Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **506,** 1896–1915 (2021) https://doi.org/10.1093/mnras/stab1799
Advance Access publication 2021 June 25

The Heraklion Extragalactic Catalogue (HECATE): a value-added galaxy catalogue for multimessenger astrophysics

K. Kovlakas ,[1,2★] A. Zezas ,[1,2,3] J. J. Andrews ,[4] A. Basu-Zych ,[5,6] T. Fragos ,[7] A. Hornschemeier ,[5,8] K. Kouroumpatzakis ,[1,2] B. Lehmer [9] and A. Ptak [5]

[1] *Physics Department, University of Crete, GR 71003, Heraklion, Greece*
[2] *Institute of Astrophysics, Foundation for Research and Technology – Hellas, GR 71110 Heraklion, Greece*
[3] *Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[4] *Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA), Northwestern University, 1800 Sherman Avenue, Evanston, IL 60201, USA*
[5] *NASA Goddard Space Flight Center, Laboratory for X-ray Astrophysics, Greenbelt, MD 20771, USA*
[6] *Department of Physics, University of Maryland Baltimore County, Baltimore, MD 21250, USA*
[7] *Geneva Observatory, University of Geneva, Chemin des Maillettes 51, CH-1290 Sauverny, Switzerland*
[8] *Department of Physics and Astronomy, The Johns Hopkins University, Homewood Campus, Baltimore, MD 21218, USA*
[9] *Department of Physics, University of Arkansas, 825 West Dickson Street, Fayetteville, AR 72701, USA*

Accepted 2021 June 22. Received 2021 May 24; in original form 2020 December 1

ABSTRACT

We present the *Heraklion Extragalactic Catalogue*, or *HECATE*, an all-sky value-added galaxy catalogue, aiming to facilitate present and future multiwavelength and multimessenger studies in the local Universe. It contains 204 733 galaxies up to a redshift of 0.047 ($D \lesssim 200$ Mpc), and it is >50 per cent complete in terms of the *B*-band luminosity density at distances in the 0–170 Mpc range. By incorporating and homogenizing data from astronomical data bases and multiwavelength surveys, the catalogue offers positions, sizes, distances, morphological classifications, star formation rates, stellar masses, metallicities, and nuclear activity classifications. This wealth of information can enable a wide range of applications, such as (i) demographic studies of extragalactic sources, (ii) initial characterization of transient events, and (iii) searches for electromagnetic counterparts of gravitational-wave events. The catalogue is publicly available to the community at a dedicated portal, which will also host future extensions in terms of the covered volume and data products.

Key words: gravitational waves – astronomical data bases: miscellaneous – catalogues – galaxies: general.

1 INTRODUCTION

With the availability of all-sky surveys across the electromagnetic spectrum (e.g. *LSST*, *ZTF*, and *eROSITA*) and the advent of the era of multimessenger observations (e.g. gravitational-wave, neutrino, and cosmic ray observatories), there is an increasing need for homogenized extragalactic catalogues that can be used for the characterization of individual sources.

Astronomical data bases like *NED* (Helou et al. 1991), *SIMBAD* (Wenger et al. 2000), and *HyperLEDA* (Makarov et al. 2014) have significantly boosted extragalactic research via the collection and organization of data such as positions, distances, photometric fluxes, and morphological classifications. However, due to the diversity of the different sources of these data, they cannot be readily used for studies requiring derived galaxy properties such as star formation rate (SFR), stellar mass ($M_\star$), metallicity, and nuclear activity, for large samples of objects. Although detailed catalogues based on focused surveys provide such information (e.g. *MPA–JHU*; Kauffmann et al. 2003; Brinchmann et al. 2004; Tremonti et al. 2004), the lack of all-sky coverage limits their usefulness for many astrophysical applications, such as the characterization of sources in multiwavelength all-sky or serendipitous surveys (e.g. X-ray surveys; Kim et al. 2007; Saxton et al. 2008).

The rapid identification of counterparts of transient sources such as gamma-ray bursts (GRBs) or rare events (e.g. high-energy cosmic rays) and the strategic planning of follow-up observations are possible with the aid of all-sky galaxy catalogues. Furthermore, the use of astrophysical information has been used to increase the effectiveness of identifying the hosts of gravitational-wave (GW) sources (e.g. Abbott et al. 2017a). To this extent, there is a growing effort to build galaxy catalogues providing information on $M_\star$ or SFR (or their proxies) aiming to aid GW follow-up observations (e.g. Kopparapu et al. 2008; White, Daw & Dhillon 2011; Gehrels et al. 2016; Dálya et al. 2018; Cook et al. 2019; Ducoin et al. 2020). However, these catalogues do not provide metallicity, which can be a key factor for the identification of GW hosts (e.g. Artale et al. 2019). Since the aforementioned galaxy catalogues were designed for applications focusing on distant galaxies (e.g. GWs and GRBs), the provided data may not be very accurate for nearby galaxies (e.g. $D < 40$ Mpc), which often require special treatment (e.g. extended versus point-source photometry, and distance measurements versus application of the Hubble–Lemaître law). Therefore, studies involving nearby galaxy samples often invest in compiling the necessary galaxy data from scratch.

★ E-mail: kkovlakas@physics.uoc.gr

In order to enable large-scale studies of transient events such as those described earlier, or multiwavelength properties of galaxies (e.g. X-ray or gamma-ray scaling relations; Ackermann et al. 2012; Komis, Pavlidou & Zezas 2019; Kovlakas et al. 2020), we require an all-sky catalogue that gives accurate locations, galaxy dimensions, distances, multiband photometry, and most importantly derived stellar population parameters. For this reason, we compiled an all-sky value-added catalogue of 204 733 nearby galaxies within a distance of 200 Mpc: the *Heraklion Extragalactic Catalogue* (*HECATE*[1]). This catalogue provides all the aforementioned quantities based on a variety of sources. Special care is taken to develop procedures that consolidate the available data, maximize the coverage of the parameters, and address possible biases and offsets between different parent catalogues. The derivation of homogenized stellar population parameters, including metallicity, and nuclear activity classifications highlight the usefulness of the *HECATE* as a reference sample for the characterization of sources in multiwavelength and/or multimessenger observations. The catalogue is publicly available at the *HECATE* portal: http://hecate.ia.forth.gr.

In Section 2, we describe the selection of galaxies from the *HyperLEDA* data base, and the incorporation of redshift (z) and size information. The assembly and combination of distance measurements, as well as the derivation of z-dependent distances for galaxies without distance measurements, are described in Section 3. The compilation of multiwavelength data and the derivation of stellar population parameters are presented in Section 4. In Section 5, we compare the *HECATE* with other galaxy catalogues, discuss its limitations, and present various applications. Finally, in Section 6 we present future extensions of the catalogue. Throughout the paper, unless stated otherwise, uncertainties correspond to 68 per cent confidence intervals.

2 SAMPLE SELECTION

As the basis of our catalogue, we use the *HyperLEDA* data base (Makarov et al. 2014), which includes, combines, and homogenizes extragalactic data in the literature, without explicit flux or volume limits. Furthermore, common problems such as misprints, duplication, poor astrometry, and wrong associations that can be found in legacy catalogues (e.g. *UGC*: Nilson 1973; *RC3*: de Vaucouleurs 1991) are generally identified and rectified by the *HyperLEDA* pipeline.

Out of the 5 377544 objects in the *HyperLEDA* (as of 2019 October), we select 3 446810 (64 per cent) that are characterized as individual galaxies (‘objtype=G’), excluding multiple systems (but not their members), groups, clusters, parts of galaxies, stars, nebulae, etc.

Since the distances for the majority of the galaxies have not been measured, we perform the selection of local Universe galaxies based on a recession velocity limit. We note that reported heliocentric radial velocity measurements typically contain the components of the peculiar motions of the Sun and the Milky Way. The peculiar motions of the galaxies are generally not known. We correct for those of the Sun with respect to the local Universe by computing the Virgo-infall corrected radial velocity, $v_{\rm vir}$, which corrects for all motions of the Sun, and Milky Way up to the level of the infall of the Local Group to the Virgo Cluster. We select all galaxies with $v_{\rm vir} < 14\,000\ \rm km\,s^{-1}$ (corresponding to $z < 0.047$ and $D \lesssim 200$ Mpc, assuming Hubble parameter $h = 0.678$; Planck Collaboration XIII 2016). The Virgo-infall corrected radial velocity in *HyperLEDA* is outdated (D. Makarov, private communication). Therefore, we compute it for all galaxies (see Appendix A for details on the computation).

204 467 objects are found in *HyperLEDA* with $v_{\rm vir} < 14\,000\ \rm km\,s^{-1}$ while 2 560816 exceeded the velocity limit. However, for the 681 527 galaxies in *HyperLEDA* without radial velocity measurements, we attempted to get measurements from *NED*. The association to *NED* is described in Appendix B1. In total, we recover the radial velocities for 1494 additional objects with $v_{\rm vir} < 14\,000\ \rm km\,s^{-1}$.

Note that in the above procedures, we performed various manual inspections to exclude duplicates in *HyperLEDA* or misclassified objects (stars, artefacts from diffraction light, ‘parts of galaxies’, etc.) In total 1228 objects were rejected in this process. The final sample consists of 204 733 galaxies. Fig. 1 shows a sky map of the *HECATE*.

Out of the 204 733 galaxies in our sample, there are 39 251 objects without size information, restricting the cross-matching capabilities of our sample. For the majority of them, the semimajor axis is complemented via cross-linking of our sample with other data bases and surveys, resulting in 199 895 galaxies with size information (97.6 per cent). The procedure is described in detail in Appendix B2.

Finally, in the *HECATE* we include additional information from *HyperLEDA* such as astrometric precision, object name, morphological classification, optical photometry, inclination, and Galactic absorption. A full list of the information provided in the *HECATE* is given in Appendix D.

3 DISTANCE ESTIMATES

Robust distance estimates for the galaxies in the *HECATE* are essential for the purposes of this catalogue, and required for estimating the stellar population parameters of the galaxies.

While redshift-derived distances can be calculated using the Hubble–Lemaître law for the majority of the galaxies in the *HECATE* (positive z), this approach is not accurate in the case of nearby galaxies for which recessional velocities are dominated by their peculiar motions. In addition, this method cannot be used for blueshifted galaxies.[2] Furthermore, at the distance range of the *HECATE*, the unknown peculiar motion of a given galaxy adds to the uncertainty on its distance, equally or more than the propagation of the uncertainties of the galaxy's z and the Hubble parameter. For this reason, we use z-independent distance measurements from *NED-D* where available (for ≈10 per cent of the galaxy sample), and combine them with the method described in Section 3.1. For the remainder of the galaxies (≈90 per cent), we estimate the distance of the galaxies using a regression method, described in Section 3.2, based on the sample of galaxies with known distances.

3.1 Redshift-independent distances

The largest resource of z-independent distances is the *NED-D* compilation, containing 326 850 measurements (as of 2020 March) for 183 062 objects, based on 96 different distance

[1] *Hekátē* (greek, Εκάτη), goddess of crossroads and witchcraft in ancient Greek mythology. Pronunciation: *hek-UH-tee*.

[2] In fact, the most blueshifted galaxy in our sample with a reliable distance measurement is a Virgo Cluster member, VCC 0815, at distance of 19.8 Mpc, which corresponds to a recession velocity of $\sim 1400\ \rm km\,s^{-1}$ but its heliocentric radial velocity is $-700 \pm 50\ \rm km\,s^{-1}$.

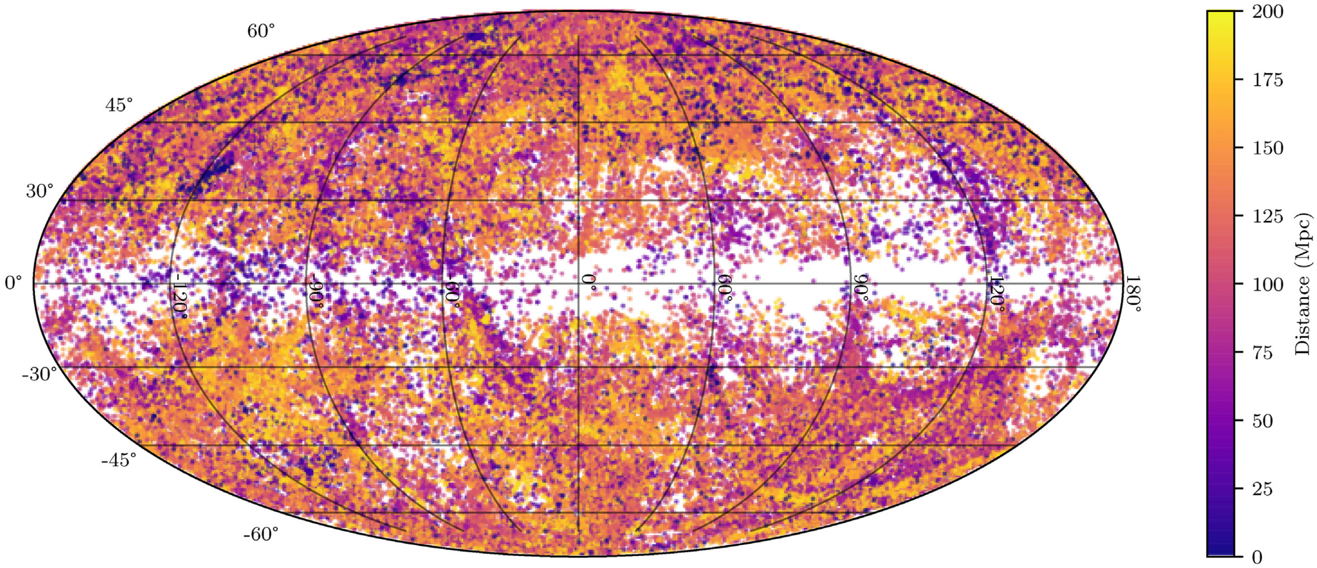


Figure 1. Sky map of the galaxies in the *HECATE* in Galactic coordinates, colour coded according to their distance.

indicators (Steer et al. 2017). However, for objects with multiple measurements, *NED-D* does not readily provide a summary of these distance estimates. On the other hand, *CosmicFlows 3.0* (*CF3*; Tully, Courtois & Sorce 2016) reports distance estimates for 17 669 galaxies at $z \lesssim 0.05$, calculated as uncertainty-weighted averages of individual measurements. Aiming at an as-large-as-possible sample of galaxies with distance determinations, we obtain distance measurements from *NED-D* in order to combine them into unique estimates for each galaxy, and use the *CF3* for consistency checks.

We reject measurements that are not based on peer-reviewed sources, and those using outdated distance moduli for the Large Magellanic Cloud (i.e. outside the 18.3–18.7 mag range; Pietrzyński et al. 2013) or distance scales calibrated for Hubble constants outside the range 60–80 km s^{-1} Mpc^{-1}. Many of the 93 distance indicators reported in *NED-D* are appropriate for objects at distances greater than the volume limit of the *HECATE* (e.g. SNIa) and therefore we do not consider them. We also avoid methods applied in less than three publications, as their systematic uncertainties or validity may be insufficiently understood. To be conservative, we select 10 commonly used indicators that are considered relatively reliable at distances <200 Mpc (e.g. Steer et al. 2017), listed in Table 1. For publications reporting for the same galaxy multiple individual measurements based on the same indicator (e.g. Cepheid distances for different stars within a galaxy), we adopt the concluding measurement in each publication. Reported zero distance uncertainties (10 cases) were treated as undefined. Preference is given to measurements with reported uncertainties over those without uncertainties. In total, we associate 43 511 distance measurements with 21 174 galaxies in the *HECATE*.

For the 13 247 galaxies with single distance measurements, we adopt them as they are, 97 per cent of which have reported uncertainties.

For the 7336 galaxies with multiple distance measurements and uncertainties, we calculate the final distances and corresponding uncertainties using a weighted Gaussian Mixture (GM) model. The weights depend on the year of publication (penalising old measurements) to reduce historical biases (e.g. older calibrations and unknown biases) and measurement uncertainties. The weight for the i-th measurement is

$$w_i = \delta^{y_i - y_{\rm ref}} \sigma_i^{-2}, \tag{1}$$

where δ is the penalty per year – we set $\delta = 2^{0.1}$ so that the weight is halved for every decade passed,[3] y_i is the year of measurement, and $y_{\rm ref}$ is an arbitrary reference year. The GM distribution of the distance modulus μ of a galaxy is derived by combining M individual measurements:

$$f_{\rm GM}(\mu) = \sum_{i=1}^{M} w_i f_i(\mu) \Big/ \sum_{i=1}^{M} w_i, \tag{2}$$

Table 1. List of distance indicators used in *HECATE* (see Steer et al. 2017, and references therein), the number of galaxies ($N_{\rm gal}$) for which the measurements (number $N_{\rm meas}$) based on each indicator were considered in the final distance estimates, and the corresponding typical uncertainty of the distance moduli, $\langle\sigma_\mu\rangle$, in mag.

Distance indicator	$N_{\rm gal}$	$N_{\rm meas}$[a]	$\langle\sigma_\mu\rangle$
Cepheids	75	1416	0.11
Eclipsing binary	4	45	0.09
Fundamental plane	10 697	26 975	0.35
Horizontal branch	29	65	0.10
Red clump	14	102	0.09
RR Lyrae	38	282	0.09
Sosies	280	280	0.29
Surface brightness fluctuations	482	1650	0.18
Tip of the red giant branch	358	1361	0.13
Tully–Fisher	10 780	11 309	0.40

[a]We note that $N_{\rm meas} \geq N_{\rm gal}$ because for many galaxies there are multiple distance measurements based on the same indicator.

[3]We chose this value because: (i) we found systematic offsets (0.05–0.2 mag) in the distance moduli measured at times with differences >20 yr, and (ii) for small values of δ (corresponding to $^1/_2$-folding time-scales of less than 5 yr), we found an increased scatter (>0.1 mag) because, effectively, only few newer measurements contribute to the distance.

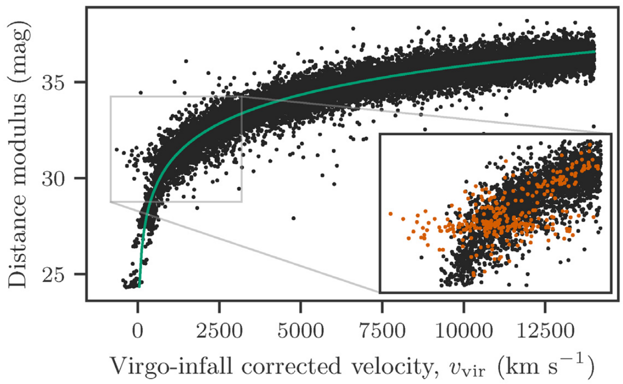


Figure 2. Hubble diagram of the galaxies in the *HECATE* with z-independent distances. For reference the green line shows the Hubble–Lemaître law with H_0=(67.8±0.1) km s^{-1} Mpc^{-1} (Planck Collaboration XIII 2016). We note that the majority of the points are following the law, albeit with significant dispersion at low values of $v_{\rm vir}$, and the existence of a 'branch' at a distance modulus of ∼31 (see the inset) caused by Virgo Cluster galaxies (orange; see Section 3.2.1) that present significant velocity dispersion (cf. fig. 10 in Tully et al. 2016).

where w_i are the weights calculated in equation (1), and f_i is the PDF of the distribution of each measurement. We consider each measurement to be Gaussian distributed, with mean and standard deviation equal to the distance modulus and its uncertainty reported in *NED-D*. We note that the mean of the distribution resulting from equation (2) is mathematically equivalent to the weighted average of the individual means (and therefore consistent with the methods for galaxies with single measurement, or without uncertainties [see below]), while its spread accounts for both the dispersion of the measurements and their uncertainties.

For the 591 galaxies with multiple measurements but no uncertainties, we use their weighted mean as the final estimate, and their weighted standard deviation as the uncertainty. In these cases, the weights are

$$w_i = \delta^{y_i - y_{\rm ref}}, \tag{3}$$

where the parameters are the same as in equation (1).

We note that for seven galaxies out of these 591, the standard deviation was 0 (possibly duplicate measurements), and therefore we do not report the uncertainty of the final distance estimate.

3.2 Redshift-dependent distances

For galaxies without distance measurements (≃90 per cent), we rely on the spectroscopic redshift information. While we could simply use the Hubble–Lemaître law and the redshift of each object in order to calculate their distances, the proximity of the galaxies in the *HECATE* sample makes them very sensitive to peculiar velocities and local deviations from the Hubble flow. For this reason, we adopt a data-driven approach where the galaxies with z-independent distances (10 per cent of the full sample) are used as the training data set in a regression model that infers the distances (and uncertainties) at similar recession velocities for the rest of the sample. The uncertainties of the radial velocities were not accounted for in the regression since, in the case of spectroscopic redshifts, they are negligible compared to the uncertainties of the distance measurements in the training sample.

Fig. 2 shows the distance modulus as a function of the radial velocity for the galaxies with z-independent distances in our sample (calculated as described in Section 3.1). We observe (i) that – unsurprisingly – the distance correlates with radial velocity even for nearby and blueshifted galaxies, albeit with higher dispersion, and (ii) a horizontal branch at a distance modulus of ≈31 mag that is caused by Virgo Cluster galaxies (see the inset of Fig. 2). In order to account for such local deviations from the Hubble flow, we employ a data-driven approach for robust distance estimates as follows:

(i) the galaxy sample is separated into two subsamples: galaxies in the Virgo Cluster and the rest (Section 3.2.1);

(ii) for each subsample, a regressor is trained using the galaxies with redshifts and z-independent distances, so that the distance and its uncertainty are predicted from the recession velocity;

(iii) the distance and its uncertainty for the galaxies without z-independent measurements are predicted using the two regressors.

3.2.1 *Virgo-cluster membership*

As we discussed in the previous paragraphs, and shown in Fig. 2, special treatment of Virgo Cluster members is necessary for estimating their distance from the recession velocity. The most up-to-date catalogue of galaxies of the Virgo Cluster, the *Extended Virgo Cluster Catalog (EVCC)*, was produced by Kim et al. (2014) using the radial velocities and a cluster infall model, as well as morphological and spectroscopic classification schemes. The EVCC is cross-matched with our sample to identify the galaxies associated with the Virgo cluster.

3.2.2 *Local average and standard deviation of Hubble diagram*

We use the Kernel Regression technique (Nadaraya 1964) to compute the intrinsic distance modulus $\mu_{\rm int}(v_{\rm vir})$ at a given Virgo-infall corrected radial velocity, $u \equiv v_{\rm vir}$. The sample is split into Virgo members (VC) and non-Virgo members (nVC). For each subsample, we compute the local (at u) distance modulus, $\mu_{\rm int}(u)$, as the weighted average of the distance moduli μ_i of the N galaxies it contains, with weights $[q_i(u)]$ given by the Gaussian kernel with bandwidth h (or 'averaging length'). Similarly, for each subsample (VC and nVC) we calculate the 'local standard deviation' in terms of the bias-corrected weighted standard deviation:

$$\sigma_i(v) = \sqrt{\frac{V_1}{V_1^2 - V_2} \sum_{i=1}^{N} q_i(u)\,[\mu_i - \mu_{\rm int}(u)]^2}, \tag{4}$$

where $V_1 = \sum_{i=1}^{N} q_i(u)$ and $V_2 = \sum_{i=1}^{N} q_i^2(u)$.

The choice of the bandwidth h effectively sets the resolution, in radial velocity, of the derived statistical quantities. Due to the significant curvature of the radial velocity versus distance modulus diagram (Fig. 2) for $u \lesssim 2000$ km s^{-1}, the resolution should be increased in this region in order to capture the shape and prevent mixing of data from regions of significantly different slopes. On the other hand, at greater radial velocities (or distances) a relatively large bandwidth would allow more data points to contribute, and hence provide an estimate that is less influenced by outliers. For these reasons, we set the bandwidth h for the Gaussian Kernel to be a function of the radial velocity, increasing with radial velocity but also kept constant in the steep part of the diagram by enforcing a minimum value, $h_{\rm min}$:

$$h(u) = h_{\rm min} \times \max\left\{1, u/2000\,{\rm km\,s^{-1}}\right\}, \tag{5}$$

where $h(u)$ is the bandwidth of the Gaussian Kernel for points evaluated at Virgo-infall corrected radial velocity u. For the Virgo Cluster model, we keep the bandwidth fixed as the distance modulus is expected to be roughly the same, rendering such considerations irrelevant.

The baseline of the bandwidth, $h_{\rm min}$, must be chosen carefully as it can easily result in 'overfitting' if too small (only a few of the data points are considered for the fit in each bin), or 'underfitting' if too large (introducing *lack-of-fit* variance). We find the optimal bandwidth, by minimizing the total regression error $\mathcal{S}$, i.e. the quadratic sum of the regression errors, $\mathcal{S}_i$, corresponding to each galaxy. $\mathcal{S}_i$ is evaluated by employing the leave-one-out cross-validation technique: The i-th galaxy is removed from the sample, and its distance is estimated using the kernel regression. The residual between the true distance modulus and the regressed one is $\mathcal{S}_i$. Additionally, when the optimal bandwidth has been found, we remove outliers based on the true distance modulus and the predicted one (and its uncertainty), by performing sigma clipping at the 3σ level, and re-optimize for $h_{\rm min}$ iteratively until no outlier is found. We applied the above procedure and found optimal minimum bandwidth $h_{\rm min} = 68.2\,{\rm km\,s^{-1}}$ for non-Virgo galaxies, after removing 149 outliers (<1 per cent of the nVC subsample). For the Virgo galaxies, the optimal bandwidth was $h_{\rm VC} = 294.5\,{\rm km\,s^{-1}}$, while only one outlier was found (<1 per cent of the VC subsample).

3.2.3 Local intrinsic dispersion

The local standard deviation we compute in equation (4) encompasses the uncertainties of the distance moduli due to measurement uncertainties and the *intrinsic scatter* of the true distance modulus. The latter is attributed to the peculiar velocities of the galaxies and the systematic uncertainties due to the distance ladder calibration. Given the model $\mu_{\rm int}(v)$ and following Kelly (2007), we formulate the error model

$$\mu_i = \mu_{\rm int}(v_i) + \epsilon_i + \epsilon_{\rm int}(v_i), \tag{6}$$

where ϵ_i is a Gaussian-distributed random variate with mean equal to 0 and standard deviation equal to the distance modulus uncertainty of the i-th galaxy, $\mu_{\rm int}$ is the local average, and $\epsilon_{\rm int}(v_i)$ is a Gaussian-distributed variate with mean equal to 0 and standard deviation $\sigma_{\rm int}(v)$, which is a *local intrinsic scatter* model. We apply a Maximum Likelihood Estimator (MLE) to calculate the local intrinsic scatter, $\epsilon_i(v)$. We note that the uncertainties on radial velocities have not been considered in our analysis as they are typically one order of magnitude smaller ($\sim 10\,{\rm km\,s^{-1}}$) than the optimal kernel bandwidth for both VC and nVC models ($\sim 100\,{\rm km\,s^{-1}}$) and the typical peculiar velocities of galaxies ($\sim 100\,{\rm km\,s^{-1}}$; e.g. Hawkins et al. 2003).

We apply the above Kernel Regression model to 617 Virgo galaxies and 182 326 nVC galaxies to derive their distances and uncertainties. Also, for 37 Virgo galaxies and 317 nVC galaxies with z-independent distances but no uncertainties, we apply the local intrinsic scatter model to estimate their uncertainty. We ensure that the two models are applied only to galaxies with radial velocities covered by the training data sets: $v_{\rm vir} \in \left[-792\,{\rm km\,s^{-1}}, 2764\,{\rm km\,s^{-1}}\right]$ for VC and $v_{\rm vir} \in \left[-481\,{\rm km\,s^{-1}}, 14\,033\,{\rm km\,s^{-1}}\right]$ for nVC in order to avoid extrapolation (note that the ranges are expanded by half optimal bandwidth, $h_{\rm min}$), leaving only 12 objects in the *HECATE* without distance estimates.

For quick reference, in Appendix C we provide empirical formulae for the distance modulus of a galaxy $\mu_{\rm int}$, and its uncertainty $\epsilon_{\rm int}$, given its Virgo-infall corrected velocity, based on the results of the aforementioned methods.

3.2.4 Validation of the regression technique

The resulting distances from the Kernel Regression technique described earlier should reflect the trends of the z-independent distances used in the *HECATE*, and converge to the Hubble–Lemaître law for large distances.

The top panel of Fig. 3 shows the distance moduli as a function of the radial velocity of the two subsamples: 273 Virgo Cluster galaxies and 15 294 non-Virgo galaxies. We see the local average and the 2σ confidence intervals in terms of local intrinsic scatter. The latter is shown independently in the middle panel of Fig. 3, where we observe that the accuracy of the non-VC model drops significantly for $v \lesssim 1500\,{\rm km\,s^{-1}}$ as expected from the domination of the peculiar velocities over the Hubble-flow component. Conversely, the VC model presents a slight increase in the distance with increasing radial velocity, which is possibly due to the contamination from background galaxies in the EVCC. For the same reason, the uncertainty of the inferred distances at high radial velocities for VC members is higher than that at the low radial velocities.

The convergence to the Hubble–Lemaître law is shown in the bottom panel of Fig. 3, where we plot the ratio of the local average model ($v_{\rm vir}$; see Section 3.2.2) to the Hubble-flow distance, for $v_{\rm vir} \in [200, 14\,000]\,{\rm km\,s^{-1}}$. Two different values of H_0 are considered: $72\,{\rm km\,s^{-1}\,Mpc^{-1}}$ (*HST* Key Project; Freedman et al. 2001) and $67.8\,{\rm km\,s^{-1}\,Mpc^{-1}}$ (Planck Collaboration XIII 2016). We see that the z-independent distances converge to the Hubble-flow distances, and agree to the local Universe estimate of the Hubble constant (*HST* Key Project).

Finally, we check the distance estimates in the *HECATE* against *CF3*. For 1949 galaxies with z-dependent distances in the *HECATE*, but z-independent in *CF3*, we find agreement within the expected scatter from the regression model for 99 per cent of the galaxies in common. Furthermore, the mean and median difference is 0.009 mag, and the scatter in the difference between the distance moduli from *CF3* and the one calculated in the regression model is 0.25 mag.

4 MULTIWAVELENGTH DATA AND STELLAR POPULATION PARAMETERS

One of the main objectives for the compilation of the *HECATE* is to provide stellar population parameters for galaxies in the local Universe. To do so, we obtain photometric and spectroscopic data by cross-correlating the *HECATE* with surveys from the infrared (IR) to the optical. In Section 4.1, we evaluate the required data to attain the most reliable galaxy properties and discuss the selection and cross-matching criteria for each survey. In Section 4.2, we describe the methodology we use for deriving the parameters from the associated multiwavelength data.

4.1 Associated photometric and spectroscopic data

SFR estimates can be obtained by photometric data from IR to UV bands (or combinations of them; for a review, see Kennicutt & Evans 2012). While optical- and UV-based SFR indicators are sensitive to dust absorption, IR indicators overcome this limitation by measuring the dust-reprocessed stellar emission. Although UV+IR composite SFR indicators (e.g. Hao et al. 2011) are now becoming more widely used (especially in the case of dwarf metal-poor galaxies), their implementation relies on the availability of integrated UV photometry. The all-sky *GALEX* UV survey does not provide integrated photometry for large, nearby galaxies, hampering the use of these SFR indicators. Therefore, we rely on mid- and far-

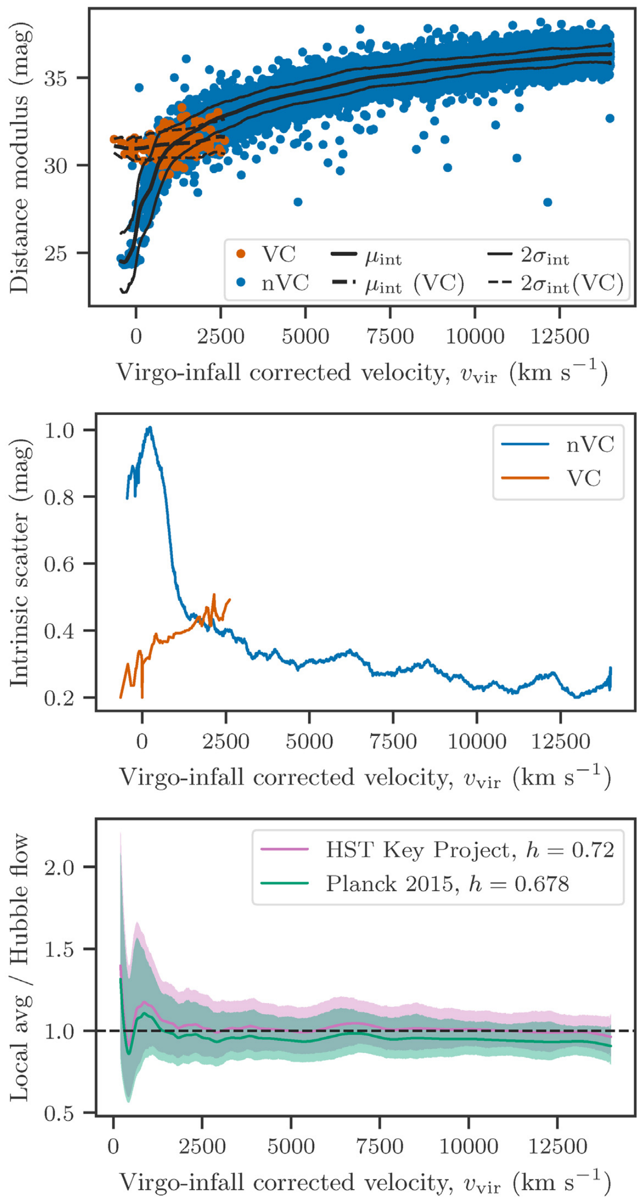


Figure 3. Assessment of the accuracy of the Kernel Regression models. The models capture the trends in the D–z in the local Universe, and provide accurate distances (0.2–0.4 mag or 10–20 per cent) particularly for $v_{\rm vir} \gtrsim 2500$ ($D \gtrsim 35$ Mpc). Top: z-independent distances in our sample (points), separated to Virgo Cluster members (orange) and nVC galaxies (blue). The black lines depict the local mean and the 2σ regions (using the local standard deviation) according to the two regressors (dashed for VC and continuous for nVC). Middle: The local standard deviation of the distance modulus of the two models. Bottom: The ratio of the local average distance (Section 3.2.2) and the distance inferred from Hubble–Lemaître law for two values of Hubble parameter: 0.72 (*HST* Key Project; magenta line) and 0.678 (*Planck* 2015; green line). For each ratio, we plot with the same colours the 68 per cent confidence region that reflects the local intrinsic scatter (Section 3.2.3).

IR indicators using *IRAS* and *WISE* photometry (see Sections 4.1.1 and 4.1.2), aiming at a homogeneous and as-complete-as-possible compilation of SFR estimates.

For the computation of the galaxy stellar masses, one of the most reliable photometric indicators is the K_s-band luminosity (e.g. Gardner et al. 1997). In order to account for the stellar-population age dependence of the mass-to-light ratio (M/L), we use calibrations that incorporate optical colours (Bell et al. 2003). For this reason, we obtain *2MASS* and *SDSS* photometry, as described in Sections 4.1.3 and 4.1.4.

Spectroscopic data can be used to estimate the metallicity of the galaxies, as well as characterize them on the basis of their nuclear activity. In Section 4.1.4, we describe the acquisition of spectroscopic data from *SDSS*.

4.1.1 *Far-infrared:* IRAS

We cross-link the *HECATE* galaxies to *IRAS* objects. For the cross-correlation with the *IRAS* catalogue, we adopt the following approach. When a galaxy is included in the *IRAS Revised Bright Galaxy Sample* (*IRAS-RBGS*), we adopt this photometric information, which is more reliable for extended galaxies (Sanders et al. 2003). In total, we associate 589 galaxies with the *IRAS-RBGS* catalogue (Appendix B3). For the remaining galaxies, we use the *Revised IRAS-FSC Redshift Catalogue* (*RIFSCz*; Wang et al. 2014) that provides a clean (excluding poor-quality and cirrus sources) sample of *IRAS* galaxies at 60 µm. This also gives more reliable positions than its parent *IRAS Faint Source Catalog* (*IRAS-FSC*; Moshir et al. 1990) via the association to more recent surveys. In total, we associate 19 082 galaxies in our sample with *RIFSCz* (Appendix B4).

4.1.2 *Mid-infrared:* WISE

The previous cross-matches with *IRAS-RBGS* and *RIFSCz* objects incorporate *IRAS* photometry for 19 671 objects in the *HECATE* (9.6 per cent). To obtain a more complete census of the IR emission of the galaxies in the *HECATE* sample, we could also use the deeper all-sky surveys (e.g. *WISE* and *AKARI*). However, at the time of compilation of the *HECATE*, there are no extended source catalogues of *WISE* and *AKARI* that can provide reliable flux measurements for nearby galaxies. For this reason, we use the 'forced photometry' catalogue by Lang, Hogg & Schlegel (2016), who extracted fluxes from *unWISE* images (Lang 2014) for *SDSS-DR10* photometric objects using the *SDSS* apertures. We cross-correlate this catalogue with the *HECATE* by matching the *SDSS* ID, which is already specified in the *HECATE* (Section 4.1.4). As the *WISE* forced photometry catalogue is organized in *unWISE* tiles, there are galaxies in overlapping regions. For these cases, we select the data from the tile in which the galaxy is closer to the tile's centre. 123 699 *HECATE* galaxies to objects are linked to *SDSS* objects with *WISE* forced photometry. We note, however, that the use of this catalogue restricts our *WISE* photometric data to the *SDSS* footprint. *WISE* photometry is available for the wider *HECATE* sample, but as mentioned earlier it is not reliable for the resolved galaxies.

4.1.3 *Near-infrared:* 2MASS

To incorporate *2MASS* photometry in our sample, we cross-match the *HECATE* with three *2MASS* catalogues in the following order of priority: (i) *Large Galaxy Atlas* (*2MASS-LGA*; Jarrett et al. 2003), (ii) *Extended Source Catalog* (*2MASS-XSC*; Skrutskie et al. 2006), and (iii) *Point Source Catalog* (*2MASS-PSC*; Cutri et al. 2012). This order ensures that for the resolved galaxies we use the most reliable measurements of their flux. Specifically, from *2MASS-LGA* and *2MASS-XSC* we obtain the *JHK* 'total' magnitudes from the extrapolated surface brightness profiles (see *2MASS-LGA* home page[4] and section 4.5.a.iv in the Explanatory Supplement[5]). From the *2MASS-PSC*, we obtain the 'default' magnitudes. We note that when no uncertainty is provided, the listed magnitudes are upper limits. We link *HECATE* galaxies to 609 objects in the *2MASS-LGA*, 117 713 in the *2MASS-XSC*, and 25 224 in the *2MASS-PSC*, overall providing *2MASS* photometry for 143 546 galaxies. More details about the cross-matching procedure can be found in Appendix B5.

4.1.4 *Optical:* SDSS

For the cross-matching of the *HECATE* and the *SDSS*, we use the DR12 photometric catalogue, and select only primary[6] objects. We use a match radius of 3 arcsec around the *HECATE* coordinates, and we select the closest match (typical separation of the matched objects is ∼0.2 arcsec), resulting in 123 711 matches.

We opt to use spectroscopic data from the *MPA–JHU DR8* catalogue (Kauffmann et al. 2003; Brinchmann et al. 2004; Tremonti et al. 2004), which are based on the emission-line component of the spectrum after subtracting the underlying stellar component, to estimate the metallicities and classify the galaxies on the basis of their nuclear activity. By cross-matching the catalogue with the *HECATE*, we obtain measurements for 93 714 out of the 123 711 *SDSS* objects in the *HECATE*.

The *GALEX–SDSS–WISE Legacy Catalog 2* (*GSWLC-2*) of Salim et al. (2016) and Salim, Boquien & Lee (2018) provides SFR and $M_\star$ estimates through optical–UV spectral energy distribution (SED) fits to galaxies within the *SDSS* footprint and distance >50 Mpc. By matching 75 672 *HECATE* galaxies to *GSWLC-2* objects on the basis of their object IDs in *SDSS*, we obtain additional SFR and $M_\star$ estimates.

4.2 Derived parameters

The following paragraphs describe the methods employed for the estimation of parameters from the acquired multiwavelength data (Section 4.1). An overview of the provided data is listed in Table D1.

4.2.1 *Stellar masses*

The stellar masses are estimated by combining the K_s-band luminosities of the galaxies with the appropriate mass-to-light ratios. The integrated K_s-band luminosities of the galaxies are calculated from their *2MASS* photometry and distances (we adopted 3.29 mag for the absolute magnitude of the Sun; Blanton & Roweis 2007). We exclude objects without uncertainties in their photometry, or uncertainty higher than 0.3 mag, resulting in L_K measurements for 133 017 (65 per cent) galaxies in the *HECATE*. The K_s-band *M*/*L* ratio ($\equiv M_\star/L_{K_s}$) is computed using the calibration of Bell et al. (2003) that accounts for differences in the stellar populations by means of the $g-r$ colour of the galaxies:

$$\log(M/L) = -0.209 + 0.197\,(g-r)\,. \tag{7}$$

$g-r$ colours are available for 53 171 (26 per cent) galaxies with reliable photometry (*SDSS* flags q_mode='+' and Q=3, and uncertainties < 0.1 mag on g and r). The mean *M*/*L* ratio of the galaxies in the *HECATE* is 0.822, while the scatter is 0.091. This mean value is used for the 79 846 (39 per cent) galaxies without *SDSS* photometry. The scatter gives us an estimation of the *M*/*L* ratio variations due to the different $g-r$ colours of the galaxies, and it can be used to assess the uncertainty on the $M_\star$ of galaxies without *SDSS* photometry. For the remainder (35 per cent) of the *HECATE* sample that does not have K_s-band measurements in *2MASS*, we do not estimate $M_\star$.

The *GSWLC-2* provides $M_\star$ derived using a different method (SED fitting using UV to IR data; Salim et al. 2016, 2018). In Fig. 4, we compare these estimates with our derived $M_\star$ using near-IR photometry. We find very good agreement (scatter of 0.21 dex), although SED-based $M_\star$ are slightly lower on average (factor of −0.11 dex), possibly due to assumptions of stellar population models, or star formation histories.

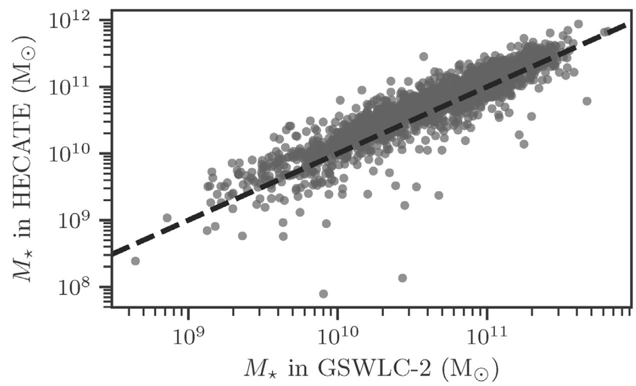


Figure 4. Comparison of the stellar mass estimates in the *HECATE* and the *GSWLC-2* for the common galaxies. The galaxies closely follow the 1:1 line, indicated by the black, dashed line, with an intrinsic scatter of 0.21 dex.

4.2.2 *Star formation rates*

The SFR estimates of the *HECATE* galaxies are based on measurements of IR luminosity from the *IRAS* or *WISE* surveys. These surveys provide the optimal combination of reliable, well-calibrated, SFR indicators (Kennicutt & Evans 2012), sensitivity, and sky coverage. Since the sensitivity of the two surveys varies depending on the band, we use a combination of SFR indicators depending on the availability of reliable measurements. For *IRAS*, we use the total-IR (TIR), far-IR (FIR), and 60 µm luminosities, depending on the bands with reliable *IRAS* fluxes ('FQUAL'≥2, i.e. excluding upper limits). For *WISE*, we use the monochromatic Band-3 (W3; 12 µm) and Band-4 (W4; 22 µm) fluxes from the *WISE* forced photometry catalogue as discussed in Section 4.1.2. *WISE* SFR estimates are not provided for objects with uncertainties greater than 0.3 mag, or those that were considered as point sources in the analysis of Lang et al. (2016).[7]

For completeness, we provide in our catalogue SFR measurements based on all indicators (Table 2) available for each galaxy (including the SED-based SFR from the *GSWLC-2* for convenience to users focusing on the *SDSS* footprint). This is particularly important since the various surveys used to derive the SFRs cover different subsets of the *HECATE*. In order to have consistent SFRs for the largest possible set of objects, and given the fact that different indicators

[4] https://irsa.ipac.caltech.edu/applications/2MASS/LGA/intro.html
[5] https://old.ipac.caltech.edu/2mass/releases/allsky/doc/explsup.html
[6] www.sdss.org/dr12/help/glossary/#surveyprimary

[7] Using the galaxies with *WISE* and *IRAS* photometry, we found that the *WISE* SFRs are significantly lower for ∼2500 sources with the flag 'treated_as_pointsource' set in the catalogue of Lang et al. (2016), 20 per cent of which have SFR estimates from the *IRAS* data.

Table 2. The five different SFR indicators used in the *HECATE*. The TIR and *WISE* indicators assume a constant star formation history (SFH), *Starburst99* stellar population models (Leitherer et al. 1999), and the Kroupa (2001) initial mass function (IMF). On the other hand, the FIR indicator assumes starbursts with time-scales of 10–100 Myr, Leitherer & Heckman (1995) stellar models, and the (Salpeter 1955) IMF, whereas the 60 μm indicator relies on a universal SFH (Rowan-Robinson 1999), the Bruzual A. & Charlot (1993) stellar models, and the Salpeter (1955) IMF. The offsets between the TIR and the other *IRAS* indicators can be explained on the basis of the different assumptions.

Survey/bands (1)	N_{gal} (2)	L calibration (3)	SFR scaling relation (4)	Scaling in SFR_{HEC} (5)	Flag; number in SFR_{HEC} (6)
IRAS 25, 60, 100 μm (TIR)	5125	Dale & Helou (2002)	Kennicutt & Evans (2012)	(Reference)	RT/FT[a]; 5125
IRAS 60, 100 μm (FIR)	5721	Helou & Walker (1988)	Kennicutt (1998)	1.15 (0.08 dex)	FF; 596
IRAS 60 μm	19 671	Rowan-Robinson (1999)		1.27 (0.10 dex)	R3/F3[a]; 13 950
WISE 12 μm	81 948	Cluver et al. (2017)		0.97 (−0.01 dex)	W3; 72 726
WISE 22 μm	46 078	Cluver et al. (2017)		0.95 (−0.02 dex)	W4; 1872

Notes. Description of columns: (1) the survey (*IRAS* or *WISE*), and the photometric bands used for the computation of the flux; (2) the number of galaxies for which the SFR indicator is computed; (3), (4) references to the definition of the composite band and SFR scaling calibration; (5) The scaling factor used only for the 'homogenization' of the SFR indicator with respect to the TIR indicator. In parenthesis we give the scatter of the 'homogenization' relation; (6) the flag (in the column 'logSFR_flag'; see Table D1) and the number of galaxies for which the homogenized SFR is based on this indicator.
[a]The first letter indicates whether the photometry was taken from the *IRAS-RBGS* (R) or *RIFSCz* (F).

often result in systematic offsets in the derived SFRs, we also provide a homogenized SFR (SFR_{HEC}), calculated as follows.

First, we account for offsets between the different SFR indicators (Table 2) by calculating their ratio with respect to the TIR-based SFR that we consider as a reference since it probes star formation regimes at different time-scales (Kennicutt & Evans 2012). The mean ratio for each indicator is adopted as the correction factor. Fig. 5 shows comparisons between the SFR indicators, also giving the scaling factor, standard deviation, and the number of galaxies used in each comparison. The homogenized SFR of an object in the *HECATE* is the TIR-SFR if available; otherwise, we use, in order of preference, the rescaled FIR-SFR and 60, 12, and 22 μm-based SFR. Although the 22 μm band (probing hot dust associated with young star-forming regions) is a better-calibrated SFR than the 12 μm-based one (probing emission from polycyclic aromatic hydrocarbons; e.g. Parkash et al. 2018), preference is given to the latter due to the higher quality of the W3 *WISE* data (e.g. Cluver et al. 2017). We note that no rescaling is performed in the individual SFR indicator columns.

The TIR luminosity includes emission in the 100 μm band, which in the case of galaxies with low specific SFR may have a non-negligible contribution from stochastically heated dust from older stellar populations (e.g. Galliano, Galametz & Jones 2018). Although this may overestimate the SFR in early-type galaxies, it is a widely used and well-understood SFR indicator that gives reliable SFR for actively star-forming galaxies, which are the majority of the *HECATE* (fig. 2 in Kovlakas et al. 2020). We note that the catalogue provides all SFR indicators (Appendix D) before rescaling, and the homogenized SFR, where indicators were rescaled according to the procedure described above. A flag is provided, denoting which indicator was used in the homogenized SFR (cf. Table 2). Therefore, based on the scaling factors reported in the table, one can translate the provided SFR to the reference indicator of their choice. Due to the selection of the TIR as a reference, the homogenized SFRs are consistent with the Kroupa (2001) IMF, constant SFH, and Leitherer et al. (1999) stellar population models.

In Fig. 6, we compare the homogenized SFRs against the SED-based SFRs from the *GSWLC-2*. We see that at SFR$\gtrsim 0.1\,M_\odot\,yr^{-1}$ (typical for star-forming galaxies), the *HECATE* provides SFRs that scale with, but are a factor of ≃2.3 larger than those of *GSWLC-2*. This could be because the SEDs used in the *GSWLC-2* do not include IR emission above 22 μm, therefore missing the dominant, relatively cold, dust component associated with star-forming activity (probed in the ∼60 μm band). Furthermore, differences in the IMFs (only 0.02 dex in this comparison), stellar population models, and SFHs might produce additional offsets (see discussion in Salim et al. 2016). As discussed above, the IR-based SFRs may overestimate the SFR in low-specific SFR galaxies, which can explain the flattening observed at low SFRs, and the difference between early-type and late-type galaxies. This is demonstrated in the right-hand panel of Fig. 6 where the $g-r$ colour is used to indicate the contribution of the older stellar populations. Galaxies with redder colours (and hence higher $g-r$) tend to have an excess of IR-based SFRs with respect to SED-based SFRs. We see qualitatively that for $g-r>0.65$, the discrepancy between the two becomes fairly significant.

4.2.3 Metallicity estimates

To measure the gas-phase metallicities for our sample, we use the optical emission-line fluxes provided in the *MPA–JHU DR8* value-added 'galSpecLine' catalogue (see details for methods in Brinchmann et al. 2004) based on the *SDSS*-DR8 data. Relevant especially for measuring accurate nebular emission lines, this catalogue applies stellar-population synthesis models to accurately fit and subtract the stellar continuum, including stellar absorption features. We calculate the gas-phase metallicities, $12+\log(O/H)$, using the Pettini & Pagel (2004) O3N2 (henceforth, PP04 O3N2) prescription, which has been shown by Kewley & Ellison (2008) to be robust (i.e. it can trace a wide range of metallicities, it has relatively low scatter, and most importantly it is less sensitive to extinction effects than other indicators). Our metallicity analysis is subject to the quality of the [O III], [N II], Hβ, or Hα emission lines and the PP04 O3N2 relation limitations. Therefore, we set the following flags (see column 'flag_metal' in Appendix D) to mark uncertain results: Sources with '1' have O3N2 > 2 ratios (670 sources), where the PP04 O3N2 relationship is invalid, and therefore the extrapolated metallicities are highly uncertain; '2' marks emission lines with low signal-to-noise ($\sigma<3$; 882 sources). Therefore, only sources with flags set to '0' have reliable metallicity measurements (62 728 sources). Objects without metallicity estimates are flagged with '−1' (140 453 sources).

4.2.4 Nuclear activity

Using the *SDSS-DR8* emission-line data (see Sections 4.1.4 and 4.2.3), we identify AGN based on the location of the galaxies in

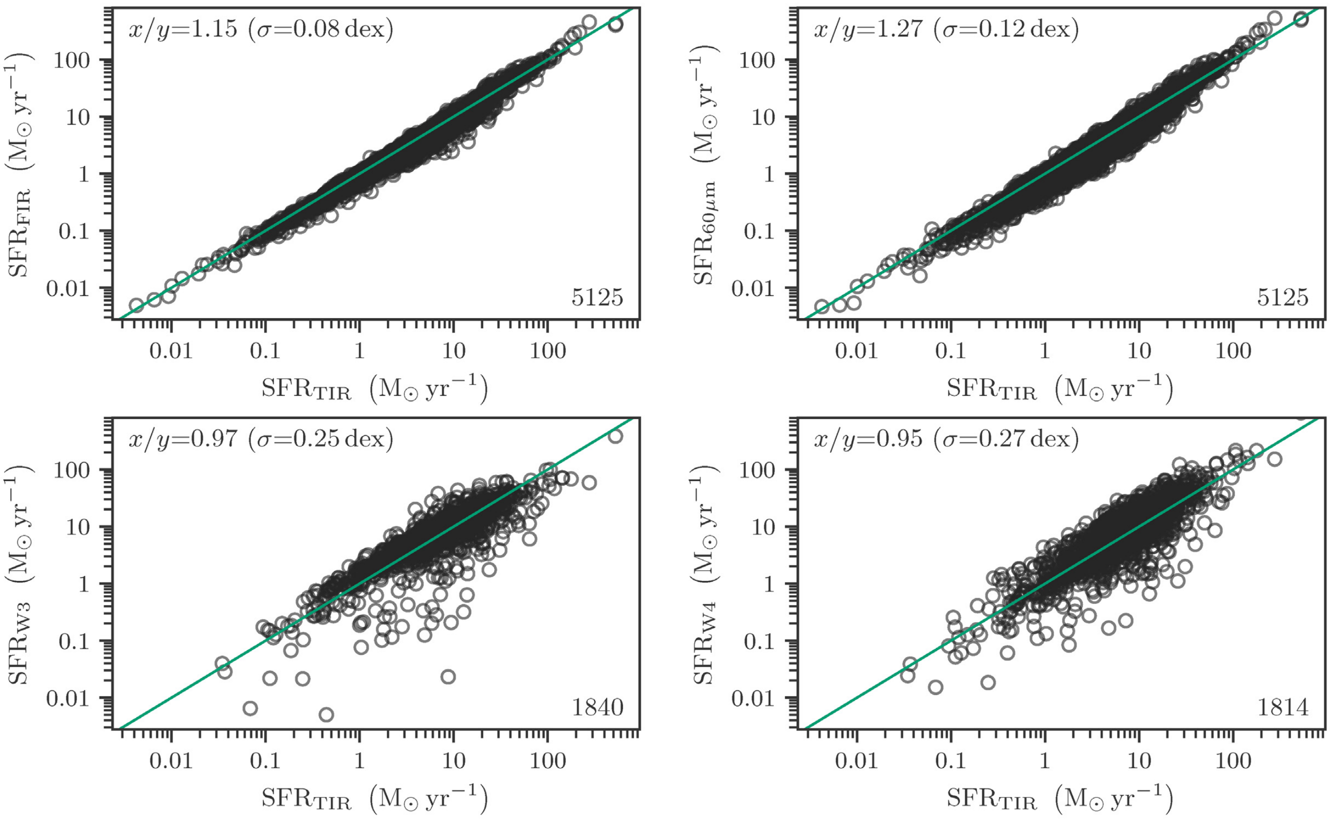


Figure 5. Comparison between the TIR-SFR indicator and the other four indicators used in the *HECATE* (not rescaled as in the computation of $\mathrm{SFR_{HEC}}$); the 1:1 line is shown as a green line. For each SFR indicator, the linear scaling factor and the scatter are reported in the top left corner, while the number of overlapping galaxies used for the scaling is reported in the bottom right corner. The four SFR indicators scale with the reference indicator (TIR) well, and present intrinsic scatter up to 0.27 dex (typical of photometric SFR estimates; Kennicutt & Evans 2012). The scaling factors are used for the computation of the homogenized SFR column in the *HECATE*.

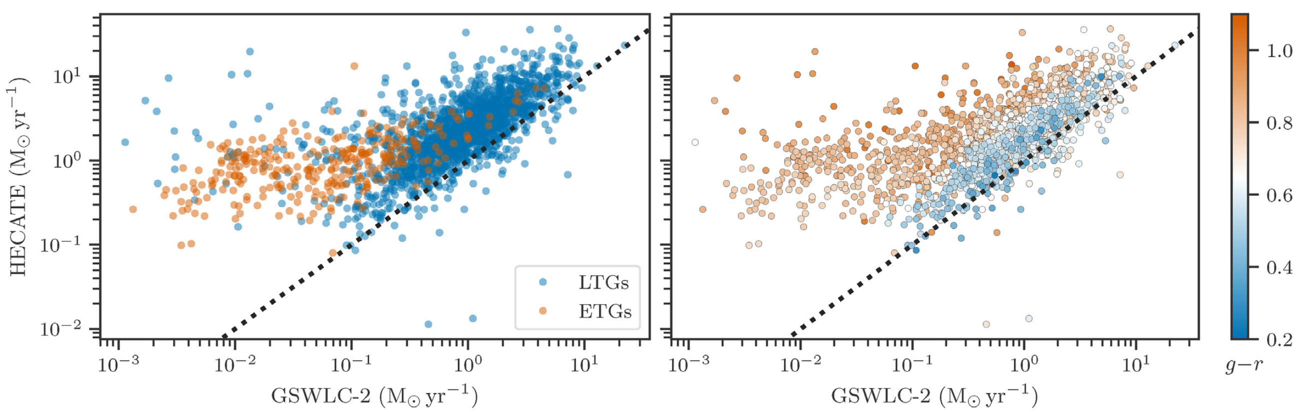


Figure 6. (a) Comparison between the homogenized SFRs and the SED-based SFRs from the *GSWLC-2* for early-type (orange) and late-type (blue) galaxies. In the bulk of the sample, mainly consisting of late-type galaxies, the *GSWLC-2* underestimates the SFR due to the lack of the dust component associated with the star-forming activity, whereas in early-type galaxies, the *HECATE* overestimates the SFR due to the dust emission caused by the stochastic heating from old stellar populations, rather than star formation. (b) Same as panel (a), but now, the galaxies are colour coded according to their *SDSS* $g-r$ colour (in mag), which is a more reliable indicator of the stellar population age of the galaxies than the morphological classifications.

the emission-line ratio diagnostic of Stampoulis et al. (2019, which has been trained on the same data set). This diagnostic takes into account all available line ratios in order to provide a single robust activity classification that avoids the contradictory classifications that can be obtained from the use of the traditional two-dimensional line-ratio diagrams. We consider the ([S II $\lambda\lambda$6717, 6731 Å]/H α, [N II λ6584 Å]/H α, [O III λ5007 Å]/H β) three-dimensional diagram and when we have reliable measurements for the [O I] λ6300 Å line we use the four-dimensional ([S II]/H α, [N II]/H α, [O I]/H α, [O III]/H β) diagram. In this way, we provide nuclear activity classification for 64 280 (31 per cent) galaxies with signal-to-noise ratio greater than 2 in the emission lines used. Out of these 64 280 galaxies, 9987

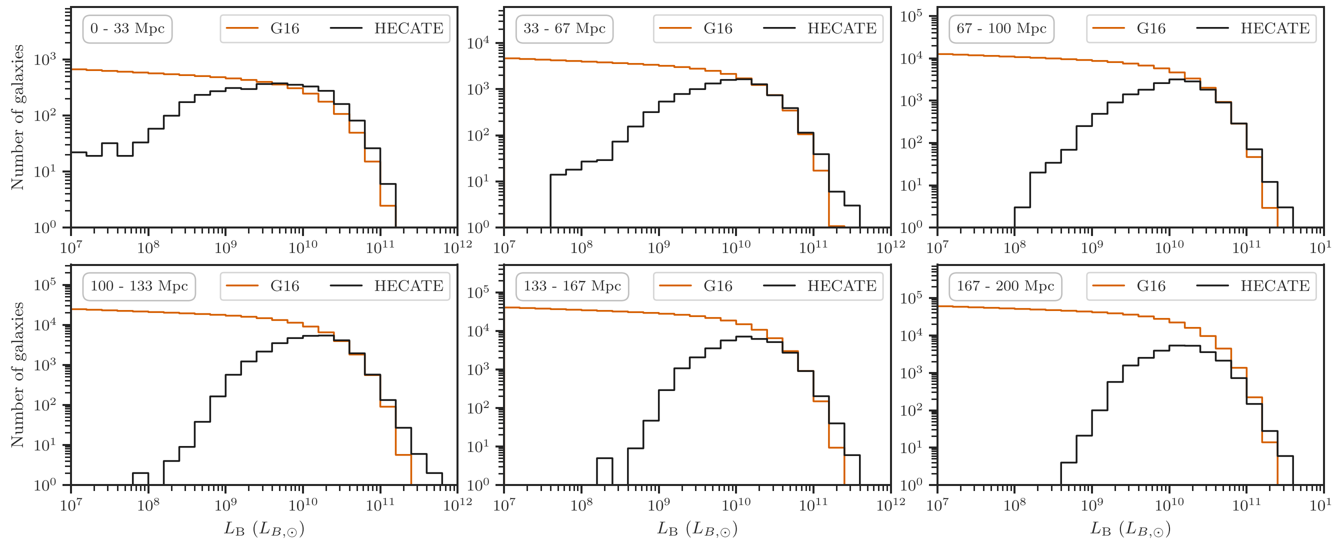


Figure 7. The distribution of the *B*-band luminosities of the galaxies in the *HECATE* at different distance ranges (top left boxes), and comparison against the expectation from the $L_{\rm B}$ LF in Gehrels et al. (2016).

(15 per cent) are characterized as AGN, leaving a non-AGN sample of 54 293 galaxies.

One of the motivations for the compilation of this galaxy catalogue was the study of X-ray source populations in 'normal' (i.e. non-AGN hosting) nearby galaxies. Therefore, we also include the AGN classifications from She, Ho & Feng (2017), who studied galaxies that had been observed with *Chandra* at distances less than 50 Mpc. In total, we obtain classifications for 716 galaxies.

Finally, we combine the classifications in a single estimate. For galaxies with classifications from only one of the two sources, we adopt them as they are. For galaxies in both the *SDSS* and She et al. (2017) sample, they are characterized as AGN if they are classified as such by either of the two sources, otherwise as non-AGN. The *SDSS*, She et al. (2017), and the combined classifications are all provided in the catalogue (for 64 280, 716, and 64 910 objects, respectively, leaving 139 823 galaxies without classification).

5 DISCUSSION

5.1 Comparison with other catalogues

Out of the available all-sky galaxy catalogues only the *Galaxy List for the Advanced Detector Era* (*GLADE*; Dálya et al. 2018), the *Mangrove* (Ducoin et al. 2020), and the *Census of the Local Universe* (*CLU*; Gehrels et al. 2016; Cook et al. 2019) are similar in scope (i.e. offer multiwavelength photometry and galaxy characterization) as the *HECATE*.

The *GLADE* galaxy catalogue provides coordinates, distances, and photometry in the *B*, *J*, *H*, and *K* bands by cross-matching five catalogues: *HyperLEDA*, *2MASS*-XSC, *GWGC*, the *2MASS photometric redshift catalogue*, and *SDSS*-DR12Q. Without an explicit limit on z, it is an ideal tool for low-redshift cosmology, and studies of distant transient events such as long GRBs. A recent extension of the *GLADE* is the *Mangrove* catalogue, which provides $M_\star$ estimates via mid-IR photometry obtained by cross-matching the *GLADE* and the *AllWISE* catalogue.

Over the distance range covered by the *HECATE*, the completeness of *GLADE* and *Mangrove* in terms of the *B*-band luminosity is similar to that of the *HECATE* (cf. fig. 2 in Dálya et al. 2018 and Fig. 8). However, the two catalogues do not include size information for the galaxies, limiting their usability for the association of host galaxies with sources from serendipitous and all-sky surveys (e.g. Webb et al. 2020). Another important difference between the *HECATE* and the *GLADE* or *Mangrove* is that the *HECATE* provides robust distances for local Universe galaxies,[8] SFRs based on a wide suite of indicators, as well as homogenized SFRs that bridge the systematic differences between the individual indicators, and integrated *2MASS* and *WISE* photometry for nearby galaxies.

The *CLU* catalogue has been progressively constructed since 2016 to aid the identification of GW hosts (Gehrels et al. 2016), and provide a census of emission-line galaxies with $D < 200$ Mpc using new observations (Cook et al. 2019). Including information from the *NED*, *HyperLEDA*, Extragalactic Distance Database, *SDSS-DR12*, 2dF Galaxy Redshift Survey, the Arecibo Legacy Fast ALFA, *GALEX*, and *WISE*, it provides multiwavelength data, SFRs and $M_\star$ based on *WISE* photometry. However, for studies of nearby galaxies, the *CLU* has the same limitations as in the case of *GLADE*: It does not provide size information on the sample galaxies, and the *WISE*-based photometry is problematic for nearby, extended objects (Section 4.1.2).

Concluding, the *HECATE* provides robust distances (an important parameter for nearby galaxies; see Section 3), and additional data that are not readily available in the other catalogues: reliable homogenized SFRs, metallicities, and morphological and AGN classifications.

5.2 Completeness

The completeness of the *HECATE* cannot be robustly calculated due to the unknown selection function of the *HyperLEDA*, which is

[8]While the use of photometric redshifts in *GLADE* and *Mangrove* provides distances estimates for distant galaxies without spectroscopic measurements, their typical uncertainty of $\Delta z = 0.015$ (Dálya et al. 2018) is prohibitive for galaxies in the local Universe ($z < 0.047$). In addition, redshift-independent distances are provided through the *GWGC* catalogue, which is limited to 100 Mpc.

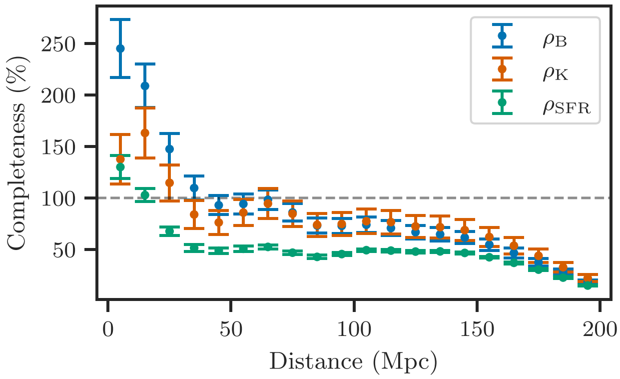


Figure 8. The completeness of the *HECATE* in terms of the included *B*-band, K_s-band, and SFR density with respect to the expectation from observational estimates. The completeness is between 50 per cent and 100 per cent within 150 Mpc. At small distances, the completeness exceeds 100 per cent because of the overdensity in the neighbourhood of the Milky Way (cf. Gehrels et al. 2016).

further complicated by the selection effects introduced by the other catalogues that it is cross-correlated with. However, we can obtain an estimate of the completeness by comparing the distribution of *B*-band luminosities with the expectation from the galaxies LF, following the approach of Gehrels et al. (2016) and Dálya et al. (2018). Using the same Schechter LF as in the aforementioned papers,[9] we compute the expected number of galaxies in different bins of luminosities and distances, shown in orange in Fig. 7, which we compare with the number of galaxies in the *HECATE* in the respective bins (black). We find that the *HECATE* is complete down to $L_B \sim 10^{9.5} L_{B,\odot}$ at distances less than 33 Mpc, and down to $L_B \sim 10^{10} L_{B,\odot}$ at $67 < D < 100$ Mpc. However, at distances greater than 167 Mpc the *HECATE* suffers incompleteness even at the high end of the LF.

Since many applications of the *HECATE* are related to the stellar content of the galaxies, we can quantify its completeness in terms of the ratio of the integrated *B*-band luminosity of galaxies at distance *D*, with respect to the mean L_B density of the local Universe. This approach has been followed in several studies of nearby samples of galaxies: Kopparapu et al. (2008), White et al. (2011), Gehrels et al. (2016), and Dálya et al. (2018). We adopt the mean L_B density of $(1.98 \pm 0.16) \times 10^8$ Mpc^{-3} (Kopparapu et al. 2008) that was used by the aforementioned works. To account for the different sources of uncertainties. we sample from the distributions of the various quantities involved in the computation (i.e. the mean L_B density, and the galaxy distances), and compute the completeness in bins of 10 Mpc. This is performed for 10 000 iterations to obtain the mean and standard deviation of the completeness as a function of the distance. The L_B completeness is shown by blue error bars in Fig. 8. We find that the *HECATE* is >75 per cent complete in terms of the blue light at $D < 100$ Mpc, and ∼50 per cent at $D \sim 170$ Mpc. The completeness above 100 per cent at small distances ($D < 30$ Mpc) is the result of the overdensity in the neighbourhood of the Milky Way.

Similarly, we calculate the completeness of the *HECATE* in terms of the $M_\star$. For this reason, we perform the same exercise with the K_s-band luminosity, which is a tracer of the $M_\star$ of the galaxies. We adopt a K_s-band luminosity density of $5.8 \times 10^8 h \mathrm{L}_{K,\odot}$ Mpc^3 (Bell et al. 2003). The result is similar to the L_B completeness as shown by orange in Fig. 8, exhibiting both the excess at small distances and the cut-off at large distances.

The completeness in terms of the SFR is calculated in the same way (shown by green points in Fig. 8), adopting a local Universe SFR density of 0.015 M$_\odot$ yr^{-1} Mpc^{-3} (Madau & Dickinson 2014), and using the homogenized SFR for the *HECATE* galaxies. In this case, the *HECATE* is incomplete at all distances in its regime, with ∼50 per cent completeness at $30 < D < 150$ Mpc. The lower completeness in SFR with respect to the other parameters (L_B and $M_\star$) stems from the fact that the *WISE*-based SFRs in the *HECATE* do not have all-sky coverage since they are based on forced photometry on *SDSS* objects. Nevertheless, due to the all-sky coverage of *IRAS* and despite its shallowness, it covers more than 50 per cent of the star-forming activity in the Galactic neighbourhood. Finally, the completeness might be overestimated with respect to estimates from SED methods, since the homogenized SFRs are systematically higher for galaxies in the low-SFR regime (see Fig. 6).

5.3 Limitations

The parent sample of the *HECATE*, the *HyperLEDA* data base, includes objects and related data, from hundreds of surveys with different sky coverage and sensitivity limits. Therefore, the selection function of the *HyperLEDA*, and as a consequence, that of the *HECATE*, is intractable (Section 5.2). Generalizations based on the provided galaxy compilation should be treated carefully.

At low distances ($D \lesssim 20$ Mpc) peculiar velocities dominate the Hubble flow (Section 3.2). This is accounted for by the regression model for estimating distances based on the recession velocities of the galaxies; however, the increased scatter reduces the accuracy of the inferred distances for velocities $v_{\mathrm{vir}} \lesssim 1500$ km s^{-1} (Fig. 3). This can be remedied by measuring *z*-independent distances for the nearby galaxies. In addition, there are a few cases where distance measurements are significantly different from the Hubble-flow distance.[10] The causes of these discrepancies are diverse and difficult to identify in most cases (e.g. problematic distances due to biases in distance indicators, wrong redshifts because of superimposed stars, typos, etc.). In the future, the methods for estimation of distances will include special treatment for such outliers.

Furthermore, the derived stellar population parameters are based on multiwavelength data from combinations of surveys and calibrations. The statistical treatments presented in this paper (e.g. homogenization of SFR estimates, fixed *M/L* ratio for galaxies without *M/L* estimate) provide estimates of stellar populations for a large fraction of galaxies in the local Universe. While this allows for statistical studies of large galaxy samples, or quick searches for objects of interest, more accurate methods ought to be preferred when focusing on individual galaxies.

The IR-based SFR estimates are based on calibrations that assume ‘normal’ star-forming galaxies. In the case of quenched, early-type galaxies, the SFRs may be overestimated (e.g. Hayward et al. 2014). Indicators based on optical–UV SED analysis could be more reliable for these galaxies.

One of the most important limitations of the *HECATE* is its non-uniform coverage in terms of the SFR and $M_\star$. In Fig. 9, we show the coverage of stellar population parameters in the *HECATE*. SFR, $M_\star$, and metallicity estimates are available for 46 per cent, 65 per cent,

[9] $\Phi = 1.6 \times 10^{-2} h^3$ Mpc^{-3}, $a = -1.07$, and $L_\star = 1.2 \times 10^{10}$ L$_{B,\odot}$ (cf. Gehrels et al. 2016). We adopt $h = 0.7$ as an intermediate value between the published H_0 calibrations.

[10] e.g. NGC 5434 is reported to have a distance of 3.8 Mpc both in *NED-D* and subsequently in the *HECATE*, but its *z* implies $D \approx 70$ Mpc.

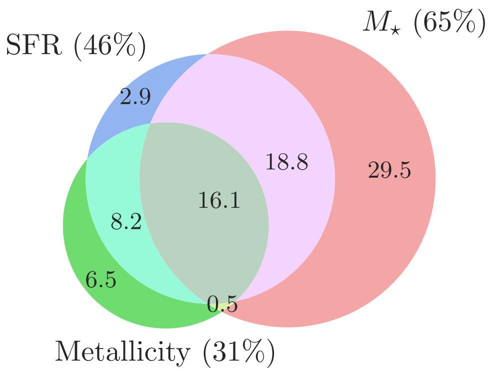


Figure 9. Venn diagram of the coverage of the stellar population parameters, SFR, $M_\star$, and metallicity, in the *HECATE*. The per cent coverage for each is reported next to its label, while the numbers in the coloured areas denote the percentage for the different combinations of parameters.

and 31 per cent of the galaxies, respectively. Currently, the *WISE* photometry is obtained through the forced photometry catalogue of Lang et al. (2016) that is limited to the *SDSS* footprint. While this is driven by the need for accurate photometry for the extended galaxies (which are the majority of the *HECATE* galaxies), it leaves a significant fraction of the sample without sensitive IR photometry that could provide reliable and uniform SFR measurements. For specific cases, this limitation can be remedied by including in the analysis data from additional catalogues. In a future version of the *HECATE*, we will apply the forced photometry method to all galaxies in the *HECATE*, thus providing robust stellar population parameters. In addition, incorporation of additional photometry and spectroscopy from other surveys (e.g. *Pan-STARSS* and *LAMOST*; Chambers et al. 2016; Deng et al. 2012) will increase the multiwavelength, AGN classification, and metallicity coverage of the *HECATE*. This will also allow the computation of SED fits, which will provide additional SFR and $M_\star$ estimates for the galaxies.

5.4 Applications

The motivation for creating an all-sky galaxy catalogue with positions, sizes, multiwavelength data, and derived parameters (e.g. SFR, $M_\star$, and metallicity) was to enable several applications relying on the initial characterization of sources in the context of the host galaxy, or identifying counterparts of transient events for follow-up observations. In this section, we outline some specific use cases.

5.4.1 Application to all-sky and serendipitous surveys

The distance limit of the *HECATE* and the large array of the information it provides make it an ideal sample for designing wide-area multiwavelength surveys or characterizing sources within. For example, it can form the baseline sample for realistic simulations of the data expected to be obtained with future surveys (see Basu-Zych et al. 2020, for an application to the *eROSITA* survey), but also it can be a reference sample for the initial characterization of newly identified sources (e.g. with Dark Energy Survey: Flaugher 2005; Large Synoptic Survey Telescope, *LSST* survey: Ivezić et al. 2019).

A demonstration of the potential of the *HECATE* is given in Kovlakas et al. (2020), a comprehensive study of ultraluminous X-ray sources in the local Universe based on the *Chandra Source Catalog 2.0* (Evans et al. 2010). The positional and size information available in *HECATE* allowed the association of X-ray sources with their host galaxies and the robust estimation of the fraction of interlopers. In addition the SFR, $M_\star$, and metallicity information was used to derive scaling relations between the ULXs and the stellar populations in their host galaxies. The special treatment of nearby galaxies (e.g. extended photometry) in the *HECATE* was essential for the science in this project since the target sample was limited in a volume out to 40 Mpc. Similarly, the combination of *HECATE* with *XMM–Newton* has been the basis for the largest study of the X-ray scaling relations of galaxies (Anastasopoulou et al., in preparation), and the largest *XMM–Newton* census of ULXs in nearby galaxies up to date (Bernadich et al., submitted).

5.4.2 Application in search of EM counterparts to GW sources

All-sky galaxy catalogues are crucial for the timely identification of electromagnetic (EM) counterparts to GW sources (e.g. Nissanke, Kasliwal & Georgieva 2013; Gehrels et al. 2016). This is a key step for constraining their nature (e.g. Abbott et al. 2017a), understanding the formation and evolution of their progenitors (e.g. Kalogera et al. 2007; Abbott et al. 2017c), or even using them as standard ‘sirens’ to measure the Hubble constant (e.g. Schutz 1986; Abbott et al. 2017b; Chen, Fishbach & Holz 2018).

The poor localization of GW sources by the contemporary GW detectors ($\gtrsim$100 deg^2; Abbott et al. 2020) makes the search for EM counterparts a daunting task. The adopted solution is to perform targeted follow-up observations of a list of potential hosts prioritized based on properties such as their distance, or the parameters of their stellar populations (e.g. Kanner et al. 2008; Nuttall & Sutton 2010; Gehrels et al. 2016; Arcavi et al. 2017; Cook et al. 2017; Dálya et al. 2017; Kasliwal et al. 2017; Del Pozzo et al. 2018; Yang et al. 2019; Salmon et al. 2020; Wyatt et al. 2020).

This approach has already led to the compilation of galaxy catalogues that provide in addition to positions and distances, photometric information (as proxies to SFR; e.g. Kopparapu et al. 2008; White et al. 2011; Gehrels et al. 2016), or directly SFR and $M_\star$ determinations (e.g. Dálya et al. 2018; Cook et al. 2019; Ducoin et al. 2020). This is driven by models that predict that GW populations scale with SFR (e.g. Phinney 1991) and/or $M_\star$ (e.g. Mapelli et al. 2017; Artale et al. 2019; Toffano et al. 2019; Adhikari et al. 2020). However, these catalogues lack information on metallicity that can be an important factor in the GW rates (e.g. O’Shaughnessy et al. 2017; Mapelli et al. 2018; Artale et al. 2019, 2020a; Neijssel et al. 2019; Bavera et al. 2020).

The *HECATE*, having a distance limit (∼200 Mpc) that is sufficient for searches of EM counterparts to GW sources from binary neutron stars (BNSs) until the mid-2020s (e.g. Buikema et al. 2020), and providing stellar population parameters, can be used for assigning likelihoods to putative GW hosts for observational follow-up campaigns. In this section, we use as an example the GW event GW170817, the only case of verified EM counterpart of a BNS, to

(i) illustrate the use of the *HECATE* in producing priority lists of galaxies for EM counterpart searches,

(ii) study the effect of the different pieces of information (direction, distance, and stellar population parameters) on the prioritization of host candidate galaxies,

(iii) assess, post facto, the ability of various schemes in giving high priority to the host galaxy of GW170817, NGC 4993.

The priority lists are the result of ordering the galaxies based on their probability of being the hosts,

$$P \propto P_{3D} \times G_{intr}, \tag{8}$$

where P_{3D} is the volume-weighted probability given the position and distance of the galaxy, and G_{intr} is a factor (or grade) that scales with the probability for a galaxy to host a GW event given its intrinsic properties (e.g. $M_\star$ or SFR proxy, or merger rate).

As a first step, we acquire the HEALPix map (Górski et al. 2005) produced by BAYESTAR (Singer et al. 2016) that contains the 2D localization probability, i.e. the probability that the GW event is on a specific direction of the sky, and the corresponding distance probability distribution. By cross-matching the *HECATE* with the HEALPix map, we find 2249 candidate host galaxies in the 99.9 per cent region of GW170817 (based on the 2D localization probability). As the 'directional', namely the 2D probability of the galaxy, P_{2D}, we assign the value of the HEALPix pixel that contains the centre of the galaxy. The 3D probability, $P_{3D} = P_{2D} \times P_d$, is computed by combining the P_{2D} with the GW event distance probability density (P_d) for the corresponding pixel in the HEALPix map, and the distance of the galaxy in the *HECATE*.

Subsequently, the 3D probabilities are multiplied by 'astrophysical' terms (G_{intr}) that are assumed to be proportional to the merger rate of BNSs, and therefore the probability of a merger. The astrophysical terms are generally parametrized in terms of the L_B (cf. Arcavi et al. 2017; Salmon et al. 2020), stellar mass (cf. Ducoin et al. 2020), and the theoretical predictions of the merger rate of BNS, as a function of different combinations of the stellar population parameters: (i) $n(M_\star)$, (ii) $n(M_\star, \mathrm{SFR})$, and (iii) $n(M_\star, \mathrm{SFR}, Z)$, based on the results of Artale et al. (2020a) for $z = 0$ (cf. their table 1), where Z is the metallicity of the galaxy.

Since the three stellar population parameters may not be known for all galaxies in the HEALPix map, we also use a 'combined' estimate, where the appropriate merger rate is used depending on the available information:

$$n_{comb} = \begin{cases} n(M_\star, \mathrm{SFR}, Z) & \text{if } M_\star, \text{SFR and } Z \text{ are defined} \\ n(M_\star, \mathrm{SFR}) & \text{if } M_\star \text{ and SFR are defined} \\ n(M_\star) & \text{if } M_\star \text{ is defined} \end{cases}. \tag{9}$$

Finally, in order to include in our analysis galaxies without $M_\star$ estimates (for which n cannot be inferred), we employ the weighting scheme of Ducoin et al. (2020):

$$P_{Du} \propto P_{3D}\,(1 + \alpha n_{comb})\,, \quad \text{where} \quad \alpha = \frac{\sum P_{3D}}{\sum P_{3D} n_{comb}}. \tag{10}$$

The quantities, namely P_{2D}, P_{3D}, $P_{3D} \times L_B$, $P_{3D} \times M_\star$, $P_{3D} \times n(M_\star)$, $P_{3D} \times n(M_\star, \mathrm{SFR})$, $P_{3D} \times n_{comb}$, and P_{Du}, are used to produce priority lists of the host galaxy candidates, to test the aforementioned schemes for prioritizing candidate host galaxies (Table 3). The scheme that accounts for the metallicity dependence of the merger rate is omitted due to lack of metallicity estimates in the sky region of the GW event.

We find that NGC 4993 is given the highest priority by the schemes involving the L_B or $M_\star$, and second priority for those also involving the SFRs.[11] Except for the priority lists based only on the 2D or 3D position, the lists feature the same top five galaxies as in the first prioritization list published after the GW170817 alert (Kasliwal et al. 2017) based on the *CLU* catalogue (Gehrels et al. 2016; Cook et al. 2017). The same holds for the top three galaxies reported in (i) Artale et al. (2020b), who use $M_\star$ and SFR estimates from the *Mangrove* catalogue (Ducoin et al. 2020), and (ii) in Yang et al. (2019), who used the *B*-band luminosity from *GLADE* (Dálya et al. 2018) as the 'astrophysical' term. The top-ranked *HECATE* galaxies based on the different prioritization schemes agree to a high degree with the results of the same schemes in Ducoin et al. (2020) using the *Mangrove* catalogue.

[11] Following Section 5.3, because GW170817 falls outside the *SDSS* footprint, for this application we supplement the *IRAS* photometry with mid-IR photometry from the *AllWISE* catalogue.

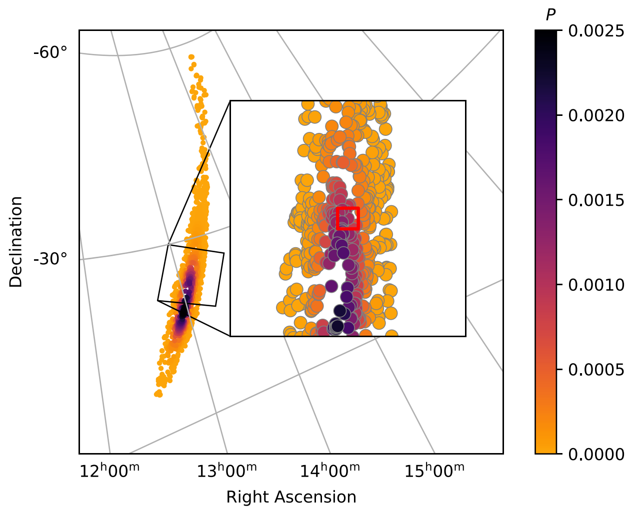
Table 3. Prioritization lists of host galaxy candidates (first five sources), and computed probabilities based on different schemes. The true host (bold text), NGC 4993, is successfully recovered as first or second most probable host galaxy once the astrophysical information is accounted for.

Galaxy[a]	P_{2D}	Galaxy[b]	P_{3D}
PGC4690279	0.002	ESO508-004	0.054
PGC3799401	0.002	ESO575-055	0.051
PGC3798804	0.002	ESO575-053	0.049
PGC4690296	0.002	PGC4692149	0.045
PGC4690280	0.002	PGC169673	0.045
Galaxy	$P_{3D} \times L_B$	Galaxy	$P_{3D} \times M_\star$
NGC 4993	0.096	**NGC 4993**	0.163
ESO508-019	0.079	NGC 4830	0.148
IC4197	0.074	IC4197	0.119
NGC 4830	0.073	NGC 4970	0.115
NGC 4970	0.072	NGC 4968	0.103
Galaxy	$P_{3D} \times n(M_\star)$	Galaxy	$P_{3D} \times n(M_\star, \mathrm{SFR})$
NGC 4993	0.164	NGC 4968	0.180
NGC 4830	0.151	**NGC 4993**	0.135
IC4197	0.121	NGC 4830	0.102
NGC 4970	0.117	IC4187	0.100
NGC 4968	0.102	NGC 4970	0.087
Galaxy	$P_{3D} \times n_{comb}$	Galaxy	P_{Du}
NGC 4968	0.180	NGC 4968	0.082
NGC 4993	0.135	**NGC 4993**	0.071
NGC 4830	0.102	NGC 4830	0.050
IC4197	0.100	IC4197	0.049
NGC 4970	0.087	NGC 4970	0.044

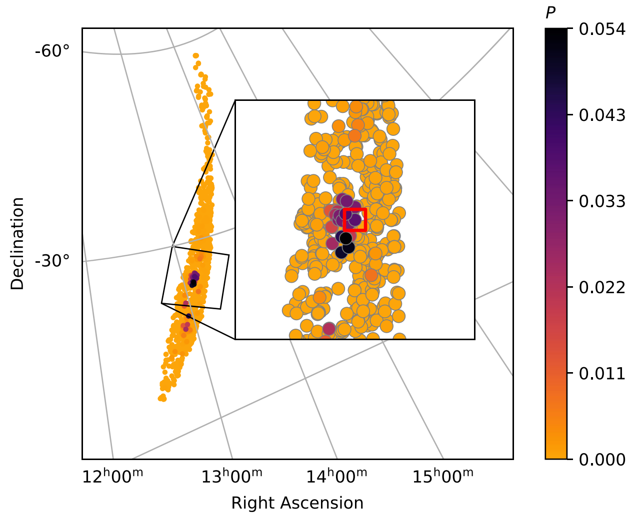
[a] The rank of NGC 4993 is 461.
[b] The rank of NGC 4993 is 7.

While in the case of GW170817 there is no significant difference between the use of *B*-band luminosity, $M_\star$ or the fits with $M_\star$ and SFR (NGC 4993 was always first or second with a small difference in the probability), refined prioritization schemes will be important for the quick identification of EM counterparts of future BNS coalescence signals with poorer localization or at higher distances. The *HECATE*, making readily available a large set of intrinsic properties for the candidate host galaxies, offers versatility in the choice, design, and assessment of different priority schemes.

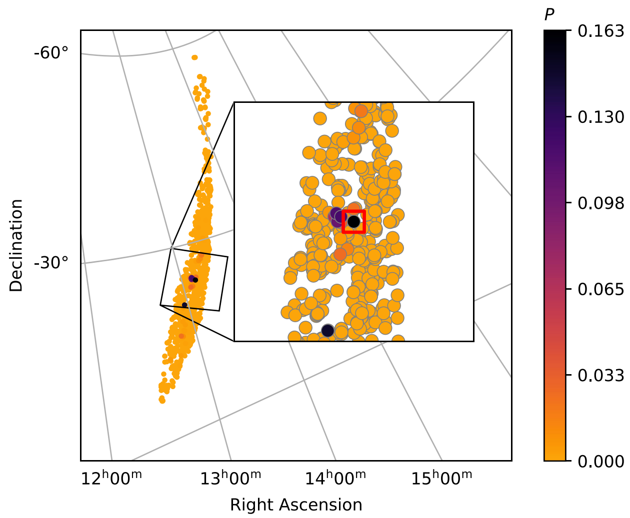
We note that the practice of initially narrowing down the galaxy sample by deciding on a confidence region based on the 2D probability increases the risk of missing the true host. This is indicated by the rank of NGC 4993 in the P_{2D}-based priority list (461), and the fact that total 2D probability of galaxies closer to the centroid (a few degrees from NGC 4993) is as high as 75 per cent. We suggest using the full galaxy catalogue together with priority schemes involving distance (and astrophysical information where possible). For example, the inclusion of the distance information in P_{3D} promotes NGC 4993 to the seventh position, and shifts the centroid by a few degrees (see panel b of Fig. 10), a consequence of the non-homogeneity of the Universe at the distance of the event (out of the 15 galaxies in Table 3



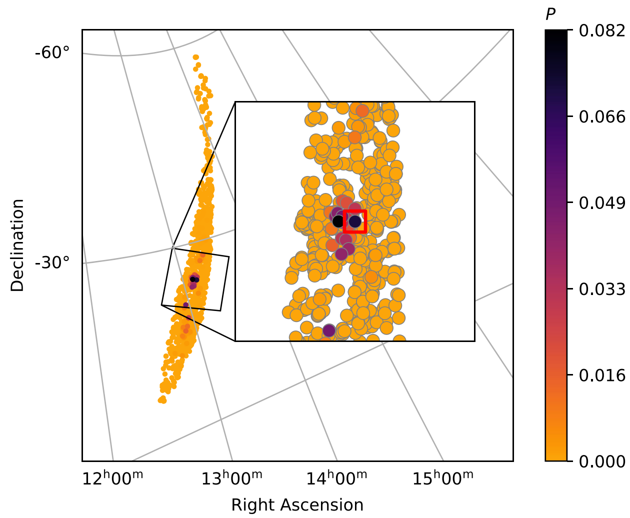
(a) 2-D probability



(b) 3-D probability



(c) 3-D probability combined with $M_\star$



(d) 3-D probability combined with merger rate as a function of $M_\star$ and SFR, using the Ducoin et al. (2020) weighting scheme.

Figure 10. (a) The sky distribution of the 2248 galaxies in the *HECATE* that lie inside the 99.9 per cent confidence region of the 2D localization map for the GW170817 event. The inset zooms in the region of the galaxy NGC 4993 (red square). The colour indicates the normalized 2D probability across all the host candidates. In (b), the 3D probability is used, accounting for the distance estimates at each direction in the sky. In (c), the 3D probability is multiplied by the stellar mass. Similarly, in (d) the 3D probability is multiplied by the merger rate as computed by the fits in Artale et al. (2020a), and the weighting scheme of Ducoin et al. (2020) is used to allow galaxies without stellar population parameter estimates to enter in the prioritization list. Note that in (a) and (b) the contrast of the highest probability galaxies to the rest is small. The introduction of the 'astrophysical' terms in (c) and (d) gives prominence to NGC 4993.

for schemes with 3D positional term, 10 are considered members of the NGC 4970 group; Kourkchi & Tully 2017).

Finally, as we show in Fig. 6, infrared estimates of the SFR (as those provided by the *HECATE*) can be overestimated up to ∼2 dex with respect to SED estimates, in the case of low-SFR galaxies (such as NGC 4993). Therefore, grading schemes that account for the expected SFR of the host galaxy may overestimate the probability of a low-SFR galaxy to host the GW event. As an example, the formula given in Artale et al. (2020a) for the number of GW events for a BNS event in a galaxy at $z = 0.1$ (which is the case most sensitive to the SFR), an overestimation by two orders of magnitude in the SFR, results in an overestimation by a factor of ∼4 in the BNS rate, and consequently the assigned probability of the galaxy as a host to an event. For GW detections in the *SDSS* footprint, the use of SED-based stellar population parameters (also given in the catalogue) is advised.

5.4.3 *Application in short GRBs*

Another manifestation of BNS mergers is short GRBs (sGRBs; e.g. Tanvir et al. 2013) as it has been shown by the association of GW170817 to GRB170817A (Goldstein et al. 2017). The identification of the host galaxies of sGRBs is important for two reasons: (a) connecting their populations with the star formation history of their host galaxies, and (b) measuring the displacement of the GRBs from their host galaxies. The former is key for modelling the evolutionary paths of sGRBs and their cosmological evolution (e.g. Leibler &

Berger 2010; Abbott et al. 2017c; Selsing et al. 2018). The latter is important for constraining the effect of kicks in the populations of sGRBs (e.g. Zevin et al. 2019) and studying the enrichment of the interstellar medium in r-process elements (e.g. Andrews & Zezas 2019). The *HECATE* can provide the initial information required to quickly associate a GRB with their host galaxy, which is also important for prompt follow-up observations.

5.4.4 Localization of neutrino and cosmic ray sources

In the case of neutrino events, the large error box of their localization poses the same challenges as the GW detections (e.g. Krauß et al. 2020). Therefore, a catalogue of galaxies, which is as complete as possible, can be a valuable resource for the identification of their origin when one could follow a similar approach as the prioritized host list developed for GW events (Section 5.4.2). Despite of the lack of distance estimates in the case of neutrino events, which can significantly affect the numbers and locations of the host candidates (compare panels a and b in Fig. 10), the availability of multiwavelength data, stellar population parameters, and nuclear activity classifications in the *HECATE* is particularly useful for weighting the candidate galaxies according to their astrophysical properties since proposed extragalactic neutrino sources may be star-forming galaxies, AGN, etc. (see Ahlers & Halzen 2014; IceCube Collaboration 2018, and references therein). In addition, future observations with a combination of catalogues such as the *HECATE* may aid in constraining the correlation of the neutrino emission and host galaxy properties.

The same holds for the case of cosmic ray detections, which also have very large error circles (e.g. Pierre Auger Collaboration 2015). Recent studies of anisotropy in the arrival direction of high-energy cosmic rays, and the existence of a dipole at high Galactic latitude, indicate that their origin is neither exclusively Galactic nor cosmological (e.g. Pierre Auger Collaboration 2017). Since the Greisen–Zatsepin–Kuzmin effect limits the propagation of high-energy cosmic rays to $\lesssim$100 Mpc (e.g. Bhattacharjee 2000), the *HECATE*, as an all-sky galaxy catalogue at this distance range, can be used for the detailed study of their origin (e.g. Pierre Auger Collaboration 2010; He et al. 2016).

5.4.5 Applications in transient astronomy

In the following paragraphs, we outline potential applications of the *HECATE* in various other fields of transient and multimessenger astrophysics.

Tidal disruption events (TDEs) are typically witnessed as an outburst in X-ray or optical wavelengths resulting from accretion of the material shredded from a star under the effect of the tidal field of a supermassive black hole (SMBH). Such events are expected to be routinely detected in the *eROSITA* all-sky X-ray survey and the *LSST* optical survey. *HECATE* can provide the basis for the quick identification of the host of such an event and its basic properties. In particular, information on the distance, the presence of an AGN, and the velocity dispersion (used to initially estimate the SMBH mass; all available in the *HECATE*) is valuable for a quick interpretation of transient events (e.g. French et al. 2020).

A value-added catalogue providing robust distances and stellar population parameters is also useful for the characterization and study of the populations of transient events observed in ongoing or future multiwavelength surveys. For example, *LSST* is expected to provide a host of supernovae every night. The association of these events with a catalogue like *HECATE* will facilitate systematic studies of their populations in the context of their host galaxies (e.g. $M_\star$, SFR, and most importantly metallicity; e.g. Greggio & Cappellaro 2019). These pilot studies can be used to effectively plan more focused follow-up observations. The same holds for the identification of hosts of fast radio bursts (e.g. Marcote et al. 2020).

6 CONCLUSIONS AND FUTURE WORK

We present a new catalogue of galaxies that includes all known galaxies within a distance limit of $D \lesssim 200$ Mpc. We

(i) base our sample on the *HyperLEDA* data base, incorporating 204 733 galaxies with a radial velocity $\lesssim$14 000 km s^{-1},

(ii) use all available distance measurements for the sample to get robust redshift-independent distances, which are preferred over recessional velocity-based estimates for galaxies in the local Universe, for as many galaxies as possible (10 per cent),

(iii) compute redshift-dependent distances for the rest of the galaxies (90 per cent) that are consistent with the redshift-independent distances (Kernel Regression method), while quantifying their uncertainties due to the unknown peculiar velocity component,

(iv) incorporate integrated multiband photometry with special treatment for nearby and/or extended galaxies,

(v) derive SFRs, $M_\star$, metallicities, and nuclear activity classifications utilizing the best available information for each galaxy,

(vi) offer five different IR-based SFR indicators, as well as a homogenized SFR indicator, while providing all the necessary information for user-defined calibrations.

Despite its limitations in terms of the completeness of the catalogue (Section 5.2), and data coverage (Section 5.3), the *HECATE* is a highly complete sample of known galaxies in the local Universe. Owing to its wealth of information, the *HECATE* can be a useful tool for a wide range of applications. By providing positions and size information, the catalogue can be used as the basis of future associations of galaxies with additional multiwavelength surveys. We discuss a wide range of applications, including the prioritization of host galaxies for follow-up searches for EM counterparts of GW sources, as well as the initial characterization of transient sources that will be critical in the era of Big Data of astronomy.

Future versions of the catalogue will expand the distance range beyond the current limit of 200 Mpc, and provide a wider coverage in terms of the stellar population parameters. SFR and $M_\star$ estimates will be improved by (i) including additional multiwavelength data, (ii) adoption of forced-photometry techniques allowing the full exploitation of existing all-sky surveys also for extended objects, and (iii) performing SED analysis. Finally, incorporation of different sources of spectroscopic data will not only extend the coverage of metallicity and nuclear activity classifications, but also more importantly will serve as a cross-validation data set for AGN classifications. This is crucial for many areas of applications (e.g. screening for AGN in X-ray studies of galaxies and identification of candidate sources of high-energy gamma-ray or cosmic rays).

ACKNOWLEDGEMENTS

We are grateful to the anonymous referee for critically and thoroughly reviewing this paper, motivating improvements to the manuscript and the catalogue.

KK thanks D. Makarov for providing support with the *HyperLEDA* data base, D. Lang for his help with the *WISE* forced photometry catalogue, and C. Berry and N. Stergioulas for discussions regarding the application on gravitational wave sources. KK also thanks P. Bonfini, A. Maragkoudakis, P. Sell, and S. J. Williams for their helpful comments on the compilation of the catalogue. We also acknowledge early users of the *HECATE*: K. Anastasopoulou, M. Colom i Bernadich, F. Haberl, A. Schwope, N. Vulic, and J. Wilms, for providing feedback that led to improvements, and identifying mistakes, limitations as well as possible extensions of the catalogue.

The research leading to these results has received funding from the European Research Council under the European Union's *Seventh Framework Programme* (FP/2007-2013)/*ERC* grant agreement number 617001, and the European Union's Horizon 2020 research and innovation programme under the *Marie Skłodowska-Curie RISE* action, grant agreement number 691164 (*ASTROSTAT*). JJA acknowledges funding from CIERA and Northwestern University through a Postdoctoral Fellowship. The material is based on work supported by NASA under award number 80GSFC21M0002.

We acknowledge the usage of the *HyperLEDA* data base, the TOPCAT astronomical software (http://www.starlink.ac.uk/topcat/), and the LIGO.SKYMAP package for PYTHON. This research made use of the cross-match service provided by CDS, Strasbourg. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI : 10.26093/cds/vizier). The original description of the VizieR service was published in 2000, A&AS, 143, 23. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Funding for *SDSS*-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The *SDSS*-III web site is http://www.sdss3.org/. This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

DATA AVAILABILITY

The data underlying this article are available at the *HECATE* Portal, at http://hecate.ia.forth.gr.

REFERENCES

Abbott B. P. et al., 2017a, Phys. Rev. Lett., 119, 161101
Abbott B. P. et al., 2017b, Nature, 551, 85
Abbott B. P. et al., 2017c, ApJ, 850, L40
Abbott B. P. et al., 2020, Living Rev. Relativ., 23, 3
Ackermann M. et al., 2012, ApJ, 755, 164
Adhikari S., Fishbach M., Holz D. E., Wechsler R. H., Fang Z., 2020, ApJ, 905, 21
Aguado D. S. et al., 2019, ApJS, 240, 23
Ahlers M., Halzen F., 2014, Phys. Rev. D, 90, 043005
Andrews J. J., Zezas A., 2019, MNRAS, 486, 3213
Arcavi I. et al., 2017, ApJ, 848, L33
Artale M. C., Mapelli M., Giacobbo N., Sabha N. B., Spera M., Santoliquido F., Bressan A., 2019, MNRAS, 487, 1675
Artale M. C., Mapelli M., Bouffanais Y., Giacobbo N., Pasquato M., Spera M., 2020a, MNRAS, 491, 3419
Artale M. C., Bouffanais Y., Mapelli M., Giacobbo N., Sabha N. B., Santoliquido F., Pasquato M., Spera M., 2020b, MNRAS, 495, 1841
Basu-Zych A. R. et al., 2020, MNRAS, 498, 1651
Bavera S. S. et al., 2020, A&A, 635, A97
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Bhattacharjee P., 2000, Phys. Rep., 327, 109
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
Buikema A. et al., 2020, Phys. Rev. D, 102, 062003
Chambers K. C. et al., 2016, preprint (arXiv:1612.05560)
Chen H.-Y., Fishbach M., Holz D. E., 2018, Nature, 562, 545
Cluver M. E., Jarrett T. H., Dale D. A., Smith J.-D. T., August T., Brown M. J. I., 2017, ApJ, 850, 68
Colless M. et al., 2001, MNRAS, 328, 1039
Cook D. O., van Sistine A., Singer L., Kasliwal M. M., Growth (Global Relay Of Observatories Watching Transients Happen) Collaboration, 2017, GCN Circ., 21519, 1
Cook D. O. et al., 2019, ApJ, 880, 7
Cutri R. M. et al., 2012, VizieR Online Data Catalog. p. II/281
Dale D. A., Helou G., 2002, ApJ, 576, 159
Dálya G., Bécsy B., Raffai P., Glade Team, 2017, GCN Circ., 21516, 1
Dálya G. et al., 2018, MNRAS, 479, 2374
de Vaucouleurs G., 1991, Science, 254, 592
de Vaucouleurs G., de Vaucouleurs A., Corwin J. R., 1976, Second Reference Catalogue of Bright Galaxies, Vol. 1976. Univ. Texas Press, Austin, TX
Del Pozzo W., Berry C. P. L., Ghosh A., Haines T. S. F., Singer L. P., Vecchio A., 2018, MNRAS, 479, 601
Deng L.-C. et al., 2012, Res. Astron. Astrophys., 12, 735
Ducoin J. G., Corre D., Leroy N., Le Floch E., 2020, MNRAS, 492, 4768
Evans I. N. et al., 2010, ApJS, 189, 37
Flaugher B., 2005, Int. J. Mod. Phys. A, 20, 3121
Freedman W. L. et al., 2001, ApJ, 553, 47
French K. D., Wevers T., Law-Smith J., Graur O., Zabludoff A. I., 2020, Space Sci. Rev., 216, 32
Galliano F., Galametz M., Jones A. P., 2018, ARA&A, 56, 673
Gardner J. P., Sharples R. M., Frenk C. S., Carrasco B. E., 1997, ApJ, 480, L99
Gehrels N., Cannizzo J. K., Kanner J., Kasliwal M. M., Nissanke S., Singer L. P., 2016, ApJ, 820, 136
Goldstein A. et al., 2017, ApJ, 848, L14
Górski K. M., Hivon E., Banday A. J., Wand elt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Greggio L., Cappellaro E., 2019, A&A, 625, A113
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Hawkins E. et al., 2003, MNRAS, 346, 78
Hayward C. C. et al., 2014, MNRAS, 445, 1598
He H.-N., Kusenko A., Nagataki S., Zhang B.-B., Yang R.-Z., Fan Y.-Z., 2016, Phys. Rev. D, 93, 043011
Helou G., Walker D. W., 1988, NASA RP-1190. Vol. 7, NASA, Washington, DC
Helou G., Madore B. F., Schmitz M., Bicay M. D., Wu X., Bennett J., 1991, in Albrecht M. A., Egret D., eds, Astrophysics and Space Science Library, Vol. 171, Databases and On-line Data in Astronomy. Springer-Verlag, Berlin, p. 89
IceCube Collaboration, 2018, Science, 361, eaat1378
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, AJ, 119, 2498
Jarrett T. H., Chester T., Cutri R., Schneider S. E., Huchra J. P., 2003, AJ, 125, 525
Kalogera V., Belczynski K., Kim C., O'Shaughnessy R., Willems B., 2007, Phys. Rep., 442, 75
Kanner J., Huard T. L., Márka S., Murphy D. C., Piscionere J., Reed M., Shawhan P., 2008, Class. Quantum Gravity, 25, 184034
Karachentsev I. D., Makarov D. A., 1996, AJ, 111, 794
Karachentsev I. D., Karachentseva V. E., Huchtmeier W. K., 2001, A&A, 366, 428
Karachentsev I. D., Karachentseva V. E., Huchtmeier W. K., 2007, Astron. Lett., 33, 512

Karachentsev I. D., Makarov D. I., Kaisina E. I., 2013, AJ, 145, 101
Kasliwal M. M. et al., 2017, Science, 358, 1559
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kelly B. C., 2007, ApJ, 665, 1489
Kennicutt R. C. Jr, 1998, ARA&A, 36, 189
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kim M., Wilkes B. J., Kim D.-W., Green P. J., Barkhouse W. A., Lee M. G., Silverman J. D., Tananbaum H. D., 2007, ApJ, 659, 29
Kim S. et al., 2014, ApJS, 215, 22
Komis I., Pavlidou V., Zezas A., 2019, MNRAS, 483, 4020
Kopparapu R. K., Hanna C., Kalogera V., O'Shaughnessy R., González G., Brady P. R., Fairhurst S., 2008, ApJ, 675, 1459
Kourkchi E., Tully R. B., 2017, ApJ, 843, 16
Kovlakas K., Zezas A., Andrews J. J., Basu-Zych A., Fragos T., Hornschemeier A., Lehmer B., Ptak A., 2020, MNRAS, 498, 4790
Krauß F. et al., 2020, MNRAS, 497, 2553
Kroupa P., 2001, MNRAS, 322, 231
Lang D., 2014, AJ, 147, 108
Lang D., Hogg D. W., Schlegel D. J., 2016, AJ, 151, 36
Leibler C. N., Berger E., 2010, ApJ, 725, 1202
Leitherer C., Heckman T. M., 1995, ApJS, 96, 9
Leitherer C. et al., 1999, ApJS, 123, 3
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Makarov D., Prugniel P., Terekhova N., Courtois H., Vauglin I., 2014, A&A, 570, A13
Mapelli M., Giacobbo N., Ripamonti E., Spera M., 2017, MNRAS, 472, 2422
Mapelli M., Giacobbo N., Toffano M., Ripamonti E., Bressan A., Spera M., Branchesi M., 2018, MNRAS, 481, 5324
Marcote B. et al., 2020, Nature, 577, 190
Moshir M. et al., 1990, IRAS Faint Source Catalogue Pasadena, JPL
Nadaraya E. A., 1964, Theory Probability Appl., 9, 141
Neijssel C. J. et al., 2019, MNRAS, 490, 3740
Nilson P., 1973, Uppsala General Catalogue of Galaxies (Uppsala Astron. Obs. Ann.). Vol. 6, Uppsala University Publications
Nissanke S., Kasliwal M., Georgieva A., 2013, ApJ, 767, 124
Nuttall L. K., Sutton P. J., 2010, Phys. Rev. D, 82, 102002
O'Shaughnessy R., Bellovary J. M., Brooks A., Shen S., Governato F., Christensen C. R., 2017, MNRAS, 464, 2831
Parkash V., Brown M. J. I., Jarrett T. H., Bonne N. J., 2018, ApJ, 864, 40
Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59
Phinney E. S., 1991, ApJ, 380, L17
Pierre Auger Collaboration, 2010, Astropart. Phys., 34, 314
Pierre Auger Collaboration, 2015, ApJ, 804, 15
Pierre Auger Collaboration, 2017, Science, 357, 1266
Pietrzyński G. et al., 2013, Nature, 495, 76
Planck Collaboration XIII, 2016, A&A, 594, A13
Rowan-Robinson M., 1999, Ap&SS, 266, 291
Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11
Salim S. et al., 2016, ApJS, 227, 2
Salmon L., Hanlon L., Jeffrey R. M., Martin-Carrillo A., 2020, A&A, 634, A32
Salpeter E. E., 1955, ApJ, 121, 161
Sanders D. B., Mazzarella J. M., Kim D.-C., Surace J. A., Soifer B. T., 2003, AJ, 126, 1607
Saxton R. D., Read A. M., Esquej P., Freyberg M. J., Altieri B., Bermejo D., 2008, A&A, 480, 611
Schutz B. F., 1986, Nature, 323, 310
Selsing J. et al., 2018, A&A, 616, A48
She R., Ho L. C., Feng H., 2017, ApJ, 835, 223
Singer L. P. et al., 2016, ApJ, 829, L15
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Springob C. M., Haynes M. P., Giovanelli R., Kent B. R., 2005, ApJS, 160, 149
Stampoulis V., van Dyk D. A., Kashyap V. L., Zezas A., 2019, MNRAS, 485, 1085
Steer I. et al., 2017, AJ, 153, 37
Tanvir N. R., Levan A. J., Fruchter A. S., Hjorth J., Hounsell R. A., Wiersema K., Tunnicliffe R. L., 2013, Nature, 500, 547
Terry J. N., Paturel G., Ekholm T., 2002, A&A, 393, 57
Toffano M., Mapelli M., Giacobbo N., Artale M. C., Ghirlanda G., 2019, MNRAS, 489, 4622
Tremonti C. A. et al., 2004, ApJ, 613, 898
Tully R. B., Courtois H. M., Sorce J. G., 2016, AJ, 152, 50
Varela J. et al., 2009, A&A, 497, 667
Verley S. et al., 2007, A&A, 470, 505
Wang L., Rowan-Robinson M., Norberg P., Heinis S., Han J., 2014, MNRAS, 442, 2739
Webb N. A. et al., 2020, A&A, 641, A136
Wenger M. et al., 2000, A&AS, 143, 9
White D. J., Daw E. J., Dhillon V. S., 2011, Class. Quantum Gravity, 28, 085016
Wolf C. et al., 2018, Publ. Astron. Soc. Aust., 35, e010
Wyatt S. D., Tohuvavohu A., Arcavi I., Lundquist M. J., Howell D. A., Sand D. J., 2020, ApJ, 894, 127
Yang S. et al., 2019, ApJ, 875, 59
Zevin M., Kelley L. Z., Nugent A., Fong W.-f., Berry C. P. L., Kalogera V., 2019, ApJ, 904, 190

APPENDIX A: COMPUTATION OF THE VIRGO-INFALL CORRECTED RADIAL VELOCITIES

Starting from the heliocentric velocity, $\upsilon_{\rm hc}$, of a galaxy at galactic coordinates (l, b), we adopt the correction of Karachentsev & Makarov (1996) for solar motion in the Local Standard of Rest (LSR) and the Milky Way's motion with respect to the Local Group (LG) centroid:

$$\upsilon_{\rm lg} = \upsilon_{\rm hc} + V_a \left[\cos b \cos b_a \cos\left(l - l_a\right) + \sin b \sin b_a\right], \tag{A1}$$

where $V_a = (316{\pm}5)$ km s^{-1} is the velocity of the Sun towards the LG centroid at galactic coordinates $l_a = (93 \pm 2)^\circ$ and $b_a = (-4 \pm 2)^\circ$. Then, we correct for the Local Group's infall to the Virgo cluster following Terry, Paturel & Ekholm (2002):

$$\upsilon_{\rm vir} = \upsilon_{\rm lg} + V_{\rm lg-infall} \cos\Theta, \tag{A2}$$

where $V_{\rm lg-infall} = (208{\pm}9)$ km s^{-1} is the infall velocity of LG to Virgo cluster, and Θ is the great-circle distance between the galaxy's supergalactic coordinates and LG's apex ($102\overset{\circ}{.}88$, $-2\overset{\circ}{.}34$).

APPENDIX B: CROSS-MATCHING PROCEDURES AND OBTAINED DATA

The following paragraphs provide additional details regarding some of the cross-matching procedures described in Section 4.1.

B1 *HyperLEDA* versus *NED*

The cross-correlation of the *HyperLEDA* and the *NED* is an essential step to (i) obtain missing radial velocities, (ii) use the associations to match *HyperLEDA* objects to z-independent distance measurements in *NED-D*, and (iii) provide quick links to *NED* entries for the galaxies. This step of the pipeline is executed before applying the recession velocity cut, since *NED* complements our sample with radial velocities. Therefore, 884 766 objects are searched, i.e. galaxies with a heliocentric velocity $<$14 500 km s^{-1} (ensuring that no object with $\upsilon_{\rm vir} < 14\,000$ km s^{-1} is excluded), and objects without radial velocity information in *HyperLEDA*. We use the PYTHON 'ASTROQUERY' package to associate the *HyperLEDA* galaxies with *NED* objects on the basis of their designation: For each object in

Table B1. The various sources of semimajor axis information incorporated in the *HECATE*. Where available, axial ratios and position angles are also obtained. The semimajor axes are rescaled to match the D_{25} isophotal one in *HyperLEDA*. The columns are: (1) the name of the source; (2) the flag in the column 'dsource' in the provided catalogue; (3) the number of objects in the *HECATE* for which the sizes were obtained from the source; (4) the number of common objects in both catalogues; (5) the scaling factor C in dex, used to homogenize the sizes r_1 from the given source to R_1 (adopted semimajor axis) as in $\log R_1 = \log r_1 + C$; (6) the scatter (in dex) between the size in *HyperLEDA* and the source for the common objects; (7) notes; (8) reference of the source.

Source (1)	Flag (2)	N (3)	$N_{\rm tot}$ (4)	Scale (5)	Scatter (6)	Notes (7)	Reference (8)
HyperLEDA	H	165 482	–	–	–	This is the reference sample. Adopted as they are.	Makarov et al. (2014)
SDSS	S	12 214	124 055	0.208	0.188	Petrosian radius in the g band from Data Release 15. The g band was selected because of the small scatter in the scaling factor, as expected due to its proximity to the B band.	Aguado et al. (2019)
2MASS	2	12 918	143 546	0.236	0.118	Supercoadd 3σ isophotal semimajor axis radius ('sup_r_3sig'). The J-band 21 mag arcsec^{-2} isophotal semimajor axis presents a slightly small scatter of 0.115 but it was not available for all objects.	Jarrett et al. (2000)
2dFGS	6	6327	21 404	1.940	0.065	Areas, eccentricities, and orientations in the B band. Computed the corresponding semimajor and semiminor axes. The scaling factor converts from pixels to arcmin.	Colless et al. (2001)
WINGS	W	740	2229	−0.056	0.104	B-band isophotal ellipses.	Varela et al. (2009)
SkyMapper	Y	1814	76 488	0.355	0.203	Data Release 1.1. Mean r-band isophotal diameters.	Wolf et al. (2018)
AMIGA-CIG	A	65	5708	−0.255	0.137	R-band isophotal major axis.	Verley et al. (2007)
UNGC	K	60	658	−0.068	0.170	B-band Holmberg isophotal semimajor axis.	Karachentsev, Makarov & Kaisina (2013)
VIII/77	V	28	8472	−0.069	0.121	Semimajor axis taken from UGC and ESO, or estimated from POSS-I.	Springob et al. (2005)
KKH2001	1	26	101	–	–	No correction applied (B-band isophotes).	Karachentsev, Karachentseva & Huchtmeier (2001)
KKH2007	7	9	90	–	–	No correction applied (B-band isophotes).	Karachentsev, Karachentseva & Huchtmeier (2007)
NED	N	212	130 586	–	–	No correction applied. Miscellaneous diameters based on the B band, mainly from *ESO-LV*	http://ned.ipac.caltech.edu

HyperLEDA, we perform two searches: based on their PGC ID (e.g. PGC000002) and principal designation (e.g. UGC12889). We perform a series of checks to identify cases where

(i) the two searches (principal designation and PGC number) return different *NED* objects (1024).
(ii) different *HyperLEDA* objects are associated with the same *NED* object (510),
(iii) positions or radial velocities disagree (1232),
(iv) the *HyperLEDA* object has a large astrometric error and size, and has been associated with a *NED* object by chance (usually Zone of Avoidance objects; 33 101),
(v) there are typographic errors in galaxy pairs (e.g. A in the place of B in *NED*; 202).

The above situations are resolved automatically (e.g. positional disagreement larger than 1 arcmin), or after manual inspection. In total, 137 586 galaxies in the *HECATE* (67 per cent) are associated with *NED* objects.

B2 Supplementary size information

HyperLEDA provides the size of the galaxies based on the D_{25} isophote in the B band. However, for 39 251 objects (19 per cent) this information is not available. Using the associations of *HyperLEDA* to *NED* objects, we find that for the majority of these objects, the diameters can be obtained from *2MASS* and *SDSS*. In addition, using the CDS XMatch service (http://cdsxmatch.u-strasbg.fr/), we find eight other catalogues that can provide diameters for the majority of the rest of these objects. The catalogues used to draw this information are listed in Table B1.

The supplementary size information is incorporated by rescaling the semimajor axis from the external catalogue, $a_{\rm ext}$, using as reference the *HyperLEDA* semimajor axis, $a_{\rm hyp}$. To do so, we

(i) associate all *HyperLEDA* objects to the external catalogue,
(ii) use the associated galaxies for which both $a_{\rm hyp}$ and $a_{\rm ext}$ are defined to compute the scaling factor $c = \langle a_{\rm hyp}/a_{\rm ext}\rangle$, and
(iii) fill in the $a_{\rm hyp}$ for the galaxies in *HECATE* without semimajor axis from *HyperLEDA*: $a_{\rm hyp} = c \times a_{\rm ext}$.

The priority of the external catalogues was based on the number of common objects in the external catalogue and the *HyperLEDA*, the proximity of the band to the B band that is available in *HyperLEDA*, and the scatter in the scaling relation (samples with smaller scatter are considered as more reliable). More details can be found in Table B1 where the external catalogues are listed in the order of their priority.

When available, semiminor axes and position angles are also taken from the external catalogues (the axial ratio in the *HECATE* is the same as the one reported by the external catalogue). In total, we complete the size information for 34 413 galaxies, leaving 4837 (2.4 per cent) galaxies without such information in the *HECATE*. Because of the different wavebands and methods used by the external catalogues, the application of a scaling factor is oversimplistic and may have introduced biases. Users of the catalogue are suggested to use the corresponding flag, 'dsource', to either filter out these galaxies or study any biases.

B3 *IRAS-RBGS*

We cross-match objects in the *HECATE* and *IRAS-RBGS* on the basis of their D_{25} elliptical regions. 589 galaxies out of the 629 objects in *IRAS-RBGS* are associated with *HECATE* galaxies. The remaining 40 objects are not cross-linked for the following reasons. 19 associations are rejected because they are galaxy pairs that are resolved in the *HECATE* but unresolved in *IRAS-RBGS*: NGC 3395/6, NGC 4038/9, NGC 4568/7, ESO 255-IG007, NGC 4922, ESO 343-IG013, NGC 7592, NGC 3994/5, NGC 5394/5, NGC 6670A/B, ESO 60-IG016, NGC 7752/3, IC 2810, IC 0563/4, IC 4518A/B, NGC 5257/8, UGC 12914/5, NGC 6052, and AM1633-682. In addition, 21 *IRAS-RBGS* objects are not found in the *HECATE* because: (i) their radial velocity exceeds our recession velocity limit $v_{\rm vir}$=14 000 km s^{-1} (18 galaxies), (ii) their object type in *HyperLEDA* is unknown (ESO 221-IG010 and ESO 350-IG038), or (iii) is identified as a star (IRAS F05170+0535).

B4 *RIFSCz*

Before the cross-matching of *RIFSCz* and *HECATE*, we corrected an object designation in *RIFSCz* that was appearing twice in the catalogue (column 'FSCNAME'): two instances of F14012+5434, one of which was corrected to F01339+1532, after manual inspection using the provided coordinates. We associate the *HECATE* objects (without associations to *IRAS-RBGS*) to *RIFSCz* objects, if the D_{25} elliptical region of the former and the 6 arcsec circle (i.e. the resolution of *IRAS*) of the latter overlap. We find thousands of multiple matches. In order to resolve the multiple matches, we apply a four-step procedure as follows:

(i) Since *RIFSCz* provides better positional accuracy than *IRAS* (through associations to other surveys such as *2MASS*), we use a matching radius of 3 arcsec to cross-link the *HECATE* and *RIFSCz*. 18 147 matches are accepted, after resolving manually six multiple matches on the basis of radial velocities and offsets of the matched sources.

(ii) For the objects in *HECATE* and *RIFSCz* that remain unmatched after step (i), we use a 6 arcsec match radius for both catalogues. We find 550 matches, after resolving manually six multiple matches with the same criteria as in (i).

(iii) The unmatched objects [after (i) and (ii)] are cross-linked using the D_{25} region in the *HECATE* and the 6 arcsec circle around the position in the *RIFSCz*. Multiple matches are resolved with the requirement that radial velocities match (<100 km s^{-1} difference). 407 matches are found, leaving only 168 unmatched objects in the *HECATE* and 175 in *RIFSCz*.

(iv) The matches of the steps (i)–(iii) are joined and inspected for ambiguous matches; i.e. galaxy pairs may be resolved in the *HECATE* but not in the *RIFSCz*. We reject 22 such associations.

The above steps provide 19 082 unique associations between *HECATE* and *RIFSCz* objects.

B5 *2MASS*

We sequentially cross-match the *HECATE* with the three catalogues providing the *2MASS* data: *2MASS-LGA*, *2MASS-XSC*, and *2MASS-PSC*. This order ensures that the associated photometric data reflect the full extent of the galaxies.

Out of the 665 objects in *2MASS-LGA*, we exclude 35 because they are not galaxies (see https://irsa.ipac.caltech.edu/data/LGA/overview.html). Out of the remaining 620 galaxies, 609 are cross-matched to *HECATE* objects. The unassociated objects were either exceeding the radial velocity criterion (seven objects), or *HyperLEDA* did not classify them as galaxies (three objects), or belonged to the galaxy pair Arp 244 that is resolved in the *HECATE* (1 object).

Then, we cross-match the *HECATE* and the *2MASS-XSC* using a 3 arcsec match radius. From this procedure, we exclude the *HECATE* objects that are already associated with the *2MASS-LGA* galaxies. We also exclude the following extended galaxies that are resolved in the *2MASS*, and would produce thousands of chance coincidence matches: Draco Dwarf, Leo B, Sextans Dwarf Spheroidal, the Magellanic Clouds, and Carina Dwarf Spheroidal. In total, we find 117 713 matches.

Finally, considering objects not associated with either the *2MASS-LGA* or the *2MASS-XSC*, we cross-match *HECATE* and *2MASS-PSC* and find 25 224 matches.

APPENDIX C: EMPIRICAL FORMULAE FOR THE DISTANCES OF THE *HECATE* GALAXIES

The intrinsic distance modulus $\mu_{\rm int}$ and its uncertainty $\epsilon_{\rm int}$ of a galaxy with Virgo-infall corrected radial velocity $v_{\rm vir}$, inferred by the Kernel Regression explained in Section 3.2.2, can be approximated by the following formulae for nVC galaxies:

$$\mu_{\rm int} \approx \begin{cases} 26.34 + 0.006\,057u, & u \le 358.5 \\ 15.74 + 5\log_{10} u, & u > 358.5 \end{cases}, \qquad \text{(C1)}$$

$$\epsilon_{\rm int} \approx 0.2611 + 0.8016 \exp\left(-\tfrac{u}{1341}\right), \qquad \text{(C2)}$$

where $u \equiv v_{\rm vir}/$ km s^{-1}. The above relations are valid for the range $u \in [-481.7, 14, 033.0]$. Similarly, for VC galaxies:

$$\mu_{\rm int,VC} \approx 31.08 + 9.177 \times 10^{-8} u^2, \qquad \text{(C3)}$$

$$\epsilon_{\rm int,VC} \approx 0.3235 + 6.464 \times 10^{-5} u, \qquad \text{(C4)}$$

valid for the range $u \in [-792.5, 2764]$. These approximating formulae for the distance modulus $\mu_{\rm int}$ and the $3\epsilon_{\rm int}$ region are plotted in Fig. C1, on top of the corresponding quantities computed using the regression models.

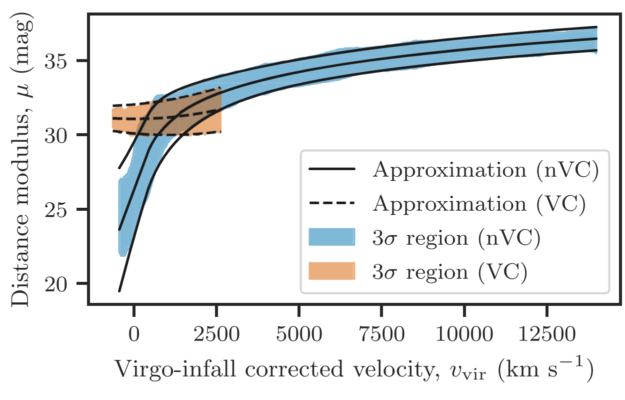


Figure C1. The local 99.7 per cent intervals of the distance modulus of the two models, orange for the Virgo cluster and blue for the rest. The black lines correspond to the mean and same interval (dashed for VC model) computed using the approximation formulae (Appendix C).

APPENDIX D: DESCRIPTION OF COLUMNS IN THE PROVIDED CATALOGUE

The columns of the *HECATE* are described in Table D1.

Table D1. Description of the columns in the machine-readable catalogue. In cases of adopted values from external catalogues, the middle column reports the source: H=*HyperLEDA*, N=*NED*, I=*IRAS*, F=*WISE* forced photometry, M=*2MASS*, S=*SDSS*, and G=*GSWLC-2*. Unflagged columns were computed by us.

Column	Flag	Description
`pgc`, `objname`	H	*Principal Catalogue of Galaxies* number, and object name in the *HyperLEDA*.
`id_ned`, `id_nedd`	N	Name in *NED* and *NED-D*, respectively.
`id_iras`	I	Name in *IRAS-RBGS*, or in *RIFSCz* if in the form Fxxxxx+xxxx.
`id_2mass`	M	ID in *2MASS-LGA*, *2MASS-XSC*, or *2MASS-PSC* (see `flag_2mass`).
`sdss_photid`, `sdss_specid`	S	*SDSS* photometric and spectroscopic IDs (consistent with DR8 and later releases).
`ra`, `dec`	H	Decimal J2000.0 equatorial coordinates (deg).
`f_astrom`	H	Astrometric precision flag. −1 for ∼0.1 arcsec; 0 for ∼1 arcsec; 1 for ∼10 arcsec; and so on.
`r1`, `r2`, `pa`	H	D_{25} semimajor and semiminor axes (arcmin), and North-to-Northeast position angle (deg).
`rsource`, `rflag`	–	Source (see Table B1) and flag of the size information: 0=missing, 1=all size information defined, 2=either `r2` or `pa` were missing and they were set equal to `r1` and 0.0, respectively (circular isophote).
`t`, `e_t`	H	Numerical Hubble type and its uncertainty. See de Vaucouleurs, de Vaucouleurs & Corwin (1976).
`incl`	H	Inclination (deg).
`v`, `e_v`	HN	Heliocentric radial velocity and its uncertainty (km s^{-1}).
`v_vir`, `e_v_vir`	–	Virgo-infall corrected radial velocity and its uncertainty (km s^{-1}).
`ndist`	–	Number of distance measurements in *NED-D* used for the computation of `d`.
`edist`	–	If True, the *NED-D* distance measurements had uncertainties.
`d`, `e_d`	–	Distance and its uncertainty (Mpc).
`d_lo68`, `d_hi68`, `d_lo95`, `d_hi95`	–	68 per cent and 95 per cent confidence intervals of the distance.
`dmethod`	–	Method for the estimation of the distance: N=from *NED-D*, Z=regressor, Zv=VC-regressor, C(v)=distance from *NED-D* but uncertainty from (VC-)regressor.
`ut`, `bt`, `vt`, `it`	H	Total *U*, *B*, *V*, and *I* apparent magnitudes (mag).
`e_ut`, `e_bt`, `e_vt`, `e_it`	H	Uncertainties on `ut`, `bt`, `vt`, `it` (mag).
`ag`, `ai`	H	Galactic and intrinsic absorption in the *B* band.
`s12`, `s25`, `s60`, `s100`	I	*IRAS* fluxes at 12, 25, 60, and 100 µm, respectively (Jy).
`q12`, `q25`, `q60`, `q100`	I	Quality flags for `s12`, `s25`, `s60`, `s100`: 0=not in *IRAS*, 1=upper limit, 2=moderate quality, 3=high quality in *FSC* or 4=flux from *RBGS*.
`wf1`, `wf2`, `wf3`, `wf4`	F	3.3, 4.6, 12, and 22 µm fluxes in the *WISE* forced photometry catalogue (mag).
`e_wf1`, `e_wf2`, `e_wf3`, `e_wf4`	F	Uncertainties on `wf1`, `wf2`, `wf3`, and `wf4` (mag).
`wfpoint`, `wftreat`	F	'True' if point source, and 'True' if treated as such, respectively, in the *WISE* forced photometry catalogue.
`j`, `h`, `k`	M	*J*-, *H*-, and K_s-band apparent magnitudes in *2MASS* (mag).
`e_j`, `e_h`, `e_k`	M	Uncertainties on `j`, `h`, and `k` (mag).
`flag_2mass`	–	Source of the *2MASS* ID and JHK magnitudes: 0=none, 1=*LGA*, 2=*XSC*, 3=*PSC*.
`u`, `g`, `r`, `i`, `z`	S	*u*-, *g*-, *r*-, *i*-, and *z*-band apparent magnitudes in *SDSS* (mag).
`e_u`, `e_g`, `e_r`, `e_i`, `e_z`	S	Uncertainties on `u`, `g`, `r`, `i`, and `z` (mag).
`logL_TIR`	–	Decimal logarithm of the TIR luminosity ($L_\odot$=3.83 × 10^{33} erg s^{-1}).
`logL_FIR`	–	Decimal logarithm of the FIR luminosity ($L_\odot$).
`logL_60u`	–	Decimal logarithm of the 60 µm-band luminosity ($L_\odot$).
`logL_12u`	–	Decimal logarithm of the 12 µm-band luminosity ($L_\odot$).
`logL_22u`	–	Decimal logarithm of the 22 µm-band luminosity ($L_\odot$).
`logL_K`	–	Decimal logarithm of the K_s-band luminosity ($L_\odot$).
`ML_ratio`	–	Mass-to-light ratio (Section 4.2.1).
`logSFR_TIR`	–	Decimal logarithm of the TIR-based SFR estimate ($M_\odot$ yr^{-1}).
`logSFR_FIR`	–	Decimal logarithm of the FIR-based SFR estimate ($M_\odot$ yr^{-1}).
`logSFR_60u`	–	Decimal logarithm of the 60 µm-based SFR estimate ($M_\odot$ yr^{-1}).
`logSFR_12u`	–	Decimal logarithm of the W3-based SFR estimate ($M_\odot$ yr^{-1}).
`logSFR_22u`	–	Decimal logarithm of the W4-based SFR estimate ($M_\odot$ yr^{-1}).
`logSFR_HEC`	–	Homogenized log SFR ($M_\odot$ yr^{-1}). Rescaling of SFR indicators is performed only here (Section 4.2.2).
`SFR_HEC_flag`	–	Flag indicating photometry source and SFR indicator used for `logSFR_HEC` (Table 2).
`logM_HEC`	–	Decimal logarithm of the $M_\star$ ($M_\odot$).
`logSFR_GSW`	G	Decimal logarithm of the SFR in *GSWLC-2* ($M_\odot$ yr^{-1}).
`logM_GSW`	G	Decimal logarithm of the $M_\star$ in *GSWLC-2* ($M_\odot$).
`min_snr`	–	Minimum signal-to-noise ratio of the emission lines used for the activity classification (`class_sp`).
`metal`, `flag_metal`	–	Metallicity [12 + log (O/H)] and its quality flag (Section 4.2.3).
`class_sp`	–	Nuclear activity classification (Section 4.2.4): 0=star forming, 1=Seyfert, 2=LINER, 3=composite, −1=unknown.
`agn_s17`	E	AGN classification in She et al. (2017): Y=AGN, N=non-AGN, ?=unknown.
`agn_hec`	–	Combination of *SDSS* and She et al. (2017) classifications (Section 4.2.4): Y=AGN, N=non-AGN, ?=unknown.

This paper has been typeset from a TEX/LATEX file prepared by the author.