A neural simulation-based inference approach for characterizing the Galactic Center γ-ray excess

Siddharth Mishra-Sharma[1,*] and Kyle Cranmer[1,2,†]

[1] *Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*

[2] *Center for Data Science, New York University, 60 Fifth Ave, New York, NY 10011, USA*

(Dated: July 26, 2021)

The nature of the *Fermi* γ-ray Galactic Center Excess (GCE) has remained a persistent mystery for over a decade. Although the excess is broadly compatible with emission expected due to dark matter annihilation, an explanation in terms of a population of unresolved point sources, *e.g.* millisecond pulsars, remains viable. The effort to uncover the origin of the GCE is hampered in particular by an incomplete understanding of diffuse emission of Galactic origin, which can lead to spurious features that make it difficult to robustly differentiate smooth emission, as expected for a dark matter origin, from more "clumpy" emission expected from a population of relatively bright, unresolved PSs. We use machine learning-based likelihood-free inference methods, in particular conditional density estimation with normalizing flows, in order to characterize the contribution of unresolved point sources to the GCE. Compared to traditional statistical techniques, our method attributes a significantly smaller fraction of the GCE flux to an unresolved point source population, and we place a lower bound of $29.7^{+9.8}_{-17.2}\%$ on such a contribution.

I. INTRODUCTION

Dark matter (DM) represents one of the major unsolved problems in particle physics and cosmology today. The traditional Weakly-Interacting Massive Particle (WIMP) paradigm envisions production of dark matter in the early Universe through freeze-out of dark sector particles weakly coupled to the Standard Model (SM) sector. In this scenario, one of the most promising avenues of detecting a dark matter signal is through an observation of excess γ-ray photons at $\sim$ GeV energies from DM-rich regions of the sky produced through the cascade of SM particles resulting from DM self-annihilation.

The *Fermi* γ-ray Galactic Center Excess (GCE), first identified over a decade ago using data from the *Fermi* Large Area Telescope (LAT) [1], is an excess of photons in the Galactic Center with properties—such as energy spectrum and spatial morphology—broadly compatible with expectation due to annihilating DM [2–16]. The nature of the GCE remains contentious however, with competing explanations in terms of a population of unresolved astrophysical point sources (PSs), in particular millisecond pulsars (MSPs), remaining viable [9, 17–25]. Analyses of the morphology of the excess have shown it to prefer a spatial distribution tracing the stellar bulge in the Galactic Center [15, 26, 27] rather than the expected distribution due to DM annihilation, although recent analysis have also shown preference for a DM-originating spatial morphology [28, 29]. Analyses leveraging the statistics of photons in the Galactic Center have shown the γ-ray data to prefer a point source origin of the excess [30–33]. Recent studies have however pointed out the potential of unknown systematics—such as the poorly understood morphology of the diffuse foreground emission and unmodeled point source populations—to affect the conclusions of these analyses [34–36].

Due to the high dimensionality of γ-ray data, a description of the photon map in terms of hand-crafted summary statistics such as the probability distribution of photon counts [30, 37] or a wavelet decomposition of the photon map [31, 38, 39] has traditionally been necessary in order to enable computationally tractable analyses. While effective, this reduced description necessarily involves loss of information compared to the original γ-ray map, increasing susceptibility to mismodeled features. On the other hand, recent developments in machine learning have given rise to analysis techniques that can extract more information from high-dimensional datasets and, consequently, more robustly hedge against unknown systematics in the data compared to traditional analyses based on specific data summaries. Machine learning methods have demonstrated promise for analyzing γ-ray data [40] and in particular for understanding the nature of the *Fermi* GCE [41, 42].

In this paper, we leverage recent developments in the field of simulation-based inference (SBI, also referred to as likelihood-free inference; see, *e.g.*, Ref. [43] for a recent review) in order to weigh in on the nature of the GCE. In particular, we employ conditional density estimation techniques based on normalizing flows [44, 45] in order to characterize the contributions of various modeled components, including "clumpy" PS-like and "smooth" DM-like emission spatially tracing the GCE, to the γ-ray photon sky at $\sim$ GeV energies in the Galactic Center region. We employ graph-based neural network architecture in order to automatically extract summary statistics from γ-ray maps optimized for the downstream task of estimating the distribution of parameters characterizing the contribution of modeled components to the GCE.

* sm8383@nyu.edu; ORCID: 0000-0001-9088-7845

† kyle.cranmer@nyu.edu; ORCID: 0000-0002-5769-7094

This paper is organized as follows. In Sec. II we describe our forward mode and the analysis framework based on neural simulation-based inference. In Sec. III we validate our analysis on various mock observations of the *Fermi* GCE. Section IV presents an application of the method to *Fermi* γ-ray data, including a discussion of systematic variations on the analysis. In Sec. V we quantify the susceptibility of the analysis to mismodeling of the signal and background templates using data-driven techniques. We conclude in Sec. VI.

II. METHODOLOGY

We being by describe the various ingredients of our forward model and the datasets used. We then detail our analysis methodology, going over in turn the general principles behind simulation-based inference, posterior estimation using normalizing flows, and learning representative summary statistics from high-dimensional γ-ray maps with graph neural networks.

A. Datasets and the forward model

We use the datasets and templates from Ref. [46] (packaged with Ref. [47]) to create the simulated maps for forward modeling. The data and templates used correspond to 413 weeks of *Fermi*-LAT Pass 8 data collected between August 4, 2008 and July 7, 2016. The top quarter of photons in the energy range 2–20 GeV by quality of PSF reconstruction (corresponding to the PSF3 event type) in the event class `ULTRACLEANVETO` are used. The recommended quality cuts are applied, corresponding to zenith angle less than 90°, `LAT_CONFIG` = 1, and `DATA_QUAL` > 0.1.[1] The maps are spatially binned using `HEALPix` [48] with `nside`=128. This dataset has been previously used in the literature for likelihood-based [32, 33, 36] as well as machine learning-based [41] analyses for characterizing the GCE.

The simulated data maps are a combination of smooth (*i.e.*, Poissonian) and PS contributions. Each PS population is completely specified by its spatial distribution, described by a spatial template, and the distribution of photon counts, specified by a source-count distribution. Two separate PS populations are modeled: *(i)* those spatially correlated with the GCE, modeled using (a line-of-sight integral of) the squared generalized Navarro-Frenk-White (NFW) [49, 50] profile with inner slope $\gamma = 1.2$ motivated by previous GCE analyses [8, 10, 51],

$$\rho(r) \propto \frac{1}{\left(r/r_\mathrm{s}\right)^{\gamma}\left(1+r/r_\mathrm{s}\right)^{3-\gamma}} \tag{1}$$

where r is the radial distance from the Galactic Center, $r_\mathrm{s} = 20\,\mathrm{kpc}$ is the Milky Way scale radius, and we take $R_\odot = 8.2\,\mathrm{kpc}$ as the distance to the Galactic Center [52, 53], and *(ii)* those correlated with the Galactic disk, modeled as a doubly-exponential profile motivated by studies of Galactic millisecond pulsar populations [54, 55],

$$n(R,z) \propto \exp(-R/5\,\mathrm{kpc})\exp(-|z|/0.3\,\mathrm{kpc}) \tag{2}$$

where R and z are the radial and vertical Galactic cylindrical coordinates. Photon counts from a generated PS population are put down on a map according to the *Fermi* PSF at 2 GeV, modeled as a King function, using the algorithm implemented in the code package `NPTFit-Sim` [56]. The source-count distribution (SCD) $\mathrm{d}N/\mathrm{d}S$ of each PS population, describing the differential number of sources per photon counts, is modeled as a doubly-broken power law,

$$\frac{\mathrm{d}N}{\mathrm{d}S} = A_\mathrm{PS}\begin{cases} \left(\frac{S}{S_{\mathrm{b},1}}\right)^{-n_1}, & S \geq S_{\mathrm{b},1} \\ \left(\frac{S}{S_{\mathrm{b},1}}\right)^{-n_2}, & S_{\mathrm{b},1} > S \geq S_{\mathrm{b},2} \\ \left(\frac{S_{\mathrm{b},2}}{S_{\mathrm{b},1}}\right)^{-n_2}\left(\frac{S}{S_{\mathrm{b},2}}\right)^{-n_3}, & S_{\mathrm{b},2} > S \end{cases} \tag{3}$$

specified by the breaks $\{S_{\mathrm{b},1}, S_{\mathrm{b},2}\}$, slopes $\{n_1, n_2, n_3\}$, and an overall normalization A_PS. We note that these parameters specify the spatially-averaged properties of the PS population—variation due to non-uniform exposure of the LAT instrument is accounted for in putting down simulated photon counts.

In addition to the PS-like emission, we account for Poissonian astrophysical emission in the simulated maps. These contributions include: *(i)* the Galactic diffuse foreground emission, *(ii)* spatially isotropic emission accounting for, *e.g.*, uniform emission from unresolved extragalactic sources *(iii)* emission from resolved PSs included in the *Fermi* 3FGL catalog [57], and *(iv)* lobe-like emission associated with the *Fermi* bubbles [58]. Finally, *(v)* Poissonian DM-like emission is included following a generalized squared-NFW profile as in Eq. (1) with inner slope $\gamma = 1.2$. The relative normalizations of these templates are regarded as parameters of the forward model. Templates for components *(ii)*-*(iv)* are obtained from Ref. [46].

The Galactic foreground component accounts for emission due to cosmic rays interacting with interstellar gas and radiation. In particular, Bremsstrahlung emission from cosmic-ray electrons scattering off of gas as well as photons produced as a result of the decay of pions produced through cosmic ray protons scattering elastically with the gas both trace the Galactic gas distributed, modulated by the incoming cosmic ray density. These components exhibit structure on smaller angular scales. Additionally, inverse Compton (up-)scattering (ICS) of the interstellar radiation field by cosmic ray electrons produces an important component of the γ-ray Galactic diffuse emission which spatially traces the Galactic

[1] https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_Data_Exploration/Data_preparation.html

charge carrier density and does not exhibit modulation on small scales. Normalizations of the gas-tracing components, subscripted 'brem/π^0', and the ICS-tracing component of the diffuse Galactic emission, subscripted 'ICS', are included separately in our forward model. Templates for these two components are described in our baseline configuration by Model O introduced in Ref. [32], which was based on the same *Fermi* dataset employed here and where it was found to be a formally better fit to the counts map in the Galactic Center region compared to templates previously employed in GCE analyses. We explore other diffuse models in Sec. IV B.

The final maps are obtained by combining a Poisson-fluctuated realization of the summed astrophysical templates with the simulated PS maps. The inner regions of the Galactic plane are masked at $|b| < 2°$, and a radial cut $r < 25°$ defines the region of interest (ROI) for the fiducial analysis. We mask resolved PSs from the 3FGL catalog at a containment radius of 0.8°, approximately corresponding to 99% confidence limit of PSF containment for the data type employed [57].

The forward model is thus specified by a total of 18 parameters—6 parameters for the overall normalizations of the Poissonian templates, and 6×2 parameters modeling the source-count distribution associated with GCE-correlated and disk-correlated PS populations. The priors used in the definition of the forward model are given in Tab. I. Priors on the Poissonian parameters $\{A_{\mathrm{brem}/\pi^0}$, A_{ICS}, A_{iso}, A_{bub}, and $A_{\mathrm{3FGL}}\}$ are motivated by a Poissonian fit to the real *Fermi* data in order to improve sample efficiency compared to using a wider prior range.

Since PS-like and Poissonian components of the model are exactly degenerate in the limit of each PS producing $\lesssim 1$ photon counts in expectation (see Refs. [33, 59] for a detailed discussion of this degeneracy, in particular in the context of previously-used likelihood-based methods), rather than placing uniform priors on the overall normalization of the PS-like and GCE-correlated Poissonian components A_{GCE} and A^{PS}, we place priors on the expected counts contributed per pixel for these components in order to mitigate biases caused by an induced prior preferring one model over the other and preventing the expression of the degeneracy. For the PS-like components, this is in practice done by placing a uniform prior on $\langle S_{\mathrm{pix}}^{\mathrm{PS}}\rangle = \int \mathrm{d}S\,\langle \mathrm{d}N/\mathrm{d}S\rangle_{\mathrm{pix}}$ where $\langle \mathrm{d}N/\mathrm{d}S\rangle_{\mathrm{pix}}$ is the mean SCD per pixel of the respective PS-like component.

Poissonian		PS-like (GCE and disk)	
Parameter	**Prior**	**Parameter**	**Prior**
$\langle S_{\mathrm{pix}}^{\mathrm{GCE}}\rangle$	[0, 2.5] ph	$\langle S_{\mathrm{pix}}^{\mathrm{PS}}\rangle$	[0, 2.5] ph
A_{brem/π^0}	[6, 12]	n_1	[10, 20]
A_{ICS}	[1, 6]	n_2	[1.1, 1.99]
A_{iso}	[0, 1.5]	n_3	[-10, 1.99]
A_{bub}	[0, 1.5]	$S_{\mathrm{b},1}$	[5, 40] ph
A_{3FGL}	[0, 1.5]	$S_{\mathrm{b},2}$	[0.1, 4.99] ph

TABLE I. Parameter priors used for the components of the forward model described in Sec. II A. All priors are uniform within the ranges specified. Priors on the Poissonian components, corresponding to overall normalization, are shown in the left table column, while those of the GCE- and disk-correlated PS components, parameterized according to Eq. (3), are shown in the right table column. The overall normalizations of the Poissonian GCE and PS-like components are parameterized through the mean number of counts contributed by the respective components in the ROI.

B. Likelihood-based inference methods

Irrespective of domain, of central interest in parameter estimation is often the probability distribution of a set of parameters of interest θ given some data x—the posterior distribution $p(\theta \mid x)$. Bayes' theorem can be used to obtain the posterior as $p(\theta \mid x) = p(\theta)\,p(x \mid \theta)/\mathcal{Z}$, where $p(x \mid \theta)$ is the likelihood and $\mathcal{Z} \equiv p(x)$ is the Bayesian evidence. In practice, parameters other than θ—latent variables z—are often involved in the data-generation process, and computing the likelihood involves marginalizing over the latent space, $p(x \mid \theta) = \int \mathrm{d}z\, p(x \mid \theta, z)$. In typical problems of interest, the high dimensionality of the latent space often means that this integral is intractable, necessitating simplifications in statistical treatment as well as theoretical modeling.

The present problem is no exception. In its simplest incarnation, traditional template fitting involves modeling the counts map as a Poisson realization of a linear combination of spatial templates, $p(x|\theta = A_i) = \mathrm{Pois}\left(\sum_i A_i T_i\right)$, where the normalizations A_i of the respective templates are parameters of interest and there are no additional latent variables. Inference on the parameters of interest in this case is easily admitted within a frequentist or Bayesian framework. In the model described in Sec. II A on the other hand, the presence of a PS population where no PS can be individually localized or characterized introduced a large number of latent variables, specifically the position of and counts emitted by each PS. Ignoring the contribution from Poissonian templates for a moment and considering only a single PS population, the likelihood in this case is formally given by

$$p(x \mid \theta_{\mathrm{PS}}) = \sum_n \int \mathrm{d}^n z\, p\left(n \mid \lambda\right)\, p(z \mid \theta_{\mathrm{PS}})\, p(x|z), \quad (4)$$

where θ are the parameters of interest that characterize the counts distribution of sources parameterized, *e.g.*, by a broken power law as in Eq. (3) and $\lambda = \lambda(\theta) = \int \mathrm{d}S\, \mathrm{d}N/\mathrm{d}S$ is the expected number of PSs in the ROI. n is the total number of PSs in the ROI, with the sum running over all possible number of PSs. This high-dimensional integral is computationally intractable, and traditional likelihood-based methods aim to simplify the problem setting in order to enable its evaluation in a practical setting.

The 1-point PDF framework, first introduced to γ-ray analyses in Ref. [60] and extended in Refs. [30, 37], considers a simplification in terms of the pixel-wise likelihood assuming each pixel to be statistically independent. This significantly reduces the latent space dimensionality by eliminating the positions of individual PSs as latent variables, localizing them within a pixel and modulating their expected number by the modeled spatial template (*e.g.*, GCE-correlated or disk-correlated in our case). We briefly outline the philosophy of this method, pointing the reader to a detailed discussed as well as numerical implementations of the method in Refs. [30, 47].

Since emission from each PS can be regarded as statistically independent, the probability of a given PS, indexed i, emitting x_i^p photons in a pixel p is given by

$$p(x_i^p \mid \theta_{\mathrm{PS}}) = \int \mathrm{d}S_i \, p(S_i \mid \theta_{\mathrm{PS}}) \, p(x_i^p|S_i), \tag{5}$$

where S_i is the expected counts from the PS following some prior probability parameterized by θ_{PS}, in this case following a broken power law as in Eq. 3 with parameters $\theta_{\mathrm{PS}} = \{A_{\mathrm{PS}}, n_1, n_2, n_3, S_{\mathrm{b},1}, S_{\mathrm{b},2}\}$, and $p(x_i^p|S_i)$ is the distribution of actual counts given a latent S_i, assumed to be given by a Poisson distribution. The probability of having a total of x_p counts in a pixel from multiple PSs is then described by a multinomial distribution, subject to the constraint that the total number of counts be equal to the observed counts.

$$p(x^p \mid \theta_{\mathrm{PS}}) = \sum_n p\left(n \mid \lambda\right) \sum_{n_j} \delta\left(\sum_j n_j j - x^p\right) \tag{6}$$

$$\times \frac{n!}{\prod_j n_j} \prod_{j=1}^{n} p(x_i^p = j \mid \theta_{\mathrm{PS}})^{n_j}. \tag{7}$$

Here n_j is the number of PSs contributing j counts, and a sum over the distribution of number of PSs n is again included and assumed to follow a Poisson distribution. In this case, the sum over n can be eliminated and the distribution of observed counts is given by

$$p(x^p \mid \theta_{\mathrm{PS}}) = \sum_{n_j} \delta\left(\sum_j n_j j - x^p\right) \tag{8}$$

$$\times \prod_j \mathrm{Pois}\left(n_j \mid \lambda \times p(x_i^p = j \mid \theta_{\mathrm{PS}})\right). \tag{9}$$

While not immediately obvious from this expression, eliminating the positions of individual PSs as latent parameters as well as the sum over the possible number of PSs n renders the per-pixel likelihood tractable, and the total data likelihood can then be computed as a product over pixels, $p(x \mid \theta_{\mathrm{PS}}) = \prod_p p(x^p \mid \theta_{\mathrm{PS}})$. We emphasize that we have only provided a brief overview of the method here, with further analytic simplifications, extensions to include the effect of an instrumental point-spread function and exposure, as well as a numerical recipe for evaluating the likelihood described in detail in Ref. [47]. We note that including a finite point spread function renders the per-pixel likelihood only approximately correct, since this introduced correlation across pixels over a range of the PSF size. Using simulations, previous studies have shown this approximation to be accurate enough for the present problem when using a pixel size of the order of the PSF size itself. Further generalizations of the method that can admit more extreme variations in the instrumental point-spread function and exposure, which are necessary for application to *e.g.* X-ray data, were introduced and studied in in Ref. [59].

Probabilistic cataloging [61] is a complementary method for characterizing the sub-threshold contribution of a PS population. This technique keeps the latent variables in Eq. (4) *i.e.*, the positions and expected fluxes of individual PSs, as parameters of interest, using trans-dimensional sampling techniques to obtain the distribution over possible catalogs of unresolved PS populations.

In this paper, we run the NPTF algorithm on *Fermi* data in order to have a comparison point to previous studies employing the method. We use the NPTF likelihood implemented in `NPTFit` [47] and construct the posterior distribution over the parameters of interest described in Sec. II A using nested sampling implemented `dynesty` [62]. The static variant of nested sampling is run in its default configuration with 1000 live points, stopping when the estimated contribution of the remaining posterior volume to the log-evidence falls below $\Delta \log \mathcal{Z} < 0.01$.

1-point PDF-based techniques have shown enormous promise in characterizing γ-ray PS populations below the *Fermi* detection limits, both in relation to the GCE [30, 32, 34, 35] and more generally, *e.g.* for characterizing the contribution of extragalactic PSs at high latitudes [63] and for searching for a DM annihilation signal from Galactic subhalos [64]. It has recently been point out however [35, 36] that signal and foreground mismodeling associated in particular with the emission in the Galactic Center region can hamper the ability to accurately characterize the PS contributions to the GCE. In particular, Refs. [30, 33, 36] pointed out that spurious residuals associated with foreground mismodeling can lead to the mischaracterization of a purely DM signal as a population of PSs. Ref. [32] showed that many of the issues pointing to the expression of such effects in *Fermi* data could be mitigated through the use of better Galactic foreground models along with affording them more large-scale degrees of freedom.

Ref. [34] further showed analytically how mismodeling, in particular an unmodeled North-South asymmetry in a DM signal, could lead to the inference of spurious PSs in NPTF analyses of the GCE. The fact that NPTF analyses rely on a per-pixel likelihood can make them especially susceptible to the effects of mismodeling. Assuming a corresponding permutation of template labels, the NPTF likelihood is invariant to a permutation of pixels within the analysis ROI. This means that residuals associated with mismodeling can mimic the effect of a PS population through the statistics of their PDF, ignoring any spatial correlations that could have an additional regularizing effect in the face of mismodeling. In the rest of this section, we will describe the components of our machine learning-based method that is able to leverage pixel-to-pixel spatial correlations with the overall aim of reducing susceptibility to signal and background mismodeling.

C. Simulation-based inference

Simulation-based inference (SBI) refers to a class of methods for performing inference when the data-generating process does not have a tractable likelihood. In this setting, a model is defined through a simulator as a probabilistic program, often knows as a forward model. Samples x from the simulator then implicitly define a likelihood, $x \sim p(x \mid \theta)$. In the simplest realizations of SBI, samples x' generated from a given prior proposal distribution $p(\theta)$ can be compared to a given dataset of interest x, with the approximate posterior defined by samples that most closely resemble x according to some similarity metric. Such methods—usually grouped under the umbrella of Approximate Bayesian Computation (ABC) [65]—are not uncommon in astrophysics and cosmology. Nevertheless, they suffer from several downsides. The curse of dimensionality usually necessitates reduction of data to representative, hand-crafted lower-dimensional summary statistics $s(x)$, resulting in loss of information. A notion and measure of distance between summaries from the implicit model and those derived from the dataset of interest is necessary, leading to inexact inference. Additionally, the ABC analysis must be performed anew for each new target dataset.

Recent methods [66–82] have leveraged advancements in machine learning, in particular the ability of neural networks to extract useful features from high-dimensional data and to flexibly approximate functions and distributions, in order to address these issues, enabling new ways of performing inference on complex models defined through simulations; see Ref. [43] for a review of recent developments.

D. Conditional density estimation with normalizing flows

In this paper, we approximate the joint posterior $p(\theta \mid x)$ through a parameterized distribution $\hat{p}_\phi(\theta \mid s)$ conditioned on *learned* summaries $s = s(x)$ from the simulated samples x. This class of simulation-based inference techniques, known as conditional neural density estimation [79], directly models the posterior distribution given a set of samples drawn from a simulator according to some prior proposal distribution $p(\theta)$.

We employ normalizing flows [44, 45], which provide an efficient way of constructing flexible probability distributions. Normalizing flows model the conditional distribution $\hat{p}_\phi(\theta \mid s)$ as a series of invertible transformations, denoted f and having a tractable inverse and Jacobian, from a base distribution $\pi_z(z)$, chosen here to be a standard Gaussian $z \sim \mathcal{N}(0, \mathbb{1})$, to the target distribution:

$$\hat{p}(\theta \mid x) = \pi_z\left(f^{-1}(\theta)\right) \left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right| \tag{10}$$

Specifically, we use Masked Autoregressive Flows (MAFs) [83] for posterior estimation. The MAF is built upon blocks of affine transformations where scaling and shifting factors for each dimension are computed with a Masked Autoencoder for Distribution Estimation (MADE) [84]. For a single block, the transformation from θ to z is expressed as

$$z_i = (\theta_i - \mu_i) \cdot \exp\left(-\alpha_i\right) \tag{11}$$

where $\mu_i = f_{\mu_i}\left(\theta_{1:i-1}; x\right)$ and $\alpha_i = f_{\alpha_i}\left(\theta_{1:i-1}; x\right)$ are scaling and shift factors modeled by a MADE according to the autoregressive condition. This allows for an analytically tractable determinant,

$$\left|\det\left(\frac{\partial f^{-1}}{\partial \theta}\right)\right| = \exp\left(-\sum_i \alpha_i\right) \tag{12}$$

and a forward pass through the flow according to Eq. (11). Multiple transformations can be composed together as $f = f_1 \circ f_2 \circ \ldots \circ f_K$ in order to model more expressive posteriors,

$$\hat{p}(\theta \mid x) = \pi_z\left(f^{-1}(\theta)\right) \prod_{i=1}^{K} \left|\det\left(\frac{\partial f_i^{-1}}{\partial z_{i-1}}\right)\right|. \tag{13}$$

The log-probability of the posterior can then be computed using Eq. (12):

$$\log \hat{p}(\theta \mid x) = \log\left[\pi_z\left(f^{-1}(\theta)\right)\right] - \sum_{i=1}^{K}\sum_{j=1}^{N} \alpha_j^i, \tag{14}$$

which acts as the optimization objective. Here, we use 8 MAF transformations, each made up of a 2-layer MADE with 128 hidden units. Each transformation is conditioned on summaries $s(x)$ extracted from the γ-ray maps (described below) by including these as inputs into each transformation block.

E. Learning summary statistics with (graph) neural networks

The curse of dimensionality makes it computationally prohibitive to condition the density estimation task on the raw dataset x *i.e.*, the γ-ray pixel counts map in the region of interest (ROI). Representative summaries $s = s_{\varphi}(x)$ of the data must therefore be extracted in order to enable a tractable analysis, where φ parameterizes the data-to-summary transformation Although many choices for data summaries are possible—*e.g.*, a Principal Component Analysis (PCA) decomposition of the photon counts map, an angular power spectrum decomposition of the photon counts map, or simply a histogram of the photon counts—in this paper, we use a neural net to automatically learn low-dimensional summaries that are efficiently suited for the specific downstream task at hand.

Graph construction and architecture

The `DeepSphere` architecture [85–87], with a configuration similar to that employed in Ref. [41], is used to extract representative summaries and is briefly outlined here. `DeepSphere` is a graph-based convolutional neural network (CNN) architecture tailored to data sampled on a sphere, and in particular is able to leverage the hierarchical structure of data in the `HEALPix` representation. This makes it well-suited for our purposes.

The `HEALPix` sphere is represented as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$. where $\mathcal{V}$ is the set of $N_{\text{pix}} = |\mathcal{V}|$ vertices, $\mathcal{E}$ is the set of edges, and A is the weighted adjacency matrix. Each pixel i is represented by a vertex $v_i \in \mathcal{V}$. Each vertex v_i is then connected to the 8 vertices v_j which represent the neighboring pixels j of pixel i, forming edges $(v_i, v_j) \in \mathcal{E}$. Given those edges, we define the weights of the adjacency matrix A over neighboring pixels following the edge weighing scheme given in Ref. [85].

We use the combinatorial graph Laplacian, defined as $L = D - A$, where D is the diagonal degree matrix, and can be used to define a Fourier basis on a graph. By construction symmetric positive semi-definite, the graph Laplacian can be decomposed as $L = U \Lambda U^T$, where U is an orthonormal eigenvector matrix and Λ is a diagonal eigenvalue matrix. The Laplacian eigenvectors then define the graph Fourier basis, with the graph Fourier transform $\tilde{x}$ of a signal x on a graph being its projection $\tilde{x} = U^T x$. Given a convolutional kernel h, graph convolutions can be efficiently performed in the Fourier basis as $h(L)x = U h(\Lambda) U^T x$.

The `DeepSphere` convolutional kernel h is defined as a linear combination of Chebychev polynomials, $h_\theta(L) = \sum_{k=0}^{K} \theta_k T_k(L)$ where T_k are the order-k Chebyshev polynomials and θ_k are the $K+1$ filter coefficients to be fit. The graph filering operation can then be expressed as

$$h_\theta(L)x = U \left(\sum_{k=0}^{K} \theta_k T_k(\Lambda) \right) U^T x = \sum_{k=0}^{K} \theta_k T_k(L) x. \quad (15)$$

We set $K = 5$ as the maximum Chebyshev polynomial order, having checked that larger values do not affect the results of the analysis.

Following Ref. [86], the feature extraction architecture is built out of layers which progressively coarsen the pixel representation of the γ-ray maps while increasing the number of filter channels at each step. The input map corresponds to the 16,384 pixels in the single pixel corresponding to `nside`=1 covering the Galactic Center region. Starting with `HEALPix` resolution `nside`=128, each graph convolution operation is followed by a BatchNorm, a ReLU nonlinearity, and a max pooling operation which downsamples the representation into a coarser resolution, starting with `nside`=128 until a single pixel channel at `nside`=1 remains after the final convolutional layer. The number of filter channels is doubled at each convolution until a maximum of 256. The output of the final convolutional layer is augmented with 2 additional auxiliary variables—the log-mean and log-standard deviation of the γ-ray map within the region of interest—and passed through a fully-connected layer with 1024 hidden units outputting a desired number of summary features, which is 128 in our fiducial configuration. Pixels outside of the ROI as well as masked PSs are set to zero in the input maps. All input maps are standardized to zero mean and unit variance across the training sample.

Using a neural network-based feature extractor, we implicitly use an approximation to the full data likelihood in Eq. (4) associated with our forward model of emission in the Galactic Center region. The method is thus able to capture pixel-to-pixel correlations in the γ-ray map, mitigating one of the major limitations of likelihood-based methods described in Sec. II B

Optimization, training, and evaluation

The graph-based and normalizing flow networks are trained simultaneously. 10^6 samples are generated using the prior proposal distribution, and models are optimized with batch size 128 using the `AdamW` [88, 89] optimizer with initial learning rate 10^{-3} and weight decay 10^{-5}, using cosine annealing to decay the learning rate across epochs. Training proceeds for up to 50 epochs with early stopping if the validation loss, evaluated on 15% of held-out samples, has not improved after 8 epochs.

After training, given a new dataset (either real or simulated *Fermi* data in our ROI), the posterior is obtained by drawing 10,000 samples from the flow within the prior using rejection sampling, conditioning each flow transformation on summaries returned by the graph-based neural network with the new dataset as input.

III. TESTS ON SIMULATED DATA

We begin by validating our pipeline on simulated *Fermi* data. We create simulated datasets by drawing parameter values from ranges motivated by a fit of the model

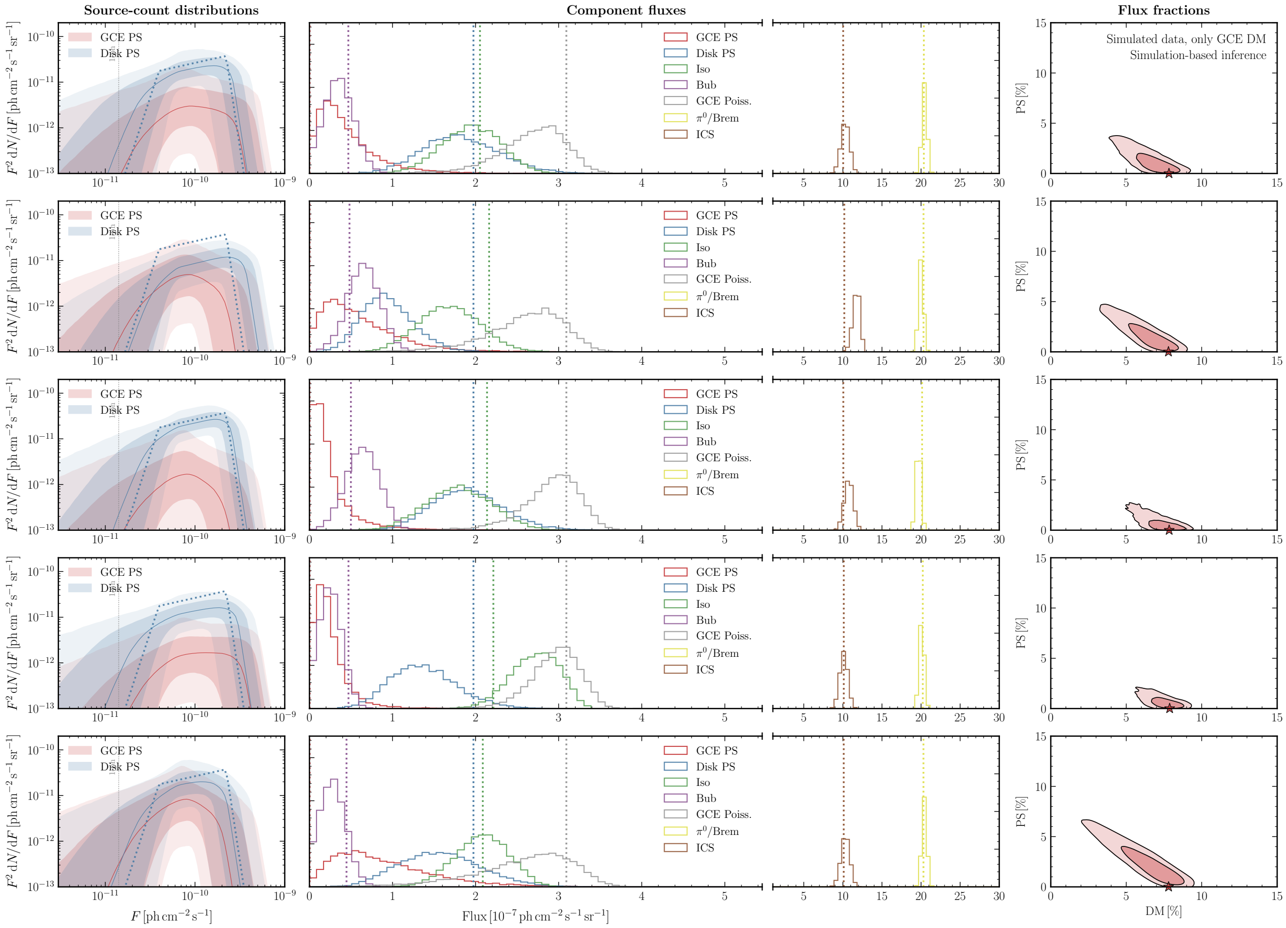

FIG. 1. Results of the analysis pipeline on simulated *Fermi* data where the GCE consists of purely DM-like emission, with different rows corresponding to five different simulated realizations. The left column shows the inferred source-count distribution posteriors for GCE-correlated (red) and disk-correlated (blue) PS. Solid lines correspond to the inferred point-wise median, with the lighter and darker bands representing the point-wise middle-68% and 95% posterior containment respectively. The middle panel shows the posteriors for the Poissonian templates. The right panel shows the joins posterior on the flux fractions of DM-like and PS-like emission. The dotted lines (in the left two columns) and the stars (in the right column) correspond to the true simulated quantities. DM-like emission is successfully inferred in each case, with the other parameter posteriors corresponding faithfully to the true simulated values.

to real *Fermi* data in our fiducial ROI, and test the ability of our model to infer the presence of either DM-like or PS-like signals on top of the modeled astrophysical background.

Figure 1 shows results of the analysis pipeline conditioned on five simulated realizations of maps where the GCE consists of purely DM-like emission. The left column shows the middle-68/95% containment of the point-wise posterior on the source-count distributions of GCE- and disk-correlated point source emission in red and blue, respectively. The middle column shows the posteriors on various modeled emission components, excluding emission from resolved 3FGL PSs as the posterior in that case is largely unconstrained owing to the fact that resolved PSs are masked out in the analysis. The right column shows the fraction of DM- and PS-like emission in proportion to the total inferred flux in the ROI. The true underlying parameter values from which the data was generated are represented by dotted lines in the left and middle columns, and by star markers in the right column. We see that, in all cases shown, the pipeline successfully recovers the presence of DM-like emission, with little flux attributed to unresolved PSs. Some PS-like emission is inferred in most cases as well however, due to a combination of degeneracy with both disk-correlated and DM-like flux. The overall flux of all components corresponds well to their true underlying values.

Figure 2 shows the corresponding results for simulated data containing PS-like emission correlated with the GCE. Here, simulations were produced such that the highest break of the GCE-correlated PS SCD was contained between 5 and 20 expected photon counts, since we found that the method cannot robustly attribute an SCD that corresponds to a peak dimmer than $\lesssim 5$ photon counts to a PS population. We see that PS-like emission is successfully inferred in each case, while at the same time exemplifying a degeneracy with the Poissonian component. Furthermore, as seen in the left column, the method is able to characterize the contribution of the two modeled PS components through the inferred source-count distribution. Some degeneracy between GCE- and disk-correlated PSs is seen, although the true SCDs are seen to lie within the 95% containment interval of the inferred point-wise SCD posteriors in each case.

IV. RESULTS ON *FERMI* DATA

We finally apply our formalism to the real *Fermi* dataset in our ROI. As a point of comparison, we also run the NPTF pipeline on the data using the same spatial templates and prior assumptions described in Sec. II A.

The results of the NPTF analysis are shown in the bottom panel of Fig. 3. Consistent with previous analyses using a similar configuration, a preference for PS-like emission is seen, with the analysis attributing the majority of the GCE to PS-like emission. [SM: More detailed description of NPTF.]

The top panel of Fig. 3 shows results using the neural simulation-based analysis pipeline introduced in this paper. Although posteriors for the astrophysical background templates are seen to be broadly consistent with the NPTF anlaysis, the preference for PSs is significantly reduced in this case. In fact, about half of the emission is attributed to PS-like and Poissonian emission each. We also note that the inferred GCE-correlated SCD peaks at values lower than those inferred from previous NPTF analyses, which have generally found the bulk of expected emission from PSs to lie just below the 3FGL PS detection threshold [30] at ~ 2–$3 \times 10^{-10}\,\mathrm{ph\,cm^{-2}\,s^{-1}}$. [SM: Quantify where the SCD peaks and how many PSs it predicts.]

A. Signal injection test on data

A crucial self-consistency test is the ability of the analysis to recover an artificial signal injected onto the real γ-ray data. As shown in Ref. [36], early applications of the 1-point PDF based methods like NPTF to the GCE would generally fail this closure test, with implications for characterizing the nature of PSs in the Galactic Center explored in Refs. [32, 33]. In particular, it was shown that the closure test can help diagnose underlying issues associated with mismodeling of the diffuse foreground emission, which have the potential to bias the characterization of PS populations. We perform a version of this test within our framework, testing the ability of our method to recover different mock signals injection onto the real *Fermi* data.

Figure 4 shows the results of this test, with the different rows corresponding to different signal configuration—purely DM, bright PSs, medium-bright PSs, and dim PSs. Bright, medium-bright, and dim PS configurations are taken to peak at 20, 10, and 5 photon counts respectively. The leftmost columns shows the fiducial analysis on *Fermi* data, with subsequent columns showing signals of progressively larger sizes injected onto the data, up to approximately the size of the original GCE signal. The dotted horizontal and vertical lines show the expected total emission on top of the median fluxes for the PS and DM components of the GCE inferred without any additional injected signal, respectively.

The additional injected signal is seen to be reconstructed correctly within the inferred 95% confidence interval in all four cases. For the DM signal (top row), the brightest tested DM signal is seen to partially reconstruct as PS-like, which could be attributed to the larger magnitude of Poisson fluctuations in this case mimicking the effect of an unresolved PS population. The injected PS signals (rows 2-4) are correctly reconstructed in all cases, with the dimmer PS signals showing a more prominent flat direction with Poissonian emission, as expected.

B. Systematic variations on the analysis

We explore several systematic variations on the fiducial analysis, shown in Fig. 5.

- *Variation of the diffuse foreground model:* In addition to diffuse Model O considered in the fiducial analysis, we consider Models A and F from Ref. [11] to model the diffuse foreground emission. Results for these variations are shown in the left panel of Fig. 5 for Models A (blue) and F (green) compared to the fiducial Model O (red). Although the overall GCE flux is seen to varying by up to a factor of ~ 2 between diffuse models, the overall conclusion regarding the relative amounts of flux attributed to PS-like and Poissonian emission remained unchanged, with similar proportions of the GCE attributed to each component. A slight preference for smooth emission is seen when using Models A and F, and the results between these two models and broadly similar.

- *Variations on the ROI size:* Although the GCE signal is concentrated predominantly in the inner 10°, the use of the larger 25° ROI in this work is motivated by the fact that a larger region may better constrain various spatially-extended modeled emission components. On the other hand, there is also the potential for more susceptibility to mismodeling effects when using a larger ROI. The right panel of Fig. 5 shows analysis results using smaller ROI sizes—$10°, 15°$ and $20°$. These are seen to be completely consistent with the fiducial analysis in the 25° ROI, with a wider posterior in the smaller ROIs as expected since these contain less information than fiducial ROI.

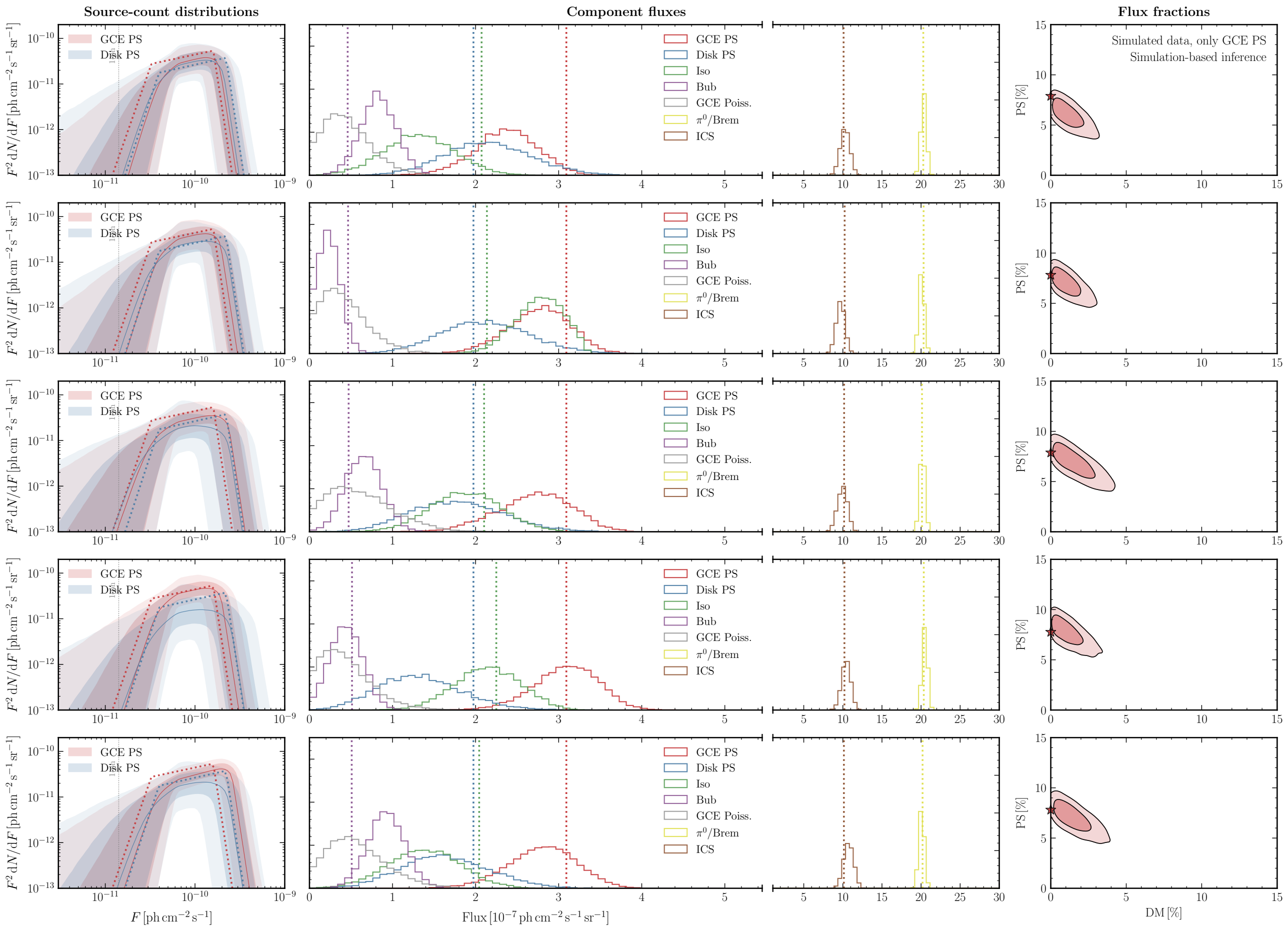

FIG. 2. Same as Fig. 1, but for five simulated realization of *Fermi* data where the GCE consists of predominantly PS-like emission. PS-like emission is inferred in each case, with the other posteriors corresponding faithfully to their true simulated quantities. The GCE-correlated source-count distribution is also seen to be successfully recovered in the left panel.

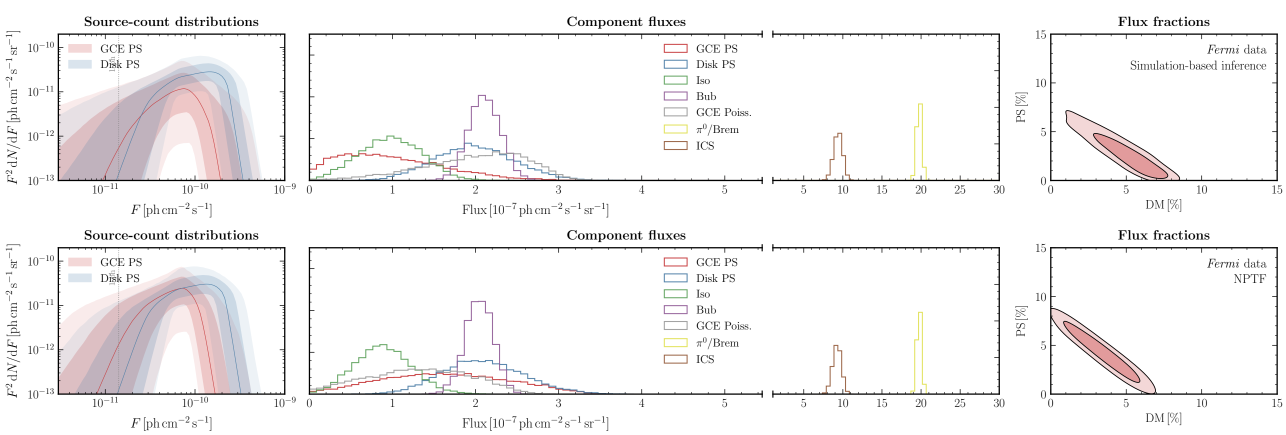

FIG. 3. Results of the fiducial analysis on real *Fermi* data. *(Top row)* Analysis using neural simulation-based inference with normalizing flows, and *(bottom row)* using the 1-point PDF likelihood implemented in the non-Poissonian template fitting (NPTF) framework. While moderate preference for a PS-like origin of the GCE is seen in the case of the NPTF analysis (bottom), the simulation-based inference analysis attributes a substantially smaller fraction of the GCE to PS-like emission (top).

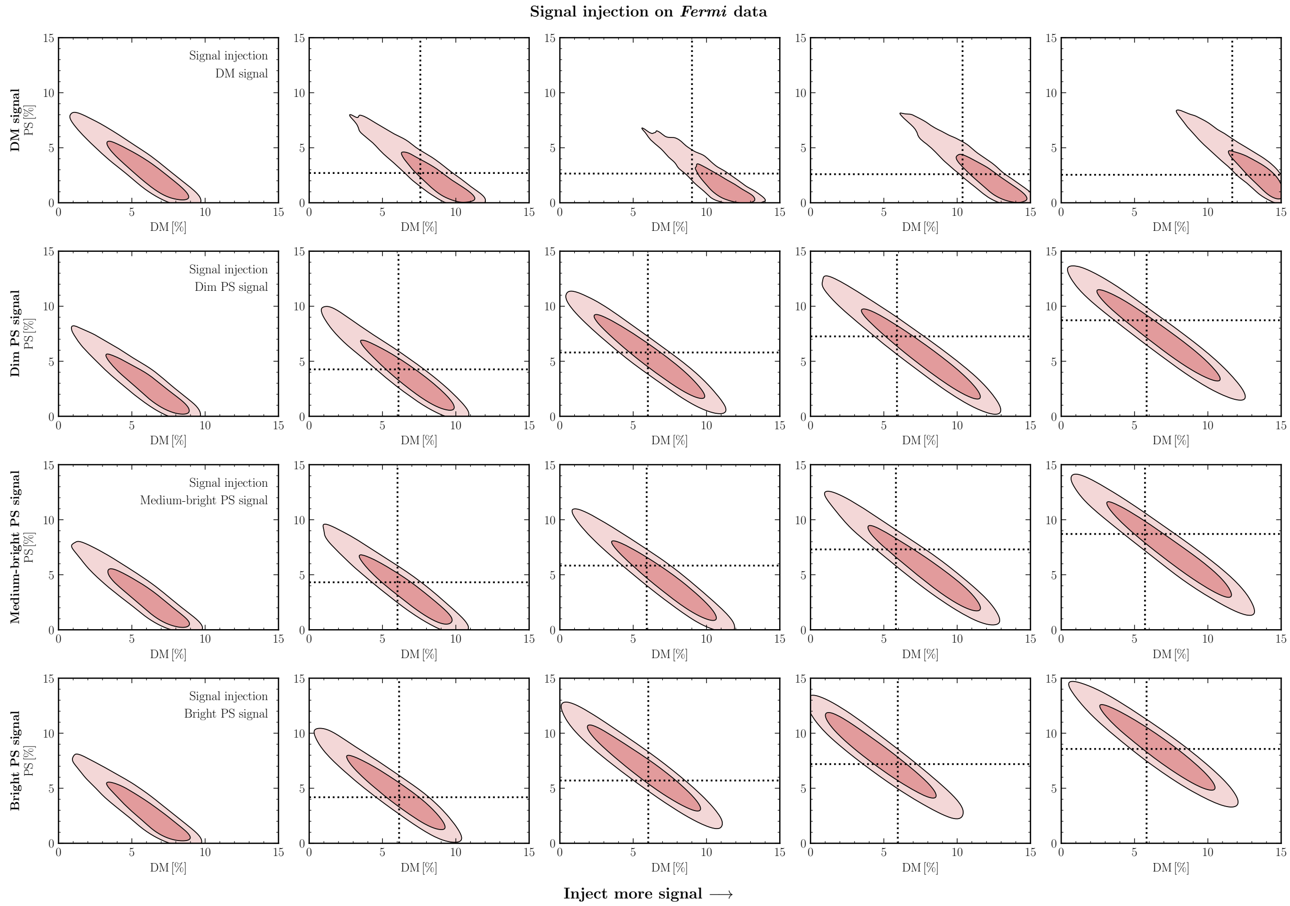

FIG. 4. Joint posterior for the flux fraction of PS-like and DM-like emission when an artificial DM signal is injected onto the real *Fermi* data. The different rows correspond to different signal types, from top to bottom, purely DM, dim PSs (peaking at 5 expected counts per PS), moderately-bright PSs (peaking at 10 expected counts per PS), and bright PSs (peaking at 20 expected counts per PS). The leftmost panels shows the fiducial analysis on *Fermi* data, with subsequent panels showing results with progressively larger signals injected onto the data. The dotted lines show the expected total emission on top of the median initial inferred flux. The additional injected DM and PS signals are seen to correctly reconstructed within the respective posterior confidence limits in all cases.

V. SUSCEPTIBILITY TO MISMODELING

A key challenge in γ-ray analyses of the Galactic Center is that associated with effects of mismodeled signal and background templates [30, 32–35]. In this section we assess the susceptibility of our simulation-based inference pipeline to these systematics, exploring the effect of background and signal mismodeling in turn.

Diffuse foreground mismodeling

In order to test the effect of large-scale foreground mismodeling, we construct a data-driven model of such mismodeling and assess the ability of our method to recover either smooth or PS-like emission in the face of such mismodeling. Following Ref. [90], we perform a Poissonian template analysis on the *Fermi* dataset x, modulating the diffuse model template T_{dif}, which describes the bremsstrahlung and neutral pion decay components of diffuse Model O, by an (exponentiated) Gaussian process (GP):

$$x \sim \mathrm{Pois}\left(\sum_{i\neq\mathrm{dif}} A_i T_i + \exp\left(f\right) A_{\mathrm{dif}} T_{\mathrm{dif}}\right). \qquad (16)$$

The other Poissonian templates T_i, including a GCE DM template and the Inverse Compton component of the diffuse foreground model, are treated as before using an overall normalization factor A_i. $f \sim \mathcal{N}(m, K)$ is the GP component with mean m set to zero, and the covariance K described through the Matérn kernel with smoothness parameter $\nu = 5/2$. We refer to Ref. [90] for further details of the analysis, as well a validation of the GP-augmented template fitting pipeline on simulated data.

The median Gaussian process describing the multiplicative mismodeling obtained relative to the real *Fermi* data when using our fiducial diffuse Model O is shown

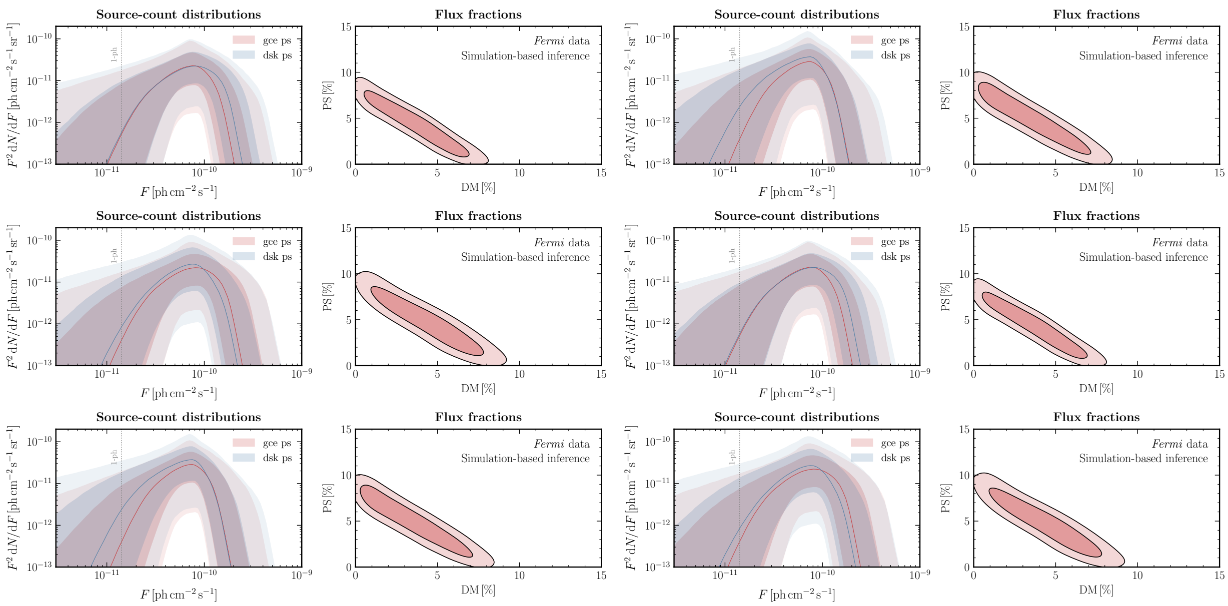

FIG. 5. Joint posterior for flux fraction of PS-like and DM-like emission on real *Fermi* data for different diffuse models *(left)* and ROI sizes *(right)*. In varying diffuse models, the fiducial Model O (red) is compared with results obtained using Models A (blue) and F (green). Although the overall GCE flux is seen to varying by up to a factor of ~ 2 between diffuse models, no evidence for PS-like emission is seen. As seen in the right panel, results remain consistent for smaller ROI sizes.

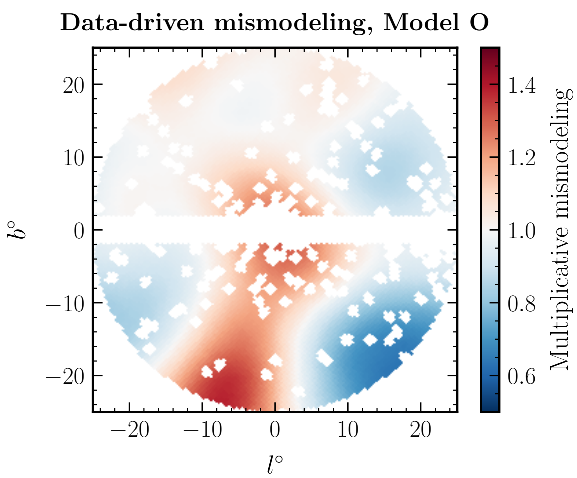

FIG. 6. The median Gaussian process description of multiplicative mismodeling associated with diffuse foreground Model O when applied to the real *Fermi* data.

in Fig. 6. The most severe mismodeling is inferred to be concentrated in the central and southern regions of the fiducial ROI. We modulate the bremsstrahlung and neutral pion decay-tracing components of Model O using samples drawn from the inferred Gaussian process, producing simulated data with the aim of mocking up the scenario of large-scale diffuse mismodeling. These simulated samples are then analyzed with our standard pipeline, using the unmodulated Model O to model the diffuse emission.

The results of this test are shown in Figs. 7 and 8, for simulated samples consisting of purely DM-like and PS-like emission, respectively. It can be seen that while large-scale mismodeling can distort the total flux attributed to either DM or PS-like emission, preference for the true underlying nature of the signal remains robust in either case. The marginalized PS flux as well as the inferred SCD is consistent with the underlying truth in all cases. The DM flux tends to be overestimated in either case however, which may be attributed to the centrally-concentrated mismodeling as seen in Fig. 6. This is also reflected in the fact that the inferred inverse Compton component flux tends to be underestimated, with the residual flux attributable to the DM template. [SM: Specify what kind of PSs are injected.]

Signal mismodeling

Besides mismodeling of the diffuse foreground emission, another major potential concern when characterizing the GCE PS population is associated with mismodeling of the signal emission itself. In particular, as pointed out in Refs. [34, 35], a North-South asymmetry in a putative dark matter signal, if unaccounted for, could lead to the inference of a spurious PS population associated with the purely smooth, asymmetric signal in the traditional NPTF framework. Refs. [34, 35] found preference for such a scenario in real *Fermi* data, with the signal in the Northern hemisphere a factor of $\sim$2 larger than that in the Southern hemisphere if the signal in the two regions is floated separately. [SM: Explain why this

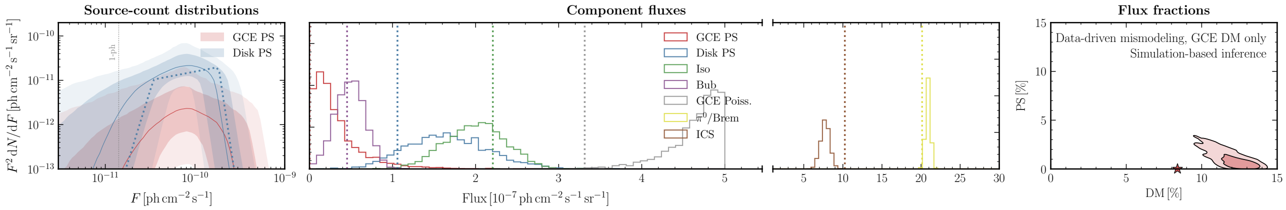

FIG. 7. Same as Fig. 1, but for simulated data where the GCE consists of purely DM-like emission and the diffuse model is modulated by draws from the Gaussian process description of diffuse mismodeling. DM-like emission is inferred in each case, although the magnitude of emission is overestimated as some of the diffuse mismodeling is absorbed into the Poissonian GCE component.

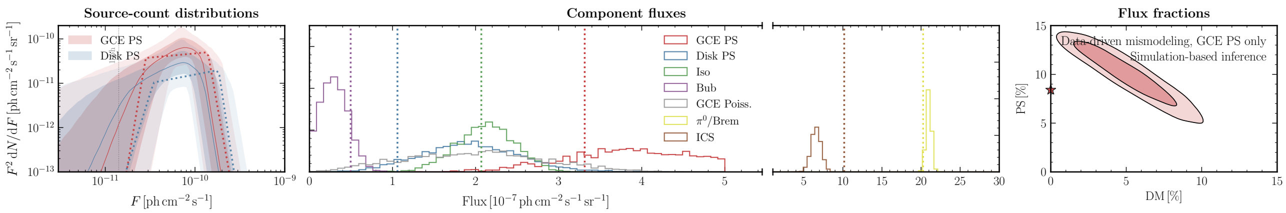

FIG. 8. Same as Fig. 1, but for simulated data where the GCE consists of a PS population peaking at 10 expected photon counts and the diffuse model is modulated by draws from the Gaussian process description of diffuse mismodeling. PS-like emission is inferred in each case, although a non-trivial DM component is inferred as some of the diffuse mismodeling is absorbed into the Poissonian GCE component.

happens.]

We test the impact of a North-South-asymmetric dark matter signal within our framework by running our pipeline on simulated datasets where the dark matter-like signal in the Northern hemisphere of the ROI is 3 times larger than that in the Southern hemisphere, larger than the ~ 2 preference in real data found in Refs. [34, 35]. The results of this test on 3 such simulated realizations is shown in Fig. 9. We see that the presence of a substantially asymmetric DM signal has only a marginal impact on the inferred posteriors, and does not lead to a spurious preference for a PS population as was found in Refs. [34, 35] in the NPTF framework. We attribute this to the fact that the `DeepSphere`-based feature extractor can account for pixel-to-pixel correlations in the γ-ray counts map, and can thus be sensitive to *local* PS-like structures. In contrast, the 1-point PDF-based NPTF framework, being agnostic to the ordering of the pixels, can notice spurious PS-like structures in the "residuals" associated with an asymmetric signal when analyzed with a symmetric template. As done in Ref. [32], we emphasize that the presence of an asymmetry in the GCE signal, if not attributed to diffuse mismodeling, would point towards astrophysical explanations of the GCE since a true dark matter signal would not be expected to be significantly asymmetric.

VI. DISCUSSION AND CONCLUSIONS

In this work, we have leveraged recent advances in neural simulation-based inference in order to characterize a putative point source population that may be responsible the observed *Fermi* Galactic Center Excess. Consistent with Ref [41] which used a Bayesian neural networks and first leveraged the `DeepSphere` graph-based network, our analysis based on conditional posterior density estimation with normalizing flows shows a significantly reduced preference for a γ-ray PS population as the explanation for the GCE compared to previous analyses based on the 1-point PDF or a wavelet decomposition of the Galactic Center photon map. In particular, we find that roughly half of the GCE emission is attributable to a PS population, with the inferred source-count distribution peaking at significantly lower expected photon counts than those found in previous analyses based on the NPTF framework [30], where the SCD is seen to peak just below the threshold for resolution of individual PSs. We have additionally shown our framework to be robust to large-scale diffuse foreground and signal mismodeling of the kind previously discussed in the literature as potential sources of bias. We have used a novel Gaussian Process-based method to construct a data-driven model of large-scale mismodeling. Our conclusions are also robust to the systematic variations we have explored, including variations on the the diffuse foreground model and size of the region of interest. As in any Galactic Center γ-ray analysis, given the poorly understood astrophysical emission in this region, we caution of the potential of unknown systematics, such as small-scale mismodeling on the scale of the size of the LAT point-spread function, to bias the results of our analysis.

Several improvements to the framework presented here are possible. The inclusion of energy-binning informa-

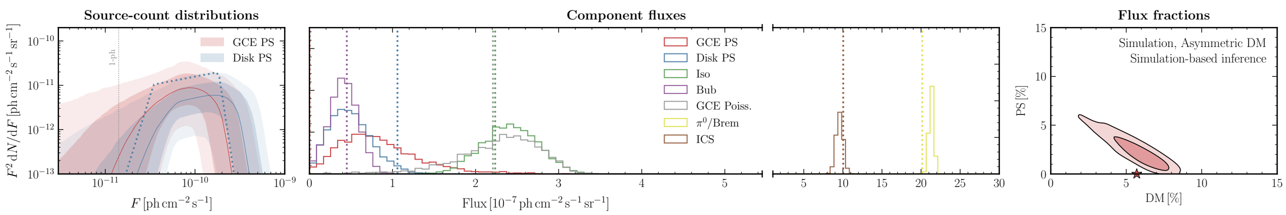

FIG. 9. Same as Fig. 1, but for simulated data where the GCE consists of purely DM-like emission with a North-South asymmetry; the signal in the Northern hemisphere is larger by a factor of 3. The mismodeled signal is seen to have marginal qualitative effect on the recovery of DM-like emission.

tion in the analysis can be implemented by splitting up the data and template maps into individual bins and inputting these as separate channels in the graph-convolutional feature extraction neural network. The use of more complex feature extraction and flow architectures can additionally improve the robustness of our results. While we have considered a simulated-based inference framework based on posterior density estimation with normalizing flows, alternative frameworks based on likelihood-ratio estimation [67–69, 71, 75, 91, 92] or flow-based likelihood estimation [93, 94] can provide complementary ways to characterize the γ-ray PS population in the Galactic Center. Additionaly, the use of sequential active-learning methods [94] and methods that extract additional latent information from the simulator [67–69, 95, 96] can significantly improve the simulation efficiency and allow for extensions to more complex latent spaces, which will be important in particular for an energy-binned analysis and if including additional degrees of freedom into the diffuse foreground model. Since diffuse mismodeling is the largest source of uncertainty in any analysis that aims to characterize the GCE, we also note the possibility of using adversarial learning methods [97] to account for systematic differences between the modeled and real *Fermi* data. Alternatively, generative modeling of the diffuse foreground either in a GP-based data-driven framework or using, *e.g.*, autoencoders trained on an ensembled of plausible diffuse model scenarios, can provide a principled way to account for the large latent space associated with the diffuse foreground. These extensions can lead to a more robust characterization of a putative PS population in the GCE, and we leave their study to future work.

The code used to obtain the results in this paper is available at `https://github.com/smsharma/fermi-flows`.

[SM: Beef up attribution to previous papers on different topics.]

ACKNOWLEDGMENTS

We thank Johann Brehmer, Florian List, Nick Rodd, and Tracy Slatyer for helpful conversations. KC is partially supported by NSF awards ACI-1450310, OAC-1836650, and OAC-1841471, the NSF grant PHY-1505463, and the Moore-Sloan Data Science Environment at NYU. SM is supported by the NSF CAREER grant PHY-1554858, NSF grants PHY-1620727 and PHY-1915409, and the Simons Foundation. This work made use of the NYU IT High Performance Computing resources, services, and staff expertise. This research has made use of NASA's Astrophysics Data System. This research made use of the `astropy` [98, 99], `dynesty` [62], `getdist` [100], `IPython` [101], Jupyter [102], `matplotlib` [103], `MLflow` [104], `nflows` [105], `NPTFit` [47], `NumPy` [106], `pandas` [107], `Pyro` [108], `PyTorch` [109], `PyTorch Geometric` [110], `PyTorch Lightning` [111], `seaborn` [112], `sbi` [113], `scikit-learn` [114], `SciPy` [115], and `tqdm` [116] software packages. We acknowledge the use of the code repository associated with Ref. [41], in particular the associated data products and templates.[2] We acknowledge the use of the use of the code repository associated with Ref. [85], in particular the implementation of the `DeepSphere` graph convolutional layer.[3]

[1] W. B. Atwood *et al.* (Fermi-LAT), Astrophys. J. **697**, 1071 (2009), arXiv:0902.1089 [astro-ph.IM].

[2] L. Goodenough and D. Hooper, (2009), arXiv:0910.2998 [hep-ph].

[3] D. Hooper and L. Goodenough, Phys. Lett. B **697**, 412 (2011), arXiv:1010.2752 [hep-ph].

[4] A. Boyarsky, D. Malyshev, and O. Ruchayskiy, Phys. Lett. B **705**, 165 (2011), arXiv:1012.5839 [hep-ph].

[5] D. Hooper and T. Linden, Phys. Rev. D **84**, 123005 (2011), arXiv:1110.0006 [astro-ph.HE].

[6] K. N. Abazajian and M. Kaplinghat, Phys. Rev. D **86**, 083511 (2012), [Erratum: Phys.Rev.D 87, 129902 (2013)], arXiv:1207.6047 [astro-ph.HE].

[2] `https://github.com/FloList/GCE_NN`
[3] `https://github.com/deepsphere/deepsphere-pytorch`

[7] D. Hooper and T. R. Slatyer, Phys. Dark Univ. **2**, 118 (2013), arXiv:1302.6589 [astro-ph.HE].

[8] C. Gordon and O. Macias, Phys. Rev. D **88**, 083521 (2013), [Erratum: Phys.Rev.D 89, 049901 (2014)], arXiv:1306.5725 [astro-ph.HE].

[9] K. N. Abazajian, N. Canac, S. Horiuchi, and M. Kaplinghat, Phys. Rev. D **90**, 023526 (2014), arXiv:1402.4090 [astro-ph.HE].

[10] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, Phys. Dark Univ. **12**, 1 (2016), arXiv:1402.6703 [astro-ph.HE].

[11] F. Calore, I. Cholis, and C. Weniger, JCAP **03**, 038 (2015), arXiv:1409.0042 [astro-ph.CO].

[12] K. N. Abazajian, N. Canac, S. Horiuchi, M. Kaplinghat, and A. Kwa, JCAP **07**, 013 (2015), arXiv:1410.6168 [astro-ph.HE].

[13] M. Ajello *et al.* (Fermi-LAT), Astrophys. J. **819**, 44 (2016), arXiv:1511.02938 [astro-ph.HE].

[14] T. Linden, N. L. Rodd, B. R. Safdi, and T. R. Slatyer, Phys. Rev. D **94**, 103013 (2016), arXiv:1604.01026 [astro-ph.HE].

[15] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi, and M. Pohl, Nature Astron. **2**, 387 (2018), arXiv:1611.06644 [astro-ph.HE].

[16] H. A. Clark, P. Scott, R. Trotta, and G. F. Lewis, JCAP **07**, 060 (2018), arXiv:1612.01539 [astro-ph.HE].

[17] K. N. Abazajian, JCAP **03**, 010 (2011), arXiv:1011.4275 [astro-ph.HE].

[18] D. Hooper, I. Cholis, T. Linden, J. Siegal-Gaskins, and T. Slatyer, Phys. Rev. D **88**, 083009 (2013), arXiv:1305.0830 [astro-ph.HE].

[19] F. Calore, M. Di Mauro, and F. Donato, Astrophys. J. **796**, 1 (2014), arXiv:1406.2706 [astro-ph.HE].

[20] I. Cholis, D. Hooper, and T. Linden, JCAP **06**, 043 (2015), arXiv:1407.5625 [astro-ph.HE].

[21] J. Petrović, P. D. Serpico, and G. Zaharijas, JCAP **02**, 023 (2015), arXiv:1411.2980 [astro-ph.HE].

[22] Q. Yuan and K. Ioka, Astrophys. J. **802**, 124 (2015), arXiv:1411.4363 [astro-ph.HE].

[23] T. D. Brandt and B. Kocsis, Astrophys. J. **812**, 15 (2015), arXiv:1507.05616 [astro-ph.HE].

[24] A. Gautam, R. M. Crocker, L. Ferrario, A. J. Ruiter, H. Ploeg, C. Gordon, and O. Macias, (2021), arXiv:2106.00222 [astro-ph.HE].

[25] H. Ploeg, C. Gordon, R. Crocker, and O. Macias, JCAP **12**, 035 (2020), arXiv:2008.10821 [astro-ph.HE].

[26] O. Macias, S. Horiuchi, M. Kaplinghat, C. Gordon, R. M. Crocker, and D. M. Nataf, JCAP **09**, 042 (2019), arXiv:1901.03822 [astro-ph.HE].

[27] R. Bartels, E. Storm, C. Weniger, and F. Calore, Nature Astron. **2**, 819 (2018), arXiv:1711.04778 [astro-ph.HE].

[28] M. Di Mauro, Phys. Rev. D **102**, 103013 (2020), arXiv:2010.02231 [astro-ph.HE].

[29] M. Di Mauro, Phys. Rev. D **103**, 063029 (2021), arXiv:2101.04694 [astro-ph.HE].

[30] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue, Phys. Rev. Lett. **116**, 051103 (2016), arXiv:1506.05124 [astro-ph.HE].

[31] R. Bartels, S. Krishnamurthy, and C. Weniger, Phys. Rev. Lett. **116**, 051102 (2016), arXiv:1506.05104 [astro-ph.HE].

[32] M. Buschmann, N. L. Rodd, B. R. Safdi, L. J. Chang, S. Mishra-Sharma, M. Lisanti, and O. Macias, Phys. Rev. D **102**, 023023 (2020), arXiv:2002.12373 [astro-ph.HE].

[33] L. J. Chang, S. Mishra-Sharma, M. Lisanti, M. Buschmann, N. L. Rodd, and B. R. Safdi, Phys. Rev. D **101**, 023014 (2020), arXiv:1908.10874 [astro-ph.CO].

[34] R. K. Leane and T. R. Slatyer, Phys. Rev. Lett. **125**, 121105 (2020), arXiv:2002.12370 [astro-ph.HE].

[35] R. K. Leane and T. R. Slatyer, Phys. Rev. D **102**, 063019 (2020), arXiv:2002.12371 [astro-ph.HE].

[36] R. K. Leane and T. R. Slatyer, Phys. Rev. Lett. **123**, 241101 (2019), arXiv:1904.08430 [astro-ph.HE].

[37] S. K. Lee, M. Lisanti, and B. R. Safdi, JCAP **05**, 056 (2015), arXiv:1412.6099 [astro-ph.CO].

[38] B. Balaji, I. Cholis, P. J. Fox, and S. D. McDermott, Phys. Rev. D **98**, 043009 (2018), arXiv:1803.01952 [astro-ph.HE].

[39] S. D. McDermott, P. J. Fox, I. Cholis, and S. K. Lee, JCAP **07**, 045 (2016), arXiv:1512.00012 [astro-ph.HE].

[40] S. Caron, K. Dijkstra, C. Eckner, L. Hendriks, G. Jóhannesson, B. Panes, R. Ruiz De Austri, and G. Zaharijas, (2021), arXiv:2103.11068 [astro-ph.HE].

[41] F. List, N. L. Rodd, G. F. Lewis, and I. Bhat, Phys. Rev. Lett. **125**, 241102 (2020), arXiv:2006.12504 [astro-ph.HE].

[42] S. Caron, G. A. Gómez-Vargas, L. Hendriks, and R. Ruiz de Austri, JCAP **05**, 058 (2018), arXiv:1708.06706 [astro-ph.HE].

[43] K. Cranmer, J. Brehmer, and G. Louppe, Proceedings of the National Academy of Sciences **117**, 30055 (2020).

[44] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, arXiv preprint arXiv:1912.02762 (2019).

[45] D. Rezende and S. Mohamed, in *International Conference on Machine Learning* (PMLR, 2015) pp. 1530–1538.

[46] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi, "Supplementary material for NPTFit," (2016).

[47] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi, Astron. J. **153**, 253 (2017), arXiv:1612.03173 [astro-ph.HE].

[48] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, Astrophys. J. **622**, 759 (2005), arXiv:astro-ph/0409513.

[49] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996), arXiv:astro-ph/9508025 [astro-ph].

[50] J. F. Navarro, C. S. Frenk, and S. D. White, Astrophys.J. **490**, 493 (1997), arXiv:astro-ph/9611107 [astro-ph].

[51] B. Zhou, Y.-F. Liang, X. Huang, X. Li, Y.-Z. Fan, L. Feng, and J. Chang, Phys. Rev. D **91**, 123010 (2015), arXiv:1406.6948 [astro-ph.HE].

[52] J. Bovy, arXiv e-prints , arXiv:2012.02169 (2020), arXiv:2012.02169 [astro-ph.GA].

[53] Gravity Collaboration, R. Abuter, A. Amorim, M. Bauböck, J. P. Berger, H. Bonnet, W. Brandner, Y. Clénet, V. Coudé Du Foresto, P. T. de Zeeuw, J. Dexter, G. Duvert, A. Eckart, F. Eisenhauer, N. M. Förster Schreiber, P. Garcia, F. Gao, E. Gendron, R. Genzel, O. Gerhard, S. Gillessen, M. Habibi, X. Haubois, T. Henning, S. Hippler, M. Horrobin, A. Jiménez-Rosales, L. Jocou, P. Kervella, S. Lacour, V. Lapeyrère, J. B. Le Bouquin, P. Léna, T. Ott, T. Paumard, K. Perraut, G. Perrin, O. Pfuhl, S. Rabien, G. Ro-

driguez Coira, G. Rousset, S. Scheithauer, A. Sternberg, O. Straub, C. Straubmeier, E. Sturm, L. J. Tacconi, F. Vincent, S. von Fellenberg, I. Waisberg, F. Widmann, E. Wieprecht, E. Wiezorrek, J. Woillez, and S. Yazici, Astron. Astrophys. **625**, L10 (2019), arXiv:1904.05721 [astro-ph.GA].

[54] D. R. Lorimer *et al.*, Mon. Not. Roy. Astron. Soc. **372**, 777 (2006), arXiv:astro-ph/0607640 [astro-ph].

[55] R. T. Bartels, T. D. P. Edwards, and C. Weniger, Mon. Not. Roy. Astron. Soc. **481**, 3966 (2018), arXiv:1805.11097 [astro-ph.HE].

[56] N. L. Rodd and M. W. Toomey, *NPTFit-Sim* (2017).

[57] F. Acero *et al.* (Fermi-LAT), Astrophys. J. Suppl. **218**, 23 (2015), arXiv:1501.02003 [astro-ph.HE].

[58] M. Su, T. R. Slatyer, and D. P. Finkbeiner, Astrophys. J. **724**, 1044 (2010), arXiv:1005.5480 [astro-ph.HE].

[59] G. H. Collin, N. L. Rodd, T. Erjavec, and K. Perez, (2021), arXiv:2104.04529 [astro-ph.IM].

[60] D. Malyshev and D. W. Hogg, Astrophys. J. **738**, 181 (2011), arXiv:1104.0010 [astro-ph.CO].

[61] T. Daylan, S. K. N. Portillo, and D. P. Finkbeiner, Astrophys. J. **839**, 4 (2017), arXiv:1607.04637 [astro-ph.IM].

[62] J. S. Speagle, Monthly Notices of the Royal Astronomical Society **493**, 3132 (2020).

[63] M. Lisanti, S. Mishra-Sharma, L. Necib, and B. R. Safdi, Astrophys. J. **832**, 117 (2016), arXiv:1606.04101 [astro-ph.HE].

[64] J. J. Somalwar, L. J. Chang, S. Mishra-Sharma, and M. Lisanti, Astrophys. J. **906**, 57 (2021), arXiv:2009.00021 [astro-ph.CO].

[65] D. B. Rubin, The Annals of Statistics **12**, 1151 (1984).

[66] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Monthly Notices of the Royal Astronomical Society , stz1960 (2019), arXiv: 1903.00007.

[67] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, Physical Review D **98**, 052004 (2018), arXiv: 1805.00020.

[68] J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, Proceedings of the National Academy of Sciences **117**, 5242 (2020), arXiv: 1805.12244.

[69] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, Physical Review Letters **121**, 111801 (2018), arXiv: 1805.00013.

[70] J. Brehmer and K. Cranmer, arXiv:2010.06439 [hep-ex, physics:hep-ph, physics:physics, stat] (2020), arXiv: 2010.06439.

[71] K. Cranmer, J. Pavez, and G. Louppe, arXiv:1506.02169 [physics, stat] (2016), arXiv: 1506.02169.

[72] K. Cranmer, J. Brehmer, and G. Louppe, arXiv:1911.01429 [cs, stat] (2020), arXiv: 1911.01429.

[73] C. Durkan, I. Murray, and G. Papamakarios, arXiv:2002.03712 [cs, stat] (2020), arXiv: 2002.03712.

[74] D. S. Greenberg, M. Nonnenmacher, and J. H. Macke, arXiv:1905.07488 [cs, stat] (2019), arXiv: 1905.07488.

[75] J. Hermans, V. Begy, and G. Louppe, arXiv:1903.04057 [cs, stat] (2020), arXiv: 1903.04057.

[76] J.-M. Lueckmann, J. Boelts, D. S. Greenberg, P. J. Gonçalves, and J. H. Macke, arXiv:2101.04653 [cs, stat] (2021), arXiv: 2101.04653.

[77] J.-M. Lueckmann, G. Bassetto, T. Karaletsos, and J. H. Macke, arXiv:1805.09294 [cs, stat] (2019), arXiv: 1805.09294.

[78] L. Pacchiardi and R. Dutta, arXiv:2104.03889 [stat] (2021), arXiv: 2104.03889.

[79] G. Papamakarios and I. Murray, arXiv:1605.06376 [cs, stat] (2018), arXiv: 1605.06376.

[80] G. Papamakarios, D. C. Sterratt, and I. Murray, arXiv:1805.07226 [cs, stat] (2019), arXiv: 1805.07226.

[81] S. Wiqvist, J. Frellsen, and U. Picchini, arXiv:2102.06522 [cs, stat] (2021), arXiv: 2102.06522.

[82] D. Zhao, N. Dalmasso, R. Izbicki, and A. B. Lee, (2021).

[83] G. Papamakarios, T. Pavlakou, and I. Murray, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (Curran Associates Inc., Red Hook, NY, USA, 2017) pp. 2335–2344.

[84] M. Germain, K. Gregor, I. Murray, and H. Larochelle, in *International Conference on Machine Learning* (PMLR, 2015) pp. 881–889.

[85] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin, arXiv preprint arXiv:2012.15000 (2020).

[86] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, Astron. Comput. **27**, 130 (2019), arXiv:1810.12186 [astro-ph.CO].

[87] M. Defferrard, N. Perraudin, T. Kacprzak, and R. Sgier, in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019) arXiv:1904.05146.

[88] D. P. Kingma and J. Ba, in *ICLR (Poster)* (2015).

[89] I. Loshchilov and F. Hutter, in *International Conference on Learning Representations* (2019).

[90] S. Mishra-Sharma and K. Cranmer, in *34th Conference on Neural Information Processing Systems* (2020) arXiv:2010.10450 [astro-ph.HE].

[91] B. K. Miller, A. Cole, G. Louppe, and C. Weniger, (2020), arXiv:2011.13951 [astro-ph.IM].

[92] B. K. Miller, A. Cole, P. Forré, G. Louppe, and C. Weniger, (2021), 10.5281/zenodo.5043707, arXiv:2107.01214 [stat.ML].

[93] C. Winkler, D. Worrall, E. Hoogeboom, and M. Welling, arXiv preprint arXiv:1912.00042 (2019).

[94] G. Papamakarios, D. Sterratt, and I. Murray, in *The 22nd International Conference on Artificial Intelligence and Statistics* (PMLR, 2019) pp. 837–848.

[95] J. Brehmer, F. Kling, I. Espejo, and K. Cranmer, Comput. Softw. Big Sci. **4**, 3 (2020), arXiv:1907.10621 [hep-ph].

[96] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, and K. Cranmer, (2018), arXiv:1808.00973 [stat.ML].

[97] G. Louppe, M. Kagan, and K. Cranmer, (2016), arXiv:1611.01046 [stat.ML].

[98] A. M. Price-Whelan *et al.*, Astron. J. **156**, 123 (2018), arXiv:1801.02634.

[99] T. P. Robitaille *et al.* (Astropy), Astron. Astrophys. **558**, A33 (2013), arXiv:1307.6212 [astro-ph.IM].

[100] A. Lewis, (2019), arXiv:1910.13970 [astro-ph.IM].

[101] F. Perez and B. E. Granger, Computing in Science and Engineering **9**, 21 (2007).

[102] T. Kluyver *et al.*, in *ELPUB* (2016).

[103] J. D. Hunter, Computing In Science & Engineering **9**, 90 (2007).

[104] A. Chen, A. Chow, A. Davidson, A. DCunha, A. Ghodsi, S. A. Hong, A. Konwinski, C. Mewald, S. Murching, T. Nykodym, *et al.*, in *Proceedings of the fourth international workshop on data management for end-to-end machine learning* (2020) pp. 1–4.

[105] C. Durkan, A. Bekasov, I. Murray, and G. Papamakar-

ios, "nflows: normalizing flows in PyTorch," (2020).

[106] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Nature **585**, 357 (2020).

[107] W. McKinney, in *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman (2010) pp. 51 – 56.

[108] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, Journal of Machine Learning Research (2018).

[109] A. Paszke *et al.*, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.

[110] M. Fey and J. E. Lenssen, in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).

[111] W. Falcon *et al.*, "Pytorchlightning/pytorch-lightning: 0.7.6 release," (2020).

[112] M. Waskom *et al.*, "mwaskom/seaborn: v0.8.1 (september 2017)," (2017).

[113] A. Tejero-Cantero *et al.*, Journal of Open Source Software **5**, 2505 (2020).

[114] F. Pedregosa *et al.*, Journal of Machine Learning Research **12**, 2825 (2011).

[115] P. Virtanen *et al.*, Nature Methods (2020), https://doi.org/10.1038/s41592-019-0686-2.

[116] C. O. da Costa-Luis, JOSS **4**, 1277 (2019).