Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **493,** 922–929 (2020)
Advance Access publication 2020 January 29

doi:10.1093/mnras/staa227

Real galaxy mergers from galaxy pair catalogues

Hugo Pfister,[1,2★†] Massimo Dotti,[3,4] Clotilde Laigle,[5] Yohan Dubois[5] and Marta Volonteri[5]

[1]*DARK, Niels Bohr Institute, University of Copenhagen, Copenhagen, 2100, Denmark*
[2]*University of Hong-Kong, Department of Physics, Hong-Kong, China*
[3]*Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano–Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[4]*INFN, Sezione Milano–Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[5]*Institut d'Astrophysique de Paris, Sorbonne Université, CNRS, UMR 7095, 98 bis bd Arago, F-75014 Paris, France*

Accepted 2020 January 23. Received 2020 January 22; in original form 2019 November 11

ABSTRACT

Mergers of galaxies are extremely violent events shaping their evolution. Such events are thought to trigger starbursts and, possibly, black hole accretion. Nonetheless, it is still not clear how to know the fate of a galaxy pair from the data available at a given time, limiting our ability to constrain the exact role of mergers. In this paper we use the light-cone of the HORIZON-AGN simulation, for which we know the fate of each pair, to test three selection processes aiming at identifying true merging pairs. We find that the simplest one (selecting objects within two thresholds on projected distance d and redshift difference Δz) gives similar results than the most complex one (based on a neural network analysing d, Δz, redshift of the primary, masses/star formation rates/aspect ratio of both galaxies). Our best thresholds are $d_{\rm th} \sim 100$ kpc and $\Delta z_{\rm th} \sim 10^{-3}$, in agreement with recent results.

Key words: galaxies: evolution – galaxies: kinematics and dynamics – methods: observational.

1 INTRODUCTION

Galaxy interactions and mergers have been advocated as one of the principal actors in galaxy evolution. Toomre (1977) proposed mergers as responsible for the fast morphological transformation of disc galaxies into spheroids or, in less dramatic cases, for the growth of massive classical bulges (Hopkins et al. 2009a,b). Although this is still a debated result (Fensch et al. 2017; Lofthouse et al. 2017), gas-rich mergers have been proposed as triggers of intense bursts of star formation (Barnes & Hernquist 1991; Mihos & Hernquist 1996; Cox et al. 2008; Calabrò et al. 2019) resulting in luminous and ultra-luminous infrared galaxies (Sanders et al. 1988; Duc, Mirabel & Maza 1997; Elbaz & Cesarsky 2003), as well as triggers of high luminosity single and double active galactic nuclei (AGNs, Di Matteo, Springel & Hernquist 2005; Capelo et al. 2015) possibly responsible for the quenching of star formation in the remnant (Sijacki & Springel 2006; Di Matteo et al. 2008; Booth & Schaye 2009; Dubois et al. 2013). Mergers can also lead to galaxy spin flip (from aligned to perpendicular) along filaments, therefore they bring diversity in the intrinsic alignment pattern (e.g. Welker et al. 2014, 2019). Finally, mergers of massive galaxies are the natural path to the formation of massive black hole pairs and binaries (Begelman, Blandford & Rees 1980; Tremmel et al. 2015, 2017, 2018; Pfister et al. 2017, 2019; Bellovary et al. 2019). If the interaction with their complex environment leads the two black holes to separation of $10^{-3}(M_{\rm binary}/10^6\,{\rm M}_\odot)^{0.75}$ pc,[1] with $M_{\rm binary}$ the total mass of the black hole binary, it can further shrink and finally coalesce in less than an Hubble time, while emitting gravitational waves detectable by current and future observational campaigns (Hobbs et al. 2010; Amaro-Seoane et al. 2013, 2017; Babak et al. 2016).

For all these reasons, galaxy mergers and their consequences have been explored thoroughly from a theoretical point of view, both analysing and post-processing the outcomes of coarse but large cosmological simulations (e.g. Steinborn et al. 2016; Volonteri et al. 2016), as well as higher resolution isolated mergers starting from idealized initial conditions (e.g. Capelo et al. 2015).

In order to confirm these results from an observational perspective, it is required to know which galaxies are going to merge. Two main methods have been used to obtain this information. The first relies on identifying perturbations in galaxy morphology due to mergers (Le Fèvre et al. 2000; Conselice et al. 2003; Lotz et al. 2008; Goulding et al. 2018). The second method, which we will study in more detail in this paper, is pair counting (Zepf & Koo

★ E-mail: hugo.pfister@nbi.ku.dk
† Sophie and Tycho Brahe Fellow.

[1]This estimate applies to close to equal mass circular binaries, for the discussion on the actual dependencies on the eccentricity and mass ratio see equation 2 in Dotti, Sesana & Decarli (2012).

1989; Le Fèvre et al. 2000; Snyder et al. 2017, 2019; Duncan et al. 2019; Ventou et al. 2019): a pair is selected as 'merging' if the relative projected distance (d) and redshift difference (Δz) of the two galaxies are smaller than given thresholds $d_{\rm th}$ and $\Delta z_{\rm th}$. Both methods have their advantages and drawbacks, in principle, the first one uses all the information in the images, but it requires very high resolution and therefore cannot be applied at high redshift. Pair counting uses 'less' information and can be applied to higher redshift, but pairs with a large real 3D separations, which will not merge nor interact within a Hubble time, could be selected. The last point naturally raises the question of the optimal thresholds as well as the dependence of these thresholds with other parameters such as the masses, the mass ratio etc.

In this study we take full advantage of the results of the HORIZON-AGN cosmological simulation (Dubois et al. 2014) to build mock catalogues of observationally selected galaxy pairs and define the best technique to select merging pairs. In Section 2, we detail how we construct this catalogue and compare it with similar catalogues (Snyder et al. 2017); in Section 3 we test three different algorithms to detect pairs and compare their efficiency; we finally give our conclusions in Section 4.

2 BUILD A NUMERICAL CATALOGUE

Our aim is to build a catalogue of galaxy pairs, as an observer would do, but knowing, for a given pair, if it will merge or not. Here we detail how we build this catalogue.[2] In Section 2.1, we present the different data available we used; we then describe our method to build the catalogue in Section 2.2; finally, in Section 2.3, we compare this catalogue with results from Snyder et al. (2017) to verify its behaviour.

2.1 Available data

We use the data from the HORIZON-AGN simulation (Dubois et al. 2014). This is one of the largest hydrodynamical cosmological simulation available, the box size is 140 Mpc at $z = 0$, with 1 kpc resolution in the most refined regions and a dark matter particle mass of 8×10^7 M$_\odot$. It has been run with the adaptive mesh refinement code RAMSES (Teyssier 2002), and contains state-of-the-art galaxy formation subgrid physics: cooling (Sutherland & Dopita 1993), background UV heating (Haardt & Madau 1996), star formation (Rasera & Teyssier 2006) resulting in a stellar particle mass of 2×10^6 M$_\odot$, feedback (stellar winds, type II and type Ia supernovae) and black hole formation, accretion, and feedback (Dubois et al. 2012). HORIZON-AGN reproduces many properties of real galaxies (Dubois et al. 2016; Volonteri et al. 2016; Kaviraj et al. 2017), therefore we can use it to produce mock catalogues, from which we can derive realistic methods observers could use to interpret the data they collect.

2.1.1 *Galaxies in the light-cone*

Concentric shells centred on a fiducial observer located at the origin of the simulation box at $z = 0$, and containing particles (dark matter, stars, and black holes) as well as gas cells, have been extracted on the fly at each coarse time-step of the simulation. This allows the creation of a light-cone (Pichon et al. 2010; Gouin et al. 2019) as sketched in Fig. 1 (top). The opening angle is 2.25 deg from $z = 0$ to $z = 1$, corresponding to the angular size of the full simulation box at $z = 1$.

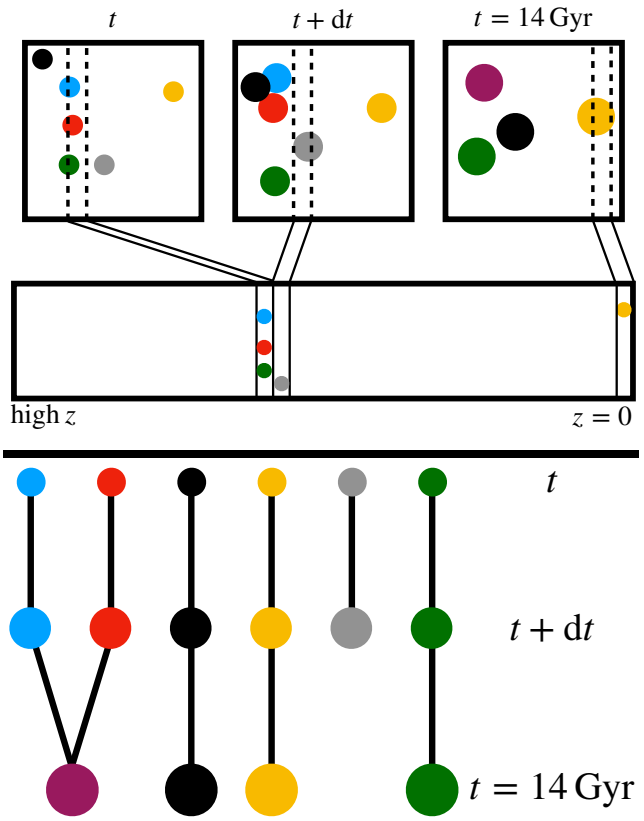

Figure 1. Top: Sketch of the construction of a light-cone. Squares on the top line represent the simulated box, which is evolved in time. Slices of the box are stored at each time-step and are then stacked to form the light-cone (bottom line). Note that some galaxies are present at all time in the box, but not in the light-cone, and that some galaxies (the black one here) can be in the box without being in the light-cone. **Bottom:** Merger tree associated with the simulation sketched, some galaxies merge (blue-red pair), some remain isolated for a long time (yellow, black, and green) and some 'dissolve' as they have no child identified (grey).

A catalogue of galaxies has been extracted from this light-cone (Laigle et al. 2017) containing, in particular, the following information for each galaxy:

(i) stellar mass M;
(ii) star formation rate SFR;
(iii) aspect ratio γ as seen in the light-cone, defined as the ratio between the semiminor and semimajor axis;
(iv) location on the sky with right ascension and declination;
(v) observed redshift, z, corresponding to the redshift an observer would measure from a spectroscopic data set.

2.1.2 *Galaxies in the box*

Galaxies in the box have been identified with ADAPTAHOP (Aubert, Pichon & Colombi 2004). The algorithm detects gravitationally bound structures containing at least 50 stellar particles, therefore having a minimum mass of 10^8 M$_\odot$. Using again the sketch in Fig. 1 (bottom), the blue, red, green, grey, and yellow dots (galaxies) are now identified both in the box and in the light-cone. However, initially, galaxies in the light-cone and in snapshots are not matched.

[2]Please contact the corresponding author if you are interested in obtaining the catalogue.

This matching is important because, for galaxies in the light-cone, similarly to galaxies in the sky, we have only an image at one particular time. Galaxies in the box are instead consistently evolved from $z = 100$ down to $z = 0$, therefore we know their history, e.g. we know how they move or how their mass evolves (as represented by the enhancement of the size of the dots). Each galaxy that can be observed in the light-cone has been associated with the same galaxy in the box (Laigle et al. 2017), connecting the 'observational view' to how a galaxy actually evolves over cosmic time.

2.1.3 *Merger tree*

The merger tree of galaxies in the box has been produced with TREEMAKER (Tweed et al. 2009). Galaxies containing particles with the same ID at different times are matched to form the history of each galaxy. With this we can follow galaxies from their birth down to $z = 0$. This is sketched in Fig. 1 (bottom) where galaxies are followed with time, some merging (blue-red pair), some remaining isolated (black, yellow, and green dots) and some 'dissolving' (grey dot, see Section 2.2).

2.2 Combining data

With all the data presented in the previous sections, for a pair of galaxies in the light-cone, we have a pair of galaxies in the box, which we can follow down to $z = 0$ with the merger tree to see if they merge, or not, and how long it takes if it is actually the case. We select all the pairs (in the light-cone) fulfilling the following criteria:

(i) Both galaxies must be observed at redshift $0.05 < z < 1$, this is the redshift range from which the light-cone of HORIZON-AGN with angular opening 2.25 deg has been produced.

(ii) The mass of the most massive galaxy has to be larger than $10^9\,M_\odot$, so that it is defined with at least 500 stellar particles, and lower that $10^{11}\,M_\odot$, so that there are more than 10 of those galaxies in the catalog. We also impose a stellar mass ratio between galaxies of $0.1 < q < 1$.

(iii) The projected distance between the two galaxies, d, measured with the angular distance assuming that the redshift is the one of the primary galaxy, has to be lower than 5 Mpc. Similarly, the redshift difference between the two galaxies, Δz, has to be lower than 0.05. These criteria are intentionally extremely loose to ensure that most merging pairs are included. Consistently with the simulation, we assume a ΛCDM cosmology with *WMAP7* parameters (Komatsu et al. 2011).

We end up with $\sim 9 \times 10^8$ pairs of galaxies, for which we know the following observational quantities: the mass of the primary and mass ratio, $M_{\rm pri}$ and q; the redshift difference, Δz; the projected distance, d; their SFR, $\mathrm{SFR_{pri}}$ and $\mathrm{SFR_{sec}}$; and their aspect ratios, $\gamma_{\rm pri}$ and $\gamma_{\rm sec}$. We also know the associated pair in the snapshots, which we can follow in the merger tree. We use the sketch in Fig. 1 (bottom) to list the possible cases:

(i) The two galaxies live at the same time, i.e. they are in the same snapshot. We can then follow their history in the merger tree and see if the two galaxies merge, and how long it takes ($\tau_{\rm merger}$). For instance, the blue-red pair merges, while the blue-green pair has not merged by $z = 0$ ($\tau_{\rm merger} = \infty$).

(ii) The two galaxies do not live at the same time, i.e. they are not in the same snapshot. We then follow the history of the galaxy with higher redshift until the two galaxies are at the same snapshot, and then apply case (i). For instance we would trace the blue galaxy in the blue-grey pair until time $t + \mathrm{d}t$ and then apply case (i). In this particular example, an additional feature happens: the grey galaxy 'dissolves', this can happen if the galaxy loses enough stars, or is so perturbed, that it is not recognized by ADAPTAHOP in one snapshot. As it is difficult to differentiate between a numerical and a physical disruption, pairs in which a galaxy 'dissolves' are discarded (this represents $\sim 3 \times 10^7$ pairs, a small fraction of the total).

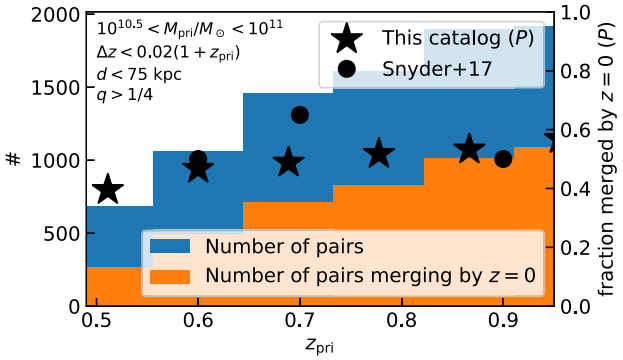

Figure 2. Left axis, histograms: Total number of seletected pairs (blue) and number of pairs actually merging by $z = 0$ (orange). We show in the top left corner the criterion to define a pair. **Right axis, markers**: purity of the selection in our catalogue and in Snyder et al. (2017).

2.3 Validation of the catalogue

To confirm that our catalogue is coherent with previous studies, we perform a similar analysis as done for fig. 2 in Snyder et al. (2017): at a given redshift $z_{\rm pri}$ for the primary, we estimate how many pairs fulfill the criterion observers use to define a merger, i.e. $\Delta z < 0.02(1 + z_{\rm pri}) = \Delta z_{\rm th}$ and $d < 75\,\mathrm{kpc} = d_{\rm th}$. Given this selection process, we can count how many selected pairs actually merge (true positive, TP) and how many selected pairs actually do not merge by $z = 0$ (false positive, FP). From this, we compute the purity P in [0, 1], corresponding to the fraction of selected pairs that actually merge:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\,. \tag{1}$$

We show our results in Fig. 2. Similarly to Snyder et al. (2017), we find that, for $z_{\rm pri} < 1$, about 50 per cent of the pairs selected with $\Delta z < 0.02(1 + z_{\rm pri}) = \Delta z_{\rm th}$ and $d < 75\,\mathrm{kpc} = d_{\rm th}$ will not have merged by $z = 0$. This confirms the robustness of this results, the goodness of our catalogue, and at the same time, shows that there is room from improvement in detecting true mergers from galaxy pairs (Cibinel et al. 2015; Snyder et al. 2019).

3 DETECTION OF REAL GALAXY MERGERS

In this section, we test three algorithms to detect real galaxy mergers from the available properties of each pairs in the catalogue. We first show in Section 3.1 what are the main problems that must be overcome to build a faithful catalogue without losing too many real pairs; we then discuss the metric we will use to judge the quality of the algorithm in Section 3.2; and we finally detail the algorithms in Sections 3.3 and 3.4.

From now on, we will not consider the 'number of pairs which have merged by $z = 0$', since it is not representative of the instantaneous merger rate, we consider instead the 'number of pairs which

Table 1. Four pairs, with similar properties but different behaviours.

Pair ID	$z_{\rm pri}$	$M_{\rm pri}$ 10^{10} M$_\odot$	$M_{\rm sec}$ 10^{10} M$_\odot$	d kpc	Δz 10^{-4}	$\tau_{\rm merger}$ Gyr
1	0.4578	2.307	2.05	123	0.4	1.977
2	0.8996	1.293	0.485	125	0.4	∞
3	0.711	3.588	0.412	349	0.4	0.738
4	0.7542	1.554	0.503	37	2.0	6.427

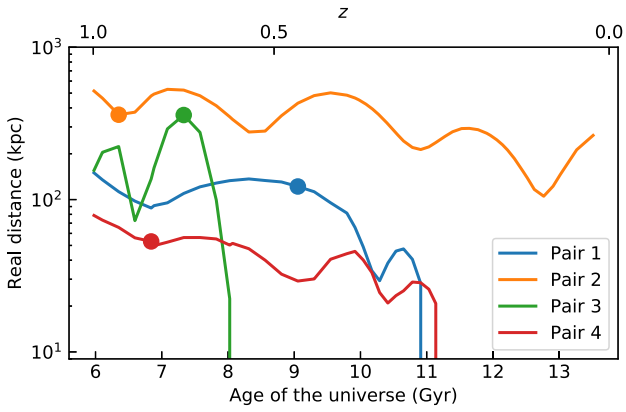

Figure 3. 3D distance as a function of time between the two galaxies of the four pairs described in Table 1. Markers indicate the time at which the pair is seen in the light-cone.

merge within 3 Gyr', meaning that $\tau_{\rm merger} < 3\,{\rm Gyr} = \tau_{\rm merger,max}$. Note that for pairs with $z < 0.25$, the time left before $z = 0$ is less than 3 Gyr, in that case we consider indeed 'pairs which have merged by $z = 0$'. The value of $\tau_{\rm merger,max} = 3$ Gyr has been chosen because it is in agreement with typical merger time-scales obtained in numerical simulations (Capelo et al. 2015), but we stress that its exact value is rather arbitrary, and partially affects the results as shown in Appendix A.

3.1 A difficult exercise

Of the 9×10^8 pairs, only $\sim 10^5$, i.e. only $\sim$0.01 per cent, merge: the problem of detecting merging pairs is extremely unbalanced. This fraction depends on the particular parameters we used to select pairs (we do not expect to have mergers for $d \sim 5$ Mpc), but it is expected to be always low, as most pairs in the sky do not merge.

In addition, the problem is also extremely degenerate. For instance, we show in Table 1 the details and fate of four specific pairs:

(i) Pairs 1 and 2 consist in two pairs, with similar properties in terms of projected distance and redshift difference, but, given our definition of 'merger', one of them merges and the other does not;

(ii) Pairs 3 and 4 consist in two pairs with, in both cases, two galaxies very close in redshift space ($\Delta z \lesssim 10^{-3}$) but, in one case, the two galaxies are far given the projected threshold usually used (projected distance d is 349 kpc) and, in the other case, they are close (d is 37 kpc); none the less, the distant pair merges whereas the close one does not.

We show in Fig. 3 the 3D distance (solid line), as measured in the simulation, between the two galaxies in the four pairs described in Table 1, as a function of time. We also indicate at which time the pair is 'seen' in the light-cone (marker). Pairs 1 and 2 have the same observed d and Δz but have in reality different orbital parameters. Galaxies in Pair 1 are separated by indeed $\sim$100 kpc and a relative speed of 100 km s^{-1}, but galaxies in Pair 2 are in fact separated by 400 kpc and a relative speed of 800 km s^{-1}. These different orbital parameters lead to a different fate, and a different merger time-scale. Pair 3 is seen at the apocenter of a very eccentric orbit resulting in a fast merger, while Pair 4 is more circular, explaining why the merger takes a longer time.

From these simple examples, it is clear that using only thresholds on projected distance and redshift difference cannot be 100 per cent accurate, other quantities such as the masses (see Section 3.3.2), the shapes, the colors etc. or relations between all these quantities (see Section 3.4) should be used (see also Snyder et al. 2017; Goulding et al. 2018; Snyder et al. 2019).

3.2 Goodness of the detection method

To compare two selection methods, and judge which one is the best, we need a metric. Purity (see Section 2.3) only is not a good metric, as a very restrictive threshold (very small $d_{\rm th}$ and $\Delta z_{\rm th}$) would result in a purity of 100 per cent, but would miss many true mergers. This is why we also consider the completeness C in [0, 1] corresponding to the fraction of true mergers selected:

$$C = \frac{TP}{TP + FN}\,, \tag{2}$$

where FN (false negative) corresponds to the number of true mergers not selected.

Clearly, purity and completeness vary in opposite directions: if the thresholds are very restrictive, as we already said, purity will be high, but completeness will be low, and vice versa. For this reason, we need a combination of P and C or, similarly, of FP, FN, TP, and TN, where TN (true negative) corresponds to the number of non-mergers non-selected. We use the *Matthews correlation coefficient*, MCC (Matthews 1975), defined as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}\,.$$

The MCC is in [−1, 1], 1 meaning that the algorithm gives perfect predictions, 0 meaning that it is random and −1 meaning it is always wrong.

3.3 Using simple thresholds on $d_{\rm th}$ and $\Delta z_{\rm th}$

3.3.1 Starting point

We begin with a simple detection method: a pair is selected and (observationally) defined as merging if its projected distance and redshift difference are lower than the thresholds $\Delta z_{\rm th}$ and $d_{\rm th}$. As discussed in Section 3.1 this method cannot be 100 per cent accurate but it still is a reasonable (and frequently used) starting point.

In this section, we search for the best $\Delta z_{\rm th}$ and $d_{\rm th}$ that optimize the MCC. For this purpose, we vary the two parameters and compute the MCC.

In Fig. 4, we show the MCC given the thresholds used. We marked with a dashed-black line when this metric is at maximum. We find $\Delta z_{\rm th} = 7 \times 10^{-4} \sim 10^{-3}$ and $d_{\rm th} = 86\,{\rm kpc} \sim 100\,{\rm kpc}$. The value of 100 kpc is similar to the threshold used by observers, and it is of the order of magnitude expected: galaxies separated by 10 kpc are very likely to undergo a merging process (the typical scale length of the disc of the Milky Way is 3.5 kpc, Binney & Tremaine 1987); and 1000 kpc would correspond to very distant, probably non-interacting, pairs. The value of 10^{-3} for the redshift difference, hardly achievable with current photometric redshift, is

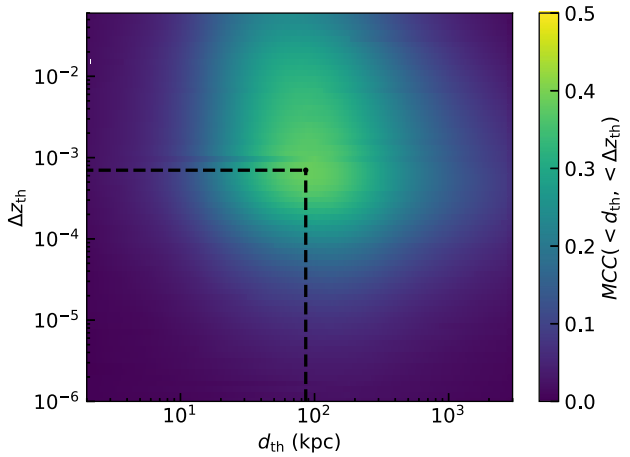

Figure 4. Estimate of the MCC as a function of the thresholds on redshift difference $\Delta z_{\rm th}$ and projected distance $d_{\rm th}$.

however about one order of magnitude lower than the threshold usually chosen, typically $10^{-2}(1 + z_{\rm pri})$. Note that Pasquet et al. (2019) suggested a method allowing to reach 10^{-3} uncertainty on photometric redshift measurement, which is encouraging for the future surveys.

The best thresholds give $P = 0.36$ and $C = 0.41$. This again confirms our first guess of Section 3.1: using only d and Δz is too degenerate to properly distinguish between mergers and non-mergers. While the value of $P = 0.36$ might seems extremely low, we recall here that the catalogue is extremely unbalanced and degenerate, with only 0.01 per cent of pairs actually merging, therefore this selection method is actually three orders of magnitude better than random selection. The maximal MCC = 0.38 is surprisingly good given the simplicity of the method. For example, Snyder et al. (2019), using Random Forest on 10 parameters on pairs from the ILLUSTRIS simulation (Vogelsberger et al. 2014), find an MCC of about 0.4, with the difference that they considered pairs up to $z \sim 9$.

It is interesting to note that, at the time we were writing this paper, Ventou et al. (2019) performed a similar independent analysis to determine the optimal thresholds to detect merging pairs. They use a different simulation (ILLUSTRIS, Vogelsberger et al. 2014), a different redshift range (up to $z = 5$ but with lower redshift resolution as they use six snapshots), and a different metric to chose their threshold (completeness of 30 per cent), none the less, they found similar values, with pairs selected as merging if $d < 50$ kpc and $\Delta z < 10^{-3}$ or 50 kpc $< d <$ 100 kpc and $d < 3 \times 10^{-4}$. This supports our and their findings.

3.3.2 *Including the dependence on* M_{pri}

More massive galaxies are expected to merge more frequently than low-mass galaxies (Fakhouri, Ma & Boylan-Kolchin 2010). This is why we expect the thresholds $\Delta z_{\rm th}$ and $d_{\rm th}$ to depend on the masses of galaxies.

We explore this with our second selection method: a pair is selected and (observationally) defined as merging if its projected distance and redshift difference are lower than the $M_{\rm pri}$ dependent thresholds $\Delta z_{\rm th}(M_{\rm pri})$ and $d_{\rm th}(M_{\rm pri})$.

With this idea in mind, we split the catalogue in sub-catalogues in which $M_{\rm pri}$ is in $[M_{\rm min}, M_{\rm max}]$, where $M_{\rm min}$ ($M_{\rm max}$) varies in equal logarithmic bins (0.25 dex) in between 10^9 and 10^{11} $M_\odot$. We then vary $d_{\rm th}$ and $\Delta z_{\rm th}$, and compute the MCC, as shown on the example

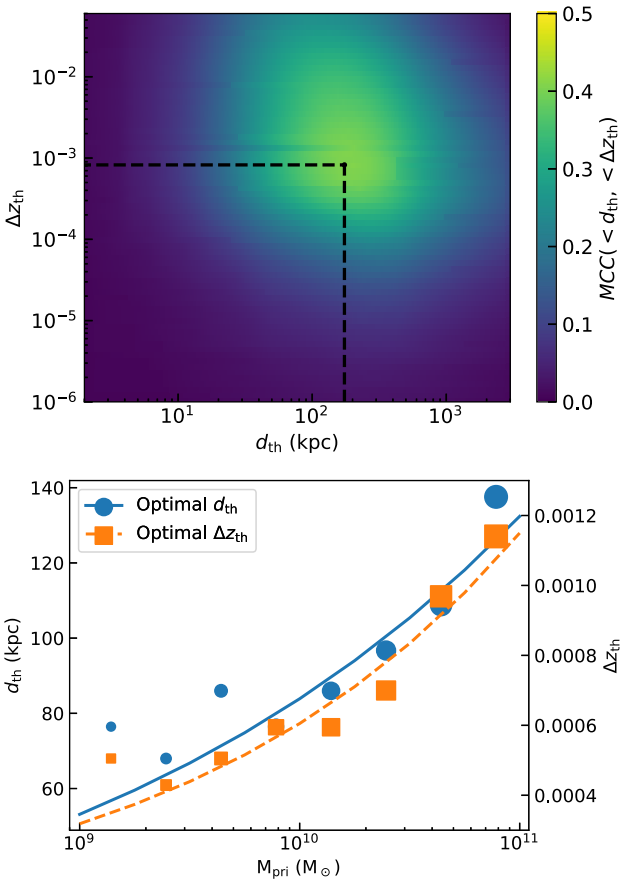

Figure 5. Top: Estimate of the MCC as a function of the thresholds on redshift difference $\Delta z_{\rm th}$ and projected distance $d_{\rm th}$. In this example, $M_{\rm pri}$ is in between 10^{10} and $10^{10.25}$ $M_\odot$. **Bottom:** Evolution of the optimal $d_{\rm th}$ (blue dots) and $\Delta z_{\rm th}$ (orange squares) for $M_{\rm pri}$ within different mass intervals equally spaced by 0.25 dex, as well as their fits reported in equations (3) and (4). The larger the dots the larger the MCC (min: 0.28, max: 0.54).

in Fig. 5 (top). In this example, for a primary mass within 10^{10} and $10^{10.25}$ $M_\odot$, the MCC peaks at 0.46 for $d_{\rm th} = 86$ kpc and $\Delta z_{\rm th} = 6 \times 10^{-4}$: the classification is much better than using only $d_{\rm th}$ and $\Delta z_{\rm th}$. However, both the MCC and the thresholds depend on $M_{\rm pri}$. We show this dependence in Fig. 5 as well as a simple power-law fit for the evolution of the thresholds:

$$d_{\rm th} = 84\,{\rm kpc}\left(\frac{M_{\rm pri}}{10^{10}\,{\rm M}_\odot}\right)^{0.20} \tag{3}$$

$$\Delta z_{\rm th} = 6 \times 10^{-4}\left(\frac{M_{\rm pri}}{10^{10}\,{\rm M}_\odot}\right)^{0.28}. \tag{4}$$

Both thresholds logarithmically scale as $\sim 1/3 \times \log(M_{\rm pri})$. This is expected, indeed, using equation (12) from Taffoni et al. (2003), the dynamical friction time-scale (Chandrasekhar 1943; Binney & Tremaine 1987), which we use as a proxy for $\tau_{\rm merger}$, scales as

$$\tau_{\rm merger} \propto \frac{r^2 v_{c,\rm pri}}{M_{\rm sec}}, \tag{5}$$

where $v_{c,\rm pri}$ is the circular velocity at the virial radius of the primary and r the real 3D distance between the galaxies. Considering constant density $\tilde{\rho}$ ($M_{\rm pri} \sim \tilde{\rho} R_{\rm pri}^3$) and virialized galaxies ($v_{c,\rm pri} \propto M_{\rm pri}^{1/2} R_{\rm pri}^{-1/2}$) immediately leads to $\tau_{\rm merger} \propto M_{\rm pri}^{-2/3} q^{-1} r^2$. In conclusions, for similar mass ratios (q is between 0.1 and 1

in our catalogue), and for a fixed dynamical friction time-scale ($\tau_{\rm merger} < \tau_{\rm merger,max} = 3$ Gyr), we have $r \propto M_{\rm pri}^{1/3}$.

The MCC is higher (∼0.6) for massive galaxies than for low-mass galaxies (∼0.3): it is easier to detect real mergers of massive pairs. However, overall, if we chose $\Delta z_{\rm th}$ and $d_{\rm th}$ given by equations (3) and (4), we find $P = 0.43$, $C = 0.36$, and MCC = 0.40. Note that we have optimistically assumed that $M_{\rm pri}$ is perfectly known, which is not true for real catalogues (uncertainty on mass is typically 0.3 dex, Davidzon et al. 2017). In conclusions, including $M_{\rm pri}$ results in a minor improvement of the classification compared with selection using only Δz and d.

3.4 Using a neural network

We have shown in Section 3.3.2 that using additional information than the projected distance or redshift difference can marginally improve the quality of the detection method. Similarly, non-linear relations between all properties of each pair could improve the quality of the detection method.

We explore this with our third selection method: we build a simple neural network with KERAS (Chollet et al. 2015), which we train so that it detects merging pairs from the properties available in our catalog. Below we describe the main features of the network and the parameters used to ensure the reproducibility of our test. The architecture of the network is somewhat similar to the one from Marchetti et al. (2017), with the following:

(i) An input layer with the nine parameters of each pairs ($d_{\rm th}$, $\Delta z_{\rm th}$, $M_{\rm pri}$, q, $z_{\rm pri}$, SFR$_{\rm pri}$, SFR$_{\rm sec}$, $\gamma_{\rm pri}$, $\gamma_{\rm sec}$).

(ii) A first hidden layer in which five neurons, i.e. five linear combinations of the nine parameters resulting in 45 weights and five bias. In order to introduce non-linearities, the results of these linear combinations are passed to an activation function for which we chose a hyperbolic tangent.[3]

(iii) A second hidden layer, again with five neurons,[4] i.e. five linear combinations of the five outputs of the previous layer resulting in 30 new free parameters. Again, the results are passed to a hyperbolic tangent.

(iv) An output layer with one neuron, i.e. a linear combination of the five outputs of the previous layer resulting in six new free parameters. In order to obtain a number that could be interpreted as a merging probability, the activation function chosen here is a sigmoid returning a real number f in [0, 1], where objects that the neural network considers secure non-mergers correspond to 0 while secure mergers are labelled 1.

The first step is the training of the network. For this task we used a sub-catalogue (referred as *training set*, with 1 per cent of the catalogue, i.e. 10^7 pairs). The training proceeds running through this catalogue multiple times (epochs), and evolving the 86 parameters of the 11 linear combinations computed in the hidden layers. As 'loss function' (the metric used to evaluated how well each set of coefficients performs) we use the binary cross-entropy:

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=0}^{N} (y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)) \ , \qquad (6)$$

where y corresponds to the real labels of the $N = 10^7$ pairs of the *training set*, $\hat{y}$ to the predicted label by the network, and index i refer to a given pair. If a pair is merging (non-merging) then its true label is 1 (0), if the predicted label is 1 (0) then L will be null and if it is 0 (1) then the loss will be infinite. During the training, the network searches for the minimum of the loss using the Adam optimizer (Kingma & Ba 2014), with an initial learning rate (the parameter that determines the size of the steps in the free parameter space) of 0.01. If the loss varies by less than 10^{-4} during 10 epochs, we divide the learning rate by 5 down to a minimum of 10^{-4}, and the training ends when the loss has varied by less than 10^{-4} during 50 epochs.

The second step is the 'validation' of the network, performed on a second *validation set* of 0.5 per cent of the catalogue (5×10^6 pairs). During this phase the network defines a threshold $f_{\rm th}$: every pair resulting in $f > f_{\rm th}$ are considered as merging, while pairs with $f < f_{\rm th}$ are dubbed as chance superpositions. The value of $f_{\rm th}$ is defined by maximizing the MCC on the validation set, consistently with the analysis discussed in the previous sections.

Finally, once the network is trained and validated, we run it on a third *test set* of 0.5 per cent of the catalogue (5×10^6 pairs[5]). The test run performed results in a very low MCC (∼0.1), due to the extreme unbalance of the catalogue (see Section 3.1), that 'teaches' the neural network to typically answer that pairs never merge.

To overcome this issue, we build a new *balanced training set* (2×10^5 pairs) containing 50 per cent of merging pairs and 50 per cent of non-merging pairs. In order to check the good behaviour of our network on a balanced catalog, we also build a *balanced test set* (3×10^4 pairs).

After training on the new balanced set and optimizing the threshold on the unbalanced validation set, we obtain an MCC on the unbalanced test set of 0.41. This is again not much of an improvement compared with naive selection using only Δz and d. The reason is that, among the large number of non-merging pairs, a small fraction have similar properties than merging pairs (we recall that the problem is degenerate, see Section 3.1). However, as the problem is also unbalanced, although this fraction is small, the resulting number of false positive can be larger than the number of merging pairs itself. We show for instance in Fig. 6 the histogram of outputs of the network, f, on mergers (solid lines) and non-mergers (dashed lines) in the case of the balanced (thick line) and unbalanced (thin line) test sets. In both cases more than 90 per cent of mergers (non-mergers) have $f > 0.95$ ($f < 0.05$): the network is perfectly capable of classifying most of pairs. In the case of the balanced test set, where degeneracy is minimized due to balancing, the network is excellent (MCC = 0.97), however, in the case of the real data set, the 0.35 per cent of non-merging pairs with $f > 0.95$, i.e. classified as 'mergers', outnumber the total number of pairs, resulting in an MCC of 0.41. This problem is inherent to the small number of input parameters in the network and, to obtain better results, more inputs should be used: one could think of parameters linked to morphology and/or disturbances, such as the ϕ-asymmetries (Conselice, Bershady & Jangren 2000), the Gini coefficient (Lotz et al. 2008) or other reduced quantities to describe the image. However, given the large size of the data set, using all the pixels of the images as input parameters in a more complex network

[3]We tested a sigmoid activation function as well and found that the neural network behaved best with the hyperbolic tangent.

[4]In both hidden layers, we also tried with 10, 20, 50, and 100 neurons, which resulted in no significant improvement. Above 20 neurons the results actually becomes worse due to overfitting. For these reasons we finally opted for five neurons.

[5]The network has not been run on the whole sample because of the computational cost of the test. Note however that the number of pairs used is large enough for this kind of architecture, Marchetti et al. (2017) typically had a sample with 2.5×10^6 objects.

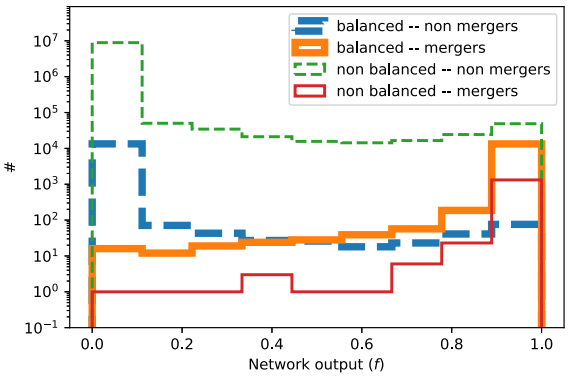

Figure 6. Histogram of outputs from the network of mergers (solid lines) and non-mergers (dashed lines) when applied to the balanced (thick line) and unbalanced (thin line) data set. In most of cases, the network is able to classify correctly the pair, but a small fraction, which in is dominant (as the problem is unbalanced), of non-merging pairs are identified as 'merging' (as the problem is degenerate) which results in poor classification.

would probably be the most efficient way of greatly improve the classification. We postpone such analysis to a future study.

4 CONCLUSIONS

In this paper, using the HORIZON-AGN simulation, we build a mock catalogue of galaxy pairs in order to infer the optimal way to determine true merging pairs. We summarize our finding below:

(i) Using only the projected distance and redshift difference cannot be 100 per cent accurate: we found two pairs with similar projected properties but with different behaviours, some merging some not.

(ii) None the less, the optimal parameters when using only the projected distance and redshift difference are $d_{\rm th} = 86$ kpc and $\Delta z_{\rm th} = 7 \times 10^{-4}$. This result is in excellent agreemenet with the recent results of Ventou et al. (2019). Note that the resulting MCC (the metric we use in this paper) is only 0.38. This is due to a combination of both the degeneracy and unbalancing of the problem.

(iii) More detailed classifiers including the mass of the primary marginally improve the MCC to 0.40. The improvement is much better for massive galaxies, because massive galaxies merge more frequently.

(iv) Including non-linear relations between the nine parameters of each pair in the catalogue (projected distance, redshift difference, masses, redshift, SFRs, and aspect ratios) through a neural network again marginally improve the MCC to 0.41. This confirms that the most relevant parameters to detect merging pairs are the projected distance and redshift difference. It also shows that, in order to be more predictive, future detection methods will need to use the full image instead of reduced quantities.

These new selection criteria can be used in large survey to refine the estimates of the evolution of the galaxy merger rate (e.g. Ventou et al. 2019), but also to study statistically the effects of mergers on the SFR (e.g. Calabrò et al. 2019) or AGNs (e.g. Koss et al. 2012).

ACKNOWLEDGEMENTS

HP is indebted to the Danish National Research Foundation (DNRF132) for support. HP, MV, and YD acknowledge support from the European Research Council (Project no. 614199, 'BLACK'). This work has made use of the Horizon Cluster hosted by the Institut d'Astrophysique de Paris; we thank Stephane Rouberol for running smoothly this cluster for us. HP thanks A. Szenicer for insightful discussions. We thank the referee, Jon Loveday, for carefully reading the manuscript and providing useful comments.

REFERENCES

Amaro-Seoane P. et al., 2013, GW Notes, 6, 4
Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Babak S. et al., 2016, MNRAS, 455, 1665
Barnes J. E., Hernquist L. E., 1991, ApJ, 370, L65
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Bellovary J. M., Cleary C. E., Munshi F., Tremmel M., Christensen C. R., Brooks A., Quinn T. R., 2019, MNRAS, 482, 2913
Binney J., Tremaine S., 1987, Galactic Dynamics, 1st edn. Princeton Series in Astrophysics. Princeton Univ. Press, Princeton, NJ
Booth C. M., Schaye J., 2009, MNRAS, 398, 53
Calabrò A. et al., 2019, A&A, 632, A98
Capelo P. R., Volonteri M., Dotti M., Bellovary J. M., Mayer L., Governato F., 2015, MNRAS, 447, 2123
Chandrasekhar S., 1943, ApJ, 97, 255
Chollet F. et al., 2015, Keras, https://keras.io
Cibinel A. et al., 2015, ApJ, 805, 181
Conselice C. J., Bershady M. A., Jangren A., 2000, ApJ, 529, 886
Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, ApJ, 126, 1183
Cox T. J., Jonsson P., Somerville R. S., Primack J. R., Dekel A., 2008, MNRAS, 384, 386
Davidzon I. et al., 2017, A&A, 605, A70
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Di Matteo P., Combes F., Chilingarian I., Melchior A.-L., Semelin B., 2008, Astron. Nachr., 329, 952
Dotti M., Sesana A., Decarli R., 2012, Adv. Astron., 2012, 940568
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662
Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297
Dubois Y. et al., 2014, MNRAS, 444, 1453
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Duc P.-A., Mirabel I. F., Maza J., 1997, A&AS, 124, 533
Duncan K. et al., 2019, ApJ, 876, 110
Elbaz D., Cesarsky C. J., 2003, Science, 300, 270
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Fensch J. et al., 2017, MNRAS, 465, 1934
Gouin C. et al., 2019, A&A, 626, A72
Goulding A. D. et al., 2018, PASJ, 70, S37
Haardt F., Madau P., 1996, ApJ, 461, 20
Hobbs G. et al., 2010, A millisecond pulsar timing array, ATNF Proposal
Hopkins P. F., Cox T. J., Dutta S. N., Hernquist L., Kormendy J., Lauer T. R., 2009a, ApJ, 181, 135
Hopkins P. F., Lauer T. R., Cox T. J., Hernquist L., Kormendy J., 2009b, ApJ, 181, 486
Kaviraj S. et al., 2017, MNRAS, 467, 4739
Kingma D. P., Ba J., 2014, Adam: A Method for Stochastic Optimization, 3rd International Conference for Learning Representations, San Diego, 2015, preprint (arXiv:1412.6980)
Komatsu E. et al., 2011, ApJS, 192, 18
Koss M., Mushotzky R., Treister E., Veilleux S., Vasudevan R., Trippe M., 2012, ApJ, 746, L22
Laigle C. et al., 2017, MNRAS, 474, 5437
Le Fèvre O. et al., 2000, MNRAS, 311, 565
Lofthouse E. K., Kaviraj S., Conselice C. J., Mortlock A., Hartley W., 2017, MNRAS, 465, 2895
Lotz J. M. et al., 2008, ApJ, 672, 177

Marchetti T., Rossi E. M., Kordopatis G., Brown A. G. A., Rimoldi A., Starkenburg E., Youakim K., Ashley R., 2017, MNRAS, 470, 1388
Matthews B., 1975, Biochim. Biophys. Acta, 405, 442
Mihos J. C., Hernquist L., 1996, ApJ, 464, 641
Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26
Pfister H., Lupi A., Capelo P. R., Volonteri M., Bellovary J. M., Dotti M., 2017, MNRAS, 471, 3646
Pfister H., Volonteri M., Dubois Y., Dotti M., Colpi M., 2019, MNRAS, 486, 101
Pichon C., Thiébaut E., Prunet S., Benabed K., Colombi S., Sousbie T., Teyssier R., 2010, MNRAS, 401, 705
Rasera Y., Teyssier R., 2006, A&A, 445, 1
Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74
Sijacki D., Springel V., 2006, MNRAS, 366, 397
Snyder G. F., Lotz J. M., Rodriguez-Gomez V., Guimarães R. D. S., Torrey P., Hernquist L., 2017, MNRAS, 468, 207
Snyder G. F., Rodriguez-Gomez V., Lotz J. M., Torrey P., Quirk A. C. N., Hernquist L., Vogelsberger M., Freeman P. E., 2019, MNRAS, 486, 3702
Steinborn L. K., Dolag K., Hirschmann M., Remus R.-S., Teklu A. F., 2016, in Sanchez S. F., Morisset C., Delgado-Inglada G., eds, The Interplay between Local and Global Processes in Galaxies, Proceeding of the conference The Interplay between Local and Global Processes in Galaxies, Cozumel, Mexico, 2016-4, p. 34
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
Taffoni G., Mayer L., Colpi M., Governato F., 2003, MNRAS, 341, 434
Teyssier R., 2002, A&A, 385, 337
Toomre A., 1977, in Tinsley B. M., Larson R. B., eds, Evolution of Galaxies and Stellar Populations, Proceedings of a Conference at Yale University, May 19-21, 1977, Yale University Observatory, New Haven, p. 401
Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, MNRAS, 451, 1868
Tremmel M., Karcher M., Governato F., Volonteri M., Quinn T. R., Pontzen A., Anderson L., Bellovary J., 2017, MNRAS, 470, 1121
Tremmel M., Governato F., Volonteri M., Pontzen A., Quinn T. R., 2018, ApJ, 857, L22
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Ventou E. et al., 2019, A&A, 631, A87
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, MNRAS, 460, 2979
Welker C., Devriendt J., Dubois Y., Pichon C., Peirani S., 2014, MNRAS, 445, L46
Welker C. et al., 2019, MNRAS, 2470
Zepf S. E., Koo D. C., 1989, ApJ, 337, 34

APPENDIX A: CHANGING OUR DEFINITION OF MERGER

In Section 3, we specified that a pair was considered as 'merging' if it merges within $\tau_{\rm merger,max} = 3$ Gyr (or within $z = 0$ if the cosmological time left is shorter than 3 Gyr). In this Appendix, we vary $\tau_{\rm merger,max}$ between 1 and 5 Gyr, and see how this affects our results. As we have found our three algorithms to have similar efficiency, we stick to the simplest one (Section 3.3.1) and study how $d_{\rm th}$ and $\Delta z_{\rm th}$ vary with $\tau_{\rm merger,max}$. We show our results in Fig. A1.

$\Delta z_{\rm th}$ is not so affected by $\tau_{\rm merger,max}$, with a mean at $\sim 10^{-3}$, as found in Section 3.3.1, and a standard deviation of 13 per cent. However, $d_{\rm th}$ varies linearly as

$$\frac{d_{\rm th}}{\rm kpc} = 26 \left(\frac{\tau_{\rm merger,max}}{\rm Gyr} \right) + 12 \,. \tag{A1}$$

The MCC is fairly constant, with 2 per cent variations and a mean at 0.38.

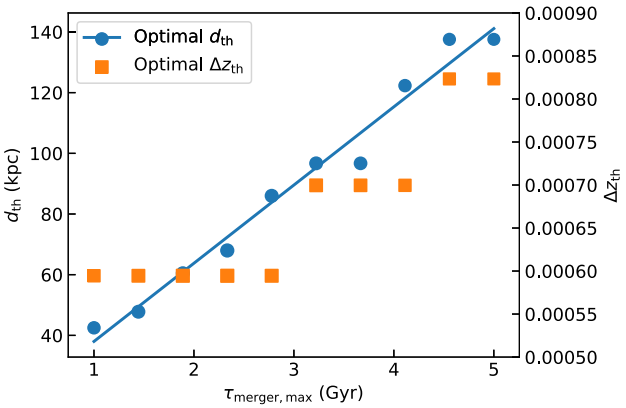

Figure A1. Evolution of $d_{\rm th}$ (blue dots) and $\Delta z_{\rm th}$ (orange squares) with $\tau_{\rm merger,max}$, as well as the fit of equation (A1).

APPENDIX B: CHANGING THE DEPTH OF THE LIGHT-CONE

In Section 2, we specified that we selected galaxies with $z < 1$, however, for some reasons it is possible that real catalogues cannot achieve this maximal redshift. Vice versa, it is possible that real catalogs achieve higher redshifts. In both case, one can wonder if the thresholds have to be changed. In this Appendix, we select galaxies with $z_{\rm max}$ in [0.2, 1] and perform the same analysis as in Section 3.3.1, assuming $\tau_{\rm merger,max} = 3$. We show our results in Fig. B1.

Both for $d_{\rm th}$ and $\Delta z_{\rm th}$, little difference is found with variation of 18 per cent.

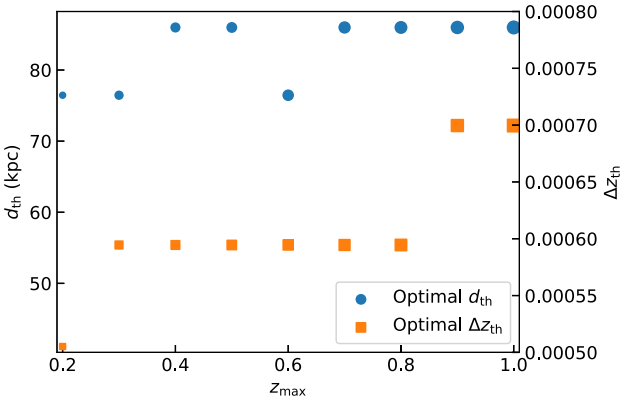

Figure B1. Evolution of $d_{\rm th}$ (blue dots) and $\Delta z_{\rm th}$ (orange squares) with $z_{\rm max}$. The larger the dots the larger the MCC (min: 0.25, max: 0.38).

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