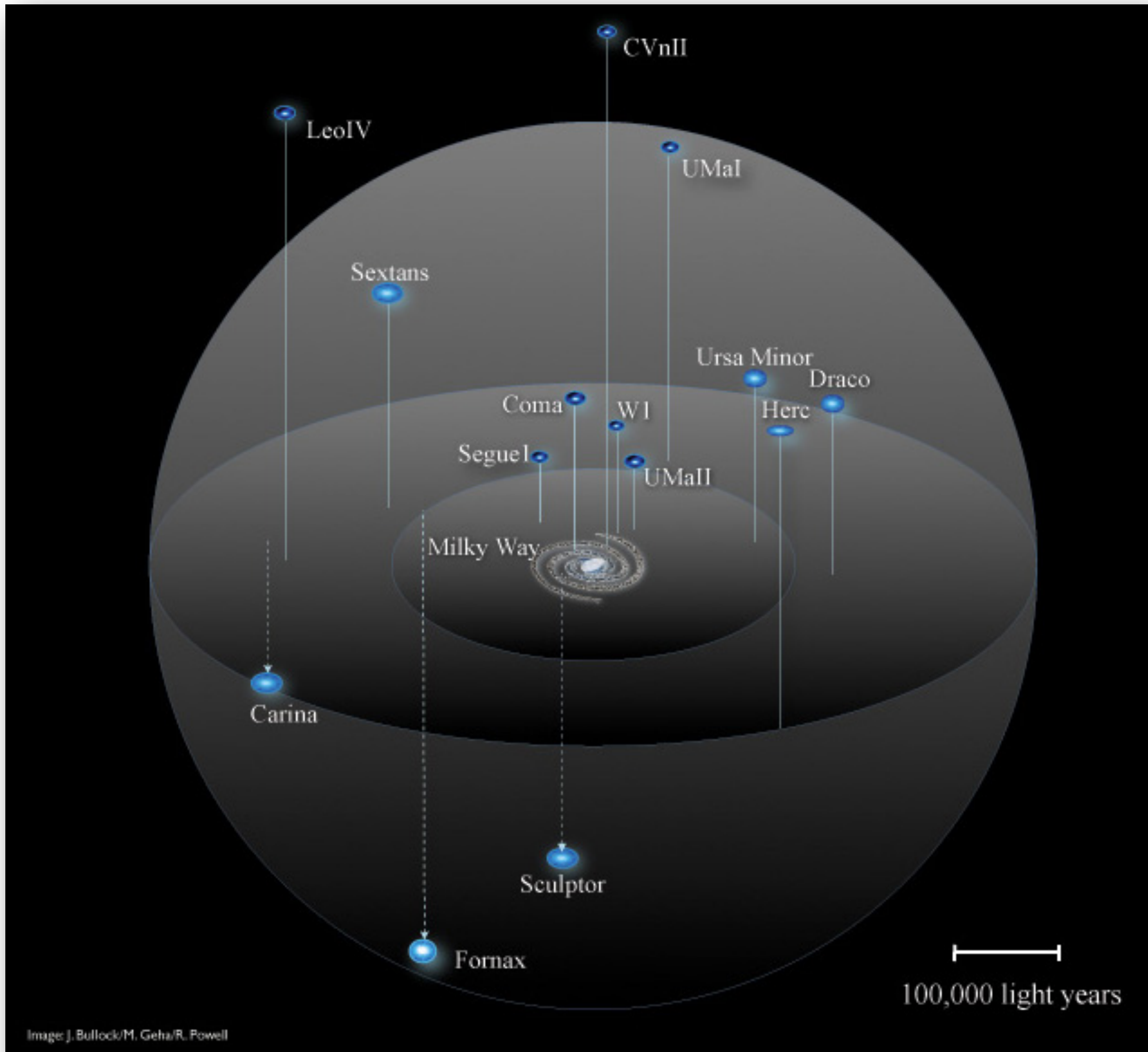
Dark matter constraints from dwarf galaxies: a data-driven analysis

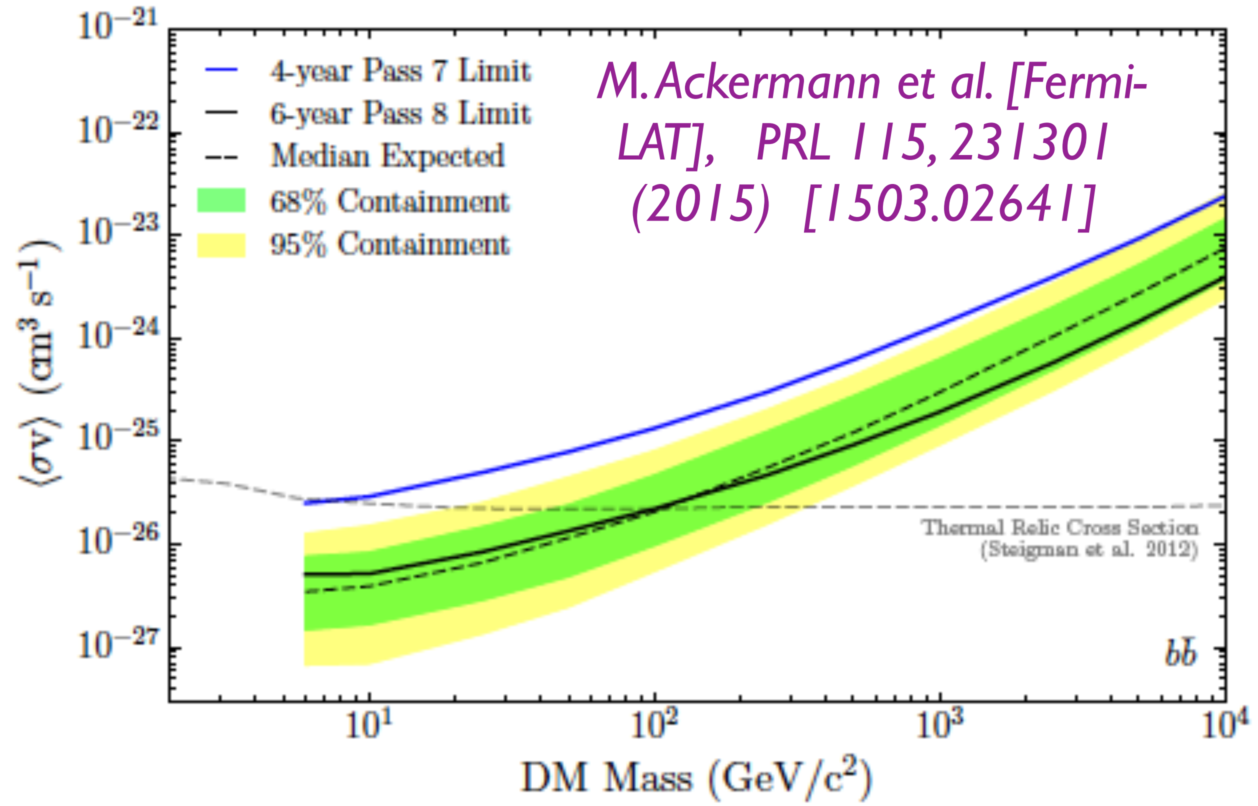
based on JCAP 10 (2018) 029 [1803.05508];
in collaboration with Francesca Calore & Bryan Zaldivar

Pasquale Dario Serpico (Annecy, France)
Fermi Symposium, Baltimore - Oct. 18, 2018

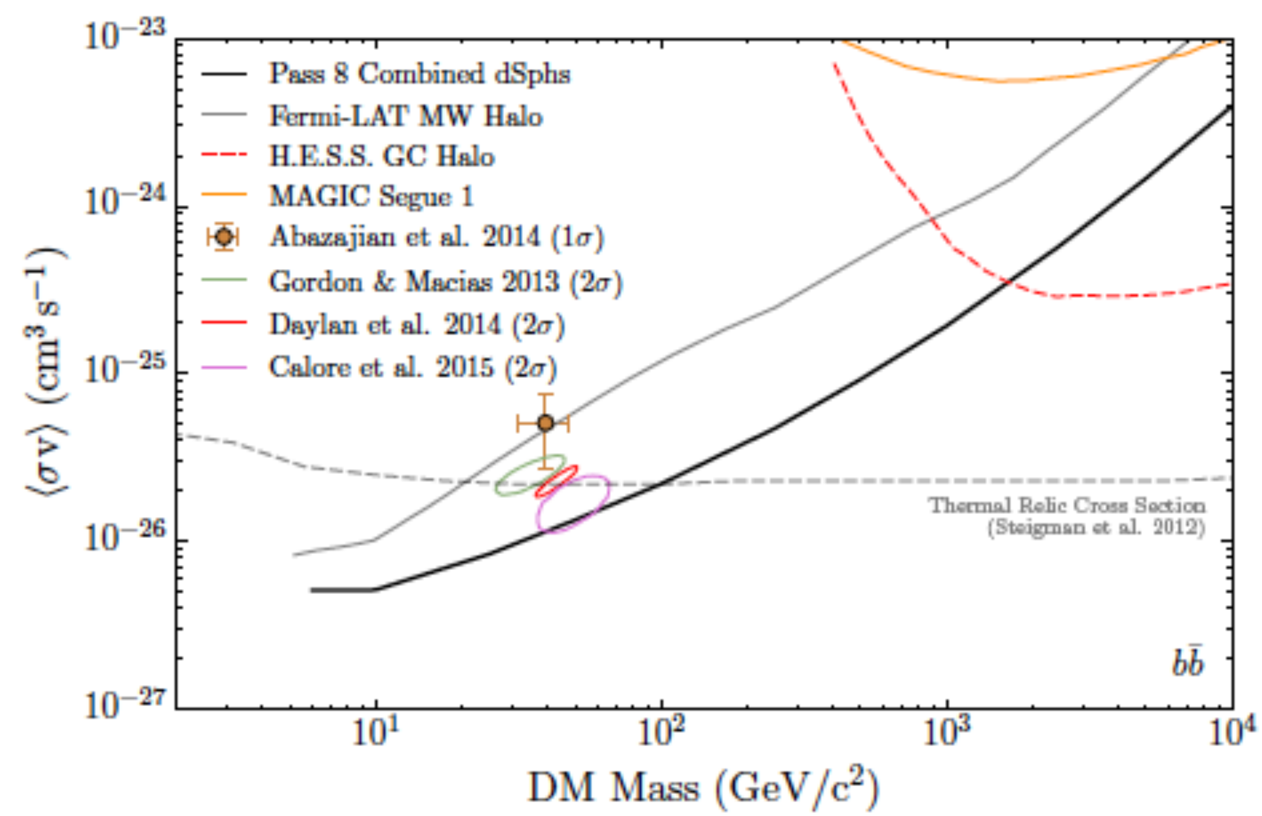
A primary WIMP DM target in γ-rays: Dwarf Spheroidals

satellites of Milky Way with high DM/baryon content, 1 to 3 orders of magnitude higher than the MW. Almost ideal Signal/Noise, even better if stacked!





Signal depends on _distance_ & _volume average of DM density²_, (so-called _J-factors_). **Can probe thermal s-wave relics annihilating into b's up to ~ 100 GeV**



On current assumptions/procedure

✓ dSphs assumed without intrinsic background, expected to be one to several o.o.m. below Fermi-LAT threshold, see *M. Winter et al. Astrophys. J. 832, no. 1, L6 (2016) [1607.06390]*

✓ Still affected by "accidental" background due to line of sight emissions (diffuse Galactic, extended Galactic sources, and both Gal. and extra-gal. point-like sources)

Current Fermi-LAT's procedure (outsider's view!)

✓ independent determination of background in a 15°x15° region around each dwarf
✓ predefined background models (diffuse and isotropic) where only normalization is fitted

Points to improve:

- new spatially-dependent contributions (unresolved sources, alternative diffusion mechanisms) may provide unequal performances in different regions of the sky
- no guarantee that background is consistently determined from one region to another
- somewhat arbitrary choice of background window
- estimation of (theoretical) systematic due to background modeling errors is hard or unclear

Alternatives

Some more "data driven" alternatives in the literature, like

A. Geringer-Sameth and S. M. Koushiappas, PRL 107, 241303 (2011) [1108.2914]

M. N. Mazziotta et al. Astropart. Phys. 37, 26 (2012) [1203.6731]

Still:

- somewhat arbitrary choice of background window
- all points within it "weighted equally"
- hard to evaluate how performing the method is
- no clear indications of paths for improvements

In

F. Calore, P. D. Serpico and B. Zaldivar, "Dark matter constraints from dwarf galaxies: a data-driven analysis," JCAP 10 (2018) 029 [1803.05508]

we propose a data-driven approach, agnostic about the (astro)physics underlying the background, tackling the above issues

This allows us to address the question of the robustness of Fermi-LAT dSphs results against (implicit or explicit) assumptions on the background

A preliminary comment

Often, machine learning applied to deterministic problems (e.g. classification):

Cat
or
cappuccino?

For details: M.-A. Fardin "On the Rheology of Cats", Rheol. Bull. 83, 16 (2014) [IgNobel prize for physics 2017]

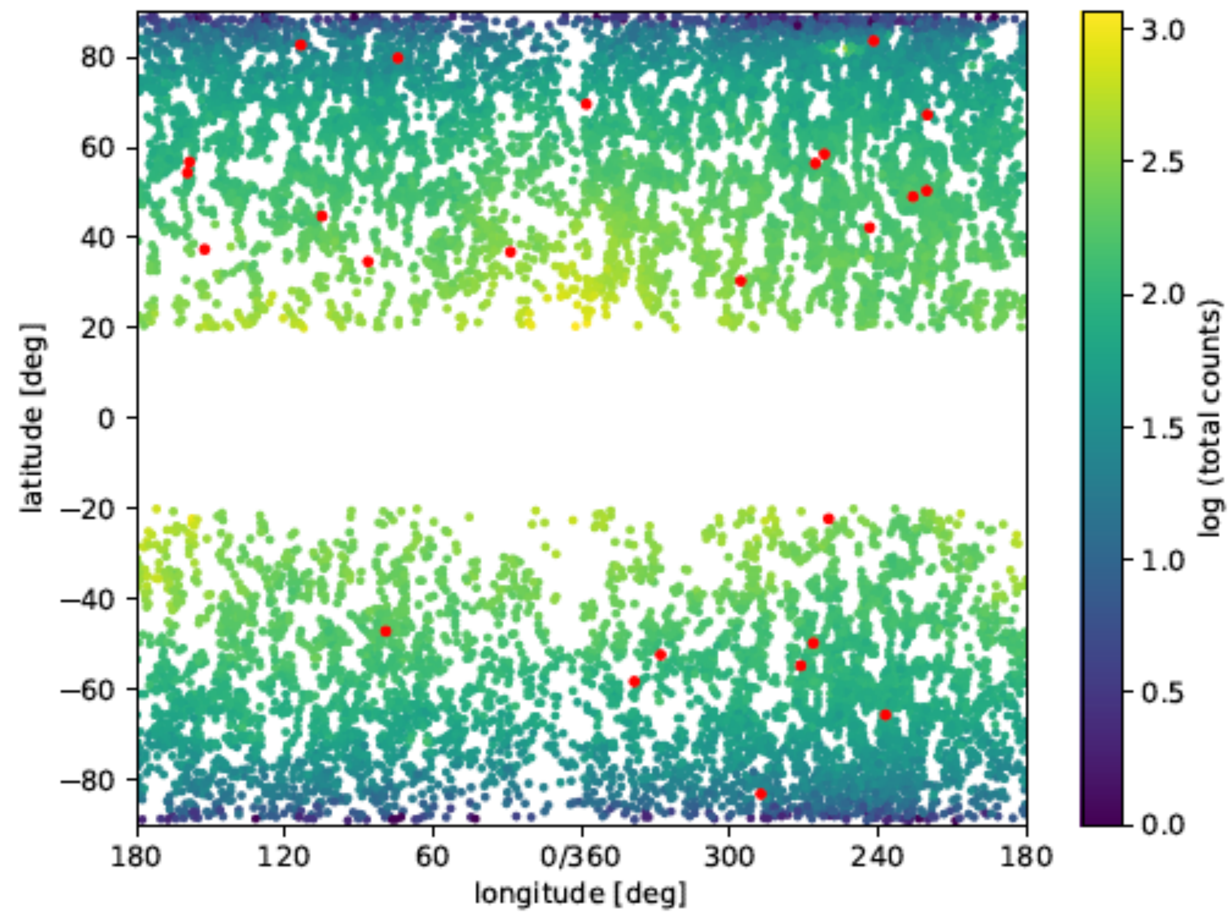
Our problem is different, since we expect that the background towards a given direction can be (at best) described in a statistical sense.

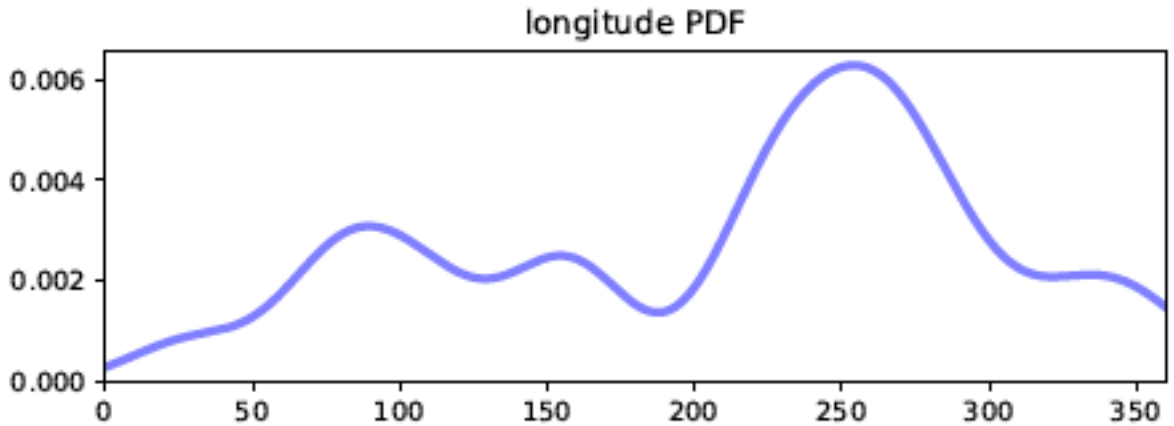
Would not very be effective to directly *interpolate* the (noisy) background in the sky

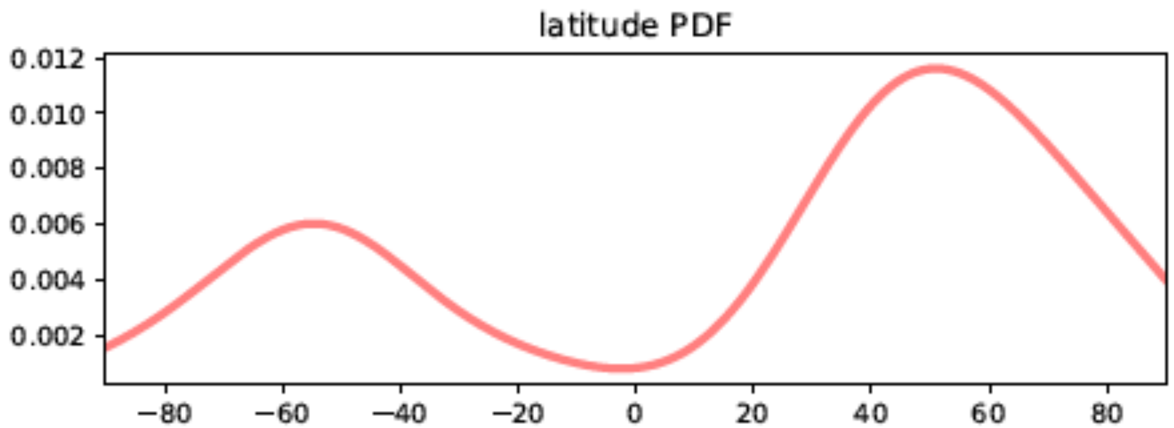
The goal will be rather to "optimize" the estimate of the background PDF (non-deterministic problem)!

A sketch of the procedure

✓ **Step 1**: remove dSphs, known pointlike and diffuse sources to build a "background sample"







to reduce chances of "hidden" systematics further cut the background by sampling it according to the smoothed distribution of dSph

turns out not to be very relevant, since the optimal estimator will be relatively local

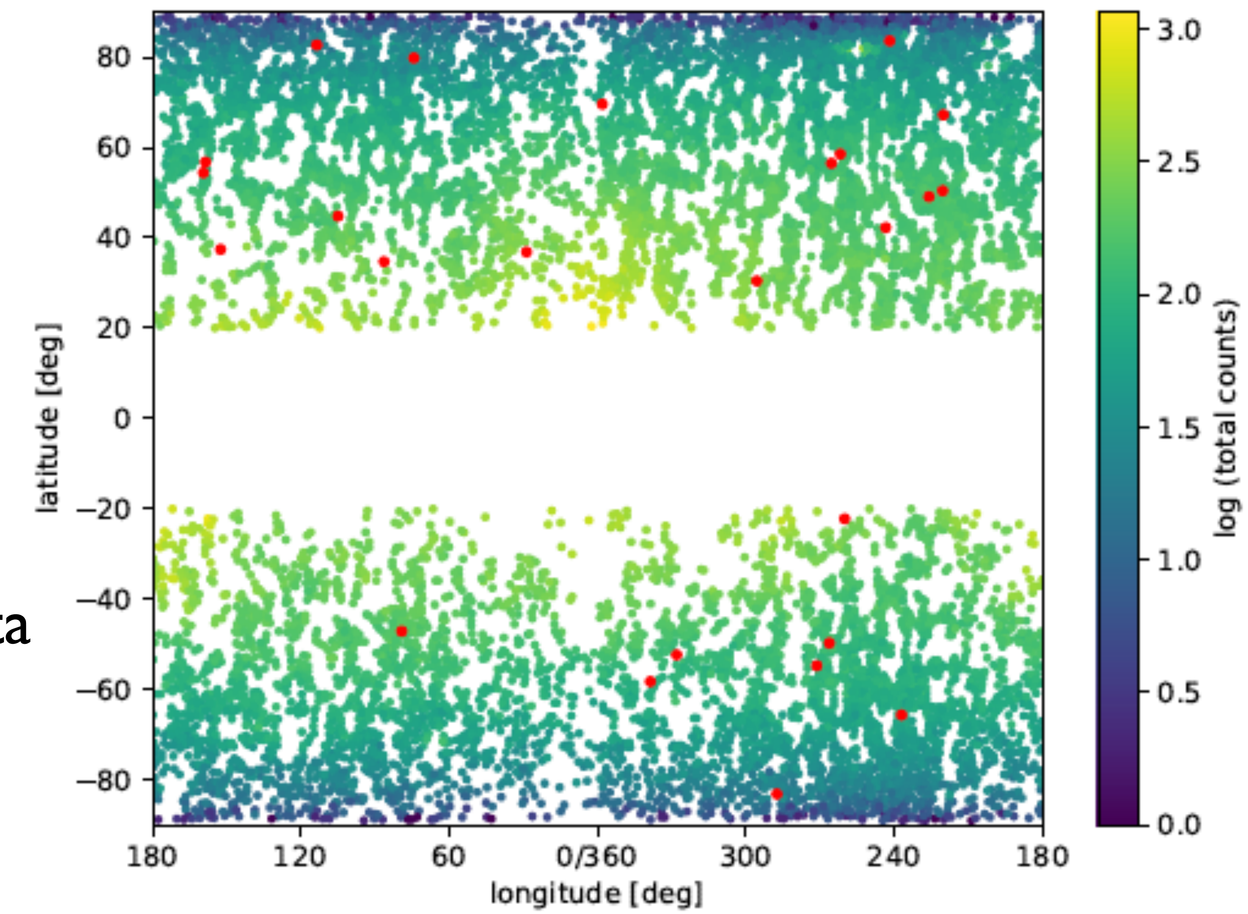
N≃9400 void regions remain

obtained with Kernel Density Estimation, python `scikit-learn` package

A sketch of the procedure

✓ **Step I**: remove dSphs, known pointlike and diffuse sources to build a "background sample"

re-sampled according to smoothed dSph distribution



✓ **Step II**: build global PDF estimator based only on data
(parameterized PDF according to E. Parzen '61, D. Specht '88,'90,'91, well-known theorems in statistics/Machine Learning community proving convergence to "true" PDF)

$$\hat{\mathscr{F}}(\overrightarrow{x}, y) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(\overrightarrow{x} - \overrightarrow{x}_i) g_\varsigma(y - y_i)$$

smooting can be "local" or "global", the latter chosen in the following for simplicity

spatial location kernel, depending on the smooting parameter σ

photon count kernel, depending on the smooting parameter ς

In particular, we take a Gaussian for K and a log-normal for g, but this choice is not essential since the convergence to the true PDF for large N is assured under weak and general hypotheses of continuity and smoothness

Optimization

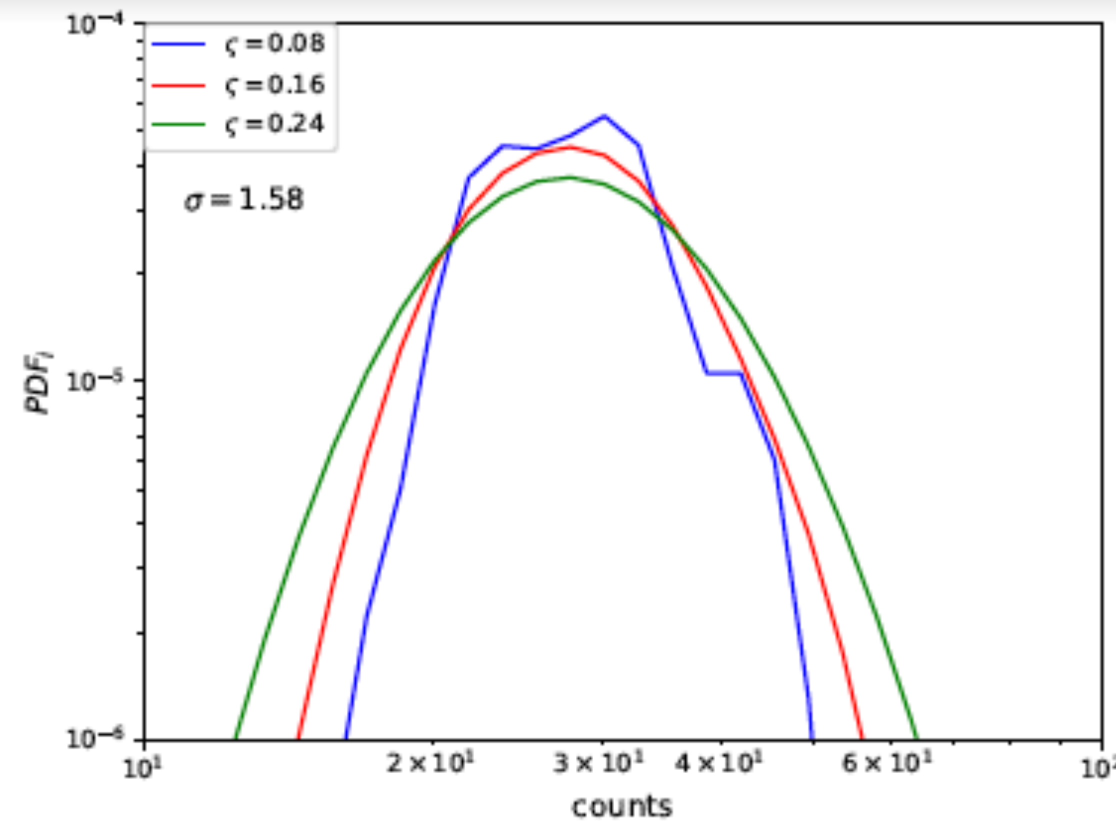
PDF at a typical position

(when varying a smoothing parameter)

In particular, low ς means low variance but large "bias" (comb-like PDF)

Large ς means high-variance (broadened PDF)

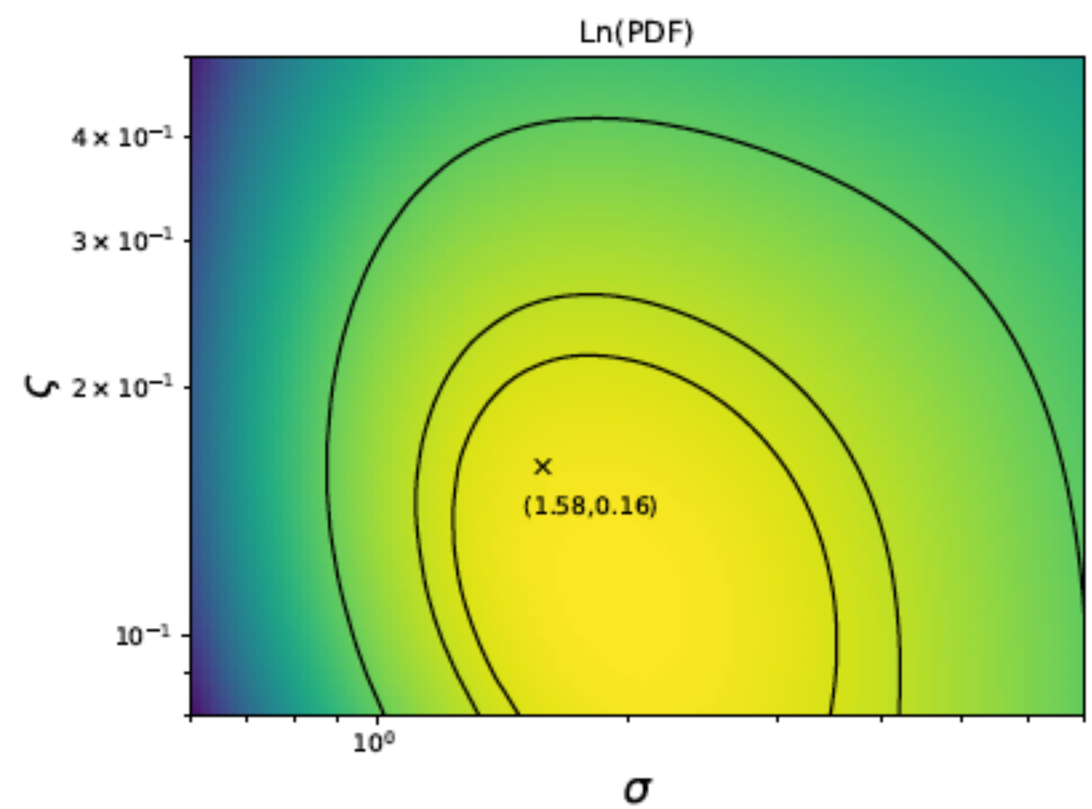
The red curve shows the "optimal" compromise



✓ **Step III**: maximize the (global) likelihood on the training background sample wrt the smoothing parameters

$$\ln \hat{\mathscr{F}}_{\mathrm{tot}}(\sigma, \varsigma) = \sum_{i=1}^{N} \ln \left[\frac{1}{N-1} \sum_{j \neq i}^{N-1} K_\sigma(\vec{x}_i, \vec{x}_j) g_\varsigma(b_i, b_j) \right]$$

$$\{\sigma^*, \varsigma^*\} = \mathrm{argmax} \;\; \ln \hat{\mathscr{F}}_{\mathrm{tot}}(\sigma, \varsigma) \approx \{1.58, 0.16\}$$



Evaluation

✓ **Step IV**: The PDF thus optimized on the "rest of the sky" can be used to evaluate the background PDF at the dwarf position, or if one wishes its statistical moments, e.g.

$$\widehat{\ln \mathrm{b}} = \frac{\sum_{i=1}^{n} K_i \ln b_i}{\sum_{i=1}^{n} K_i} \qquad \widehat{\mathrm{Var}}(\ln b)_{\sigma,\varsigma} = \varsigma^2 + \left[\frac{\sum_{i=1}^{n} K_i (\ln b_i)^2}{\sum_{i=1}^{n} K_i} - \left(\frac{\sum_{i=1}^{n} K_i \ln b_i}{\sum_{i=1}^{n} K_i} \right)^2 \right]$$

(beware, the PDF is far from Gaussian!)

✓ **Step V**: We construct the likelihood for dwarf *d*, energy bin *e*, accounting for Poisson statistics of the counts (including both background and signal), the PDF of the astrophysical inferred *J*-factors, and the Background PDF

$$\mathscr{L}_{d,e}(\lambda_{d,e}, \log_{10} J_d, \ln b_{d,e}) = \frac{\lambda_{d,e}^{c_{d,e}} e^{-\lambda_{d,e}}}{c_{d,e}!} \mathscr{N}(\log_{10} J_d) \boxed{\mathscr{B}(\ln b_{d,e})}$$

$$\lambda_{d,e} = \lambda_{d,e}(\langle \sigma v \rangle, m_{\mathrm{DM}}, \log_{10} J_d, \ln b_{d,e}) = 10^{\log_{10} J_d} \langle \sigma v \rangle f_{d,e}(m_{\mathrm{DM}}) + e^{\ln b_{d,e}}$$

from here on, standard "Fermi-like" (stacking and) profile likelihood method, as described in

M. Ackermann et al., Phys. Rev. D 89, 042001 (2014) [arXiv:1310.0828]

just extended to profiling over the J-factor and background PDF

Our choice of dSph & J-factors

Sample of 25 dSphs, all of which have J-factors estimated from spectroscopic measurements
J-factors within circular regions of 0.5°, with log-normal distribution

$$\mathcal{N}(\log_{10} J_d) = \frac{1}{\sqrt{2\pi}\sigma_d^J} \exp\left[-\left(\frac{\log_{10} J_d - \overline{\log_{10} J_d}}{\sqrt{2}\sigma_d^J}\right)^2\right]$$

19/25 taken from Table I of

A. Albert et al. [Fermi-LAT and DES Collaborations], Astrophys. J. 834, no. 2, 110 (2017) [1611.03184]

in turn based on the analysis

A. Geringer-Sameth, S. M. Koushiappas and M. Walker, Astrophys. J. 801, no. 2, 74 (2015) [1408.0002]

5/25 (Horologium I, Hydra II, Pisces II, Willman I and Grus I) from

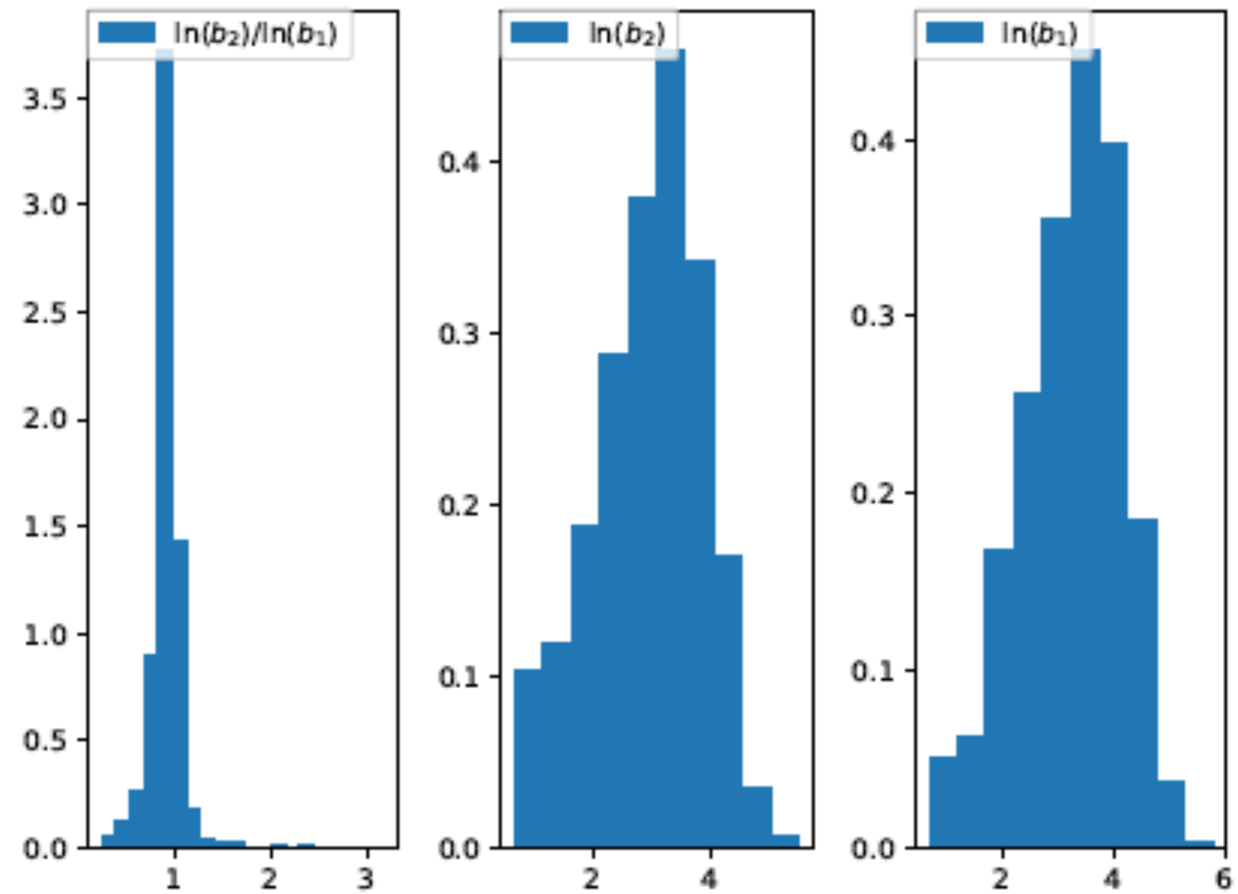
N.W. Evans, J. L. Sanders and A. Geringer-Sameth,
"Simple J-Factors and D-Factors for Indirect Dark Matter Detection," Phys. Rev. D 93, no. 10, 103512 (2016) [1604.05599]

1/25 (Tucana II) from

M. Walker et al. Astrophys. J. 819, 53 (2016) [1511.06296]

Energy dependence: simplification

The counts in different energy bins (same location) are strongly correlated, so that the PDF of their ratios is much narrower than the single bin PDF



The computations speed up significantly if we assume that the background E-distribution is fixed, and only the overall number is subject to profiling.

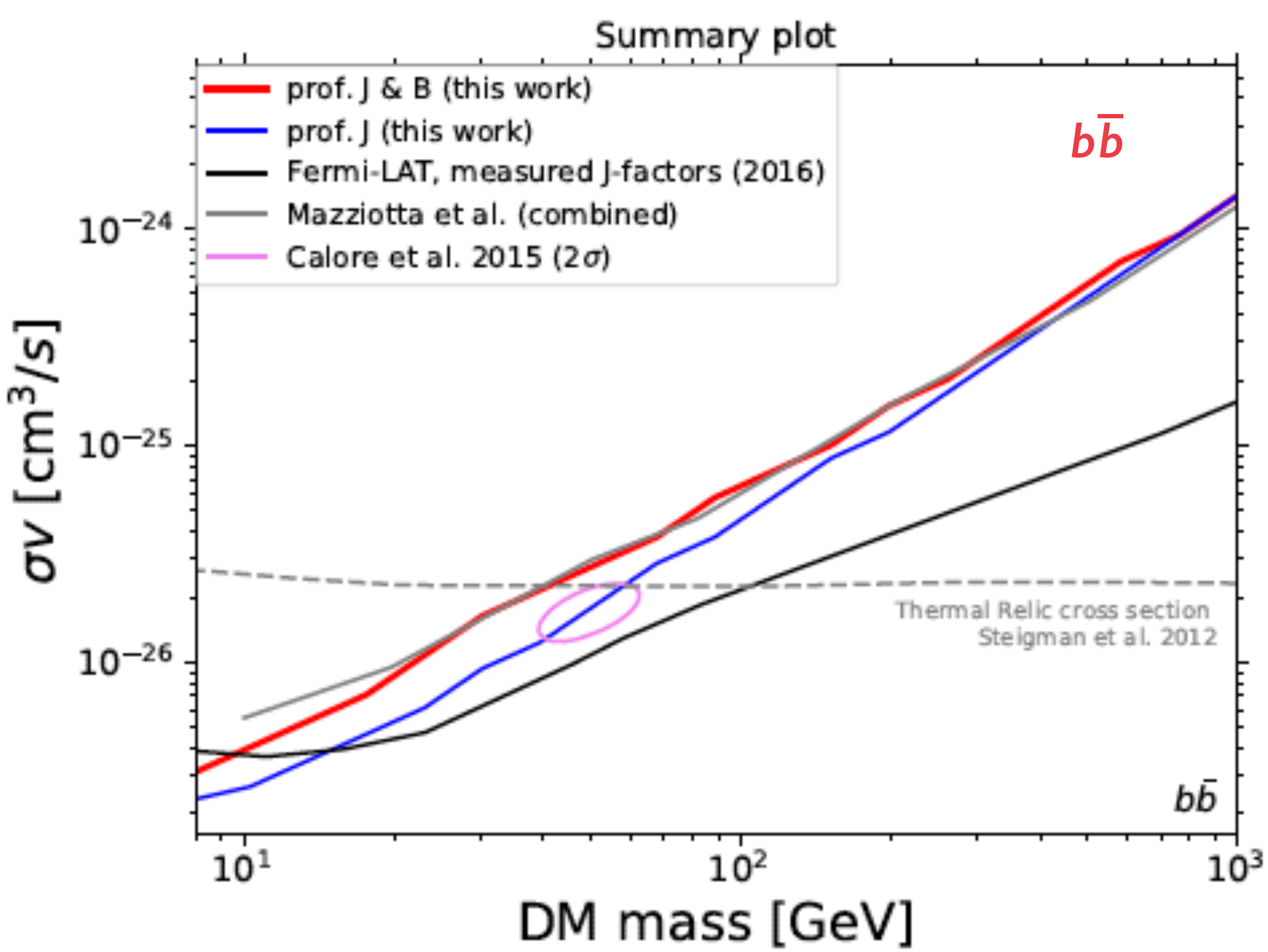
not a limitation of the method, just a useful trick to reduce the # of distributions to profile over.
Real bounds should be actually a little bit weaker because of this neglected effect…

Main results

✓If we do not profile over background PDF, we improve over previously proposed data-driven techniques and obtain results comparable to Fermi-LAT ones at small DM masses

(At large masses, lower sensitivity wrt Fermi, but we have all high-E data in a single bin to comply with our choice of "log" background estimator for comparable performance all-over the sky; loss of sensitivity expected and results more similar to "E-integrated" analyses)

✓When background uncertainty accounted for, bounds degrade up to a factor 2! Impact on cross-checks and multimessenger studies (Gal. Center Excess, antiproton excess, etc.)



Summary, and the road ahead

- We proposed an alternative, data-driven method to analyse Fermi-LAT results to derive a bound on DM from dSphs, to test the robustness of the bound wrt background assumptions and simplify the estimate of the theoretical error on the background modeling.

- The results obtained are comparable to Fermi-LAT ones when performing a similar analysis. However, profiling over the background distribution weakens the bounds up to a factor 2. This effect should be thus taken into account when assessing more precisely the bounds from dSphs.

- Our goal was mainly methodological. Room remains for improvements (green ongoing)

  - a local (vs. global) optimization for the kernel parameters, σ, ς and then inter/extrapolate those
  - account for more realistic J-factor PDF
  - test different kernels
  - introduce more kernel optimization parameters
  - account for energy-bin dependent background PDFs
  - "hybrid" strategies, including some theory prior on the description of the background?
  - apply to non-DM problems
  - (you name it!)

Thank you!

Backup

Technical details on how the bounds are obtained

Neyman-Pearson lemma

The log likelihood quantity gives the test statistics (TS) with maximum statistical power

Wilks theorem

If data are distributed according to $\mathscr{L}(\theta_1, \ldots, \theta_N)$, the maximum log likelihood ratio TS

parameters of interest

"nuisance" par. (profiling)

$$\lambda(\theta_1, \theta_2, \ldots, \theta_k) = -2 \ln \frac{\mathscr{L}(\theta_1, \ldots, \theta_k, \theta^*_{k+1}, \ldots, \theta^*_N)}{\mathscr{L}(\theta^*_1, \ldots, \theta^*_N)}$$

where θ_i^* are max likelihood estimators of the parameters, i.e.

$$(\theta^*_{k+1}, \ldots, \theta^*_N) = \underset{\theta_{k+1},\ldots,\theta_N}{\mathrm{argmax}} \{\mathscr{L}(\theta_1, \ldots, \theta_k, \theta_{k+1}, \ldots, \theta_N)\}$$

tends to a χ^2 distribution with k degrees of freedom, provided that θ_i^* are normally distributed

(appropriateness checked in *M.Ackermann et al. [Fermi-LAT], PRL 115, 231301 (2015)*)

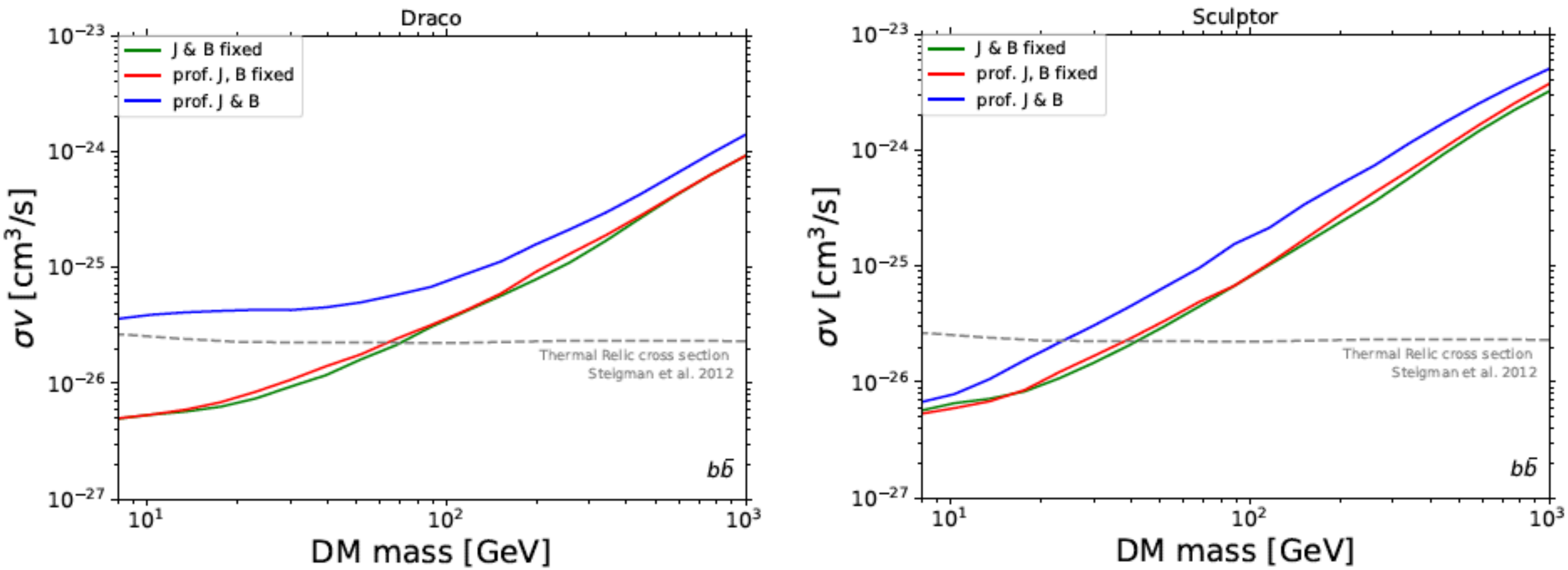
for each mass, 95% CL of <σ v> estimated from standard χ^2 distribution intervals

Numerical results

dwarf	name	$\log_{10} J \pm \sigma^J$	$c \pm \sqrt{c}$	$\bar{b}$	$\widehat{\ln b} \pm \Delta(\ln b)$
1	Boötes I	18.2 ± 0.4	70 ± 8	71.99	4.28 ± 0.2
2	Canes Venatici I	17.4 ± 0.3	22 ± 5	29.29	3.38 ± 0.25
3	Canes Venatici II	17.6 ± 0.4	19 ± 4	20.14	3.0 ± 0.31
4	Carina	17.9 ± 0.1	248 ± 16	259.13	5.56 ± 0.16
5	Coma Berenices*	19.0 ± 0.4	27 ± 5	27.61	3.32 ± 0.51
6	Draco*	18.8 ± 0.1	221 ± 15	292.8	5.68 ± 0.17
7	Fornax	17.8 ± 0.1	76 ± 9	70.14	4.25 ± 0.2
8	Hercules	16.9 ± 0.7	348 ± 19	308.32	5.73 ± 0.17
9	Horologium I	18.64 ± 0.95	125 ± 11	105.21	4.66 ± 0.21
10	Hydra II	16.56 ± 1.85	255 ± 16	275.46	5.62 ± 0.18
11	Leo I	17.8 ± 0.2	164 ± 13	152.69	5.03 ± 0.18
12	Leo II*	18.0 ± 0.2	49 ± 7	62.67	4.14 ± 0.21
13	Leo IV	16.3 ± 1.4	118 ± 11	121.44	4.8 ± 0.2
14	Leo V	16.4 ± 0.9	129 ± 11	112.39	4.72 ± 0.22
15	Pisces II	17.9 ± 1.14	178 ± 13	174.88	5.16 ± 0.18
16	Reticulum II	18.9 ± 0.6	124 ± 11	120.12	4.79 ± 0.21
17	Sculptor*	18.5 ± 0.1	14 ± 4	23.18	3.14 ± 0.34
18	Segue I*	19.4 ± 0.3	158 ± 13	138.5	4.93 ± 0.19
19	Sextans	17.5 ± 0.2	179 ± 13	166.76	5.12 ± 0.2
20	Tucana II	18.7 ± 0.9	122 ± 11	109.55	4.7 ± 0.18
21	Ursa Major I	17.9 ± 0.5	123 ± 11	104.78	4.65 ± 0.19
22	Ursa Major II	19.4 ± 0.4	314 ± 18	231.32	5.44 ± 0.17
23	Ursa Minor*	18.9 ± 0.2	187 ± 14	172.49	5.15 ± 0.18
24	Willman 1	19.29 ± 0.91	96 ± 10	93.87	4.54 ± 0.18
25	Grus I	17.96 ± 1.93	108 ± 10	80.37	4.39 ± 0.19

A few objects dominate the combined limit (*)
Background and counts agree within "~2 sigmas" (background PDF non-gaussian, a bit better approximated by log-normal… still not very well!)

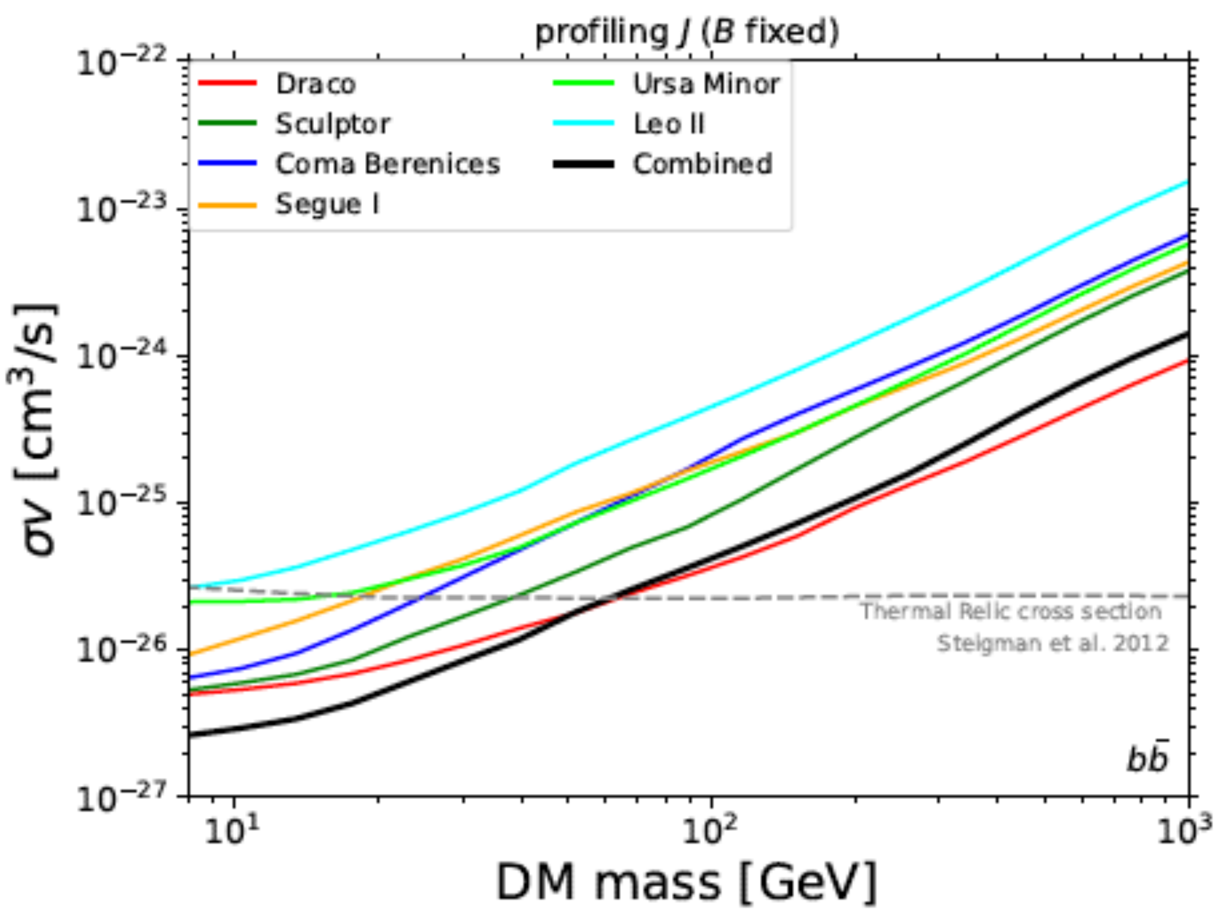
Example for single dwarfs

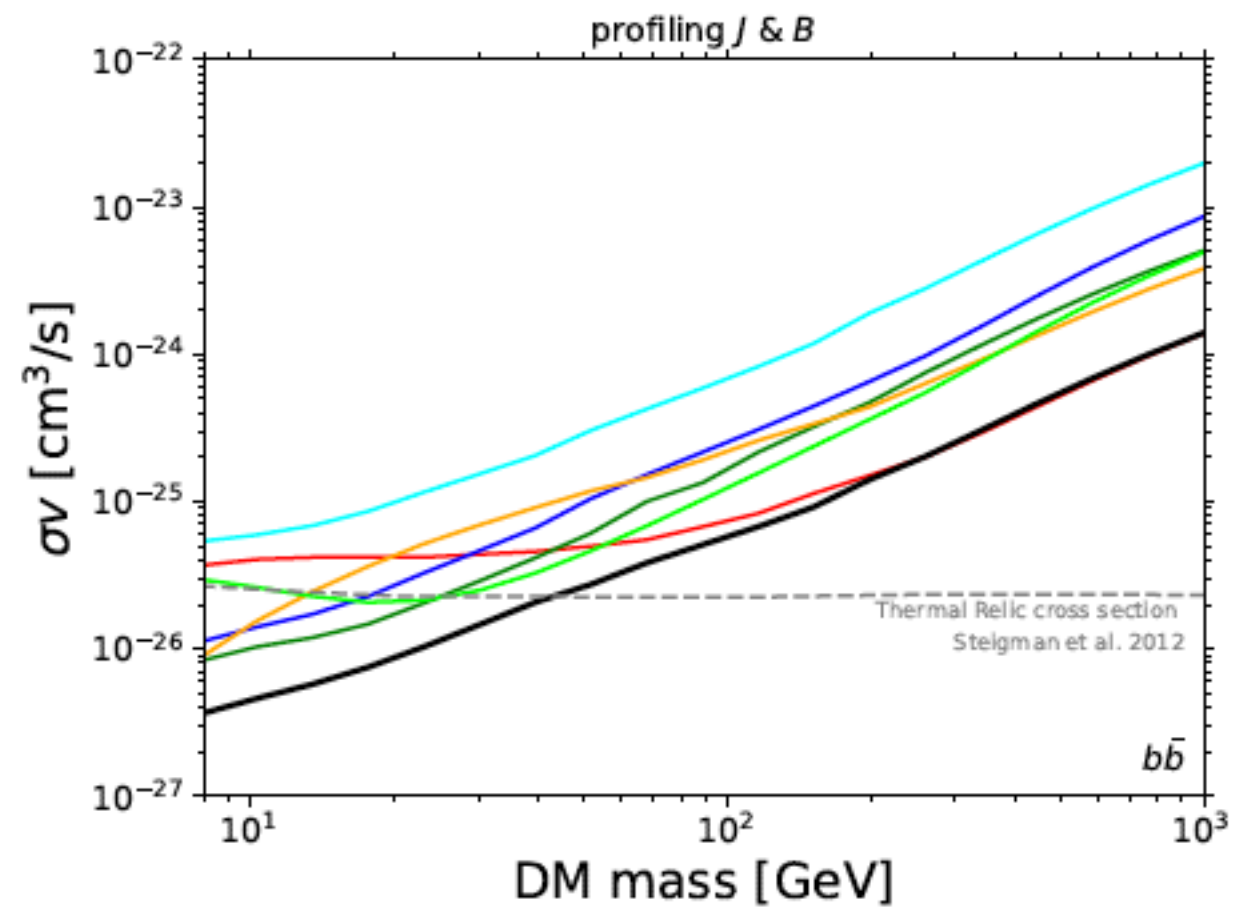


varying the background allows for a larger DM signal, lowering the bounds

Example for multiple dwarfs

The stacking usually improves the bound, but marginal worsening possible if excesses are present…





The excesses "go away" (globally speaking) once profiling over background; the stacked bound is better than any single one, but worse than the one obtained by profiling over J-only.