PROCEEDINGS OF SCIENCE

The impact of model realism on interpretations of the Galactic center excess

Christopher Eckner[a,b,*] **for the Fermi-LAT collaboration**

[a] *LAPTh, CNRS, F-74000 Annecy, France*
[b] *LAPP, CNRS, F-74000 Annecy, France*

E-mail: eckner@lapth.cnrs.fr

More than a decade ago the *Fermi* Large Area Telescope detected a GeV gamma-ray excess originating from the Galactic center. The origin of this Galactic Center Excess (GCE) remains a topic of scientific debate. The leading hypotheses explaining its nature are 1) a population of dim and unresolved millisecond pulsars or 2) dark matter annihilation. Each gamma-ray analysis depends on assumptions about background modelling. Hence, the question arises: how does the model used for the analysis affect the consequent interpretations? Are instrumental and background model uncertainties taken into account? When different models lead to different conclusions, there may be a general gap between the model space and reality that influences our conclusions. In this talk, we report the results of our study, showing that DeepEnsemble Networks can robustly detect the background components and the GCE in all model iterations while the predicted emission associated with the background components is consistent with the outcome of a traditional likelihood analysis. However, the reconstructed composition of the GCE is model-dependent. It is likely biased by the presence of a reality gap. We assess the severity of such a gap for each model instance using the One-Class Deep Support Vector Data Description method, and we show that it persists across all iterations. Our work clearly demonstrates the limitations of analyses aiming at characterizing the GCE's nature as soon as a reality gap is an intrinsic issue.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

*Speaker

PoS(ICRC2023)1401

1. Introduction

The central region of the Milky Way (MW) is one of the most promising targets for gamma-ray searches for thermal dark matter (DM) in relative proximity to Earth. Data taken by the *Fermi* Large Area Telescope (LAT) during its first mission year revealed an excess of GeV gamma rays over the known astrophysical backgrounds towards the Galactic centre (GC) [1]. It was dubbed the Galactic centre excess (GCE), which persists until today. While proposed interpretations in terms of conventional astrophysics are by far not excluded, a large fraction of the recent literature on the GCE focuses on two opposing hypotheses: The signature of pair-annihilating DM or a population of unresolved point-like sources, in particular, millisecond pulsars (MSPs) (see [2] and references therein for a review of the history of the GCE and its interpretations).

The properties of the GCE, such as its spatial and spectral properties, are still uncertain due to large uncertainties in the modelling of diffuse background emissions. Recent studies have debated the preferred spatial morphology of the GCE, with some suggesting spherical symmetry [3, 4] and others favouring an asymmetric stellar density profile tracing old stellar populations in the MW's bulge [5–9]. The preference for a stellar bulge component is seen as corroborative evidence for the point-source nature of the GCE. The MW's bulge is a favourable environment for the formation of MSPs. Early studies employed wavelet transforms [10] and non-Poissonian template fitting methods [11] to investigate the point-source nature of the GCE, but the robustness of these techniques has been questioned. Background mis-modelling plays a crucial role in examining the nature of the GCE, and various methods have been proposed to mitigate uncertainties, including nuisance parameters [12], Gaussian processes [13], and adaptive template fitting [6].

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have been proposed to extract small-scale photon clustering in images of the GC [13–16] in order to probe the point-source nature of the excess. These works have demonstrated the potential of CNNs in distinguishing between the two emission origins while their results do not provide strong, robust evidence in favour of either a DM or point-source interpretation.

The challenge in understanding the GCE is reconciling the discrepancy between our imperfect background modelling with the actual reality, known as the "reality gap". We examine the impact of the chosen gamma-ray emission model (γEM) on the final interpretation of the GCE in terms of DM or point-source emission by creating increasingly complex γEMs and developing a fast inference pipeline using Deep Ensemble Networks (DENs) and template-based training data generation. We discover a persistent reality gap between our model space and real data, leading to additional out-of-distribution uncertainty in our results, which may also affect previous GCE studies. Here, we present a selection of critical aspects from our comprehensive work [17].

2. Data selection, model preparation and parameter inference framework

Our analysis is based on Pass 8 (R3) data taken from the 4th of August 2008 to the 2nd of April 2018 (~10 years). We select events in the `ULTRACLEANVETO` event class of the `FRONT` type. We apply the standard zenith angle cut ($< 100^\circ$) as well as the time cut filters `DATA_QUAL==1 && LAT_CONFIG==1`. We define our Region Of Interest (ROI) as a $30^\circ \times 30^\circ$ square centred on the GC binned into square spatial bins of size $0.25^\circ \times 0.25^\circ$ using the plate carrée projection. The

analysis takes into account reconstructed photon energies from 500 MeV to 500 GeV split into 5 energy bins (0.5–1 GeV, 1–2 GeV, 2–7 GeV, 7–20 GeV, 20–500 GeV). We implement the LAT's energy dispersion information for all simulated data utilising the flag `edisp_bins=-1` in all relevant routines of the *Fermi Science Tools*[1] (version 2.0.8).

2.1 Building nested γEMs with increasing complexity

We devise three nested γEMs, labelled as *Model 1*, *Model 2* and *Model 3*, to achieve a set of models with increasing complexity based on a selection of background and signal templates fixed in spatial and partially in spectral morphology. While the subset of model templates describing the GCE (signal components) stays the same in all γEM iterations, the background components describing conventional, known astrophysical contributors to the gamma-ray sky are treated with varying degrees of freedom. *Model 2* can be thought of as the benchmark γEM whereas *Model 1* is a simplistic version of this benchmark case and *Model 3* an extension of it. The gamma-ray sky derived from a model iteration is a linear combination of the templates being part of it, i.e. for each template exists an associated global normalisation parameter scaling the fractional contribution of this template to the overall emission. We train our DENs on data generated from gamma-ray sky images obtained by varying these global normalisation parameters thereby sampling the respective γEM's parameter space. We refer to [17] for an explicit definition of all parameters' prior distributions.

Background template setup. Our astrophysical background setup in *Model 2* consists of the components listed below, which is shown in the left panel of Fig. 1 in terms of expected photon counts in the 1 - 2 GeV energy bin. The other model iterations are discussed afterwards.

(i) Localised gamma-ray sources: Our data selection coincides with the time period used to derive the second iteration (DR2) of the *Fermi*-LAT collaboration's 4FGL catalogue [18]. We adopt the 4FGL-DR2 source positions and spectra of the listed gamma-ray sources to prepare two templates: sources with an angular distance to the GC $\leq 5^\circ$ and those with $\geq 5^\circ$.

(ii) Low-latitude Fermi Bubbles (FBs): This extended diffuse emission component exhibiting an hour-glass shape above and below the Galactic plane at higher latitudes is spectrally and spatially modelled according to their characterisation in [19].

(iii) Isotropic gamma-ray background (ISO): We adopt the characterisation implemented within the *Fermi* Science Tools corresponding to the chosen event class and type of the LAT data.

(iv) Interstellar emission (IE): IE is the consequence of interactions of primary (charged) cosmic rays with the interstellar medium (gas, dust, radiation fields) of the MW. We adopt one of the IE models created for the "1st Fermi LAT Supernova Remnant Catalog" in spatial and spectral morphology [20]. The gamma-ray emission follows the distribution of pulsars in the Milky Way as reported in [21]. The typical height of the cosmic-ray propagation halo is set to $z = 10$ kpc, while the spin temperature of the interstellar medium is taken to be $T_s = 150$ K. It is split into 4 Galactocentric annuli (0–4 kpc, 4–8 kpc, 8–10 kpc and 10–30 kpc) for two gas components (atomic hydrogen HI and molecular hydrogen as traced by carbon monoxide CO), respectively. The associated inverse Compton (IC) emission (not split into annuli) due

[1] https://github.com/fermi-lat/Fermitools-conda

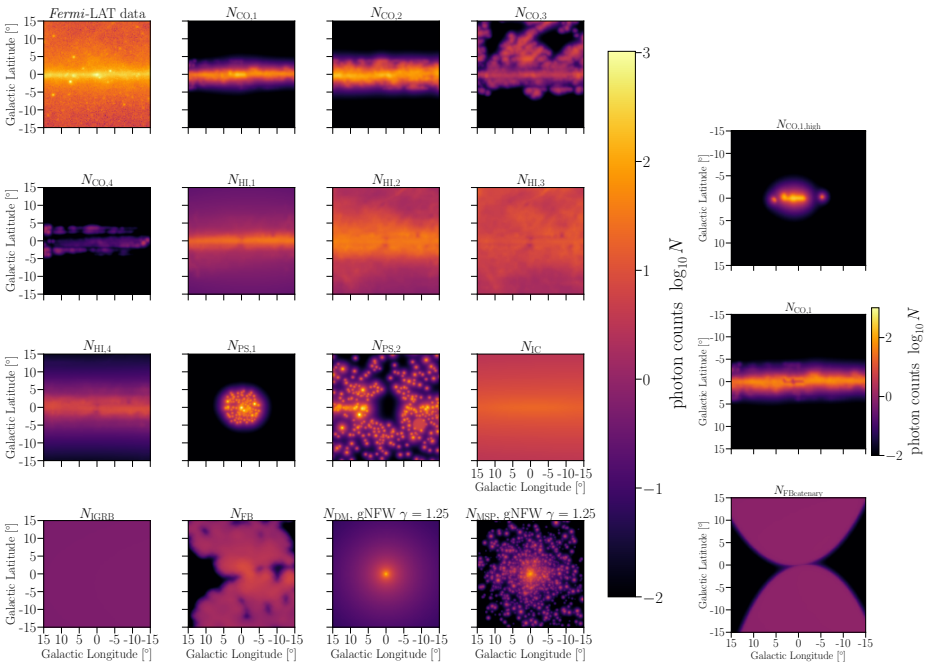

Figure 1: (*Left*:) Spatial morphology of all gamma-ray templates used in our benchmark setup of *Model 2*. The upper panel's first image displays the *Fermi*-LAT data in our ROI between 1 GeV and 2 GeV, which is the same energy bin chosen for the remaining templates. The templates are the output of the Fermi Science Tools routine `gtmodel` and hence display the expected events from the respective flux model for the given *Fermi*-LAT observation time in the infinite statistics limit. The colour indicates the base-10 logarithm of the number of expected gamma-ray events per spatial pixel. We fix the spatial morphology of the signal components to a gNFW profile with $\gamma = 1.25$. (*Right*:) Additional background templates added in the most complex γEM iteration *Model 3*.

to charged leptons interacting with ambient radiation fields is added as a further template. This component yields 9 free parameters in *Model 2*.

Model 1 definition: We reduce the number of background components to two by first performing a maximum likelihood fit (a short description is provided in Sec. 2.2) of *Model 2* to the selected LAT dataset. The obtained best-fitting parameters are used to cast all diffuse background components into a single template as well as the two 4FGL-DR2 templates into a single component for point-like and extended sources.

Model 3 definition: We extend the number of background templates by splitting the CO template of the first ring (0–4 kpc) into two templates: Pixels with at least 25% of the maximal flux value found in the full template and anything below this threshold. We also add a second FBs template featuring uniform emission within the spatial boundaries defined in [22] while adopting the same spectrum assumed for the first FBs template. These templates are shown in the right panel of Fig. 1. Lastly, we introduce a power-law correction δ for the initial IE model spectra; one for each gas component ring (identical for CO and HI), respectively, and one for the IC template. The value of δ may run from -0.3 to 0.3 in discrete steps of 0.1.

Signal template setup. We inject two signal templates into our γEMs: one template describing DM pair-annihilation emission and one containing the cumulative emission from a dim MSP population in the MW's bulge. The spatial profile for both signal components is the same for each generated gamma-ray image from the respective model iteration. We select the generalised Navarro-Frenk-

White (gNFW) $\rho_{\rm gNFW}(r) = \rho_s/(\frac{r}{r_s})^{\gamma}/(1+\frac{r}{r_s})^{3-\gamma}$ with a variable inner slope parameter γ as the underlying spatial morphology. γ is a parameter inferred by the networks with values between 0.8 and 1.3. The DM spectrum is assumed to follow the expectations for pair-annihilation into $b\bar{b}$ final states as implemented in the *Fermi* Science Tools. The MSP template is generated by first drawing random source positions from the respective gNFW profile and adding individual MSPs to the template until a randomly drawn total cumulative MSP luminosity $F_{\rm MSP}$ (100 MeV to 100 GeV) is reached. $F_{\rm MSP}$ is another signal parameter inferred by our neural networks. We refer for more details about the DM and MSP template generation to [17]. We devise two scenarios for the MSP templates: *Scenario A* – all simulated MSPs are below the 3PC [23] detection threshold of the LAT in our ROI; *Scenario B* – even potentially detectable MSPs are added to the template (but do not count towards the $F_{\rm MSP}$ bound).

2.2 Data analysis

Neural network setup. The deep ensemble technique is used in this study to estimate uncertainty in network predictions. To this end, we create a set of five DENs; each with a random initialisation of the network parameters and a random shuffling of the training data following the prescription in [24]. The individual networks are constructed as follows: *(I)* The batch normalisation algorithm is employed to standardise the input. *(II)* Afterwards, we use five convolutional "blocks". One block comprises the following layers: a convolutional layer with X channels and a kernel size of 3, a convolutional layer with $2X$ channels, a max-pooling layer with a pooling value of 2 and the number of channels X of the first convolutional layer is 8. *(III)* The last layer (an image that is 3x3 pixels and has 256 channels) is flattened to a vector with 2304 elements. *(IV)* For every output prediction, a sub-network contains three dense layers of 128, 64 and 32 neurons and then two output neurons: one that predicts the output mean, and one that predicts the logarithm of the output variance.

By assuming that the data come from a Gaussian distribution, we work with the following loss function during the training $\mathcal{L} = \frac{1}{2}\cdot\sum_i^T e^{-\hat{z}_i}\cdot(\hat{y}_i - y_i)^2 + \frac{1}{2}\cdot\hat{z}_i$, where T is the number of output predictions, $\hat{z}$ is the predicted logarithm of the variance, $\hat{y}$ is the predicted output value, and y is the true output value. The final mean predicted value $\hat{y}_*$ and its variance σ_*^2 – arising due to the probabilistic initialisation of the network itself – is quantified via $\hat{y}_* = N^{-1}\sum_j \hat{y}_j$ and $\sigma_*^2 = N^{-1}\sum_j(\sigma_j^2 + \hat{y}_j^2) - \hat{y}_*$, where $N = 5$ in our case.

Maximum likelihood method. To compare the parameter inference of our DENs, we employ the more traditional maximum likelihood method to obtain the best-fitting parameters per model iteration. We use a Poisson likelihood function $\mathcal{L}(\boldsymbol{\mu}|\,\boldsymbol{n}) = \prod_{i,j}\mu_{ij}^{n_{ij}}e^{-\mu_{ij}}/(n_{ij})!$ for binned γEM data (our input templates, except for the MSP one) $\boldsymbol{\mu}$ and observational data $\boldsymbol{n}$ (where the index i runs over the energy bins while the index j enumerates the spatial pixels of our templates).

3. Results

Performance of DENs regarding background and signal parameter inference. In the left panel of Fig. 2 we exemplify the results of our parameter inference pipeline based on DENs and the maximum likelihood method with respect to *Model 2, Scenario B*. In each panel of this figure, the x-axis displays the ground truth of the parameter while the y-axis shows the network's prediction for it. Black data points and error bars refer to predictions on individual images of the training dataset

that the network has never seen during the training (validation dataset). The grey bands represent the 1&2 σ uncertainty of the DENs averaged over the full validation dataset while the black line is the corresponding median prediction. In contrast with respect to the real LAT data, the orange points and error bars denote the results of the maximum likelihood fit whereas the red points with uncertainty estimate are the inference result of the DENs. We find that the maximum likelihood template-based fit is in good agreement with the network's predictions on LAT data regarding the astrophysical background components. Background components that extend further than the considered ROI are harder to constrain like the fourth gas ring (CO and HI) as well as the isotropic gamma-ray background. The spatial distribution parameter of the GCE γ, as well as the DM ($A_{\rm DM}$) and MSP normalisation parameter, are detected consistently *Model 2* and among all γEM iterations considered in this work. The DENs yield a larger uncertainty regarding the components modelling the GCE as compared to the astrophysical components of the γEM.

Inferred composition of the GCE. Our DENs trained on gamma-ray images of the GC inside the range spanned by the respective γEM iteration predict the value of the fraction of the MSP emission to the total GCE luminosity $f_{\rm src} = F_{\rm MSP}/(F_{\rm MSP} + F_{\rm DM})$. While $F_{\rm MSP}$ is a training parameter, $F_{\rm DM}$ is calculated from the predicted value of γ and the DM template normalisation $A_{\rm DM}$ (for details, we refer to [17]). The results per model iteration are displayed in the right panel of Fig. 2. We demonstrate that the inferred composition of the GCE depends on the complexity of the employed γEMwhile the difference between the MSP template scenarios at fixed model iteration is less pronounced. As a general trend, the more complex the γEM becomes, the more the data seem to prefer the MSP origin of the excess. The γEM iterations labelled *X-1D* refer to networks whose only output parameters are $f_{\rm src}$ and its associated uncertainty in contrast to our typical setup that predicts all γEM parameters. In such a setting, we are able to render the predictions of the network very precise as illustrated by the small uncertainty compared to *Model 3*.

Quantifying the model iterations' realism – The Reality Gap. Our findings illustrate that we are able to predict essentially all values of $f_{\rm src}$ between 0 and 1 with the right amount of γEM complexity. Now, we want to quantify how well a γEM iteration can describe the real LAT data, i.e. attributing a degree of realism. To this end, we employ the One-Class Deep Support Vector Data Description method (Deep SVDD) [25], which solely requires us to alter the last layer of our already built DENs turning it into an anomaly detector. Given the input map x, the Deep SVDD network should predict a vector O_n^d of a certain length d filled with an identical scalar value n. Thus, we train on the loss function $s(x) = \left[O_n^d - \text{Model }(x)\right]^2$ to minimise the distance between such a vector and the output of the Deep SVDD. We apply the Deep SVDD to the validation dataset of *Model 3, Scenario B* and show the distance (arbitrary normalisation) of each gamma-ray image of the GC from the output vector O_n^d as the blue histogramme in Fig. 3. The distance of the real LAT data (red line) is much farther from the space spanned by our γEM, indicating that the real data is not encompassed by even our most complex model iteration; a *reality gap*. The green line – being extremely far from both our γEM and LAT data – in the same figure illustrates the distance to pure Gaussian noise confirming the predictive power of our Deep SVDD networks. This result implies that our assessment of $f_{\rm src}$ lacks an additional uncertainty contribution from the fact that the real data is out-of-distribution regarding our γEM spaces.

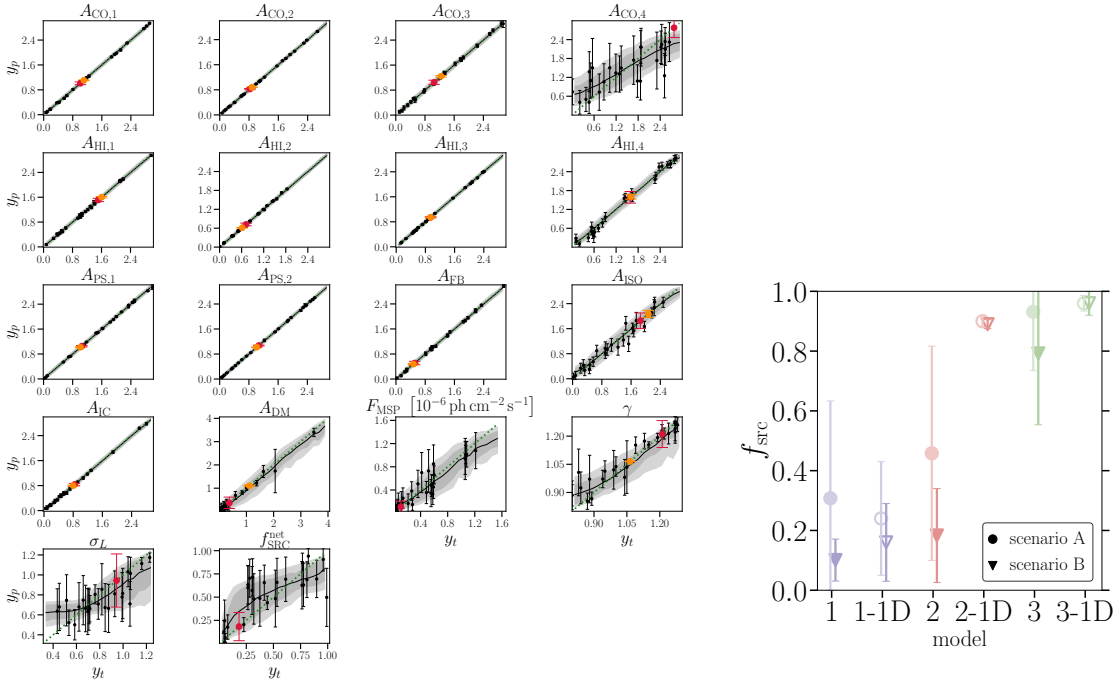

Figure 2: (*Left*:) Per panel: Prediction of the neural network y_P (y-axis) compared to the true parameter value y_t (x-axis) based on the validation data set as part of the training data generated for *Model 2, Scenario B*. A full description of the shown quantities is provided in the main text. (*Right*:) Comparison of the networks' prediction for f_{src} (see main text) per γEM and scenario. 3-1D shows the result of networks trained only on one (f_{src}) parameter (treating the rest as nuisance parameters) when applied to *Model 3*.

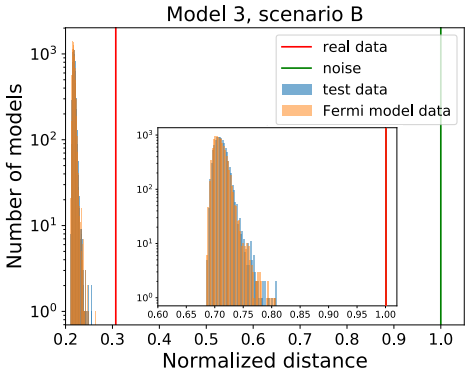

Figure 3: Deep SVDD prediction of the distance between the encoded vector O_n^d for the test data *Model 3B* (blue), the Fermi diffuse model (orange, see [17] for a definition), the real data (red line) and an image of the same dimensions filled only with Gaussian noise (green line).

4. Conclusions

We use DENs to perform a detailed gamma-ray analysis of the complex GC region. The trained networks successfully recover background emission components and detect the presence of the GCE consistently across different γEMs. However, when applying the networks outside of their training domain, the nature of the GCE is not robust, indicating the presence of a reality gap. The reality gap is further quantified using the Deep SVDD architecture, which confirms that the real

data is out-of-distribution even for the most complex γEM.

We provide the first detailed study of the limitations and reality gap of nested γEMsin extracting small-scale photon clustering in gamma-ray data. We highlight the importance of studying and quantifying the reality gap before claiming robust results, particularly when using neural networks trained on theoretical or data-driven models. Our work derives a quantitative statement about the contribution of point sources to the GCE and suggests *Model 3* as the most reliable description of reality. However, the additional uncertainty arising from the real data being outside the investigated γEM space is not captured in the network's predictions. We argue that bridging the reality gap or understanding its limitations and biases is crucial before drawing conclusions about the nature of the GCE; a task necessary in all future characterisation attempts of the GCE.

Acknowledgements. The *Fermi*-LAT Collaboration acknowledges support for LAT development, operation and data analysis from NASA and DOE (United States), CEA/Irfu and IN2P3/CNRS (France), ASI and INFN (Italy), MEXT, KEK, and JAXA (Japan), and the K.A. Wallenberg Foundation, the Swedish Research Council and the National Space Board (Sweden). Science analysis support in the operations phase from INAF (Italy) and CNES (France) is also gratefully acknowledged. This work performed in part under DOE Contract DE-AC02-76SF00515. C. E. acknowledge the financial support from the Slovenian Research Agency (grants P1-0031, I0-0033, J1-1700 and the Young Researcher program). C. E. further acknowledges support by the "Agence Nationale de la Recherche", grant n. ANR-19-CE31-0005-01 (PI: F. Calore). The work of C. E. has been supported by the EOSC Future project which is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020, Grant Agreement 101017536.

References

[1] L. Goodenough and D. Hooper, 0910.2998.
[2] S. Murgia, *Ann. Rev. Nucl. Part. Sci.* **70** (2020) 455–483.
[3] M. Di Mauro, *Phys. Rev. D* **103** (2021) 063029, [2101.04694].
[4] I. Cholis, Y.-M. Zhong, S. D. McDermott and J. P. Surdutovich, *Phys. Rev. D* **105** (2022) 103023, [2112.09706].
[5] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi et al., *Nature Astron.* **2** (2018) 387–392, [1611.06644].
[6] R. Bartels, E. Storm, C. Weniger and F. Calore, *Nature Astron.* **2** (2018) 819–828, [1711.04778].
[7] K. N. Abazajian, S. Horiuchi, M. Kaplinghat, R. E. Keeley and O. Macias, *Phys. Rev. D* **102** (2020) 043012, [2003.10416].
[8] F. Calore, F. Donato and S. Manconi, *Phys. Rev. Lett.* **127** (2021) 161102, [2102.12497].
[9] M. Pohl, O. Macias, P. Coleman and C. Gordon, *Astrophys. J.* **929** (2022) 136, [2203.11626].
[10] R. Bartels, S. Krishnamurthy and C. Weniger, *Phys. Rev. Lett.* **116** (2016) 051102, [1506.05104].
[11] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer and W. Xue, *Phys. Rev. Lett.* **116** (2016) 051103, [1506.05124].
[12] M. Buschmann, N. L. Rodd, B. R. Safdi, L. J. Chang, S. Mishra-Sharma, M. Lisanti et al., *Phys. Rev. D* **102** (2020) 023023, [2002.12373].
[13] S. Mishra-Sharma and K. Cranmer, *Phys. Rev. D* **105** (2022) 063017, [2110.06931].
[14] S. Caron, G. A. Gómez-Vargas, L. Hendriks and R. Ruiz de Austri, *JCAP* **05** (2018) 058, [1708.06706].
[15] F. List, N. L. Rodd, G. F. Lewis and I. Bhat, *Phys. Rev. Lett.* **125** (2020) 241102, [2006.12504].
[16] F. List, N. L. Rodd and G. F. Lewis, *Phys. Rev. D* **104** (2021) 123022, [2107.09070].
[17] S. Caron, C. Eckner, L. Hendriks, G. Jóhannesson, R. Ruiz de Austri and G. Zaharijas, *JCAP* **06** (2023) 013, [2211.09796].
[18] Fermi-LAT collaboration, S. Abdollahi et al., *Astrophys. J. Suppl.* **247** (2020) 33, [1902.10045].
[19] Fermi-LAT collaboration, M. Ackermann et al., *Astrophys. J.* **840** (2017) 43, [1704.03910].
[20] Fermi-LAT collaboration, F. Acero et al., *Astrophys. J. Suppl.* **224** (2016) 8, [1511.06778].
[21] D. R. Lorimer et al., *Mon. Not. Roy. Astron. Soc.* **372** (2006) 777–800, [astro-ph/0607640].
[22] Fermi-LAT collaboration, F. Acero et al., *Astrophys. J. Suppl.* **223** (2016) 26, [1602.07246].
[23] B. Limyansky, in *AAS/High Energy Astrophysics Division*, p. 109.32, Mar., 2019.
[24] B. Lakshminarayanan, A. Pritzel and C. Blundell, *arXiv e-prints* (Dec., 2016) arXiv:1612.01474, [1612.01474].
[25] S. Caron, L. Hendriks and R. Verheyen, *SciPost Physics* **12** (feb, 2022) .