PoS(ICRC2017)600

Looking for infrared counterparts of Fermi/LAT blazar candidates

J. Lefaucheur[*a], **C. Boisson**[a], **P. Goldoni**[b], **S. Pita**[b]

[a]*LUTH, Observatoire de Paris, PSL Research University, CNRS, Université Paris Diderot*
5 Place Jules Janssen, 92190 Meudon, France
[b]*APC, AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/Irfu,*
Observatoire de Paris, Sorbonne Paris Cité
10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
E-mail: julien.lefaucheur@obspm.fr

The Fermi/LAT telescope is an efficient blazar-detector in the MeV/GeV range. More than 1100 (900) blazars detected above 100 MeV (10 GeV) are clearly associated to BL Lacertae or Flat Spectrum Radio Quasar objects in the Fermi/LAT 3FGL catalogue. This number could significantly increase if multi-wavelength counterparts could be identified for the 573 3FGL blazars with unknown type, or even for the 1010 3FGL unassociated sources which are thought to be dominated by blazars, at least at high galactic latitude. Unfortunately, the size of the Fermi/LAT error box makes multi-wavelength follow-ups difficult.

We propose a method to associate "blazar-like" infrared counterparts, having coordinates with a precision of a few arcseconds, to Fermi/LAT blazars and unassociated sources. To reach this goal, we built machine-learning classifiers based on the statistical differences of magnitude measurements obtained by the WISE satellite, between a sample of well-identified infrared blazars and samples of other types of infrared sources located in regions of the sky where no known blazar is present. We provide a list of potential infrared counterparts for 3FGL blazar candidates, along with the associated number of expected false positives. This study contributes to increase the number of well-identified extragalactic blazars and also provides promising blazar targets for the Cherenkov Telescope Array.

35th International Cosmic Ray Conference — ICRC2017
10–20 July, 2017
Bexco, Busan, Korea

---

[*]Speaker.

1. Introduction

Blazars dominate the extragalactic sky above 100 MeV. They are radio-loud active galactic nuclei (AGN) whose jet is quasi-aligned to the line of sight. An important Doppler effect blue-shifts their spectra and increases their observed luminosity. Their spectral energy distribution is characterised by a first bump peaking between the infrared and the X-ray domain which is associated to synchrotron emission of relativistic electrons. The high energy bump, in the MeV/TeV energy range, is usually associated to inverse-Compton radiation in a leptonic scenario but might also be explained with hadronic scenarios. In the very high energy range ($E \geq 100\,$GeV), the current understanding of their population and the improvement of the diffuse extragalactic background light measurement are nowadays limited by the small number of detected blazars.

Since 2008, the Fermi/LAT telescope maps the sky above 100 MeV with unmatched sensitivity and angular resolution in this frequency domain. The LAT collaboration reported the detection of 3034 γ-ray sources [1] among which 1717 are blazars, including 660 BL Lacertae (BL Lacs), 484 flat spectrum radio quasars (FSRQs) and 573 blazars of undetermined type (BCUs). In addition, 1010 sources are still of unknown nature because of the lack of firmly identified counterparts at other wavelengths. The identification of these sources is not an easy task considering the multiple candidate associations due to the large error localisation of the Fermi/LAT and the incompleteness of counterpart catalogues. It is expected that a significant fraction of the unassociated sources are blazars since they are the dominant class of sources detected by the LAT. Several studies [2, 3] looked for blazar candidates among the unassociated sources with the help of machine-learning classification methods and separation power between different classes of sources extracted from Fermi/LAT catalogues.

The determination of possible counterparts to the unassociated sources or the BCUs will help to reveal their nature and simplify their identification at other wavelengths. Massaro et al. [4] used the assumption that blazars occupy a special position in the colour-colour diagram constructed with the magnitudes measured with the WISE satellite [5]. By building "blazar regions" with a selected sample of infrared blazars, and by comparing the distance of the unassociated sources in the colour space to these regions, one can identify potential candidates for blazar-like counterparts [6]. However, this method is based only on a selected sample of known blazars and does not consider the behaviour of other infrared source classes. Therefore, it makes it hard to estimate the number of false positives.

In this contribution, we propose a method to associate infrared counterparts to high galactic latitude ($b \geq 10°$) γ-ray blazar candidates, along with the number of expected false positives for each association. We used three colors obtained from the four magnitudes extracted from the WISE catalogue and a newly defined parameter miming the blazar efficiency to produce infrared photons by synchrotron emission to discriminate between "blazar-like" and non-blazar infrared counterparts. Afterwards, we built a classifier with a sample of well-identified infrared counterparts of blazars against a sample of infrared sources selected in regions of the sky where no γ-ray blazar is present. In order to estimate the number of false associations, we defined different classes of associations according to the number of expected false positives.

In Section 2 we introduce the different data samples and the discriminant parameters. Section 3 presents the classifier construction and the procedure to determine the infrared associations. Results
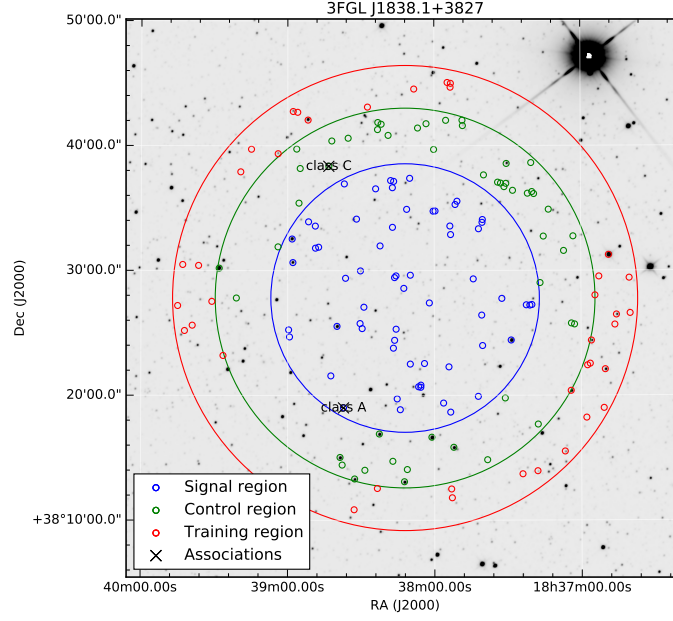
Figure 1: Illustration of the different regions used in this study to select the infrared counterparts of a Fermi/LAT unassociated source. The image is a mosaic of individual frames from the WISE W_1 filter around the source.

are shown in Section 4 and we conclude with a short discussion in Section 5.

2. Data samples and discriminant parameters

We used different samples of infrared sources, provided by the AllWISE Source Catalogue[1], to build a binary classifier to search for counterparts of high galactic latitude ($b > 10\,\mathrm{deg}$) blazar candidates. We corrected the magnitudes of all sources for infrared Galactic extinction in the two shorter wavelength filters W_1 (3.4 μm) and W_2 (4.6 μm) using the Schlegel et al. [7] data and the Indebetouw et al. [8] extinction law. In order to use reliable magnitude measurements, we only kept infrared sources with:

- a signal/noise ratio greater or equal to 2 in all filters

- a contamination and confusion flag only equal to "0", "o" or "h" in all filters[2]

- an "extended" flag less or equal to 1 corresponding to point-like source

To create a sample of well-identified infrared blazars, we selected all the sources from the 3LAC catalogue [9] labelled as BL Lac or FSRQ belonging to the 3LAC clean sample (1018). We further selected their infrared counterparts (754) in a circular region of radius $1''$ centered on the source position given by the 3LAC catalogue. For the sample of non-blazar infrared sources, composed mainly of stars, normal galaxies and QSOs, we selected and stacked all the sources
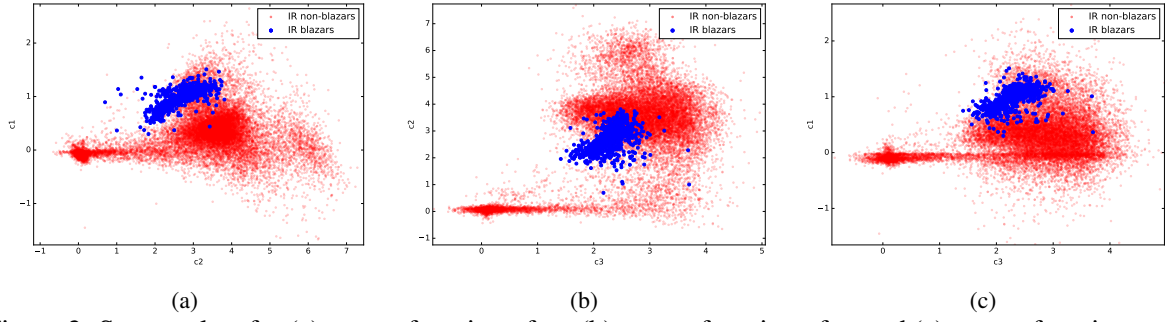

Figure 2: Scatter plots for (a) c_1 as a function of c_2, (b) c_2 as a function of c_3 and (c) c_1 as a function of c_3, for the infrared blazars (red) and the infrared non-blazars (blue).
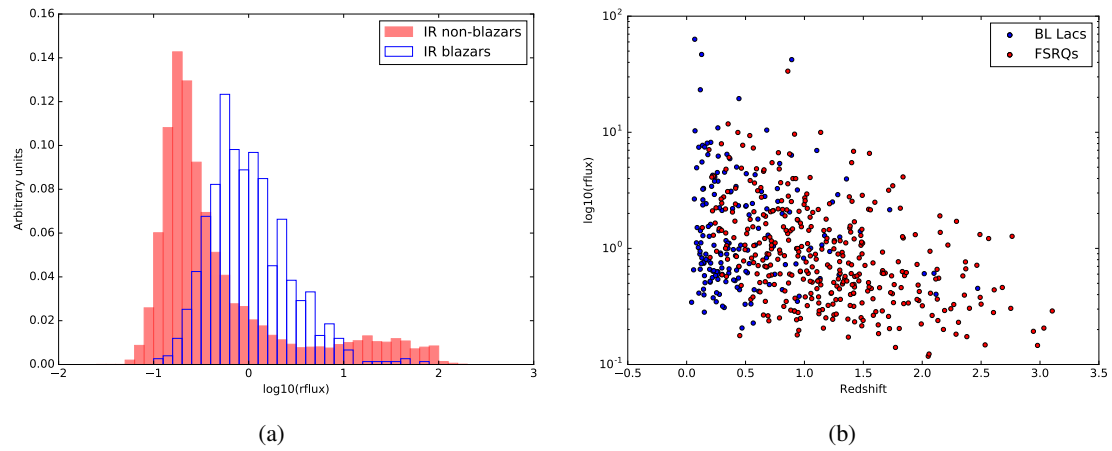

Figure 3: (a) Distribution of the *rflux* parameter for the infrared blazars (red) and the infrared non-blazars (blue). (b) *rflux* parameter as a function of the redshift for infrared blazars having a measured of redshift. BL Lacs and FSRQs are respectively shown in blue and in red.

in every annular regions surrounding the 531 high-latitude unassociated sources (see Figure 1), except the so-called c-sources[3], of the 3FGL catalogue. The inner and outer radius of the annular regions were respectively set to $2\sqrt{\theta_{95}}$ and $3\sqrt{\theta_{95}}$, where θ_{95} is the 95 % confidence level on the localisation of a Fermi/LAT source. The stack, composed of $\sim$18900 sources, will be used as an estimate for the infrared background sources. The target samples were defined as the high-latitude unassociated sources and the high-latitude BCUs, without the c-sources. For each of the sources in these two samples, we selected all the infrared counterparts in the circular region centered on the LAT position , called the "source region", of radius θ_{95}. Furthermore, we selected the sources in an annular region surrounding each of the sources in the two target samples, called the "control regions" and of the same area as the source region, to search for potential blazar-like counterparts which are further away than the Fermi/LAT error boxes. The inner and outer radius of the control regions were respectively set to $\sqrt{\theta_{95}}$ and $2\sqrt{\theta_{95}}$.

To discriminate between the different classes of sources we used the three colors $c_1 = W_1 - W_2$,

---

[3]The c-sources are considered to be potentially confused with galactic diffuse emission.

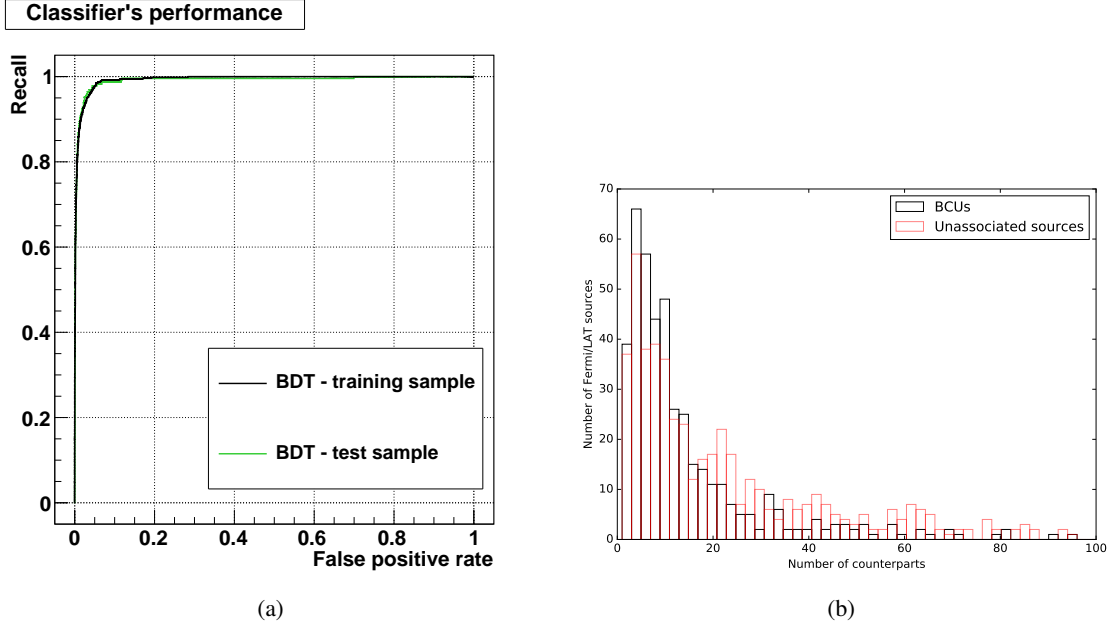(a)   (b)

Figure 4: (a) The ROC curves of the boosted decision tree classifier estimated on the training sample using a ten-fold cross-validation method (black) and estimated on the test sample (green). (b) The number of infrared counterparts in the "source" region for the Fermi/LAT BCUs (black) and unassociated sources (red). The x-axis have been truncated above a number of 100 counterparts for readability.

$c_2 = W_2 - W_3$ and $c_3 = W_3 - W_4$ as defined in [6]. A sample of scatter plots is shown on Figure 2. In addition, we use a fourth parameter, called $rflux$, which quantify, as a first order approximation, the blazars' exceptionally efficient production mechanism of infrared photons compared to other classes of source. This parameter is defined as the integrated flux estimated according to [10] divided by a value of reference corresponding to the average blazar integrated flux. To study the Compton Dominance[4] (CD), the ratio of luminosities between the inverse-Compton and the synchrotron bumps, of a sample of blazars detected by WISE and the Fermi/LAT, D'Abrusco et al. [10] estimated the integrated flux in the WISE energy range by summing the fluxes in the four WISE filters. Figure 3a shows the distribution of $rflux$. The separation power between blazar and non-blazar infrared sources is manifest but it is always tricky to use a flux as a discriminant parameter as it is distance-dependent. However, Figure 3b shows the $rflux$ parameter as a function of the redshift of 533 infrared blazars having an estimation of redshift and we notice a small decrease of the flux which stays relatively small compared to the bulk of the distribution.

3. Classifier construction and association procedure

We considered several machine-learning algorithms to identify potential infrared blazar-like counterparts of the Fermi/LAT sources. We chose the boosted decision tree (BDT) algorithm for its capacity to obtain good performance "out of the box", its robustness against over-training and

---

[4]Here we can not use CD as a discriminant parameter since we do not have an estimate of the γ-ray flux for the non-blazar infrared counterparts.

also for the smaller time needed to build a model compared to other methods such as the neural-networks or the support vector machine based-methods. We selected a single split of infrared blazars and non-blazars, respectively 70 % and 30 % for the training and the test samples. The training/test split was obtained according to the proximity of its performance compared to the average behaviour of the BDT method estimated on a large (100) number of training/test phases with random splits (see [2]). To build a model from the training sample we used a ten-fold cross validation method. The receiver operating characteristic (ROC) curves of the classifier estimated on the training and on the test samples are shown on Figure 4a. We determined the cutoff, called $\zeta^\star$, on the score distribution, by requiring a true positive rate of 90 %. The performance metrics, the recall TP and the false positive rate FP, estimated on the test sample are respectively equal to 91.19 % and 1.66 %. In the following, all infrared counterparts which score (called ζ) is greater or equal to $\zeta^\star$ will be considered as potential blazar-like counterparts.

In order to control the number of expected false positives for each association, we defined classes of sources according to the expected contamination. Each of the Fermi/LAT sources in the target samples has its own number of infrared counterparts N_c, see Figure 4b, which we further considered as dominated by non-blazar sources. For each of the potential blazar-like counterparts one can define an estimation of the expected number of false positives $\eta = N_c \times \text{FP}$, where the FP value is estimated with the score ζ of the source. From this, we defined four classes of sources A, B, C and D which distribute the infrared counterparts according to the expected number of false positives η:

- Class A, $\eta \leq 5/100$

- Class B, $5/100 < \eta \leq 10/100$

- Class C, $10/100 < \eta \leq 25/100$

- Class D, $25/100 < \eta \leq 50/100$

In the following, the infrared candidates with an expected number of source contamination greater than 0.5 will not be considered for further analysis.

4. Application on the target samples

The application of the procedure to the 444 BCUs of the 3FGL catalogue gives a total of 315 infrared blazar-like counterparts for 265 Fermi/LAT sources ($\sim$1.2 counterparts per γ-ray source). The infrared sources are distributed among 197 Class A, 54 Class B, 39 Class C and 25 Class D corresponding to a number of expected false positives less than 0.05, 0.10, 0.25 and 0.5, respectively. The same procedure applied to the 531 unassociated sources of the 3FGL catalogue gives a total of 188 infrared blazar-like counterparts for 155 Fermi/LAT sources ($\sim$1.2 counterparts per γ-ray source), distributed among 50 Class A, 35 Class B, 54 Class C and 49 Class D.

5. Discussion and conclusions

With our approach, building a classifier with a well-identified sample of infrared blazars and a sample of infrared non-blazar sources, we propose blazar counterparts for the Fermi/LAT sources

and we can estimate for each association an expected number of false positives. Considering only the most promising associations, the source belonging to class A or class B and correspond to a number of false positives less than 0.05 and 0.01, respectively, we find 251 potential blazar-like counterparts for 235 BCUs. The sum of the expected numbers of false positive is less than 7. Furthermore, only 11 associations are found for these sources in the control regions for which an equal or a better association exists (10 for class A and one for class B). For the Fermi/LAT unassociated sources, we find 85 blazar-like counterparts of class A or B for 82 γ-ray sources with an expected number of false positives less than 4. In the corresponding control region, 9 counterparts have an equal or a better association than the blazar-like counterparts in the signal region.

To assess the true nature of the selected infrared sources, as candidates for an infrared counterpart of γ-ray blazar, a multi-wavelength study is necessary. In addition to provide astrometric coordinates to simplify the follow-up at other wavelengths, it can help to prioritise the search for blazars. For example, by crossing the list of blazar candidates for the unassociated sources proposed in [2] and [3] we found out that 130 infrared sources proposed in this work have a match, including 60 infrared sources of class A or B.

In this work we focused on the high latitude Fermi/LAT sources from the 3FGL catalogue. Dedicated classifiers should be used to tackle γ-ray sources of lower galactic latitude, as the infrared source population differs from the high latitude ones. Finally, the procedure could be applied to the Fermi/LAT 3FHL catalogue [11] in order to help the science preparation with the upcoming of the Cherenkov Telescope Array.

Acknowledgments

References

[1] Acero et al., *Fermi Large Area Telescope Third Source Catalog*, ApJS **218** (June, 2015) 23, [1501.02003].

[2] J. Lefaucheur and S. Pita, *Research and characterisation of blazar candidates among the Fermi/LAT 3FGL catalogue using multivariate classifications*, A&A **602** (June, 2017) A86, [1703.01822].

[3] P. M. Saz Parkinson, H. Xu, P. L. H. Yu, D. Salvetti, M. Marelli and A. D. Falcone, *Classification and Ranking of Fermi LAT Gamma-ray Sources from the 3FGL Catalog using Machine Learning Techniques*, ApJ **820** (Mar., 2016) 8, [1602.00385].

---

[5]http://tmva.sourceforge.net/

[6]http://www.starlink.ac.uk/stilts/

[7]http://www.astropy.org/

[8]https://pyvo.readthedocs.io/en/latest/

[4] F. Massaro, R. D'Abrusco, M. Ajello, J. E. Grindlay and H. A. Smith, *Identification of the Infrared Non-thermal Emission in Blazars*, ApJ **740** (Oct., 2011) L48, [1203.0304].

[5] E. L. Wright et al., *The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance*, AJ **140** (Dec., 2010) 1868–1881, [1008.0031].

[6] F. Massaro, R. D'Abrusco, A. Paggi, N. Masetti, M. Giroletti, G. Tosti et al., *Unveiling the Nature of Unidentified Gamma-Ray Sources. II. Radio, Infrared, and Optical Counterparts of the Gamma-Ray Blazar Candidates*, ApJS **206** (June, 2013) 13, [1303.3585].

[7] D. J. Schlegel, D. P. Finkbeiner and M. Davis, *Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds*, ApJ **500** (June, 1998) 525, [astro-ph/9710327].

[8] R. Indebetouw et al., *The Wavelength Dependence of Interstellar Extinction from 1.25 to 8.0 µm Using GLIMPSE Data*, ApJ **619** (Feb., 2005) 931–938, [astro-ph/0406403].

[9] M. Ackermann et al., *The Third Catalog of Active Galactic Nuclei Detected by the Fermi Large Area Telescope*, ApJ **810** (Sept., 2015) 14, [1501.06054].

[10] R. D'Abrusco, F. Massaro, M. Ajello, J. E. Grindlay, H. A. Smith and G. Tosti, *Infrared Colors of the Gamma-Ray-detected Blazars*, ApJ **748** (Mar., 2012) 68, [1203.0568].

[11] The Fermi-LAT Collaboration, *3FHL: The Third Catalog of Hard Fermi-LAT Sources*, *ArXiv e-prints* (Feb., 2017) , [1702.00664].

[12] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, *ArXiv Physics e-prints* (Mar., 2007) , [physics/0703039].

[13] M. B. Taylor, *STILTS - A Package for Command-Line Processing of Tabular Data*, in *Astronomical Data Analysis Software and Systems XV* (C. Gabriel, C. Arviset, D. Ponz and S. Enrique, eds.), vol. 351 of *Astronomical Society of the Pacific Conference Series*, p. 666, July, 2006.

[14] Astropy Collaboration, *Astropy: A community Python package for astronomy*, A&A **558** (Oct., 2013) A33, [1307.6212].