

This is the accepted manuscript made available via CHORUS. The article has been published as:

Robust method for treating astrophysical mismodeling in dark matter annihilation searches of dwarf spheroidal galaxies

Tim Linden

Phys. Rev. D **101**, 043017 — Published 20 February 2020

DOI: [10.1103/PhysRevD.101.043017](https://doi.org/10.1103/PhysRevD.101.043017)

A Robust Method for Treating Astrophysical Mismodeling in Dark Matter Annihilation Searches of Dwarf Spheroidal Galaxies

Tim Linden*

*Center for Cosmology and AstroParticle Physics (CCAPP), and
Department of Physics, The Ohio State University Columbus, OH, 43210*

Fermi-LAT observations have strongly constrained dark matter annihilation through the joint-likelihood analysis of dwarf spheroidal galaxies (dSphs). These constraints are expected to be robust because dSphs have measurable dark matter content and produce negligible astrophysical emission. However, each dSph is dim, with a predicted flux that typically falls below the accuracy of the background model. We show that this significantly diminishes the reliability of previous joint-likelihood algorithms, and develop an improved analysis that directly accounts for the effect of background mismodeling. This method produces more robust limits and detections of dark matter in both real and mock data. We calculate improved limits on the dark matter annihilation cross-section, which differ by nearly a factor of two from previous analyses — despite examining identical data.

Over the last decade, Fermi-LAT observations have strongly constrained the dark matter annihilation cross-section. In many cases, these observations have met, or exceeded, the thermal annihilation cross-section — at which dark matter particles in thermal equilibrium in the early universe would freeze out to produce the observed dark matter abundance [1]. Intriguingly, several excesses are consistent with a ~ 30 – 100 GeV dark matter particle annihilating near this cross-section [2–6], most notably the γ -ray excess observed surrounding the Milky Way’s galactic center [7–12]. These facts place us in a new “precision” regime, where the accuracy of dark matter annihilation constraints is critical.

The most robust annihilation constraints stem from the Milky Way’s dwarf spheroidal galaxies (dSphs) [13–19]. While dSphs are unlikely to be the brightest annihilation sources, three factors produce strong constraints. First, stellar dispersion measurements constrain their dark matter content [20–22]. Second, they lie far from the complex Milky Way plane. Third, they produce negligible astrophysical emission [23]. These factors are unique to dSphs and validate their status as the “silver bullet” target for indirect detection.

However, there are two complications in setting annihilation constraints with dSphs. First, while their dark matter content is measured, their relative annihilation rate is uncertain due to variations in the dark matter radial profile, as well as effects such as asphericity and triaxiality. These complexities are translated into statistical errors in the dSph “J-factor”, calculated as the square of the dark matter density integrated over the line of sight. Significant efforts are underway to quantify and model these uncertainties [20–22, 24–28]. However, at the present time, the systematic uncertainty in these J-factor estimations is still large, as different research groups and algorithms have often assigned very different J-factors to individual dSphs (see e.g. [29–31]). Further optical observations will be necessary to constrain the dark matter content in each individual dSph and constrain this systematic uncertainty in setting the dark matter annihilation limits.

Second, the γ -ray emission near each dSph is not modeled to the level of Poisson noise. Errors stem from inaccura-

cies in diffuse Galactic emission models and the large population of point sources with fluxes too small to be identified. Studies have focused on the latter effect by correlating γ -ray hotspots with multi-wavelength catalogs [32–34]. However, likelihood-fitting algorithms inexorably link these uncertainties by allowing a combination of point sources and diffuse emission to fit the data over a large region-of-interest (ROI). In Ref. [35, hereafter, L16] (see also Appendix A) we showed that these issues become acute for sources detected at the $\sim 2\sigma$ level. In this regime, the density of sub-threshold sources approaches the Fermi-LAT angular resolution, and the best-fit model produces a distribution of negative and positive residuals stemming from over- and under-subtraction.

The Fermi-LAT Collaboration’s joint-likelihood analysis of dSphs is the quintessential example of an algorithm that operates in this low-significance regime [16–18]. In calculating a 2σ exclusion limit evaluated over an ensemble of dSphs, the analysis sets constraints based on low-significance information about each object. While the joint-likelihood algorithm is well-defined when each dSph likelihood profile can be interpreted statistically, the standard algorithm does not account for the possibility that mismodeling induces significant positive or negative residuals coincident with dSph locations. The Fermi-LAT collaboration examines these uncertainties by repeating their analysis on thousands of “blank sky” locations that do not contain dSphs, finding that the calculated 2σ limit can vary by an order of magnitude at 95% confidence [18]. This provides an indication for the level of uncertainty produced by diffuse mismodeling, but does not reveal how the true dSph constraint has been affected.

Several data-driven methods have been proposed to constrain the dSph flux without relying on a particular background model [13, 15, 19, 34, 36–40]. In general, these techniques choose numerous “blank sky locations” to predict (either through averaging or machine learning) the γ -ray photon count or flux within a given ROI surrounding each dSph. Comparing the predicted flux with the true dSph flux, they constrain any additional dark matter signal. These analyses benefit because they directly integrate the uncertainty in astrophysical mismodeling into the predicted dSph flux — accounting for the increased prevalence of low-significance excesses in real data. However, they lose sensitivity in two ways.

* linden.70@osu.edu

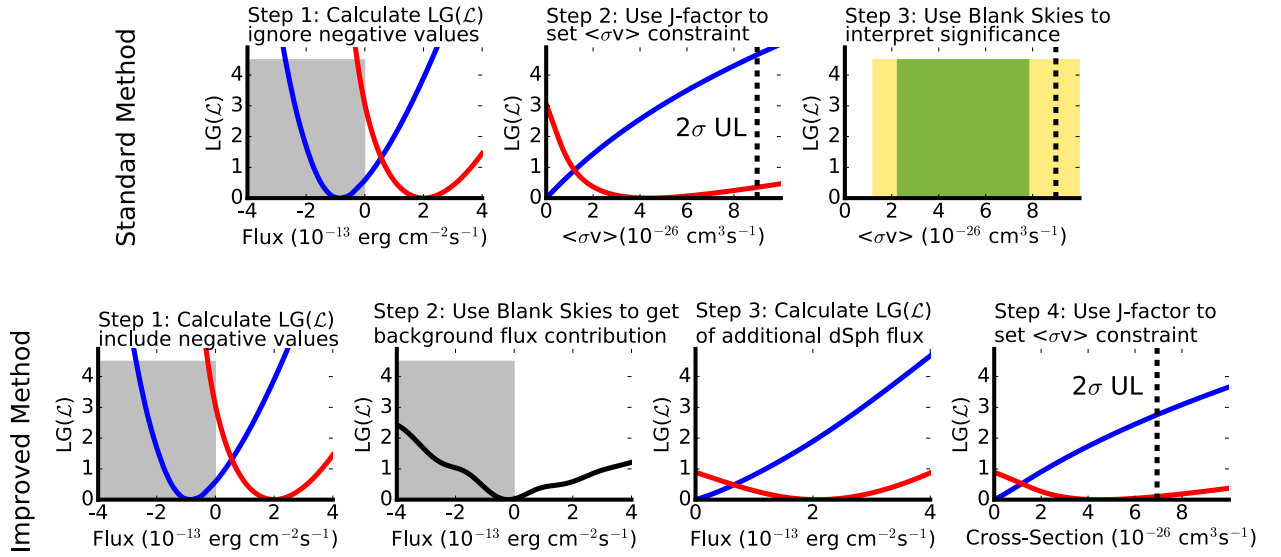


FIG. 1. The standard (top) and improved (bottom) joint-likelihood analyses, applied to a two dSph stack of Ursa Major II (blue) and Willman 1 (red). Results are shown for 100 GeV dark matter annihilating to $b\bar{b}$. Step 1 is identical, and involves calculating the improvement in the fit to the γ -ray data when a point source with an arbitrary flux is added at the location of a given dSph. The key difference is Step 2 and 3, where the new analysis marginalizes the best-fit source flux over the ensemble of possible contributions from systematic errors in the background model. Finally, in each model the J-factor is used to convert the individual dSph fluxes into constraints on the annihilation cross-section, and a joint-likelihood is calculated by adding the individual likelihoods. In this example analysis of two systems, the 2σ limit is improved by $\sim 30\%$. More details are provided in the text.

First, while they utilize γ -ray data to predict the astrophysical flux near each dSph, they do not model known features (e.g. gas clouds or bright point-sources) that have been observed (and integrated into) astrophysical models. Second, they typically do not take into account the full information provided by the energy-dependent point-spread function in differentiating source events from background.

In this paper, we develop an improved analysis that combines the advantages of the Fermi-LAT joint-likelihood analyses with the accurate treatment of fluctuations provided by data-driven methods. The key conceptual improvement is as follows: while the standard joint-likelihood analysis tests the null-hypothesis that: *the distribution of observed dSph fluxes (weighted by J-factors) is consistent with 0*, our analysis tests the null hypothesis: *the distribution of observed dSph fluxes (weighted by J-factors) is consistent with the distribution of “fake” point-source fluxes imparted by astrophysical mismodeling*. This integrates the “blank sky” results directly into the calculated limit. We show that this method produces more resilient detections and constraints on dark matter annihilation, and the results of the new analysis match statistical expectations. Intriguingly, this method can affect the annihilation constraint by a factor of 2 compared to previous work, even when analyzing identical data. We use this model to robustly constrain dark matter annihilation.

Joint-Likelihood Analysis — Our modeling proceeds similarly to L16. Figure 1 illustrates each step in our analysis. For each dark matter model, a probability distribution of point source fluxes induced by astrophysical mismodeling is calcu-

lated from an ensemble of blank sky locations as:

$$P'_{bg}(\phi_{bg}) = \frac{1}{N} \sum_i A_i \mathcal{L}_i(\phi_{bg}) \quad (1)$$

where \mathcal{L}_i is the likelihood function of a single source evaluated at flux ϕ_{bg} , A_i normalizes the integral of the likelihood function over all fluxes and N (the number of blank sky locations) normalizes the value of P'_{bg} to have an integral of one over all blank sky locations. In Bayesian language, P' can be described as a conjugate prior to the likelihood function. We average over 1000 blank sky positions located at $|b| > 30^\circ$, and $> 1^\circ$ from any 3FGL source. The flux can be either positive or negative, because these blank skies account for the impact of mismodeling. The function P'_{bg} is spectrum specific, and is computed for each dark matter mass and final state.

This definition of P'_{bg} includes contributions from both astrophysical mismodeling and Poisson fluctuations in the photon count. Poisson fluctuations in the number of expected background counts are supposed to exist in any model that explains the data up to Poisson noise, while the mismodeling contributions correspond to fluctuations that exceed the Poisson expectation. In order to isolate the effect of mismodeling, we calculate the expected flux distribution $\phi_{bg, Poisson}$, expected from statistical fluctuations, and then deconvolute the two effects in order to isolate a function P_{bg} , which describes the additional systematic uncertainty due to background mismodeling. In this paper, we employ a forward-modeling technique designed to guess a functional form for P_{bg} , which can be convolved with the Poisson noise term $P_{Poisson}$ to yield the calculated value of P'_{bg} . Using an iterative function-finding technique, we obtain functions for P_{bg} that fit the observed P'_{bg} distribution to within a few percent. We then use P_{bg} for the

remainder of the analysis. We note that removing the Poisson noise term at all produces only a $\sim 10\%$ change in the results, as the effect of mismodeling is dominant. In Appendices A and B we describe this procedure in detail and show both P'_{bg} and P_{bg} for a 100 GeV dark matter particle annihilating to $b\bar{b}$.

The true γ -ray flux from a dSph is given by:

$$\phi_{s,i} = \frac{\langle\sigma v\rangle\phi_\chi 10^J}{8\pi m_\chi^2} \quad (2)$$

where $\langle\sigma v\rangle$ is the annihilation cross-section, J is the logarithmic J-factor, and ϕ_χ is the model-specific energy flux per annihilation. For a given $\langle\sigma v\rangle$, the probability that a dSph emits a flux $\phi_{s,i}$ is given by the offset between the calculated J-factor and photometric J-factor:

$$P(\phi_{s,i}, \langle\sigma v\rangle, J_i) = P(J_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(J - J_i)^2}{2\sigma_i^2}\right] \quad (3)$$

where J_i and σ_i are the photometric J-factors and uncertainties for dSph i , expressed logarithmically. Given this definition, the joint-likelihood of an ensemble of dSphs is calculated by first integrating out the background contribution, and then maximizing the resulting likelihood for some value of $P(\phi_{s,i})$:

$$P(\langle\sigma v\rangle) = \prod_i \max_{\phi_{s,i}} \int_{-\infty}^{\infty} \mathcal{L}(\phi_{s,i} + \phi_{bg}) P_{bg}(\phi_{bg}) \times \\ \times P_{s,i}(\phi_{s,i}, \langle\sigma v\rangle, J_i, \sigma_i) d\phi_{bg} \quad (4)$$

where each dSph flux is a combination of a true source flux ($\phi_{s,i}$) and a flux induced by background mismodeling (ϕ_{bg}). The source flux is non-negative (it represents a physical source), while the background flux can be negative (it represents the flux induced by mismodeling). The results are weighted by the probabilities (P_{bg} and $P_{s,i}$) that each flux is consistent with the blank sky and dark matter fluxes, respectively. The joint-likelihood is the product of each dSph likelihood. In this model, the probability is maximized at the best-fit value of $\phi_{s,i}$. In Appendix C we show results obtained by marginalizing over $\phi_{s,i}$. At their core, these models both integrate the information from blank sky positions into the dSph flux calculation *before* the joint-likelihood is performed. We now demonstrate the advantages of this approach.

Fermi-LAT Analysis — We analyze 9.5 yr of Fermi-LAT data (MET: 239557417 – 548377407) in a $14^\circ \times 14^\circ$ ROI around each dSph, restricting our analysis to photons recorded with a zenith angle smaller than 90° . We divide P8R2_SOURCE_V6 photons into four bins based on their angular reconstruction (event types: PSF0–PSF3), and 24 logarithmic energy bins spanning 500 MeV to 500 GeV. We analyze each bin separately and sum the results to obtain the total log-likelihood. Each set of events is binned into 140×140 pixels with a bin size of 0.1° , centered on the position of a dSph (or blank sky) source. The ROI is slightly larger than that chosen for the Fermi-LAT analysis, which utilized a 100×100 grid with identical 0.1° pixels. This difference likely explains the different sensitivity of our low-energy Monte Carlo constraints

compared to the Fermi-LAT results [17]. However, it is also possible that differences in the dark matter spectral models at low-masses affect the results. The instrumental exposure is calculated in 0.5° increments across the spatial map in the same 24 energy bins utilized for the analysis. The exposure is calculated using the EDGES option, which calculates the instrumental exposure at both edges of the energy bin and interpolates the result.

For each dSph, we first calculate the likelihood fit of a background model in each sky region. This fit includes all 3FGL sources, as well as the `gll_iem_v06.fits` diffuse model and the `iso_P8R2_SOURCE_v6.PSF_X_v06.txt` isotropic background model, where X represents the PSF class of each analysis. We calculate the likelihood function (and flux) of each source independently in each energy bin. As such, we allow only the total normalization of each source to vary, with the spectrum of each source fixed to the best-fit 3FGL properties. Because we are using small energy bins, this only marginally affects our analysis. In this stage, we do not include a source at the tentative position of the dSph or blank sky position. The source model is produced with `gtsrcmaps` and the likelihood fit evaluated with `gtlike`, with energy dispersion disabled.

We then evaluate the expected flux from a point source added at each dSph or blank sky position. We utilize `gtmodel` to generate the expected number of counts as a function of position and energy for a given source normalization, and then calculate the improvement to the total log-likelihood as a function of the source flux in each energy bin. During this stage the background model is fixed. We allow the flux of the putative point source to be either positive or negative, representing the additional contribution to the point-source flux from background mismodeling.

Unlike previous work, we evaluate both positive and negative fluxes in each energy bin to calculate the dSph likelihood profile ($\Delta LG(\mathcal{L})$), where $LG(\mathcal{L})$ is defined as the log of the likelihood function, which is calculated by comparing the number of observed and modeled photon counts in each angular and energy bin (using Poisson statistics). In this work, we utilize a two-step fitting algorithm. First, the background diffuse and point source emission is fit to data over the ROI following the standard formalism. We subsequently fit the dSph to the data, allowing only the dSph normalization to float. We have tested alternative methods, and found that they negligibly affect our results when the dSph flux is small, as is the case here. Further analysis details are provided in L16.

We note that Poisson likelihoods cannot be used when the total model prediction (including both the dSph and the background model) is negative. This scenario is very rare, and present only in the highest energy bins, which contribute negligibly to the total $LG(\mathcal{L})$. However, to prevent numerical issues in likelihood scans, we treat this scenario as follows. If the model predicts a negative number of counts in a specific angular and energy bin, we first check the number of observed counts in the data. If the number of observed counts is non-zero, we assign an infinite likelihood due to continuity (the log-likelihood of predicting 0 counts would also be infinite). If the number of observed counts is 0, we allow the predicted

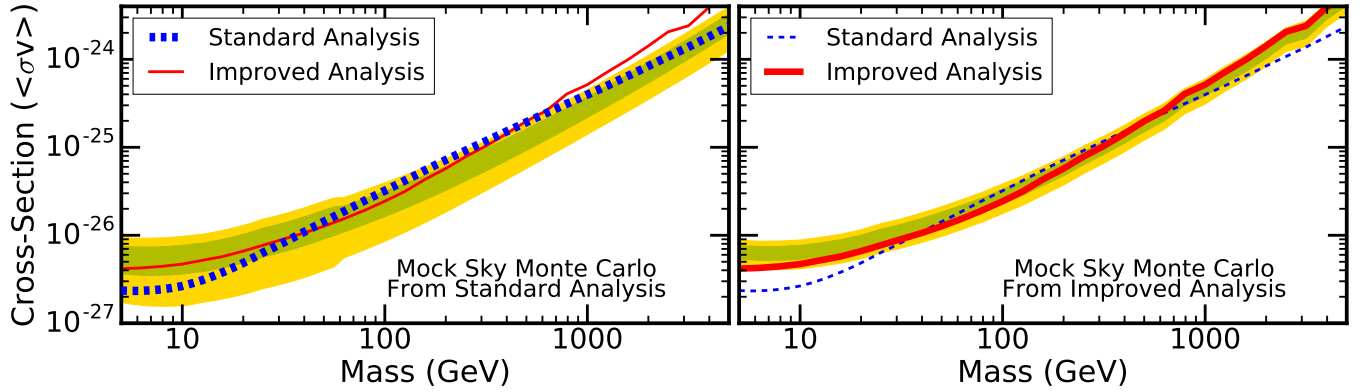


FIG. 2. The constraints on dark matter annihilation calculated using the ensemble of 15 dSphs utilized in [16] for the standard analysis (left) and our improved analysis (right). Results from the opposing analysis are shown with thin lines to aid the eye. We find that the resulting constraints can differ by a factor of two at both low and high masses. In the improved analysis, the uncertainty bands decrease drastically, because deviations produced by astrophysical mismodeling are absorbed by P_{bg} . This illustrates the robustness of the approach.

flux to be negative, and assign a log-likelihood based on the absolute value of the predicted model counts. This maintains a local minimum at the best-fit prediction of zero. We stress that this scenario rarely arises near the best-fit values of the flux, and is important only for the stability of algorithms.

In addition to analyzing each dSph, we additionally analyze 1000 “blank sky locations” at latitude $|b| > 30^\circ$ and separated by $> 1^\circ$ from any 3FGL source using the exact same procedure. We assume that each dSph is represented by a point-source – rather than a spatially extended profile. Ref. [16] showed that this only marginally affects the calculated flux – and we stress that the goal of this paper is to focus on the likelihood algorithms. Thus, whenever possible, we have chosen to keep our analysis choices equivalent to those chosen by the Fermi-LAT collaboration.

Results — We reproduce two Fermi collaboration studies, comparing the standard joint-likelihood analysis with our method. We first focus on the 15 dSphs examined by [16, 17] and then the 45 dSphs studied by [18]. We utilize the reported locations, distances and J-factors from previous work, but use 9.5 yr of data. We make these choices to isolate the effect of the joint-likelihood analysis on the annihilation constraint. *We stress that any change in the constraint stems directly from the analysis method. The datasets are identical, thus the results cannot be considered to be “statistically consistent.”*

In Figure 2, we constrain the annihilation cross-section using the standard (left) and improved (right) methods. The green (1σ) and yellow (2σ) uncertainty bands represent the limits that would be derived by re-assigning each of the 15 dSphs to a random blank sky location (out of the 1000 blank sky locations chosen for this analysis). Thus, these error bands represent the limits that would be expected based on Fermi-LAT analyses if no dark matter signal were present. We note three effects. First, the uncertainty in the expected limit decreases significantly in the improved analysis, demonstrating the resilience of the method. In previous work, the limits varied based on the diffuse mismodeling near individual dSphs. Our model accounts for these fluctuations, decreasing the uncertainty from outlier dSphs.

Second, while the limits are similar between ~ 50 – 500 GeV,

we find weaker constraints (by a factor of 2) at low masses. This is primarily due to the oversubtraction of low-energy γ -rays near Segue I, which is the highest J-factor dSph in the sample. The best-fit flux for 5 GeV dark matter in Segue I is -7×10^{-14} erg cm $^{-2}$ s $^{-1}$. Because the standard model does not allow mismodeling to induce this flux, it adopts a likelihood function that is skewed towards 0. This forces the low-mass constraint to fall nearly 2σ below expectations. In the improved analysis, the negative flux from Segue I is absorbed into ϕ_{bg} . Our analysis still obtains a strong limit from Segue I, because $\phi_{s,i}$ cannot be too big without ϕ_{bg} becoming unreasonably negative. Indeed the improved constraint is still $\sim 2\sigma$ below expectations, but the amplitude of the shift is smaller.

Third, the high-mass limits are also a factor of ~ 2 weaker in our analysis. This is not a fluctuation, but a robust feature (note that the expected error bands predict weaker constraints). The energy flux from TeV dark matter peaks at ~ 100 GeV, where few photons are present. It is likely that 0 photons are observed near the dSph, while the background model predicts $\mathcal{O}(0.1)$ events. Thus, the average dSph has a slightly negative best-fit flux. In the standard analysis, this significantly strengthens the combined limit. However, the improved analysis is normalized to blank-sky positions that also skew towards negative fluxes, correcting for this effect.

In Figure 3 (blue), we show constraints from the 45 dSphs analyzed in [18]. In this case, we find that our analysis produces stronger constraints (by $\sim 25\%$) below ~ 500 GeV. The stronger limits derive from the small excesses found in several dSphs (e.g. Reticulum II), which cause the standard analysis to prefer the annihilation of 150 GeV dark matter to $b\bar{b}$ at a test-statistic (TS) of TS = 8.9 (similar to Ref. [18]). Nominally, this would translate to a 3σ detection of dark matter. However, an excess of this significance is relatively common in analyses of blank-sky locations, decreasing our confidence in this result. The improved analysis also finds a slight excess at 150 GeV (because these dSphs have real upward fluctuations). However, after accounting for the probability that background mismodeling induces similar upward fluctuations, it finds the significance of this excess to be only TS=0.4.

In Figure 3 (red-dashed), we show a scenario that

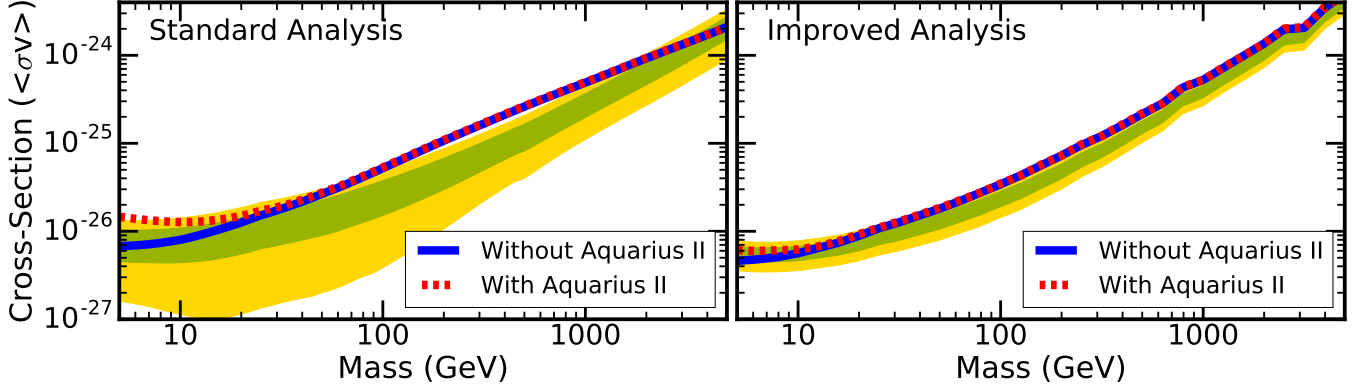


FIG. 3. The constraint on dark matter annihilation from the ensemble of 45 dSphs utilized in [18]. Constraints are shown without (blue solid) or with (red dashed) the addition of the Aquarius II dwarf, which has a significant excess at low γ -ray energies that is unlikely to be caused by dark matter annihilation. This inclusion of this system significantly affects the standard analysis (weakening the combined limit from 45 dSphs by more than a factor of 2), while our improved analysis accurately treats the tension between the high flux and the low J-factor of Aquarius II.

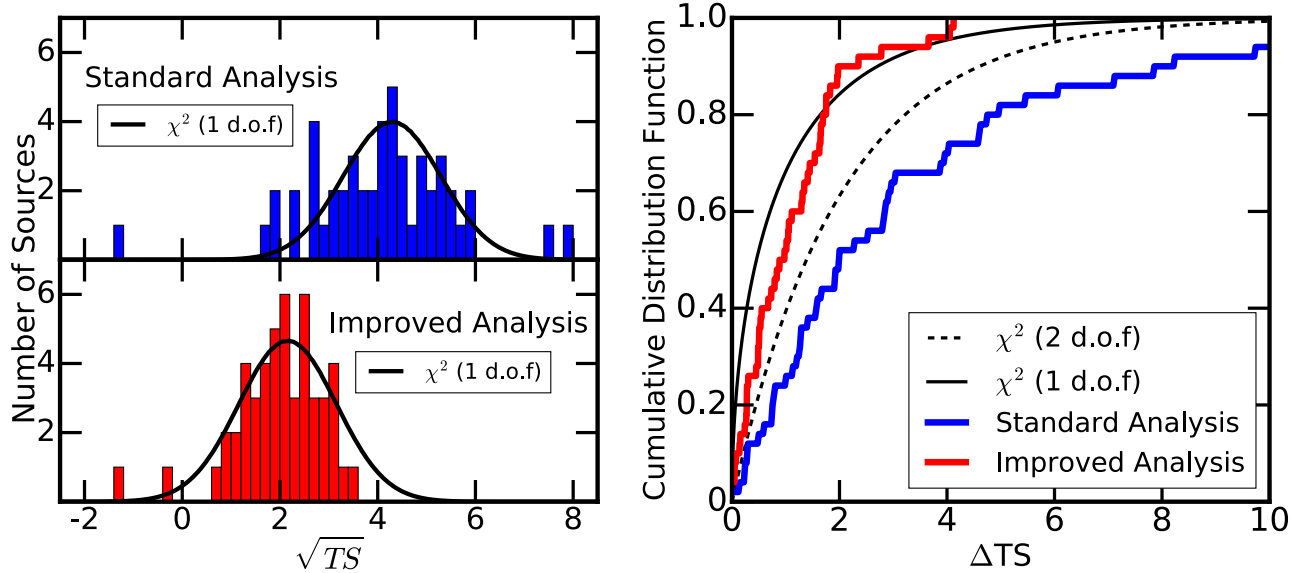


FIG. 4. (Left) - The recovered statistical significance from 50 trials including an injected 100 GeV dark matter signal with cross-section of $3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$. In the default analysis (top left), the statistical significance includes several significant upward and downward fluctuations that would not be expected from a χ^2 distribution surrounding the average significance (such a distribution is rejected at 5.1σ). The improved analysis (bottom left) closely follows a normal distribution. (Right) - The cumulative distribution function of the ΔTS between the injected dark matter mass and cross-section and the recovered best-fit dark matter mass and cross-section in the standard (blue) and improved (red) analyses. Results are compared χ^2 -distributions with 1 and 2 d.o.f. (as the mass is only scanned coarsely). The improved analysis follows this expectation, while the standard analysis prefers an incorrect mass and cross-section at a significance $TS > 4$ in $> 30\%$ of cases.

challenges the standard analysis, induced by adding the Aquarius II dSph into the stack. Aquarius II has a bright, soft-spectrum γ -ray excess. Modeling this excess with 5 GeV dark matter (to $b\bar{b}$) improves the log-likelihood by 10.7, statistically representing a 4.6σ detection of dark matter. However, Aquarius II is not expected to be bright, its logarithmic J-factor is only 18.27 ± 0.62 [41]. The standard algorithm resolves this discrepancy by treating the J-factor measurement and likelihood function equivalently, assuming that the excess is statistical in nature. Thus, the algorithm increases the Aquarius II J-factor to 19.96, paying a log-likelihood cost of 6.9. This produces a $TS=14$ detection of 5 GeV dark matter in the joint-

likelihood analysis and weakens the limit by a factor of 2.

The improved algorithm includes the possibility that the Aquarius II flux is produced by ϕ_{bg} . Blank sky analyses indicate that there is a $\sim 0.7\%$ chance that a background fluctuation would induce an excess compatible with 5 GeV dark matter at this significance. Thus, the Aquarius II result is significant at $\lesssim 2.4\sigma$. Combined with its low J-factor, this prevents the improved algorithm from overreacting to the excess. This test illuminates two lessons. First, the improved analysis allows the statistical uncertainties in each dSph J-factor to be directly compared to the significance of any γ -ray excess. Second, the analysis can accurately treat

outliers in the Fermi-LAT data, whereas the standard method would require manual intervention to remove Aquarius II.

Mock Signal — Thus far, one should worry that our analysis can only constrain dark matter, and may misinterpret a true signal as a fluctuation in ϕ_{bg} . Here, we show that the analysis accurately recovers signals that are injected into γ -ray data. We produce 50 simulations of 15 blank sky locations at latitude $|b| > 30^\circ$ lying $> 1^\circ$ from any 3FGL source. We assign each blank sky to be one of the 15 dSphs analyzed in [16] and use `gtobssim` to inject a signal corresponding to the best-fit J-factor of each dSph and a 100 GeV dark matter model that annihilates to $b\bar{b}$ with a cross-section of $3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$. This lies near the 2σ constraint shown in Figure 2 for both analyses. Thus, we expect a typical detection at $\sim 2\sigma$.

In Figure 4, we show that standard analyses produce results that vary from a -1.4σ exclusion of the signal to a 7.9σ detection. The hypothesis that this dispersion is consistent with Gaussian fluctuations around the mean signal significance is rejected at 5.1σ (still 3.3σ even when the worst outlier is removed). Scanning the 2D mass/cross-section space, we find that the incorrect mass and cross-section are preferred at $\text{TS} > 4$ in $> 30\%$ of cases. While blank skies analyses could characterize this uncertainty, the standard method would have difficulty indicating the correct parameters.

Our analysis produces a distribution of significances following a χ^2 distribution centered near the predicted value of $\sim 2\sigma$. While the distribution of $\sqrt{\text{TS}}$ is slightly more clustered than expected, this is significant at only 0.9σ . In the two-dimensional reconstruction, the variance between the injected and recovered parameters also follows a χ^2 distribution. We note that the dark matter mass and the γ -ray energy are scanned coarsely, and the mass and cross-section parameters are not linearly independent. We thus expect to observe a χ^2 distribution with between 1—2 d.o.f. We conclude that our method is sensitive to dark matter signals, and moreover, that the significance of any discovered excess can be interpreted using standard statistical methods.

Dark Matter Constraints — We now evaluate the annihilation constraints shown in Figures 2 and 3, finding that our analysis rules out (at 95% confidence) dark matter masses below 120 GeV and 87 GeV annihilating to $b\bar{b}$ at a cross-section of $3 \times 10^{-26} \text{ cm}^3 \text{ s}^{-1}$. The constraints do weaken after the inclusion of 30 additional dSphs (and some J-factor changes between [16] and [18]). This is a real effect, based on the low-significance detection of excesses in Reticulum II and a handful of other dSphs. Our results produce maximum TS values of ~ 0.04 at a mass of 5 TeV with 15 dSphs, and 1.06 at a mass of ~ 31 GeV with 45 dSphs. These TS values can be interpreted as true statistical preferences for dark matter, though at a very low significance. We argue that these results, utilizing 9.5 yr of data and an improved statistical method, set robust constraints on dark matter annihilation.

Conclusion — In this paper we introduced a simple, but sig-

nificantly more powerful, method to evaluate the constraints on annihilating dark matter through the joint-likelihood analysis of dwarf spheroidal galaxies. We have demonstrated several advantages to this technique using real Fermi-LAT data: (1) the technique is resilient to systematic uncertainties in Fermi-LAT background modeling and does not produce overly strong limits or unreasonably strong excesses due to diffuse over- or under-subtraction, (2) the technique allows the statistical significance of excesses to be directly compared to the statistical significance of the J-factor measurement, without biasing results towards the Fermi-LAT measurements that are not represented by Poisson statistics, (3) the technique accurately measures the uncertainty in both the cross-section and mass of an injected dark matter signal, and does not produce high precision (but low accuracy) measurements.

We note that this method integrates seamlessly with existing analyses conducted by the Fermi-LAT collaboration, only requiring the use of Eq. 4 to evaluate the joint-likelihood function. We thus recommend its integration in future dSph analyses. Moreover, as noted in L16, the result can be trivially adapted to any search for low-significance excesses in Fermi-LAT data, potentially improving our sensitivity to new phenomena on the threshold of detection. In Appendix D, we discuss several technical considerations required in the implementation of these analysis techniques.

Finally, the result of this analysis provides of the most robust constraints on annihilating dark matter. Using 9.5 yr of Fermi-LAT data, we rule out the thermal annihilation cross-section up to masses of ~ 87 GeV. However, several simplifications have been made (due to our focus on statistical techniques), which will be improved in an upcoming publication. Most notably, we have not utilized the most up to date photometric J-factors produced by numerous groups, instead choosing values consistent with previous studies. Second, we have modeled dSphs as point sources, rather than carefully considering their spatial extension. Third, we have not used the very recently released 4FGL catalog, which appeared subsequent to the computations shown in this paper. The inclusion of these improvements into the current framework is straightforward, and will be considered in an upcoming publication.

ACKNOWLEDGEMENTS

I would like to thank John Beacom, Francesca Calore, Jan Conrad, Alex Drlica-Wagner, Dan Hooper, Jason Kumar, Simona Murgia, and Annika Peter for providing feedback that significantly improved the quality of this paper. I especially would like to thank Alex Geringer-Sameth for a number of valuable comments that provided significant insight into the statistical formalism described here. Portions of this project were completed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. This project also benefited from significant technical support from the Ohio Supercomputing Center.

-
- [1] G. Steigman, B. Dasgupta, and J. F. Beacom, *Phys. Rev.* **D86**, 023506 (2012), arXiv:1204.3622 [hep-ph].
 - [2] D. Hooper, A. V. Belikov, T. E. Jeltema, T. Linden, S. Profumo, and T. R. Slatyer, *Phys. Rev.* **D86**, 103003 (2012), arXiv:1203.3547 [astro-ph.CO].
 - [3] B. Bertoni, D. Hooper, and T. Linden, *JCAP* **1605**, 049 (2016), arXiv:1602.07303 [astro-ph.HE].
 - [4] A. Cuoco, M. Krmer, and M. Korsmeier, *Phys. Rev. Lett.* **118**, 191102 (2017), arXiv:1610.03071 [astro-ph.HE].
 - [5] M.-Y. Cui, Q. Yuan, Y.-L. S. Tsai, and Y.-Z. Fan, *Phys. Rev. Lett.* **118**, 191101 (2017), arXiv:1610.03840 [astro-ph.HE].
 - [6] I. Cholis, T. Linden, and D. Hooper, (2019), arXiv:1903.02549 [astro-ph.HE].
 - [7] L. Goodenough and D. Hooper, (2009), arXiv:0910.2998 [hep-ph].
 - [8] K. N. Abazajian and M. Kaplinghat, *Phys. Rev.* **D86**, 083511 (2012), [Erratum: *Phys. Rev.* **D87**, 129902 (2013)], arXiv:1207.6047 [astro-ph.HE].
 - [9] C. Gordon and O. Macias, *Phys. Rev.* **D88**, 083521 (2013), [Erratum: *Phys. Rev.* **D89**, no.4, 049901 (2014)], arXiv:1306.5725 [astro-ph.HE].
 - [10] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, *Phys. Dark Univ.* **12**, 1 (2016), arXiv:1402.6703 [astro-ph.HE].
 - [11] M. Ajello *et al.* (Fermi-LAT), *Astrophys. J.* **819**, 44 (2016), arXiv:1511.02938 [astro-ph.HE].
 - [12] M. Ackermann *et al.* (Fermi-LAT), *Astrophys. J.* **840**, 43 (2017), arXiv:1704.03910 [astro-ph.HE].
 - [13] A. Geringer-Sameth and S. M. Koushiappas, *Phys. Rev. Lett.* **107**, 241303 (2011), arXiv:1108.2914 [astro-ph.CO].
 - [14] M. Ackermann *et al.* (Fermi-LAT), *Phys. Rev. Lett.* **107**, 241302 (2011), arXiv:1108.3546 [astro-ph.HE].
 - [15] A. Geringer-Sameth and S. M. Koushiappas, *Phys. Rev.* **D86**, 021302 (2012), arXiv:1206.0796 [astro-ph.HE].
 - [16] M. Ackermann *et al.* (Fermi-LAT), *Phys. Rev.* **D89**, 042001 (2014), arXiv:1310.0828 [astro-ph.HE].
 - [17] M. Ackermann *et al.* (Fermi-LAT), *Phys. Rev. Lett.* **115**, 231301 (2015), arXiv:1503.02641 [astro-ph.HE].
 - [18] A. Albert *et al.* (DES, Fermi-LAT), *Astrophys. J.* **834**, 110 (2017), arXiv:1611.03184 [astro-ph.HE].
 - [19] S. Hoof, A. Geringer-Sameth, and R. Trotta, (2018), arXiv:1812.06986 [astro-ph.CO].
 - [20] A. Charbonnier, C. Combet, M. Daniel, S. Funk, J. A. Hinton, D. Maurin, C. Power, J. I. Read, S. Sarkar, M. G. Walker, and M. I. Wilkinson, *MNRAS* **418**, 1526 (2011), arXiv:1104.0412 [astro-ph.HE].
 - [21] V. Bonnivard, C. Combet, D. Maurin, M. G. Walker, and A. Geringer-Sameth, *Proceedings, 14th International Conference on Topics in Astroparticle and Underground Physics (TAUP 2015): Torino, Italy, September 7-11, 2015*, *J. Phys. Conf. Ser.* **718**, 042005 (2016).
 - [22] L. E. Strigari, *Rept. Prog. Phys.* **81**, 056901 (2018), arXiv:1805.05883 [astro-ph.CO].
 - [23] M. Winter, G. Zaharijas, K. Bechtol, and J. Vandenbroucke, *Astrophys. J.* **832**, L6 (2016), arXiv:1607.06390 [astro-ph.HE].
 - [24] V. Bonnivard, C. Combet, D. Maurin, and M. G. Walker, *MNRAS* **446**, 3002 (2015), arXiv:1407.7822 [astro-ph.HE].
 - [25] N. W. Evans, J. L. Sanders, and A. Geringer-Sameth, *Phys. Rev.* **D93**, 103512 (2016), arXiv:1604.05599 [astro-ph.GA].
 - [26] K. Hayashi, K. Ichikawa, S. Matsumoto, M. Ibe, M. N. Ishigaki, and H. Sugai, *MNRAS* **461**, 2914 (2016), arXiv:1603.08046.
 - [27] A. Chiappo, J. Cohen-Tanugi, J. Conrad, L. E. Strigari, B. Anderson, and M. A. Sanchez-Conde, *Mon. Not. Roy. Astron. Soc.* **466**, 669 (2017), arXiv:1608.07111 [astro-ph.CO].
 - [28] A. B. Pace and L. E. Strigari, *ArXiv e-prints* (2018), arXiv:1802.06811.
 - [29] M. Niederste-Ostholt, V. Belokurov, N. W. Evans, G. Gilmore, R. F. G. Wyse, and J. E. Norris, *Mon. Not. Roy. Astron. Soc.* **398**, 1771 (2009), arXiv:0906.3669 [astro-ph.GA].
 - [30] J. L. Sanders, N. W. Evans, A. Geringer-Sameth, and W. Dehnen, *Phys. Rev.* **D94**, 063521 (2016), arXiv:1604.05493 [astro-ph.GA].
 - [31] P. Ullio and M. Valli, *JCAP* **1607**, 025 (2016), arXiv:1603.07721 [astro-ph.GA].
 - [32] E. Carlson, D. Hooper, and T. Linden, *Phys. Rev.* **D91**, 061302 (2015), arXiv:1409.1572 [astro-ph.HE].
 - [33] D. Hooper and T. Linden, *JCAP* **1509**, 016 (2015), arXiv:1503.06209 [astro-ph.HE].
 - [34] A. Geringer-Sameth, S. M. Koushiappas, M. G. Walker, V. Bonnivard, C. Combet, and D. Maurin, (2018), arXiv:1807.08740 [astro-ph.HE].
 - [35] T. Linden, *Phys. Rev.* **D96**, 083001 (2017), arXiv:1612.03175 [astro-ph.HE].
 - [36] M. N. Mazziotta, F. Loparco, F. de Palma, and N. Giglietto, *Astropart. Phys.* **37**, 26 (2012), arXiv:1203.6731 [astro-ph.IM].
 - [37] A. Geringer-Sameth, S. M. Koushiappas, and M. G. Walker, *Phys. Rev. D* **91**, 083535 (2015), arXiv:1410.2242.
 - [38] A. Geringer-Sameth, M. G. Walker, S. M. Koushiappas, S. E. Koposov, V. Belokurov, G. Torrealba, and N. W. Evans, *Physical Review Letters* **115**, 081101 (2015), arXiv:1503.02320 [astro-ph.HE].
 - [39] K. Boddy, J. Kumar, D. Marfatia, and P. Sandick, *Phys. Rev.* **D97**, 095031 (2018), arXiv:1802.03826 [hep-ph].
 - [40] F. Calore, P. D. Serpico, and B. Zaldivar, (2018), arXiv:1803.05508 [astro-ph.HE].
 - [41] A. B. Pace and L. E. Strigari, (2018), arXiv:1802.06811 [astro-ph.GA].
 - [42] F. Acero *et al.* (Fermi-LAT), *Astrophys. J. Suppl.* **218**, 23 (2015), arXiv:1501.02003 [astro-ph.HE].
 - [43] M. Ajello, M. S. Shaw, R. W. Romani, C. D. Dermer, L. Costamante, O. G. King, W. Max-Moerbeck, A. Readhead, A. Reimer, J. L. Richards, and M. Stevenson, *ApJ* **751**, 108 (2012), arXiv:1110.3787.
 - [44] A. Berlin and D. Hooper, *Phys. Rev.* **D89**, 016014 (2014), arXiv:1309.0525 [hep-ph].

Appendix A: The Density of Sub-Threshold Sources as a Function of TS

In this section, we discuss the distribution of positive and negative residuals in the Fermi-LAT data, and their influence on the TS distribution of γ -ray point sources. We also refer the reader to L16, which provides additional information regarding our statistical methods. The influence of unidentified point-sources is significant in Fermi-LAT analyses due to a combination of its high sensitivity ($\sim 10^{-12}$ erg cm $^{-2}$ s $^{-1}$) and modest angular resolution ($\sim 1^\circ$ for individual 1 GeV photons). This combination results in the detection of a handful of sources observed at the $\sim 2\sigma$ level within every angular resolution element – and implies that unidentified sources contribute to the photon count everywhere in the γ -ray sky.

To quantify these statements, we note that there is a strong correlation between the calculated flux of a γ -ray source and the TS of the same source. In regions that are background dominated (the number of source photons is much smaller than the number of background photons), one would expect the source TS to scale as the square of the number of photons attributed to the source. This is due to the fact that (so long as there are many photons) the result is normally distributed:

$$\sigma_s = \sqrt{\text{TS}} = \frac{(n_s + n_{bg} - n_{bg})}{(\sqrt{n_{bg} + n_s})} \approx \frac{n_s}{\sqrt{n_{bg}}} \quad (\text{A1})$$

where n_{bg} is the number of events expected (and observed) from the background, while n_s is the number of events produced by the source. In Figure 5, we show the results of a direct calculation of this correlation over our ensemble of blank sky positions. We find that for a spectrum fitting a dark matter mass of 100 GeV annihilation to $b\bar{b}$, the best-fit powerlaw spectrum is $\text{TS} \propto \phi^{1.88}$. Intriguingly, this correlation holds for both positive and negative values of the γ -ray flux.

To determine the density of sources that contribute such a flux to the γ -ray data, we note that the 3FGL catalog includes 1307 sources detected at $|b| > 30^\circ$ with a $\text{TS} > 25$. Furthermore, the flux-distribution of these sources (at fluxes far from the Fermi-LAT detection threshold) is approximately $dN/df \propto f^{-2}$ [42]. The majority of these sources are extragalactic AGN, which have power-law flux distributions that continue to very low fluxes [43]. Thus, the observation of 1307 sources in our ROI, along with the TS-flux correlation identified above, implies the existence of ~ 5000 sources contributing at the level $\text{TS} > 2$, for detections using only the 4 years of data included in the 3FGL catalog. Moving to 9 years of data (noting that the TS of a true source should scale linearly with exposure), indicates that the density of $\text{TS} \sim 5$ sources is approximately ~ 0.25 deg $^{-2}$, similar to the $\sim 1^\circ$ PSF obtained when only a handful of γ -rays are observed.

Because there is $\mathcal{O}(1)$ source that contributes at $\text{TS} \sim 5$ in each PSF element scanned by the Fermi-LAT, and furthermore, because this source is poorly localized, these sub-threshold sources produce $\sim 2\sigma$ excesses at a much higher rate than expected from Poisson statistics. Moreover, regions with ~ 1 such source will have positive residuals, while regions with 0 sources will have negative residuals. Figure 5 shows

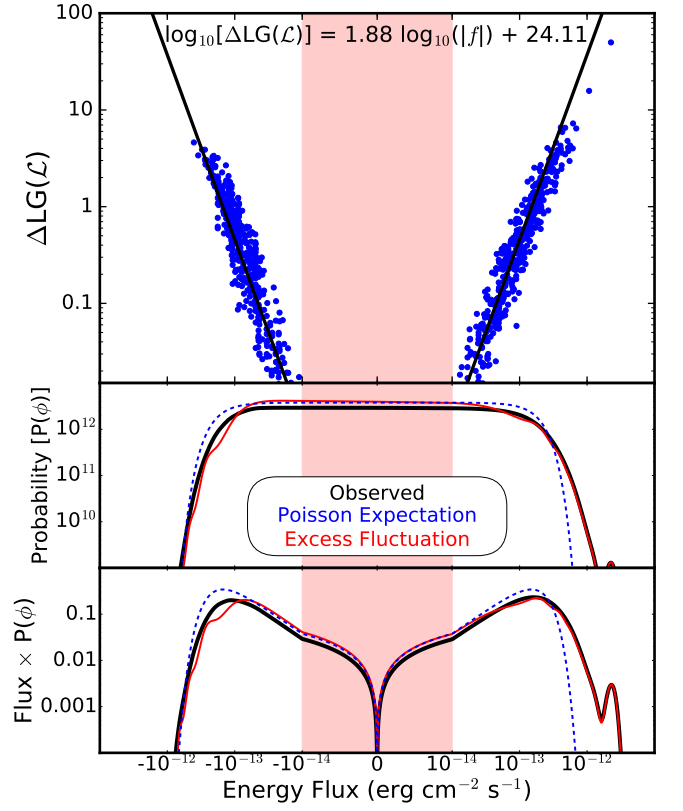


FIG. 5. (Top) The relationship between the point source flux and the improvement of the log-likelihood fit to the data. Points are taken from 1000 blank sky positions, assuming a γ -ray spectrum produced by the annihilation of 100 GeV dark matter to $b\bar{b}$. The improvement in the log-likelihood and the best-fit flux are tightly correlated, as would be expected given that the $\Delta\text{LG}(\mathcal{L})$ is based on the photon count in each energy and angular bin. Importantly, this relationship holds for γ -ray sources with both positive and negative fluxes, illustrating the importance of accurately quantifying negative fluctuations produced by modeling uncertainties. (Middle) The relative probability that a blank sky position has a particular point-source flux for the same dark matter model (black solid). The distribution of positive and negative fluxes is nearly symmetric up to a flux of nearly $\sim \pm 2 \times 10^{-13}$ erg cm $^{-2}$ s $^{-1}$, which roughly translates to $\Delta\text{LG}(\mathcal{L}) = 2$. The flux is broken down into the expected distribution from Poisson fluctuations (blue dotted), and the residual which is due to the mismodelling of the background (red solid). (Bottom) Same as the Middle, but multiplied by the γ -ray flux to show the regions where sources are most likely to be located. A large, significant residual is seen, especially at high positive fluxes.

this distribution, finding a nearly 50% chance that an analysis of a random sky position has its log-likelihood improved when a negative point source is added to the analysis. This is due to a combination of both modeling errors, as well as Poisson fluctuations in the photon count.

Appendix B: Separation of Poissonian and Systematic Errors

Because the blank sky analysis operates on real Fermi-LAT data, the likelihoods that enter into Equation 1 include both systematic errors (which we would like to isolate and account

for in P_{bg}) and Poisson fluctuations (which should not be included in the calculation of P_{bg} , because we want the calculated fluxes of dSphs to include statistical fluctuations). To deconvolve these two terms, we note that the Poissonian error nearly follows a Gaussian distribution, and we calculate the distribution of fluxes induced by Poissonian fluctuations using the correspondence between the γ -ray flux and the log-likelihood fit to the γ -ray data (Figure 5 (top)). In this paper, we build a model for the Poissonian fluctuations by calculating the average Poisson error in all blank sky locations. We note that this technique could be slightly improved by instead calculating the individual Poisson errors in each blank sky location (since the photon counts differ slightly near each blank sky position). However, because we use blank sky locations that lie at high Galactic latitude and far from any point sources, the background fluxes for all sources are similar.

In theory, many deconvolution algorithms can be used to separate these two components – but preventing numerical issues can be tricky (due to the large dynamic range in P_{bg}). In this paper, we utilize forward modeling. We first guess at the systematic distribution P_{bg} and then convolve this distribution with the Poissonian noise term to fit the measured value of P'_{bg} . We iteratively improve this fit and approach a “good” solution. After each iteration we use a high-pass filter to remove high-frequency noise. We obtain solutions that fit to within 5% (usually within 1%) over the range where the P_{bg} is more than 10^{-5} of its maximum value. This fit could be improved with additional iterations. The result for 100 GeV dark matter is shown in the middle and lower panels of Figure 5. We find that the amplitude of the excess is much larger for positive flux values – representing the appearance of real sources that produce fluctuations above the expected level of Poisson noise. However, there are important contributions to P_{bg} at negative fluxes as well. In particular, we note that the similarity between the Poisson expectation of P'_{bg} and the observed value of P'_{bg} is incidental. If only positive systematic errors existed in the Fermi-LAT data, we would expect P'_{bg} to have fewer negative fluctuations than expected by our Poisson model.

Appendix C: Marginalizing Over J-Factor Uncertainties

In our default analysis, we determine the flux from each individual dSph by integrating over the possible background contributions to the total flux to determine the best-fit value of the source flux $\phi_{s,i}$. However, we then minimize to find the best-fit value of $\phi_{s,i}$. Because we are testing a specific value of $\langle\sigma v\rangle$ in a specific dSph, this is equivalent to finding a best-fit J-factor in each dSph (where best-fit includes the penalty from choosing a J-factor that deviates from the observed value). In this sense, the method is pseudo-Bayesian, as it builds a prior for the systematic errors, but minimizes the fit for some values of the dSph J-factors. Alternatively, we could institute a more Bayesian approach, where we marginalize our results over all possible J-factors that could contribute to the fit, as follows:

$$P(\langle\sigma v\rangle) = \prod_i \int_{-\infty}^{\infty} \int_0^{\infty} \mathcal{L}(\phi_{s,i} + \phi_{bg}) P_{bg}(\phi_{bg}) \times \\ \times P_{s,i}(\phi_{s,i}, \langle\sigma v\rangle, J_i, \sigma_i) d\phi_{s,i} d\phi_{bg} \quad (C1)$$

where the likelihood and probability distributions are defined as in the main text, while the integral over $\phi_{s,i}$ is assumed to have a logarithmically-flat prior. The integral over $\phi_{s,i}$ includes only non-negative values, as the emission represents a true source that must be non-negative. This is identical to a constraint that the dSph J-factor must be non-negative.

This technique potentially improves the sensitivity of dark matter searches, as it includes additional information regarding the distribution of J-factor uncertainties. For example, while Equation 4 will not impose any costs to a scenario where the J-factor of each dSph falls exactly at the predicted value, Equation C1 imposes likelihood costs for any scenario where a portion of the J-factor parameter space that is consistent with optical velocity dispersion measurements begins to be ruled out. On the other hand, this means that Equation C1 assumes detailed knowledge of not only the best-fit J-factors, but also of the uncertainty in each J-factor calculation. Moreover, Equation C1 assumes that the uncertainties in each dSph J-factor are uncorrelated, which may or may not hold depending on the analyses utilized in the J-factor calculations.

In Figure 6, we show the constraints from each method, applied to the Monte Carlo analyses of the 45 dSphs analyzed in [18] (the differences in the cross-section constraint in the case of 15 dSphs are significantly smaller, with an amplitude of $\sim 20\%$). We find that utilizing Equation C1 produces a constraint on the annihilation cross-section that is nearly a factor of three stronger than our standard analysis (given by Equation 4). This is similar to the recent result of [19], who found that Bayesian analyses of dSphs produced significantly stronger limits. This result is due to the fact that the J-factor uncertainties are logarithmic, and significant upward fluctuations in the J-factor are strongly constrained by the γ -ray analysis. Intriguingly, such a method would rule out the thermal annihilation cross-section for dark matter annihilation to $b\bar{b}$ final states up to masses of more than 200 GeV.

However, marginalizing over the J-factor uncertainties introduces peculiar statistical properties into the calculated limit. It is clear that in the case where the J-factor uncertainty goes to 0, the two limits must agree. In Figure 6, we show the calculated dark matter annihilation limit in our default analysis behaves as expected – as the J-factor uncertainty decreases, the limits become stronger as downward fluctuations of the dSph J-factors become inconsistent with the data. However, in the marginalized case, the limits become weaker as the uncertainty in the J-factor calculation decreases. This is due to the fact that the strength of the marginalized constraint derives primarily from the strong exclusion of upward fluctuations in the J-factor of numerous dSphs. As the fraction of the total J-factor probability that includes large upward fluctuations increases, the constraint on the dark matter annihilation cross-section becomes stronger.

The legitimacy of each interpretation is open for debate.

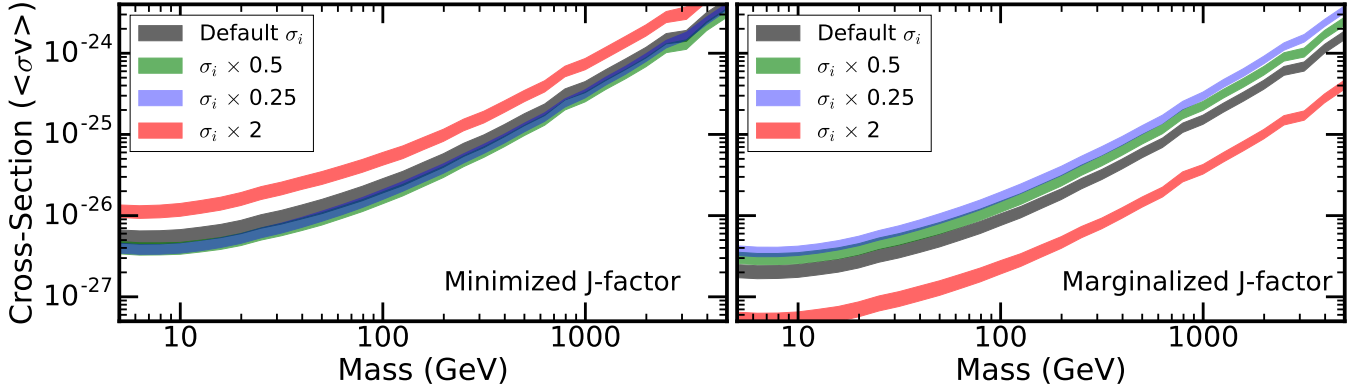


FIG. 6. The 1σ uncertainties in the expected constraint on dark matter annihilation obtained from the stacked analyses of 45 dSphs described in [18]. Results are shown using two different methods for including J-factor uncertainties into the joint-likelihood analysis, including the analysis described by Equation 4 (left), and an analysis that marginalizes over the J-factor uncertainty following Equation C1 (right). In both cases, we show the default limits (black), as well as limits from scenarios where we multiply the J-factor uncertainty in each individual dSph by 2 (red), or divide the J-factor uncertainty by 2 (green) and 4 (blue). In our default analysis, the limit behaves as expected, producing stronger constraints as the uncertainty on the J-factor of individual dSphs decreases. In the marginalized analysis, the limit demonstrates perverse behavior, becoming stronger as the J-factor uncertainty increases. This is due to the very strong constraints on large upward fluctuations in the dSph J-factor provided by γ -ray analyses.

It appears reasonable that the marginalized model should be used in scenarios where the uncertainty in each individual dSph is understood and is statistically independent from the uncertainty in other dSphs. In this case, the stacked analysis of 45 dSphs should rule out scenarios where *none* of the 45 dSphs in the analysis were consistent with upward fluctuations in the observed J-factor. However, these assumptions are unlikely to hold – as the calculation of the J-factor depends on the careful evaluation of a number of systematic uncertainties that are likely to be correlated between individual dSphs. Thus, we find the performance of our default J-factor model to provide a more accurate representation of our best constraints on the dark matter annihilation cross-section.

Appendix D: Further Technical Details

In this section, we communicate several more technical details from our analysis method. This includes detailing several simplifications utilized to conserve computational time during this study, as well as several caveats and warnings to researchers who may wish to utilize this technique.

We first note several simplifications in this analysis. In all cases, we believe that these negligibly affect our results. First, following the choice made by the Fermi-LAT collaboration, we have only investigated blank sky positions with latitude $|b| > 30^\circ$, despite the fact that the population of 45 dSphs analyzed in [18] includes several objects that are closer to the Galactic plane. We have made this choice in order to produce an “expected constraint” plot which is similar to the Fermi-LAT results, but note that the correct choice of blank sky locations is somewhat more important in our analysis — as the blank sky locations are directly utilized in order to calculate the dark matter annihilation limit. In future work, we will select an ensemble of blank sky locations which is more closely

matched to the analyzed dSph population.

Second, in generating the expected dark matter limits, we have utilized the same blank sky population that was employed to produce the distribution P_{bg} in the joint-likelihood analysis. This technically removes some statistical independence from the blank sky locations used to produce the expected dark matter constraints, since each dwarf contributes at the 0.1% level to the P_{bg} calculation. Thus, this is expected to narrow the range of uncertainties in our calculated limits at the order of a few tenths of a percent (since a handful of dSphs dominate the result in any dwarf stacking analysis). This was done to save computational time, and will be tested in detail in future analyses.

Third, in our analysis of the expected limits for 45 dSphs, we have doubled the number of Monte Carlo trials by recalculating the expected limit when each blank sky location is re-assigned from representing the highest J-factor dSph to representing the lowest J-factor dSph (and vice versa). This decreases the statistical independence of our results by a small amount — because the 22 dSphs with the lowest J-factors contribute negligibly to the limit. This was done to double the number of trials in the dSph Monte Carlo, since more than 22 trials were necessary to generate reasonable 1σ and 2σ errors. We used an entirely independent set of Fermi-LAT sky positions to calculate the expected signal when dark matter is injected into the data, keeping that analysis independent.

We additionally note several technical details for those wishing to replicate these results. First, the utilization of negative values for the point source flux can produce mathematical issues whenever the expected counts from the entire model (source + astrophysical background) becomes negative, as the Poisson statistic is no longer well-defined. This happens very rarely in this analysis, and only at energies above ~ 100 GeV. These bins contribute negligibly to the total limit. To address these issues, we adopt the following two-prong approach. If

the number of observed counts in the energy and angular bin is non-zero, we set the resulting log-likelihood to infinity. This is straightforward, since the log-likelihood expectation of producing 0 counts when ≥ 1 count is observed is also infinity. If the number of observed counts in the bin is instead 0, we set the number of predicted accounts to be equal to its absolute value. This is sensible, as it produces a best-fit point at the prediction of 0 counts. We note that these issues contribute very negligibly to the calculated limits, as energy bins above 100 GeV contribute negligibly compared to the lower-energy photon data.

Finally, while the techniques described in this letter are widely applicable to investigations of low-significance excesses throughout the Fermi-LAT data, their interpretation may be tricky in scenarios where they are applied to studies

of Fermi-LAT blazars. This is due to the fact that blazars are responsible for a significant fraction of the low-significance ($TS \sim 5$) excesses observed in the Fermi-LAT data. Thus the calculation of P_{bg} will include substantial contributions from the source population it is attempting to constrain, and will associate these sources with background fluctuations. This is unlikely to affect other source classes, such as star-forming galaxies (as studied in L16), or dSphs, though we note that dark matter annihilation from subhalos can potentially produce a few percent of the underlying low-statistics fluctuations [44]. We note that in the case of blazars, variations in the standard analysis may be utilized to chose blank sky locations that significantly differ from the class of blazars being analyzed in the stacking analysis – but care must be employed when utilizing this method.