# An Extended List of Galaxies for Gravitational-Wave Searches in the Advanced Detector Era

G. Dálya, Z. Frei, G. Galgóczi, P. Raffai, R. S. de Souza
Eötvös University, Institute of Physics, 1117 Budapest, Hungary
E-mail: dalyag@caesar.elte.hu
Last updated: 11/04/2015

## I. INTRODUCTION

Spatial resolution of advanced GW detectors is still low, expected localisation areas are ≈ 10-100 deg$^2$ [1], what makes the identification of electromagnetic (EM) counterparts of gravitational wave (GW) trigger events challenging even for wide-field telescopes. As it has been shown in multiple papers (eg. Abadie et al. [2], Hanna et al. [1]) the usage of a galaxy catalog could greatly increase the chance of catching the EM counterpart. The current catalog used by the LIGO/Virgo Collaboration (GWGC) is complete to 40 Mpc, however, in the advanced detector era a much higher completeness is needed (even to ≈ 300 Mpc [3]).

We are introducing GLADE (Galaxy List for the Advanced Detector Era), a value-added full-sky galaxy catalog with high completeness in order to support future EM follow-up projects of the Collaboration. The catalog has been constructed (combined and matched) from four existing galaxy catalogs: GWGC[1], 2MPZ[2], 2MASS XSC[3] and HyperLEDA[4]. It contains 1,918,147 galaxies, which is two orders of magnitude greater than the number of galaxies in the GWGC catalog alone (53,312), which is currently in use by the collaboration. Furthermore, we have associated B-band magnitudes and photometric redshifts for 548,876 2MASS galaxies which lacked these properties with a regression algorithm teached on a subsample of the 2MPZ catalog. Our catalog is complete to 73 Mpc and even at 300 Mpc has a relatively high completeness (53%). It contains measured or estimated distance and B-band magnitude for all of the galaxies. Naturally, our catalog could be used in a broad range of various astrophysical projects besides EM follow-up efforts.

In section II we introduce the methods used in the construction of the catalog such as finding duplicate galaxies and filling in missing data. In section III we investigate its completeness and in section IV we summarize the conclusions.

## II. CATALOG COMPILATION AND STATISTICS

Our main goal was to extend the currently used Gravitational Wave Galaxy Catalogue (GWGC) both in numbers of galaxies, completeness and available information about each galaxy. We achieved this by merging the GWGC with three other catalogs: the 2 Micron All-Sky Survey Extended Source Catalog (2MASS XSC), the 2MASS Photometric Redshift Catalog (2MPZ) and the HyperLEDA catalog.

The GWGC [4] has been merged from four existing catalogs: the Tully Nearby Galaxy Catalog, the Catalog of Neighboring Galaxies, the V8k catalog and HyperLEDA, however has a cut at 100 Mpc. It also includes a list of 150 known Milky Way globular clusters.

The 2MASS XSC [5] contains position and photometry in J, H and K bandpasses for more than 1.6 million objects, from which approximately 97% are galaxies. Unfortunately, B magnitudes and redshifts which are important for the EM follow-up are not included in the catalog thus we had to estimate them using a regression algorithm.

The 2MPZ catalog [6] has been constructed by cross-matching 2MASS XSC, WISE and SuperCOSMOS, and employing an artificial neural network algorithm trained on redshift surveys. It contains both B magnitudes and

[1] http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=GWGC
[2] http://ssa.roe.ac.uk/TWOMPZ.html
[3] http://www.ipac.caltech.edu/2mass/
[4] http://leda.univ-lyon1.fr/

photometric redshifts for its more than 900,000 galaxies.

HyperLEDA [7] was created by merging LEDA and Hypercat, and contains over 3 million objects from which we kept only those who were identifiable as galaxy with a high level of confidence. That means more than 800,000 galaxies.

We aimed to reduce the number of false entries in our catalogue as much as possible, therefore we only categorized objects from HyperLEDA as being galaxies if they satisfied all these following criteria:

1. The object must have an object type G. This criterion reduces the HyperLEDA objects to galaxies and quasars.

2. The object must have a logd25 value, which means, that the object is an extended source. We believe this excludes all the quasars from our sub-sample.

3. The object must have a modz and bt value. These criteria were added based on the requirement that we would like to have distance estimates and B magnitudes for all our entries in the GLADE catalogue.

We found it crucial to reduce the level of redundancy in our catalog by eliminating duplicate galaxies. In order to achieve this goal in a fast and reliable way we have employed a k-dimensional (k-d) tree. It is a space-partitioning data structure, where nearest neighbor search could be done effectively, with an $\mathcal{O}(\log n)$ time complexity. We designated two galaxies the same if the offset between their positions was smaller than $0.001°$. This criterion was justified by two reasons. First, by applying a higher value there were several cases where more than one galaxies were identyfied from one catalog to be the same as a single galaxy in the other catalog. Second, when we chose two catalogs and calculated the distances between the galaxies of the first catalog and their nearest neighbors in the second we got histograms like the one presented in Figure 1. The histogram indicates that there are lots of galaxies which has a pair in the other catalog nearer than $0.001°$, and there is a sharp cutoff at approximately this value, so we characterized the galaxies with nearest neighbor farther than the cutoff as different ones.

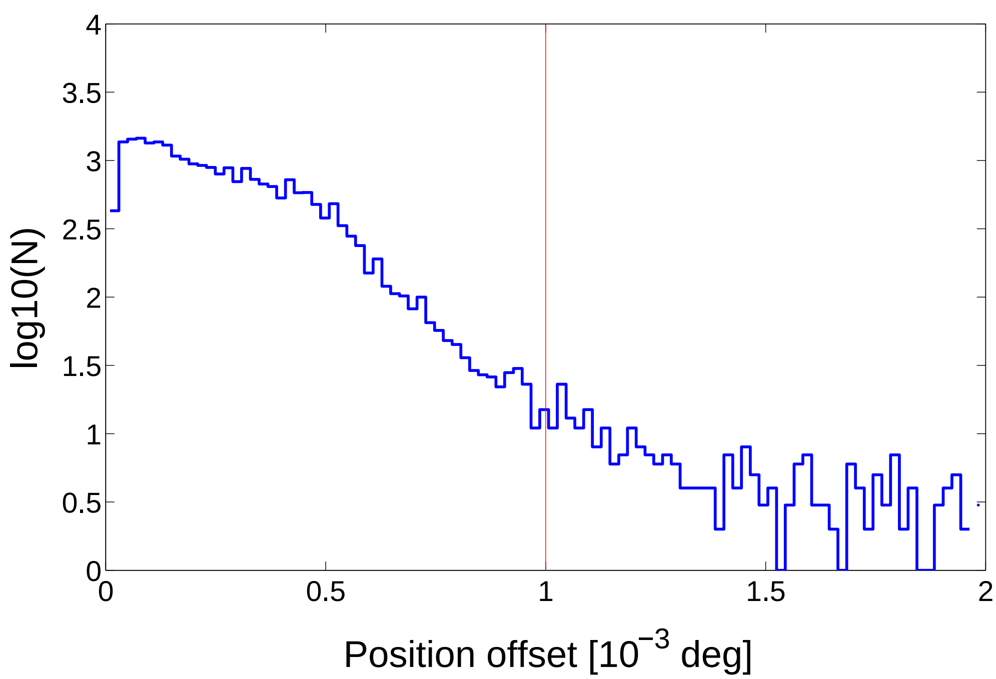


FIG. 1. Histogram of position difference between GWGC galaxies and their nearest neighbor in the 2MPZ catalog

After applying the k-d tree with the aforementioned selection criterion 2,068,841 galaxies were left in our catalog. Their distribution on the sky could be seen in Figure 2 in the form of a density plot. As it could be observed in the figure, our catalog is non-uniform. Some strips are clearly visible where more observations have been performed. These are arising from the HyperLEDA catalog. The plane of the Milky Way is also noticeable, as its gas and dust blocks our view.

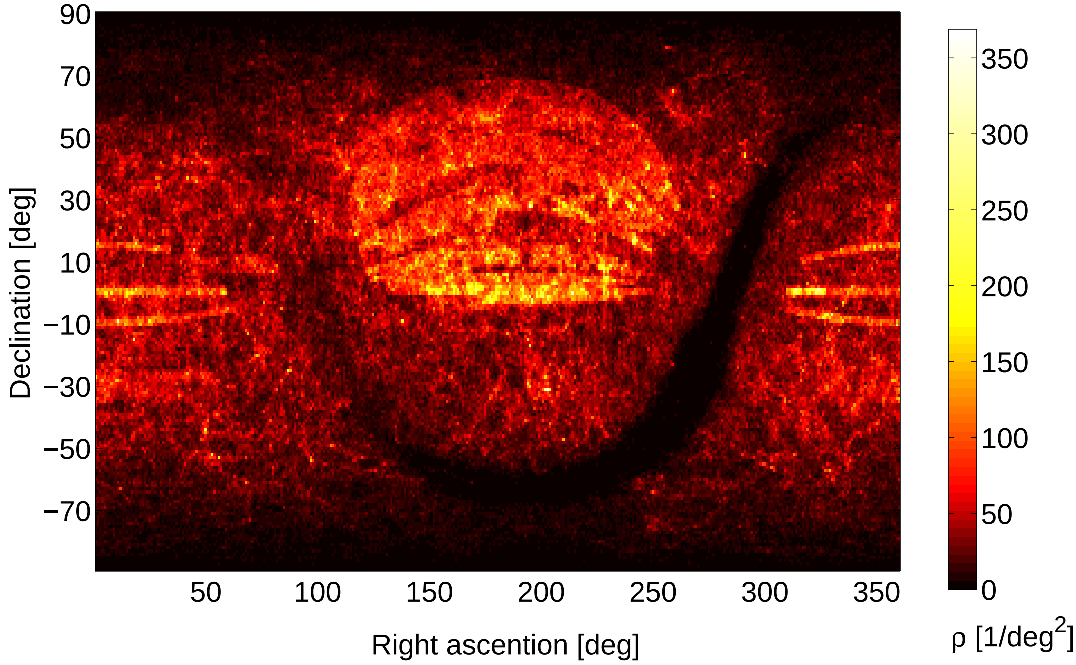


FIG. 2. The density of galaxies in the GLADE catalog

Having B-band magnitudes and (photometric) redshifts for all of the galaxies was considered as a crucial requirement while creating the catalog. These parameters were not available in the 2MASS XSC catalog, however, since 2MPZ galaxies are a subsample of 2MASS XSC galaxies and they contain the aforementioned parameters we have implemented a machine learning algorithm in order to fill in the missing values. We have chosen to use quantile regression forest, since it is a robust methodology and requires little fine tuning. It is a popular and powerful tool for regression problems [8].

We have derived the redshifts and B-band magnitudes for only 26% of the 2MASS XSC catalog, since the estimation errors for these parameters given by the random forest algorithm are highly dependent on the number of magnitude bands used. The algorithm was not accurate enough when we used only two of the 2MASS J, H, and K bands, therefore we applied it only when for an extended source (galaxy), all the J, H, and K magnitudes were available in the 2MASS XSC catalog. Furthermore, we have not given any estimates on the photo z and B band magnitude values for 2MASS XSC sources that had these values readily available from another catalogue.

We have investigated the precision of the regression forest as a function of the size of the teaching sample and concluded that over 10,000 galaxies the precision enhances slowly, however, it becomes computationally costly to use larger samples. That is why we used a teaching sample of 10,000 2MPZ galaxies. The learning was based on J, H and K magnitudes. We have associated B magnitude and redshift values for 548,876 galaxies, which is 26% of the whole catalog. It is worth to notice, that those galaxies are typically farter away, their mean distance is $\approx 340$ Mpc. We have also taken into consideration the fact that probably not everyone would like to use the predicted values in their work, hence we marked the galaxies containing the associated values.

## III. COMPLETENESS

In order to determine the completeness of our catalog we have executed a similar investigation as in White et al. [4]. We have compared the cumulative blue luminosity of the galaxies in our catalog to the expected value assuming a blue luminosity density of $(1.98 \pm 0.16) \times 10^{-2}$ $L_{10}$ Mpc$^{-3}$ from Kopparapu et al. [9], where $L_{10} = 10^{10}$ $L_{B,\odot}$ and

$L_{B,\odot}$ is the B-band luminosity of the Sun, namely $L_{B,\odot} = 2.16 \times 10^{33}$ ergs/s. The value was calculated to a redshift of $z = 0.1$ (420 Mpc). Our catalog exceeds this limit, however, according to Faber et al. [10] the blue luminosity density at $z = 0.3$ (1.2 Gpc) is $(1.92 \pm 0.23) \times 10^{-2}\ L_{10}\ \mathrm{Mpc}^{-3}$, so we can conclude that its value does not change between $z = 0.3$ and $z = 0.1$, so for simplicity we can use the value from Kopparapu et al. [9] for determining the completeness even to 1.2 Gpc.

Figures 3-6 shows the completeness of the different catalogs used, while in figure 7 you can see the completeness percentage of the final catalog against the distance, from where it could be determined, that the catalog is complete to 73 Mpc.

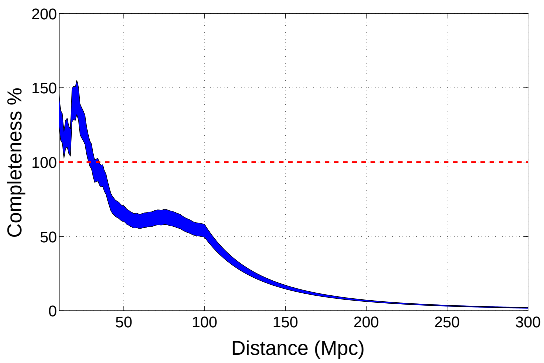


FIG. 3. Completeness of GWGC

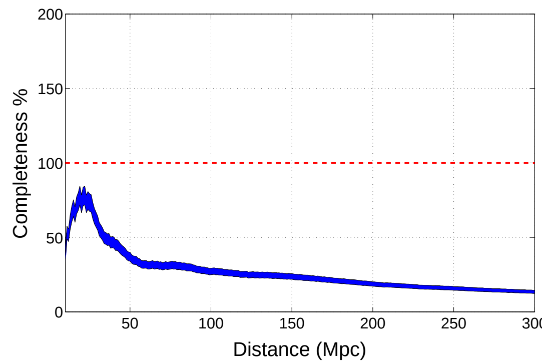


FIG. 4. Completeness of HyperLEDA

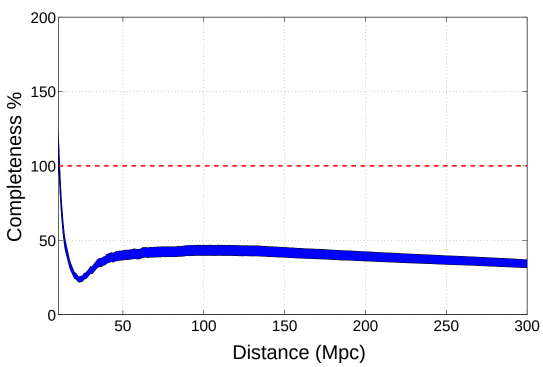


FIG. 5. Completeness of 2MPZ

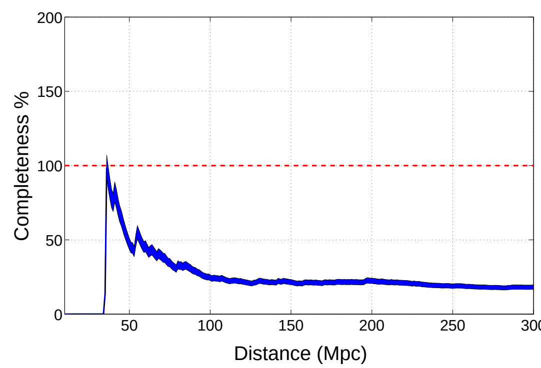


FIG. 6. Completeness of 2MASS

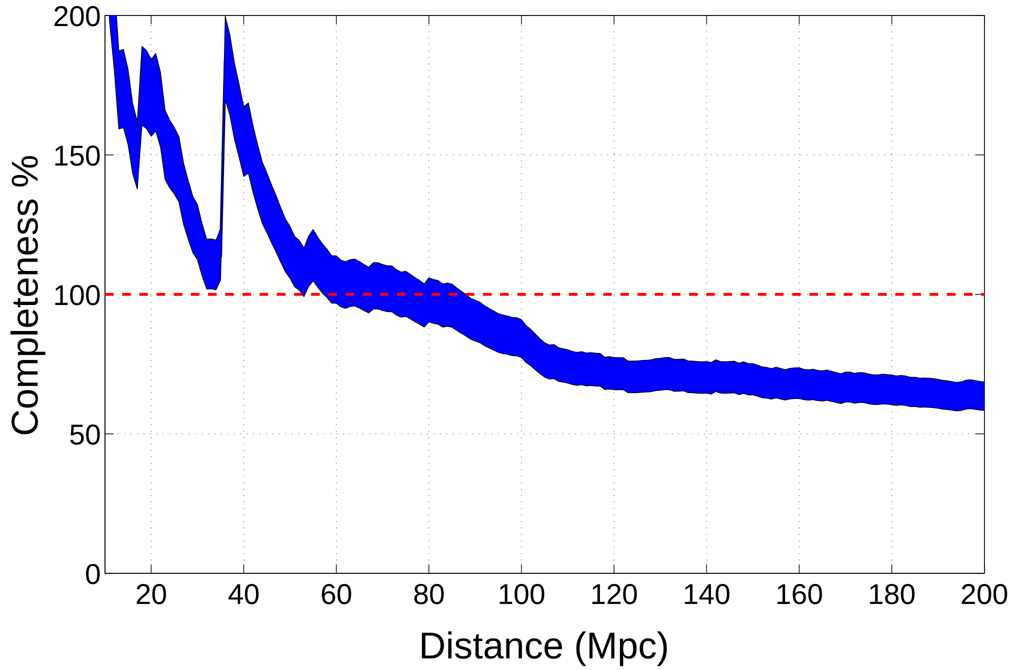


FIG. 7. Completeness percentage of the catalog against the distance. The width of the curve represents its error.

## IV. CONCLUSIONS

The GLADE catalog contains nearly 2 million galaxies and complete to 73 Mpc. Its completeness does not decreases rapidly at farther distances, eg. at 300 Mpc its completeness is still above 50%, which makes it a suitable catalog for supporting EM follow-up in the forthcoming advanced detector era.

The catalog is value-added with B magnitudes and photometric redshifts. The same pipeline which we used to associate these values can be used to predict other relevant parameters for the galaxies. We are going to continuously upgrade and expand our catalog to maintain or even improve its suitability for the future projects. We are currently working on determining masses for the galaxies which would help rank them based on the probability of NS-NS and NS-BH mergers.

The catalog can be downloaded from the following website: aquarius.elte.hu/glade

[1] Hanna, C., Mandel, I., Vousden, W., 2014, ApJ, **784** 7
[2] J. Abadie, B. P. Abbott, R. Abbott et al., 2012, AAP, **539** A124
[3] Evans, P. A, Osborne, J. P., Kennea, J. A. et al., 2015, astro-ph/1506.01624
[4] White, D. J., Daw, E. J. & Dhillon, V. S., 2011, Class. Quantum Grav., **28** 085016
[5] Skrutskie, M. F., Cutri, R. M., Stiening, R. et al., 2006, AJ, **131** 2
[6] Bilicki, M., Jarrett, T. H., Peacock, J. A. et al., 2014, ApJS, **210** 16
[7] Makarov, D., Prugniel, P., Terekhova, N. et al., 2014, A&A, **570** A13
[8] Meinshausen, N., 2006, Journal of Machine Learning Research, **7** 983
[9] Kopparapu, R. K., Hanna, C., Kalogera, V., et al. 2008, ApJ, **675** 1459
[10] Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, ApJ, **665** 265