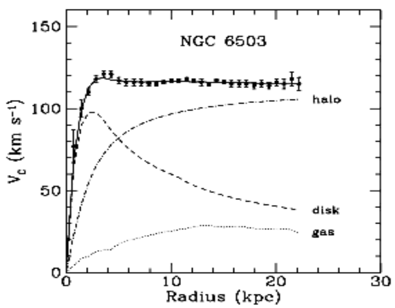
Dark matter constraints from dwarf galaxies with machine-learning

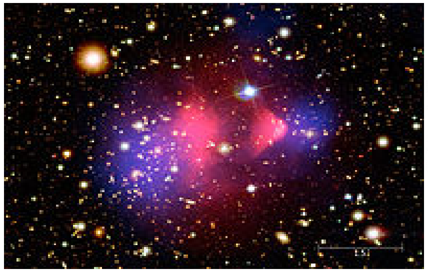
Bryan Zaldívar
IFT Madrid

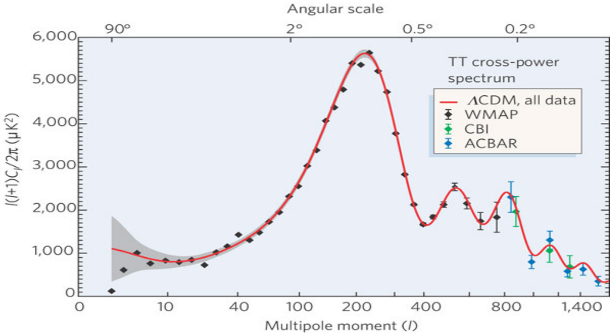
Based on:
1803.05508 (in collaboration with Francesca Calore & Pasquale D. Serpico)

DM & Indirect Detection

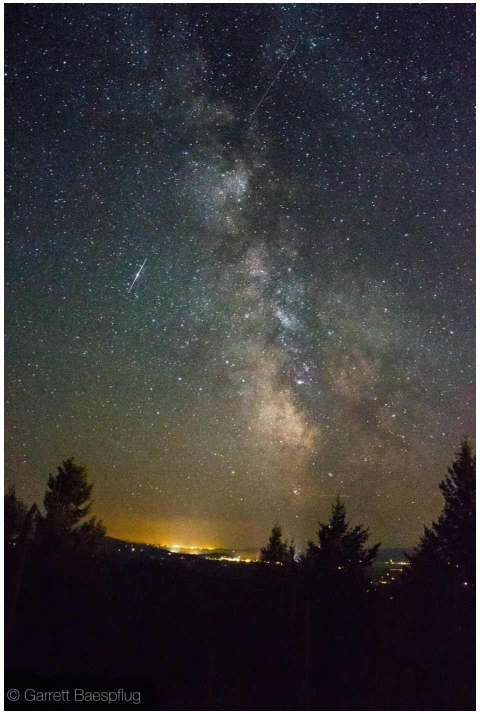
- compelling evidence for existence of DM (only gravitational)



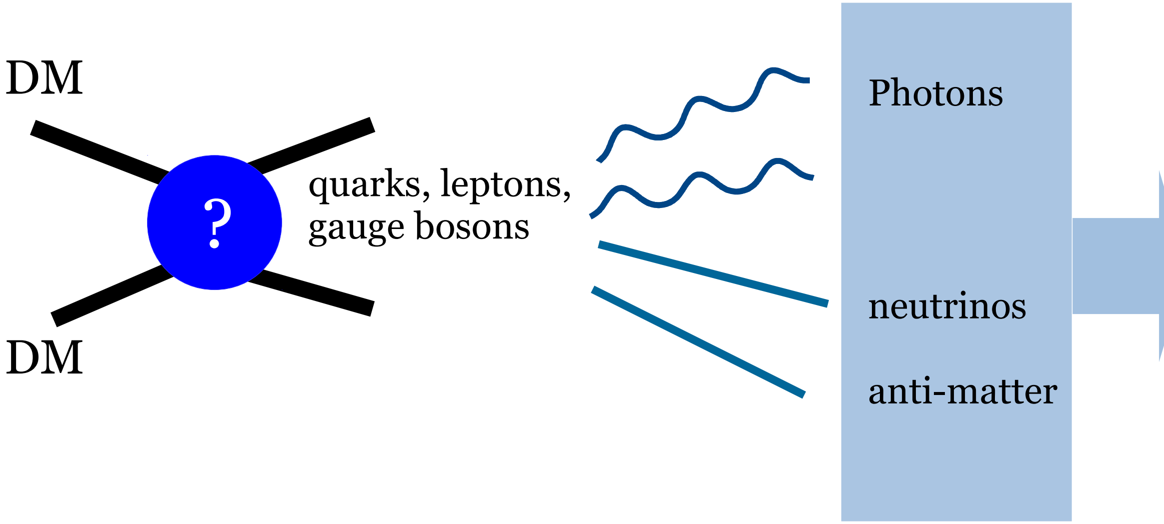




- we are searching for non-gravitational interactions of DM with baryonic matter (us)

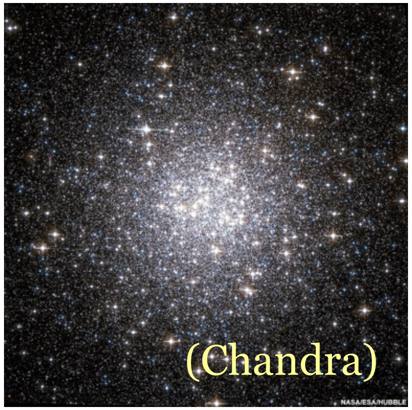


- indirect detection: *search for excess in cosmic ray flux from dense regions in the sky*



dwarf Spheroidal galaxies (dSphs)

Coma Berenices dSph



- Milky Way galaxy satellites

- $\mathcal{O}(100)$ kpc away from the Galactic centre

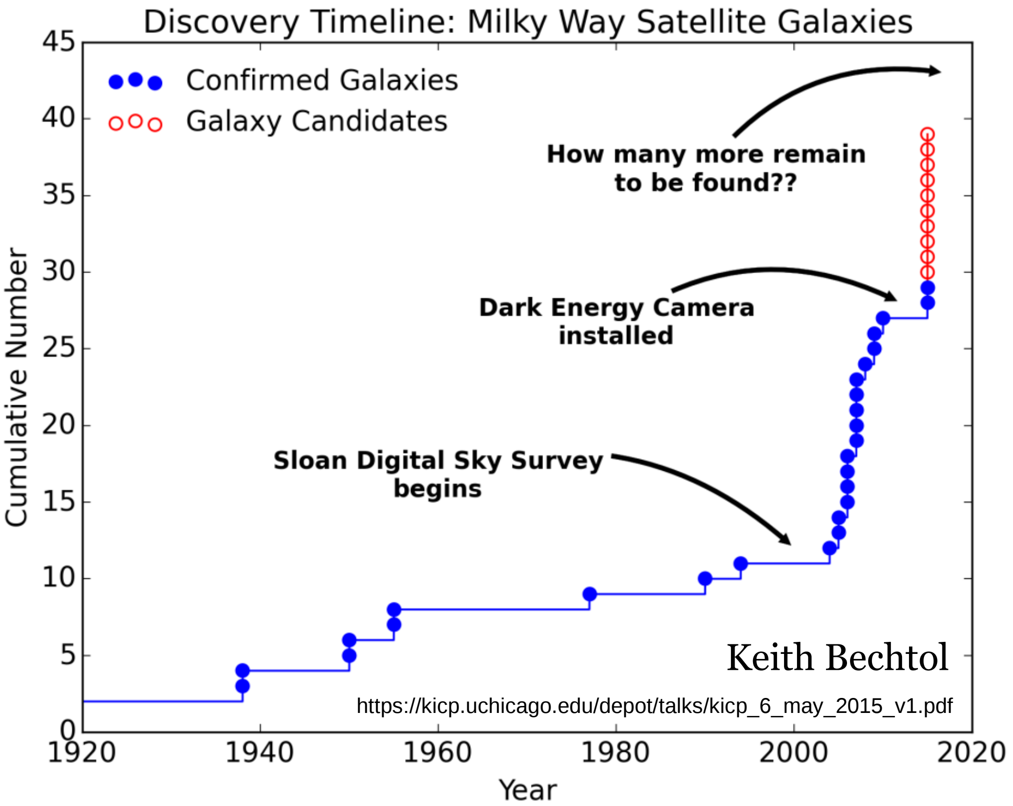
- DM dominated objects as shown by the kinematics of their stars

Mateo, astro-ph/9810070

Expected to shine in gamma-rays mostly from their DM content

(If DM has non-gravitational interactions with the SM)

A priori very clean laboratories for DM searches!



Constraints on DM from dSphs

J-factors can be measured independently

$$\frac{d\Phi}{dE} = \frac{\langle\sigma v\rangle}{8\pi m_{\rm DM}^2}\frac{dN_\alpha}{dE}J$$

Photon counts at dwarf d →

$$\lambda_{d,e}^{\rm DM} = \int_e \frac{d\Phi}{dE}\mathcal{E}(E)dE$$

} dark matter contribution (known, fixed hypothesis to test)

total expected no. of events:

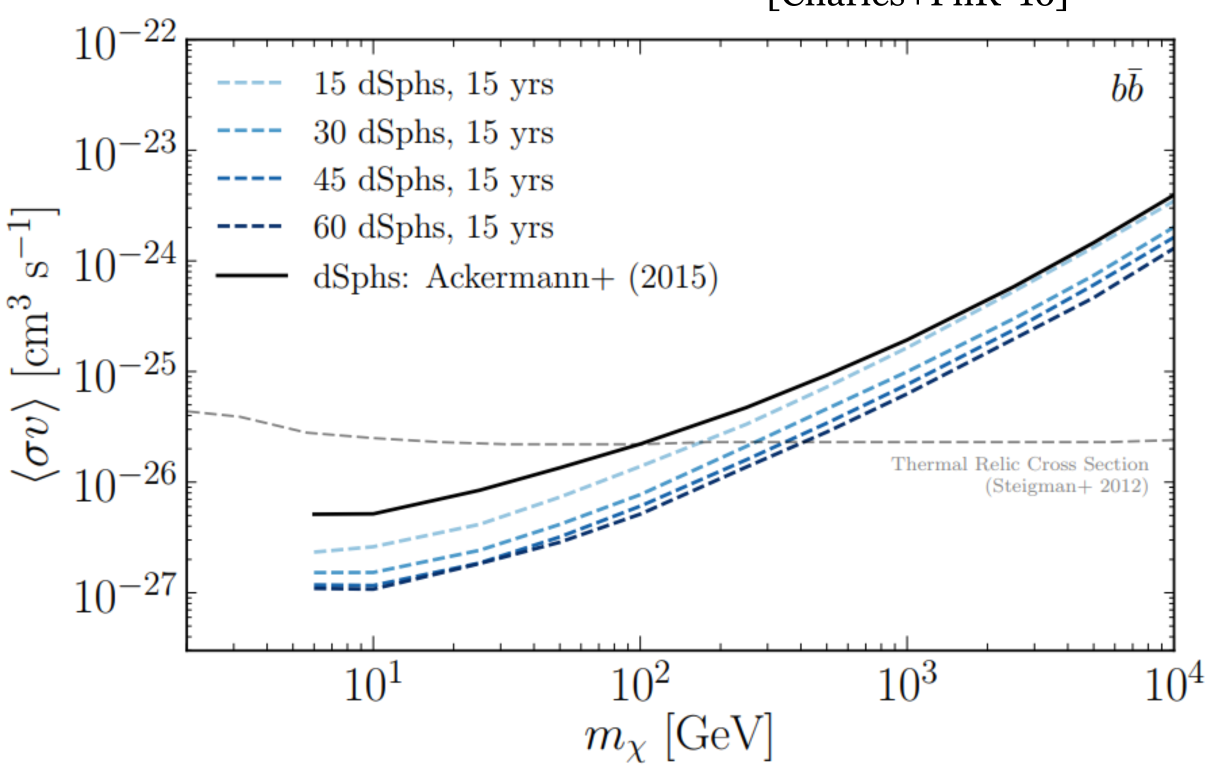
$$\lambda_{d,e}^{\rm tot} = \lambda_{d,e}^{\rm bkg} + \lambda_{d,e}^{\rm DM}$$

$\lambda_{d,e}^{\rm bkg}$: background (everything not being DM)

make the measurement, build your likelihoods,

If no significant excess put limits on DM contrib.

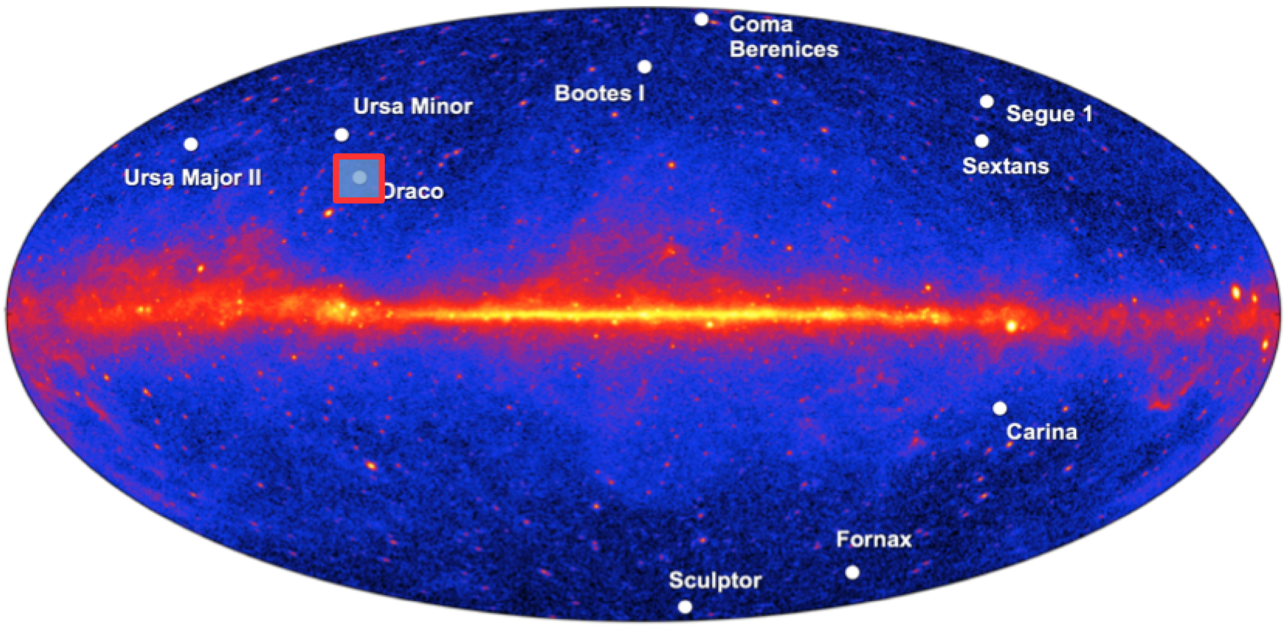
[Charles+PhR '16]



Current Fermi-LAT's procedure

(from non-expert opinion)

- predefined background models (diffuse, isotropic, PLS) where *only normalisation is fitted*
- *independent* determination of background in a 15^{o}x15^o region around each dwarf



Points to improve:

- new (unresolved) spatially-dependent contributions may provide unequal performances in different regions of the sky
- no guarantee that background is consistently determined from one region to another
- Estimation of (theoretical) systematic errors is unclear

A data-driven alternative for bckg estimation

- Be agnostic about a possibly underlying physics as for background is concerned

- *Build a global estimator* based only on data, from reasonably well-defined control regions

- Extrapolation to estimate the background contribution on dwarfs (signal regions)

- Include background uncertainties in the statistical analysis

---

Regression problem

ML approach

("Supervised Learning")

Suppose data is: $(\vec{x}_i; y_i)$ $\quad i = 1, ..., N$ points

$\vec{x}_i$: D-dimensional input vector (fixed)

y_i : output (random variable)

Parametric methods:

- *assume a specific shape for the distribution (Likelihood) of data:* y
  e.g. **Gaussian**
- assume identical-variance for all data points: $\sigma^2(\vec{x}) = \sigma^2$
- model the mean at each point

e.g. basis function expansion: $\bar{y}_i(\vec{x}_i) = \sum_i^p \beta_i h_i(\vec{x}_i)$ (but also other non-linear models as neural networks)

Maximum-Likelihood-Estimator equivalent to minimise:

$$\mathcal{C} = \sum_{i=1}^{N} (y_i - \bar{y}_i)^2$$

Training phase: fitting the parameters for a given model complexity (value of p)
Testing phase: choosing the model complexity that best fit new data

ML approach

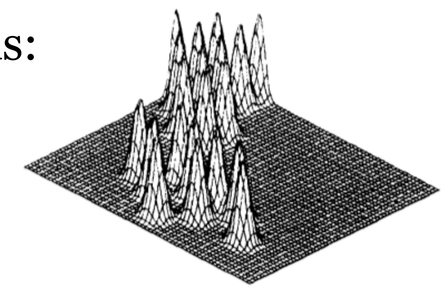
("Supervised Learning")

Non-parametric methods: (followed in this work)

- DOES NOT assume a specific shape for the Likelihood

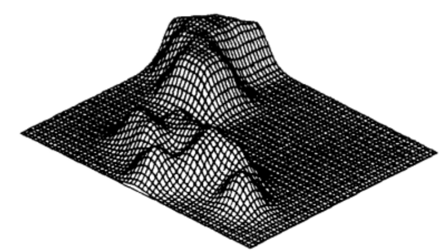
Parametrise the likelihood using *kernel density estimation* methods:

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



a. A small value of σ.

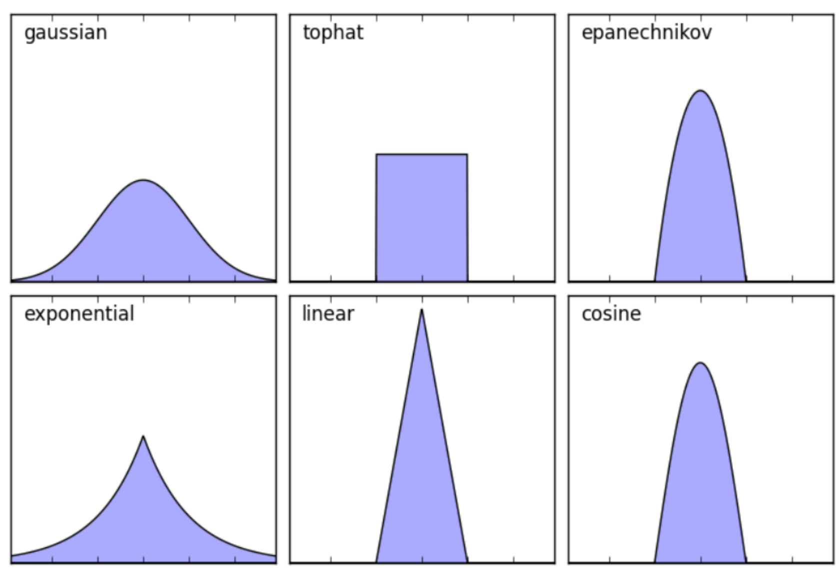
[Parzen '62] Under quite weak hypothesis (continuity, smoothness) *it provides an unbiased estimator of the true underlying PDF, $\mathcal{F}$*

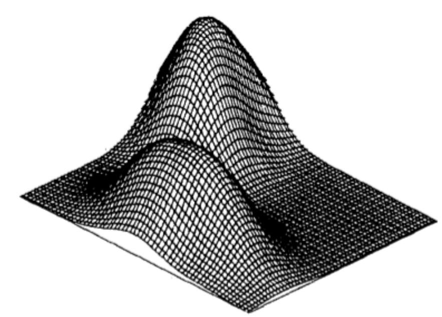


b. A larger value of σ.

Choice of kernels not unique

[Bishop '06]





c. An even larger value of σ.

[Specht '89]

Stating the problem

Aims:

1) *A 100% data-driven estimation of the (PDF of the) background emission at dwarf positions,*
2) *A consistent treatment of background uncertainties, when setting limits on DM annihilation cross section*

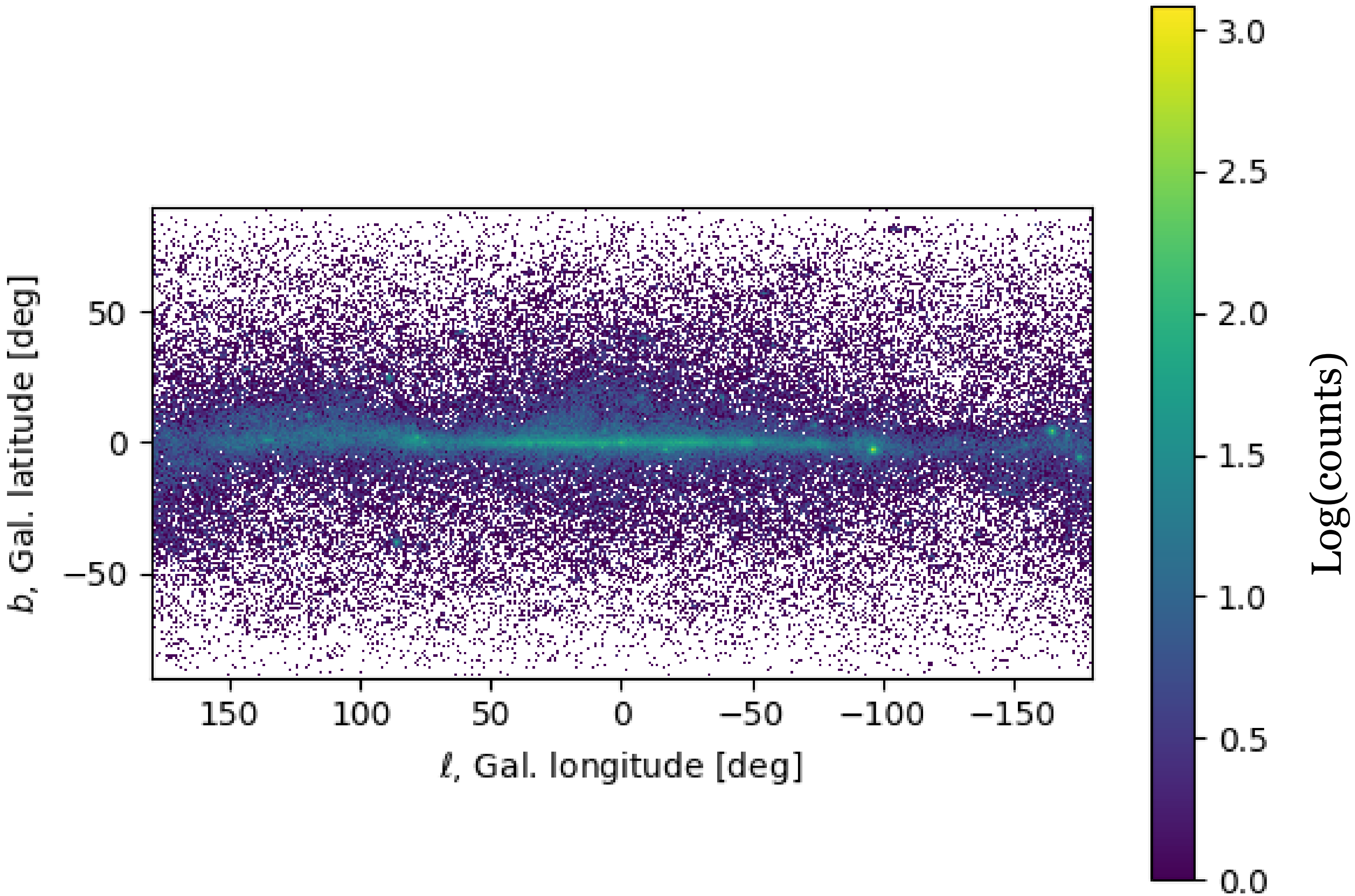
What is needed:

- *Reasonable definition of the "control region" (background-only)*
- *Concrete ML method for estimating the background distribution*
- *Likelihood construction*

Which data is used:

- *Fermi-LAT public data (2008-2015)* `SOURCE Fermi-LAT P8R2` *class data*
  All-sky, binned in Cartesian coordinates with pixel size 0.1 deg
  500MeV – 500GeV energy range in 24 bins

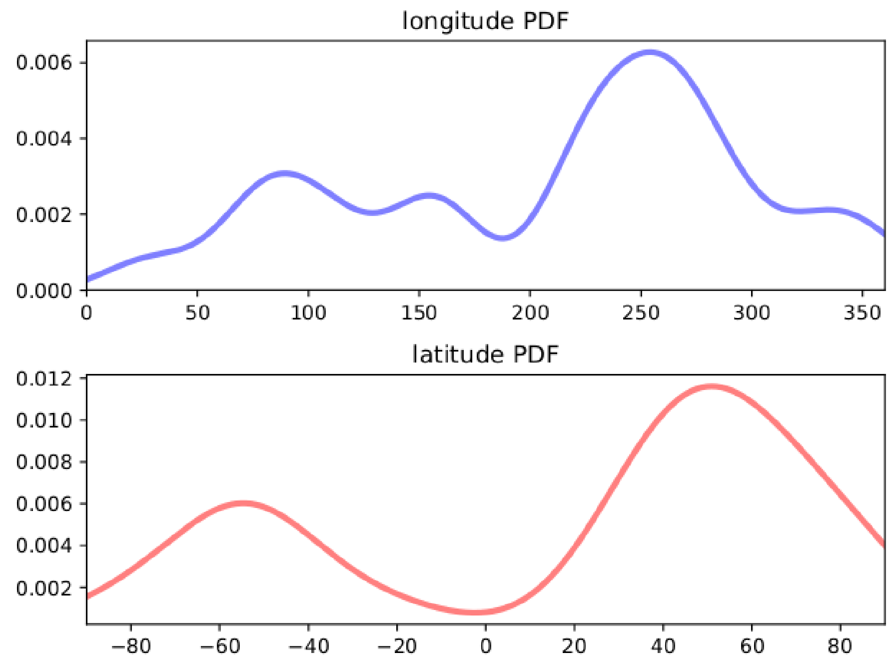
How does data looks like?

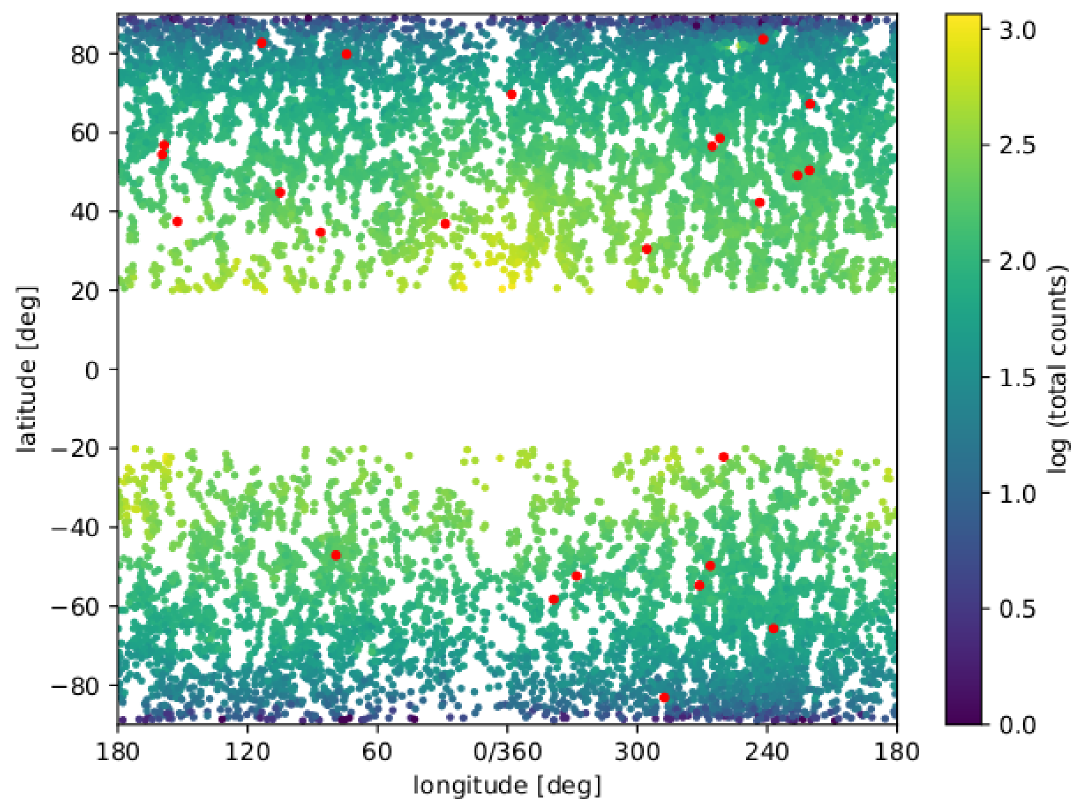


Control regions

As many non-overlapping circular regions as possible(*),
identical in size as dwarfs (r= 0.5°) and spatially distributed as the dwarfs

Dwarfs distribution





(*) removing:

- dwarfs themselves
- galactic disk $(|b| \leq 20°)$
- point-like sources from the catalog 3FGL
- extended sources ” ” ”

Note:
dwarfs and control regions are both superposed to the background DM halo

Optimum PDF parameters

Maximum-Likelihood-Estimator

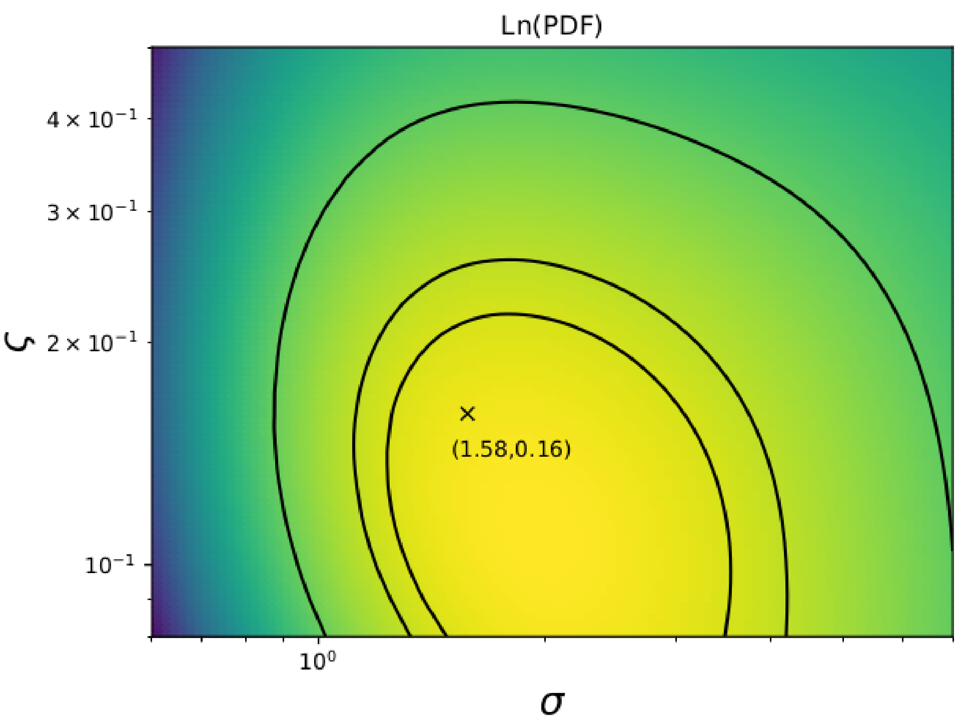
at data point $\vec{x}_i = (\ell_i, \flat_i)$, the PDF is built from the N-1 remaining data points

The total (log) Likelihood is:

$$\ln \mathcal{F}(\sigma, \varsigma) = \sum_{i=1}^{N} \ln \left[\frac{1}{N-1} \sum_{j \neq i}^{N-1} K_\sigma(\vec{x}_i - \vec{x}_j) g_\varsigma(y_i, y_j) \right]$$

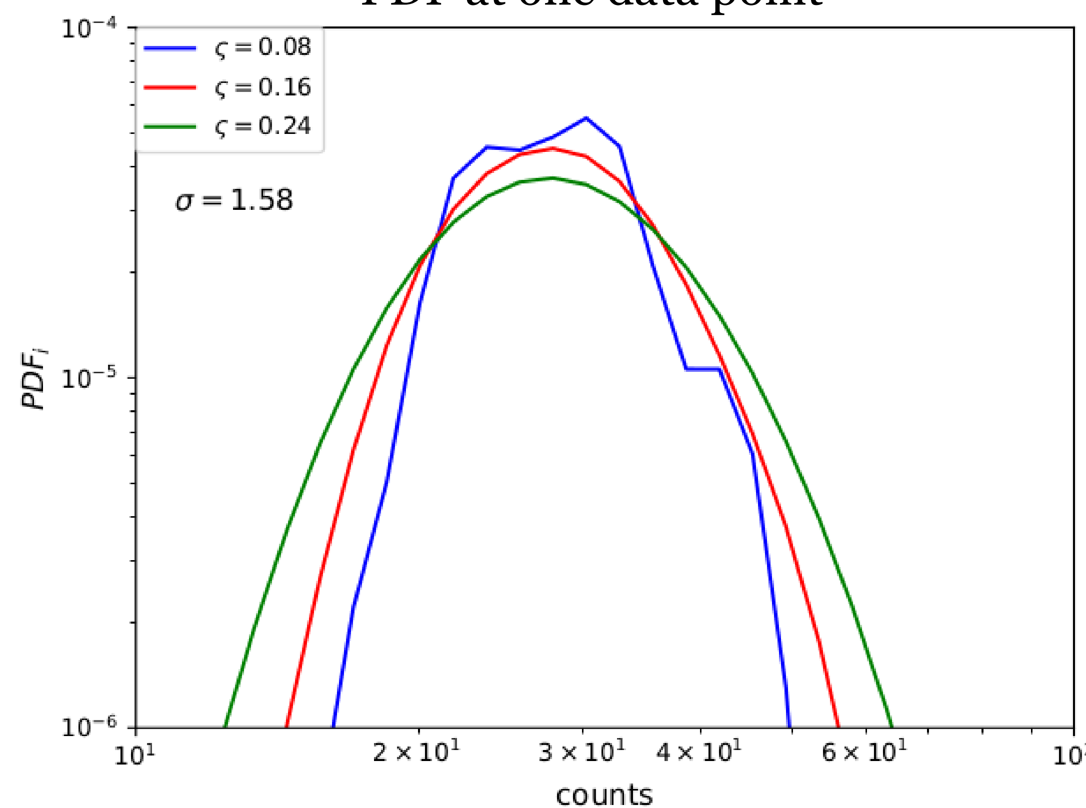
K_σ : Gaussian $\quad$ g_ς : Log-normal

y_i : photon counts



$$\{\sigma_*, \varsigma_*\} \approx \{1.58 \text{ deg}, 0.16\}$$

Data is noisy!



Bckg predictions at dSphs

The PDF for dSph d is:

$$\mathcal{F}_d(\vec{x}_d, \ln b) = \frac{1}{N}\sum_{i=1}^{N} K_{\sigma_*}(\vec{x}_d - \vec{x}_i) \cdot g_{\varsigma_*}(\ln b - \ln b_i)$$

$\ln b$: logarithm of bckg counts

N : the whole sample of bckg-only regions

Expected value of bckg counts:

$$\langle \ln b_d \rangle = \frac{\int_{-\infty}^{\infty} \ln b \; \mathcal{F}_d db}{\int_{-\infty}^{\infty} \mathcal{F}_d db} = \frac{\sum_{i=1}^{N} K_{\sigma_*}(\vec{x}_d - \vec{x}_i) \ln b_i}{\sum_{i=1}^{N} K_{\sigma_*}(\vec{x}_d - \vec{x}_i)}$$

(in practice, a weighted average over all neighbours)

For example...

dwarf	Obs. counts	Exp. counts
Segue I	158	138.5
Sculptor	14	23.2
Coma Berenices	27	27.6
Ursa Minor	187	172.5
Leo II	49	62.7
Draco	221	292.8

Likelihood analysis

- In general the Likelihood for dSph d and energy bin e is:

new

$$\mathcal{L}_{d,e}(\lambda_{d,e}, \log J_d, \ln b_{d,e}) = \frac{\lambda_{d,e}^{n_{d,e}} e^{-\lambda_{d,e}}}{n_{d,e}!} \mathcal{N}(\log J_d) \; \mathcal{F}(\ln b_{d,e})$$

with: $\lambda_{d,e} = J_d \langle \sigma v \rangle f_{d,e}(m_{\rm DM}) + \ln b_{d,e}$

- We take the Likelihood for dSph d as:

$$\mathcal{L}_d(\langle \sigma v \rangle, \log J_d, \ln b_{d,1}) = \mathcal{N}(\log J_d) \mathcal{F}(\ln b_{d,1}) \prod_e^{N_e} \mathcal{P}(\langle \sigma v \rangle, \log J_d, \ln b_{d,e})$$

- The total (stacked) Likelihood is thus:

$$\mathcal{L}(\langle \sigma v \rangle, \; \log \mathbf{J}, \ln \mathbf{b}_1) = \prod_d^{N_d} \mathcal{L}_d(\langle \sigma v \rangle, \log J_d, \ln b_{d,e})$$

Nuissance parameters

$$\vec{\theta} \equiv (\, \log \mathbf{J}, \ln \mathbf{b_1})$$

- Profile likelihood method:

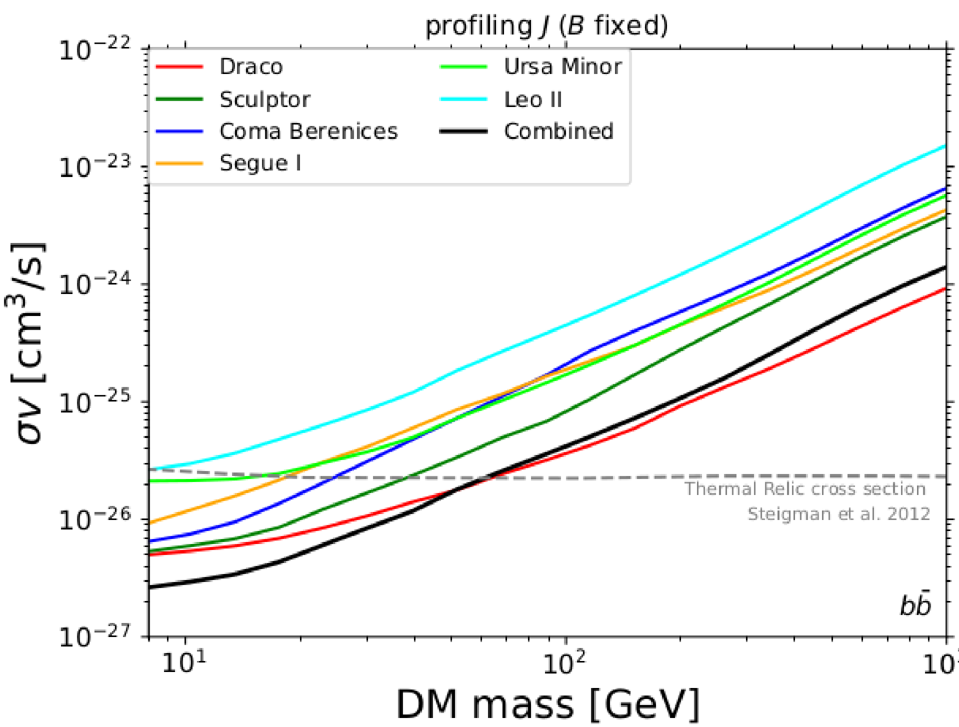
(for fixed DM mass)

$$\mathrm{TS}(\langle \sigma v \rangle) = -2 \ln \frac{\mathcal{L}(\langle \sigma v \rangle, \vec{\theta}^*(\langle \sigma v \rangle))}{\sup(\mathcal{L})}$$

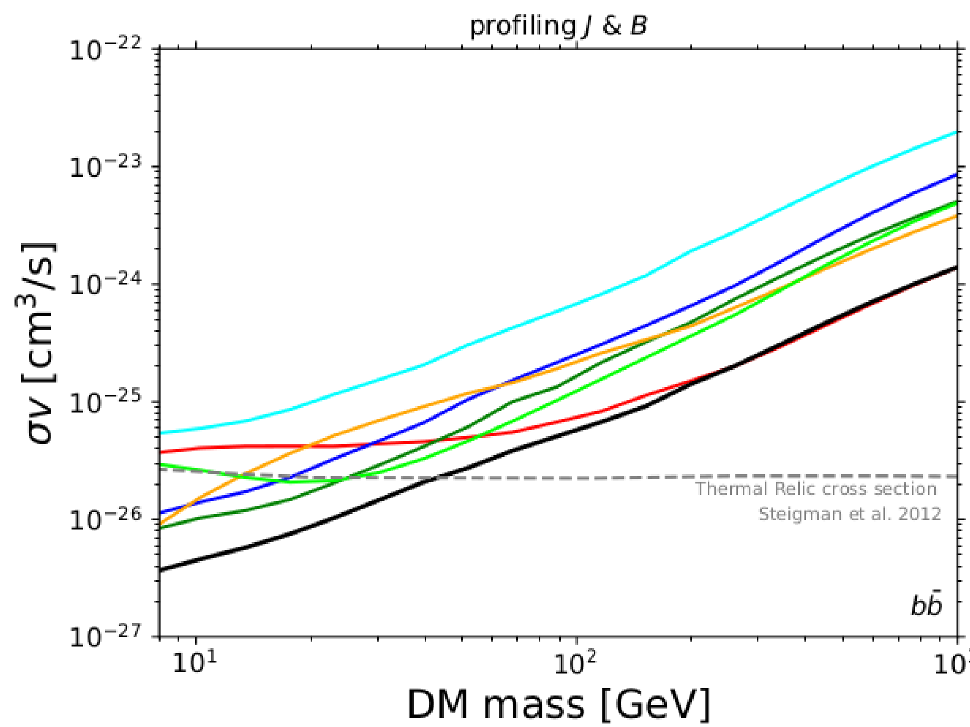
- Note we are only profiling over first energy bin (just for computational simplicity) since we assume maximum-correlation between bins (after some checks)

Dark matter limits

Profiling over J-factor,
bckg fixed to central value

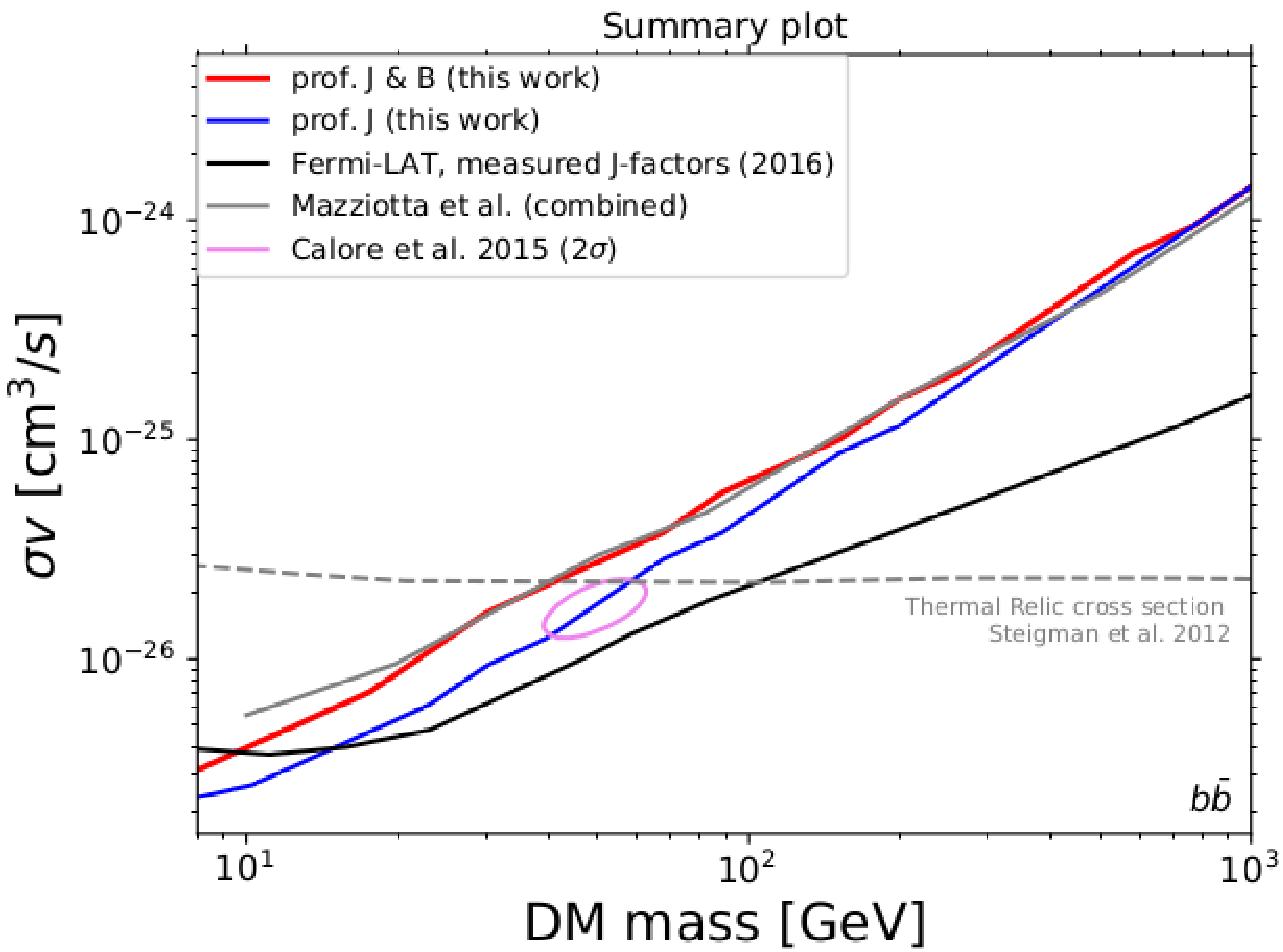


Profiling over J-factor
and bckg simultaneously



- Only showing the 6 dSphs giving strongest limits
- dSph ranking can change w.r.t. case of profiling only over J
- For a given dSph, difference between two limits change, depending on PDF moments

Dark matter limits



- Fixing the bckg, limits are stronger (weaker) than Fermi-LAT's for low (high) masses (optimised bckg modelling vs. limited statistics)
- Profiling also over the bckg, limits weaker by factor 1.5 (relevant e.g. for GC excess)

Conclusions

- Indirect detection of DM hampered by limited knowledge of astrophysical bckg

- Physical modelling of such bckg is imperfect; may not cover all information data provides

- Machine-learning (ML) methods can definitely help, if our aim is to understand DM

- Accounting for background uncertainties have an important impact on DM limits !

Thanks!

bckp

Existing applications of ML to DM

- *"Analyzing gamma-rays of the Galactic Center with Deep Learning"*
  Caron + , arXiv: 1708.06706

- *Improving the positional offset [of DM clusters] (Gravitational lensing)*
  Harvey +, arXiv: 1311.0704

- *DM-vs-neutrino discrimination in electromagnetic showers (data from OPERA)*
  https://www.kaggle.com/c/dmse-2-1

... and many more to come!

https://indico.cern.ch/event/664842/overview

https://indico.cern.ch/event/687473/overview

A model-independent analysis of the Fermi Large Area Telescope gamma-ray data from the Milky Way dwarf galaxies and halo to constrain dark matter scenarios

M. N. Mazziotta,[1,*] F. Loparco,[1,2,†] F. de Palma,[1,‡] and N. Giglietto[1,2]

[1] *Istituto Nazionale di Fisica Nucleare, Sezione di Bari, 70126 Bari, Italy*

[2] *Dipartimento di Fisica "M. Merlin" dell'Università e del Politecnico di Bari, I-70126 Bari, Italy*

(Dated: April 2, 2012)

We implemented a novel technique to perform the collective spectral analysis of sets of multiple gamma-ray point sources using the data collected by the Large Area Telescope onboard the Fermi satellite. The energy spectra of the sources are reconstructed starting from the photon counts and without assuming any spectral model for both the sources and the background. In case of faint sources, upper limits on their fluxes are evaluated with a Bayesian approach. This analysis technique is very useful when several sources with similar spectral features are studied, such as sources of gamma rays from annihilation of dark matter particles. We present the results obtained by applying this analysis to a sample of dwarf spheroidal galaxies and to the Milky Way dark matter halo. The analysis of dwarf spheroidal galaxies yields upper limits on the product of the dark matter pair annihilation cross section and the relative velocity of annihilating particles that are well below those predicted by the canonical thermal relic scenario in a mass range from a few GeV to a few tens of GeV for some annihilation channels.

https://arxiv.org/pdf/1108.2914.pdf