ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

MLFermiDwarfs

A data-driven framework for dark matter constraints from dwarf galaxies

C. Eckner, F. Calore, P. D. Serpico, B. Zaldívar

E-OSSR Onboarding Presentation

1st October 2021

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

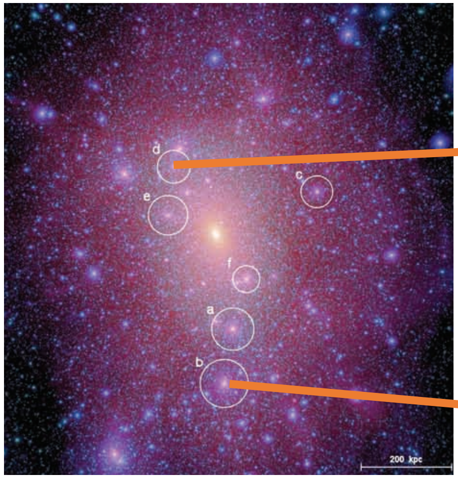
“MLFermiDwarfs” in a nutshell

- **Framework to derive constraints on the (velocity-independent) dark matter pair-annihilation cross-section utilising Fermi-LAT gamma-ray data of Milky Way dwarf spheroidal galaxies (indirect detection), which features a machine learning-based assessment of astrophysical background emission (intrinsic + extrinsic) from these objects.**
- Maximum likelihood approach with dark matter content of dwarfs (J-factor) and/or ML estimated astrophysical background as nuisance parameters
- For some classical dwarfs, data-driven (data on stellar kinematics) J-factor probability density functions are provided.
- Code entirely written in python 3 and relatively well-known packages like *scikit-learn.*
- **Use cases**:
  - Deriving an estimate for the astrophysical gamma-ray background of a Milky Way dwarf spheroidal galaxy in the Fermi LAT energy range that goes beyond the usual assumptions defined by the Fermi LAT collaboration.
  - Deriving constraints on, e.g. WIMP or user-defined dark matter model (dark matter parameters including a user-selected list of dwarf galaxies.
- Performance of the approach already demonstrated in two publications:
  - F. Calore et al. JCAP10 (2018) 029
  - A. Alvarez et al. JCAP09 (2020) 004

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

Why “MLFermiDwarfs”?

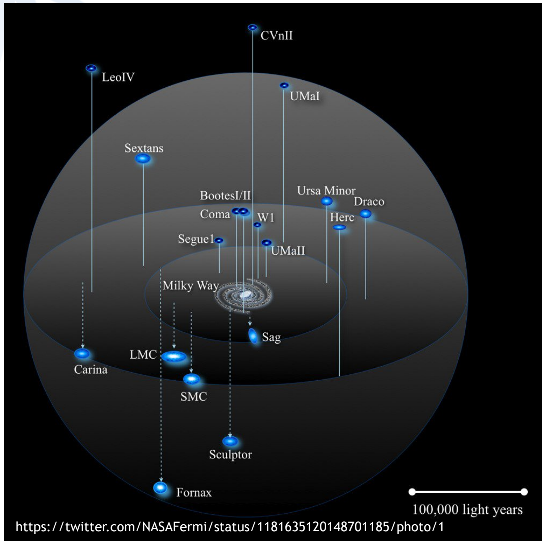
❖ Connection between dark matter, gamma rays and dwarf galaxies:



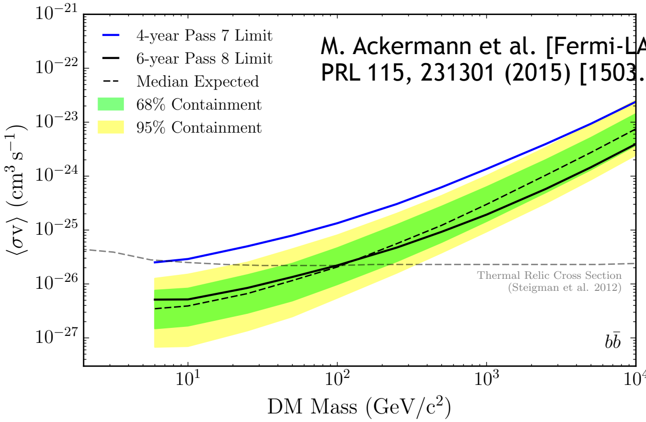
CDM-only, Aquarius project

$$\Phi_\gamma \propto \frac{1}{d^2} \int_{\mathrm{l.o.s.}} \rho_{\mathrm{DM}}^2 \mathrm{d}l$$

J-factor



❖ ratio of DM/baryonic matter in dwarfs almost 1 to 3 order(s) of magnitude larger than in the Milky Way + low background
—> “clean DM laboratories”: can constrain thermal WIMPy DM up to O(100 GeV)



M. Ackermann et al. [Fermi-LAT], PRL 115, 231301 (2015) [1503.02641]

Funded by the European Union’s Horizon 2020 - Grant N° 824064

Why “MLFermiDwarfs”?

- **General assumptions for such analyses**:
  — dwarfs have a negligible intrinsic gamma-ray background
  — “accidental” background might occur along the line of sight (diffuse emission from the MW’s plane, nearby extended Galactic sources, sub-threshold galactic and extragalactic sources

- **Fermi LAT collaboration’s solution**:
  —> independent determination of background in a $15^{\circ} \times 15^{\circ}$ region around each dwarf using pre-defined models for Galactic diffuse emission, isotropic background, re-fit of known localised gamma-ray sources, etc.

- **Is that sufficient, where 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 uncertainty due to background modelling errors is hard or unclear

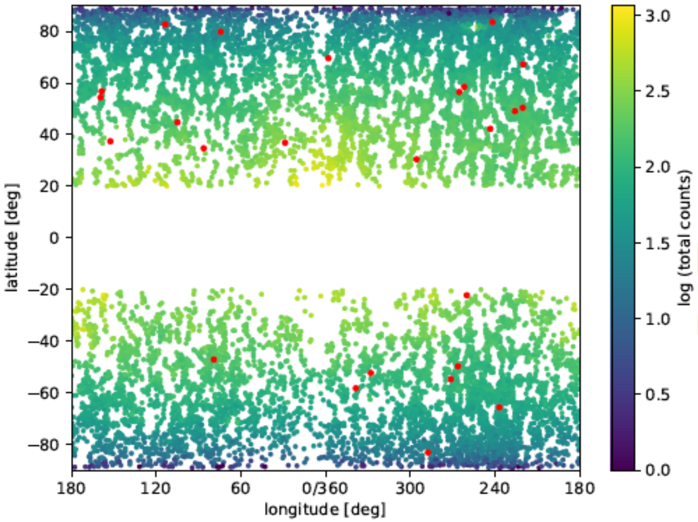
Funded by the European Union’s Horizon 2020 - Grant N° 824064

How does “MLFermiDwarfs” work?

❖ **We propose an alternative approach:**
—> data-driven and agnostic about the physics underlying the background

GOAL: create a probability density function describing the gamma-ray background at each position in the sky based on real data samples and optimised with machine learning tools

Step 1: define the average angular size of dwarf (r ~ 0.5°), tile the entire sky in samples non overlapping with known dwarfs, bright point sources (3FGL catalog) and the Galactic plane and each other (isotropic selection)



A bit less than 10.000 sample positions as “training data” for the pdf optimisation (here: 7-year LAT data set)

Funded by the European Union’s Horizon 2020 - Grant N° 824064

How does "MLFermiDwarfs" work?

Step 2: build global PDF estimator $\hat{\mathscr{F}}$ based only on data
(parameterised PDF according to E. Parzen '61, D. Specht '88,'90,'91, well-known theorems in statistics/Machine Learning community proving convergence to "true" PDF)

$$\hat{\mathscr{F}}(\vec{x}, b) = \frac{1}{N} \sum_{i=1}^{N} K_\sigma(\vec{x} - \vec{x}_i)\, g_\zeta(b - b_i)$$

spatial location kernel,
smoothing parameter σ

photon count kernel,
smoothing parameter ζ

—> Smoothing parameters may be local (depending on position) but here chosen to be global for simplicity.
—> We take a **Gaussian for *K*** and a **log-normal for *g*** (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).

Step 3: maximise the (global) likelihood on the training background sample w.r.t. the smoothing parameters

$$\ln \hat{\mathscr{F}}_{\text{tot}}(\vec{x}, b) = \sum \ln \left[\frac{1}{N-1} \sum_{i \neq j} K_\sigma\left(\vec{x}_i - \vec{x}_j\right) g_\zeta\left(b_i - b_j\right) \right] \qquad \{\hat{\sigma}, \hat{\zeta}\} = \operatorname{argmax} \left\{ \ln \hat{\mathscr{F}}_{\text{tot}}(\vec{x}, b) \right\}$$

Funded by the European Union's Horizon 2020 - Grant N° 824064

How does "MLFermiDwarfs" work?

Step 4: Use the optimised pdf to predict the expected background counts (and higher moments) at each dwarf position (CAVEAT: the pdfs will not be Gaussian)

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

Step 5: Constructing (stacked) Poisson likelihood function (terms per dwarf d and energy bin e) and deriving upper limits on the dark matter annihilation cross-section via the log-likelihood ratio test statistic (adopted from standard Fermi-LAT analyses, e.g. arXiv:1310.0828)

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

background and J-factor as nuisance

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

signal parameter

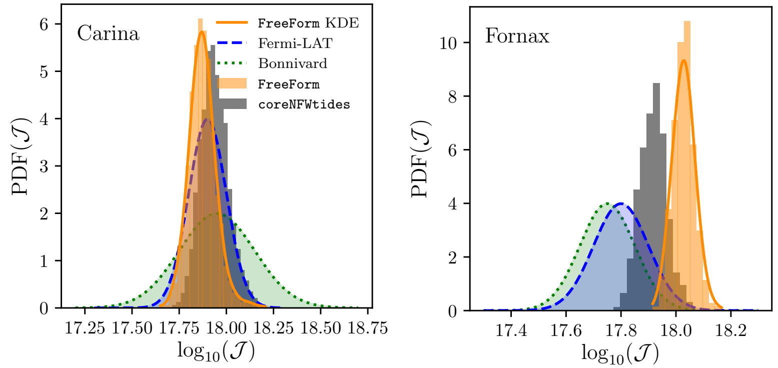
Funded by the European Union's Horizon 2020 - Grant N° 824064

Additional technical details

1. Pdf of dwarf J-factors in two ways (user may decide which to use):

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

Log-normal; spectroscopic measurements (0.5° circular ROI)
values adopted from previous publications: [1611.03184, 1604.05599, 1511.06296]



Largest dwarfs have been reanalysed in the light of their stellar kinematics (Spencer et al. 2017, 2018) and photometry (Flewelling et al. 2016)
—> J-factor pdfs directly from data-analysis, no imposed NFW profile

2. Background pdf parameter optimisation per energy bin?

The optimisation of σ and ζ is only done with respect to the first energy bin (0.67 GeV to 0.89 GeV) —> rather mild dependence + low photon count at energies > 20 GeV

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 …

Funded by the European Union's Horizon 2020 - Grant N° 824064

Code structure and workflow

❖Package is **optimised for use from the command line enabling a quick check of the viability of a user-defined DM model** but functions can be imported and used within scripts.

❖Version 1 is based on the Fermi-LAT data set prepared and utilised in A. Alvarez et al. JCAP09 (2020) 004, i.e. ~10 years (SOURCE event class, Pass8, 500 MeV to 500 GeV, logarithmic energy binning into 24 bins: later summed to form 6 macro bins)
—> **background pdf is already optimised and kernel parameters are fixed**
—> a user only needs to select the dwarf spheroidal galaxies and DM model to derive upper limits on the model parameters

❖**Future versions of the code will allow the user to provide their own LAT data** set and to perform the pdf optimisation w.r.t. to this data set
—> already in the making
—> needs some additional stability checks (not any LAT data will work, some minimal requirements must be met)

Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE

Code structure and workflow

❖Successive call of scripts which perform the likelihood profiling and calculation of upper limits.

Required input:

dwarf_profiling.py

output created by
used for

dwarf_setlimits.py

output created by
used for

dwarf_plot.py

— machine-readable table of dwarf spheroidal galaxies and their properties (position, J-factor + uncertainty, LAT exposure per energy bin)
* comes with code: `default_dwarf_summary_table.dat`
* changes possible (however, LAT data set is fixed!)
* user declares a selection of these dwarfs by their name, joined with '+'

— a list of DM masses to scan: .txt file with one mass value per line (for each entry, the likelihood profiling will be done with the spectrum from the input table)

— $\mathrm{d}N_\gamma/\mathrm{d}E$ for the favourite DM model of the user: table format with three columns; DM mass [GeV] — Energy [GeV] — differential spectrum [GeV^-1]

— a definition of the confidence level of the upper limit (test statistic value) default: 95% (TS = 3.84; for one parameter)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Code structure and workflow

❖Apart from the necessary primary input, minor adjustments may be coordinated by additional **parser arguments.** The main input files are also specified as parser arguments.

Available switches:

dwarf_profiling.py

output created by
used for

dwarf_setlimits.py

output created by
used for

dwarf_plot.py

`-c`: either 'J' or 'JB' —> likelihood profiling is either w.r.t. J-factors or J factors + background

`-j`: Boolean —> if True use custom pdf for classical dwarfs else use the Gaussian pdf with values from the input table

`-svmin/svmax`: min/max annihilation cross-section for the evaluation of the likelihood (needs to capture the expected range)

A few additional switches to point to the path of the required data files; only relevant if their default path must be changed.

`--type`: "all"/"single" —> plots either a single line of the calculated upper limits as a function of mass or includes the limits from each individual dwarf used for stacking (individual limits have to be computed first)

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Software Development

- **Software development on gitlab:**
  https://gitlab.in2p3.fr/francesca.calore/mlfermidwarfs
- **Documentation:**
  README jupyter script explaining the functionality with examples
- **Software license:** MIT License
- **Test and CI/CD:**
  —> examples in the jupyter notebook
  —> default provided DM annihilation spectra include a table with PPPC spectra
- **Operating system, compilation environment:**
  —> Linux, MacOS, Windows (later versions that accept custom LAT data sets may require healpy, which does not work well within Windows)
  —> python 3, setup.py available
- **Dependencies:**
  —> several python packages are required: scikit-learn, scipy, iminuit, astropy
  —> package requirements defined in setup.py
- **Hardware requirements:** none
- **Interface options:** envisaged interface to micrOMEGA

Funded by the European Union's Horizon 2020 - Grant N° 824064

OSSR Integration

- **What is available?**

  source code, python setup script, tutorial jupyter notebook, default data to run the analysis and to reproduce the figures of A. Alvarez et al. JCAP09 (2020) 004

- **What will be on-boarded (source code, container, test workflow incl. data)?**

  all of the above

- **Are there open points?**

  —> Again, future versions will allow the user to perform the full analysis based on a user-defined LAT data set.

  —> Refined treatment of background optimisation (e.g. based on physically motivated models) may become available.

- **What is the user story?**

  Typical story: A theoretical physicists develops a new particle dark matter model which happens to produce — through some interactions — a flux of gamma rays. They may want to check which part of their theories parameter space is excluded by LAT dwarf data. With a table of gamma-ray spectra, they can obtain the desired answer.

Funded by the European Union's Horizon 2020 - Grant N° 824064

Time for a short demo

- The mentioned jupyter notebook serves as an ideal starting point for a demo …

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Discussion Time

- Thank you for your attention!

- Are there any questions?

Funded by the European Union's
Horizon 2020 - Grant N° 824064