Deep learning models of the discrete component of the Galactic interstellar math xmlns="http://www.w3.org/1998/Math/MathML" display="inline">mi>γ/mi>/math>-ray emission

Alexander Shmakov, Mohammadamin Tavakoli, Pierre Baldi, Christopher M. Karwin, Alex Broughton, and Simona Murgia

Phys. Rev. D **107**, 063018 — Published 17 March 2023

DOI: 10.1103/PhysRevD.107.063018

Deep Learning Models of the Discrete Component of the Galactic Interstellar γ-ray Emission

Alexander Shmakov,[1,*] Mohammadamin Tavakoli,[1,†] Pierre Baldi,[1,‡]
Christopher M. Karwin,[2,3,§] Alex Broughton,[3,¶] and Simona Murgia[3,**]

[1] *Department of Computer Science, University of California, Irvine, California 92697, USA*
[2] *Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634, USA*
[3] *Department of Physics and Astronomy, University of California, Irvine, California 92697, USA*

(Dated: November 4, 2022)

A significant point-like component from the small-scale (or discrete) structure in the H_2 interstellar gas might be present in the *Fermi*–LAT data, but modeling this emission relies on observations of rare gas tracers only available in limited regions of the sky. Identifying this contribution is important to discriminate γ-ray point sources from interstellar gas, and to better characterize extended γ-ray sources. We design and train convolutional neural networks to predict this emission where observations of these rare tracers do not exist, and discuss the impact of this component on the analysis of the *Fermi*–LAT data. In particular, we evaluate prospects to exploit this methodology in the characterization of the *Fermi*–LAT Galactic center excess through accurate modeling of point-like structures in the data to help distinguish between a point-like or smooth nature for the excess. We show that deep learning may be effectively employed to model the γ-ray emission traced by these rare H_2 proxies within statistical significance in data-rich regions, supporting prospects to employ these methods in yet unobserved regions.

The Galactic γ-ray interstellar emission (IE) traces interactions of cosmic rays (CRs) with the interstellar gas and radiation field. In a companion paper [1], we showed that interstellar H_2 gas is more structured and point-like than current IE models assume, and the related γ-ray emission might be a statistically significant component of the *Fermi*–LAT data. If this structure is not adequately captured by the IE model, it can impact the identification of resolved point-like sources as well as the characterization of extended components in the γ-ray sky. We demonstrated that unidentified sources in the the fourth *Fermi*–LAT catalog [2] could indeed be originating from it. In addition, we have argued that this component could artificially inflate the unidentified and/or unresolved point source component in the data and, depending on its morphology, contribute to confounding the interpretation of the the Galactic center (GC) excess observed by *Fermi*–LAT [1, 3–12] (see [13] for a review). Improved modeling of small-scale gas related emission in the *Fermi*–LAT data is therefore necessary to robustly characterize γ-ray sources.

Interstellar H_2 gas is traced indirectly via the emission lines of other molecules that are found concurrently in gas clouds. Carbon monoxide (^{12}CO, or CO hereafter) is used as a proxy, **however, its line emission is typically optically thick in the denser cores of molecular clouds**, and therefore it underestimates the H_2 column density there. **The line emission of rarer isotopologues of** CO **also present in the clouds, such as** ^{13}CO**, remains optically thin to larger column densities** and are therefore more reliable to probe dense H_2 cloud cores, and therefore the H_2 small-scale structure. We briefly summarize the methodology we developed in [1] to model this emission. We employ observations of the $J = 1$–0 transitions of CO and ^{13}CO from the Mopra Southern Galactic Plane CO Survey [14], which cover a 50 square degree region, spanning Galactic longitudes l=300–350° and latitudes $|b| \leq 0.5°$. **As described in [1], the data reduction process by the Mopra team yields data cubes of the brightness temperature for each spectral line, as a function of Galactic coordinates and local standard of rest velocity. As described in detail in the rest of this paper, we employ these data to train machine learning (ML) models to predict** ^{13}CO **in regions of the sky where observations are not available. To ensure the highest quality data in the rare isotopologues, only pixels for which the brightness temperature exceeds the** 1σ **noise provided by Mopra are used. The performance of ML models may by biased by noise, and this conservative** 1σ **noise threshold allows us to retain the majority of the data for training purposes while still rejecting very low quality data.** We calculate H_2 column densities corresponding to Mopra's CO and ^{13}CO, referred to

* ashmakov@uci.edu
† mohamadt@uci.edu
‡ pfbaldi@uci.edu
§ ckarwin@clemson.edu
¶ abrought@uci.edu
** smurgia@uci.edu

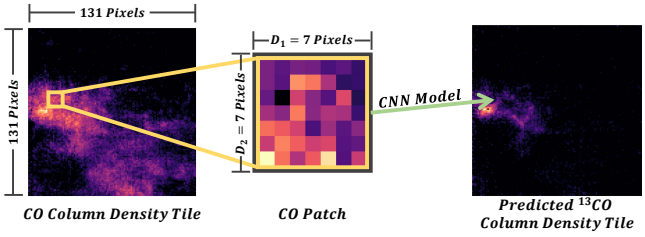

FIG. 1. A Graphical overview of our CO $\rightarrow^{13}$CO modeling process for a single velocity bin. A CO column density **tile** of size $1^\circ \times 1^\circ = 131 \times 131$ pixels (*left*) is split into into overlapping smaller **patches** (*center*) of size $0.053^\circ = 7$ pixels. Patches are fed to a CNN model to produce the corresponding predicted ^{13}CO column density **tile** (*right*) by estimating the ^{13}CO column density at the center of each CO patch.

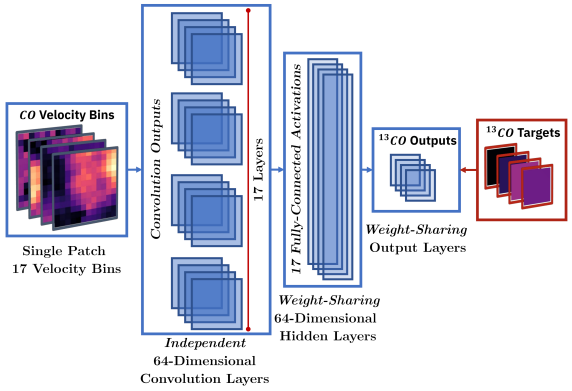

FIG. 2. A diagram of the internal layers of the network. We maintain independent weights for the 17 convolution layers, one for each velocity bin. However, the hidden and output fully-connected layers share a single set of weights and the same layers are applied to every velocity bin.

as $N(\mathrm{H_2})_{\mathrm{CO}}$ and $N(\mathrm{H_2})_{\mathrm{CO13}}$ respectively. **We integrate the column density to produce 17 velocity bins, corresponding to increasing radial distances from the galactic center. The column density computation and integration procedure is described in [1]. The final data is stored as measurements in $1^\circ \times 1^\circ \times 17$ tiles.** With this as input, we build a Modified Map by replacing $N(\mathrm{H_2})_{\mathrm{CO12}}$ with $N(\mathrm{H_2})_{\mathrm{CO13}}$ for pixels where $N(\mathrm{H_2})_{\mathrm{CO13}} > N(\mathrm{H_2})_{\mathrm{CO12}}$. The "Modified Excess Template", which **corresponds to** the small-scale $\mathrm{H_2}$-related γ-ray emission traced by ^{13}CO, is determined from the difference between the Modified Map and the baseline CO map.

The central idea of this paper is to harness ML, in particular deep learning [15], to predict the distribution of ^{13}CO based on the CO observation, and therefore infer the $\mathrm{H_2}$ small-scale structure in regions where ^{13}CO observations do not exist. Since a straightforward and robust analytical mapping between the two distributions is not available, ML would estimate this mapping from data. We train a deep learning model to map between CO and ^{13}CO column densities in the **Mopra region**. To simplify this regression problem and enlarge our effective data set, we subdivide the **tiled** Mopra data into small $D_1 \times D_2$ **patches** of the sky to their respective center points (Fig. 1). The validity of this simplification requires the gas column density to be locally correlated. That is, we assume it is unlikely that column densities outside of our chosen patch will significantly change the ^{13}CO estimate. This simplification greatly reduces the model size and provides us with an effective data set of over 50 million patches when $D_1 = D_2 = 0.0581^\circ$, or 7 pixels at the Mopra resolution. We find that larger patch sizes do not improve model accuracy, justifying this locality assumption (see [16]). We also apply smoothing techniques to eliminate noise from the data (see [16]). The CO$\rightarrow^{13}$CO modeling problem may now be formally written as finding a parameterized function f_θ mapping a source patch $S \in \mathbb{R}^{D_1 \times D_2}$ to an estimate target column density at the center of that same patch $T \in \mathbb{R}$. The source S_i corresponds to Mopra CO column densities and the targets T_i to the corresponding ^{13}CO column densities at the center of the patch (Fig. 1). We evaluate our estimates by splitting the Mopra region into independent sub-regions which we designate for training and testing. We employ different splitting choices (see [16]), but we report on one, the *Alternate Tiled* split, throughout the remainder of this paper. **This split consists of alternating $1^\circ \times 1^\circ$ tiles, training on longitudes $\{301-302, \ldots, 349-350\}$ and testing on longitudes $\{300-301, \ldots, 348-349\}$. This allows us to evaluate the model's performance on neighboring data, reducing the variance between training and testing data distributions.**

We employ convolutional neural networks (CNNs) to model and predict the ^{13}CO column density in a translation-invariant manner. CNNs use neurons with shared connection parameters in order to implement convolution operations that provide the basis for building translation invariant architectures [15]. Each convolution operation is associated with a *kernel*, or filter, corresponding to a set of connection weights that are shared by all the neurons in the corresponding layer. For each of the 17 velocity bins, we apply K learnable convolution filters of size $D_1 \times D_2$, operating on an entire patch and independently for each velocity bin, producing 17 K-dimensional latent vectors for each patch. Afterwards, we apply parametric rectified linear units ($PReLU$) [17–19] with a learnable slope, a batch normalization layer [20], and a random dropout layer [21, 22]. The latent vec-

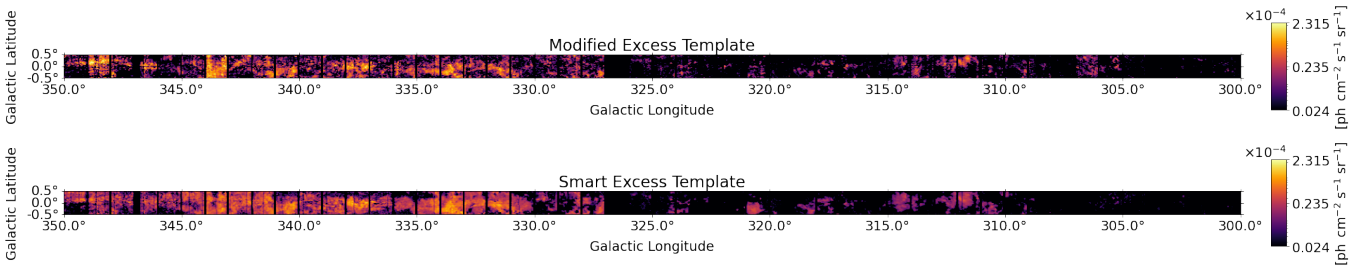

FIG. 3. *Top*: the Modified Excess Template (from [1]). *Bottom*: the Smart Excess Template. The color scale (logarithmic) indicates the γ-ray intensity per 0.03125^2 degree pixel. The first tile to the left (closest to the GC, and covering longitudes $349 - 350$) is testing data, and the neighboring one is training data for the CNN. The interleaving pattern of testing and training tiles is adopted for the entire Mopra region.

tors are then processed by several fully-connected layers, each with their own $PReLU$, batch-normalization, and dropout. Unlike the convolution layer, these hidden layers are shared between velocity bins, learning identical weights for every bin. This design allows spatial components (CNN) of the network to be specialized for each velocity bin while allowing latent higher level components to be shared between bins. The resulting latent vectors are fed through a final fully-connected layer which produces the ^{13}CO concentration point estimates for each patch and velocity bin. A diagram of the network architecture is presented in Fig. 2.

The Mopra dataset contains widely varying column densities from bright to very dim. To effectively learn this high-spread distribution, we model the incidence of photons on the Mopra detector as a Poisson process and use a Poisson log-likelihood loss. We further re-weight this loss based on the ^{13}CO density to elevate the importance of bright pixels, prioritizing accuracy in hot-spots over a slight degradation in the background. This increases the importance of accurate measurement in the bright regions, encouraging the network to focus on the accuracy of these regions first. As the network trains, we anneal this weighting back to uniform (with all target values having equal weighting) in order to minimize any bias introduced by this loss. The design of this loss is guided by the overall goal of finding small angular scale features while limiting the amount of overestimation throughout the Mopra region. A formal description of the loss function is available in [16].

We tune the CNN's hyperparameters using the Sherpa hyperparameter optimization library [23, 24], testing 2000 network variations, scoring each parameterization with Poisson likelihood, and using Gaussian Process optimization to suggest parameters. Our final parameterization is presented in [16], which is evaluated after training the network for 200 iterations. We generate the predicted ^{13}CO density by stitching the network output for every patch in a source image. When performed on a single Nvidia Titan XP GPU, and batching 8192 patches at once, this inference process requires approximately 11 seconds to cover the entire Mopra region.

Following the same procedure as in [1], we employ the CR propagation code GALPROP (v56)[25–35] to calculate γ-ray sky maps in 17 Galactocentric radial bins for the H_2-related emission, where the latter is traced by CO and ^{13}CO. The CO and ^{13}CO related H_2 column densities, $N(H_2)_{CO13}$ and $N(H_2)_{CO}$ respectively, are used to determine the H_2 that is missed in dense regions when only CO is used as a tracer. We follow the procedure from [1] to determine the Modified Map and "Modified Excess Template" (Fig. 3, top panel) to construct their ML analogs. The "Smart Map" is determined similarly to the Modified Map, but instead of using the true Mopra ^{13}CO data, we use the ^{13}CO estimates from the CNN. We use this to determine the "Smart Excess Template", defined as the difference between the Smart map and the baseline CO map. The Smart Excess Template, also integrated over all annuli and energies, is shown in Fig. 3, bottom panel. The map covers the full Mopra region, which includes alternating training and testing tiles for the CNN.

To assess how closely the CNN predicts the γ-ray emission inferred by Mopra observations, we compare the excess templates with three different metrics. We quantify the differences with the following three metrics which are determined by using the predicted photon counts per pixel for each of the templates and scaled to $\sim$12 years of *Fermi*–LAT data (we assume the same *Fermi*–LAT observation parameters and event selection as for the simulations described in [1] and in this work): absolute difference (Smart-Modified), fractional difference ((Smart-Modified)/Modified), difference in units of the standard deviation, σ, for the Modified counts ((Smart-Modified)/$\sqrt{\text{Modified}}$). The latter metric allows us to compare the magnitude of the differences due to the CNN

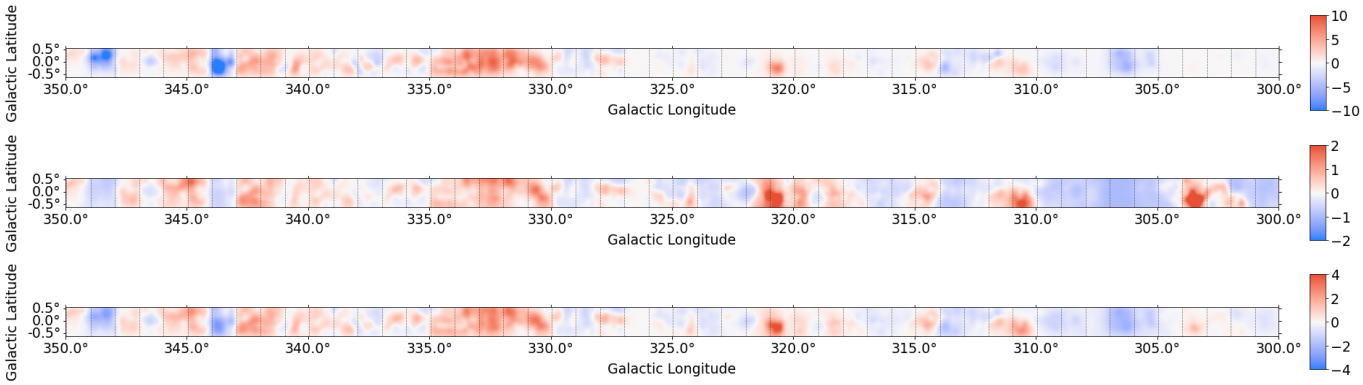

FIG. 4. The pixel-wise residuals for the full region between the predicted counts from the Smart Excess Template and the predicted counts from the Modified Excess Template. *Top*: residual in counts (Smart-Modified), *Middle*: fractional residuals ((Smart-Modified)/Modified), *Bottom*: Residual in units of 1 standard deviation for the Modified Excess Template ((Smart-Modified)/$\sqrt{\text{Modified}}$). The color scale is per 0.08^2 degree pixel.

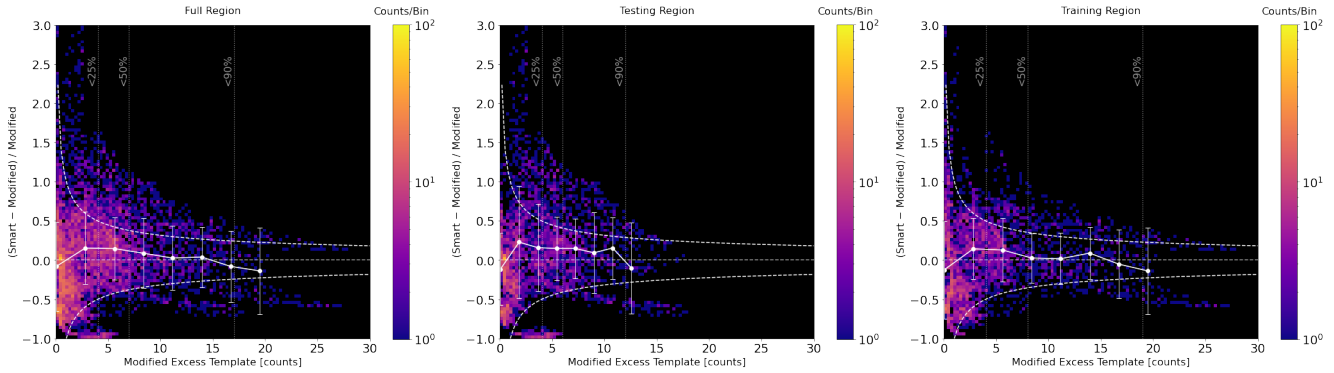

FIG. 5. The fractional residuals between the nominal Smart Excess Template (*Smart*) and Modified Excess Template (*Modified*) predicted counts per pixel as a function of Modified Excess Template pixel counts for all tiles (*left*), testing tiles (*center*), and training tiles (*right*). The points and error bars represent the median and median absolute deviation of the fractional difference for the pixels in 8 domain intervals (we note that the scatter of the points within each interval is not simply characterized by these quantities and does not always indicate a most probable outcome.) We also plot the boundaries of the $\pm 1\sigma$ of the Poisson noise fluctuations of the Modified Excess Template. We overlay the 25%, 50%, 95% flux percentiles to indicate the flux fraction encompassed by pixels below that threshold.

performance in light of the statistical power of the data. All results are shown in Fig. 4 as a function of longitude and latitude for the Mopra region (see [16] for distributions of residuals integrated over the region.) We observe that for the vast majority of the region (83.6% of the pixels), the predicted counts for the Smart Excess Template are within $\pm 1\sigma$ of the Modified Excess Template counts, and therefore the difference is generally within the statistical uncertainty of the data. This result indicates that the CNN performance is adequate for modeling the small-scale structure in H_2-related γ-ray emission traced by Mopra, for the statistics achieved by *Fermi*–LAT.

The CNN performance varies with respect to longitude. Most tiles display either an overall underprediction by the CNN across the entire tile or an overprediction, rather than a comparable mixture of under/over predictions across the same tile. This may be explained by the design of the loss function, placing increased importance towards brighter pixels while also biasing the network to prefer underprediction via Poisson regression. Prominent examples include the training tile $343° < l < 344°$ and testing tile $348° < l < 349°$, the brightest in each dataset, where the CNN underpredicts the gas column density, as shown in Fig. 4. The dependence of the residuals as a function of pixel brightness (in counts/pixel), is shown in Fig. 5 for the full Mopra region, and for the training and testing **sub-regions** separately. Generally, better agreement between Mopra (Modified Excess Template) and the CNN prediction (Smart Excess Template) is found for the brighter pixels. There is a broad spread in the fractional residuals, but it is confined to be between $\pm 50\%$ for the vast majority of the pixels. We overlay con-

tours to indicate the $\pm 1\sigma$ statistical fluctuations in the Modified Excess Template, which encompasses 83.6% of the pixels. We also overlay the median and median absolute deviation of the fractional difference in bins of the Modified Excess Template counts to guide the eye, but emphasize that the scatter of the points within each interval cannot be simply characterized by these quantities and does not always indicate a most probable outcome. As expected, we find that the CNN performs (somewhat) better in the training tiles. Approximately 58% of pixels where the CNN prediction is beyond the $\pm 1\sigma$ level are in the testing **sub-region**. The CNN is more likely to overpredict the emission for the dimmer pixels (60.8% of pixels across the full region are below the 25% flux percentile), consistently with the design. The CNN overprediction in the dim pixels essentially spreads out the brightness of hot-spots over a larger area. We also note that small statistics causes the fractional difference to increase dramatically for the faintest pixels. Above the 90% flux percentile, the distribution of the fractional residuals bifurcates into two separate distribution. The underpredicted pixels are in the $343^\circ < l < 344^\circ$ tile, and the overpredicted pixels are in $333^\circ < l < 334^\circ$ tile, both in the training **sub-region**. This is likely caused by our choice to combine independent convolution layers for each velocity bin with a shared hidden layer. Since certain bins have higher overall brightness than others, the shared hidden layer will attempt to average the error between the two extremes.

We determine the significance of the Smart Excess Template in the *Fermi*–LAT data, similarly to [1] for the Modified Excess Template. The simulations cover the same observations and event selection (∼12 years, $1 - 100$ GeV, P8R3 CLEAN FRONT+BACK). We only simulate the H_2-related γ-ray emission in the Mopra region, excluding all other components, since our goal is to establish the performance of the Smart Excess Template in the optimistic scenario where all other components are known. The simulated events trace the H_2-related γ-ray emission modeled with the Modified Map from Mopra. The simulated data are fit based on a binned maximum likelihood method to a model that includes the baseline CO map (as observed by Mopra) and the Smart Excess Template. The normalization of the Smart Excess Template is free to vary in the fit. The energy spectrum, calculated by GALPROP, is assumed, and held fixed during the likelihood fit. The normalization and spectral index of the CO baseline contribution is also free to vary. The 17 radial bins from GALPROP are combined into 4 radial annuli, which we refer to as A1, A2, A3, and A4, with the same partitioning as in [1]. Also consistently to [1], the normalization of A4 is held constant to a normalization of 1.0, contributing a partial flux of 4.98×10^{-9} ph cm^{-2} s^{-1}.

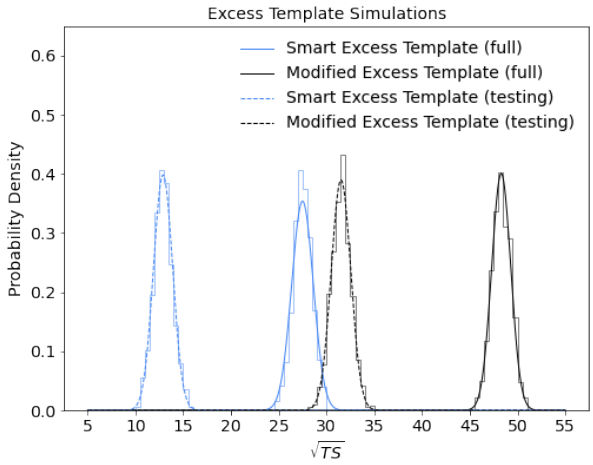

FIG. 6. Statistical significance of the Smart Excess Template for the full Mopra region (solid lines), and for the testing **sub-region** (dashed lines). Each contains 1000 realizations of 12 years of *Fermi*–LAT data. A fit with a Gaussian distribution is overlaid to each distribution. These distributions are compared to the results for the Modified Excess Template from [1].

We simulate 1000 realizations and calculate the Test Statistics (TS) for the nested models ($-2\log(L_0/L)$, where L_0 is the null hypothesis (CO baseline), and L is the alternative hypothesis (CO baseline and Smart Excess Template.) The statistical significance is approximated by $\sigma \approx \sqrt{\mathrm{TS}}$. The distribution of the $\sqrt{TS}$ for the simulations is shown in Fig. 6 for the full Mopra region (solid line), and for the testing **sub-region** only (dashed line). The Smart Excess Template corresponds to $\sqrt{TS} = 27.5 \pm 1.1$ (mean and standard deviation) in the full Mopra region and 12.9 ± 1.0 in the testing **sub-region**. The significance in the testing **sub-region** is lower, not only because of the smaller statistics, but also because the full region contains the training tiles where the CNN more closely matches the Mopra data. For comparison, the Modified Excess Template has a $\sqrt{TS} = 48.3 \pm 1.0$ and $\sqrt{TS} = 31.5 \pm 1.0$ for the full region and testing **sub-region**, respectively. In Fig. 7, we show the distributions for the best-fit flux of the Smart Excess Template overlaid with those for the Modified Excess Template, as well as the flux for the CO baseline emission, separated into annuli. The integrated flux for the Smart Excess Template is approximately 83.7% of the flux of the Modified Excess Template, indicating an overall underprediction by the CNN. The fractional count residuals ((Data-Model)/Model) as a function of energy for the best fit model are shown in Fig. 8 for the fits that include Smart Excess Template, in addition

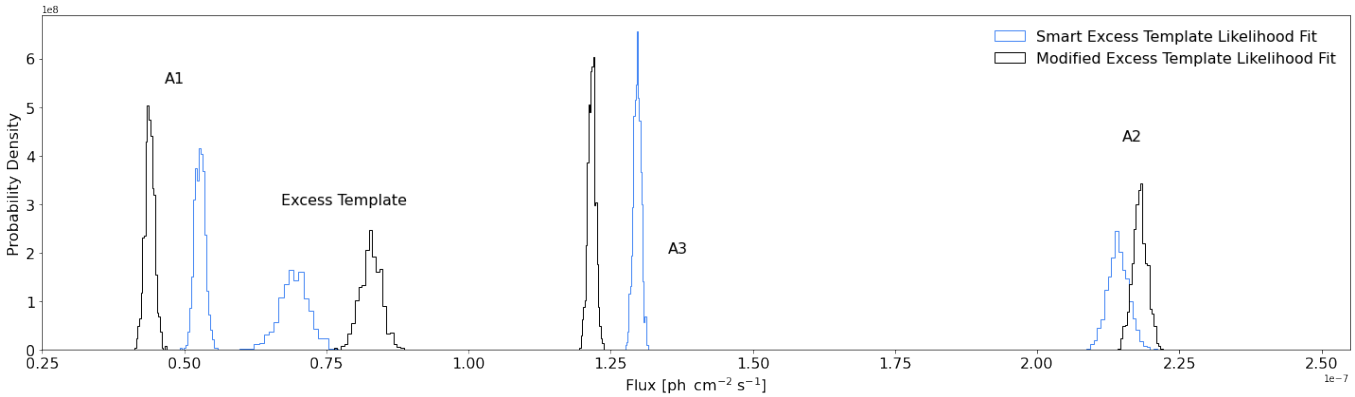

FIG. 7. Distribution of fluxes for each component of the best-fit model using the Smart Excess Template maps over the full region for 1000 simulations of *Fermi*–LAT data. The results for the Modified Excess Template maps from [1] are overlaid. Note that one of the annuli (A4), not shown, has negligible contribution and the normalization was fixed in the fitting process.

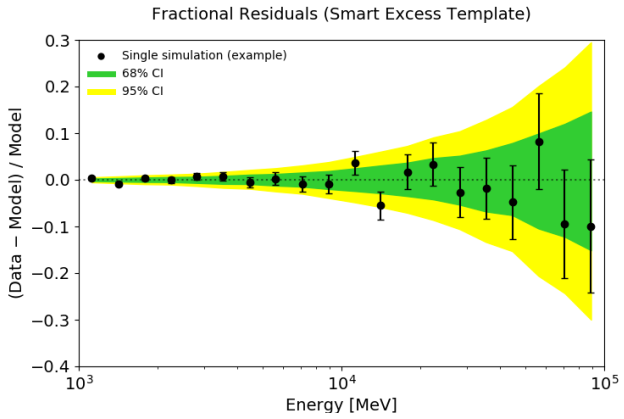

FIG. 8. Fractional count residuals for the fit with the Smart Excess Template. The green and yellow bands show the 68% and 95% confidence regions from 1000 simulations, respectively. We also plot the results for a single simulation as an example in each case, which is shown with black data points.

to the other CO components. They are consistent with zero. The best fit spectra of the CO components agree with the GALPROP prediction. Finally, Fig. 9 shows the fractional count residuals in latitude, longitude. The residuals, which incorporate differences between the (best fit) Smart Excess Template and the (simulated) Modified Excess Template, are smaller than those shown in Fig. 4, (middle panel) indicating the other CO model components partially compensate for the discrepancies between the Excess Templates. We have performed this analysis also using Gaussian Processes and find that the performance is worse compared to the CNN [16].

The ML methodology presented here must be refined to be extended to other regions of the sky. The available all sky CO J=1–0 observations have significantly poorer spatial resolution ($8'$ from [36]) compared to Mopra ($0.6'$). The CNN model will therefore require additional transfer learning to operate on lower resolution maps. The resolution of the *Fermi*–LAT data is worse (for most energies and event types), suggesting that the impact of the poorer CO resolution on the CNN predictions for the γ-rays might ultimately be less pronounced. Another consideration is that the available ^{13}CO training data for the CNN (from Mopra, and other observations [37–41] that could also be included), are confined to the Galactic plane. The H_2 scale height is low, but its small-scale structure contribution at higher latitudes might not be negligible because of the contribution of more local H_2. However, because of the lack of adequate training data the CNN prediction for this component could be more uncertain and this is relevant for the characterization of the GC excess, which extends to high latitudes. Finally, this analysis inherits the limitations and uncertainties in modeling the H_2 component with traditional methods, including the loss of kinematic resolution for the gas in the direction of the GC [42]. And it adds another: the H_2 column densities from CO and ^{13}CO have been treated independently in this work. Their estimates can be related analytically, but rely on a assumptions pertaining to the optical depth, beam filling factor, spatial variation, etc. (e.g. [43, 44]). More ambitiously, ML may be used to constrain some of these uncertainties by using the γ-ray data. Finally, small-scale structure in the γ-ray data could also arise from other components of the IE, e.g. related to HI, and shall be included in a more comprehensive study.

Conclusions.— We have presented a methodology that harnesses ML to predict the small-scale component of the interstellar H_2 gas and its related γ-ray emission for the first time. **Our work is motivated by the the fact that while CO is widely employed as a**

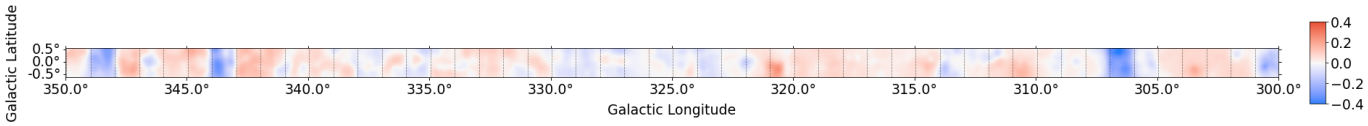

FIG. 9. Fractional count residuals, (Data-Model)/Model, in latitude and longitude for the fits with the Smart Excess Template, in addition to the other CO components. The residuals are calculated using the mean over 1000 simulations.

tracer of H_2, its emission alone does not reliably capture the more highly structured component of H_2. Incorporating this small-scale component in γ-ray IE models is crucial for determining γ-ray point sources and could affect the characterization of extended γ-ray sources, e.g. the *Fermi*–LAT GC excess. Here, we have improved H_2 modeling by employing observations from the Mopra survey which includes isotopologues of CO such as ^{13}CO that do not suffer from the same limitations and can, therefore, more reliably trace the H_2 small-scale structure (the survey spans Galactic longitudes l=300–350° and latitudes $|b| \leq 0.5°$.) In particular, we designed and trained a CNN to predict the concentration of ^{13}CO in the Mopra region. We have tested the performance of this methodology in predicting the contribution of the H_2 related γ-ray small-scale structure to the *Fermi*–LAT data. We find that the accuracy of the predicted γ-ray emission, which incorporates the CNN-predicted contribution, is within the statistical uncertainty of the *Fermi*–LAT data for the vast majority of the region. Moreover, the predicted γ-ray contribution can have a large statistical significance, $\sim \sqrt{TS} = 28$ if we assume other components of the γ-ray IE are known. We designed a dynamically-weighted loss function for the CNN's training to prioritize emission accuracy from brighter spots in the Mopra data while avoiding over-predicting the ^{13}CO concentration in these regions. This is a conservative choice aimed at minimizing the injection of spurious γ-ray sources in the IE model. As a consequence of this design, over-prediction is more likely for the dimmer spots, which we show to be less important. Based on these results, we conclude that deep learning can improve the modeling of γ-ray emissions in data-rich regions, supporting prospects to extend this methodology to other regions of the sky. In this work, we restrict our analysis to the Mopra observations of CO and ^{13}CO which are confined to the Galactic plane. Extending this methodology to other regions will require the CNN model to model ^{13}CO from the poorer resolution CO maps available for the majority of the sky, and will require extrapolating predictions to higher latitudes.

I. ACKNOWLEDGEMENTS

We thank Troy Porter for many helpful discussions and insights. The work of AS, MT, and PB is in part supported by grants NSF NRT 1633631 and ARO 76649-CS to PB. The work of AB was supported in part by the UCI Physical Sciences Machine Learning NEXUS Fellowship. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611.

[1] C. Karwin, A. Broughton, S. Murgia, A. Shmakov, M. Tavakoli, and P. Baldi, "Improved modeling of the discrete component of the galactic interstellar γ-ray emission and implications for the *Fermi*–lat galactic center excess," (2022), arXiv:2206.02809 [astro-ph.HE].

[2] J. Ballet, T. H. Burnett, S. W. Digel, and B. Lott, arXiv e-prints , arXiv:2005.11208 (2020), arXiv:2005.11208.

[3] L. Goodenough and D. Hooper, "Possible evidence for dark matter annihilation in the inner milky way from the fermi gamma ray space telescope," (2009), arXiv:0910.2998 [hep-ph].

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

[5] D. Hooper and T. Linden, Phys. Rev. D **84**, 123005 (2011).

[6] K. N. Abazajian, Journal of Cosmology and Astroparticle Physics **2011**, 010 (2011).

[7] K. N. Abazajian and M. Kaplinghat, Phys. Rev. D **86**, 083511 (2012).

[8] K. Abazajian, N. Canac, S. Horiuchi, and M. Kaplinghat, Physical Review D **90** (2014), 10.1103/PhysRevD.90.023526.

[9] C. Gordon and O. Macías, Phys. Rev. D **88**, 083521 (2013).

[10] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. Portillo, N. L. Rodd, and T. R. Slatyer, Physics of the Dark Universe **12**, 1 (2016).

[11] F. Calore, I. Cholis, and C. Weniger, Journal of Cosmology and Astroparticle Physics **2015**, 038 (2015).

[12] M. Ajello, A. Albert, W. Atwood, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, E. Bissaldi, R. Blandford, E. Bloom, *et al.*, The Astrophysical Journal **819**, 44 (2016).

[13] S. Murgia, Ann. Rev. Nucl. Part. Sci. **70**, 455 (2020).

[14] C. Braiding, G. F. Wong, N. I. Maxted, D. Romano, M. G. Burton, R. Blackwell, M. D. Filipović, M. S. R. Freeman, B. Indermuehle, J. Lau, and et al., Publications of the Astronomical Society of Australia **35** (2018), 10.1017/pasa.2018.18.

[15] P. Baldi, *Deep Learning in Science* (Cambridge University Press, Cambridge, UK, 2021).

[16] A. Shmakov, M. Tavakoli, P. Baldi, C. Karwin, A. Broughton, and S. Murgia, "Deep learning models of the discrete component of the galactic interstellar γ-ray emission: Supplemental material," (2022), See Supplemental Material at [URL will be inserted by publisher] for details on the data cleaning of the Mopra maps, choice of the training and testing regions, description of the Gaussian Processes approach, and additional comparisons between smart and modified maps.

[17] F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, arXiv preprint arXiv:1412.6830 (2014).

[18] K. He, X. Zhang, S. Ren, and J. Sun, CoRR **abs/1502.01852** (2015), 1502.01852.

[19] M. Tavakoli, F. Agostinelli, and P. Baldi, Neural Networks **140**, 1 (2021).

[20] S. Ioffe and C. Szegedy, in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15 (JMLR.org, 2015) p. 448–456.

[21] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Journal of Machine Learning Research **15**, 1929 (2014).

[22] P. Baldi and P. Sadowski, Artificial Intelligence **210C**, 78 (2014).

[23] L. Hertel, J. Collado, P. Sadowski, and P. Baldi (openreview.net, 2018).

[24] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, SoftwareX (2020), also arXiv:2005.04048. Software available at: https://github.com/sherpa-ai/sherpa.

[25] I. V. Moskalenko and A. W. Strong, Astrophys. J. **493**, 694 (1998), arXiv:astro-ph/9710124 [astro-ph].

[26] I. V. Moskalenko and A. W. Strong, Astrophys. J. **528**, 357 (2000), arXiv:astro-ph/9811284 [astro-ph].

[27] A. W. Strong and I. V. Moskalenko, Astrophys. J. **509**, 212 (1998), arXiv:astro-ph/9807150 [astro-ph].

[28] A. W. Strong, I. V. Moskalenko, and O. Reimer, Astrophys. J. **537**, 763 (2000), [Erratum: ApJ 541,1109(2000)], arXiv:astro-ph/9811296 [astro-ph].

[29] V. S. Ptuskin, I. V. Moskalenko, F. C. Jones, A. W. Strong, and V. N. Zirakashvili, Astrophys. J. **642**, 902 (2006), astro-ph/0510335.

[30] A. W. Strong, I. V. Moskalenko, and V. S. Ptuskin, ARNPS **57**, 285 (2007), arXiv:astro-ph/0701517 [astro-ph].

[31] A. E. Vladimirov, S. W. Digel, G. Johannesson, P. F. Michelson, I. V. Moskalenko, P. L. Nolan, E. Orlando, T. A. Porter, and A. W. Strong, Comput. Phys. Commun. **182**, 1156 (2011), arXiv:1008.3642 [astro-ph.HE].

[32] G. Jóhannesson *et al.*, Astrophys. J. **824**, 16 (2016), arXiv:1602.02243 [astro-ph.HE].

[33] T. A. Porter, G. Johannesson, and I. V. Moskalenko, Astrophys. J. **846**, 23pp (2017).

[34] G. Johannesson, T. A. Porter, and I. V. Moskalenko, Astrophys. J. **856**, 45 (2018), arXiv:1802.08646 [astro-ph.HE].

[35] Y. Génolini, D. Maurin, I. V. Moskalenko, and M. Unger, PhRvC **98**, 034611 (2018).

[36] T. M. Dame, D. Hartmann, and P. Thaddeus, Astrophys. J. **547**, 792 (2001), arXiv:astro-ph/0009217 [astro-ph].

[37] J. S. Urquhart, C. König, A. Giannetti, S. Leurini, T. J. T. Moore, D. J. Eden, T. Pillai, M. A. Thompson, C. Braiding, M. G. Burton, T. Csengeri, J. T. Dempsey, C. Figura, D. Froebrich, K. M. Menten, F. Schuller, M. D. Smith, and F. Wyrowski, Monthly Notices of the Royal Astronomical Society **473**, 1059 (2018), arXiv:1709.00392 [astro-ph.GA].

[38] D. J. Eden *et al*, Monthly Notices of the Royal Astronomical Society **498**, 5936 (2020), arXiv:2009.05073 [astro-ph.GA].

[39] P. M. W. Kalberla, J. Kerp, and U. Haud, Astronomy & Astrophysics **639**, A26 (2020), arXiv:2004.14630 [astro-ph.GA].

[40] R. Blackwell, M. Burton, and G. Rowell, in *The Multi-Messenger Astrophysics of the Galactic Centre*, Vol. 322, edited by R. M. Crocker, S. N. Longmore, and G. V. Bicknell (2017) pp. 164–165.

[41] E. M. Di Teodoro, N. M. McClure-Griffiths, F. J. Lockman, S. R. Denbo, R. Endsley, H. A. Ford, and K. Harrington, Astrophys. J. **855**, 33 (2018), arXiv:1802.02152 [astro-ph.GA].

[42] M. Ackermann, M. Ajello, W. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, *et al.*, The Astrophysical Journal **750**, 3 (2012).

[43] M. G. Burton, C. Braiding, C. Glueck, P. Goldsmith, J. Hawkes, D. J. Hollenbach, C. Kulesa, C. L. Martin, J. L. Pineda, G. Rowell, *et al.*, Publications of the Astronomical Society of Australia **30** (2013).

[44] D. Cormier, F. Bigiel, M. J. Jiménez-Donaire, A. K. Leroy, M. Gallagher, A. Usero, K. Sandstrom, A. Bolatto, A. Hughes, C. Kramer, M. R. Krumholz, D. S. Meier, E. J. Murphy, J. Pety, E. Rosolowsky, E. Schinnerer, A. Schruba, K. Sliwa, and F. Walter, Monthly Notices of the Royal Astronomical Society **475**, 3909 (2018), https://academic.oup.com/mnras/article-

pdf/475/3/3909/23965800/sty059.pdf.