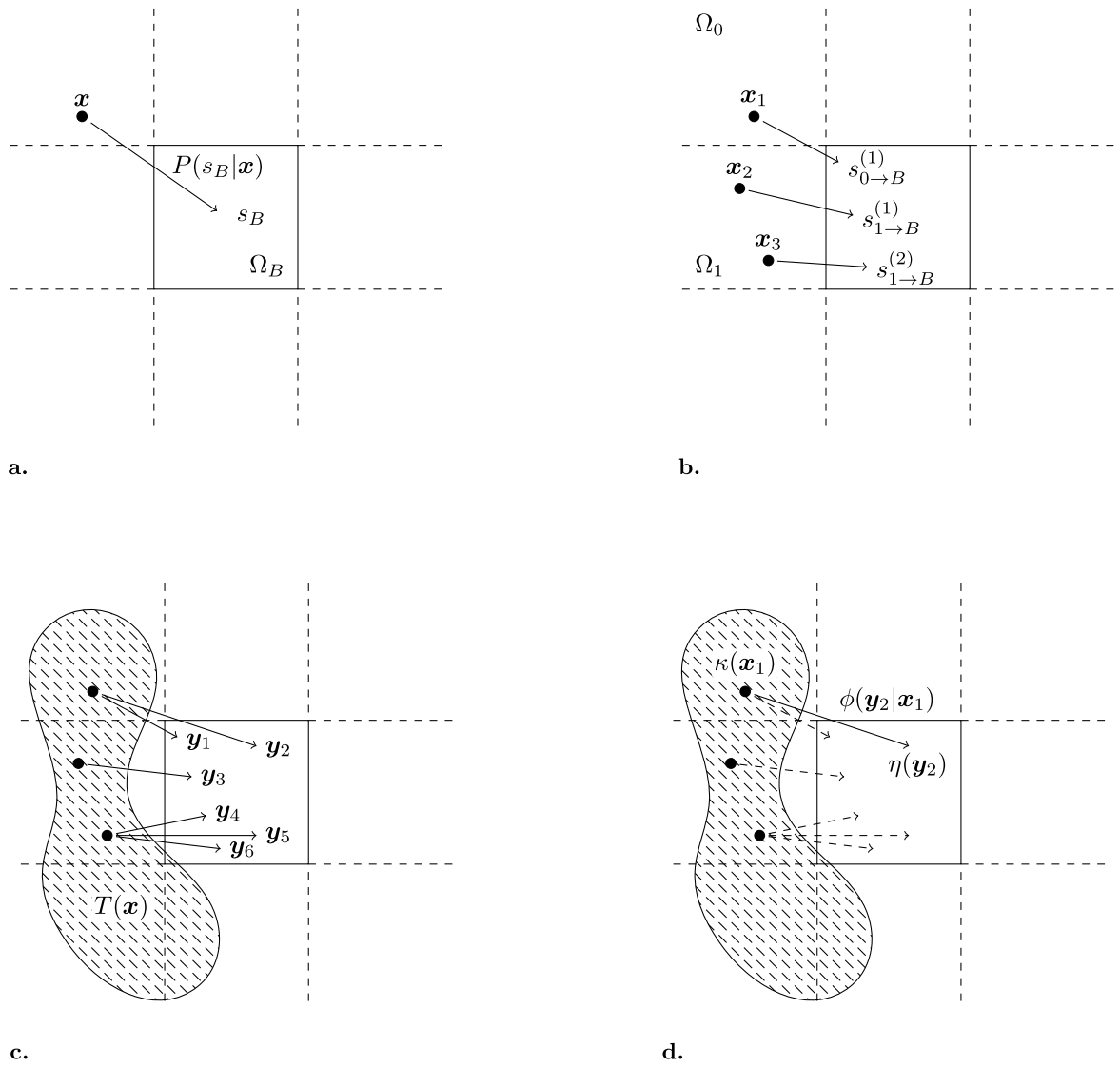

Figure 4. A relational diagram of the terms defined and used in Section 4 and Appendix B, demonstrating the relationship between the terms graphically. (a) A single source located at $\boldsymbol{x}$ contributes s_B counts to bin B as defined by the bin extents Ω_B (solid square). The number of counts is determined by the Poisson distribution $P(s_B|\boldsymbol{x})$ as defined in the text. (b) Multiple sources located at $\{\boldsymbol{x}_l\}$ each contribute $\{s^{(j)}_{i\to B}\}$ counts to bin B, creating a total of k_B counts in the bin. (c) A Poisson process describes sources spatially distributed according to $T(\boldsymbol{x})$ (sketched by shaded region). Each source may contribute multiple photons to the bin, which are recorded as counts at locations $\{\boldsymbol{y}_j\}$. (d) A single photon created by a source at $\boldsymbol{x}_1$ is converted to an expected number of counts using the following detector effects: the detector response (effective area and exposure time), $\kappa(\boldsymbol{x}_1)$, that is evaluated at the source location; the PSF distribution $\phi(\boldsymbol{y}_2|\boldsymbol{x}_1)$ evaluated at the location where the photon lands, $\boldsymbol{y}_2$, and conditioned on the source location; the detector efficiency, $\eta(\boldsymbol{y}_2)$, that is evaluated where the photon lands.

sources, the source intensity function is $NT(\boldsymbol{x})$, so that the integral of the source intensity function is equal to the mean number of sources, N.[13] In more detail, $T(\boldsymbol{x})$ carries units of [sr^{-1}], so that $NT(\boldsymbol{x})$ has dimension [sources sr^{-1}].

Now, let the template

$$T_B = \int_{\Omega_B} d\boldsymbol{x}\; T(\boldsymbol{x}) \tag{16}$$

be the fraction of the spatial distribution in bin B with bin extents Ω_B. We will use this to construct a generating function that accounts for sources located in surrounding bins that contribute flux to the current bin.

We define $k_{i\to B}$ as the number of counts that all sources located in bin i contributes to bin B. As there can be more than one source in bin i, $k_{i\to B}$ is a sum of M_i random variates,[14] $s^{(j)}_{i\to B}$:

$$k_{i\to B} = \sum_{j=1}^{M_i} s^{(j)}_{i\to B}, \tag{17}$$

as shown in Figure 4(b).

[13] We leave the region over which $T(\boldsymbol{x})$ is normalized unspecified. In general, it need not correspond to the region within which the analysis is being performed. For example, when studying an extragalactic source class, it may be convenient to normalize $T(\boldsymbol{x})$ over the full sky, even if this is larger than the region over which the data is collected.

[14] Here, we are calculating the likelihood exclusively for the bin B, and as such M_i is formally the number of sources in bin i that contribute counts to B. Accordingly, M_i is implicitly defined as a random variable in the context of calculating the likelihood for B. As this particular approach to the construction does not take into account correlations, there is a separate and independent random variate of M_i for each B.

Now, note that the spatial distribution for a source, when conditioned on it being located in bin i, is

$$T(\boldsymbol{x}|i) = \begin{cases} \dfrac{T(\boldsymbol{x})}{T_i} & \boldsymbol{x} \in \Omega_i \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

Then, each $s_{i\to B}^{(j)}$ is identically distributed according to the generating function $G_{s_{i\to B}}$. This generating function is formed by taking the expectation of the generating function[15] $G_{s_{i\to B}|\boldsymbol{x}}$ over $T(\boldsymbol{x}|i)$—the spatial distribution conditioned on the source being located in bin i:

$$G_{s_{i\to B}}(z) = \int d\boldsymbol{x}\; T(\boldsymbol{x}|i) G_{s_{i\to B}|\boldsymbol{x}}(z) \tag{19}$$

$$= \int_{\Omega_i} d\boldsymbol{x}\; \frac{T(\boldsymbol{x})}{T_i} G_{s_{i\to B}|\boldsymbol{x}}(z) \tag{20}$$

$$= \int dS_{i\to B}\; e^{S_{i\to B}(z-1)} \times \int dF\; p_{i\to B}(S_{i\to B}|F) p(F), \tag{21}$$

where

$$p_{i\to B}(S_{i\to B}|F) = \int_{\Omega_i} d\boldsymbol{x}\; \frac{T(\boldsymbol{x})}{T_i} p(S_{i\to B}|F, \boldsymbol{x}) \tag{22}$$

is now the instrument response for bin i contributing to bin B. This object can be intuitively thought of as the average response in bin B, from sources located in bin i, marginalized over the probability of a source appearing at any specific location within bin i.

As emphasized already, the number of sources, M_i, is itself a random variate distributed according to a Poisson distribution with mean NT_i, recalling that N is the mean number of sources for the total population. Thus, the generating function for M_i is

$$G_{M_i}(z) = e^{NT_i(z-1)}. \tag{23}$$

Now, the generating function for the sum $k_{i\to B}$ is the composition of the generating functions for M_i and $s_{i\to B}^{(j)}$:

$$G_{k_{i\to B}}(z) = G_{M_i}(G_{s_{i\to B}}(z)). \tag{24}$$

The total number of counts in bin B, denoted by k_B, is then the sum of all the $k_{i\to B}$ across all bins:

$$k_B = \sum_i k_{i\to B}. \tag{25}$$

Note that this sum must range over the full domain of the template, T_i. Primarily, this ensures that the recovered value of N is correctly normalized to the template, but it can also be understood to ensure that the contribution from any source—no matter where it may be in the spatial distribution of the population—is counted correctly, as the PSF generally allows contributions to k_B from anywhere in the spatial distribution.

Thus, the generating function for k_B is the product of the generating functions for $k_{i\to B}$:

$$G_{k_B}(z) = \prod_i G_{k_{i\to B}} = \exp\left[N\left(\left[\sum_i T_i G_{s_{i\to B}}(z)\right] - 1\right)\right], \tag{26}$$

as $\sum_i T_i = 1$.

At this stage, we could substitute into this expression $G_{s_{i\to B}}(z)$ as given in Equation (21). Before doing so, however, let us capture into a single function the detector response as it would appear in Equation (26):

$$p(S_B|F) = \sum_i T_i p_{i\to B}(S_{i\to B}|F) \tag{27}$$

$$= \int d\boldsymbol{x}\; T(\boldsymbol{x}) p(S_B|F, \boldsymbol{x}). \tag{28}$$

Using this, we can write the full generating function for k_B as

$$G_{k_B}(z) = \exp\left[N\left(\int dS_B e^{S_B(z-1)} \times \int dF\; p(S_B|F)p(F) - 1\right)\right]. \tag{29}$$

The distribution for k_B is known as a compound Poisson distribution (Daley & Vere-Jones 2003), and so G_{k_B} is a compound Poisson generator. As shown in Appendix B, this compound Poisson distribution actually describes a compound Poisson point process. The process gives the probability density of finding a detected count at location $\boldsymbol{y}$ as shown in Figure 4(c). Equation (29) generates the probability of finding k_B such detected counts, $\{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k_B}\}$, anywhere in Ω_B.

4.3. Detector Effect Correction

In the discussion thus far, we have determined the generating function associated with a population of sources, specified by a differential source-count function. The final aspect of Equation (29) we have avoided confronting is the instrument response, which we turn to now. Recall that we codified the instrumental effects as follows:

$$p(S_B|F) = \int d\boldsymbol{x}\; T(\boldsymbol{x}) p(S_B|F, \boldsymbol{x}), \tag{30}$$

where, as defined above, $T(\boldsymbol{x})$ is the spatial point-source template, and $p(S_B|F, \boldsymbol{x})$ accounts for how the instrument converts a flux F from a source at location $\boldsymbol{x}$ into an expected number of counts in bin B. In general, the instrument response and the spatial template cannot be fully factorized, as they are in the NPTF approach to the problem. We will see this explicitly in the discussion that follows.

Our treatment will consider four separate detector effects: the exposure time, which converts flux to time-integrated flux; the effective area, which converts time-integrated flux to expected photons incident on the detector; the PSF, which gives the probability density for the deviation of a photon's recorded direction of arrival from its true incident direction; and the detector efficiency, which gives the probability of a single incident photon being detected. Both the exposure time and effective area are merged into a single detector response value, $\kappa(\boldsymbol{x})$, which converts flux to mean counts for a point-source at location $\boldsymbol{x}$, and was discussed in Section 3. The PSF is a probability density, $\phi(\boldsymbol{y}|\boldsymbol{x})$, for a count detected at location $\boldsymbol{y}$ conditioned on its parent point source at location $\boldsymbol{x}$. The detector efficiency, $\eta(\boldsymbol{y})$, is the probability of a count being detected conditioned on the location of detection $\boldsymbol{y}$. The geometry of each function is shown in Figure 4(d), and examples of how these quantities combine for mosaiced images are shown in Figure 5.

[15] The index j does not parameterize this generator, as each $s_{i\to B}^{(j)}$ is a specific realization of the random variate $s_{i\to B}$.

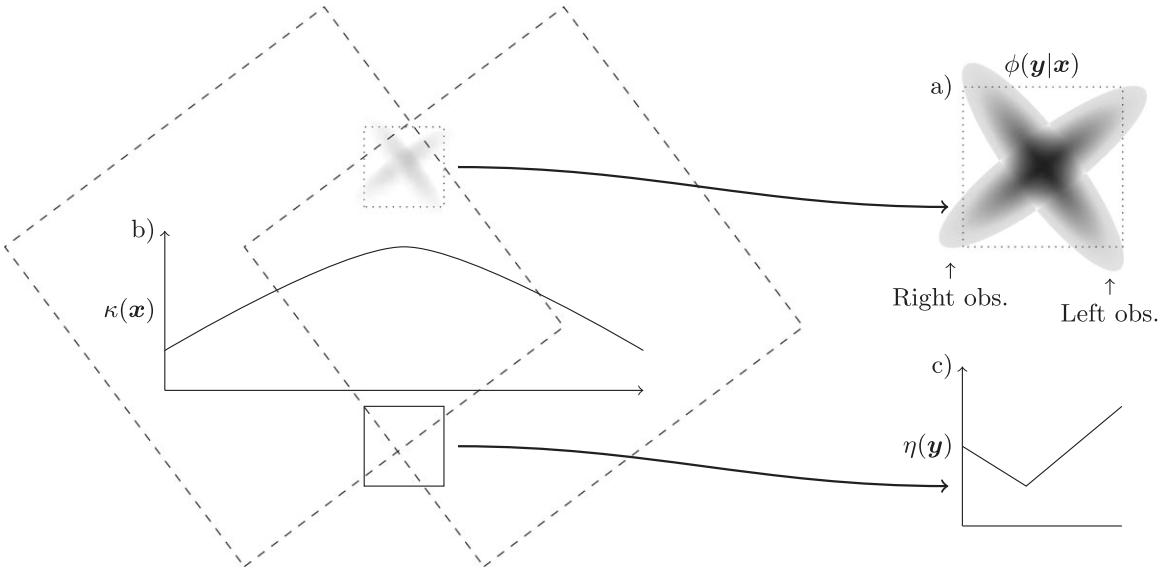

Figure 5. Two separate observations with dashed boundaries form a mosaic. (a) In areas where the observations overlap, the PSF will be an appropriate mixture of the PSF for each observation. Near the edge of the observation, we show an example of how the PSF can be highly distorted in the radial direction, as expected for NuSTAR observations. The mixture forms a cross, with the PSF from the left observation lying diagonally NW–SE and the PSF from the right observation lying diagonally SW–NE. (b) The detector response, $\kappa(\boldsymbol{x})$, is a sum of the response for each observation, and varies smoothly across the mosaic as the response extends outside the observation boundaries. (c) In contrast, the detector efficiency, $\eta(\boldsymbol{y})$, has a sharp transition from unity to zero at the observation boundary, and will generally be nonsmooth across any bin that straddles these boundaries.

To give a concrete example of these terms in the context of NuSTAR (Fermi, IceCube): a source is located in the direction of $\boldsymbol{x}$ and emits a primary X-ray (γ-ray, neutrino) from the direction of $\boldsymbol{x}$. This primary interacts with the optics (detector, ice) and produces a secondary X-ray (electron, lepton) that is scattered in the direction of $\boldsymbol{y}$ according to the PSF $\phi(\boldsymbol{y}|\boldsymbol{x})$. This secondary interacts with the active detector, causing a count to be recorded with probability $\eta(\boldsymbol{y})$.

In terms of these individual detector responses, the expected number of incident photons produced by a point source at location $\boldsymbol{x}$ with flux F is $S_{\boldsymbol{x},F} = \kappa(\boldsymbol{x})F$, and then the mean detected photon density at location $\boldsymbol{y}$ will be $S_{\boldsymbol{x},F}(\boldsymbol{y}) = \eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})S_{\boldsymbol{x},F}$. As a result, the mean number of detected photons in bin B is

$$S_B = \int_{\Omega_B} d\boldsymbol{y}\, S_{\boldsymbol{x},F}(\boldsymbol{y}), \tag{31}$$

and the associated distribution, $p(S_B|F, \boldsymbol{x})$, must be

$$p(S_B|F, \boldsymbol{x}) = \delta\left(S_B - F\kappa(\boldsymbol{x}) \int_{\Omega_B} d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})\right), \tag{32}$$

which then determines the marginal form as given in Equation (30).

Direct substitution of this result into Equation (29) is not immediately helpful, as it results in a double integral over spatial coordinates that must be evaluated during the computation of the likelihood. Instead, all of the detector effects can be encoded into a single measure, which we denote $\mu_B(\varepsilon)$, over an effective detector response variable, ε, such that $S_B = \varepsilon F$. In particular, we define

$$\mu_B(\varepsilon) = \int d\boldsymbol{x} T(\boldsymbol{x})\delta\left(\varepsilon - \kappa(\boldsymbol{x}) \int_{\Omega_B} d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})\right). \tag{33}$$

In terms of the new measure, we can restate Equation (30) as

$$p(S_B|F) = \int d\varepsilon \mu_B(\varepsilon)\delta(S_B - \varepsilon F). \tag{34}$$

Then, substitution of this form of $p(S_B|F)$ into Equation (29) reframes the CPG as

$$G_{k_B}(z) = \exp\left[N\left(\int dF \int d\varepsilon e^{\varepsilon F(z-1)}\mu_B(\varepsilon)p(F) - 1\right)\right], \tag{35}$$

which only involves integrals over the effective detector response ε and source flux F.

Importantly, the detector correction function, $\mu_B(\varepsilon)$, can be precalculated, thus saving considerable computation time during likelihood evaluation. As it is rare to have closed form expressions for $\kappa(\boldsymbol{x})$ and $\phi(\boldsymbol{y}|\boldsymbol{x})$ in most experiments, $\mu_B(\varepsilon)$ will almost always need to be numerically estimated. This estimation can proceed by effectively evaluating the integrals over $\boldsymbol{x}$ and $\boldsymbol{y}$ in Equation (33) through Monte Carlo integration. Samples are drawn from $T(\boldsymbol{x})$, and then—conditioned on these values of $\boldsymbol{x}$—samples are drawn from $\phi(\boldsymbol{y}|\boldsymbol{x})$. The values of $\boldsymbol{y}$ are accumulated to create a value of ε, and the resulting ε values are histogrammed to create a density estimate over ε—which is $\mu_B(\varepsilon)$. This process, and an explicit algorithm for construction $\mu_B(\varepsilon)$, is detailed in Appendix E. As emphasized at the outset, the detector response involves the source template intimately. This is distinct from the handling of the detector response in NPTF, which occurs through the function $\rho(f)$, and we will explore the differences between these two approaches in Section 6.

4.4. Calculation of the Single-pixel Likelihood

Once $\mu(\varepsilon)$ has been constructed, evaluation of the CPG in Equation (35) requires the integrals over ε and F to be performed. As $\mu(\varepsilon)$ is numerically constructed, the integral over ε will also be performed numerically. The remaining integral

over F may then be performed numerically, or analytically, if the assumed form of $p(F)$ is amenable to such treatment. For the examples in this investigation, $p(F)$ is assumed to follow a broken power-law distribution, in which case the evaluation can be performed analytically as detailed in Appendix C. In this section, the only assumption required is that this evaluation produces a power series in z:

$$G_{k_B}(z) = \exp\left[\sum_{m=0}^{\infty} \frac{a_B^{(m)}}{m!} z^m\right]. \tag{36}$$

This is a fairly mild assumption, as it only requires that the expression within the square brackets of Equation (35) be an analytic function of z. This is easily satisfied when the moments of $p(F)$ and $\mu_B(\varepsilon)$ are finite. For now, this power series will be assumed to be infinite in order; later, it will be shown that only a finite order is required in practice.

The goal is to find $P(k_B)$, the probability of measuring k_B detected photons, from this generating function. Recall that a generating function is defined as

$$G(z) = \sum_{k=0}^{\infty} P(k) z^k. \tag{37}$$

The relationship between the power series in Equation (36) and the power series in Equation (37) is given by the Bell polynomials (Comtet 1974):

$$\exp\left[\sum_{m=0}^{\infty} \frac{a_B^{(m)}}{m!} z^m\right] = \sum_{k=0}^{\infty} \frac{B_k(a_B^{(0)},\ldots,a_B^{(k)})}{k!} z^k. \tag{38}$$

Thus, by inspection, we have in this case $P(k_B) = B_k(a_B^{(0)},\ldots,a_B^{(k)})/k!$, and note that from Equation (37), for a bin with k counts, only the first k terms of the power series in Equation (36) need to be calculated. The evaluation of the Bell polynomials can be performed using recurrence relations, as we demonstrate in Appendix D.

4.5. Whole-image Likelihood

Through Equation (38) and the results preceding it, we have achieved our aim of writing the single-bin likelihood, $P(k_B|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the model parameters for the population, as encoded in dN/dF. A likelihood for the whole image, $\mathcal{I} = \{k_B: \forall B\}$, can be constructed as a simple product over the bins,

$$P(\mathcal{I}|\boldsymbol{\theta}) = \prod_{k_B \in \mathcal{I}} P(k_B|\boldsymbol{\theta}). \tag{39}$$

Importantly, though, this construction does not take into account the correlations between the bins, which may be induced by the PSF. Indeed, it assumes that the k_B are statistically independent, which will be not be true unless the PSF is a delta-function distribution. Such a delta distribution ensures that sources at the edge of a pixel only deposit flux in a single bin.

The effect of this broken assumption is that the resulting posterior distribution on $\boldsymbol{\theta}$ will be narrower than the true posterior if these correlations were accounted for—by treating every pixel as independent, we have assumed there is more available information than is actually present in the image. This will underestimate the uncertainty on the model parameters, as sources that overlap bins are effectively counted as multiple independent observations of the same source, instead of the true single observation—a similar effect to double-counting data. The degree to which the posterior is narrowed will depend on the chosen bin size, as smaller bins—relative to the PSF—will be more strongly correlated. The test cases in Section 6 show that this is not a significant effect for the bin sizes chosen in this investigation, which are several times larger than the PSF.

In principle, there are other ways correlations between pixels can be induced that would invalidate the factorization in Equation (39)—for instance, an incorrect template for one of the Poisson models, or perhaps other instrumental effects. The reason we single out the PSF is that it is never truly a delta-function distribution in any real instrument, and this represents the most significant correlation that can only be mitigated by appropriate choice of bin sizes.

While this is an unambiguous deficiency of the above derivation, so far no computationally feasible method has been proposed to account for the correlations in a binned analysis. Indeed, the most obvious extensions of the above construction require the computation of every possible combination of how a source can distribute its counts to all bins in the image. As stated, in the present work, we will work with a binning such that this shortcoming is suppressed, but we caution that, for a general binning, the biases associated with this effect must be considered.

4.6. Multiple-emission Models

It is a common occurrence for an image to have contributions from both populations of point sources as well as purely Poisson emission or background components. As such, it is important to be able to accommodate this reality in the likelihood, as we do so in this subsection. Let $k_{B,j}$ be the number of counts that population j contributes to bin B, and $\varpi_{B,l}$ be the number of counts that Poisson component l contributes to the same bin. Then the total number of counts in this bin is simply a sum over the contribution from each component,

$$K_B = \sum_j k_{B,j} + \sum_l \varpi_{B,l}. \tag{40}$$

In turn, the generating function for the combined emission, K_B, is

$$G_{K_B}(z) = \left(\prod_j G_{k_{B,j}}(z)\right)\left(\prod_l G_{\varpi_{B,l}}(z)\right). \tag{41}$$

For point-source populations, the generating function is as derived earlier in this section, whereas $G_{\varpi_{B,l}}(z) = \exp[\lambda_{B,l}(z-1)]$ is the generating function for $\varpi_{B,l}$, and

$$\lambda_{B,l} = \int_{\Omega_B} d\boldsymbol{y}\, I_l(\boldsymbol{y}) \tag{42}$$

is the integral of the intensity function $I_l(\boldsymbol{y})$ for Poisson component l in bin B, which parameterizes the mean of $\varpi_{B,l}$. Recall that a single point source in the population is specified by a flux F, which carried dimensions of [photons cm^{-2} s^{-1}]. In comparison, the Poisson diffuse-emission component is an extended source and has a differential flux of $F_P T_P(\boldsymbol{x})$ with dimensions [photons cm^{-2} s^{-1} sr^{-1}]. Here, $T_P(\boldsymbol{x})$ is the template for the Poisson emission, which may or may not be the

same as the spatial distribution of the source population, $T(\boldsymbol{x})$. It does, however, have the same units of [sr^{-1}], so that F and F_P will also carry the same dimensions. In terms of these quantities, the Poisson intensity is given by

$$I(\boldsymbol{y}) = \eta(\boldsymbol{y}) \int d\boldsymbol{x} \phi(\boldsymbol{y}|\boldsymbol{x}) \kappa(\boldsymbol{x}) F_P T_P(\boldsymbol{x}), \tag{43}$$

as photons from the diffuse component are also scattered by the PSF. The mean can then be written more compactly as

$$\lambda_{B,l} = F_{P,l} \int d\varepsilon \; \varepsilon \; \mu_{B,l}(\varepsilon), \tag{44}$$

where $F_{P,l}$ is the flux for Poisson component l, and $\mu_{B,l}$ is the detector correction function for template $T_{P,l}$. As for the point-source population, we envision the spatial template as fixed, which leaves a single model parameter for the emission, F_P.

Combining these results, the generating function for K_B can be written in the same form as Equation (36):

$$\begin{aligned} G_{K_B}(z) = \exp\Bigg[& \left(\sum_j a_{B,j}^{(0)} - \sum_l \lambda_{B,l} \right) \\ & + \left(\sum_j a_{B,j}^{(1)} + \sum_l \lambda_{B,l} \right) z \\ & + \sum_{m=2}^{\infty} \left(\sum_j \frac{a_{B,j}^{(m)}}{m!} \right) z^m \Bigg], \end{aligned} \tag{45}$$

and the same likelihood evaluation method of Section 4.4 can be employed, thereby completely specifying the CPG likelihood.

5. Biases Induced by Common Prior Parameterizations

In this section, the priors on the population model parameters —necessary for a Bayesian analysis—are discussed. Our focus will be to demonstrate that poorly chosen priors on a common combination of the population flux and background flux can lead to misleading posterior distributions. We further introduce a set of priors where these issues can be reduced, and advocate for their use more generally in population studies.

Let us make a general point at the outset. In Bayesian analyses, one is free to choose any set of priors. Priors can be adopted that reflect a preference toward either hypothesis. The Poisson hypothesis may be preferred for its simplicity, or alternatively one may wish the results to reflect an underlying bias toward the point-source model, given that in many situations it is known that unresolved point sources must be present. Whatever set of priors is adopted, however, the preference they reflect should be considered. As we will show in this section, taking simple priors on the parameters that describe the point-source model can induce a complex bias in the question of which model is generating the flux. The priors we will introduce instead place the Poisson and point-source models on equal footing at the outset, and if not adopted directly, at the very least represent a starting point for adopting a principled set of priors for Bayesian point-source inference.

For the remainder of this investigation, we will restrict the model under question to at most one population of point sources and one Poisson component, that both share a common template, i.e., $T(\boldsymbol{x}) = T_P(\boldsymbol{x})$. Realistic analyses are more complicated than this restriction; for instance, existing NPTF Fermi analyses involve two (or three) point-source population models and multiple Poisson components, while the NPTF IceCube analysis used one population with multiple Poisson components. However, common to both analyses was the use of a point-source population model and Poisson component with an identical spatial distributions. This is the situation where the particular bias we will discuss can emerge, justifying our restriction.

Fundamentally, such a scenario arises when the underlying nature of flux distributed according to $T(\boldsymbol{x})$ is unknown, and the question of interest is whether it is due to a measurable population of sources, purely diffuse emission, or instead a mixture of the two. By using a common template for the two possible emission sources, the flux could be assigned to either, or some fraction to both. For example, in the case of Fermi, the fundamental question was determining whether the anomalous flux at the Galactic Center was attributable to a population of sources, such as millisecond pulsars, or instead to diffuse dark matter emission that would be Poisson-distributed. For the example of IceCube, the goal has been to determine whether a measurable fraction of the astrophysical neutrino flux can be assigned to a population of sources.

To begin the investigation, the differential source-count function must be parameterized in order to evaluate the power-series terms, $a_B^{(m)}$ as given in Equation (45), and accordingly the CPG likelihood. Note that, while we will use the CPG likelihood in order to demonstrate the potential prior bias, the effects we reveal are equally applicable to other point-source likelihoods such as the NPTF. The use of the CPG also furnishes us with an example to demonstrate the application of the likelihood. In the Fermi and IceCube analyses, the following standard form was used for the source-count distribution:

$$\frac{dN}{dF} = A \begin{cases} \left[\prod_{j=2}^{m} \left(\frac{F_{b(j+1)}}{F_{b(j)}} \right)^{-n_j} \right] \left(\frac{F}{F_{b(m)}} \right)^{-n_m} & F \in [0, F_{b(m)}] \\ \vdots & \\ \left[\prod_{j=2}^{i} \left(\frac{F_{b(j+1)}}{F_{b(j)}} \right)^{-n_j} \right] \left(\frac{F}{F_{b(i)}} \right)^{-n_i} & F \in (F_{b(i+1)}, F_{b(i)}] \\ \vdots & \\ \left(\frac{F}{F_{b(2)}} \right)^{-n_2} & F \in (F_{b(3)}, F_{b(2)}] \\ \left(\frac{F}{F_{b(2)}} \right)^{-n_1} & F > F_{b(2)}. \end{cases} \tag{46}$$

This defines a broken power law with $m-1$ breaks in flux, specified by $F_{b(m)},\ldots,F_{b(2)}$; m power-law indices, parameterized by $n_m,\ldots,n_1$; and a scale factor A, which is related to the expected number of sources in the population. The break parameters have the same units as F, while A has units of sources per inverse units of F. Note this is a generalization of the singly broken power law given in Equation (7).

With this parameterization, Fermi and IceCube analyses—see, for example, Lee et al. (2016) and Aartsen et al. (2020), respectively—have used a Bayesian inference framework. In both cases, uniform priors on the indices, n_i, and log-uniform priors on A were chosen. For the Fermi analysis, the flux breaks, $F_{b(2)}$, and Poisson component flux F_P were given uniform and log-uniform priors, respectively, whereas in the IceCube analysis, $F_{b(2)}$ was given a log-uniform prior while F_P was given a uniform prior.

5.1. A Sketch of Poor Prior Parameterization

Let us first outline where the issues associated with the poor prior parameterization originate. For simplicity, we will concentrate on a single-break differential source-count function; however, the arguments here generalize to multiple breaks. It is important to note that the total flux of the point-source population, $F_{\rm PS}$, is not proportional to $F_{b(2)}$, and also depends on A,

$$F_{\rm PS} = AF_{b(2)}^2\left(\frac{1}{n_1 - 2} + \frac{1}{2 - n_2}\right). \quad (47)$$

Already from this expression we can make the following observation: a uniform prior chosen for $F_{b(2)}$ will not, in general, uniformly weight values of $F_{\rm PS}$ after the change in coordinates.

Consider a situation in which this model is used to analyze data that have no distinct population of sources, such that the data are entirely consistent with a Poisson distribution. Clearly, the model can explain these data by assigning the entirety of the flux to the Poisson component; however, this is not the only solution for this inference problem. In particular, if we have a population of dim sources, there is a limit in which the sources are so dim that this distribution becomes indistinguishable from Poisson emission. To provide an explicit example, if we had a population with N expected sources, each of which produces μ counts on average, then the mean number of counts produced by the population is $\lambda = N\mu$. In the limit where $\mu \ll 1$, such that all sources produce either 0 or 1 counts only, then the point-source likelihood exactly reduces to the Poisson distribution with mean λ, and the two hypotheses are formally indistinguishable in the data. Note that, for λ to stay finite, we require a large N when $\mu \ll 1$. Returning to our single-break scenario, note that

$$N = AF_{b(2)}\left(\frac{1}{n_1 - 1} + \frac{1}{1 - n_2}\right), \quad (48)$$

and so for this point-source Poisson degeneracy regime to be achieved here, the prior on A must be sufficiently large. But so long as it is, then in principle data associated with Poisson emission can be equally well described by the diffuse or point-source hypothesis.

Ideally, in this situation, the posterior should show complete uncertainty on F_P and $F_{\rm PS}$, with a perfect anticorrelation that corresponds to the total flux in the image.[16] However, if the same kind of prior is chosen for $F_{b(2)}$ and F_P—for example, if both are log-uniform—then the corresponding prior on $F_{\rm PS}$ may have a different form than the prior on F_P. In this case, the posterior will show a preference to assigning flux to either the population or the Poisson component. If the difference between the priors is substantial, essentially all of the flux will be assigned to one of the two components, despite the data having no power to distinguish them.

Unless this subtlety in choice of prior is accounted for, the result may be unexpected, potentially leading an experimenter to erroneously conclude that the data support a population of sources where there are none, or vice versa. This is the bias we will explore in more detail below, and then outline priors that can be chosen such that the two hypotheses remain indistinguishable given uninformative data.

[16] Of course, in principle one may wish to choose a prior that reflects a preference for one of the two hypotheses. Nevertheless, this preference should be placed in the priors in a principled manner—the point we are seeking to emphasize in the present discussion is that existing priors adopted in the literature represent a complicated transformation away from the natural coordinate system we introduce, thereby introducing biases that could be unintended.

5.2. Prior Effect Demonstration

In order to investigate in detail the potential bias induced by the choice of prior parameterization as discussed in the previous subsection, we consider simulated NuSTAR data sets (for details of the NuSTAR simulation see Appendix A). For this scenario, the vignetting is disabled so that the detector response is uniform across the image. In addition, the PSF is locked to the on-axis PSF of NuSTAR, so that the PSF does not change as a function of source location. These simplifications are chosen to ensure that the effect of prior parameterization is not obscured.

The spatial distribution for both the population model and the Poisson component is specified as a uniform distribution concentric with the image and with a width and height twice that of the field of view. As this demonstration requires data that are indistinguishable from a Poisson distribution, no point sources are injected and only a uniform Poisson background is used to generate the image. Further details are given in Table 1.

Given this scenario, we perform a Bayesian analysis of the resulting images using the CPG, but taking four variations on the choice of prior, considering variations for the prior of the amplitude A and the flux parameters $F_{b(2)}$ and F_P. The same form of prior is used for $F_{b(2)}$ and F_P. We consider the four combinations resulting from a uniform linear or log-uniform prior on the amplitude and flux parameter. The detailed prior ranges are given in Table 2. We take the same lower limit for both flux priors. The upper limit of the F_P prior is extended because this parameter captures the total flux of the Poisson component, while $F_{b(2)}$ is more closely related to the average flux of a single source in the population, so we allow for a lower value. The upper and lower limits for the priors on each parameter are identical across variations. Uniform priors are chosen for the power-law indices, n_i. Of course, the results from running point-source inference on one simulated image may not be representative, as the simulation is a Monte Carlo procedure that may produce an outlier image. To capture the variation in the simulation, for each choice of priors, six images are generated and a posterior is sampled for each of these trials. The posteriors are sampled using the `emcee` Affine Invariant MCMC (Foreman-Mackey et al. 2013), discarding the first 90% of samples as burn-in.

From the resulting posteriors, we focus our attention on a single parameter of interest: the proportion of the flux that is assigned to the population model. In particular, we define the fraction of flux assigned to the population as $\omega = F_{\rm PS}/(F_P + F_{\rm PS})$. For each of the trial images, a coordinate transformation is applied to the posterior samples, to yield samples in ω. These samples are then histogrammed, and from the set of trials the median, 16% and 84% quantiles over the histogram bins are computed and shown in Figure 6. The clear trend observed here is a highly bimodal posterior in ω—the model assigns essentially all of the flux to only the population or the Poisson component in a situation where from the perspective of the data, the two are indistinguishable. Unless this behavior is anticipated, it could generate misleading conclusions. Again, the data support both the population model

Table 1
All Scenario Configuration Options

Setting	Prior Demo	Nonuniform Dist	Anisotropic PSF	Sub-bin Eff. Area	$\rho(f)$ Scenario	Realistic
Spatial dist.	Uniform	5 Delta	Uniform	Uniform	Uniform	Uniform
Vignetting	Off	Off	Off	Checkerboard	Off	On
PSF	Isotropic	Isotropic	NuSTAR	Delta distribution	Isotropic	NuSTAR
Binning	9×9	9×9	9×9	4×4	9×9	10×10
Energy range	3–10 keV	3–10 keV	3–10 keV	3–10 keV	3–10 keV	3–10 keV
Energy spectrum	$E^{-1.5}$	$E^{-1.5}$	$E^{-1.5}$	$E^{-1.5}$	$E^{-1.5}$	$E^{-1.5}$
Background	2×10^{-2} s^{-1}	None	None	None	None	8.73×10^{-2} s^{-1}
CXB	None	None	None	None	None	1.5×10^{-3}
Exposure time	200 ks	200 ks	200 ks	200 ks	200 ks	200 ks
Exposures	Single	Single	4×2 adjacent tiled	Single	Single	4×2 tiled
A	N/A	1.5×10^{9}	1.5×10^{12}	N/A	8×10^{10}	10^{12}
F_b	N/A	1.2×10^{-8}	3×10^{-9}	N/A	6×10^{-9}	6×10^{-9}
n_1, n_2	N/A	3, −2	3, −2	N/A	3, −2	3, −2
N	N/A	N/A	N/A	1.2×10^{4}	N/A	N/A
F	N/A	N/A	N/A	3×10^{-10}	N/A	N/A

Note. Flux parameters are in units of [counts cm^{-2} keV^{-1} s^{-1}]. A parameters are in units of the inverse of the flux units. CXB flux is given in units of [photons s^{-1} cm^{-2} deg^{-2}]. For scenarios with mosaiced exposures, the binning value gives the equivalent bin size for one observation.

and the Poisson component, so one would expect the posterior on ω to be close to uniform.

Recently, in the context of application of NPTF to the Fermi GCE, concerns have been raised with regard to the possibility that flux from a diffuse dark matter component could be mistakenly attributed to the point-source population model. Leane & Slatyer (2019) raised these concerns in relation to mismodeling of the spatial distribution of sources, and the authors show that a strong flux misattribution effect can result from spatial mismodeling. The authors also examine the attribution of flux when the spatial distribution is correctly modeled. In particular, Figure S7 shows that, even with correct modeling, some diffuse dark matter flux is attributed to the point-source model—although it should be clarified that the effect is considerably weaker than the spatial mismodeling effect observed in the rest of the study. Figure S7 also shows that, when the true flux of the diffuse dark matter component exceeds the flux of the point-source population, the flux is then attributed to the diffuse component of the model. We can understand that this behavior is very likely impacted by the choice of priors in that work, given their importance as demonstrated above. In particular, Leane & Slatyer (2019) placed a log-uniform prior on the diffuse emission and A, but a uniform prior on $F_{b(2)}$. The use of a uniform prior for the point-source model flux break will place greater weight on this model to describe the diffuse flux—when the dN/dF for this model can accommodate both the diffuse flux and the brighter point-source emission favored by the data. Accordingly, when small amounts of diffuse flux are injected, we would expect it to be absorbed by the point-source model for this choice of priors. However, once the diffuse flux is comparable to or larger than the point-source flux, it becomes difficult to explain both the population *and* the diffuse flux using the same power-law dN/dF. Thus, the model reverts to attributing the diffuse flux to the Poissonian model.

The prior effect can be observed more clearly in Chang et al. (2020). Figure 3 of this study shows the posterior distribution for the flux of a point-source and Poissonian component for a dark matter GCE scenario. Although this study concludes that all flux is correctly attributed to the dark matter component in the posterior, in fact, based on the previous arguments and

Table 2
Prior Ranges for the Demonstration Using the Standard Differential Source-count Function Parameterization

Parameter	Lower Limit	Upper Limit
A	10^{8} cm^{2} s	10^{14} cm^{2} s
$F_{b(2)}$	10^{-15} cm^{-2} s^{-1}	10^{-3} cm^{-2} s^{-1}
F_P	10^{-15} cm^{-2} s^{-1}	10^{8} cm^{-2} s^{-1}
n_1	2	5
n_2	−3	0

Figure 6, the unbiased posterior would uniformly assign the flux between the point-source and dark matter components. The observed posterior is instead likely driven entirely by the chosen priors and the constraint that the flux of the population and Poisson component must sum to the total flux in the image.

More generally, if we denote the total flux by F_T, so that $F_T = F_{\rm PS} + F_P$—a diagonal line on the plane of $F_{\rm PS}$ and F_P—it is clear that the only choice of flux priors that will result in the expected behavior are ones that assign equal probability to all values of $F_{\rm PS}$ and F_P that lie on this diagonal. Two choices of priors satisfy this requirement: either both uniform priors or both exponential priors on $F_{\rm PS}$ and F_P.[17] This may appear to suggest that a flat posterior on ω should be observed in Figure 6 for the "Uniform F" cases. The nonuniform posterior in these cases comes from a second effect: although the prior on the flux break, $F_{b(2)}$, is specified to be uniform, this does not mean the prior on the total point-source flux is uniform, as emphasized below Equation (47) above. Recall that the total flux is proportional to $AF_{b(2)}^2$, so a uniform prior on A and $F_{b(2)}$ is effectively an F_T^{-1} prior on the total flux, resulting in the observed concentration toward $\omega = 0$.

A log-uniform prior on A with a uniform prior on F effectively generates an N^{-2} prior on the mean number of sources. Although this appears to be a uniform prior on the total flux, one must consider how the boundary of the prior transforms. The effect of the prior boundaries can be incorporated by marginalizing out N from the joint prior,

[17] As for an exponential prior, $p(F) \propto \exp(-F)$, so the product of such priors for $F_{\rm PS}$ and F_P is $\propto \exp(-F_{\rm PS} - F_P) = \exp(-F_T)$.

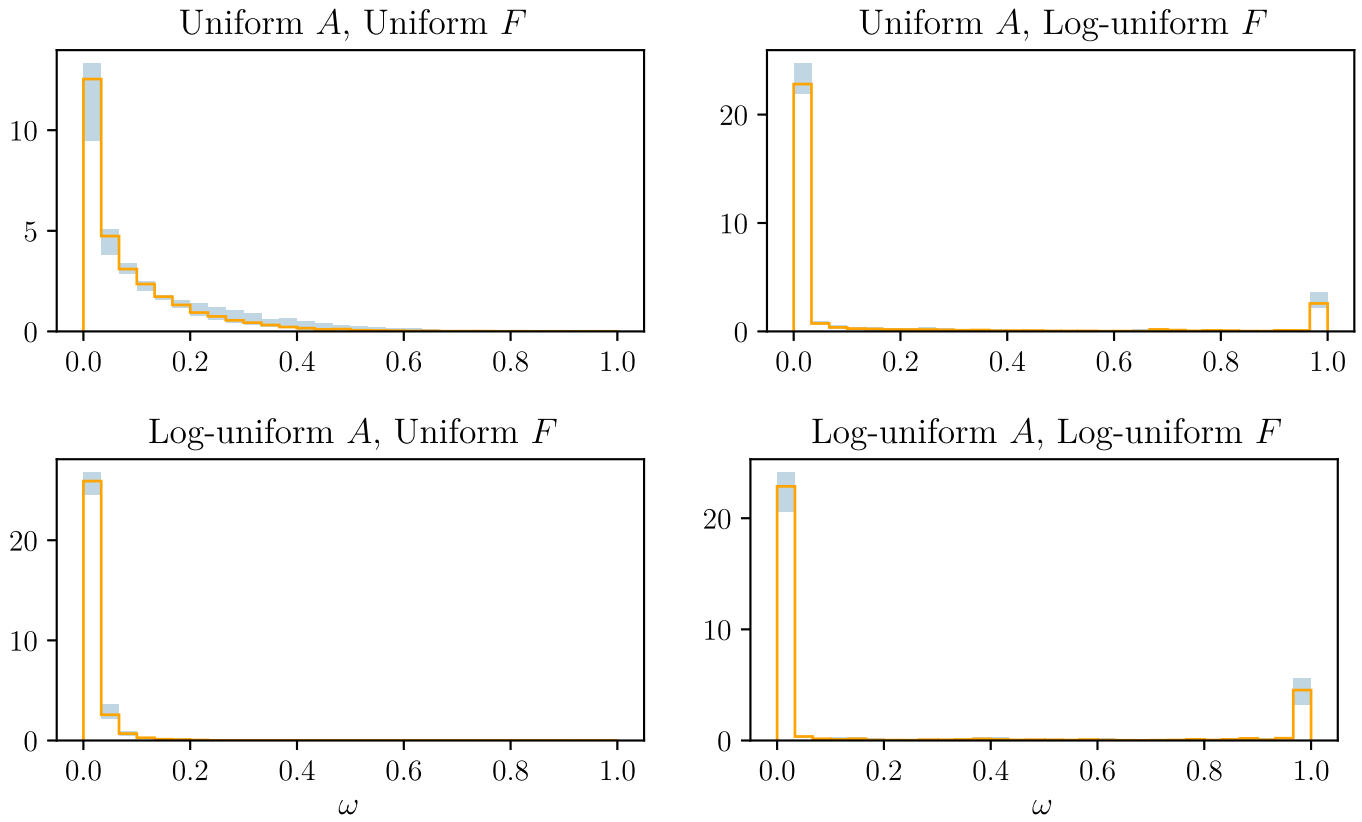

Figure 6. The posterior for the fraction of flux assigned to the point-source population, $\omega = F_{\rm PS}/(F_P + F_{\rm PS})$, when analyzing data sets generated from purely a Poisson distribution, but analyzed with both a Poisson and point-source template using the CPG likelihood. The median posterior (over multiple random trials, for each bin) is shown by the solid line, while the 16% to 84% quantile range is shown by the fill. The four panels correspond to different prior choices, representative of common choices in the literature. None of the choices of prior result in a uniform posterior for ω; instead, the posterior is biased toward assigning all flux to one model component or another, which could easily be misleading unless accounted for.

which results again in an F_T^{-1} prior on the total flux. Accordingly, none of the commonly existing prior choices result in the desired flat ω distributions. Of course, the particular priors existing choices achieve may be desirable in certain circumstances. Nevertheless, when considering the question of whether emission is fundamentally Poisson or point-source, it is undoubtedly useful to have a system of priors that generates a posterior where both are treated equally when the data are uninformative. With this in mind, in the next section we will introduce a new approach that involves reparameterizing the priors entirely with a new coordinate system.

5.3. Parameterizing Priors in a Natural Coordinate System

Much of the previous discussion has already suggested an alternative approach. A natural way to describe the combination of both population and Poisson component is through the coordinates of ω and F_T, the relative and total flux, respectively, so that the population model is specified in terms of $F_{\rm PS}$. From ω and F_T, the flux of either component is straightforward to calculate, although we caution that, given the inherent degeneracy, the individual fractions must be interpreted carefully. Nevertheless, as it is considerably easier to define and compute the likelihood using the A, n_i, and $F_{b(i)}$ coordinates, a coordinate transform is needed from $F_{\rm PS}$.

First, the complete coordinate system for the population model must be defined. A natural complement to $F_{\rm PS}$ is N, the expected number of sources. From this, the average flux per source can be readily determined. Next, the position of the breaks must be defined. For $m-1$ breaks, $m-1$ coordinates are required. The $F_{\rm PS}$ coordinate is necessarily involved, leaving $m-2$ remaining coordinates to specify. These are defined as a series of fractions, β_i, that give the location of each break relative to the previous break:

$$\beta_i = \frac{F_{b(i+1)}}{F_{b(i)}}. \tag{49}$$

These coordinates define a system of equations from which the break locations can be solved. The full transformation is detailed in Appendix G.

Finally, the power-law indices must be specified. These could be left as the $\{n_i\}$ in Equation (46); however, we take the opportunity to correct another subtle issue with the priors. The most common choice of prior on n_i is a uniform prior. Suppose that such a uniform prior is chosen for n_1 in the range of 2–100. The uniform prior assigns nearly 10 times as much prior probability to $10 < n_1 < 100$ as it does to $2 < n_1 < 10$. This is despite the fact that most observed power-law indices are less than 10; thus, a uniform prior is contrary to our prior knowledge of these physical systems. Instead, the index can be specified as an angle, ψ_i, and the index defined as $n_i = \tan\psi_i$. A uniform prior on ψ_i now places as much probability to $2 < n_i < 4$ as it does to $4 < n_i < 38$. The intuition is also clear, on a log–log plot of the source-count function, the prior is uniform on the angle of the line formed by the power law. This should not be taken as an objectively better choice,

and there may be scenarios where a uniform prior on the power-law index is a preferable. However, the uniform prior is all too often used without much consideration, and the intention here is to provide a principled alternative.

In summary, we propose defining the priors on the broken power law with $m-1$ flux breaks, as defined in Equation (46), using the coordinate system $\{N, F_{\rm PS}, \beta_2,\ldots,\beta_{m-1}, \psi_1,\ldots,\psi_m\}$, rather than $\{A, F_{b(2)},,\ldots, F_{b(m)}, n_1,\ldots,n_m\}$. A Poisson component may be added to these coordinates through the ω flux fraction parameter. The $F_{\rm PS}$ parameter is then replaced by an F_T parameter that represents the combined flux of the point-source population and the Poisson component. Then, during the likelihood evaluation, a coordinate transform is applied using $F_{\rm PS}=\omega F_T$ and $F_{\rm Poiss}=(1-\omega)F_T$. When considering a population model plus Poisson component, the new coordinate system we advocate for has parameters $\{\omega, N, F_T, \beta_2,\ldots,\beta_{m-1}, \psi_1,\ldots,\psi_m\}$, which may be compared to the equivalent in the standard coordinate system: $\{F_{\rm Poiss}, A, F_{b(2)},,\ldots, F_{b(m)}, n_1,\ldots,n_m\}$.

In the next subsection, we will demonstrate that, within this arguably more natural coordinate system for point-source distributions, priors can be chosen where the degeneracy inherent in the physics is faithfully represented in the posteriors.

5.4. Demonstration of Bias Removal

We consider a simulation scenario identical to that defined in Section 5.2, except we now approach it using priors defined in the natural coordinate system. A unit uniform prior was chosen for ω, and uniform priors are chosen for the ψ_i coordinates. As before, we consider four prior variations, which result from considering combinations of linear or log-flat priors on N and F_T. Specifics are provided in Table 3.

The results are shown in Figure 7, in the same format as those in Figure 6. As the prior on ω is uniform, we observe that the posterior on ω is now generally uniform when the data has no preference for the population or Poisson component.

There is, however, one exception to this behavior that occurs for the combination of a log-uniform prior on N and a uniform prior on F. In that case, the results demonstrate a clear bias in the posterior toward assigning all of the flux to the point-source template. The cause is a combination of effects from both priors. The log-uniform prior on N allows for small values of N. In particular, when $N<1$, the probability that there will be no sources in the image becomes significant. If there are no sources, the flux on the source population cannot be constrained, and any value on F is allowed. In such a case, if the prior on F is also log-uniform, then large values of F are relatively less weighted than they would be with a uniform prior, resulting in this lack of constraint having little effect on the posterior. However, when the prior on F is uniform, large values of F are encouraged, and the posterior assigns significant probability to $N<1$ and $F_{\rm PS}\gg F_P$, as shown by Figure 8. This issue may be avoided in two ways: choose either a uniform prior on N or a log-uniform prior on F, or alternatively, set the lower bound of the prior on N to be larger than one.

Regardless, beyond this specific case, the desired diffuse and point-source degeneracy can be readily be achieved in this coordinate system, and as such, we advocate for its use generally over the existing choices.

Table 3
Prior Ranges for the Demonstration Using the Natural Coordinate System for Specifying the Priors

Parameter	Lower Limit	Upper Limit
N	0.1	10^{20}
F_T	10^{-15} cm^{-2} s^{-1}	10^{-3} cm^{-2} s^{-1}
ψ_1	arctan(2)	arctan(5)
ψ_2	arctan(−3)	arctan(0)
ω	0	1

5.5. Degeneracies between Multiple Components

This section has concentrated on the effect of the prior parameterization on a Poisson and point-source population component with identical spatial distributions. Certainly, one can expect similar problems if two Poisson components have identical spatial distributions, or if the spatial distributions for two point-source population components are identical, and we briefly comment on these scenarios here.

Caution should even be taken even for components that do not share a spatial distribution. If the spatial distributions for two components—Poisson or point-source—are similar to the degree that the distributions cannot be distinguished from each other given the available data, then we can also expect a prior effect to manifest. Even if the spatial distributions for each component in the model are highly distinct, a degeneracy between the distributions can arise if the set of distributions is not linearly independent. This degeneracy allows multiple solutions for the given data, and so the prior effect may also manifest. If the spatial distributions are nearly linearly dependent—as measured by the statistical power to distinguish them—then they should also be considered potentially problematic.

Therefore, when constructing a model, care should be taken to avoid near-linear dependence between the spatial distributions. If the hypothesis in question requires such linear dependence, then the prior parameterization we have introduced provides a solution that ensures that any physical degeneracy is faithfully represented in the posterior for the flux assigned to each source.

6. CPG Performance and Comparison with Existing Methods

In this section, a mathematical connection is drawn between the CPG construction and previous approaches to the problem of parametric point-source inference. In particular, we will consider a number of scenarios that highlight expected problems with the existing methods, and a performance comparison with the CPG is made using simulations. To highlight that these limitations arise from the likelihood construction, the natural coordinate system described in Section 5 is employed for all simulations shown here.

As the NPTF method is the current leading parametric point-source inference method in high-energy astrophysics, this comparison will focus on the essentials of how the NPTF likelihood relates to the CPG construction. For a complete explanation of the NPTF method, we refer the reader to Mishra-Sharma et al. (2017). The NPTF likelihood is specified in Mishra-Sharma et al. (2017) as a generating function. The NPTF generator, $\hat{G}$, is written as an exponential of a power

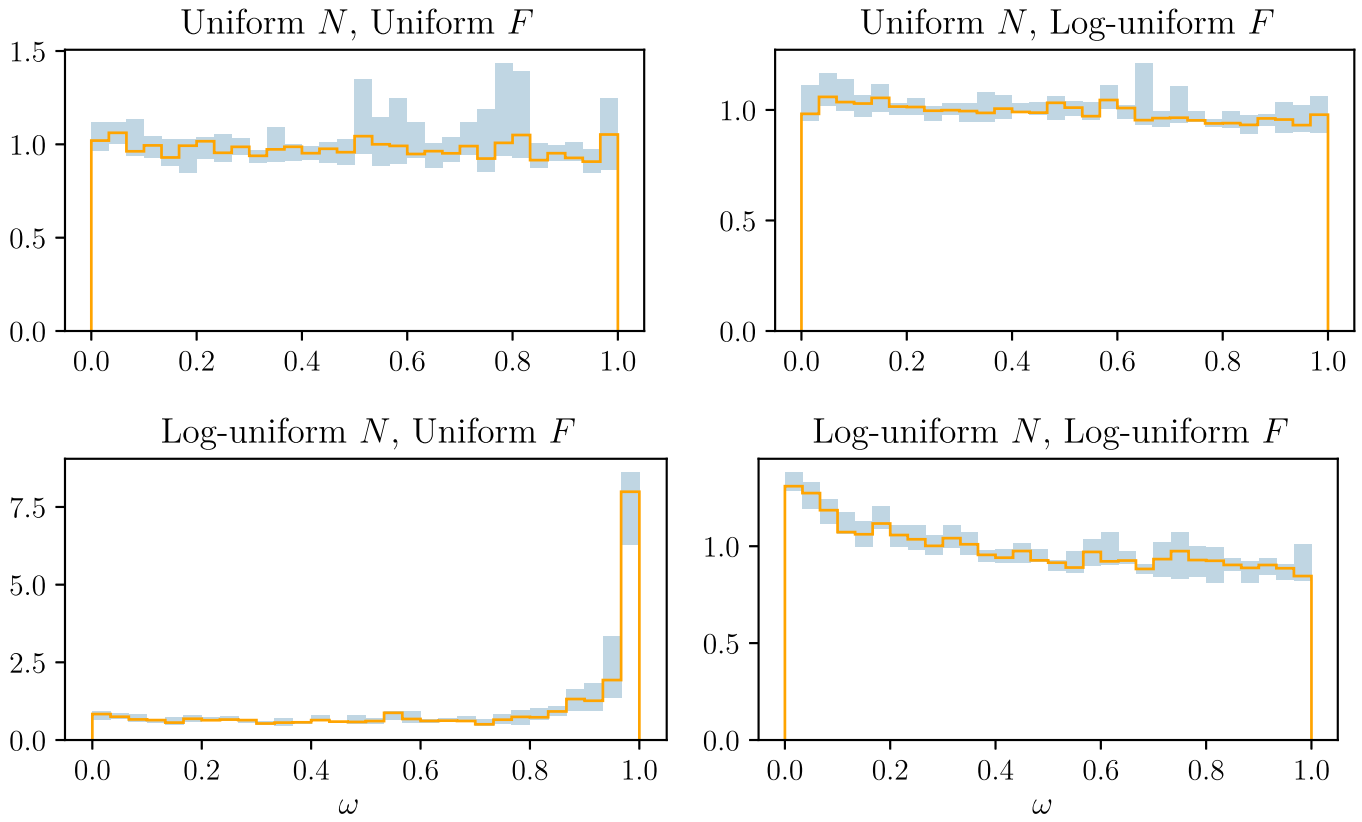

Figure 7. The posterior for the fraction of flux assigned to the point-source population, $\omega = F_{\rm PS}/(F_P + F_{\rm PS})$, under the natural coordinate system. The median posterior (over multiple random trials, for each bin) is shown by the solid line, while the 16% to 84% percentile range is shown by the fill. Ideally, the posterior for ω should be uniform, as the diffuse emission cannot be distinguished from a below-threshold point-source population, and unlike the standard system shown in Figure 6, the new natural coordinate system recovers a uniform posterior—with the exception of the log-uniform N, uniform F priors, as discussed further in the text.

series, but can be equivalently written as

$$\hat{G}_{k_B}(z) = \exp\left[NT_B\left(\int dF \int_0^1 df e^{\bar{\kappa}_B f F(z-1)}\rho(f)p(F) - 1\right)\right] \quad (50)$$

where

$$\bar{\kappa}_B = \frac{1}{|\Omega_B|}\int_{\Omega_B} d\boldsymbol{x}\kappa(\boldsymbol{x}), \quad (51)$$

$$T_B = \int_{\Omega_B} d\boldsymbol{x}\boldsymbol{T}(\boldsymbol{x}), \quad (52)$$

which are the average detector response and template in the bin of interest, respectively.

The NPTF generator in Equation (50) also depends on $\rho(f)$, a function introduced to correct for the PSF. This PSF correction encodes the fraction, f, of point-source flux that migrates out of or into a bin due to the finite angular resolution of the instrument. To build intuition for the effect this correction function aims to address, consider a single point source that is located in a bin, B. Not all of the flux generated with this point source will fall within B: one can find the captured flux by integrating the PSF over the extent of the bin. The expected observed counts from the point source within B is then the product of this flux fraction and the total expected counts from the source. If another point source is in an adjacent bin, then the PSF can cause a fraction of the flux to leak into B. The leakage would cause the flux of the original point source to be overestimated unless accounted for, so the PSF needs to be

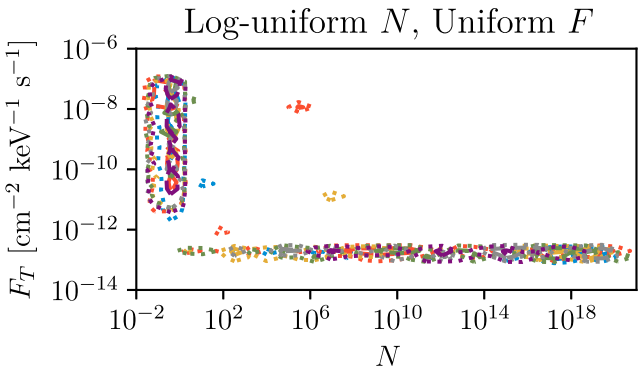

Figure 8. The posterior for the mean number of sources N, and total flux F_T, under the natural coordinate system for the log-uniform N, uniform F prior scenario. The total flux is tightly constrained for $N > 1$; however, it becomes completely unconstrained once the mean number of sources is less than one in the whole population ($N < 1$).

integrated again—now centered somewhere in the adjacent bin—to find the fraction of flux that leaks into B. This fraction is again multiplied by the appropriate counts of this second source—as determined by the dN/dF that is specified for the entire population—so that this extra flux is properly accounted for. The value of $\rho(f)$ is proportional to the frequency of occurrence for a point source contributing the fraction, f, of flux to a bin. The correction function is not a probability, however. The normalization of $\rho(f)$ is fixed so that $\int_0^1 f\rho(f)df = 1$, which ensures both that the flux of the population is conserved and that the number of sources in the population is not overestimated.

The statistical justification for $\rho(f)$ is not given in Malyshev & Hogg (2011), where it was introduced, nor any subsequent works. Instead, $\rho(f)$ is defined as an infinitesimal limit of a numerical estimation for the fraction of flux that a point source will contribute to a pixel after accounting for the PSF. This procedure is constructed as a simulation, and the complete details of the algorithm are given in Appendix F. To provide a brief description, the algorithm places a point source somewhere on the sky, and counts associated with this source are drawn from the PSF and placed in the image bins. All bins are then divided by the total number of simulated counts, so that each bin now contains the fraction of flux that the source contributes to that bin. These flux fractions are then histogrammed, and the final estimate is an average over multiple repetitions of this procedure, so that the histogrammed flux fractions capture multiple possible source locations. The final result is a numerical estimate of $\rho(f)$ that is normalized such that $\int_0^1 f\rho(f)df = 1$.

A direct comparison of the CPG and NPTF generating functions, as given in Equations (35) and (50), reveals that a clear difference lies in the quantities $\bar{\kappa}_B$, T_B, and $\rho(f)$, which only appear for the NPTF. In the NPTF construction, the detector effects are captured in $\bar{\kappa}$ and $\rho(f)$. One might imagine that a connection to $\mu_B(\varepsilon)$ of Equation (35) could be made through $d\varepsilon = \bar{\kappa}_b df$. However, $\rho(f)$ is not specified per bin; instead, it is an average over all bins. Indeed, the CPG construction reveals that the detector effects cannot, in general, be averaged over bins in this way.

As an example, consider data that are binned irregularly (for instance, into bins of significantly differing size). The construction of $\rho(f)$ does not produce a correction function that represents any bin under consideration, and as a result, the likelihood does not describe the statistics of the data. In the case of NuSTAR, the field of view for the instrument is narrow enough that a Euclidian coordinate system can be used as a good approximation to the angular sky coordinates—allowing a regular binning. However, instruments such as Fermi and IceCube measure the entire sky. As a regular tiling of a sphere is impossible, data from these experiments must be irregularly binned, usually with the use of a `HEALPix` map (Górski et al. 2005). As NuSTAR does not suffer from this issue, an in-depth examination of it is outside the scope of this investigation. However, one should not expect this to have a particularly large effect on the Fermi GCE analysis or most of the results from the recent IceCube analysis. Both analyses generally prioritize the region of the sky around the Galactic Center, using templates that place the most weight in this region. The `HEALPix` maps employed were centered on the galactic coordinate system, placing the Galactic Center at the center of the `HEALPix` maps, where the maps have the most regular tiling.[18] Thus, for both analyses, the tiling is close to regular where the templates under consideration have the greatest weight. (One IceCube analysis considered a uniform all-sky template, for which the irregular `HEALPix` binning could produce issues.) In the case of the IceCube analysis, the construction of the $\rho(f)$ was further weighted by the spatial template under consideration, ensuring it more closely reflected the binning in this region. It should be noted that the irregularity of the HEALPix binning is reduced—both in terms of average bin shape and average number of neighbors—for large numbers of bins. However, the potential for the regularity to be an issue must still be considered—especially for analyses where important information is present in the poles, or where bin sizes are large.

Extending NPTF to use a per-bin PSF correction, $\rho_B(f)$, would address this problem, but limitations in such an NPTF-like method remain. The primary limitation of the NPTF construction is the use of an integrated value for the spatial distribution—the template. The CPG construction shows that the spatial distribution and detector response cannot, in general, be factorized into two separate terms. We will demonstrate this explicitly by considering a number of examples in the following subsections.

6.1. Nonuniform Spatial Distribution

To begin with, we consider an edge case that illustrates the problems that arise from the factorization of the template from the detector effects. Consider a spatial population that occupies only a single bin, $l = 0$, such that $T_l = \delta_{l0}$. In addition, suppose the PSF is broad enough that a non-negligible amount of flux migrates out of bin 0. As the population does not exist outside of bin 0, out-migration is the only effect of the PSF for this bin. This implies that flux is not conserved for this bin: the flux captured by this bin is always less than the flux of the population. The other bins, in turn, are only affected by in-migration. According to the spatial template, $T_l = 0$ for these bins, and so the population fluxes for these bins are also equal to zero. Thus, flux is not conserved for these bins either, as they receive flux from adjacent bins. Both of these effects are illustrated in Figure 9.

Conservation of flux is only a property of the entire image, and cannot be enforced on a bin-by-bin basis through $\rho(f)$. The distribution of flux fractions that these bins receive is clearly a function of the bin in question, and so the average over all bins, $\rho(f)$, will not describe any bin in the image. In addition, the distribution of flux fractions for each bin clearly depends on the distance of that bin from bin 0, demonstrating manifestly why the spatial distribution of the population cannot be factorized out of the detector response.

Although this edge case is artificially extreme—in order to bring out the effect—it is not uncommon for scenarios to have much of the spatial distribution concentrated in a single bin. Any sharply peaked spatial distribution will suffer this problem to an extent; for example, the GCE spatial profile has a sharp peak at the Galactic Center.

A scenario was constructed in the NuSTAR simulation (described in detail in Appendix A.1) in order to investigate the potential bias this may induce. In this scenario, five bins have equally nonzero share of the population. Multiple bins are necessary for point-source population reconstruction, and the bins are spaced far enough apart that the demonstrated effect will be largely similar to the single-bin scenario. The edge case as described above can only be meaningfully analyzed with the CPG likelihood. The NPTF method will fail entirely and result in zero probability for all parameter values. The reason for this is that, as it can be seen from Equation (50), the NPTF predicts zero flux in any bin with $T_B = 0$, for any value of the model parameters. If any of these bins record counts due to the finite PSF, then these are events the NPTF simply cannot reconstruct. In order to allow for a comparison with the NPTF, we rescale the template so that no bins are exactly zero, but rather have

[18] The `HEALPix` tiling is most irregular near the poles of the maps, and has greater regularity near the equator, which contains the Galactic Center when galactic coordinates are mapped onto the `HEALPix` coordinate system.

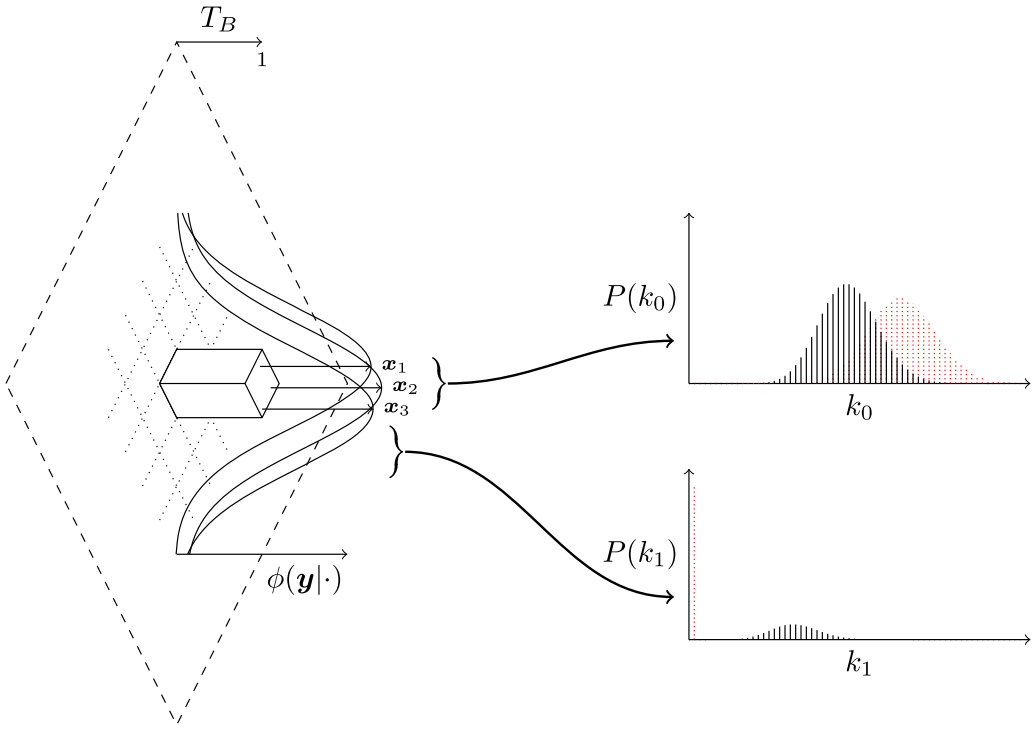

Figure 9. Left: A template has all of its probability concentrated in a single bin, shown as the 2D histogram on the left. As such, all three sources, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$, are located within this bin. Each has an associated PSF shown as the solid distribution centered on the source location. Top Right: The probability distribution for the number of counts in the central bin. The distribution for conserved flux, as in the NPTF construction, is shown in dotted red. The actual distribution is shown in black, with a smaller mean. Bottom Right: The distribution for an adjacent bin. The red dotted line shows the distribution for the NPTF construction with a zero mean. The black shows the actual distribution, which has a nonzero mean.

small nonzero values of T_B that cumulatively add up to no more than 1% of the template. Further details on the scenario setup are given in Table 1.

The NPTF method was implemented by estimating $\rho(f)$ according to the algorithm in Appendix F, and then transforming it to an equivalent CPG representation through $\rho_B(\varepsilon) = T_B\rho(\varepsilon/\bar{\kappa}_B)/\bar{\kappa}_B$. In this way, the exact same computational implementation of the likelihood and Bayesian analysis procedure can be used to evaluate both the NPTF and CPG methods. Thus, any differences observed can be entirely attributed to the NPTF $\rho(f)$ construction, and are not due to any subtle differences in the software implementation of the likelihood and MCMC. The generation of $\mu(\varepsilon)$ required approximately 100 CPU hours;[19] however, in practice, excellent results can be achieved with significantly less time —as little as five CPU hours. As $\rho(f)$ is shared between bins, it requires substantially less time to generate: in this case, less than one CPU hour total. These times are highly application-dependent, as they are almost an entirely a function of the computational complexity of the detector simulation, and in the case of $\mu(\varepsilon)$, the chosen binning scheme.

The natural coordinate system employed in the prior effect demonstrations of Section 5.4 was used, with the exception of the Poisson component. Thus, the priors used here are also described by Table 3, other than for ω, which is not needed, and the flux parameter is referred to as $F_{\rm PS}$ instead of F_T. The posteriors were sampled using the `emcee` Affine Invariant MCMC (Foreman-Mackey et al. 2013). A total of 200,000 steps were taken with 64 walkers, and the first 90% of samples were discarded as burn-in.

The results are shown in Figure 10(a). The true source-count function parameters are shown by a solid red line. The dotted and dashed lines describe the recovered posterior, where each color represents one trial image. The dashed lines are the 1σ Highest Posterior Density (HPD) regions, while the dotted lines define the 2σ HPD regions. The NPTF posterior is biased toward a high number of sources with an approximately correct total flux. In comparison, CPG does not exhibit the same bias. Here, bias is defined as an overall shift in the recovered parameters across multiple trials. We can see that the yellow CPG trial is well outside the true parameters, while the purple NPTF trial is inside the true parameters. This is to be expected due to the random variations between trials. However, as a group, the NPTF trials are clearly recovering a number of sources greater than the true N. In Figure 10(b), we further show the difference between the CPG $\mu_B(\varepsilon)$ and the equivalent NPTF $\rho_B(\varepsilon)$. Each color represents the $\mu_B(\varepsilon)$ measure for a bin in the scenario. As NPTF averages the PSF correction over bins and the effective area is isotropic, the equivalent $\rho_B(\varepsilon)$ is the same for all bins up to the two unique template values of T_B.

The averaging approximation used in the NPTF construction is clearly deficient in this edge case. The inference is driven by the five bins that contain the population. The $\mu_B(\varepsilon)$ for these bins is heavily weighted toward high ε. The NPTF $\rho(\varepsilon)$ is heavily weighted toward low ε, due to averaging over the larger number of bins that do not contain the population. The normalization condition, $\int_0^1 df\, f\rho(f) = 1$, imposed on $\rho(f)$ modifies the overall normalization of the function so that flux is conserved in all bins. However, as described above, the flux is manifestly not conserved on a bin-by-bin basis in this scenario.

[19] Under 2 hr wall-clock time using an AMD Ryzen Threadripper 3990X 64-core processor.

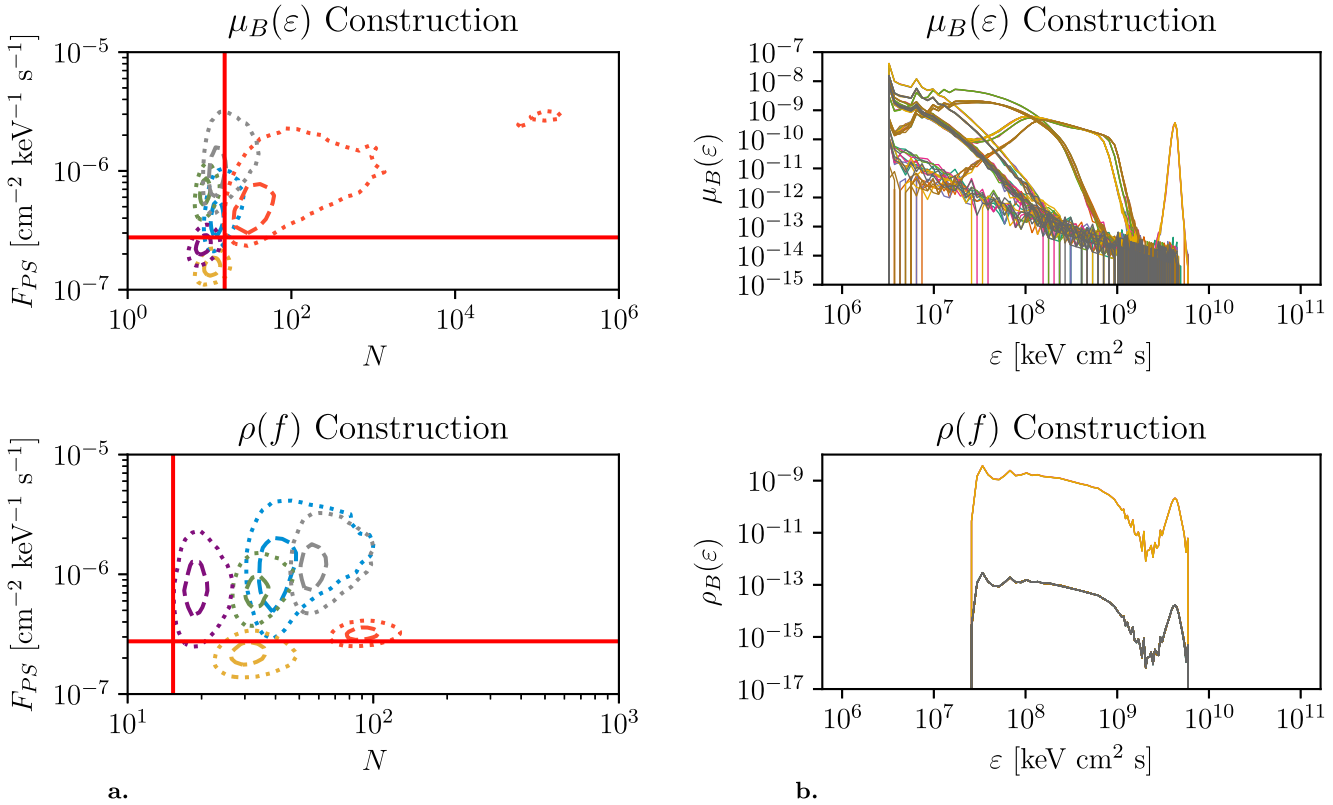

Figure 10. The results of applying the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction to the nonuniform spatial distribution scenario. The CPG posteriors are clustered around the injected population, while the NPTF posteriors show a bias toward higher N. The averaging procedure in constructing $\rho(f)$ results in a detector correction that is not representative of any of the bins in the image, as shown by the $\mu_B(\varepsilon)$ functions. Left: The recovered posterior for both the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction. The solid red line shows the (true) injected source-count function. Each of the colored dashed and dotted lines are the 1σ and 2σ HPD regions, respectively, for separate trials. Right: The detector effect correction function for both the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction (converted to $\rho_B(\varepsilon)$ for comparison as described in the text). Here, each colored line corresponds to the function for a bin in the image.

In Equation (50), a change in the normalization of $\rho(f)$ is equivalent to a change in N, as this generator can also be written as

$$\hat{G}_{k_B}(z) = \exp\left[NT_B\int_0^1 df\rho(f)\left(\int dF e^{F_{k_B}fF(z-1)}p(F) - f\right)\right]. \quad (53)$$

Thus, the normalization effectively drives $\rho_B(\varepsilon)$ at high ε down, causing the posterior on N to be driven up to compensate. This leads to the observed overestimation bias in the number of sources.

The posteriors in Figure 10(a) show a between-trial variation that is on the same scale as the posteriors themselves. We can rule out statistical fluctuations as the dominant cause, as the variations are an order of magnitude in N for the NPTF posteriors. Instead, these variations are due to the small amount of information that is present in only five bins, which leads to poor recovery of the total number of sources. We should expect the posteriors to overlap (and they mostly do for CPG), but as the NPTF likelihood does not describe the data, we cannot expect the posteriors to be well-behaved.

6.2. Anisotropic PSF

In the standard NPTF construction, $\rho(f)$ is common to all bins, and thus it can, at best, capture the average PSF throughout the image. The following scenario examines the bias that this approximation may introduce to the inference of the source-count function model parameters when an anisotropic PSF is present—as arises, for instance, in NuSTAR.

We consider a scenario largely identical to that in Section 5.2, with details given in Table 1. In this scenario, multiple exposures are overlaid into a mosaic, forming a single binned image. This increases the amount of anisotropy in the PSF, as sources can bleed across the boundaries of the exposures. When a source contributes to two or more exposures, an appropriate mix of PSFs is required, leading to more complicated PSF distributions. For this scenario, we wish to only test the effect of an anisotropic PSF. Thus, the exposures are perfectly aligned edge-to-edge in a grid with no gaps or overlaps. This, along with the lack of vignetting, ensures that the effective area is uniform across the entire image. We note that this is not a natural arrangement—most mosaics involve a degree of overlap between exposures.

The results of analyzing the resulting data sets are shown in Figure 11(a). A high degree of bias toward a low number of sources is observed in the NPTF posterior, while CPG is largely consistent with the true population parameters. Figure 11(b) further shows that the equivalent NPTF $\rho_B(\varepsilon)$ is much closer to the CPG $\mu_B(\varepsilon)$ in this scenario as compared to the previous nonuniform spatial template scenario. However, the shape of $\rho_B(\varepsilon)$ is significantly different from $\mu_B(\varepsilon)$ near the highest ε. Unlike before, the normalization of $\rho_B(\varepsilon)$ will be correct here: the uniform spatial distribution ensures that flux is conserved on a bin-by-bin basis. While this normalization ensures that the average ε is correct, the distribution of ε is not.

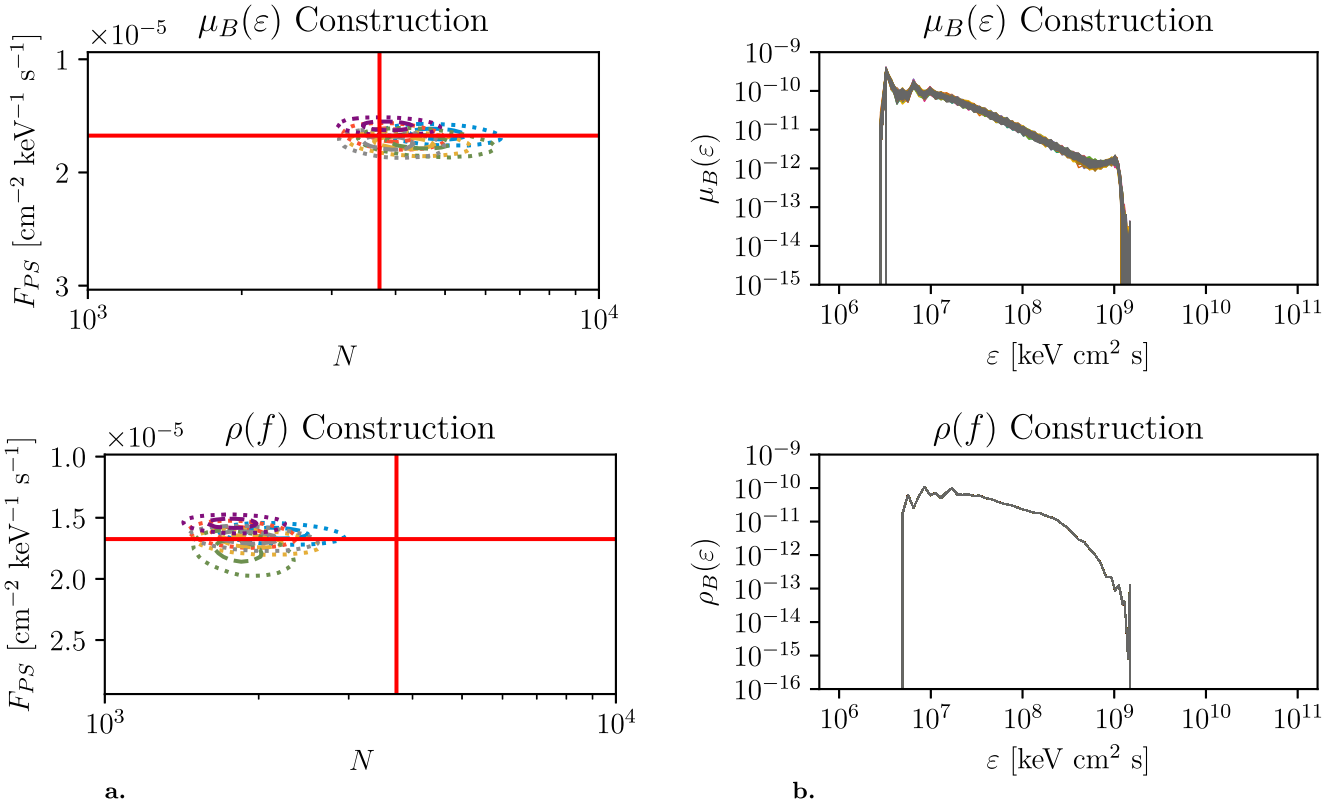

Figure 11. As in Figure 10, but for the anisotropic PSF scenario. The injected population is well recovered by CPG, while the NPTF posteriors are clearly biased toward low N. The cusp structure visible in each $\mu_B(\varepsilon)$ is smeared out in $\rho_B(\varepsilon)$. Left: the recovered posterior for both the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction. Right: the detector effect correction function for both the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction.

Observe that $\mu_B(\varepsilon)$ generally has a cusp at high ε, while $\rho_B(\varepsilon)$ shows no cusp. This lack of cusp causes the NPTF likelihood to be driven by lower ε in comparison, and so the posterior drives up the average flux per source and drives down the number of sources in order to maintain the conservation of flux. Note that the quantity shown in Figure 11(a) is the total flux of the entire population, so when the number of sources, N, reduces, it is the average flux per source, roughly proportional to $F_{\rm PS}/N$, that increases.

6.3. Sub-bin Effective Area

The prescribed construction of $\rho(f)$ in the NPTF method does not include any accounting for the effective area, exposure time, or detector efficiency of the instrument and observation. This may lead to an incorrect construction of the likelihood distribution, as the following scenario illustrates.

Consider an instrument where the effective area varies sharply within a bin, and a population that is spatially uniform. Thus, the distribution of fluxes for a source is not a function of position within this bin. However, the distribution of expected counts will be a function of position within the bin, as the flux must be converted to a number of photons using the effective area. In particular, take a scenario where there is no PSF, $\phi(\boldsymbol{y}|\boldsymbol{x}) = \delta(\boldsymbol{y} - \boldsymbol{x})$; and the injected source population has no variation in flux, $p(F) = \delta(F - \bar{F}_{\rm PS})$. In order to generate the effect of interest, we then let the detector response change discontinuously at some point within the bin, between the values of κ_0 and κ_1. If a source is located where $\kappa(\boldsymbol{x}) = \kappa_i$, then the distribution of the detected number of counts at this location is $p(S_B|\boldsymbol{x}) = \delta(S_B - \kappa_i \bar{F}_{\rm PS})$, with $i = 1$ or 2. The distribution for the whole bin is found by integrating the position-dependent distribution over $T(\boldsymbol{x})$. Let the spatial template be uniform. Then the whole bin distribution is a mixture of the previous two distributions:

$$p(S_B) = C\delta(S_B - \kappa_0 \bar{F}_{\rm PS}) + (1 - C)\delta(S_B - \kappa_1 \bar{F}_{\rm PS}), \quad (54)$$

where C is some mixture fraction that depends on what area of the bin has a detector response of κ_0.

For this scenario, the NPTF calculates the average detector response across the bin, $\bar{\kappa}$, and uses this to find $p(S_B) = \delta(S_B - \bar{\kappa}\bar{F}_{\rm PS})$. This will not result in the correct distribution of photon number; for example, if $\kappa_0 = 0$, then the NPTF construction preserves the total flux of the bin, but will overestimate the flux of sources within that bin, as $\bar{\kappa} < \kappa_1$. A comparison between the exact and mean distributions is shown in Figure 12. It should be noted here that, due to computational constraints, the effective area is rarely considered on a bin-by-bin basis in the NPTF construction. Instead, it is averaged over larger "exposure regions," as the computational complexity of the NPTF power-series calculation depends on the number of bins. Although these noncontiguous regions are chosen by the similarity of the effective area of each bin within the regions, this necessarily reduces the accuracy of the likelihood evaluation further. In contrast, CPG requires no exposure regions, as the computational complexity of the CPG power series calculation does not depend on the number of bins and is computed only once for each likelihood evaluation. As exposure regions only enter into NPTF due to computational constraints, they are not, by themselves, an intrinsic limitation of NPTF; therefore, we do not consider the effect of taking a number of exposure regions smaller than the number of pixels in this investigation.

To test this scenario, the simulated vignetting was modified into a checkerboard pattern, taking values of either κ_0 or κ_1. Each square of the checkerboard was defined to be equal to the bin size, but with an offset so that the checkerboard boundaries lie within the bins of the image. This causes the effective area to change discontinuously across most bins of the image. The remaining scenario parameters are largely identical to those in Section 5.2, with details given in Table 1.

The results are shown in Figure 13(a). Here, NPTF fares significantly better when compared to the previous scenarios. However, a recognizable bias is still present in the recovered posteriors. CPG, on the other hand, is considerably closer to the true population parameters. The detector effect correction functions are shown in Figure 13(b). Clearly, the CPG $\mu_B(\varepsilon)$ has the two expected modes corresponding to the sub-bin effective area variation. In contrast, the equivalent NPTF $\rho_B(\varepsilon)$ has a single mode corresponding to the average effective area. Thus, the likelihood is driven by a lower ε, and so the average flux per source is driven up while the number of sources is driven down to compensate.

The CPG posteriors in Figure 13(a) display a long tail toward high N. This same behavior can be observed to a lesser extent in Figure 11(a) (anisotropic PSF scenario). The summary in Figure 1 shows this effect most clearly. As discussed in Section 5, the discrimination ability of the CPG likelihood is reduced at high N as the population model approaches the diffuse limit. In contrast, population models with N less than the true value will be highly disfavored, as they produce sources with high flux. In these scenarios, the prior on N is log-uniform and does not discourage models with high N. Therefore, an asymmetry in the posterior toward high N is expected. This effect is also present in Figure 10(a) (nonuniform template scenario), but the small number of bins with useful data causes a wide variation in the posteriors that largely hides the effect.

6.4. A Scenario where ρ(f) Is Valid

The three previous scenarios demonstrate the biases introduced into NPTF by the failure to correctly model the effective area, PSF, and the spatial distribution of sources, as well as how the comprehensive approach of CPG resolves these issues. When these complications are not present, however, it is expected that the NPTF construction will be equivalent to CPG. To confirm this, we consider a scenario where the population is spatially uniform, the PSF is isotropic, and the effective area is constant throughout the image. The details are given in Table 1, and the results are shown in Figure 14. Both NPTF and CPG display no bias, and each trial is equivalent—up to slight variations—between the two methods. This shows that, under these conditions, the more general CPG reduces to NPTF, and NPTF does indeed work correctly.

6.5. Realistic Scenario

Finally, we turn to a realistic scenario, where all NuSTAR detector effects are enabled and a Poisson background is injected. This results in a realistic simulation of a NuSTAR observation. Eight observations are combined into a mosaic that exhibits the overlapping diamond pattern that is common to composite NuSTAR images—see Hong et al. (2016) for one such instance.

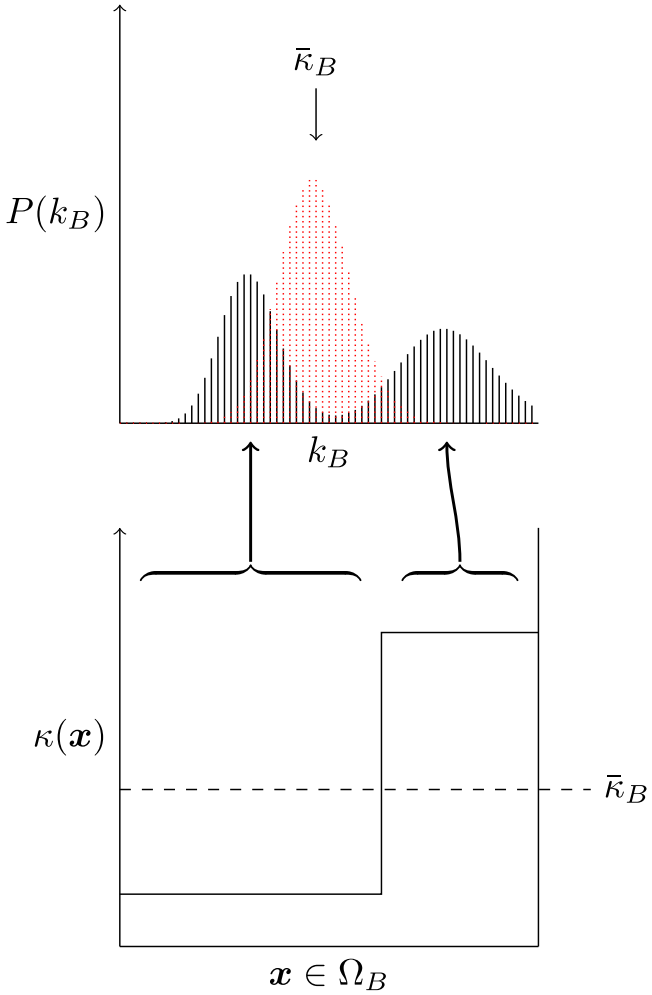

Figure 12. Bottom: The detector response, $\kappa(\boldsymbol{x})$, varies sharply over bin b. The average detector response for this bin, $\bar{\kappa}_b$, is shown by the dashed line. Top: If the average detector response is used to convert flux to counts, the probability distribution follows the red dotted comb; however, the within-bin variation of the detector response must create two modes in the distribution, shown by the black comb.

For this scenario, there are two Poisson background components: the intrinsic detector background, which is uniform within each observation, and the the Cosmic X-ray Background (CXB), which is a flux that is multiplied by the effective area of the instrument. Both components scale with the exposure time, and their respective rates and fluxes are detailed in Table 1. The background rate was chosen to be representative of an observation with NuSTAR. The CXB flux was derived from Krivonos et al. (2021) by integrating over the 3–10 keV energy window.

The CXB is incorporated into the model using a shared flux parameter in the natural coordinate system discussed in Section 5. The intrinsic background, in contrast, is not specified as part of the natural coordinate system. The natural coordinate system divides a total amount of flux between multiple models, but this background component is not described in terms of a flux, as it is not astrophysical in origin. Thus, it was given a separate uniform prior around the known background rate, and is parameterized by $\lambda_{\rm bkg}$, a count rate for the entire detector, which is multiplied by the exposure time and divided by the relative bin size to get a mean number of counts per bin. The injected source-count function was selected to be similar to the

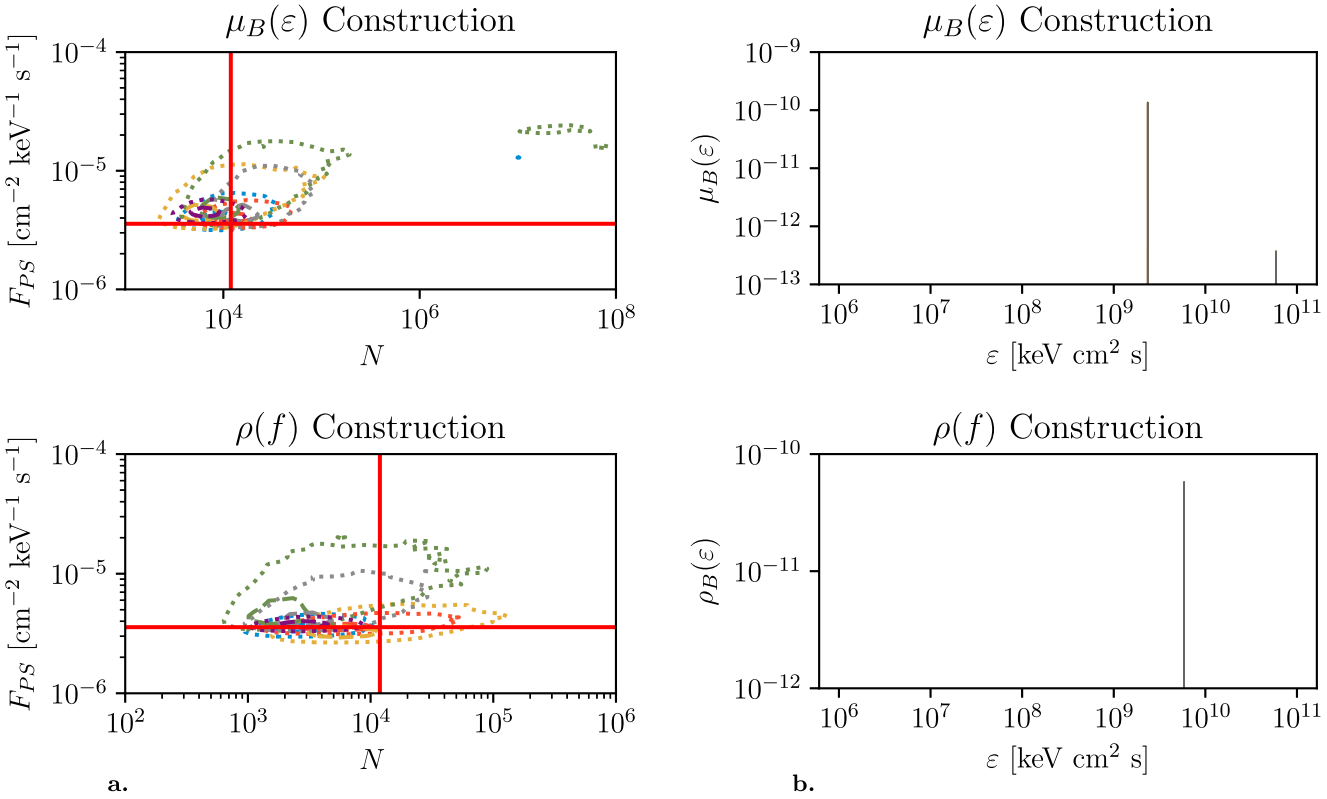

Figure 13. As in Figure 10, but for the variation in sub-bin effective area scenario. The CPG posteriors are clustered around the injected population, while the NPTF posteriors exhibit a small bias toward low N. The bimodal effective area is shown clearly by $\mu_B(\varepsilon)$, while for $\rho_B(\varepsilon)$ this structure is removed by the averaging the effective area, resulting in a single mode. Left: the recovered posterior for both the CPG $\mu(\varepsilon)$ construction and the NPTF $\rho(f)$ construction. Right: the detector effect correction function for both the CPG $\mu(\varepsilon)$ construction and the NPTF $\rho(f)$ construction.

observed population of X-ray sources near the Galactic Center (Hong et al. 2016). Further details are given in Table 1, and the prior ranges are shown in Table 4.

The results—derived with both the CPG $\mu_B(\varepsilon)$ and NPTF $\rho(f)$ constructions—are shown in the upper half of Figure 15 as a source-count density function. Each color represents one of six trial images, and the dashed lines show the 16% and 84% percentiles for the source-count function recovered from the posterior. The solid red line shows the true population source-count function. The top of the gray fill is the mean of the 84% percentiles for each trial, while the bottom of the gray fill is the mean of the 16% percentiles.

For the CPG construction, there is a large uncertainty on the location of the flux break itself, and below the break, there is an underestimation in the number of sources. This is unsurprising, as the uniform prior on the flux fraction parameter between the point-source and CXB component, ω, gives weight to models where a significant amount of flux is carried by the CXB, which corresponds to the below-threshold point-source flux being assigned to the CXB component.

This effect can be reduced by placing a more informative prior on ω, but this is not as easy at it may first appear. Even if the CXB flux is well-known, placing a prior on ω requires prior knowledge of the relative total flux between the population and the CXB. If the prior is informed only by an estimate of this relative flux, then an incorrect value could exacerbate the effect. Alternatively, the prior on ω could be conditioned on the total flux F_T such that the inferred total flux informs the prior on the estimated flux of the point-source population.

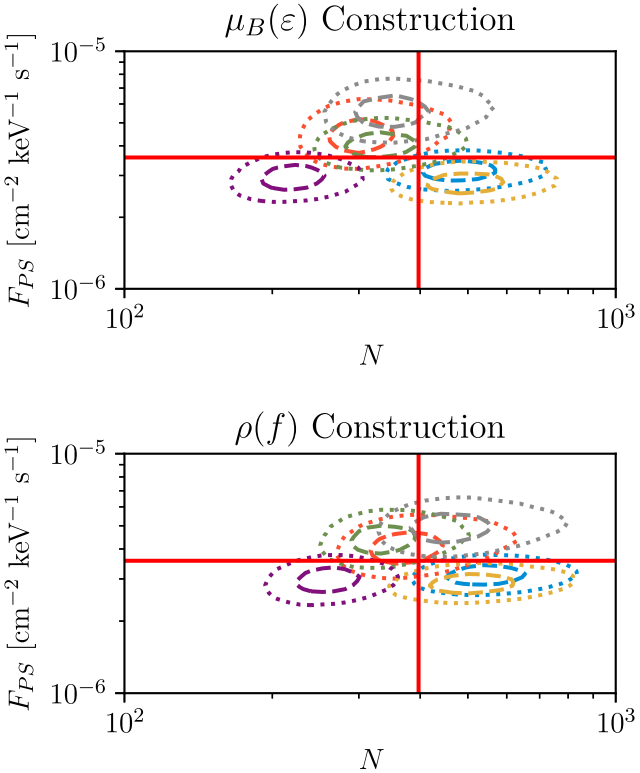

Figure 14. The recovered source-count function for both the CPG $\mu_B(\varepsilon)$ construction and the NPTF $\rho(f)$ construction in the $\rho(f)$ scenario, described in the text. When all conditions for the construction of $\rho(f)$ are met, both CPG and NPTF produce posteriors clustered around the injected population.

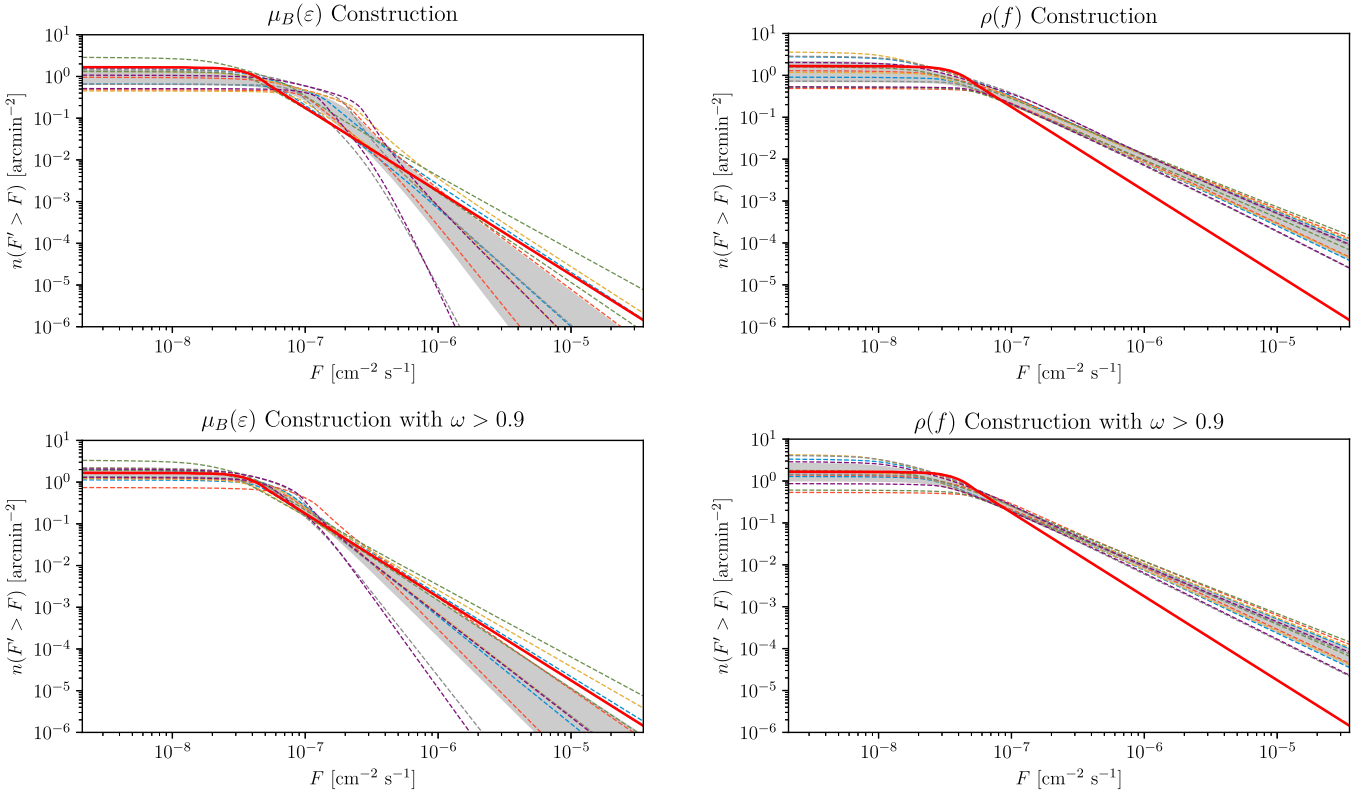

Figure 15. The recovered source-count function (as defined in Equation (9)) for the CPG $\mu_B(\varepsilon)$ construction (top left) and the NPTF $\rho(f)$ construction (top right) in the realistic NuSTAR scenario. Each color represents a different random trial, with the 16% and 84% percentiles shown as dashed lines derived from the posterior. The true source-count function is shown in solid red. The gray fill shows the average percentile band, as a guide. The $\rho(f)$ construction clearly recovers the incorrect source-count function, as the slope of the power-law is well outside all percentile bands. In comparison, the CPG construction is generally more accurate. The underestimation of sources in both methods is further discussed in the text. The same results, with the posterior conditioned to $\omega > 0.9$, are shown in the bottom left and bottom right for CPG and NPTF, respectively. With this conditioning, the recovery of the total number of sources is drastically improved for the CPG.

Table 4
Prior Ranges for the Realistic Scenario

Parameter	Lower Limit	Upper Limit
N	0.1	10^8
F_T	10^{-15} cm^{-2} s^{-1}	10^{-3} cm^{-2} s^{-1}
ψ_1	arctan(2)	arctan(10)
ψ_2	arctan(-5)	arctan(0)
ω	0	1
$\lambda_{\rm bkg}$	6.984×10^{-2} s^{-1}	10.476×10^{-2} s^{-1}

We can also imagine a similar problem occurring for the intrinsic detector background, in that below-threshold sources could be confused with the intrinsic background model. The solution here, which has been employed in this scenario, is to use a tight prior on the intrinsic detector background based on previous measurements. This is simple to implement, as the intrinsic detector background is not specified in the natural coordinate system—and thus has an independent prior.

In the end, this is not an issue with the CPG construction, it is simply the nature of Bayesian analyses. We can see this more clearly by conditioning the posterior on $\omega > 0.9$. This, in effect, largely removes the CXB from the model. The result is shown in the lower half of Figure 15. The uncertainty on the flux break is greatly reduced, and the total number of sources is now accurately recovered to within the percentiles.

However, the power index, n_1, sits just outside the percentiles. This turns out to be a generally unavoidable effect. The large bin sizes cause the brightest sources to be washed out by the large number of dimmer sources. For this reason, the model cannot distinguish between a few bright sources and many dim ones, and so the posterior is extended to larger n_1, thereby pushing the true n_1 outside of the percentile interval. This effect can be reduced by reducing the bin size, so that bright sources make up a greater fraction of the total flux in each bin. Unfortunately, as discussed in Section 4.5, smaller bins will cause the statistics to be overcounted, making the posteriors artificially small.

Thus, when applying CPG to a realistic scenario with a Poisson background that has a poorly informed prior, care must be taken in interpreting the number of very low-flux sources. In addition, the prior on the detector background should be chosen to incorporate as much information as possible that is known about the background rate. Finally, one must take care in the selection of the bin size. When the goal is to recover the population parameters, smaller bins should be chosen in order to retain more information on the high-flux sources. Any reported results should then caution of the between-bin correlation effect elaborated on in Section 4.5. If model selection is the aim and marginal likelihoods are in use, it may be better to use larger bins, as a high degree of overcounting of statistics may result in an erroneous marginal likelihood.

In comparison, the NPTF $\rho(f)$ construction has poorer performance. The same underestimation of the number of sources below the break is observed, but to a larger degree. In addition, the location of the break and index of the power law above the break is incorrectly estimated, and the true source-

count function lies well outside all of the recovered source-count functions in the region above the break. Conditioning the posterior on $\omega > 0.9$ does not improve the situation appreciably: the total number of sources now lies within the percentiles (although the uncertainty is much larger than when using CPG), but the region above the break is still incorrectly recovered. In this instance, while the total number of sources is approximately correct, the total flux of the population is overestimated.

7. Conclusion

In this paper, we have introduced a new approach to the age-old problem of point-source population inference. In particular, we constructed a new likelihood using CPG functionals. To complement this likelihood, a new natural coordinate system has been developed for parameterizing Bayesian priors on point-source population models where a Poisson background component is also present. In particular, we revealed that existing prior choices often result in posteriors that assign all flux to either the point-source population or the Poisson background model. The new natural coordinate system produces posterior distributions that correctly capture the uncertainty on the population model in this circumstance. Combined with this new prior formulation, the CPG likelihood has been shown to correctly handle a series of test scenarios that each exhibit a particular edge case that is common in astronomical instrumentation. For these scenarios, a comparison was made to non-Poissonian template fitting (NPTF)—a current leading parametric point-source inference method. In contrast to CPG, the NPTF method has been shown to produce biases in the recovered posterior distributions, specifically the mean number of sources. These biases are attributed to limitations in the construction of the NPTF likelihood, which factorizes the spatial distribution, PSF, and effective area. The CPG derivation shows how these contributions to the likelihood cannot be factorized in general. An implementation of the CPG is made publicly available here.[20]

Our focus in the present work has been on the construction of the CPG and on demonstrating the improvements it brings. However, given the general importance of point-source studies, applying this new method to actual data is an important open direction, and may even be able to finally shed light on open questions such as the nature of the GCE.

We thank M. Lisanti, F. List, S. Mishra-Sharma, B. M. Roach, B. Safdi, and T. Slatyer for useful discussions, as well as the wider community in the NSF AI Institute for Artificial Intelligence and Fundamental Interactions, the Laboratory for Information and Decision Systems, and the MIT Statistics and Data Science Center. N.L.R. is supported by the Miller Institute for Basic Research in Science at the University of California, Berkeley. K.P. and G.H.C. were supported by the Cottrell Scholar Award, Research Corporation for Science Advancement (RCSA), ID 25928. This work made use of resources provided by the National Energy Research Scientific Computing Center, a U.S. Department of Energy Office of Science User Facility supported by Contract No. DE-AC02-05CH11231.

[20] https://github.com/ghcollin/cpg_likelihood

Appendix A
NuSTAR Observatory and Simulation Procedure

The NuSTAR observatory comprises of two adjacent aligned telescopes. Each telescope consists of a grazing incidence X-ray optics module and a focal plane module (FPM) that houses four pixelated CdZnTe semiconductor detectors. For both telescopes, these elements are separated by a single shared extended mast. For this investigation, only one of these telescopes will be considered, in order to simplify the simulation.

The optics focus X-rays via grazing reflections from two sets of concentric shells arranged in an approximate Wolter-I geometry. When an X-ray is reflected by both shells, it will be correctly focused onto the detector plane. An X-ray may only undergo one reflection in the optics. These X-rays are poorly focused and are called ghost rays. The observatory does not have an enclosed barrel, and the majority of the length of the telescope is open to space. A series of aperture stops collimate incoming X-rays, preventing most X-rays from entering the FPM through the sides of the telescope. However, a small window remains, allowing X-rays that pass close to the optics modules to strike the detector without passing through the optics. Known as stray light, X-rays that enter the FPM this way are unfocused.

The four detectors in each FPM are arranged in a 2×2 grid. Each detector is composed of 32×32 pixels, and subpixel positional information is available. The gap between the detectors is small, and is ignored in this study. Further details on the observatory can be found in Harrison et al. (2013), Wik et al. (2014), and Madsen et al. (2015, 2017).

A.1. Simulation

Use of NuSTAR as a test case requires a simulation of the detector in order to create observations where the parameters of the injected point-source population are known exactly. The simulation draws a number of point sources from a given differential source-count function using a Poisson distribution with mean N. The flux of each source is also drawn from the $p(F)$, and the location of each source is drawn from a given spatial distribution specified by the template. The simulation constructs an image according to the NuSTAR field of view, and with an adjustable bin size. The detector response is extracted from the NuSTAR CALDB (Harrison et al. 2013) FPM data using the vignetting-corrected effective area. The energy spectrum is assumed to take the power-law form given in Equation (4) with $E_0 = 1\,\mathrm{keV}$ and $\gamma = 1.5$, whereas the energy range and exposure time are scenario-dependent.

Once the mean counts for a given source, S, has been determined, a number of photons are then drawn from a Poisson distribution with mean S, and these photons are placed into the image according to locations drawn from the CALDB PSF. This PSF is generated by a physics-based simulation of the X-ray optics. If a photon lies outside the field of view, the photon is discarded. The NuSTAR PSF is nonisotropic, and varies according to the distance of the source from the optical axis of the telescope. At the optical axis, the PSF is radially symmetric. As the source moves toward the edge of the image, it is distorted radially, creating a fish-eye effect.

Apart from the point-source population, multiple other effects can cause the real or apparent detection of photons. Astrophysical backgrounds and foregrounds may include

extended sources such as gas clouds, or very bright point sources that are not part of the population such as magnetars. These can be modeled as Poisson distributions with appropriate flux maps. As they are highly scenario-dependent, these effects are not modeled in this investigation. Ghost rays and stray light are usually only visible when bright sources are in or around the FOV; assuming the sources can be located, their contribution can be modeled by a Poisson distribution with an appropriate flux map. The effect of ghost rays from population sources is incorporated into the PSF by construction. The CXB is a significant source of stray-light-induced background. This stray-light contribution can be modeled much like other astrophysical backgrounds—with a Poisson distribution using an appropriate CXB flux. As with the other sources of astrophysical background, the effect of stray-light CXB is not considered in this investigation. NuSTAR also suffers from a detector background caused by spaceborne radiation. This background enters through the FPM shielding, and is not associated with the X-ray optics. Thus, the background is uniformly distributed across each detector, and varies across each of the detectors in the FPM. It can be simulated by drawing counts from a Poisson distribution according to an appropriate background count rate.

Position information of photons beyond the detector pixelization is not available; however, in this simulation, the effect of pixelation is not considered and the photons are directly processed into bins. The size of these bins is scenario-dependent, but they are generally much larger than the detector pixels, thus the effect of pixelation is considered small and the additional complications involved in the pixel subdivisions are avoided.

If the scenario calls for multiple observations, then the above procedure is repeated for each observation minus the final binning. This creates a list of photons from all observations, which are then transformed into a shared coordinate system for all of the observations and then binned into a shared binning scheme. An example of this is shown in Figure 3, with a binning scheme chosen to be approximately the same as the NuSTAR subpixelation, in order to show the PSF.

Appendix B
Generating Functions and Functionals

In this appendix, a full derivation of the CPG generating function without approximations will be presented. From this generating function, an unbinned likelihood and a binned likelihood can be derived, both taking into account correlations between pixels. Unfortunately, these likelihoods are intractable, and so the derivation in Section 4 is shown to be a tractable mean-field approximation for the likelihoods here.

Generating functions are a useful representation for a discrete probability distribution when complex constructions are required. In this appendix, we provide a more expansive discussion of these objects than appeared in the main body. The generating function, $G(z)$, is defined as the z-transform of a discrete distribution, $P(n)$, over its support $\mathcal{N}$:

$$G(z) = \sum_{n\in\mathcal{N}} P(n) z^n = \mathbb{E}_P[z^n]. \tag{B1}$$

Typically, $\mathcal{N} \subset \mathbb{N}_0$, with $\inf \mathcal{N} = 0$. The probability distribution can be recovered from the generating function via higher-order derivatives evaluated at $z = 0$:

$$P(n) = \frac{1}{n!}\frac{d^n G}{dz^n}(0). \tag{B2}$$

Chief among the generating function's useful properties is the sum-of-random-variates rule. Let each X_i be one of M random variates, drawn from the distribution $P_i(x)$ with generating function $G_i(z)$. The generating function for the sum of these variates,

$$K = \sum_{i=1}^{M} X_i, \tag{B3}$$

is given by

$$G_K(z) = \prod_{i=0}^{M} G_i(z). \tag{B4}$$

This property has an analogy in the characteristic function of continuous distributions; where the distribution for a sum of two random variates is the convolution of the two distributions, the characteristic function is defined as the Fourier transform of the distribution function, so that the characteristic function for the sum of the variates is the product of the characteristic functions via the convolution theorem.

If the X_i are identically distributed—that is, drawn from the same distribution $P_X(x)$—and the number of variates to sum, M, is itself a random variate drawn from a distribution $P_M(m)$ with generating function $G_M(z)$, then the generating function for the sum

$$K = \sum_{i=1}^{M} X_i, \tag{B5}$$

is provided by the nested expression,

$$G_K(z) = G_M(G_X(z)). \tag{B6}$$

When $P_M(m)$ is a Poisson distribution, $G_K(x)$ is known as a compounded generator.

As a simple example, for the Poisson distribution,

$$P(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}, \tag{B7}$$

with λ the mean of the distribution, the generating function is

$$G(z) = e^{\lambda(z-1)}. \tag{B8}$$

Poisson distributions count the number of events that occur in a defined interval—most often in time, but we are also interested in space intervals. The above properties work well when the underlying Poisson process—the mechanism that generates these events—is homogeneous in this interval.

When the Poisson process is inhomogeneous, a generating functional can prove more useful. An inhomogeneous Poisson process is defined by an intensity function $\Lambda(\boldsymbol{x})$ with support Ξ. The integral of this intensity function over the support gives the expected number of events, $\lambda = \int_\Xi d\boldsymbol{x}\, \Lambda(\boldsymbol{x})$. Thus, the distribution for the number of events is a Poisson distribution with mean λ. The generating functional for the process in this case is defined as

$$G[f] = e^{\int_\Xi d\boldsymbol{x}\, \Lambda(\boldsymbol{x})(f(\boldsymbol{x})-1)}, \tag{B9}$$

where $f(\boldsymbol{x})$ is a functional argument defined over the same support Ξ. The probability density function for finding an event at $\boldsymbol{y}$ is given by the variational derivative of this functional:

$$p(\boldsymbol{y}) = \frac{dG}{df(\boldsymbol{y})}[\mathbb{0}] = \Lambda(\boldsymbol{y})e^{-\int_{\Xi} d\boldsymbol{x}\,\Lambda(\boldsymbol{x})}, \tag{B10}$$

where $0(\boldsymbol{x}) = 0$. The probability density for a set of events $\{\boldsymbol{y}_i\}$ is the higher-order variational derivative

$$p(\boldsymbol{y}_i) = \left(\left[\prod_i \frac{d}{df(\boldsymbol{y}_i)}\right]G\right)[\mathbb{0}] = [\prod_i \Lambda(\boldsymbol{y}_i)]e^{-\int_{\Xi} d\boldsymbol{x}\,\Lambda(\boldsymbol{x})}. \tag{B11}$$

The Poisson process is one kind of process that generates events—known generally as a point process. The concept of a generating functional can be extended to all point processes.

The generating functional has a compounding property similar to that of the generating functions. Let G_X be the generating functional for a point process, while G_M is the generating functional for another point process. If G_M, rather than directly generating events, instead generates point processes according to G_X, then draws from G_M will generate multiple events according to G_X. The generating functional for this compounded process is $G_K[f] = G_M[G_X[f]]$.

For a point-source population, the intensity function is

$$\Lambda(\boldsymbol{x}) = NT(\boldsymbol{x}), \tag{B12}$$

such that the intensity is distributed according to the spatial distribution, $T(\boldsymbol{x})$, and the total of the intensity function over the support of the spatial distribution is exactly the mean number of sources:

$$N = \int d\boldsymbol{x}\Lambda(\boldsymbol{x}). \tag{B13}$$

Now, let $G_{\boldsymbol{x}|N,T}$ be the generating functional for sources from the population:

$$G_{\boldsymbol{x}|N,T}[f] = \exp\left(N\int d\boldsymbol{x}T(\boldsymbol{x})(f(\boldsymbol{x}) - 1)\right), \tag{B14}$$

so that events drawn from this process define individual sources located at location $\boldsymbol{x}$. Each source, accordingly, defines its own generating functional:

$$G_{\boldsymbol{y}|F,\boldsymbol{x}}[h] = \exp\left(\kappa(\boldsymbol{x})F\int d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})(h(\boldsymbol{y}) - 1)\right), \tag{B15}$$

from which events define counts detected at location $\boldsymbol{y}$, given a source of flux F located at $\boldsymbol{x}$. The population averaged source functional is thus

$$G_{\boldsymbol{y}|p(F),\boldsymbol{x}}[h] = \mathbb{E}_{p(F)}[G_{\boldsymbol{y}|F,\boldsymbol{x}}[h]] = \int dFp(F)\exp\left(\kappa(\boldsymbol{x})F\int d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})(h(\boldsymbol{y}) - 1)\right). \tag{B16}$$

The total count generating functional is now a compound of these two generators (Daley & Vere-Jones 2003):

$$\begin{aligned} G_{\boldsymbol{y}|N,T,p(F)}[h(\boldsymbol{y})] &= G_{\boldsymbol{x}|N,T}[G_{\boldsymbol{y}|p(F),\boldsymbol{x}}(h(\boldsymbol{y}))] \\ &= \exp\left[N\int d\boldsymbol{x}T(\boldsymbol{x})\left(\int dFp(F)\exp\left[\kappa(\boldsymbol{x})F\int d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})(h(\boldsymbol{y}) - 1)\right] - 1\right)\right]. \end{aligned} \tag{B17}$$

The unbinned likelihood can be calculated using $G_{\boldsymbol{y}|N,T,p(F)}$; here, however, we are interested in the binned likelihood.

Consider the trivial generating function $G(z) = z$—a distribution with $p(0) = 0$ and $p(1) = 1$. Notice that

$$G_{\boldsymbol{x}|N,T}[G(z)] = e^{N(z-1)}; \tag{B18}$$

that is, the compound of the generating functional $G_{\boldsymbol{x}|N,T}$ with the generating function $G(z)$ gives the generating function for the number of events generated by the Poisson process. We can interpret this as a compounding of a point process with a counting process which records 100% of events, as $p(1) = 1$.

Direct substitution of this trivial generating function into $G_{\boldsymbol{y}|N,T,p(F)}$ would give the total number of counts in the entire image, as the integral $\int d\boldsymbol{y}$ runs over the entire support of $\eta(\boldsymbol{y})$. To get the number of counts, k_B, in bin B, we could perform this substitution, and then alter the support of $\eta(\boldsymbol{y})$ to Ω_B. An alternative is to consider the slightly less trivial generating function

$$G_{k_B|\boldsymbol{y}}(z) = 1 + \mathbf{1}_{\Omega_B}(\boldsymbol{y})(z - 1), \tag{B19}$$

where **1** is the indicator function, which is equal to one when $\boldsymbol{y} \in \Omega_B$ and zero otherwise. Thus, the probability of recording an event conditioned on the photon location $\boldsymbol{y}$ is $p(1|\boldsymbol{y}) = \mathbf{1}_{\Omega_B}(\boldsymbol{y})$, which is equal to one when the photon is inside the bin and zero otherwise. The conditional probability of discarding an event is $p(0|\boldsymbol{y}) = 1 - \mathbf{1}_{\Omega_B}(\boldsymbol{y})$, as expected. Using this, the generating function for the number of counts in bin B is

$$\begin{aligned} G_{k_B|N,T,p(F)}(z) &= G_{\boldsymbol{y}|N,T,p(F)}[G_{k_B|\boldsymbol{y}}(z)] \\ &= \exp\left[N\int d\boldsymbol{x}T(\boldsymbol{x})\left(\int dFp(F)\exp\left[\kappa(\boldsymbol{x})F\int_{\Omega_B} d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})(z-1)\right]-1\right)\right], \end{aligned} \tag{B20}$$

and after rearranging,

$$G_{k_B|N,T,p(F)}(z) = \exp\left[N\left(\int dFp(F)\int d\boldsymbol{x}T(\boldsymbol{x})\exp\left[\kappa(\boldsymbol{x})F\int_{\Omega_B} d\boldsymbol{y}\eta(\boldsymbol{y})\phi(\boldsymbol{y}|\boldsymbol{x})(z-1)\right]-1\right)\right]. \tag{B21}$$

From here, the $\mu_B(\varepsilon)$ measure defined in Equation (33) can be used to bring this into the expected form:

$$G_{k_B}(z) = \exp\left[N\left(\int dFp(F)\int d\varepsilon\mu_B(\varepsilon)e^{\varepsilon F(z-1)}-1\right)\right]. \tag{B22}$$

For multiple bins, the indicator generating function is

$$G_{\{k_B\}|\boldsymbol{y}}(\{z_B\}) = 1 + \sum_B \mathbf{1}_{\Omega_B}(\boldsymbol{y})(z_B - 1). \tag{B23}$$

Substitution of this into $G_{\boldsymbol{y}|N,T,p(F)}$ gives the multiple-bin generating function. However, evaluation of the probabilities for such a generating function would not only require a multidimensional $\mu(\{\varepsilon_i\})$ with an effective detector response for each bin, but would also require computing many cross terms between the z_i variables. For more than a few bins, this quickly becomes computationally untenable. For this reason, the whole-image likelihood is constructed by taking the product of the single-bin likelihoods as in Equation (39)—essentially a mean field approximation, where each bin is treated as statistically independent, removing the correlations.

Appendix C
Evaluation of the Power Series

Here, we demonstrate how to analytically evaluate the CPG generating function when the source-count function takes the form of a multiply broken power-law. First, we write the generating function for a point-source population as

$$G_{k_B}(z) = \exp\left[\left(\int d\varepsilon g(\varepsilon, z)\mu_B(\varepsilon) - N\right)\right], \tag{C1}$$

where

$$g(\varepsilon, z) = \int dFe^{\varepsilon F(z-1)}Np(F). \tag{C2}$$

As stated, we assume a broken power-law differential source-count function:

$$\frac{dN}{dF} = A\begin{cases} c_m\left(\dfrac{F}{F_{b(m)}}\right)^{-n_m} & F < F_{b(m)} \\ c_i\left(\dfrac{F}{F_{b(i)}}\right)^{-n_i} & F_{b(i+1)} < F < F_{b(i)} \\ c_1\left(\dfrac{F}{F_{b(2)}}\right)^{-n_1} & F_{b(2)} < F. \end{cases} \tag{C3}$$

This is equivalent to the broken power-law definition in Equation (46), with the c_i introduced to account for the normalization factors.

From the relation $Np(F) = dN/dF$, we have

$$g(\varepsilon, z) = \int dFe^{\varepsilon F(z-1)}\frac{dN}{dF}. \tag{C4}$$

The solution to this integral has a closed form for this choice of source-count function:

$$g(\varepsilon, z) = A\left(c_1 F_{b(2)}\sigma_1^0(\varepsilon, z) + \sum_{i=2}^{m-1} c_i F_{b(i)}\sigma_i^0(\varepsilon, z) + c_m F_{b(m)}\sigma_m^0(\varepsilon, z)\right), \tag{C5}$$

where, using $J_i(\varepsilon, z) = -\varepsilon F_{b(i)}(z-1)$ to shorten the notation,

$$\sigma_1^j(\varepsilon, z) = J_2(\varepsilon, z)^{n_1-1}\Gamma(1+j-n_1, J_2(\varepsilon, z)), \tag{C6}$$

$$\sigma_i^j(\varepsilon, z) = J_{i+1}(\varepsilon, z)^{n_i-1}\Gamma(1+j-n_i, J_{i+1}(\varepsilon, z))\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - J_i(\varepsilon, z)^{n_i-1}\Gamma(1+j-n_i, J_i(\varepsilon, z)), \tag{C7}$$

$$\sigma_m^j(\varepsilon, z) = J_m(\varepsilon, z)^{n_m-1}\gamma(1+j-n_m, J_m(\varepsilon, z)), \tag{C8}$$

where $\Gamma(n, x)$ and $\gamma(n, x)$ are the upper and lower incomplete gamma functions. The form of these σ functions has been chosen such that the following identities hold:

$$\frac{d\sigma_1^j}{dz}(\varepsilon, 0) = \sigma_1^{j+1}(\varepsilon, 0), \tag{C9}$$

$$\frac{d\sigma_i^j}{dz}(\varepsilon, 0) = \sigma_i^{j+1}(\varepsilon, 0), \tag{C10}$$

$$\frac{d\sigma_m^j}{dz}(\varepsilon, 0) = \sigma_m^{j+1}(\varepsilon, 0). \tag{C11}$$

The above results imply that an identification can be made with the power-series coefficients in Equation (36). For $j > 0$, we have

$$a_B^{(j)} = \int d\varepsilon g^j(\varepsilon, 0)\mu_b(\varepsilon), \tag{C12}$$

where

$$g^j(\varepsilon, 0) = A\left(c_1 F_{b(2)}\sigma_1^j(\varepsilon, 0) + \sum_{i=2}^{m-1} c_i F_{b(i)}\sigma_i^j(\varepsilon, 0) + c_m F_{b(m)}\sigma_m^j(\varepsilon, 0)\right). \tag{C13}$$

We postpone a discussion of the case of $j = 0$ for now. Note that $g^j(\varepsilon, 0)$ is not a function of B, the bin number. This allows the z-series of g^j to be computed once for the entire image; all of the bin-specific detector effects are contained in μ_B, which requires a simple numerical integration against this series for each bin. This can be many orders of magnitude more computationally efficient than NPTF, which requires the z-series to be computed for every bin.

The evaluation of upper and lower incomplete gamma functions is computationally expensive, and must be performed using numerical approximations. Thus, it is desirable to avoid evaluating these special functions for every term in the z-series. There exists a recurrence relation for the incomplete gamma functions, which is best exploited in terms of the scaled σ functions, $\hat{\sigma}$:

$$\hat{\sigma}_1^{j+1}(\varepsilon) = \frac{\sigma_1^{j+1}(\varepsilon, 0)}{(j+1)!} = \left(1 - \frac{n_1}{j+1}\right)\hat{\sigma}_1^j(\varepsilon) + \exp\left[j\ln J_2(\varepsilon, 0) - J_2(\varepsilon, 0) - \ln\Gamma(j+2)\right], \tag{C14}$$

$$\begin{aligned}\hat{\sigma}_i^{j+1}(\varepsilon) = \frac{\sigma_i^{j+1}(\varepsilon, 0)}{(j+1)!} &= \left(1 - \frac{n_i}{j+1}\right)\hat{\sigma}_i^j(\varepsilon) + \exp\left[j\ln J_i(\varepsilon, 0) - J_i(\varepsilon, 0) - \ln\Gamma(j+2)\right] \\ &\times \left(\left[\frac{F_{b(i)}}{F_{b(i+1)}}\right]^{n_i-1-j}\exp\left[J_i(\varepsilon, 0) - J_{i+1}(\varepsilon, 0)\right] - 1\right),\end{aligned} \tag{C15}$$

$$\hat{\sigma}_m^j(\varepsilon) = \frac{\sigma_m^j(\varepsilon, 0)}{j!} = \left(1 - \frac{n_m}{j}\right)^{-1}(\hat{\sigma}_m^{j+1}(\varepsilon) + \exp[(j-1)\ln J_m(\varepsilon, 0) - J_m(\varepsilon, 0) - \ln\Gamma(j+1)]). \tag{C16}$$

For $\hat{\sigma}_1$ and $\hat{\sigma}_i$, the base case of $j = 0$ is first computed from

$$\sigma_1^0(\varepsilon, z) = J_2(\varepsilon, z)^{n_1-1}\Gamma(1 - n_1, J_2(\varepsilon, z)) = E_{n_1}(J_2(\varepsilon, z)), \tag{C17}$$

$$\sigma_i^0(\varepsilon, z) = J_{i+1}(\varepsilon, z)^{n_i-1}\Gamma(1 - n_i, J_{i+1}(\varepsilon, z))\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - J_i(\varepsilon, z)^{n_i-1}\Gamma(1 - n_i, J_i(\varepsilon, z))$$

$$= E_{n_i}(J_{i+1}(\varepsilon, z))\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - E_{n_i}(J_i(\varepsilon, z)), \tag{C18}$$

where the exponential integral, $E_n(x)$, is used to stabilize the calculation. Then, the higher-order $\hat{\sigma}_1$ and $\hat{\sigma}_i$ terms are found using the above recurrence relations.

For $\hat{\sigma}_m$, the initial case starts from the highest-order term required, which for the whole image is $j = \max_B k_B$. This is computed using Equation (C8), then the lower-order terms are found using the recurrence relations. The power-series coefficient for $j = 0$ also includes the constant factor of N, which appeared in Equation (C1):

$$a_B^{(0)} = \int d\varepsilon g^0(\varepsilon, 0)\mu_b(\varepsilon) - N. \tag{C19}$$

If g^0 is calculated using the above procedure, this factor of N will cause a catastrophic cancelation for very dim (i.e., below-threshold) populations. This cancelation arises from the left-hand term of the right-hand side of this equation taking a value very close to N, such that it is equal to N up to the machine precision of the floating-point data type used in the computation. A resolution to this cancelation is achieved by incorporating the constant factor into the calculation of the exponential integral. The expected number of sources, N, is first evaluated in closed form in terms of the source-count function parameters:

$$N = A\left(\frac{c_1 F_{b(2)}}{n_1 - 1} + \sum_{i=2}^{m-1}\frac{c_i F_{b(i)}}{n_i - 1}\left[\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - 1\right] + \frac{c_m F_{b(m)}}{1 - n_m}\right). \tag{C20}$$

Each term in this equation is incorporated into the corresponding $\hat{\sigma}^0$ function, to give the modified σ terms

$$\tilde{\sigma}_1^0(\varepsilon) = E_{n_1}(J_2(\varepsilon, 0)) - \frac{1}{n_1 - 1}, \tag{C21}$$

$$\tilde{\sigma}_i^0(\varepsilon) = \left(E_{n_i}(J_{i+1}(\varepsilon, 0)) - \frac{1}{n_i - 1}\right)\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - \left(E_{n_i}(J_i(\varepsilon, 0)) - \frac{1}{n_i - 1}\right). \tag{C22}$$

In the very dim below-threshold regime, $J_i(\varepsilon, 0) \ll 1$, and the exponential integral becomes dominated by a leading constant factor:

$$E_n(x) = \frac{1}{n - 1} + \mathcal{O}(x). \tag{C23}$$

Thus, the cause of the numerical instability is revealed. Define a modified exponential integral function, $\tilde{E}_n(x, y) = y + E_n(x)$, so that

$$\tilde{E}_n\left(x, -\frac{1}{n - 1}\right) = \mathcal{O}(x). \tag{C24}$$

Now the modified σ terms can be written

$$\tilde{\sigma}_1^0(\varepsilon) = \tilde{E}_{n_1}\left(J_2(\varepsilon, 0), -\frac{1}{n_1 - 1}\right), \tag{C25}$$

$$\tilde{\sigma}_i^0(\varepsilon) = \tilde{E}_{n_i}\left(J_{i+1}(\varepsilon, 0), -\frac{1}{n_i - 1}\right)\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i-1} - \tilde{E}_{n_i}\left(J_i(\varepsilon, 0), -\frac{1}{n_i - 1}\right). \tag{C26}$$

The numerical algorithm for approximating the exponential integral is based on Navas-Palencia (2018). This algorithm was modified to calculate $\tilde{E}_n(x, y)$. For $x \ll 1$, series calculations are used by Navas-Palencia (2018); during evaluation of the first series term, y is added so that the modification occurs before subsequent terms are added. The result is that the first term of the series is canceled, allowing subsequent terms to accumulate without being lost in the machine precision. For other scales of x, non-series calculations are used and so y is simply added to the final result—an acceptable solution, as catastrophic cancelation does not occur in these regimes.

When $J_i(\varepsilon, 0) \gg 1$, the exponential integral algorithm may overflow. This corresponds to populations with unrealistically high flux per source. If this happens, a NaN is propagated through the likelihood-evaluation algorithm and is returned as the final probability. This allows this failure mode to be caught and reported. In these investigations, the correct course of action was to simply reduce the range of the flux prior, as the problem was always caused by the MCMC initialization drawing these unrealistically large flux values from the uniform prior.

With these modifications in place, the scaled power-series coefficients can be defined as

$$\nu_B^{(0)} = \int d\varepsilon A \left(c_1 F_{b(2)} \tilde{\sigma}_1^0(\varepsilon) + \sum_{i=2}^{m-1} c_i F_{b(i)} \tilde{\sigma}_i^0(\varepsilon) + c_m F_{b(m)} \tilde{\sigma}_m^0(\varepsilon) \right) \mu_B(\varepsilon), \tag{C27}$$

$$\nu_B^{(j)} = \int d\varepsilon A \left(c_1 F_{b(2)} \hat{\sigma}_1^j(\varepsilon) + \sum_{i=2}^{m-1} c_i F_{b(i)} \hat{\sigma}_i^j(\varepsilon) + c_m F_{b(m)} \hat{\sigma}_m^j(\varepsilon) \right) \mu_B(\varepsilon). \tag{C28}$$

We will use these expressions in Appendix D.

Appendix D
Evaluation of Bell Polynomials

The probability for k_b counts from the CPG likelihood was shown to be given by

$$P(k_B) = \frac{B_{k_B}(a_B^{(1)},\ldots,a_B^{(k_B)})}{k_B!}. \tag{D1}$$

This expression is written in terms of the Bell polynomials, which obey the recurrence relation

$$B_i(x_0,\ldots,x_i) = \sum_{j=0}^{i-1} \frac{(i-1)!}{j!(i-j-1)!} B_{i-j-1}(x_1,\ldots,x_{i-j-1}) x_{j+1}, \tag{D2}$$

with $B_0 = 1$. Given this, we can write

$$\begin{aligned} P(k_B) &= \sum_{j=0}^{k_B-1} \frac{(k_B-1)!}{j!k_B!} \frac{B_{k_B-j-1}(a_B^{(1)},\ldots,a_B^{(k_B-j-1)})}{(k_B-j-1)!} a_B^{(j+1)} \\ &= \sum_{j=0}^{k_B-1} a_B^{(j+1)} \frac{(k_B-1)!}{j!k_B!} P(k_B-j-1). \end{aligned} \tag{D3}$$

Now, let $l = k_B - j - 1$, so that

$$\begin{aligned} P(k_B) &= \sum_{l=0}^{k_B-1} \frac{a_B^{(k_B-l)}}{(k_B-l)!} \frac{(k_B-l)}{k_B} P(l) \\ &= \sum_{l=0}^{k_B-1} \nu_B^{(k_B-l)} \frac{(k_B-l)}{k_B} P(l), \end{aligned} \tag{D4}$$

where $\nu_B^{(k_b-l)}$ are the scaled power-series coefficients from Appendix C. This recurrence relation is equivalent to that found in Mishra-Sharma et al. (2017).

The numerical calculation of $P(k_B)$ is performed logarithmically,

$$\ln P(k_B) = \ln \sum_{l=0}^{k_B-1} \exp\left[\ln(\nu_B^{(k_B-l)}) + \ln(k_B-l) - \ln k_B + \ln P(l)\right], \tag{D5}$$

so as to ensure that numerical underflow does not occur if the scale of the likelihood is below that of the machine precision of the floating-point data type used. A standard exponential sum algorithm is employed, which determines the scale of the sum terms in advance; then, the terms are rescaled to avoid underflow before the summation is computed; last, the final result is scale-corrected.

Appendix E
Numerical Evaluation of $\mu_B(\varepsilon)$

In this appendix, we outline how to numerically compute $\mu_B(\varepsilon)$. The process is detailed in algorithm 1, which constructs a histogram density estimate $(\mu_B)_i$. In detail, a simulated source is drawn from $T(\boldsymbol{x})$, and the location $\boldsymbol{x}$ of this source is used to find the detector response $\kappa(\boldsymbol{x})$ and to select the appropriate PSF, $\phi(\boldsymbol{y}|\boldsymbol{x})$. This draw will form a sample from $\mu_B(\varepsilon)$, but as μ_B is a function of ε, the corresponding value of ε must be determined.

According to Equation (33), this requires an integral of the PSF and detector efficiency over the bin extents. This integral will also be performed through Monte Carlo sampling: a number, $N_{\text{sim counts}}$, of samples are drawn from the PSF. Let $\{\boldsymbol{y}_l\}$ be the set of these samples. The detector response for bin b is calculated as

$$\varepsilon = \kappa(\boldsymbol{x}) \frac{1}{N_{\text{sim counts}}} \sum_{\boldsymbol{y} \in \{\boldsymbol{y}_l\}} \eta(\boldsymbol{y}) \mathbf{1}_{\Omega_B}(\boldsymbol{y}), \tag{E1}$$

where $\mathbf{1}_{\Omega_B}(\boldsymbol{y}) = 1$ if $\boldsymbol{y}$ is within the bin extent Ω_B and is zero otherwise.

Thus, a set $\{\boldsymbol{x}_j\}$ of $N_{\text{sim sources}}$ simulated sources yields a set of detector responses, $\{\varepsilon_j\}$. The empirical estimate of $\mu_b(\varepsilon)$ can now be written as

$$\mu_B(\varepsilon) = \frac{1}{N_{\text{sim sources}}} \sum_{\varepsilon' \in \{\varepsilon_j\}} \delta(\varepsilon - \varepsilon'). \tag{E2}$$

Although not strictly required, conversion of this empirical measure to a histogram, $(\mu_B)_i$, will substantially speed up calculation of the likelihood when $N_{\text{sim sources}}$ and $N_{\text{sim counts}}$ are large. For histogram bin i, the value of the bin is

$$(\mu_B)_i = \int_{\hat{\varepsilon}_i, \hat{\varepsilon}_{i+1}} d\varepsilon \mu_B(\varepsilon), \tag{E3}$$

where $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_{i+1}$ are the histogram bin edges.

Algorithm 1. Numerical $\mu_B(\varepsilon)$

$(\mu_B)_i \leftarrow 0 \quad \forall B, i$
loop $N_{\text{sim sources}}$ times
  $\varepsilon_B \leftarrow 0 \quad \forall B$
  $\boldsymbol{x} \leftarrow$ sample from $T(\cdot)$
  loop $N_{\text{sim counts}}$ times
    $\boldsymbol{y} \leftarrow$ sample from $\phi(\cdot|\boldsymbol{x})$
    $B \leftarrow B'$ such that $\boldsymbol{y} \in \Omega_{B'}$
    $\varepsilon_B \leftarrow \varepsilon_B + \kappa(\boldsymbol{x})\eta(\boldsymbol{y})/N_{\text{sim counts}}$
  end loop
  $i_B \leftarrow i_B'$ such that $\varepsilon_B \in [\hat{\varepsilon}_{i_B'}, \hat{\varepsilon}_{i_B'+1}) \quad \forall B$
  $(\mu_B)_{i_B} \leftarrow (\mu_B)_{i_B} + 1/N_{\text{sim sources}} \quad \forall B$
end loop

Appendix F
Numerical Estimation of $\rho(f)$

For completeness, we also provide the algorithm for computing $\rho(f)$ as it appears in the NPTF. Again, this is computed numerically and not analytically, using algorithm 2. This algorithm creates a density estimate for $\rho(f)$ in the form of a histogram ρ_i, with bin edges $\hat{f}_i$.

Algorithm 2. Numerical $\rho(f)$

$\rho_i \leftarrow 0 \quad \forall i$
loop $N_{\text{sim sources}}$ times
  $f_B \leftarrow 0 \quad \forall B$
  $\boldsymbol{x} \leftarrow$ sample from Ω
  loop $N_{\text{sim counts}}$ times
    $\boldsymbol{y} \leftarrow$ sample from $\phi(\cdot|\boldsymbol{x})$
    $B \leftarrow B'$ such that $\boldsymbol{y} \in \Omega_{B'}$
    $f_B \leftarrow f_B + 1/N_{\text{sim counts}}$
  end loop
  $i_B \leftarrow i_B'$ such that $f_B \in [\hat{f}_{i_B'}, \hat{f}_{i_B'+1}) \quad \forall B$
  $\rho_{i_B} \leftarrow \rho_{i_B} + 1/N_{\text{sim sources}} \quad \forall B$
end loop
$\tilde{f}_i \leftarrow (\hat{f}_i + \hat{f}_{i+1})/2 \quad \forall i$
$\Delta_i \leftarrow (\hat{f}_{i+1} - \hat{f}_i) \quad \forall i$
$\rho_i \leftarrow \rho_i / \sum_j \tilde{f}_j \rho_j \delta_j \quad \forall i$

Appendix G
Natural Coordinate System

In this appendix, we provide additional details regarding the natural coordinate system introduced in Section 5 of the main text. The mean number of sources is taken to be N, while here, $F_{\rm PS}$ will be the total amount of flux emitted by the population of sources. There will be $m-1$ breaks, the locations of which will be defined by $m-2$ numbers $\beta_i \in (0, 1)$, where

$$\beta_i = \frac{F_{b(i+1)}}{F_{b(i)}}, \tag{G1}$$

is the ratio of flux location of the $(i+1)$th break to the same for the ith break.

The conversion between this coordinate system and the standard coordinate system of Equation (46) requires finding A and all $F_{b(i)}$ in terms of N, $F_{\rm PS}$, and all β_i. To do so, let

$$\alpha_i = \frac{F_{b(i)}}{F_{b(2)}} = \prod_{j=2}^{i-1} \beta_j, \tag{G2}$$

be a similar ratio to β_i but to the flux location of the first break, $F_{b(2)}$, instead. Now we can write the mean number of sources as

$$N = A\left(\frac{c_1 F_{b(2)}}{n_1 - 1} + \sum_{i=2}^{m-1} \frac{c_i F_{b(i)}}{n_i - 1}\left[\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i - 1} - 1\right] + \frac{c_m F_{b(m)}}{1 - n_m}\right), \tag{G3}$$

$$AF_{b(2)} = N\left(\frac{c_1 \alpha_2}{n_1 - 1} + \sum_{i=2}^{m-1} \frac{c_i \alpha_i}{n_i - 1}[(\beta_i^{-1})^{n_i - 1} - 1] + \frac{c_m \alpha_m}{1 - n_m}\right)^{-1}. \tag{G4}$$

There is a similar expression for the total flux of the population:

$$F_{\rm PS} = A\left(\frac{c_1 F_{b(2)}^2}{n_1 - 2} + \sum_{i=1}^{m-1} \frac{c_i F_{b(i)}^2}{n_i - 2}\left[\left(\frac{F_{b(i)}}{F_{b(i+1)}}\right)^{n_i - 2} - 1\right] + \frac{c_m F_{b(m)}^2}{2 - n_m}\right), \tag{G5}$$

$$AF_{b(2)}^2 = F_{\rm PS}\left(\frac{c_1 \alpha_2^2}{n_1 - 2} + \sum_{i=1}^{m-1} \frac{c_i \alpha_i^2}{n_i - 2}[(\beta_i^{-1})^{n_i - 2} - 1] + \frac{c_m \alpha_m^2}{2 - n_m}\right)^{-1}. \tag{G6}$$

From the ratio of Equations (G6) and (G4), we can find $F_{b(2)}$, then from either of these equations, we can find A. Finally, all remaining flux break locations, $F_{b(i)}$, can be found through Equation (G2).

ORCID iDs

Gabriel H. Collin https://orcid.org/0000-0003-1032-6496
Nicholas L. Rodd https://orcid.org/0000-0003-3472-7606
Tyler Erjavec https://orcid.org/0000-0002-9524-5101
Kerstin Perez https://orcid.org/0000-0002-6404-4737

References

Aartsen, M., Ackermann, M,, Adams, J., et al. 2020, ApJ, 893, 102
Abazajian, K. N., Canac, N., Horiuchi, S., & Kaplinghat, M. 2014, PhRvD, 90, 023526
Abazajian, K. N., Canac, N., Horiuchi, S., Kaplinghat, M., & Kwa, A. 2015, JCAP, 07, 013
Abazajian, K. N., & Kaplinghat, M. 2012, PhRvD, 86, 083511
Ajello, M., Albert, A., Atwood, W. B., et al. 2016, ApJ, 819, 44
Barcons, X. 1992, ApJ, 396, 460
Bartels, R., Krishnamurthy, S., & Weniger, C. 2016, PhRvL, 116, 051102
Buschmann, M., Rodd, N. L., Safdi, B. R., et al. 2020, PhRvD, 102, 023023
Calore, F., Cholis, I., & Weniger, C. 2015, JCAP, 03, 038
Calore, F., Donato, F., & Manconi, S. 2021, PhRvL, 127, 161102
Chang, L. J., Mishra-Sharma, S., Lisanti, M., et al. 2020, PhRvD, 101, 023014
Clark, H. A., Scott, P., Trotta, R., & Lewis, G. F. 2018, JCAP, 07, 060
Comtet, L. 1974, Advanced Combinatorics (Dordrecht: Reidel)
Daley, D. J., & Vere-Jones, D. 2003, An Introduction to the Theory of Point Processes, Probability and Its Applications (New York: Springer)
Daylan, T., Finkbeiner, D. P., Hooper, D., et al. 2016, PDU, 12, 1
Daylan, T., Portillo, S. K. N., & Finkbeiner, D. P. 2017, ApJ, 839, 4
Feyereisen, M. R., Ando, S., & Lee, S. K. 2015, JCAP, 09, 027
Feyereisen, M. R., Tamborra, I., & Ando, S. 2017, JCAP, 03, 057
Fleishman, G. D., & Bietenholz, M. F. 2007, MNRAS, 376, 625
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gelman, A., & Meng, X.-L. 1998, StaSc, 13, 163
Goodenough, L., & Hooper, D. 2009, arXiv:0910.2998
Gordon, C., & Macias, O. 2013, PhRvD, 88, 083521
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Green, P. J. 1995, Biometrika, 82, 711
Harrison, F. A., Craig, W. W., Christensen, F. E., et al. 2013, ApJ, 770, 103
Hong, J., Mori, K., Hailey, C. J., et al. 2016, ApJ, 825, 132
Hooper, D., & Goodenough, L. 2011, PhLB, 697, 412
Hooper, D., & Linden, T. 2011, PhRvD, 84, 123005
Hooper, D., & Slatyer, T. R. 2013, PDU, 2, 118
Krivonos, R., Wik, D., Grefenstette, B., et al. 2021, MNRAS, 502, 3966
Leane, R. K., & Slatyer, T. R. 2019, PhRvL, 123, 241101
Leane, R. K., & Slatyer, T. R. 2020a, PhRvL, 125, 121105
Leane, R. K., & Slatyer, T. R. 2020b, PhRvD, 102, 063019
Lee, S. K., Ando, S., & Kamionkowski, M. 2009, JCAP, 07, 007
Lee, S. K., Lisanti, M., & Safdi, B. R. 2015, JCAP, 05, 056
Lee, S. K., Lisanti, M., Safdi, B. R., Slatyer, T. R., & Xue, W. 2016, PhRvL, 116, 051103
Linden, T., Rodd, N. L., Safdi, B. R., & Slatyer, T. R. 2016, PhRvD, 94, 103013
Lisanti, M., Mishra-Sharma, S., Necib, L., & Safdi, B. R. 2016, ApJ, 832, 117
List, F., Rodd, N. L., Lewis, G. F., & Bhat, I. 2020, PhRvL, 125, 241102
Macias, O., Gordon, C., Crocker, R. M., et al. 2018, NatAs, 2, 387
Madsen, K. K., Christensen, F. E., Craig, W. W., et al. 2017, JATIS, 3, 044003
Madsen, K. K., Harrison, F. A., Markwardt, C. B., et al. 2015, ApJS, 220, 8
Malyshev, D., & Hogg, D. W. 2011, ApJ, 738, 181
Manconi, S., Korsmeier, M., Donato, F., et al. 2020, PhRvD, 101, 103026
Masias, M., Freixenet, J., Lladó, X., & Peracaula, M. 2012, MNRAS, 422, 1674
Mishra-Sharma, S., Rodd, N. L., & Safdi, B. R. 2017, AJ, 153, 253

Miyaji, T., & Griffiths, R. E. 2002, ApJL, 564, L5
Mukai, K. 2017, PASP, 129, 062001
Navas-Palencia, G. 2018, NuAlg, 77, 603
Portillo, S. K. N., Lee, B. C. G., Daylan, T., & Finkbeiner, D. P. 2017, AJ, 154, 132
Scheuer, P. A. G. 1957, PCPS, 53, 764
Skilling, J. 2004, in AIP Conf. Proc. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer et al. (Melville, NY: AIP), 395
Wik, D. R., Hornstrup, A., Molendi, S., et al. 2014, ApJ, 792, 48
Zechlin, H.-S., Cuoco, A., Donato, F., Fornengo, N., & Regis, M. 2016a, ApJL, 826, L31
Zechlin, H.-S., Cuoco, A., Donato, F., Fornengo, N., & Vittino, A. 2016b, ApJS, 225, 18
Zechlin, H.-S., Manconi, S., & Donato, F. 2018, PhRvD, 98, 083022
Zhong, Y.-M., McDermott, S. D., Cholis, I., & Fox, P. J. 2020, PhRvL, 124, 231103