



On the Physical Association of Fermi-LAT Blazars with Their Low-energy Counterparts

Raniere de Menezes^{1,2} , Raffaele D’Abrusco³ , Francesco Massaro^{2,4,5,6} , Dario Gasparrini^{7,8} , and Rodrigo Nemmen¹

¹ Universidade de São Paulo, Departamento de Astronomia, Rua do Matão, 1226, São Paulo, SP 05508-090, Brazil; raniere.m.menezes@gmail.com

² Dipartimento di Fisica, Università degli Studi di Torino, via Pietro Giuria 1, I-10125 Torino, Italy

³ Center for Astrophysics—Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 20138, USA

⁴ Istituto Nazionale di Fisica Nucleare, Sezione di Torino, I-10125 Torino, Italy

⁵ INAF-Osservatorio Astrofisico di Torino, via Osservatorio 20, I-10025 Pino Torinese, Italy

⁶ Consorzio Interuniversitario per la Fisica Spaziale (CIFS), via Pietro Giuria 1, I-10125, Torino, Italy

⁷ Istituto Nazionale di Fisica Nucleare, Sezione di Roma “Tor Vergata,” I-00133 Roma, Italy

⁸ Space Science Data Center—Agenzia Spaziale Italiana, Via del Politecnico, snc, I-00133, Roma, Italy

Received 2020 February 21; revised 2020 April 20; accepted 2020 April 21; published 2020 May 28

Abstract

Associating γ -ray sources to their low-energy counterparts is one of the major challenges of modern γ -ray astronomy. In the context of the Fourth Fermi Large Area Telescope Source Catalog (4FGL), the associations rely mainly on parameters such as apparent magnitude, integrated flux, and angular separation between the γ -ray source and its low-energy candidate counterpart. In this work, we propose a new use of the likelihood ratio (LR) and a complementary supervised learning technique to associate γ -ray blazars in 4FGL, based only on spectral parameters such as the γ -ray photon index, mid-infrared colors, and radio-loudness. In the LR approach, we crossmatch the Wide-field Infrared Survey Explorer Blazar-Like Radio-Loud Sources catalog with 4FGL and compare the resulting candidate counterparts with the sources listed in the γ -ray blazar locus to compute an association probability (AP) for 1138 counterparts. In the supervised learning approach, we train a random forest algorithm with 869 high-confidence blazar associations and 711 fake associations and then compute an AP for 1311 candidate counterparts. A list with all 4FGL blazar candidates of uncertain type associated by our method is provided to guide future optical spectroscopic follow-up observations.

Unified Astronomy Thesaurus concepts: [Gamma-ray astronomy \(628\)](#); [Active galactic nuclei \(16\)](#); [Astronomical methods \(1043\)](#)

Supporting material: machine-readable table

1. Introduction

The association of γ -ray sources with their low-energy counterparts is a long-standing challenge since the beginning of modern γ -ray astronomy (Fichtel et al. 1994). Its major underlying difficulty is related to the large positional uncertainty of γ -ray observations. Even in the era of the Fermi Large Area Telescope (LAT; Atwood et al. 2009), the positional uncertainty of γ -ray sources ranges from a couple of arcminutes up to $\sim 1^\circ$ (Abdollahi et al. 2020). Then, according to the recent release of the Fermi-LAT Fourth Source Catalog (4FGL; Abdollahi et al. 2020), $\sim 25\%$ of its sources lack an assigned low-energy counterpart and thus have an uncertain nature. Although the fraction of unassociated γ -ray sources (UGSs) is still large, 4FGL presents a modest improvement in comparison with its previous releases, which had $\sim 30\%$ of unassociated sources (Abdo et al. 2010; Nolan et al. 2012; Acero et al. 2015). The largest population of associated γ -ray sources is dominated by pulsars, pulsar wind nebulae, and supernova remnant in the Galactic plane and by blazars in the extragalactic sky (Abdollahi et al. 2020).

Several studies searching for the counterparts of γ -ray sources have been performed in the past decade (Abdo et al. 2010; Nolan et al. 2012; Acero et al. 2015). Dedicated follow-up observations at radio, infrared, and X-rays allowed for the identification of potential counterparts for several Fermi-LAT sources, for which optical spectra were then collected to establish their nature (Paggi et al. 2014; Landoni et al. 2015; Massaro et al. 2015b, 2016; Álvarez Crespo et al. 2016a; Peña-Herazo et al. 2017, 2019; Marchesini et al. 2019;

de Menezes et al. 2020). Once a candidate counterpart is found, it can be considered in the association methods of the Fermi-LAT catalogs. All association methods used by Abdollahi et al. (2020) consist of computing the association probability (AP) based mainly on the angular separation between the center of the γ -ray source and the position of its candidate counterpart, thus being mostly a geometrical approach and neglecting physical properties of candidate counterparts, such as colors, spectral shapes, and radio-loudness to name a few. When a candidate counterpart has $AP > 80\%$ in 4FGL, the source is considered associated.

In this work, we propose two different association procedures, both independent of angular separation and relying mainly on the γ -ray–mid-infrared (MIR) connection of candidate blazar counterparts selected from the Wide-field Infrared Survey Explorer (WISE) Blazar-Like Radio-Loud Sources catalog (WIBRaLS; D’Abrusco et al. 2019). WIBRaLS is a catalog of radio-loud candidate blazars whose WISE MIR colors are selected to be consistent with the MIR colors of confirmed γ -ray emitting blazars (Massaro et al. 2011; D’Abrusco et al. 2012; Massaro & D’Abrusco 2016). We compute the APs with two different methods: (i) the likelihood ratio (LR; Sutherland & Saunders 1992), already adopted by the Fermi-LAT Collaboration, but with a different setup, and (ii) a random forest algorithm (RF; Breiman 2001). By assigning an AP to each candidate counterpart, we can schedule/program future optical spectroscopic follow-ups on these results, prioritizing those targets with a higher AP.

Computing an AP for each counterpart candidate is crucial because all methods used to select candidate blazars are

statistical in nature and do not take into account the specific γ -ray properties of each Fermi source to be associated but do take into account only collective features of a population of sources (in our case, the γ -ray emitting confirmed blazars in the blazar locus; see Section 2). Thus, association methods bridge the gap between collective behavior and each specific case, giving an intra-source prioritization—should there be multiple candidates for the same γ -ray source—and an inter-source prioritization to maximize the effectiveness of spectroscopic follow-ups.

Machine-learning techniques have been used before in the context of Fermi-LAT catalogs to (i) predict the spectral class of UGSs (Doert & Errando 2014; Parkinson et al. 2016; Lefaucheur & Pita 2017; Salvetti et al. 2017) based only on the γ -ray properties available in the Second and Third Fermi Source Catalogs (2FGL and 3FGL, respectively; Nolan et al. 2012; Acero et al. 2015), (ii) predict the nature of blazar candidates of an uncertain type (BCUs; Hassan et al. 2012; Chiaro et al. 2016; Kovačević et al. 2019), and (iii) even spot candidates of dark matter Galactic subhalos (Mirabal et al. 2012). This is the first time, however, that machine-learning is used to associate Fermi-LAT sources with their low-energy counterparts.

The idea underlying this work is that γ -ray blazars have some specific radio–MIR characteristics that can tell them between, e.g., two counterpart candidates lying in the same elliptical uncertainty region of a γ -ray source. Indeed, some sources associated in 4FGL have a very high AP with counterparts that are not the closest ones if compared with sources listed in the latest version of WIBRaLS (D’Abrusco et al. 2019). This can happen simply because the latest version of WIBRaLS was not taken into account when associating 4FGL sources, but it also highlights a possible bias in a method that associate sources based mainly on angular separation: if the real γ -ray source counterpart is not listed in one of the catalogs used by the association algorithms, then the method will simply choose the closest candidate.

The paper is organized as follows. In Sections 2 and 3, we describe the samples used to carry out our analysis, providing basic details on both WIBRaLS and 4FGL. In Sections 4 and 5, we describe the adopted methods followed by the achieved results in Sections 6 and 7. We conclude and summarize our work in Section 8. The WISE magnitudes are in the Vega system and are not corrected for the Galactic extinction, since, as shown in D’Abrusco et al. (2014), such correction only affects the magnitude at $3.4\ \mu\text{m}$ for sources lying close to the Galactic plane and it ranges between 2% and 5% of the magnitude, thus it is not significantly affecting the results. WISE bands are indicated as W1, W1, W3, and W4 and correspond, respectively, to the nominal wavelengths at 3.4, 4.6, 12, and $22\ \mu\text{m}$, while the colors are defined as $c_{12} = W1 - W2$, $c_{23} = W2 - W3$, and $c_{34} = W3 - W4$. The adopted MIR radio-loudness parameter is defined as $q_{22} = \log(S_{22\ \mu\text{m}}/S_{\text{radio}})$, where $S_{22\ \mu\text{m}}$ is the flux density in the WISE W4 band, and S_{radio} is the radio flux density at 1.4 GHz or at 843 MHz, depending on the radio survey in which the WIBRaLS radio counterpart is identified (see D’Abrusco et al. 2019, for more details).

2. Samples Used for the LR Method

We apply the LR method to all 1311 sources listed in 4FGL with at least one counterpart in WIBRaLS, comparing their

MIR colors and radio-loudness with those of the γ -ray blazars lying within the blazar locus (D’Abrusco et al. 2013).

The blazar locus is defined by a sample of confirmed Fermi-LAT blazars listed in Roma-BZCAT (Massaro et al. 2015a) and associated with WISE counterparts detected in all four filters: W1, W2, W3, and W4. The locus is modeled in a three-dimensional space generated by the principal components of the MIR color–color–color distribution and describes the typical MIR colors of γ -ray blazars, thus being ideal for our purposes (see, e.g., D’Abrusco et al. 2013, 2019).

3. Samples Used for the Supervised Learning Method

The training set of supervised learning algorithms needs to be representative of the expected distribution of outcomes in the test set or the sample on which the trained algorithm will be applied. In this case, we train an RF algorithm by gathering high-confidence associations and not associated (fake) counterparts. We do this by selecting all associated sources in 4FGL that have a counterpart in WIBRaLS and with γ -ray statistical significance above 10σ . Such high-confidence γ -ray detections tend to have smaller error ellipses, and for this reason, they are easier to correctly associate with low-energy counterparts. The final number of high-confidence associations in the training set was 869. The fake associations were selected as the WIBRaLS sources not associated in 4FGL but lying within 0.5° from 4FGL sources that are already associated with a low-energy counterpart, resulting in a total of 711 fake counterparts for the training sample. It is possible that a few of these fake associations are the real counterparts of the γ -ray sources.

The panels in Figure 1 show the differences between high-confidence associations (contours) and fake associations (orange dots) in the training sample. Based on these γ -ray–MIR characteristics, supervised learning algorithms can be trained to tell how likely a WIBRaLS source is associated as the counterpart of a γ -ray source. Once trained, we applied the RF method to a test sample made up of the 1311 4FGL–WIBRaLS crossmatches, which is the same sample used in the LR method.

4. The LR Method

The LR technique was first used in the context of source association by Richter (1975) and has been introduced and developed to look for possible counterparts among faint radio, infrared, and X-ray sources (Wolstencroft et al. 1986; Sutherland & Saunders 1992; Masci et al. 2001). In the context of Fermi-LAT, the LR between a candidate counterpart, i , within the error ellipse of a 4FGL source, j , is computed as (Ackermann et al. 2011; Ajello et al. 2019)

$$\text{LR}_{ij} = \frac{e^{-r_{ij}^2/2}}{NA}, \quad (1)$$

where $r_{ij} = \theta/(\sigma_i^2 + \sigma_j^2)^{1/2}$ is the normalized angular separation between the γ -ray source j and candidate counterpart i , with θ being the angular separation between the counterpart and the center of the γ -ray source, σ_i being the low-energy positional uncertainty, and $\sigma_j \equiv \sigma_{95\%}/2.4477$ being the normalized geometric mean of the semimajor and semiminor axes of the 95% 4FGL confidence error ellipse. N is the surface density of objects brighter than the candidate i and A is the solid angle encompassed by the 95% confidence LAT error

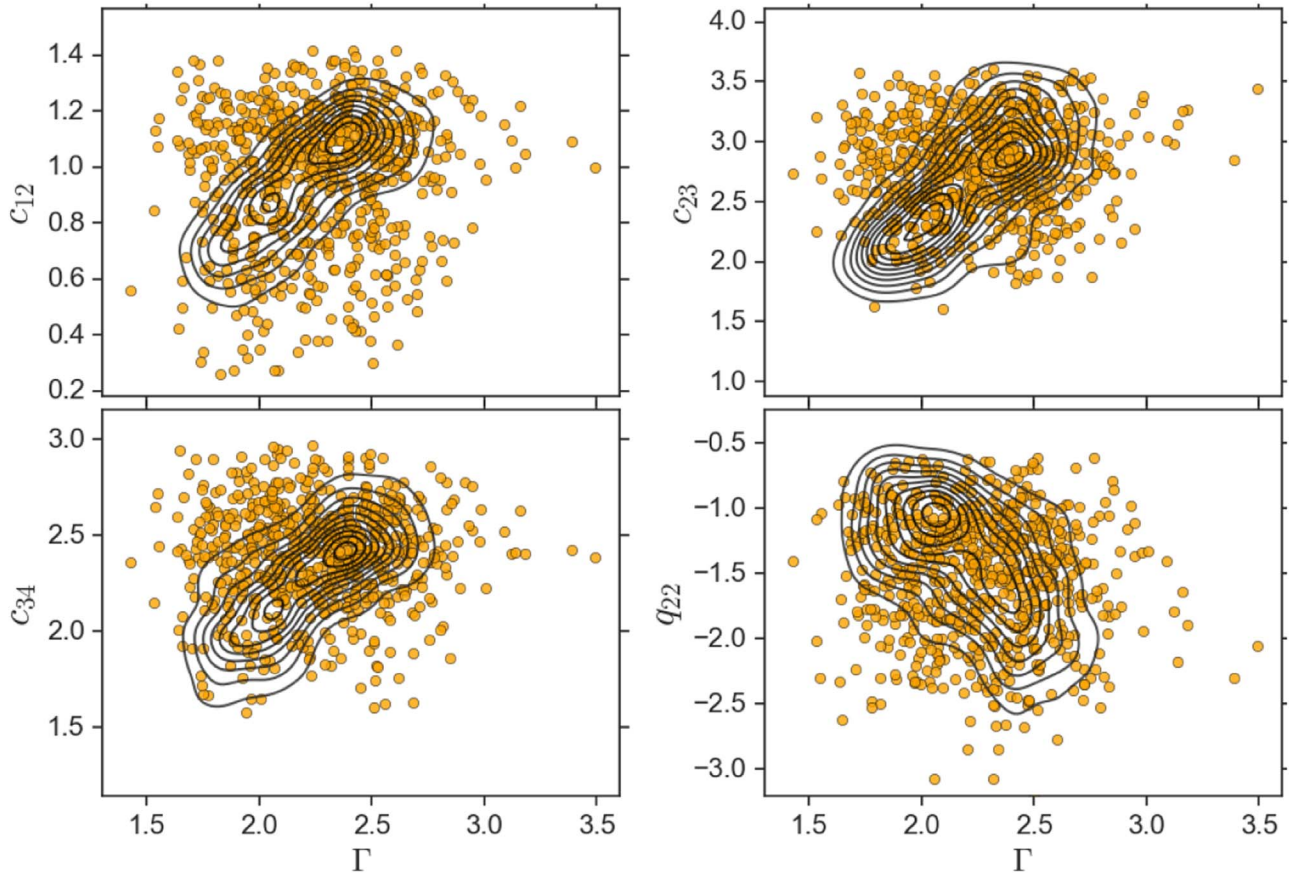


Figure 1. Training sample for the RF algorithm. The contours represent the distribution of the 869 high-confidence associations, while the orange dots represent the fake associations. The spectral parameters for the training sample are shown in terms of the γ -ray photon index, Γ ; the three MIR colors, c_{12} , c_{23} , and c_{34} (as defined in the last paragraph of Section 1); and the MIR radio-loudness, q_{22} .

ellipse. Below, we modify this method by taking into account only spectral properties of the candidate counterparts, as MIR colors and radio-loudness, and comparing these properties with what is expected for γ -ray blazars lying in the blazar locus (D’Abrusco et al. 2013). In Section 6.1, we add a dependence on the angular separation to this method to measure the impact of angular separation in our results.

The LR method we adopt to estimate the AP is a modification of the LR method described in Sutherland & Saunders (1992) and Ackermann et al. (2011). The adopted steps are as follows.

1. We crossmatch WIBRaLS with 4FGL to create the list of all γ -ray blazar candidates that lie within the positional uncertainty region at 95% level of confidence of each Fermi-LAT object. With a total of 1311 potential counterparts, the result of this crossmatch corresponds to all of the 4FGL candidate counterparts listed in WIBRaLS.
2. For each 4FGL source, we consider the γ -ray photon index interval $\Delta\Gamma = [\Gamma - \sigma_\Gamma, \Gamma + \sigma_\Gamma]$, where σ_Γ is the uncertainty on the photon index as listed in 4FGL. Then, given the correlations between Γ , the MIR colors c_{12} and c_{34} , and the MIR radio-loudness q_{22} for the blazars in the γ -ray blazar locus (D’Abrusco et al. 2012), we select from WIBRaLS only those sources with parameters within the blazar locus ranges of $\Delta c_{12} = [c_{12}^{\min}, c_{12}^{\max}]$, $\Delta c_{34} = [c_{34}^{\min}, c_{34}^{\max}]$, and $\Delta q_{22} = [q_{22}^{\min}, q_{22}^{\max}]$ corresponding to the $\Delta\Gamma$ interval, as

shown in Figure 2. This step reduces the total number of candidate counterparts from 1311 to 1138.

3. For each MIR counterpart, i , of a 4FGL source, j , we compute the LR using the following equation:

$$\log \text{LR}_{i,j} = \log \left[\frac{1}{2\pi\sigma_1\sigma_2} \frac{Q(\Delta c_{12}, \Delta c_{34}, \Delta q_{22})}{N(\Delta c_{12}, \Delta c_{34}, \Delta q_{22})} \right],$$

where $Q(\Delta c_{12}, \Delta c_{34}, \Delta q_{22})$ is the probability of finding a WIBRaLS source with MIR colors and radio-loudness within the ranges of Δc_{12} , Δc_{34} and Δq_{22} over the entire WIBRaLS catalog (i.e., it is the ratio between the total number of sources lying within the specified intervals and the total number of WIBRaLS sources), and $N(\Delta c_{12}, \Delta c_{34}, \Delta q_{22})$ is the local surface density of WIBRaLS sources within a circle of 8° radius centered in the 4FGL source and that has MIR colors and radio-loudness within the specified ranges (i.e., it is the surface density of the background objects at the appropriate Galactic latitude). The positional uncertainty errors of the 95% confidence regions of 4FGL sources are given by σ_1 and σ_2 . Angular separation between the γ -ray source and its counterpart is not taken into account.

4. To compute the AP based on the LR calculated above, we first generate 500 fake γ -ray catalogs by shifting the sky positions of each 4FGL source by a random value between 0.3° and 3° in a random direction of the sky. These small shifts in R.A. and decl. guarantee that we preserve the inhomogeneity of the γ -ray sky, which has

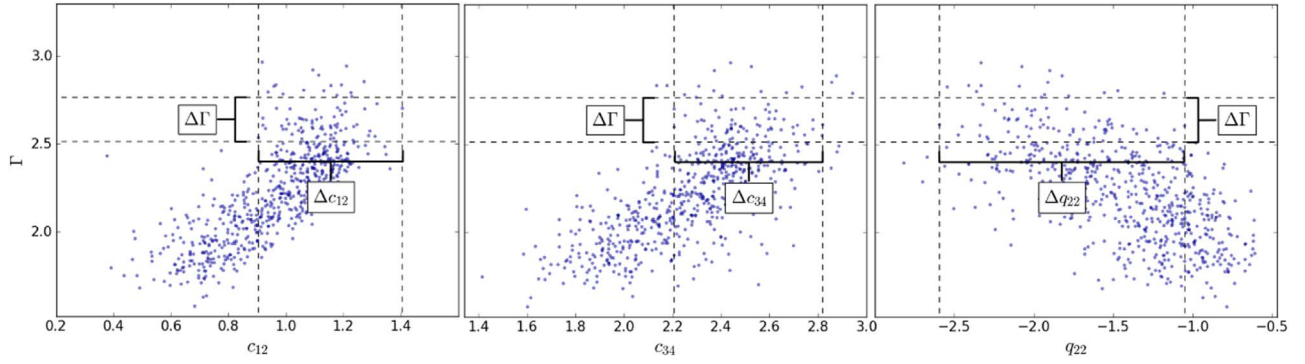


Figure 2. Region selected in the blazar locus for a source with the γ -ray photon index of $\Gamma = 2.65 \pm 0.10$. Given an interval $\Delta\Gamma$, only the sources within the ranges of Δc_{12} , Δc_{34} , and Δq_{22} can be selected as possible counterparts of γ -ray sources.

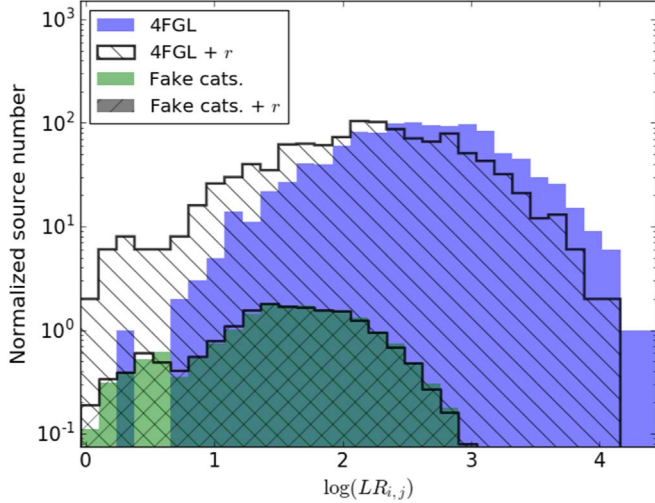


Figure 3. Distribution of the $\log(LR_{i,j})$ for the crossmatches between WIBRaLS and 4FGL (G_{real}) is shown in blue; the average distribution for the crossmatches between WIBRaLS and the fake γ -ray catalogs (G_{fake}) is shown in green. The hatched histograms represent the same distributions when taking angular separation, r , into account with the exponential behavior shown in Equation (1).

more sources concentrated toward the Galactic plane. We then repeat steps 2 and 3 above for all matches between the fake sources (belonging to the 500 generated catalogs) and WIBRaLS sources to compute the average distribution function ($\langle G_{\text{fake}}(LR_{i,j}) \rangle$) of fake $\log LR_{i,j}$. The $\log LR_{i,j}$ distributions for the real and fake matches are shown in blue and green, respectively, in Figure 3. As a matter of comparison, we overplot the $\log LR_{i,j}$ distributions when adding the exponential dependence of angular separation, r , shown in Equation (1), to our method (hatched distributions).

5. We then compare the real and the fake distributions of $\log LR_{i,j}$ to determine the reliability for the real associations by computing

$$\rho(LR_{i,j}) = 1 - \frac{\langle G_{\text{fake}}(LR_{i,j}) \rangle}{G_{\text{real}}(LR_{i,j})},$$

where $\langle G_{\text{fake}}(LR_{i,j}) \rangle$ is the average $\log LR_{i,j}$ distribution for fake catalogs and $G_{\text{real}}(LR_{i,j})$ is the $\log LR_{i,j}$ distribution for the real γ -ray source catalog, as shown in Figure 3. The reliability, ρ , computed according to this equation represents an approximate measurement of the

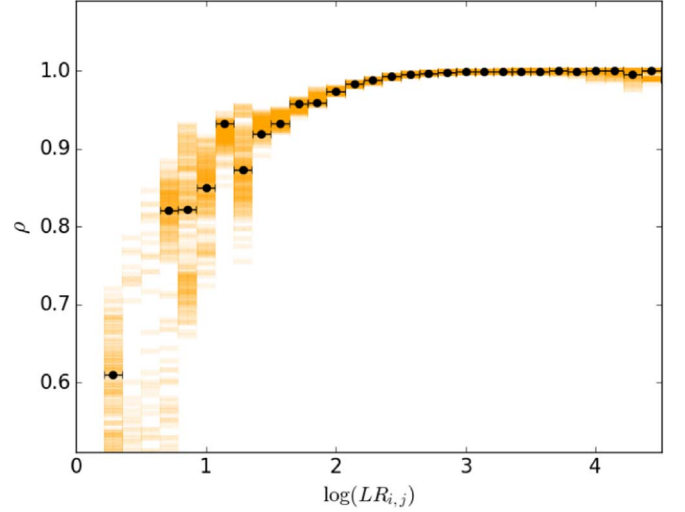


Figure 4. Reliability ρ derived from the LR method as a function of $\log(LR_{i,j})$. The black dots are the points adopted in this work, while the orange shadow represents the uncertainty region due to different choices of radii when computing the local surface density, $N(\Delta c_{12}, \Delta c_{34}, \Delta q_{22})$. In the region $\log(LR_{i,j}) > 1.5$, the impact of choosing different radii is negligible.

AP for a potential counterpart having a given $\log LR_{i,j}$. In Figure 4, we show the dependence of ρ on $\log(LR_{i,j})$. Some fluctuations are observed for $\log(LR_{i,j}) < 1.5$, but the overall behavior of the curve is clear: for $\log(LR_{i,j}) > 1.5$, basically all counterpart candidates have very high ($>95\%$) APs. The orange shadow in Figure 4 represent the uncertainty region caused by choosing different radii when computing the local surface density (see step 3 above). This region was generated by ranging the radii from 3.5° up to 15° in 200 linearly spaced intervals. We observe that, in the region defined by $\log(LR_{i,j}) > 1.5$, the APs do not really depend on how we choose the background region radius.

The APs obtained with this method range from 60% to 100%. Among the sources listed as identified in 4FGL—those for which the association to the low-energy counterpart is guaranteed—the APs are always above 99%.

5. The RF Method

In the context of supervised learning algorithms, we use a RF (Breiman 2001) to compute the AP of 4FGL counterparts. The RF method is an ensemble classifier that uses decision trees as

Table 1
List of Sources with Associations Provided with the LR Method Developed Here and that Have a Different Association in 4FGL

4FGL name	Association WIBRaLS	Association 4FGL	AP LR WIB	AP LR 4FGL
J0119.9+4053	J011947.90+405418.4	CRATES J012018+405314	0.97	...
J0211.1-0646	J021116.95-064419.9	LEDA 1029376	0.99	0.96
J0237.7+0206	J023737.97+020742.5	PKS 0235+017	0.93	...
J0402.9+6433	J040254.43+643510.0	1RXS J040301.8+643446	0.99	0.83
J0640.9-5204	J064111.26-520232.2	ERC217 G261.25-22.62	0.99	...
J0725.6-3530	J072548.16-353041.8	NVSS J072547-353039	0.99	...
J0801.3-0617	J080141.07-060535.2	CRATES J080140-054037	0.61	...
J0807.7-1206	J080730.69-120608.9	CRATES J080736.06-120745.9	0.93	0.92
J1122.0-0231	J112213.71-022914.0	2QZ J112156-0229	0.96	...
J1136.3-0501	J113547.40-050804.8	NVSS J113607-050156	0.86	...
J1145.7+0453	J114631.75+045819.2	PKS 1142+052	0.93	0.89
J1300.4+1416	J130020.92+141718.4	OW 197	0.96	0.95
J1305.3+5118	J130645.87+511655.0	IERS B1303+515	0.83	...
J1516.8+2918	J151641.71+291816.5	RGB J1516+293	0.97	0.89
J1518.6+0614	J151847.70+061259.0	TXS 1516+064	0.96	0.96
J1550.8-1750	J155053.18-175618.4	TXS 1548-177	0.97	0.88
J1734.0+0805	J173510.44+080831.0	2MASS J17340287+0805237	0.86	0.83
J1807.1+2822	J180712.91+282059.5	WISE J180634.08+281908.0	0.86	...
J1953.0-7025	J195306.71-702428.7	PKS 1947-705	0.99	0.96
J2110.2-1021	J211038.30-103337.2	PKS 2107-105	0.86	0.95
J2209.8-5028	J221040.80-502652.5	PMN J2210-5030	0.96	...
J2329.7-2118	J232940.19-211345.0	PKS 2327-215	0.97	0.88

Note. As expected, most (but not all) of the associations listed in 4FGL are closer to the γ -ray emission center than the WIBRaLS candidate counterparts selected based on MIR colors. Starting from the left side, the columns are (i) the name of the γ -ray source in 4FGL, (ii) the WISE name of the WIBRaLS candidate counterpart selected based on MIR colors, (iii) the candidate counterpart listed in 4FGL, (iv) the AP obtained with the LR method developed here, and (v) the AP obtained with the LR used in 4FGL.

building blocks for classification (James et al. 2013). For classifying a new object, each tree in the forest chooses one class and, by aggregating the predictions of all decision trees, the RF makes a final prediction based on the choice made by the majority of the trees, thus improving the predictive capability and reducing the tendency of standard decision trees to over-fit the training sample.

We train and estimate the accuracy of the RF method with the sample described in Section 3 using cross-validation: the RF algorithm has been trained with 10 subsets, each containing 90% of the training sample (where the sources are randomly chosen) and has been tested on the complementary 10 subsets containing 10% of the training sample, in a way that the final accuracy of $78.2\% \pm 3.3\%$ is simply the average ratio of correctly predicted observations to the total observations. The following attributes listed in 4FGL and WIBRaLS were used during the learning process: the γ -ray power-law photon index, Γ ; WISE colors c_{12} , c_{23} , and c_{34} ; and radio-loudness, q_{22} , computed with respect to the infrared flux in band W4. We tested several other parameters available in 4FGL and WIBRaLS, as γ -ray variability index and ratios between γ -ray flux densities (power-law and logparabola spectral models) and the radio flux density, but all of them presented a negligible impact on the results, with improvements in the algorithm accuracy of $\sim 3\%$ in the best case (i.e., γ -ray variability index) and sometimes only adding noise. Among the used parameters, the one with largest impact in the accuracy when combined with Γ is c_{12} , followed by c_{34} , q_{22} , and c_{23} , respectively. As discussed in Section 1, we are neglecting angular separation between the center of the γ -ray sources and the position of the MIR sources listed in WIBRaLS, as we are proposing an association method that relies only on spectral properties.

We use the RF classifier available in the Python library `sklearn` (Pedregosa et al. 2011). The code fits several decision trees on subsamples of the data set and average them to improve the predicted accuracy and prevent over-fitting. To guarantee the stability of our results, we use 5000 trees and let the nodes grow until all leaves contain less than the minimum number of samples required to split an internal node. The APs obtained with the RF algorithm range from 5% to 100% and the lowest AP obtained for blazars listed as identified in 4FGL (all of them included in the training sample) is 70%.

6. Results from the LR Approach

The total number of associations with AP $>99\%$ (i.e., as high as the AP obtained for the identified 4FGL sources; see Section 4) is 743 out of the 1138 original 4FGL-WIBRaLS crossmatches within the intervals Δc_{12} , Δc_{34} , and Δq_{22} . This number increases to 1051 when considering counterparts with AP $>95\%$. Among all the 1138 sources for which we compute an AP, 283 are associated in 4FGL as BCUs, 473 as BL Lacs, 350 as flat spectrum radio quasars, and the rest of them are divided into a few non-blazar extragalactic and unknown objects. We found five UGSs with a counterpart in WIBRaLS and with APs ranging from 93% up to 99%. The BCUs and UGSs associated here—especially those with APs as high as the identified 4FGL sources—are promising targets for future optical spectroscopic follow-up missions. In fact, optical spectroscopic observations for some of them are available in de Menezes et al. (2020) and H. A. Peña-Herazo et al. (2020, in preparation), where all of the sources have blazar-like optical spectra. All BCUs and UGSs associated here are listed in Tables A1 and A2 in the Appendix.

Among the γ -ray sources associated here (with any AP), 22 of them have different counterparts in 4FGL that are not listed in WIBRaLS (see Table 1). The 22 counterparts listed in 4FGL

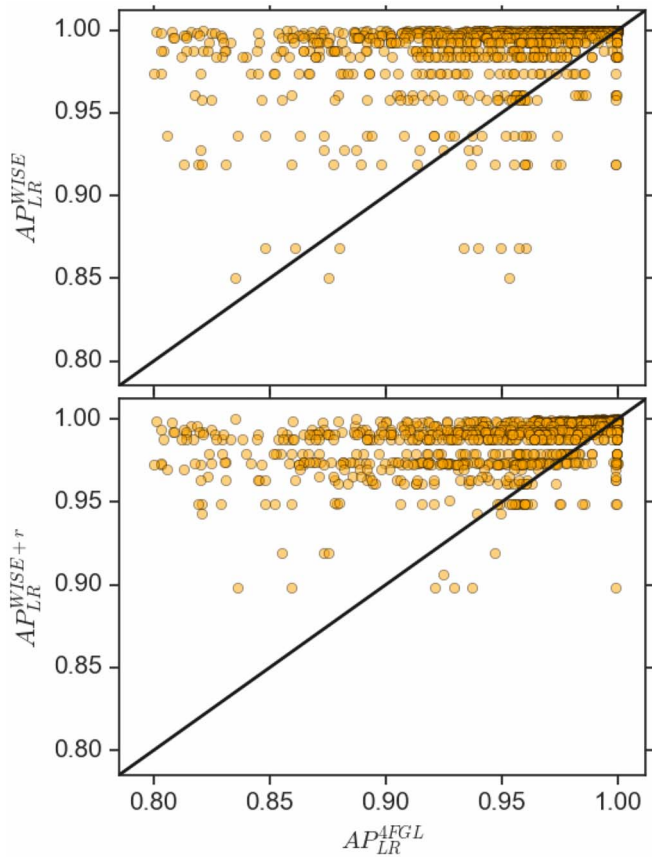


Figure 5. Comparison between the APs obtained with our LR method neglecting (upper panel) and considering (bottom panel) angular separation and the APs obtained with the LR method used in 4FGL. The APs computed by our method tend to be higher than those listed in 4FGL paper.

are generally closer to the center of the γ -ray sources than the WIBRaLS counterparts selected based on MIR colors, which is probably due to the exponential dependence with angular separation (see Equation (1)) adopted in 4FGL. Furthermore, 165 of the 1138 WIBRaLS counterpart candidates are not associated by the LR method adopted in 4FGL. These sources were associated in 4FGL mainly via the Bayesian method (Abdollahi et al. 2020). Additionally, no correlation is observed between the APs computed here and the APs computed with the LR method adopted in 4FGL, indicating that the methods are completely independent.

6.1. Considering Angular Separation

When we consider angular separation (with an exponential dependence as shown in Equation (1)) together with WISE MIR colors and q_{22} , the LR distributions look like the hatched histograms in Figure 3. As the distributions are very similar, the APs resulting when considering angular separation are also similar. The total number of associated counterparts in this case is 580 with $AP > 99\%$, or 1071 when considering $AP > 95\%$. Furthermore, all sources with $AP < 95\%$ have low γ -ray statistical significance ($< 10\sigma$), which generally implies in large positional and γ -ray photon index errors.

In the upper panel of Figure 5, we compare the APs obtained with the LR method developed here with those obtained with the LR adopted in 4FGL. In the bottom panel, we show the same comparison when taking into account the angular

separation, r , in the LR method. In both cases, our method tends to give higher APs.

7. Results from the RF Approach

We found a total of 960 associations with $AP > 70\%$ (i.e., in the same range of the identified 4FGL sources; see Section 5) out of 1311 test sources (see Section 3). Furthermore, the number of counterparts with $AP > 50\%$ increases to 1109. Due to the low accuracy presented by the RF method (see Section 5), we use these results only as a comparison to the LR method described in the previous section. The APs for the BCUs and UGSs associated with the RF are shown in Tables A1 and A2 in the Appendix.

When comparing the APs obtained with the RF algorithm and those obtained with the LR methods used here and in 4FGL, we found no significant correlation, although the APs computed with the RF approach tend to be smaller and distributed over a larger range. The major difference between both methods is the way we feed the two algorithms with the data. In Figure 1, it is clear that several fake associations (orange dots) occupy the same region as the high-confidence associations, thus counterparts lying within the high-confidence association region (black contours) still have a high chance of being a fake association in the RF approach. The same does not happen with the LR method. Furthermore, based on how we defined the training set and the fake associations, we are solving an intrinsically more complicated task with the RF approach than with the LR method.

8. Discussion and Conclusions

Associating γ -ray sources with their low-energy counterparts is challenging (Massaro et al. 2013). Given the absence of multiwavelength monitoring and optical spectroscopic data within the error ellipses of many 4FGL sources, we have to recur to statistical methods when associating γ -ray sources to their counterparts.

In this work, we propose an alternative to the LR association method adopted by the Fermi-LAT Collaboration (Abdollahi et al. 2020). Here, the APs are independent of angular separation and apparent magnitude and rely only on spectral parameters such as MIR colors and radio-loudness. As a complementary association method, we trained an RF algorithm to associate γ -ray sources to their counterparts. Both methods, however, are suited only for associating blazar candidates to γ -ray sources. Other types of γ -ray sources, like pulsars and starburst galaxies (to name a few), cannot be directly associated with the proposed methods.

In the LR method adopted here, we naturally loose some true blazar associations that have MIR colors just outside the blazar locus (i.e., the region of the WISE color-color-color diagram populated by confirmed γ -ray blazars). The sources associated here, however, are very good γ -ray blazar candidates and excellent targets for future optical spectroscopic follow-ups. Furthermore, as all sources in WIBRaLS are radio-loud with respect to the q_{22} parameter (D’Abrusco et al. 2019), it is likely to find many BL Lac-galaxy dominated among the sources excluded by our method. It is really unlikely that these sources have no sign of nuclear activity, as the presence of non-active galactic nuclei in WIBRaLS is estimated to be $< 5\%$ (de Menezes et al. 2019). The results from our analysis are as follows.

1. The total number of associations computed with the LR method is 743 with $AP > 99\%$, or 1051 associations with $AP > 95\%$. We associate a total of 283 BCUs and five UGSs (listed in Tables A1 and A2 in the Appendix).
2. Adding a dependence on angular separation to the LR method does not significantly change the results, indicating that WIBRaLS sources are generally in good positional agreement with 4FGL sources.
3. A supervised learning method is used for the first time to associate counterparts to γ -ray sources. As the performance of the algorithm is $\sim 80\%$, the results derived from this method are considered complementary to those obtained with the LR method and should not be interpreted as giving definitive associations to 4FGL sources.

This is the first work where spectral properties of blazars are used to associate γ -ray sources to their low-energy counterparts. Previous methods rely basically on parameters like the apparent magnitude and angular separation between the γ -ray source center and the position of its candidate counterpart.

The lists of BCUs and UGSs available in the Appendix give us the excellent opportunity to test the LR method presented here. In Tables A1 and A2, the targets are sorted alphabetically and can be prioritized by choosing those with higher APs. Optical spectroscopy follow-ups of these targets may be crucial to test if the association based on MIR colors is indeed an appropriate approach.

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) under grants 2016/25484-9, 2018/24801-6 (R.M.) and 2017/01461-2 (R.N.). The work of F. M. is partially supported by the “Departments of Excellence 2018–2022” grant awarded by the Italian Ministry of Education,

University and Research (MIUR) (L. 232/2016) and by the Ministry of Education, Universities and Research for the grant MASFFABR_17_01. F.M. also acknowledges financial contribution from the agreement ASI-INAF n.2017-14-H.0. R.D.A. is supported by National Aeronautics and Space Administration (NASA) contract NAS8-03060 (Chandra X-ray Center). This investigation is supported by NASA grants GO6-17081X, GO9-20083X, and GO0-21110X. We thank the anonymous referee for the constructive comments allowing us to improve the manuscript.

In this publication, we made use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. We also used TOPCAT⁹ (Taylor 2005) and astropy¹⁰ (Astropy Collaboration et al. 2013, 2018) for preparation and manipulation of the data.

Appendix

List of Targets for Optical Spectroscopic Follow-up

A list with all BCUs associated with the LR method is provided in Table A1, where the sources are promising targets for future optical spectroscopic follow-ups (Massaro et al. 2015c; Ricci et al. 2015; Álvarez Crespo et al. 2016b). The columns are the name of the source in γ -rays and MIR, the AP obtained with the LR method developed here, the AP obtained with the RF approach, the AP obtained with the Bayesian method used in 4FGL, and the AP computed with the LR method based on the angular distance as in 4FGL. Similarly, Table A2 lists the five UGSs described in Sections 6 and 7 and their APs.

Table A1
BCUs Associated by the LR Method Based Only on Spectral Parameters

4FGL Name	WISE Name	LR WISE	RF	Bayesian	LR 4FGL
J0001.6-4156	J000132.74-415525.2	0.99	0.97	1.0	0.85
J0008.0-3937	J000809.17-394522.8	0.87	0.61	0.92	0.88
J0010.8-2154	J001053.64-215704.2	0.96	0.42	0.99	0.95
...	...	...	...	...	...
J1238.1-4541	J123806.03-454129.6**	0.99	1.0	1.0	0.94
J1239.4+0728	J123924.58+073017.2*	1.0	0.96	1.0	0.94
...	...	...	...	...	...

Note. The BCUs listed here are good targets for future optical spectroscopic follow-ups. In the first two columns, we have the names of the γ -ray source and its WISE counterpart. The last four columns show the APs as given by the LR and RF methods in this work and by the Bayesian and LR methods used in 4FGL. Sources tagged with “*” and “**” are confirmed blazars, with optical spectra available in de Menezes et al. (2020) and H. A. Peña-Herazo et al. (2020, in preparation), respectively. (This table is available in its entirety in machine-readable form.)

⁹ <http://www.star.bris.ac.uk/~mbt/topcat/>

¹⁰ <https://www.astropy.org/index.html>

Table A2
UGSs Associated by the LR Method

4FGL Name	WISE Name	LR WISE	RF
J0202.4+2943	J020239.70+294325.0	0.99	0.75
J0620.7-5034	J062046.13-503350.9	0.97	0.68
J1028.1-3550	J102751.56-355210.7	0.98	0.31
J1309.3+8035	J130628.16+803142.3	0.93	0.61
J2013.9-8717	J201440.97-871707.8	0.98	0.90

Note. These sources are good targets for future optical spectroscopic follow-ups. In the first two columns, we have the names of the γ -ray source and its WISE counterpart. The last two columns show the AP as given by the LR and RF methods.

ORCID iDs

Raniere de Menezes  <https://orcid.org/0000-0001-5498-4925>

Raffaele D'Abrusco  <https://orcid.org/0000-0003-3073-0605>

Francesco Massaro  <https://orcid.org/0000-0002-1704-9850>

Dario Gasparini  <https://orcid.org/0000-0002-5064-9495>

Rodrigo Nemmen  <https://orcid.org/0000-0003-3956-0331>

References

- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, *ApJS*, **188**, 405
- Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, *ApJS*, **247**, 33
- Acero, F., Ackermann, M., Ajello, M., et al. 2015, *ApJS*, **218**, 23
- Ackermann, M., Ajello, M., Allafort, A., et al. 2011, *ApJ*, **743**, 171
- Ajello, M., Angioni, R., Axelsson, M., et al. 2019, arXiv:1905.10771
- Álvarez Crespo, N., Massaro, F., & Milisavljevic, D. 2016a, *AJ*, **151**, 95
- Álvarez Crespo, N., Masetti, N., & Ricci, F. 2016b, *AJ*, **151**, 32
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Atwood, W., Abdo, A. A., Ackermann, M., et al. 2009, *ApJ*, **697**, 1071
- Breiman, L. 2001, *Mach. Learn.*, **45**, 5
- Chiaro, G., Salvetti, D., la Mura, G., et al. 2016, *MNRAS*, **462**, 3180
- D'Abrusco, R., Crespo, N. Á., Massaro, F., et al. 2019, *ApJS*, **242**, 4
- D'Abrusco, R., Massaro, F., Ajello, M., et al. 2012, *ApJ*, **748**, 68
- D'Abrusco, R., Massaro, F., Paggi, A., et al. 2013, *ApJS*, **206**, 12
- D'Abrusco, R., Massaro, F., Paggi, A., et al. 2014, *ApJS*, **215**, 14
- de Menezes, R., Amaya-Almazán, R. A., Marchesini, E. J., et al. 2020, *Ap&SS*, **365**, 12
- de Menezes, R., Peña-Herazo, H. A., Marchesini, E. J., et al. 2019, *A&A*, **630**, A55
- Doert, M., & Errando, M. 2014, *ApJ*, **782**, 41
- Fichtel, C., Bertsch, D., Chiang, J., et al. 1994, *ApJS*, **94**, 551
- Hassan, T., Mirabal, N., Contreras, J., & Oya, I. 2012, *MNRAS*, **428**, 220
- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, *An Introduction to Statistical Learning* (Berlin: Springer)
- Kovačević, M., Chiaro, G., Cutini, S., & Tosti, G. 2019, *MNRAS*, **490**, 4770
- Landoni, M., Massaro, F., Paggi, A., et al. 2015, *AJ*, **149**, 163
- Lefaucheur, J., & Pita, S. 2017, *A&A*, **602**, A86
- Marchesini, E., Peña-Herazo, H., Crespo, N. Á., et al. 2019, *Ap&SS*, **364**, 5
- Masci, F. J., Condon, J., Barlow, T., et al. 2001, *PASP*, **113**, 10
- Massaro, E., Maselli, A., Leto, C., et al. 2015a, *Ap&SS*, **357**, 75
- Massaro, F., Crespo, N. Á., D'Abrusco, R., et al. 2016, *Ap&SS*, **361**, 337
- Massaro, F., & D'Abrusco, R. 2016, *ApJ*, **827**, 67
- Massaro, F., D'Abrusco, R., Ajello, M., Grindlay, J., & Smith, H. A. 2011, *ApJL*, **740**, L48
- Massaro, F., D'Abrusco, R., Landoni, M., et al. 2015b, *ApJS*, **217**, 2
- Massaro, F., D'Abrusco, R., Paggi, A., et al. 2013, *ApJS*, **209**, 10
- Massaro, F., Landoni, M., & D'Abrusco, R. 2015c, *A&A*, **575**, 124
- Mirabal, N., Frias-Martinez, V., Hassan, T., & Frias-Martinez, E. 2012, *MNRAS*, **424**, L64
- Nolan, P. L., Abdo, A., Ackermann, M., et al. 2012, *ApJS*, **199**, 31
- Paggi, A., Milisavljevic, D., Masetti, N., et al. 2014, *AJ*, **147**, 112
- Parkinson, P. S., Xu, H., Yu, P., et al. 2016, *ApJ*, **820**, 8
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Peña-Herazo, H., Massaro, F., Chavushyan, V., et al. 2019, *Ap&SS*, **364**, 85
- Peña-Herazo, H. A., Marchesini, E. J., Crespo, N. Á., et al. 2017, *Ap&SS*, **362**, 228
- Ricci, F., Massaro, F., Landoni, M., et al. 2015, *AJ*, **149**, 160
- Richter, G. 1975, *AN*, **296**, 65
- Salvetti, D., Chiaro, G., la Mura, G., & Thompson, D. J. 2017, *MNRAS*, **470**, 1291
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, **259**, 413
- Taylor, M. B. 2005, in *ASP Conf. Ser. 347, TOPCAT & STIL: Starlink Table/VO Table Processing Software*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Wolstencroft, R., Savage, A., Clowes, R., et al. 1986, *MNRAS*, **223**, 279