

Self-Supervised Learning for Modeling Gamma-ray Variability in Blazars

Aryeh Brill

NASA Postdoctoral Program Fellow, NASA Goddard Space Flight Center
aryeh.brill@nasa.gov

Abstract

Blazars are active galactic nuclei with relativistic jets pointed almost directly at Earth. Blazars are characterized by strong, apparently stochastic flux variability at virtually all observed wavelengths and timescales, from minutes to years, the physical origin of which is still poorly understood. In the high-energy gamma-ray band, the Large Area Telescope aboard the *Fermi* space telescope (*Fermi*-LAT) has conducted regular monitoring of thousands of blazars since 2008. Deep learning can help uncover structure in gamma-ray blazars' complex variability patterns that traditional methods based on parametric statistical modeling or manual feature engineering may miss. In this work, we propose using a self-supervised Transformer encoder architecture to construct an effective representation of blazar gamma-ray variability. Measurement errors, upper limits, and missing data are accommodated using learned encodings. The model predicts a set of quantiles for the flux probability distribution at each time step, an architecture naturally suited for describing data generated by a stochastic process. As a proof of concept for how the model output can be analyzed to extract scientifically relevant information, a preliminary search for weekly-timescale time-reversal asymmetry in gamma-ray blazar light curves was conducted, finding no significant evidence for asymmetry.

Introduction

Blazars are the most extreme class of active galactic nuclei (AGN), energetic phenomena powered by accretion onto supermassive black holes in the centers of a few percent of galaxies. A fraction of AGN host relativistic jets that accelerate particles close to the speed of light. AGN with jets oriented close to the line of sight towards Earth are called blazars (Urry and Padovani 1995). Blazars are the most luminous long-lasting sources of electromagnetic radiation in the gamma-ray sky. Blazar emission extends across the electromagnetic spectrum, including gamma rays in the high-energy (HE; ~ 0.1 -100 GeV) band and even higher energies.

HE gamma-ray observations of blazars have been conducted since 2008 by the Large Area Telescope on board the *Fermi* Gamma-Ray Space Telescope (*Fermi*-LAT). *Fermi*-LAT detects gamma rays from 20 MeV to above 500 GeV using a pair-conversion technique (Atwood et al. 2009). Located in low Earth orbit, *Fermi*-LAT primarily operates in survey mode, during which it scans the entire sky every 3 hr. Among other data products, *Fermi*-LAT generates binned

gamma-ray flux light curves (time series) for each detected source. The time bin duration (typically chosen between 1 day and several months) is a trade-off, with a shorter duration improving time resolution but worsening sensitivity.

Blazars can be divided into two main classes based on their optical spectra, flat spectrum radio quasars (FSRQs) and BL Lacertae-type objects (BL Lacs). The brightest and most variable objects in HE gamma rays are typically FSRQs. Over 3000 blazars have been detected by *Fermi*-LAT (Ajello et al. 2022). Of the *Fermi*-LAT blazars, 755 are FSRQs, 1379 are BL Lacs, and 1208 are of an unknown type.

The physical processes that cause gamma-ray blazar variability are poorly understood. Long-timescale ($\gtrsim 1$ year) gamma-ray variability may be connected to processes in the accretion disk, while short-timescale ($\lesssim 1$ day) variability may result from emission occurring in compact structures in the jet (Rieger 2019). Some gamma-ray blazars exhibit apparent trends (e.g. Valverde et al. 2020) or (quasi-)periodic oscillations (e.g. Peñil et al. 2020; Rueda, Glicenstein, and Brun 2022) on timescales of years.

Of particular physical and observational interest is intermediate-timescale variability, ranging from days to months. Most strikingly, on these timescales blazars can undergo flares, or short-lived flux increases of as much as two orders of magnitude (e.g. Adams et al. 2022). The physical origin of gamma-ray flares in blazars, or even whether flares should be understood as a physical process distinct from ordinary gamma-ray emission, is not known. For example, a model of intermediate-timescale variability as the interaction between an unresolved short-timescale burst process and long-timescale stochastic variations can yield flares as an emergent property (Brill 2022). An understanding of to what extent flares are asymmetric in time is also valuable, as their rise and decay timescales may be connected to the timescales of the particle acceleration and cooling processes in the emission region (e.g. Abdo et al. 2010). For these reasons, a better understanding of blazar variability at intermediate timescales, including the shape of the flux probability distribution and whether it changes over time, can give crucial insight into physical processes.

Blazar emission typically has a power spectral density (PSD) with a power-law shape (e.g. Abdo et al. 2010), possibly with spectral breaks, indicating that it is wholly or partly stochastic. Characteristic gamma-ray variability timescales

can be discovered through time-series modeling (e.g. Kelly, Bechtold, and Siemiginowska 2009; Ryan et al. 2019; Brill 2022). However, variability analysis based on PSD fitting or autoregressive modeling has several important limitations. Model parameters, such as the shape of the flux distribution, must be imposed. Depending on the method used, it can be difficult to deal with missing time bins or upper limits. Methods based on second-order statistics are restricted to modeling time-symmetric autocorrelation. A key assumption underlying many time-series models is stationarity, or time invariance of statistical properties (possibly after filtering out trends or periodicities). The extent to which blazar light curves can be properly described as stationary is still under investigation. For example, Duda and Bhatta (2021) reported transient non-stationarity features in gamma-ray blazar light curves. One potential form of apparent nonstationarity could be a tendency for flaring activity to increase (or suppress) further flux variability, manifesting in time-asymmetric conditional heteroskedasticity. In this work, we explore how the expressive power of deep neural networks can complement traditional analysis tools and enable novel studies by surfacing variability patterns in a model-independent way.

Several different approaches have been used to analyze gamma-ray blazar variability with deep neural networks. Some works have focused on the supervised task of classifying sources of unknown spectral type as FSRQs or BL Lacs using variability features manually extracted from light curves (e.g. Doert and Errando 2014; Kaur et al. 2019). In most contexts, however, light curves are unlabeled, making it natural to adopt an unsupervised approach capable of automatically extracting information of scientific interest for further analysis, classification, and interpretation. Furthermore, light curves may contain rich variability patterns that manually determined features do not adequately describe.

One possible approach, increasingly applied in optical astronomy (Huertas-Company and Lanusse 2022), has used recurrent autoencoders in a semi-supervised or anomaly detection framework to search for transient events such as supernovae (e.g. Villar et al. 2020, 2021) or anomalous variability behavior in AGN (Sánchez-Sáez et al. 2021). Using a recurrent autoencoder, Tachibana et al. (2020) found evidence of temporal asymmetry in the optical variability of quasars, though such an asymmetry could have been produced by a selection bias induced at the time of sample selection (Shen and Burke 2021). Autoencoders can efficiently represent a light curve’s overall long-term variability structure. However, because their compression objective leads them to discard short-term fluctuations, they are less well suited for modeling variability on shorter time scales.

Instead, we propose an approach to uncovering variability patterns in gamma-ray light curves based on self-supervised learning (SSL), in which some of the data points in the input light curves are randomly masked, and a deep neural network is trained to predict the missing data based on its context. We employ a model closely based on Bidirectional Encoder Representations from Transformers (BERT; Devlin et al. 2018), widely used in natural language processing. BERT is built around the Transformer architecture (Vaswani et al. 2017), which employs a self-attention mechanism to

efficiently model complex long-term dependencies in data. In astronomy, SSL using Transformers has been used to remove noise and outliers from optical light curves (Morvan et al. 2022) and to perform pre-training for classifying the light curves of variable stars (Donoso-Oliva et al. 2022).

In this work, we have applied SSL to extract variability information from gamma-ray light curves of blazars. In order to naturally represent the variability structure of stochastic light curves, we introduce a training objective in which the network predicts a probability distribution, rather than a scalar value, for each missing data point, represented non-parametrically as a set of quantiles. After training, the resulting probability distributions can be analyzed to obtain quantities of scientific interest. As a proof of concept, we applied SSL to search for weekly-timescale time-reversal asymmetry in gamma-ray blazar light curves.

Data

For our gamma-ray dataset, we used the publicly available light curves from the *Fermi*-LAT Light Curve Repository¹ (LCR; Fermi Large Area Telescope Collaboration 2021). The LCR consists of light curves for 1525 sources estimated to have a less than 1% chance of being a steady source over 10 time intervals in the source catalog of the first 10 years of *Fermi*-LAT observations (Ballet et al. 2020). Of these, about 93% are blazars.

The LCR provides light curves with time bin sizes of 3 days, 7 days, and 30 days. The light curves are generated by a maximum likelihood analysis. For each time bin, the significance is estimated by a likelihood test statistic (TS ; Mattox et al. 1996). For time bins with a significant detection, the flux F and 1σ errors σ_F are provided; otherwise a 95% photon flux upper limit UL is reported. Time bins may be missing entirely if insufficient data were collected during a time bin or if the analysis otherwise failed to converge. We downloaded the 7 day photon flux light curves corresponding to a fixed spectral index for all sources for the time interval from August 5, 2008 through March 14, 2022, for a maximum of 710 time bins per source.

Guided by the LCR usage notes², we applied a number of quality cuts to the data. Potentially non-convergent bins were removed by excluding bins with a nonzero return code; $TS \leq 0$; $UL < 0$; $(F/\sigma_F)^2/TS \geq 1$; or $(F/\sigma_F)^2/TS \leq 0.1$. Upper limits were selected for all time bins with $TS < 4$, and low-quality data points with $F \leq \sigma_F$ were excluded.

On March 16, 2018, the *Fermi* spacecraft experienced an anomaly in its drive system, after which it adopted a modified orbital profile³. This resulted in occasional periods of low exposure for some sources, leading to erroneous photon flux estimates. To remove any data points potentially affected by this anomaly, for each source i , data points with

¹<https://fermi.gsfc.nasa.gov/ssc/data/access/lat/LightCurveRepository/>

²<https://fermi.gsfc.nasa.gov/ssc/data/access/lat/LightCurveRepository/about.html>

³<https://fermi.gsfc.nasa.gov/ssc/observations/types/post.anomaly/>

$\sigma_F/\mu_i < 1.5$ and upper limits with $UL/2\mu_i < 1.5$ were excluded, where μ_i is that source’s mean photon flux.

Finally, any source with fewer than 100 flux data points; fewer than 300 combined data points and upper limits; or not classified as an FSRQ, BL Lac, or blazar candidate of unknown type was excluded, leaving 689 sources. Fig. 1 shows the light curves after all cuts of the most and least variable sources in the final dataset, as determined by the 4FGL-DR2 variability index (Ballet et al. 2020). Each time bin has either an associated flux F and error σ_F , or an upper limit UL , all in units of photons $\text{cm}^{-2} \text{s}^{-1}$, as well as an associated time index $0 \leq t < 710$.

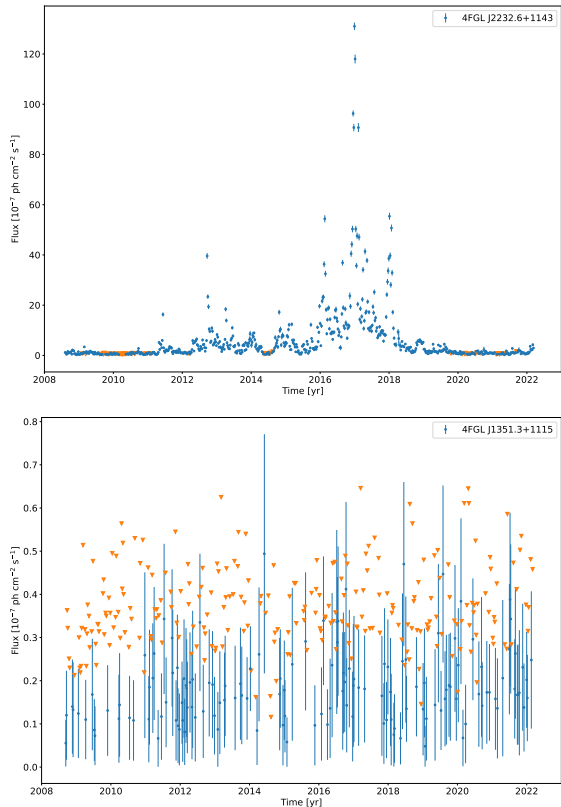


Figure 1: Light curves of the most (top) and least (bottom) variable sources in the LCR dataset after all cuts. Upper limits are indicated by orange triangles.

In order to develop and validate the network, we also created a simulated dataset intended to roughly emulate the general statistical features of the LCR dataset. 1378 light curves were simulated, of which 689 were used for training and 689 were reserved for validation. Each light curve consisted of 710 time steps and was generated following a first-order autoregressive (AR(1)) process,

$$y_{t+1} = \phi y_t + \epsilon_t, \quad (1)$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma)$. The light curves were exponentiated to obtain a lognormal flux distribution and scaled by a random normalization parameter μ_N . The AR(1) parameters of

each light curve were randomly selected following the distribution parameters in Table 1. Simulated measurement errors were applied to the light curves and time bins were randomly removed using a flux/error correspondence and missing-bin probability empirically estimated from the LCR data. Time bins with $F < 2\sigma_F$ were replaced with upper limits such that $UL = \max(2\sigma_F + F, 2\sigma_F)$. A comparison of some statistical properties of the real and simulated datasets is shown in Fig. 2.

Param.	Distribution	μ_{dist}	σ_{dist}	Min	Max
ϕ	Trunc. Norm.	0.70	0.30	0	0.975
σ	Trunc. Norm.	0.45	0.05	0	-
μ_N	Lognormal	-17.4	0.8	-	-

Table 1: Distributions of parameters used to generate the simulated light curves.

Methodology

The core of our model consisted of a BERT-style Transformer encoder network with $L = 4$, $H = 64$, and $A = 4$, with L the number of layers, H the hidden size, and A the number of self-attention heads. The feedforward dimension is set to $4H$ and dropout was used with probability 0.1. Adam was used to train the model with a learning rate of 1×10^{-3} , $\beta_1 = 0.9$, and $\beta_2 = 0.999$ (Kingma and Ba 2014).

We applied several preprocessing steps to transform the data into an input format suitable for the network. First, for each light curve, a segment of 101 time bins was extracted, with the start point chosen randomly. The time indices were shifted to make the index of the central data point 0. Time steps were randomly selected for masking with 20% probability. Of those selected, a mask was actually applied with 90% probability, otherwise, the data point was unaltered. The natural logarithm was applied to the flux points, 1σ flux lower and upper bounds, and flux upper limits. The mean of the unmasked flux points in the segment was subtracted. To create the input vector for the network, the time indices and flux values were converted to vectors of length 32 using fixed trigonometric encodings (Vaswani et al. 2017). Since the flux values and time index both used fixed encodings, the vectors were concatenated rather than summed, yielding input vectors of length 128. For upper limits, the flux upper and lower limits were replaced by a single learned upper limit encoding, while for masked time steps all three flux values were replaced by a single learned mask encoding (superseding any upper limit encoding). This process was repeated in batches, with a batch size of 64.

To generate a non-parametric probability distribution for the flux at each time step, the output encoding was fed through a Quantile Head, illustrated in Fig. 3. Three independent stacks of fully-connected networks computed the median; quantiles less than 0.5; and quantiles greater than 0.5, respectively. The Quantile Head efficiently prevents quantile crossing, i.e., non-monotonicity of the estimated quantiles, by applying a softplus activation followed by a cumulative sum to the estimated quantile vectors. The vector for quantiles less than 0.5 was flipped and negated, the

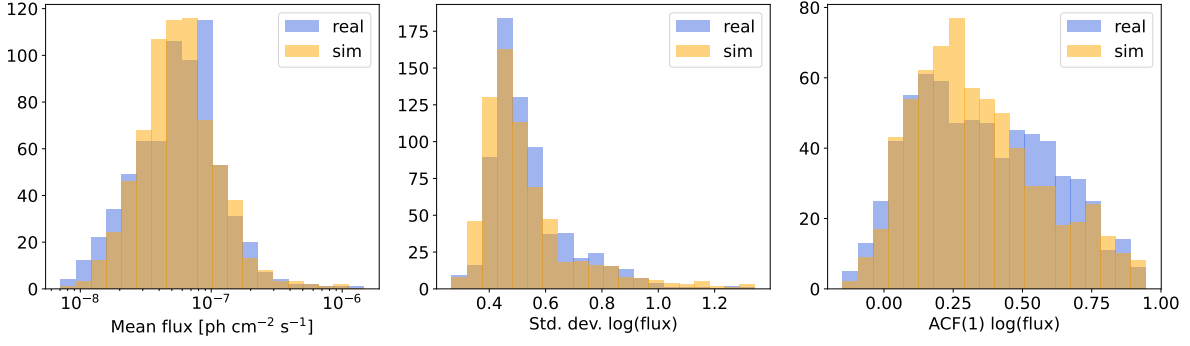


Figure 2: A comparison of statistical properties of the real and simulated datasets, showing, from left to right, the mean photon flux, the standard deviation of the natural logarithm of the flux, and the lag-1 autocorrelation function of the natural logarithm of the flux. Upper limits were excluded from all calculations. No adjustments were made to account for missing time bins.

median added to the other quantiles, and all of the quantiles finally concatenated. We used the set of 19 quantiles for probabilities 0.05 to 0.95, in increments of 0.05.

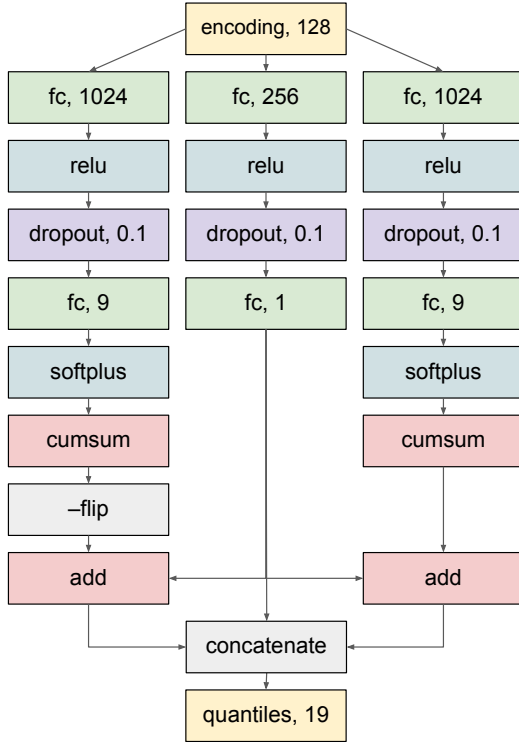


Figure 3: Diagram of the Quantile Head applied to the output embedding at each time step.

The SSL objective was the quantile score or pinball loss function,

$$\mathcal{L} = 2 \sum_p \max(-p(q_p - y), (1 - p)(q_p - y)), \quad (2)$$

where q_p is the quantile corresponding to probability $p \in (0, 1)$ and y is the measured value of $\log(F)$. For $p = 0.5$, the quantile score reduces to the absolute loss. Only masked flux points were included in the loss; any masked upper limits were ignored.

Results

The network was first trained on the simulated dataset for 1000 epochs, monitoring the validation loss throughout training to confirm that it remained consistent with the training loss. The network was then retrained from scratch on the real LCR dataset using the same procedure. Each model was subsequently used to generate predictions on its respective corresponding dataset only.

An example of the output of the model trained on real data is shown in Fig 4. For each data point of each source, predicted quantiles were generated by using as input the interval of length 101 centered on that data point, replacing the data for that point but no others with a mask encoding. No quantiles were generated for the first and last 50 data points.

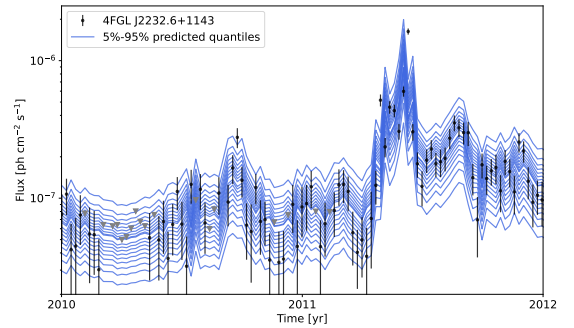


Figure 4: Example of quantiles generated by the trained model, for a portion of the light curve of the most variable LCR source (Fig. 1). Gray triangles indicate upper limits.

The trained models were then used to perform a search for time-reversal asymmetry in flux variability on weekly

timescales. We considered the simple case in which the distribution of $\log(F)$ is assumed to be normal with a standard deviation that may vary at different time steps. To do so, a time-reversed set of predicted quantiles was generated as above, except that the time indices within each interval were reversed. The distributions were centered at 0 by subtracting their respective median. The standard deviation of $\log(F)$ was estimated from each set of forward and time-reversed quantiles as the mean of their values divided by the respective values for a normal distribution. For each data point, the relative difference Δ_{rel} between the forward standard deviation σ_{for} and time-reversed one σ_{rev} was defined as

$$\Delta_{\text{rel}} = 2 \frac{\sigma_{\text{for}} - \sigma_{\text{rev}}}{\sigma_{\text{for}} + \sigma_{\text{rev}}}. \quad (3)$$

The distributions of the mean relative difference $\mu_{\Delta_{\text{rel}}}$ for each source in the simulated and real datasets are shown in Fig. 5. The average value of $\mu_{\Delta_{\text{rel}}}$ for the simulated dataset was -0.002 ± 0.003 and the average for the real dataset was 0.002 ± 0.004 . No significant deviation from 0 was observed for either dataset. By construction, the mean relative difference for the simulated dataset should be 0 for every source. One possible explanation for a negative relative difference could be the network overfitting on the time orientation it was trained on, causing it to be slightly more confident in the forward direction. For the real data, we therefore adopted 0.002 as an additional systematic uncertainty on the mean relative difference. Incorporating this systematic uncertainty, a preliminary upper bound was placed on the time asymmetry of the amplitude of weekly variability of $|\mu_{\Delta_{\text{rel}}}| \lesssim 0.01$ at the 95% confidence level.

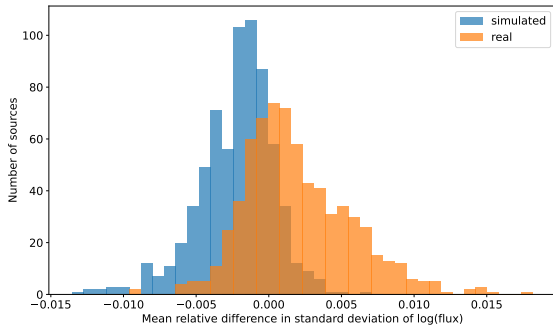


Figure 5: Distributions of the mean relative difference in standard deviation of $\log(F)$ for the simulated and real datasets.

Discussion

In this work, we investigated self-supervised learning, a potentially powerful new tool for astrophysical data analysis. As a proof of concept, we used predicted flux distributions generated by a self-supervised Transformer encoder network to perform a novel search for weekly-timescale time-reversal asymmetry in gamma-ray blazar light curves. No significant time asymmetry in the amplitude of the flux distribution was

found, suggesting that the amplitude of gamma-ray variability in blazars does not itself vary on weekly timescales. This finding is relevant for many statistical methods commonly used in high-energy astrophysics that assume time-reversal symmetry, such as autoregressive modeling.

More generally, the self-supervised framework has a number of advantages relative to standard time series analysis methods used with astrophysical data. Upper limits, measurement errors, and missing data can be naturally accommodated using learned encodings. The neural network output is independent of any statistical model or parameterization and can be used for a variety of tasks, making it an unbiased basis for comparison between models or for exploratory searches. In particular, although we did not rely on this property in this work, the Quantile Head generates non-parametric probability distributions. It can therefore be used as a foundation for a variety of studies, such as comparing different intermediate-timescale flux probability distributions. In addition, the trained network encodes knowledge of and generalizes variability patterns learned from an entire dataset, as opposed to a statistical model which must be fit to one light curve at a time. Finally, although we have focused exclusively on gamma-ray data in this work, the network architecture presented here would naturally extend to accommodate data from multiple wavebands at once, potentially allowing novel multiwavelength variability analyses to be conducted.

Multiple avenues exist for future studies to build on the method discussed here. The flux distributions predicted by the network in the present architecture mix together several sources of variability, incorporating not only variability due to physical processes in the blazar but also measurement error; nuisance parameters such as luminosity and redshift; and estimation error due to the finite length of the light curves. One potential way to eliminate the variability caused by measurement error from the predicted distributions would be to add a secondary self-supervised objective to predict the measurement error at each time step. The predicted error distribution could then be deconvolved from the predicted flux distribution using a Monte Carlo or other approach. In addition, physically-motivated data augmentations, such as stretch and scale transformations to mimic the effect of changing redshift, could be applied to help the network disentangle the effects of cosmological nuisance parameters, as well as to enhance the effective size of the dataset. However, because measurement error is nonlinearly correlated with flux, among other factors, doing so would likely require a more detailed model of the *Fermi*-LAT sensitivity. These methods may provide an improved estimation of the intrinsic variability, enabling more detailed studies of the predicted flux probability distributions. Unsupervised clustering and correlation studies using the generated output encodings could also be performed.

References

Abdo, A. A.; Ackermann, M.; Ajello, M.; Antolini, E.; Baldini, L.; Ballet, J.; Barbiellini, G.; Bastieri, D.; Bechtol, K.; Bellazzini, R.; Berenji, B.; Blandford, R. D.; Bloom, E. D.; Bonamente, E.; Borgland, A. W.; Bouvier, A.; Bre-

geon, J.; Brez, A.; Brigida, M.; Bruel, P.; Buehler, R.; Burnett, T. H.; Buson, S.; Caliandro, G. A.; Cameron, R. A.; Caraveo, P. A.; Carrigan, S.; Casandjian, J. M.; Cavazzuti, E.; Cecchi, C.; Çelik, Ö.; Chekhtman, A.; Cheung, C. C.; Chiang, J.; Ciprini, S.; Claus, R.; Cohen-Tanugi, J.; Cominsky, L. R.; Conrad, J.; Costamante, L.; Cutini, S.; Dermer, C. D.; de Angelis, A.; de Palma, F.; Silva, E. d. C. e.; Drell, P. S.; Dubois, R.; Dumora, D.; Farnier, C.; Favuzzi, C.; Fegan, S. J.; Focke, W. B.; Fortin, P.; Frailis, M.; Fukazawa, Y.; Funk, S.; Fusco, P.; Gargano, F.; Gasparrini, D.; Gehrels, N.; Germani, S.; Giebels, B.; Giglietto, N.; Giommi, P.; Giordano, F.; Glanzman, T.; Godfrey, G.; Grenier, I. A.; Grondin, M. H.; Grove, J. E.; Guiriec, S.; Hadasch, D.; Hayashida, M.; Hays, E.; Healey, S. E.; Horan, D.; Hughes, R. E.; Itoh, R.; Jóhannesson, G.; Johnson, A. S.; Johnson, W. N.; Kamae, T.; Katagiri, H.; Kataoka, J.; Kawai, N.; Knödlseeder, J.; Kuss, M.; Lande, J.; Larsson, S.; Latronico, L.; Lemoine-Goumard, M.; Longo, F.; Loparco, F.; Lott, B.; Lovellette, M. N.; Lubrano, P.; Madejski, G. M.; Makeev, A.; Marsaro, E.; Mazziotta, M. N.; McEnery, J. E.; Michelson, P. F.; Mitthumsiri, W.; Mizuno, T.; Moiseev, A. A.; Monte, C.; Monzani, M. E.; Morselli, A.; Moskalenko, I. V.; Mueller, M.; Murgia, S.; Nolan, P. L.; Norris, J. P.; Nuss, E.; Ohno, M.; Ohsugi, T.; Omodei, N.; Orlando, E.; Ormes, J. F.; Ozaki, M.; Panetta, J. H.; Parent, D.; Pelassa, V.; Pepe, M.; Pesce-Rollins, M.; Piron, F.; Porter, T. A.; Rainò, S.; Rando, R.; Razzano, M.; Reimer, A.; Reimer, O.; Ritz, S.; Rodriguez, A. Y.; Romani, R. W.; Roth, M.; Ryde, F.; Sadrozinski, H. F. W.; Sander, A.; Scargle, J. D.; Sgrò, C.; Shaw, M. S.; Smith, P. D.; Spandre, G.; Spinelli, P.; Starck, J. L.; Strickman, M. S.; Suson, D. J.; Takahashi, H.; Takahashi, T.; Tanaka, T.; Thayer, J. B.; Thayer, J. G.; Thompson, D. J.; Tibaldo, L.; Torres, D. F.; Tosti, G.; Tramacere, A.; Uchiyama, Y.; Usher, T. L.; Vasileiou, V.; Vilchez, N.; Vitale, V.; Waite, A. P.; Wallace, E.; Wang, P.; Winer, B. L.; Wood, K. S.; Yang, Z.; Ylinen, T.; and Ziegler, M. 2010. Gamma-ray Light Curves and Variability of Bright Fermi-detected Blazars. *The Astrophysical Journal*, 722(1): 520–542.

Adams, C. B.; Batshoun, J.; Benbow, W.; Brill, A.; Buckley, J. H.; Capasso, M.; Cavins, B.; Christiansen, J. L.; Coppi, P.; Errando, M.; Farrell, K. A.; Feng, Q.; Finley, J. P.; Foote, G. M.; Fortson, L.; Furniss, A.; Gent, A.; Giuri, C.; Hanna, D.; Hassan, T.; Hervet, O.; Holder, J.; Houck, M.; Humensky, T. B.; Jin, W.; Kaaret, P.; Kertzman, M.; Kieda, D.; Krennrich, F.; Kumar, S.; Lundy, M.; Maier, G.; McGrath, C. E.; Moriarty, P.; Mukherjee, R.; Nieto, D.; Nievas-Rosillo, M.; O'Brien, S.; Ong, R. A.; Oppenheimer, A.; Otte, A. N.; Patel, S.; Pfrang, K.; Pohl, M.; Prado, R. R.; Pueschel, E.; Quinn, J.; Ragan, K.; Reynolds, P. T.; Rhatigan, A.; Ribeiro, D.; Roache, E.; Ryan, J. L.; Santander, M.; Sembroski, G. H.; Williams, D. A.; Williamson, T. J.; Valverde, J.; Horan, D.; Buson, S.; Cheung, C. C.; Ciprini, S.; Gasparrini, D.; Ojha, R.; van Zyl, P.; and Sironi, L. 2022. Variability and Spectral Characteristics of Three Flaring Gamma-Ray Quasars Observed by VERITAS and Fermi-LAT. *The Astrophysical Journal*, 924(2): 95.

Ajello, M.; Baldini, L.; Ballet, J.; Bastieri, D.; Becerra Gon-

zalez, J.; Bellazzini, R.; Berretta, A.; Bissaldi, E.; Bonino, R.; Brill, A.; Bruel, P.; Buson, S.; Caputo, R.; Caraveo, P.; Cheung, T.; Chiaro, G.; Cibrario, N.; Ciprini, S.; Crnogorčević, M.; Cutini, S.; D'Ammando, F.; De Gaetano, S.; Di Lalla, N.; Di Venere, L.; Dominguez, A.; Fallah Ramazani, V.; Ferrara, E.; Fiori, A.; Fukazawa, Y.; Funk, S.; Fusco, P.; Gammaldi, V.; Gargano, F.; Garrappa, S.; Gasparrini, D.; Giglietto, N.; Giordano, F.; Giroletti, M.; Green, D.; Grenier, I.; Guiriec, S.; Horan, D.; Hou, X.; Kayanoki, T.; Kuss, M.; Larsson, S.; Latronico, L.; Lewis, T.; Li, J.; Liodakis, I.; Longo, F.; Loparco, F.; Lott, B.; Lovellette, M.; Lubrano, P.; Madejski, G.; Maldera, S.; Manfreda, A.; Marti-Devesa, G.; Mazziotta, M. N.; Mereu, I.; Michelson, P.; Mirabal, N.; Mitthumsiri, W.; Mizuno, T.; Monzani, M. E.; Morselli, A.; Moskalenko, I.; Negro, M.; Ojha, R.; Orienti, M.; Orlando, E.; Ormes, J. F.; Pei, Z.; Pena-Herazo, H. A.; Persic, M.; Pesce-Rollins, M.; Petrosian, V.; Pillera, R.; Poon, H.; Porter, T.; Principe, G.; Raino, S.; Rando, R.; Rani, B.; Razzano, M.; Razzaque, S.; Reimer, A.; Reimer, O.; Scargle, J.; Scotton, L.; Serini, D.; Sgro, C.; Siskind, E. J.; Spandre, G.; Spinelli, P.; Suson, D.; Tajima, H.; Torres, D. F.; Valverde, J.; Yassin, H.; and Zaharijas, G. 2022. The Fourth Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope – Data Release 3. *arXiv e-prints*, arXiv:2209.12070.

Atwood, W. B.; Abdo, A. A.; Ackermann, M.; Althouse, W.; Anderson, B.; Axelsson, M.; Baldini, L.; Ballet, J.; Band, D. L.; Barbiellini, G.; Bartelt, J.; Bastieri, D.; Baughman, B. M.; Bechtol, K.; Bédérède, D.; Bellardi, F.; Bellazzini, R.; Berenji, B.; Bignami, G. F.; Bisello, D.; Bissaldi, E.; Blandford, R. D.; Bloom, E. D.; Bogart, J. R.; Bonamente, E.; Bonnell, J.; Borgland, A. W.; Bouvier, A.; Bregeon, J.; Brez, A.; Brigida, M.; Bruel, P.; Burnett, T. H.; Busetto, G.; Caliandro, G. A.; Cameron, R. A.; Caraveo, P. A.; Carius, S.; Carlson, P.; Casandjian, J. M.; Cavazzuti, E.; Ceccanti, M.; Cecchi, C.; Charles, E.; Chekhtman, A.; Cheung, C. C.; Chiang, J.; Chipaux, R.; Cillis, A. N.; Ciprini, S.; Claus, R.; Cohen-Tanugi, J.; Condamoor, S.; Conrad, J.; Corbet, R.; Corucci, L.; Costamante, L.; Cutini, S.; Davis, D. S.; Decotigny, D.; DeKlotz, M.; Dermer, C. D.; de Angelis, A.; Digel, S. W.; do Couto e Silva, E.; Drell, P. S.; Dubois, R.; Dumora, D.; Edmonds, Y.; Fabiani, D.; Farnier, C.; Favuzzi, C.; Flath, D. L.; Fleury, P.; Focke, W. B.; Funk, S.; Fusco, P.; Gargano, F.; Gasparrini, D.; Gehrels, N.; Gentit, F. X.; Germani, S.; Giebels, B.; Giglietto, N.; Giommi, P.; Giordano, F.; Glanzman, T.; Godfrey, G.; Grenier, I. A.; Grondin, M. H.; Grove, J. E.; Guillemot, L.; Guiriec, S.; Haller, G.; Harding, A. K.; Hart, P. A.; Hays, E.; Healey, S. E.; Hirasawa, M.; Hjalmarsdotter, L.; Horn, R.; Hughes, R. E.; Jóhannesson, G.; Johansson, G.; Johnson, A. S.; Johnson, R. P.; Johnson, T. J.; Johnson, W. N.; Kamae, T.; Katagiri, H.; Kataoka, J.; Kavelaars, A.; Kawai, N.; Kelly, H.; Kerr, M.; Klamra, W.; Knödlseeder, J.; Kocian, M. L.; Komin, N.; Kuehn, F.; Kuss, M.; Landriu, D.; Latronico, L.; Lee, B.; Lee, S. H.; Lemoine-Goumard, M.; Lionetto, A. M.; Longo, F.; Loparco, F.; Lott, B.; Lovellette, M. N.; Lubrano, P.; Madejski, G. M.; Makeev, A.; Marangelli, B.; Massai, M. M.; Mazziotta, M. N.; McEnery, J. E.; Menon, N.; Meurer, C.; Michelson, P. F.; Minuti, M.; Mirizzi, N.;

- Mitthumsiri, W.; Mizuno, T.; Moiseev, A. A.; Monte, C.; Monzani, M. E.; Moretti, E.; Morselli, A.; Moskalenko, I. V.; Murgia, S.; Nakamori, T.; Nishino, S.; Nolan, P. L.; Norris, J. P.; Nuss, E.; Ohno, M.; Ohsugi, T.; Omodei, N.; Orlando, E.; Ormes, J. F.; Paccagnella, A.; Paneque, D.; Panetta, J. H.; Parent, D.; Pearce, M.; Pepe, M.; Perazzo, A.; Pesce-Rollins, M.; Picozza, P.; Pieri, L.; Pinchera, M.; Piron, F.; Porter, T. A.; Poupard, L.; Rainò, S.; Rando, R.; Rapposelli, E.; Razzano, M.; Reimer, A.; Reimer, O.; Reposeur, T.; Reyes, L. C.; Ritz, S.; Rochester, L. S.; Rodriguez, A. Y.; Romani, R. W.; Roth, M.; Russell, J. J.; Ryde, F.; Sabatini, S.; Sadrozinski, H. F. W.; Sanchez, D.; Sander, A.; Sapozhnikov, L.; Parkinson, P. M. S.; Scargle, J. D.; Schalk, T. L.; Scolieri, G.; Sgrò, C.; Share, G. H.; Shaw, M.; Shimokawabe, T.; Shrader, C.; Sierpowska-Bartosik, A.; Siskind, E. J.; Smith, D. A.; Smith, P. D.; Spandre, G.; Spinelli, P.; Starck, J. L.; Stephens, T. E.; Strickman, M. S.; Strong, A. W.; Suson, D. J.; Tajima, H.; Takahashi, H.; Takahashi, T.; Tanaka, T.; Tenze, A.; Tether, S.; Thayer, J. B.; Thayer, J. G.; Thompson, D. J.; Tibaldo, L.; Tibolla, O.; Torres, D. F.; Tosti, G.; Tramacere, A.; Turri, M.; Usher, T. L.; Vilchez, N.; Vitale, V.; Wang, P.; Watters, K.; Winer, B. L.; Wood, K. S.; Ylinen, T.; and Ziegler, M. 2009. The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission. *The Astrophysical Journal*, 697(2): 1071–1102.
- Ballet, J.; Burnett, T. H.; Digel, S. W.; and Lott, B. 2020. Fermi Large Area Telescope Fourth Source Catalog Data Release 2. *arXiv e-prints*, arXiv:2005.11208.
- Brill, A. 2022. Variability Signatures of a Burst Process in Flaring Gamma-Ray Blazars. *The Astrophysical Journal*, 936(2): 147.
- Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv e-prints*, arXiv:1810.04805.
- Doert, M.; and Errando, M. 2014. Search for Gamma-ray-emitting Active Galactic Nuclei in the Fermi-LAT Unassociated Sample Using Machine Learning. *The Astrophysical Journal*, 782(1): 41.
- Donoso-Oliva, C.; Becker, I.; Protopapas, P.; Cabrera-Vives, G.; Vishnu, M.; and Vardhan, H. 2022. ASTROMER: A transformer-based embedding for the representation of light curves. *arXiv e-prints*, arXiv:2205.01677.
- Duda, J.; and Bhatta, G. 2021. Gamma-ray blazar variability: new statistical methods of time-flux distributions. *Monthly Notices of the Royal Astronomical Society*, 508(1): 1446–1458.
- Fermi Large Area Telescope Collaboration. 2021. 10-year Fermi LAT point source catalog. *The Astronomer's Telegram*, 15110: 1.
- Huertas-Company, M.; and Lanusse, F. 2022. The DAWES review 10: The impact of deep learning for the analysis of galaxy surveys. *arXiv e-prints*, arXiv:2210.01813.
- Kaur, A.; Falcone, A. D.; Stroh, M. D.; Kennea, J. A.; and Ferrara, E. C. 2019. Classification of New X-Ray Counterparts for Fermi Unassociated Gamma-Ray Sources Using the Swift X-Ray Telescope. *The Astrophysical Journal*, 887(1): 18.
- Kelly, B. C.; Bechtold, J.; and Siemiginowska, A. 2009. Are the Variations in Quasar Optical Flux Driven by Thermal Fluctuations? *The Astrophysical Journal*, 698(1): 895–910.
- Kingma, D. P.; and Ba, J. 2014. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, arXiv:1412.6980.
- Mattox, J. R.; Bertsch, D. L.; Chiang, J.; Dingus, B. L.; Digel, S. W.; Esposito, J. A.; Fierro, J. M.; Hartman, R. C.; Hunter, S. D.; Kanbach, G.; Kniffen, D. A.; Lin, Y. C.; Macomb, D. J.; Mayer-Hasselwander, H. A.; Michelson, P. F.; von Montigny, C.; Mukherjee, R.; Nolan, P. L.; Ramana-murthy, P. V.; Schneid, E.; Sreekumar, P.; Thompson, D. J.; and Willis, T. D. 1996. The Likelihood Analysis of EGRET Data. *The Astrophysical Journal*, 461: 396.
- Morvan, M.; Nikolaou, N.; Hou Yip, K.; and Waldmann, I. 2022. Don't Pay Attention to the Noise: Learning Self-supervised Representations of Light Curves with a Denoising Time Series Transformer. *arXiv e-prints*, arXiv:2207.02777.
- Peñil, P.; Domínguez, A.; Buson, S.; Ajello, M.; Otero-Santos, J.; Barrio, J. A.; Nemmen, R.; Cutini, S.; Rani, B.; Franckowiak, A.; and Cavazzuti, E. 2020. Systematic Search for γ -Ray Periodicity in Active Galactic Nuclei Detected by the Fermi Large Area Telescope. *The Astrophysical Journal*, 896(2): 134.
- Rieger, F. 2019. Gamma-Ray Astrophysics in the Time Domain. *Galaxies*, 7(1): 28.
- Rueda, H.; Glicenstein, J.-F.; and Brun, F. 2022. Search for Periodicities in High Energy AGNs with a Time Domain Approach. *The Astrophysical Journal*, 934(1): 6.
- Ryan, J. L.; Siemiginowska, A.; Sobolewska, M. A.; and Grindlay, J. 2019. Characteristic Variability Timescales in the Gamma-Ray Power Spectra of Blazars. *The Astrophysical Journal*, 885(1): 12.
- Sánchez-Sáez, P.; Lira, H.; Martí, L.; Sánchez-Pi, N.; Arredondo, J.; Bauer, F. E.; Bayo, A.; Cabrera-Vives, G.; Donoso-Oliva, C.; Estévez, P. A.; Eyheramendy, S.; Förster, F.; Hernández-García, L.; Arancibia, A. M. M.; Pérez-Carrasco, M.; Sepúlveda, M.; and Vergara, J. R. 2021. Searching for Changing-state AGNs in Massive Data Sets. I. Applying Deep Learning and Anomaly-detection Techniques to Find AGNs with Anomalous Variability Behaviors. *The Astronomical Journal*, 162(5): 206.
- Shen, Y.; and Burke, C. J. 2021. A Sample Bias in Quasar Variability Studies. *The Astrophysical Journal Letters*, 918(1): L19.
- Tachibana, Y.; Graham, M. J.; Kawai, N.; Djorgovski, S. G.; Drake, A. J.; Mahabal, A. A.; and Stern, D. 2020. Deep Modeling of Quasar Variability. *The Astrophysical Journal*, 903(1): 54.
- Urry, C. M.; and Padovani, P. 1995. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107: 803.
- Valverde, J.; Horan, D.; Bernard, D.; Fegan, S.; Fermi-LAT Collaboration; Abeysekara, A. U.; Archer, A.; Benbow, W.;

Bird, R.; Brill, A.; Brose, R.; Buchovecky, M.; Buckley, J. H.; Christiansen, J. L.; Cui, W.; Falcone, A.; Feng, Q.; Finley, J. P.; Fortson, L.; Furniss, A.; Gent, A.; Gillanders, G. H.; Giuri, C.; Gueta, O.; Hanna, D.; Hassan, T.; Hervet, O.; Holder, J.; Hughes, G.; Humensky, T. B.; Kaaret, P.; Kelley-Hoskins, N.; Kertzman, M.; Kieda, D.; Krause, M.; Krennrich, F.; Lang, M. J.; Maier, G.; Moriarty, P.; Mukherjee, R.; Nieto, D.; Nieves-Rosillo, M.; O'Brien, S.; Ong, R. A.; Otte, A. N.; Park, N.; Petrashyk, A.; Pfrang, K.; Pichel, A.; Pohl, M.; Prado, R. R.; Pueschel, E.; Quinn, J.; Ragan, K.; Reynolds, P. T.; Ribeiro, D.; Richards, G. T.; Roache, E.; Sadeh, I.; Santander, M.; Scott, S. S.; Sembroski, G. H.; Shahinyan, K.; Shang, R.; Sushch, I.; Vasiliiev, V. V.; Weinstein, A.; Wells, R. M.; Wilcox, P.; Wilhelm, A.; Williams, D. A.; Williamson, T. J.; VERITAS Collaboration; Noto, G.; Edwards, P. G.; Piner, B. G.; Fallah Ramazani, V.; Hovatta, T.; Jormanainen, J.; Lindfors, E.; Nilsson, K.; Takalo, L.; Kovalev, Y. Y.; Lister, M. L.; Pushkarev, A. B.; Savolainen, T.; Kiehlmann, S.; Max-Moerbeck, W.; Readhead, A. C. S.; Lähteenmäki, A.; and Tornikoski, M. 2020. A Decade of Multiwavelength Observations of the TeV Blazar 1ES 1215+303: Extreme Shift of the Synchrotron Peak Frequency and Long-term Optical-Gamma-Ray Flux Increase. *The Astrophysical Journal*, 891(2): 170.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. *arXiv e-prints*, arXiv:1706.03762.

Villar, V. A.; Cranmer, M.; Berger, E.; Contardo, G.; Ho, S.; Hosseinzadeh, G.; and Lin, J. Y.-Y. 2021. A Deep-learning Approach for Live Anomaly Detection of Extragalactic Transients. *The Astrophysical Journal Supplement Series*, 255(2): 24.

Villar, V. A.; Hosseinzadeh, G.; Berger, E.; Ntampaka, M.; Jones, D. O.; Challis, P.; Chornock, R.; Drout, M. R.; Foley, R. J.; Kirshner, R. P.; Lunnan, R.; Margutti, R.; Milisavljevic, D.; Sanders, N.; Pan, Y.-C.; Rest, A.; Scolnic, D. M.; Magnier, E.; Metcalfe, N.; Wainscoat, R.; and Waters, C. 2020. SuperRAENN: A Semisupervised Supernova Photometric Classification Pipeline Trained on Pan-STARRS1 Medium-Deep Survey Supernovae. *The Astrophysical Journal*, 905(2): 94.

Acknowledgments

A.B. is supported by the NASA Postdoctoral Program at NASA Goddard Space Flight Center, administered by Oak Ridge Associated Universities. Thanks to Daniel Kocevski, Michela Negro, and Janeth Valverde for helpful discussions regarding the *Fermi*-LAT Light Curve Repository. This research has made use of NASA's Astrophysics Data System Bibliographic Services.