

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: May 24, 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 from annihilating dark matter, an explanation in terms of a population of unresolved point sources remains viable. The effort to uncover the origin of the GCE is hampered in particular by a poor understanding of the diffuse emission of Galactic origin, which can lead to spurious residuals that make it difficult to robustly differentiate smooth emission, as expected due to a dark matter origin, from more “clumpy” emission expected from relatively bright, unresolved point sources. We use neural simulation-based inference methods, in particular conditional density estimation with normalizing flows, in order to extract more information from γ -ray maps of the Galactic Center with the aim of characterizing the contribution of unresolved point sources to the GCE. We do not find a preference in favor of a point source origin of the GCE within our framework.

I. INTRODUCTION

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

The *Fermi* γ -ray Galactic Center Excess (GCE), first identified over a decade ago using data from the *Fermi* Large Area Telescope (LAT), is an excess of photons in the Galactic Center with properties—such as energy spectrum and spatial morphology—broadly compatible with that expected due to annihilating DM. The nature of the GCE remains contentious however, with competing explanations in terms of a population of unresolved astrophysical point sources (PSS) remaining viable. For example, analyses of the spatial morphology of the excess have shown it to be more compatible with that tracing a stellar bulge distribution in the Galactic Center. Furthermore, analyses based on the statistics of photons in the Galactic Center have shown the data to prefer a point source origin of the excess. Recent analyses have however pointed out the potential of unknown systematics—such as the poorly understood morphology of the diffuse foreground emission—to bias these conclusions.

Due to the high-dimensional nature of γ -ray data, traditional analyses techniques have resorted to a hand-

crafted reduction of the photon map into summaries such as the probability distribution of photon counts or a wavelet decomposition of the photon map. Novel analysis techniques that can extract more information from the photon maps could do a better job of hedging against unknown systematics in the data compared to traditional analyses based on specific statistics of the map. Machine learning methods in particular have demonstrated promise for general analyses of γ -ray data and in particular focusing on understanding the nature of the GCE. In this paper, we leverage recent developments in the field of simulation-based inference (SBI) in order to weigh in on the nature of the GCE.

This paper is organized as follows. In Sec. II we describe the various components of our analysis method based on simulation-based inference. In Sec. III we describe our simulation pipeline and verify our analysis on simulated data. Section IV presents results on *Fermi* γ -ray data. In Sec. V we explore the susceptibility of the analysis to mismodeling of the background and signal templates, and present systematic variations on our analysis in Sec. IV B. We conclude in Sec. VI.

II. METHODOLOGY

We begin by describing 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 neural networks.

A. Simulation-based inference

Of central interest in parameter estimation is often the probability distribution of a set of parameters of interest

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

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

θ given some data x —the posterior distribution $p(\theta | x)$. Bayes’ theorem can be used to obtain the posterior as $p(\theta | x) = p(\theta) p(x | \theta) / \mathcal{Z}$, where $p(x | \theta)$ is the likelihood and \mathcal{Z} 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, $p(x | \theta) = \int dz p(x | \theta, z)$. 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.

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 known as a forward model. Samples x from the simulator then implicitly define a likelihood, $x \sim p(x | \theta)$. 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 approximate x according to some distance metric. Such methods—usually grouped under the umbrella of Approximate Bayesian Computation (ABC)—are not uncommon in astrophysics and cosmology. Nevertheless, they suffer from several downsides. The curse of dimensionality usually necessitates reduction of data to representative 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 have leveraged advancements in machine learning, in particular the ability of neural networks to extract useful features from high-dimensional data and 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. [1] for a review of recent developments.

B. Conditional density estimation with normalizing flows

In this paper, we approximate the joint posterior $p(\theta | x)$ through a parameterized distribution $\hat{p}_\phi(\theta | s)$ 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, directly models the posterior distribution given a set of samples drawn from simulator according to some prior proposal distribution $p(\theta)$.

We employ normalizing flows, which provide an efficient way of constructing flexible probability distributions. Normalizing flows model the conditional distri-

bution $\hat{p}_\phi(\theta | s)$ 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, to the target distribution:

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

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

$$z_i = (\theta_i - \mu_i) \cdot \exp(-\alpha_i) \quad (2)$$

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

$$\left| \det \left(\frac{\partial f^{-1}}{\partial \theta} \right) \right| = \exp \left(- \sum_i \alpha_i \right) \quad (3)$$

and forward sampling according to Eq. 2. Multiple transformations can be composed together as $f = f_1 \circ f_2 \circ \dots \circ f_K$ in order to model more expressive posteriors,

$$\hat{p}(\theta | x) = \pi_z (f^{-1}(\theta)) \prod_{i=1}^K \left| \det \left(\frac{\partial f_i^{-1}}{\partial z_{i-1}} \right) \right| \quad (4)$$

The log-probability of the posterior can be computed using Eq. 3:

$$\log \hat{p}(\theta | x) = \log [\pi_z (f^{-1}(\theta))] - \sum_{i=1}^K \sum_{j=1}^N \alpha_j^i, \quad (5)$$

which acts as the optimization objective.

C. Learning summary statistics with neural networks

The curse of dimensionality makes it computationally prohibitive to condition the density estimation task on the the raw dataset x *i.e.*, the γ -ray pixel counts map in the region of interest (ROI). Representative summaries $s = s_\phi(x)$ of the data must be extracted in order to enable a tractable analysis. 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 task at hand.

The *DeepSphere* [2, 3] architecture, with a configuration similar to that introduced and employed in Ref. [4], is used and summarized briefly 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.

- Graph construction
- Graph Fourier basis
- Convolutions on graphs and Chebyshev filters
- Specific architecture choices

D. Datasets and the forward model

We use the datasets and templates from Ref. [5] (packaged with Ref. [6]) 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 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 [7] with nside=128.

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. PS populations spatially correlated with (*i*) the GCE, modeled using a squared-NFW profile with inner slope $\gamma = 1.2$ motivated by previous GCE analysis,

$$\rho(r) \propto \frac{1}{(r/r_s)^\gamma (1 + r/r_s)^{3-\gamma}} \quad (6)$$

where $r_s = 20$ kpc is the Milky Way scale radius, and (*ii*) the Galactic disk, modeled as a doubly-exponential profile,

$$n(R, z) \propto \exp(-R/5 \text{ kpc}) \exp(-|z|/1 \text{ kpc}) \quad (7)$$

where R and z are the radial and vertical Galactic cylindrical coordinates. Variations on these choices are explored in the Appendix. 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 [8]. The source-count distribution dN/dS of each PS population is modeled as a doubly-broken power law,

$$\frac{dN_p}{dS} = A \begin{cases} \left(\frac{S}{S_{b,1}}\right)^{-n_1}, & S \geq S_{b,1} \\ \left(\frac{S}{S_{b,1}}\right)^{-n_2}, & S_{b,1} > S \geq S_{b,2} \\ \left(\frac{S_{b,2}}{S_{b,1}}\right)^{-n_2} \left(\frac{S}{S_{b,2}}\right)^{-n_3}, & S_{b,2} > S \end{cases} \quad (8)$$

specified by the breaks $\{S_{b,1}, S_{b,2}\}$, slopes $\{n_1, n_2, n_3\}$, and an overall normalization A . 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 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, described in the baseline case by Model 0 from Ref. [9], (*ii*) DM-like emission following a generalized squared-NFW profile with $\gamma = 1.2$, (*iii*) isotropic emission, (*iv*) resolved PSs from the 3FGL catalog, and (*v*) emission from the *Fermi* bubbles [10]. The latter three templates are obtained from Ref. [5]. The final maps are obtained by combining a Poisson-fluctuated realization of the combined astrophysical templates with the PS maps. The inner regions of the Galactic plane are masked at $|b| < 2^\circ$, and a radial cut $r < 25^\circ$ defines the region of interest for the fiducial analysis. We mask resolved PSs from the 3FGL catalog at a containment radius of 0.8° .

The forward model is thus specified by 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. Simulated samples are produced using parameters drawn from priors motivated by a Poissonian fit to the real *Fermi* data in order to improve sample efficiency. Since PS-like and Poissonian components of the model are exactly degenerate in the limit of dim sources, we place priors on the expected counts contribution per pixel for the GCE-correlated PS-like and Poissonian emission in order to mitigate biases caused by an induced prior preferring one model over the other and preventing the expression of the degeneracy.

E. Optimization and training

10^6 training samples are generated. Models are trained with batch size 128 using the AdamW optimizer with initial learning rate 10^{-3} and weight decay 10^{-5} . Cosine annealing is used to decay the learning rate, training for up to 100 epochs with early stopping if the validation loss has not improved after 10 epochs.

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

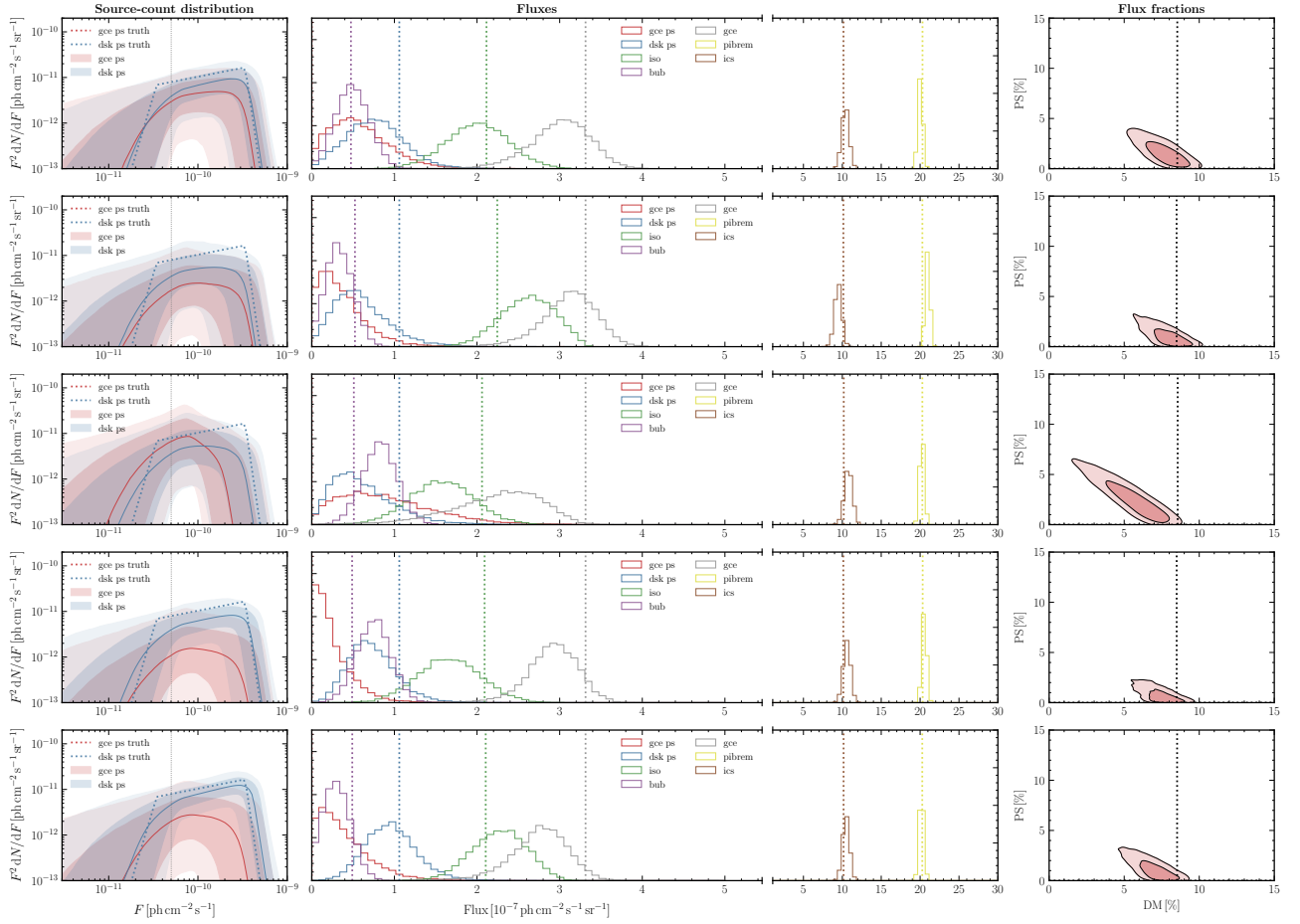


FIG. 1. Results on simulated data where the GCE consists of purely DM-like emission, with different rows corresponding to different simulated realizations. The left column shows the inferred source-count distribution for GCE-correlated (red) and disk-correlated (blue) PS. The middle panel shows the posteriors for the Poissonian templates. The right panel shows the joint posterior on DM-like and PS-like emission. The dotted lines corresponding to the true simulated quantities in each panel. DM-like emission is inferred in each case, with the other posteriors corresponding well to their true simulated values.

III. TESTS ON SIMULATED DATA

We begin by validating our trained model on simulated data. We create simulated datasets using the best-fit astrophysical model obtained from a non-Poissonian fit 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 this background.

Figure 1 shows results on 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 posterior on emission of all components in our model. 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 quantities from which the data was generated are represented by dotted lines. We see that, in all cases shown, we are successfully able to infer the presence of

DM-like emission. Some PS-like emission is inferred in most cases as well, 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. We see that PS-like emission is successfully inferred in each case. Furthermore, the model is able to characterize the source-count of PSs through the source count distribution. Some degeneracy with disk-like source is seen.

IV. RESULTS ON *FERMI* DATA

We finally apply our formalism to real *Fermi* data. 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. IID. The NPTF likelihood implemented in NPTFit from Ref. [6] was used, and the

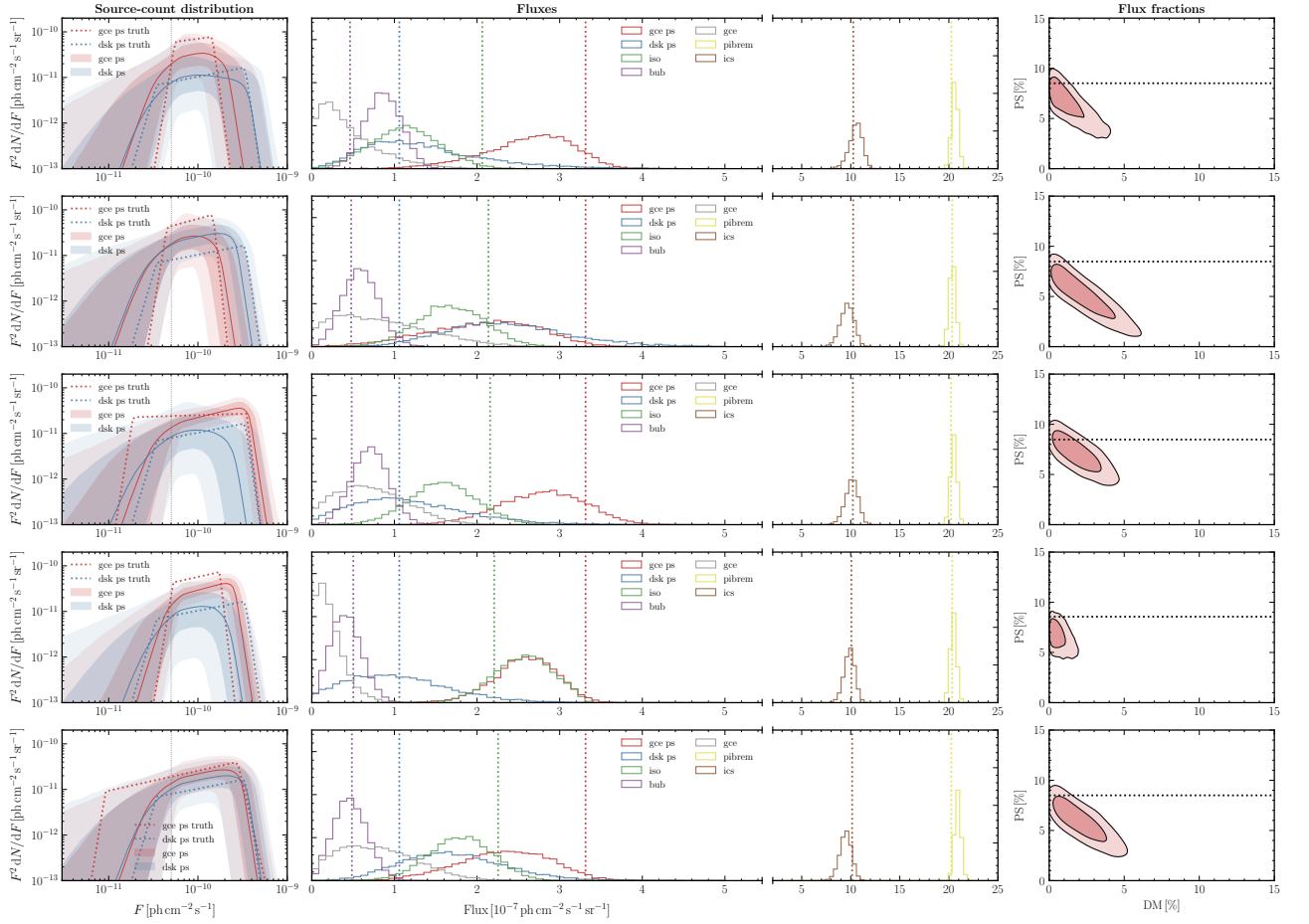


FIG. 2. Same as Fig. 1, but for simulated data where the GCE consists of purely PS-like emission. PS-like emission is inferred in each case, with the other posteriors corresponding well to their true simulated values.

posterior distribution was constructed through nested sampling using **dynesty** with 500 live points. The results of this 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.

The top panel of Fig. 3 shows results using the neural simulation-based analysis pipeline. Although posteriors for the astrophysical background templates are seen to be broadly consistent with the NPTF analysis, No strong preference for PS-like emission is seen in this case. In fact, a near-exact degeneracy between DM-like and PS-like emission is seen.

A. Signal injection test

A crucial self-consistency test is the ability of the analysis to recover an artificial signal injected onto the real γ -ray data. Figure 4 shows the joint posterior for flux fraction of PS-like and DM-like emission when an artificial DM signal is injected onto the real data. The left-most panel shows the fiducial analysis on *Fermi* data, with subsequent panels showing DM of varying normal-

izations injected onto the data. The dotted lines show the expected total emission on top of the median initial inferred flux. The additional DM is seen to be reconstructed correctly.

B. Systematic variations on analysis

We explore several systematic variations of 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. [] 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, no evidence for PS-like emission is seen using the additional diffuse models. A slight preference for smooth emission is seen for Models A and F, and the results between these

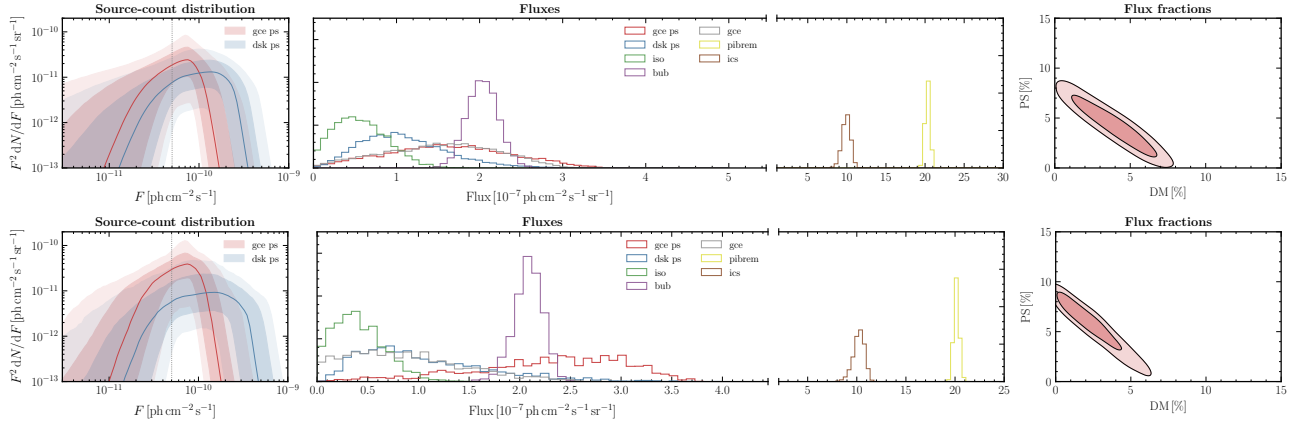


FIG. 3. Results of the fiducial analysis on data, with the top row corresponding to those obtained using the neural simulation-based inference pipeline, and the bottom row using NPTF. While moderate preference for a PS-like origin of the GCE is seen in the case of the NPTF analysis, the SBI analysis finds a nearly perfect degeneracy between smooth and PS-like emission.

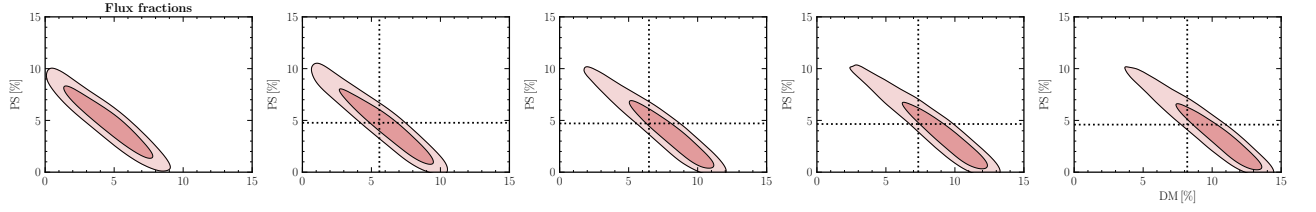


FIG. 4. Joint posterior for flux fraction of PS-like and DM-like emission when an artificial DM signal is injected onto the real data. The leftmost panel shows the fiducial analysis on *Fermi* data, with subsequent panels showing DM of varying normalizations injected onto the data. The dotted lines show the expected total emission on top of the median initial inferred flux. The additional DM is seen to be reconstructed correctly.

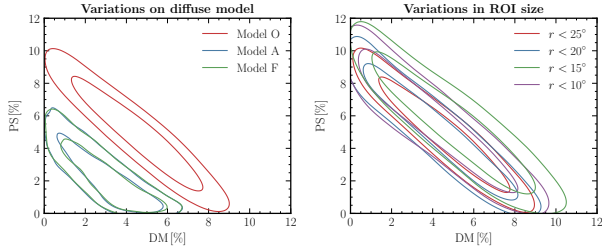


FIG. 5. Joint posterior for flux fraction of PS-like and DM-like emission on real *Fermi* data for different diffuse models (left) and ROI sizes (right). In varying diffuse models, the fiducial Model O (red) is compared with results obtained using Models A (blue) and F (green). Although the overall GCE flux is seen to be 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.

two models and broadly similar.

- *Variations on the ROI size:* Although the GCE signal is concentrated predominantly in the inner 10° , the larger 25° ROI used in this work may better constrain various emission components. However, 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 small ROI sizes— 10° , 15° and 20° . These are seen to be completely consistent with the fiducial analysis in the 25° ROI.

V. SUSCEPTIBILITY TO MISMODELING

A key challenge associated with Galactic Center analyses is that associated with effects of mismodeled signal and background templates. In this section we assess the susceptibility of our simulation-based inference pipeline towards these systematics by constructing a data-driven model of large-scale mismodeling and assessing the ability of our method to recover either smooth or PS-like emission in the face of such mismodeling.

Following Ref. [11], we perform a Poissonian template analysis on the *Fermi* dataset d , modulating the diffuse model template T_{dif} as described by the bremsstrahlung and neutral pion decay components of diffuse Model O by an (exponentiated) Gaussian process (GP):

$$d \sim \text{Pois} \left(\sum_{i \neq \text{dif}} A_i T_i + \exp(f) A_{\text{dif}} T_{\text{dif}} \right). \quad (9)$$

The other Poissonian templates T_i are treated as before

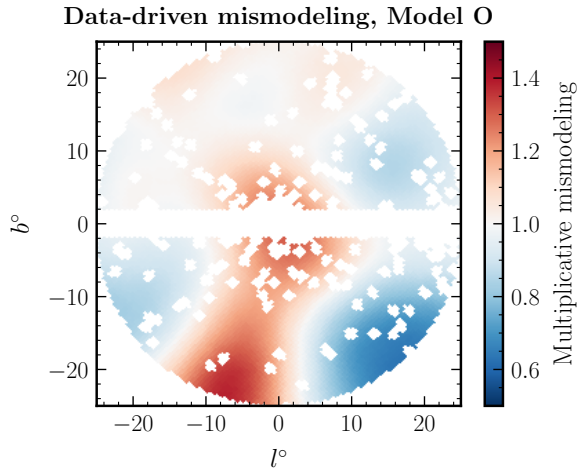


FIG. 6. The median Gaussian process description of multiplicative mismodeling associated with diffuse foreground Model O when applied to the real *Fermi* data.

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. [11] for further details of the analysis, as well a validation of the 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 in Fig. 6. The most severe mismodeling is seen 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 by this map, producing simulated data with the aim of mocking up large-scale mismodeling. These simulated samples are then analyzed with our standard pipeline, using the unmodulated diffuse model.

The results for 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. We note that the test with data-driven mis-

modeling tends to overestimate the DM flux, which may be attributed to the centrally-concentrated mismodeling as seen in Fig. 6.

VI. DISCUSSION AND CONCLUSIONS

In this work, we have leveraged recent advances in neural simulation-based inference in order to weigh in on the nature of the *Fermi* Galactic Center Excess. Consistent with the previous machine learning-based analysis of Ref [4], our analysis based on conditional posterior density estimation with normalizing flows shows no substantive preference for a PS-like origin for a GCE signal across the systematic variations considered. We caution however of the potential of unknown systematics, such as small-scale diffuse mismodeling, mismodeling of the GCE signal itself, and unmodeled PS populations, to bias the conclusions of our study.

The code used to obtain in this paper is available at <https://github.com/smsharma/sbi-fermi>.

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* [12, 13], *dynesty* [14], *IPython* [15], *Jupyter* [16], *matplotlib* [17], *mlflow*, *NPTFit* [6], *NumPy* [18], *pandas* [19], *Pyro*, *PyTorch* [20], *PyTorch Geometric* [21], *PyTorch Lightning* [22], *seaborn* [23], *sbi* [24], *scikit-learn* [25], *SciPy* [26], and *tqdm* [27] software packages. We acknowledge the use of the code repository associated with Ref. [4], in particular the associated data products and templates.²

-
- [1] K. Cranmer, J. Brehmer, and G. Louppe, *Proceedings of the National Academy of Sciences* **117**, 30055 (2020).
 - [2] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin, *arXiv preprint arXiv:2012.15000* (2020).
 - [3] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier,

Astron. Comput. **27**, 130 (2019), [arXiv:1810.12186](https://arxiv.org/abs/1810.12186) [[astro-ph.CO](https://arxiv.org/abs/1810.12186)].

- [4] F. List, N. L. Rodd, G. F. Lewis, and I. Bhat, *Phys. Rev. Lett.* **125**, 241102 (2020), [arXiv:2006.12504](https://arxiv.org/abs/2006.12504) [[astro-ph.HE](https://arxiv.org/abs/2006.12504)].
- [5] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi, “*Supplementary material for NPTFit*,” (2016).
- [6] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi, *Astron. J.* **153**, 253 (2017), [arXiv:1612.03173](https://arxiv.org/abs/1612.03173) [[astro-ph.HE](https://arxiv.org/abs/1612.03173)].

² https://github.com/FloList/GCE_NN

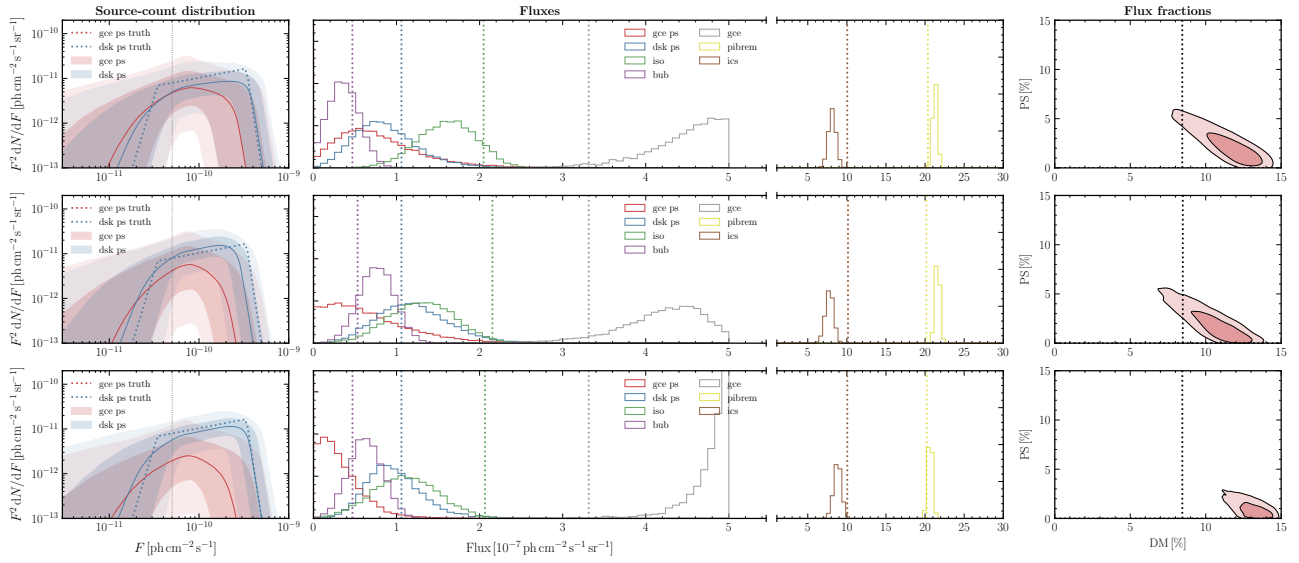


FIG. 7. Same as Fig. 1, but for simulated data where the GCE consists of purely DM-like emission. DM-like emission is inferred in each case, with the other posteriors corresponding well to their true simulated values.

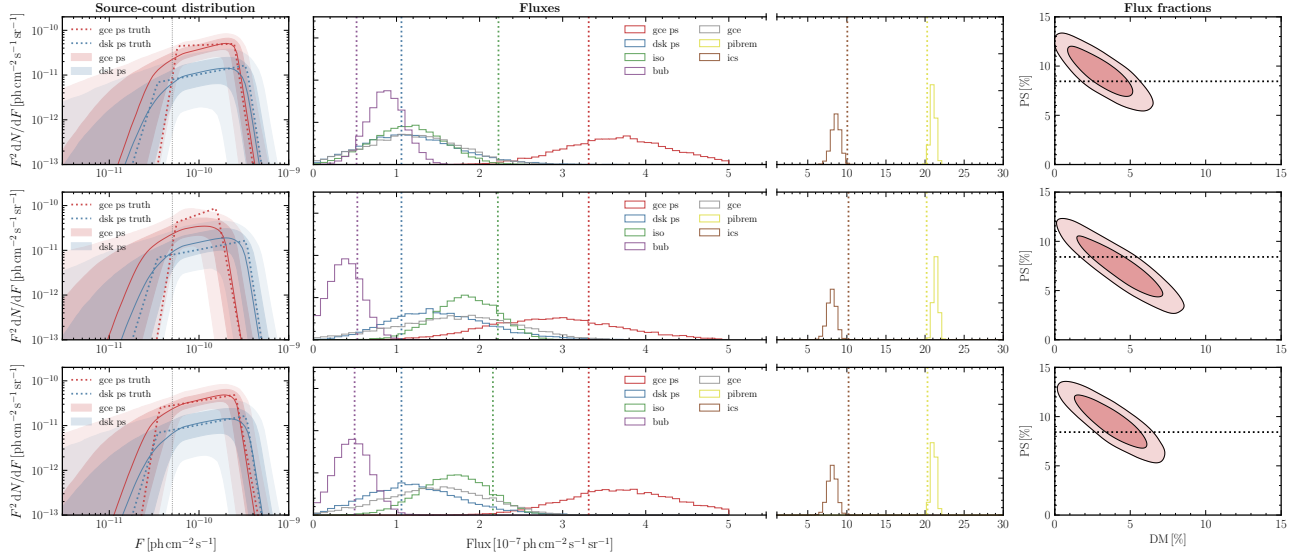


FIG. 8. Same as Fig. 1, but for simulated data where the GCE consists of purely PS-like emission. PS-like emission is inferred in each case, with the other posteriors corresponding well to their true simulated values.

- [7] 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](#).
- [8] N. L. Rodd and M. W. Toomey, *NPTFit-Sim* (2017).
- [9] 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 \[astro-ph.HE\]](#).
- [10] M. Su, T. R. Slatyer, and D. P. Finkbeiner, *Astrophys. J.* **724**, 1044 (2010), [arXiv:1005.5480 \[astro-ph.HE\]](#).
- [11] S. Mishra-Sharma and K. Cranmer, in *34th Conference on Neural Information Processing Systems* (2020) [arXiv:2010.10450 \[astro-ph.HE\]](#).
- [12] A. M. Price-Whelan *et al.*, *Astron. J.* **156**, 123 (2018), [arXiv:1801.02634](#).
- [13] T. P. Robitaille *et al.* (Astropy), *Astron. Astrophys.* **558**, A33 (2013), [arXiv:1307.6212 \[astro-ph.IM\]](#).
- [14] J. S. Speagle, *Monthly Notices of the Royal Astronomical Society* **493**, 3132 (2020).
- [15] F. Perez and B. E. Granger, *Computing in Science and Engineering* **9**, 21 (2007).
- [16] T. Kluyver *et al.*, in *ELPUB* (2016).
- [17] J. D. Hunter, *Computing In Science & Engineering* **9**, 90 (2007).
- [18] S. van der Walt, S. C. Colbert, and G. Varoquaux, *Computing in Science and Engineering* **13**, 22 (2011), [arXiv:1102.1523 \[cs.MS\]](#).
- [19] W. McKinney, in *Proceedings of the 9th Python in Sci-*

- ence Conference, edited by S. van der Walt and J. Millman (2010) pp. 51 – 56.
- [20] 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.
 - [21] M. Fey and J. E. Lenssen, in *ICLR Workshop on Representation Learning on Graphs and Manifolds* (2019).
 - [22] W. Falcon *et al.*, “Pytorchlightning/pytorch-lightning: 0.7.6 release,” (2020).
 - [23] M. Waskom *et al.*, “mwaskom/seaborn: v0.8.1 (september 2017),” (2017).
 - [24] A. Tejero-Cantero *et al.*, *Journal of Open Source Software* **5**, 2505 (2020).
 - [25] F. Pedregosa *et al.*, *Journal of Machine Learning Research* **12**, 2825 (2011).
 - [26] P. Virtanen *et al.*, *Nature Methods* (2020), <https://doi.org/10.1038/s41592-019-0686-2>.
 - [27] C. O. da Costa-Luis, *JOSS* **4**, 1277 (2019).