Non-parametric γ-ray modeling with Gaussian processes and variational inference

Siddharth Mishra-Sharma, Kyle Cranmer
New York University
{sm8383, kyle.cranmer}@nyu.edu

Abstract

Uncertainties in modeling diffuse emission of Galactic origin can seriously bias the characterization of γ-ray data, particularly in the Inner Milky Way where such emission makes up $\sim$ 80–90% of the observed photon counts at $\sim$GeV energies. We introduce a novel class of methods that use Gaussian processes and variation inference to build flexible background and signal models for γ-ray analyses with the goal of enabling a more robust interpretation of the make-up of the γ-ray sky, particularly focusing on characterizing potential signals of dark matter in the Galactic Center.

1 Introduction

The nature of dark matter remains a major persisting mystery in particle physics and cosmology today. One of the primary avenues to search for dark matter is through astrophysical indirect detection—looking for visible byproducts of dark matter annihilation or decay in dark matter-rich regions of the sky (see Refs. [44, 29] for a review). Indeed such a putative signal was identified in the inner regions of the Milky Way—the Galactic Center—over a decade ago using γ-ray data from the *Fermi*-LAT space telescope [12]. The origin of this signal, known as the Galactic Center Excess (GCE), has been hotly debated and remains contentious. The spatial and spectral properties of the signal were shown early on to be compatible with that expected from dark matter annihilation [18, 9]. More recently, the statistical properties of the signal were shown to prefer an explanation in terms of an unresolved population of γ-ray point sources (PSs) rather than annihilating dark matter using the statistics of 1-point photon counts [27] and a wavelet decompositions of the GCE signal [4]. The spatial morphology of the signal was subsequently also shown to prefer an astrophysical explanation [31, 32, 5].

Robustly characterizing the GCE signal is however complicated by the fact that the Galactic Center region contains a large background of emission of diffuse Galactic origin sourced by cosmic rays interacting with the gas, dust, and charged particle populations in the Milky Way disk. While this Galactic background emission can be modeled using cosmic-ray propagation codes such as Galprop [46, 47] and Dragon [10], uncertainties in the properties of 3-D cosmic-ray transport, as well as the underlying distribution of interstellar gas and dust means that large uncertainties in the determination of this background contribution remain.

While the existence of the GCE has been shown to be generally robust to reasonable variation of the modeled diffuse emission [9, 8, 3, 28], its statistical interpretation as originating from a population of unresolved PSs can be be susceptible to mismodeling of the Galactic diffuse emission [27, 23], to the extent of potentially mischaracterizing a dark matter signal as arising from a population of PSs. Robustly characterizing the origin of the GCE therefore requires an improved understanding and modeling of the Galactic diffuse contribution. Complementary to building better models of this emission are methods that aim to augment existing models by introducing extra spatial and/or spectral degree of freedom. In particular, Ref. [45, 5] used regularized likelihoods to build adaptive templates for the Galactic diffuse emission. More recently, Ref. [7] showed that many of the issues associated

Preprint. Under review.

with imperfect background modeling can be alleviated by marginalizing over the large-scale structure of the diffuse model in the basis of spherical harmonics. In this work, we present a complementary approach that uses Gaussian processes and recent advances in variation inference to build flexible models of the Galactic diffuse emission with the goal of robustly characterizing the make-up of the γ-ray sky.

2 Model and inference

Template regression Template regression is a standard technique in astrophysics and cosmology where spatially and/or spectrally binned data is described through a set of spatial templates, binned the same way as the data, each representing the contribution of a particular modeled component, either signal or background. For counts data, Poisson template regression is often employed, where the data is assumed to be a Poisson realization of the modeled emission. Focusing on a single spectral bin with p indexing spatial bins, we have the pixel-wise counts data $d^p \sim \mathrm{Pois}\left(\mu^p(\boldsymbol{\theta})\right)$, where $\boldsymbol{\theta}$ represents the template parameters. Most commonly, the template parameters correspond to overall normalizations A_i of the individual spatial templates T_i^p, $\mu^p(\boldsymbol{\theta}) = \sum_i A_i T_i^p$. This framework is often employed to search for the evidence of a particular signal (*e.g.*, emission of dark matter origin) in counts data while accounting for uncertainties in the normalizations of background template parameters in either a Bayesian [17] of frequentist [30] setting.

Beyond Poisson template regression, methods based on the 1-point PDF of photon counts can model the contribution of populations of unresolved PSs to the counts data, where each PS is too dim to be resolved individually but the collective emission from the population can be detected statistically [33, 26]. These methods can be used in a model comparison setting to differentiate between PS-like ('clumpy') and DM-like ('smooth') origins of a signal. While we focus on the simpler case of Poisson regression in this work, extensions to non-Poissonian template fitting are easily admitted by introducing additional parameters characterizing the contribution of unresolved PS populations [34].

Gaussian processes Gaussian processes (GPs) define a prior distribution over the space of functions such that any collection of function values, evaluated at any collection of points, has a multivariate Gaussian distribution, $\left[f\left(x_1\right), \ldots, f\left(x_n\right)\right] \sim \mathcal{N}(m, K)$, where the covariance function $K_{ij} = k\left(x_i, x_j\right)$ controls the inductive biases of the GP model *e.g.,* its smoothness and periodicity, and the mean function m is often set to zero (see Refs. [42, 50] for a review). Here we employ the Matérn kernel, which generalizes the more common exponential quadratic kernel, and fix the smoothness parameter $\nu = 5/2$, which was found to work well in the present context. The lengthscale and variance of the kernel are hyperparameters of the GP. Since we are interested in modeling processes on the celestial sphere, we use the great-circle distance (in angular units) as our distance measure.

Here, the goal is to use Gaussian processes to give additional freedom to certain background templates—in particular, the Galactic diffuse emission template—in order to enable more robust characterization of other γ-ray contributors. Treating the other templates as before using an overall normalization factor, following Ref. [7] we modulate the Galactic diffuse template (notationally subscripted as 'dif') by a Gaussian process so that the overall model is described by

$$d \sim \mathrm{Pois}\left(\sum_{i \neq \mathrm{dif}} A_i T_i + \exp\left(\mathrm{GP}\right) A_{\mathrm{dif}} T_{\mathrm{dif}}\right) \tag{1}$$

where we have left the pixel indices implicit and used an exponential link function to ensure positivity of the zero-mean GP component. We fix the non-GP multiplicative coefficient of the diffuse template A_{dif} to the best-fit value found through a maximum-likelihood fit of the model without the GP component in our region of interest. We note that a setting the GP mean to zero does not imply a zero *predictive* mean, and the GP component has ample freedom to model departures from the mean obtained in the initial fit.

Our ultimate goal is latent function inference, and there is not a distinction between training and test data points as is often the case in applications of GPs to predictive models. In addition to the fact that the log-marginal/predictive likelihoods are not analytically tractable in this setting with a non-Gaussian likelihood, the datasets in question can be relatively large with $n_{\mathrm{pix}} \sim \mathcal{O}(10^4)$ data points. We thus make use of sparse variational Gaussian process (SVGP) methods [41], implemented through GPyTorch [11, 38] and Pyro [6].

Variational inference Our ultimate goal is to characterize the contribution of various templates and models to γ-ray data in parallel to learning the structure of the Gaussian process that models large scale uncertainties in the Galactic diffuse template. Even in the realm of sparse GPs, sampling the posterior distribution of various template parameters in conjunction with learning the GP (hyper)parameters is computationally expensive. Variational inference tackles this issue by approximating the posterior over parameters of interest θ with a simpler parameterized distribution $q_\phi(z)$, known as the variational distribution. The approximate density is then fit by maximizing a lower bound on the marginal log-likelihood, the evidence lower bound, $\mathrm{ELBO} \equiv \mathbb{E}_{q_\phi(z)}\left[\log p_\theta(d, z) - \log q_\phi(z)\right]$. It can be shown [51] that the difference between the ELBO and the log-evidence is given by the KL divergence between the posterior and variational distribution, $\log p_\theta(d) - \mathrm{ELBO} = D_{\mathrm{KL}}\left(q_\phi(z) \| p_\theta(z \mid d)\right)$. Taking stochastic gradient steps in $\{\theta\text{-}\phi\}$ space then results in the variational distribution approximating the 'true' posterior more and more closely [16].

For the Gaussian process component, the GP posterior over a subset $m \leq n_{\mathrm{pix}}$ of inducing points u is approximated using a multivariate Gaussian distribution $q(u) = \mathcal{N}(m_u, K_u)$ with a learned mean m_u and covariance K_u, with the inducing point locations themselves being learnable parameters as well. The strategy of Ref. [15], implemented in GPyTorch [11], is used to marginalize out the inducing points by transforming a variational distribution over the GP values at the inducing points to a variational distribution at the pixel locations x_p, $q(f(x_p)) = \int \mathrm{d}u\, p(f(x_p), u)\, q(u)$.

Besides the variational treatment of the GP component, a parametric form for the variational distribution over the model parameters characterizing the other spatial templates—in this case, the template normalizations—has to be chosen. A common choice is to describe the joint variational distribution over all the parameters as a multivariate Gaussian with learnable mean and covariance, similar to how the inducing point posterior is described. This choice has several drawbacks for the present task—it restricts the form of the template parameter posteriors to (correlated) Gaussian distributions, and is unable to model correlations between the GP component and template parameters. The ability to model such correlations and more complicated posterior distributions may be especially important in applications beyond the Poisson regression case we consider here—one can imagine, *e.g.*, residuals associated to PS populations to be highly correlated with the GP latent function.

We take the following approach to allow for richer posterior distributions and include correlation effects between the GP and template parameter posteriors. Largely following Ref. [41], starting with a unit Gaussian base distribution with diagonal covariance $\mathcal{N}(0, I)$ we apply a series of inverse autoregressive flow (IAF) transformation—normalizing flows containing a masked autoregressive neural networks (NNs). The autoregressive NNs are augmented to take in additional context variables as inputs in order to condition the template posteriors on variables summarizing samples from the GP variational distribution (see Ref. [37] for details). In this work, we use a setup with 4 IAF transformations and autogressive NNs with 2 hidden layers and a number of nodes per layer equal to 10 times the number of input (template) parameters. The per-pixel-normalized dot product of the (exponentiated) GP sample with each spatial template as well as the mean and root variance of the GP draw are used as context variables for conditioning the transformations in order to capture correlations between the GP and template parameters. We note that additional context variables characterizing the GP draw, or even the entire map (or a downgraded version of it), may be used as inputs to capture subtler correlations with the GP component as required.

3 Tests on simulated data

We use simulated *Fermi*-LAT γ-ray data to validate our pipeline using the example dataset and templates provided with Refs. [34, 7], corresponding to 413 weeks of data in the 2–20 GeV energy range. In addition to the γ-ray counts data, templates corresponding to resolved PSs from the *Fermi* 3FGL catalog [1], isotropically-uniform emission, emission from the *Fermi* bubbles [48], emission from dark matter annihilation following an NFW^2 [36, 35] spatial profile, and Galactic diffuse emission modeled using either the `p6v11` model or the more recent Model O [7] are provided. The maps are spatially binned using HEALPix [13, 52] with `nside=128`. A region of interest (ROI) with latitude cut $|b| > 2^\circ$ and radial cut $r < 20^\circ$ is chosen, and resolved PSs from the 3FGL catalog

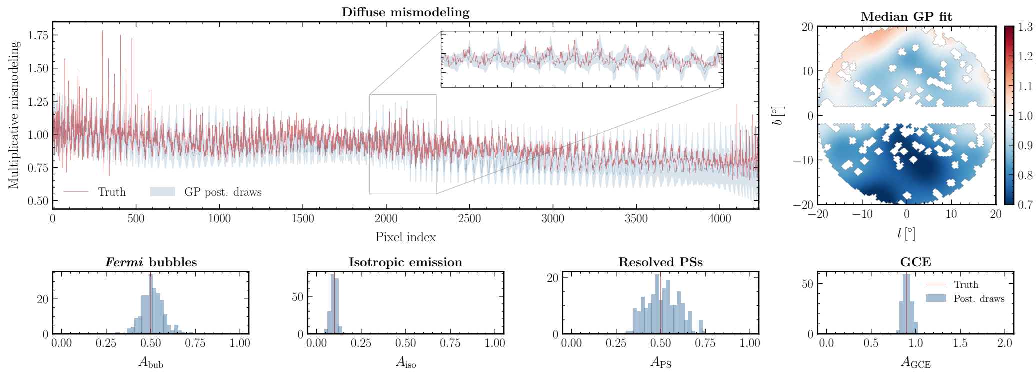

Figure 1: *(Top left)* Pixel-wise 95% highest-posterior density interval (blue band) and median (blue line) of the GP posterior-predictive distribution, along with the true multiplicative modeling between the Model O (in simulation) and `p6v11` (in fit) diffuse background models (red line). The inset shows a zoomed-in region closest to the Galactic Center $(l, b) = (0^\circ, 0^\circ)$. The Gaussian process is seen to faithfully describe uncertainty in the diffuse model on larger scales. *(Top right)* The median inferred map of multiplicative mismodeling in the analysis region of interest. *(Bottom row)* Samples from the posterior-predictive distributions of the Poissonian template normalizations (blue histograms) and the corresponding ground truths (vertical red lines).

masked at 0.8°. This corresponds to a typical Inner Galaxy analysis ROI with $n_{\rm pix} = 4234$ total pixels.

By creating simulated data using one Galactic diffuse model and analyzing it with another diffuse model, we can get a sense of how well we are able to recover the 'ground truth' diffuse mismodeling introduced in the simulation. We create simulated data as a Poisson realization of the sum of templates in the analysis ROI best-fit to the real *Fermi* data, using Model O to model the Galactic diffuse emission. We then analyze it with the `p6v11` diffuse model, using Poisson regression with a GP modulating the diffuse emission as in Eq. (1). The GP hyperparameters—its lengthscale and variance—are fixed to respective mean values obtained by performing an exact GP regression of the diffuse model used in the fit with a suite of other diffuse models from Ref. [2] (not including diffuse Model O, which was used to create the simulated data) in the analysis ROI.

Since each pixel is conditionally independent given the parameters, data subsampling can be used to significantly speed up inference while serving as an additional source of stochasticity in the training. The model is trained using the Adam optimizer [21] with learning rate $\alpha = 10^{-3}$ and other parameters set to their default values, and run for 50,000 iterations with a subsample size of 1500. Fig. 1 shows the posterior-predictive distributions for the template normalization parameters (bottom row) as well as the pixel-wise median and 95% highest-posterior density interval of the exponentiated Gaussian process (top left, shown as a function of HEALPix pixel index within the analysis ROI—pixel indices cross the map left to right, starting from the top), obtained by taking 200 samples from the respective variational distributions conditioned on the simulated data. Ground truth values for the template normalizations and per-pixel multiplicative mismodeling are shown in red. We see that the GP is able to faithfully model the residual large scale mismodeling in the diffuse emission, and the emission modeled by the respective Poissonian templates is correctly recovered. The median inferred GP map is shown in the top right panel.

4 Conclusions and outlook

In this ongoing work, we have introduced a novel method to account for residual mismodeling in analyses of astrophysical counts data with a particular emphases on modeling uncertainties in emission of Galactic diffuse origin in γ-ray analyses. An immediate application of our method is to characterize the statistical nature of the *Fermi* Galactic Center Excess while accounting for large-scale diffuse mismodeling uncertainties, which would require extending our framework to use a likelihood based on the the 1-point PDF instead of Eq. (1). Additionally, given the uncertainty in the spatial distribution of the GCE signal itself and its potential to bias statistical inference of its nature [24, 25], our method can be extended to infer the signal morphology in a data-driven manner.

Acknowledgments and Disclosure of Funding

We thank Lukas Heinrich for collaboration at the early stages of this work. KC is partially supported by NSF grant PHY-1505463m, NSF awards ACI-1450310, OAC-1836650, and OAC-1841471, 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 was also supported through the NYU IT High Performance Computing resources, services, and staff expertise. We thank the *Fermi* collaboration for making publicly available the γ-ray data used in this work. This research has made use of NASA's Astrophysics Data System. This research made use of the Astropy [43, 40], GPyTorch [11], HEALPix [13, 52], IPython [39], Jupyter [22], Matplotlib [19], NumPy [14], Pyro [6], PyTorch [38], SciPy [20], and Seaborn [49] software packages.

References

[1] F. Acero et al. Fermi Large Area Telescope Third Source Catalog. *Astrophys. J. Suppl.*, 218(2):23, 2015.

[2] M. Ackermann et al. The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV. *Astrophys. J.*, 799:86, 2015.

[3] M. Ajello et al. Fermi-LAT Observations of High-Energy γ-Ray Emission Toward the Galactic Center. *Astrophys. J.*, 819(1):44, 2016.

[4] R. Bartels, S. Krishnamurthy, and C. Weniger. Strong support for the millisecond pulsar origin of the Galactic center GeV excess. *Phys. Rev. Lett.*, 116(5):051102, 2016.

[5] R. Bartels, E. Storm, C. Weniger, and F. Calore. The Fermi-LAT GeV excess as a tracer of stellar mass in the Galactic bulge. *Nat. Astron.*, 2(10):819–828, 2018.

[6] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.*, 20:28:1–28:6, 2019.

[7] M. Buschmann, N. L. Rodd, B. R. Safdi, L. J. Chang, S. Mishra-Sharma, M. Lisanti, and O. Macias. Foreground Mismodeling and the Point Source Explanation of the Fermi Galactic Center Excess. *Phys. Rev. D*, 102(2):023023, 2020.

[8] F. Calore, I. Cholis, and C. Weniger. Background model systematics for the Fermi GeV excess. *JCAP*, 1503:038, 2015.

[9] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer. The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. *Phys. Dark Univ.*, 12:1–23, 2016.

[10] C. Evoli, D. Gaggero, A. Vittino, G. Di Bernardo, M. Di Mauro, A. Ligorini, P. Ullio, and D. Grasso. Cosmic-ray propagation with $DRAGON2$: I. numerical solver and astrophysical ingredients. *JCAP*, 02:015, 2017.

[11] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

[12] L. Goodenough and D. Hooper. Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope. 2009.

[13] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman. HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophys. J.*, 622:759–771, 2005.

[14] 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. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020.

[15] J. Hensman, A. Matthews, and Z. Ghahramani. Scalable variational gaussian process classification. *arXiv preprint arXiv:1411.2005*, 2014.

[16] M. Hoffman, D. M. Blei, C. Wang, and J. Paisley. Stochastic Variational Inference. *arXiv:1206.7051 [cs, stat]*, Apr. 2013. arXiv: 1206.7051.

[17] S. Hoof, A. Geringer-Sameth, and R. Trotta. A Global Analysis of Dark Matter Signals from 27 Dwarf Spheroidal Galaxies using 11 Years of Fermi-LAT Observations. *JCAP*, 02:012, 2020.

[18] D. Hooper and L. Goodenough. Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope. *Phys.Lett.*, B697:412–428, 2011.

[19] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[20] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed August 26, 2019].

[21] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014.

[22] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and et al. Jupyter notebooks - a publishing format for reproducible computational workflows. In *ELPUB*, 2016.

[23] R. K. Leane and T. R. Slatyer. Revival of the Dark Matter Hypothesis for the Galactic Center Gamma-Ray Excess. *Phys. Rev. Lett.*, 123(24):241101, 2019.

[24] R. K. Leane and T. R. Slatyer. Spurious Point Source Signals in the Galactic Center Excess. 2 2020.

[25] R. K. Leane and T. R. Slatyer. The Enigmatic Galactic Center Excess: Spurious Point Sources and Signal Mismodeling. 2 2020.

[26] S. K. Lee, M. Lisanti, and B. R. Safdi. Distinguishing Dark Matter from Unresolved Point Sources in the Inner Galaxy with Photon Statistics. *JCAP*, 1505(05):056, 2015.

[27] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue. Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy. *Phys. Rev. Lett.*, 116(5):051103, 2016.

[28] T. Linden, N. L. Rodd, B. R. Safdi, and T. R. Slatyer. High-energy tail of the Galactic Center gamma-ray excess. *Phys. Rev.*, D94(10):103013, 2016.

[29] M. Lisanti. Lectures on Dark Matter Physics. In *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings*, pages 399–446, 2017.

[30] M. Lisanti, S. Mishra-Sharma, N. L. Rodd, B. R. Safdi, and R. H. Wechsler. Mapping Extragalactic Dark Matter Annihilation with Galaxy Surveys: A Systematic Study of Stacked Group Searches. *Phys. Rev.*, D97(6):063005, 2018.

[31] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi, and M. Pohl. Galactic bulge preferred over dark matter for the Galactic centre gamma-ray excess. *Nat. Astron.*, 2(5):387–392, 2018.

[32] O. Macias, S. Horiuchi, M. Kaplinghat, C. Gordon, R. M. Crocker, and D. M. Nataf. Strong Evidence that the Galactic Bulge is Shining in Gamma Rays. *JCAP*, 1909(09):042, 2019.

[33] D. Malyshev and D. W. Hogg. Statistics of gamma-ray point sources below the Fermi detection limit. *Astrophys. J.*, 738:181, 2011.

[34] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi. NPTFit: A code package for Non-Poissonian Template Fitting. *Astron. J.*, 153(6):253, 2017.

[35] J. F. Navarro, C. S. Frenk, and S. D. White. A Universal density profile from hierarchical clustering. *Astrophys.J.*, 490:493–508, 1997.

[36] J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *Astrophys. J.*, 462:563–575, 1996.

[37] B. Paige and F. Wood. Inference Networks for Sequential Monte Carlo in Graphical Models. *arXiv:1602.06701 [stat]*, Mar. 2018. arXiv: 1602.06701.

[38] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.

[39] F. Pérez and B. E. Granger. IPython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, May 2007.

[40] A. Price-Whelan et al. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astron. J.*, 156(3):123, 2018.

[41] J. Quiñonero-Candela and C. E. Rasmussen. A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research*, 6(Dec):1939–1959, 2005.

[42] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[43] T. P. Robitaille et al. Astropy: A Community Python Package for Astronomy. *Astron. Astrophys.*, 558:A33, 2013.

[44] T. R. Slatyer. Indirect Detection of Dark Matter. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics*, pages 297–353, 2018.

[45] E. Storm, C. Weniger, and F. Calore. SkyFACT: High-dimensional modeling of gamma-ray emission with adaptive templates and penalized likelihoods. *JCAP*, 1708(08):022, 2017.

[46] A. W. Strong and I. V. Moskalenko. Propagation of cosmic-ray nucleons in the Galaxy. *Astrophys. J.*, 509:212–228, 1998.

[47] A. W. Strong and I. V. Moskalenko. The GALPROP program for cosmic-ray propagation: new developments. 1999.

[48] M. Su, T. R. Slatyer, and D. P. Finkbeiner. Giant Gamma-ray Bubbles from Fermi-LAT: AGN Activity or Bipolar Galactic Wind? *Astrophys.J.*, 724:1044–1082, 2010.

[49] M. Waskom, O. Botvinnik, D. O'Kane, P. Hobson, S. Lukauskas, D. C. Gemperline, T. Augspurger, Y. Halchenko, J. B. Cole, J. Warmenhoven, J. de Ruiter, C. Pye, S. Hoyer, J. Vanderplas, S. Villalba, G. Kunter, E. Quintero, P. Bachant, M. Martin, K. Meyer, A. Miles, Y. Ram, T. Yarkoni, M. L. Williams, C. Evans, C. Fitzgerald, Brian, C. Fonnesbeck, A. Lee, and A. Qalieh. mwaskom/seaborn: v0.8.1 (september 2017), Sept. 2017.

[50] A. G. Wilson. *Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes*. PhD thesis, 2014.

[51] D. Wingate and T. Weber. Automated Variational Inference in Probabilistic Programming. *arXiv:1301.1299 [cs, stat]*, Jan. 2013. arXiv: 1301.1299.

[52] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski. healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in python. *Journal of Open Source Software*, 4(35):1298, Mar. 2019.