This article has been accepted for publication in MNRAS © 2017 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **473,** 5221–5236 (2018)
Advance Access publication 2017 October 17

doi:10.1093/mnras/stx2701

AMICO: optimized detection of galaxy clusters in photometric surveys

Fabio Bellagamba,[1,2]★ Mauro Roncarelli,[1,2] Matteo Maturi[3] and Lauro Moscardini[1,2,4]

[1]*Dipartimento di Fisica e Astronomia, Università di Bologna, via Gobetti 93/2, I-40129 Bologna, Italy*
[2]*Istituto Nazionale di Astrofisica (INAF) – Osservatorio Astronomico di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy*
[3]*Zentrum für Astronomie, Universitatät Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany*
[4]*INFN – Sezione di Bologna, viale Berti-Pichat 6/2, I-40127 Bologna, Italy*

Accepted 2017 October 13. Received 2017 October 12; in original form 2017 May 8

ABSTRACT

We present Adaptive Matched Identifier of Clustered Objects (AMICO), a new algorithm for the detection of galaxy clusters in photometric surveys. AMICO is based on the Optimal Filtering technique, which allows to maximize the signal-to-noise ratio (S/N) of the clusters. In this work, we focus on the new iterative approach to the extraction of cluster candidates from the map produced by the filter. In particular, we provide a definition of membership probability for the galaxies close to any cluster candidate, which allows us to remove its imprint from the map, allowing the detection of smaller structures. As demonstrated in our tests, this method allows the deblending of close-by and aligned structures in more than 50 per cent of the cases for objects at radial distance equal to $0.5 \times R_{200}$ or redshift distance equal to $2 \times \sigma_z$, being σ_z the typical uncertainty of photometric redshifts. Running AMICO on mocks derived from *N*-body simulations and semi-analytical modelling of the galaxy evolution, we obtain a consistent mass–amplitude relation through the redshift range of $0.3 < z < 1$, with a logarithmic slope of ∼0.55 and a logarithmic scatter of ∼0.14. The fraction of false detections is steeply decreasing with S/N and negligible at S/N > 5.

Key words: methods: data analysis – galaxies: clusters: general – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

Clusters of galaxies were historically recognized as overdensities of galaxies in optical images (e.g. Shapley 1933; Abell 1958). Although large samples of clusters have also been drawn from X-ray observations or through the detection of the Sunyaev-Zel'dovich (SZ) effect on the CMB, photometric observations remain the most promising source of discovery of new clusters of galaxies. This applies especially to the wide ongoing and upcoming photometric surveys such as the Kilo Degree Survey (KiDS; de Jong et al. 2013), the Dark energy Survey (DES; The Dark Energy Survey Collaboration 2005), Pan-STARRS (Kaiser et al. 2002), the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009) and the European Space Agency Cosmic Vision mission *Euclid*[1] (Laureijs et al. 2011). On the other hand, the detection of clusters through photometric data requires non-obvious procedures to distinguish genuine physical groups from clumps generated by chance alignments and non-collapsed overdensities due to the matter distribution on larger scales.

Because of these challenges, and following the evolution in the quality and the amount of data to be analysed, a multitude of automated methods has been developed in the last decades. A possible (but not rigid) classification can be made according to the kind of information the algorithm uses to perform the detection. A first class of methods is purely geometrical; that is, they try to group galaxies from their positions (including photometric redshifts, when available). These methods do not consider explicitly the galaxy observational properties (i.e. magnitudes and colours) in the detection process, although they are accounted for in the photo-z measurement. The first automated approaches, such as those by Shectman (1985) and Lumsden et al. (1992), performed a Gaussian smoothing of galaxy counts in cells. More recent and sophisticated attempts include friends-of-friends algorithms (e.g. Botzler et al. 2004; Trevese et al. 2007; Wen, Han & Liu 2012) and those based on Voronoi tessellation (e.g. Ramella et al. 1999; Soares-Santos et al. 2011).

Another group of methods is instead based on known observational properties of the galaxy population in clusters, and in particular on the presence of a dominant passive component. In practice, clusters are searched as overdensities of galaxies with a similar (red)

★ E-mail: fabio.bellagamba2@unibo.it
[1] http://sci.esa.int/euclid/

© 2017 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

colour. The evolution of the observed colour for this galaxy population encodes the redshift information and makes photo-*z*s redundant in this kind of analysis. This way of searching for clusters was named Cluster Red Sequence method by Gladders & Yee (2000), and then evolved in the MaxBCG algorithm (Hansen et al. 2005; Koester et al. 2007) and finally in redMaPPer (Rykoff et al. 2014, 2016) and RedGOLD (Licitra et al. 2016).

A different approach is the one of Matched Filters (MFs), which dates back to Postman et al. (1996). This approach requires an a-priori definition of the cluster model, usually made up of a radial profile and a luminosity function, which may vary also with redshift: clusters are searched as patterns in the data that resemble the model and are not likely to be produced by random fluctuations. This method evolved in different directions: the Adaptive Matched Filter (Kepner et al. 1999; White & Kochanek 2002; Dong et al. 2008), 3D-MF (Milkeraitis et al. 2010) and the Bayesian Cluster Finder (Ascaso, Wittman & Benítez 2012) that, differently from the others, includes galaxy colours and the presence of a BCG in the likelihood evaluation.

Optimal Filtering is a general technique to extract and measure an expected signal in a data set affected by a noisy background. In astronomy, it was originally applied to weak lensing detection of galaxy clusters by Maturi et al. (2005) and Pace et al. (2007), and to X-ray and SZ detections in Pace et al. (2008). In Bellagamba et al. (2011), it was shown that in case of photometric detections, if one assumes that the background is homogeneous, the Optimal Filter actually corresponds to an MF. A more recent application to photometric data of the same method can be found in Radovich et al. (2017). Differently from other methods, the formalism of Optimal Filtering is very generic and allows us to consider (or neglect) in the model any galaxy property, such as magnitudes or colours, the presence of a BCG, or even photometric and morphological classifications. It is the algorithm itself that selects the most relevant properties for the cluster detection in the available data set and thus gives appropriate weights to the galaxies in the catalogue. The extreme flexibility of the Optimal Filtering method makes it suitable for surveys that span a wide range in redshift, differently from other methods that are tailored to a specific kind of data. Moreover, this formalism allows to combine in a consistent way optical data with other observables such as weak gravitational lensing, SZ and X-rays observations.

In this paper, we present Adaptive Matched Identifier of Clustered Objects (AMICO), an improved Optimal Filtering algorithm for cluster detection. Apart from many improvements over the version described in Bellagamba et al. (2011) in terms of usability, robustness and efficiency, it presents new features in the detection procedure itself. In particular, in this work, we will focus on a new method to translate the map derived from the application of the filter to a list of cluster detections, which is the ultimate and the most important output of such an algorithm. The basic idea is to iteratively remove the imprint of candidate clusters from the map, allowing for the detection of other objects at lower signal-to-noise ratio (S/N) in the surroundings. This becomes a critical point when dealing with surveys that are deep enough to detect tens of clusters/groups per square degree, making cluster detection severely limited by confusion. A crucial quantity in this procedure is the membership probability that is calculated for all possible cluster members after each detection. In practice, each galaxy is initially considered part of the background distribution and, as the detection procedure goes on, may be gradually attributed probabilistically to one or more detections. A deblending technique allowing the removal of the candidate clusters has already been applied to SZ detections, but only through the cluster template, instead of using the likely members of the cluster (Pace et al. 2008). This new approach leads to more accurate and self-consistent results. Another important feature of AMICO is the local estimation of the background population which, as we will see, is a key ingredient in getting an unbiased mass proxy for the detections.

A critical point in all filtering methods is the definition of the model or template that is the expected appearance of the clusters in the given data set. On one side, using the knowledge about clusters helps the detection increasing the S/N of poor or distant structures. On the other side, when more and more information is included, it may induce an 'overspecialization' of the filter which is going to target a very specific type of objects leading to a non-straightforward selection function for the cluster sample, or it may produce sub-optimal results if the filter is applied in regions of wavelength or redshift where knowledge about clusters is limited. As stated above, AMICO employs a very general structure, and the cluster model is considered an input parameter of the algorithm. AMICO has the potential to iteratively perform the extraction and the refinement of the model for the cluster from the available data starting from simple assumptions, but we leave this topic for a future paper. In the following, we will assume to know the model that describes the expected galaxy distribution in clusters at each redshift.

Many of the key ingredients in the definition of the detection performance of AMICO and of other cluster finders depend, of course, on survey specific quantities, such as its depth, the magnitude errors, the quality of photometric redshifts and the accuracy of their error modelling. Since in this paper we do not aim to make any forecast for specific applications of AMICO to current or future data sets, we instead focus in providing a general presentation ought to be used as a benchmark for results on real data. For this reason, and to keep the analysis as simple as possible, in the catalogues used in this paper the photometric redshifts are modelled as simple Gaussian distributions with a realistic width. Nevertheless, as described in the following section, AMICO can deal with arbitrary probability redshift distributions as provided by photometric redshift algorithms.

The paper is organized as follows. In Section 2, we give an overview of the basic principles of Optimal Filtering and we describe the relevant quantities for its application to photometric data. In Section 3, we outline the newly implemented procedure to iteratively extract clusters and their members from the map produced by the application of the filter to the data. In Sections 4, we perform different tests on fields with clusters in specific configurations to highlight the capability of the algorithm in disentangling clusters that are close-by or aligned along the line of sight. In Section 5, we perform the cluster detection in a more realistic field, where clusters have a more diverse population and are embedded in a realistic, non-homogeneous, background. Finally, in Section 6, we summarize the results of our work and make some concluding remarks.

2 OPTIMAL FILTERING OF THE DATA

Optimal Filtering is a general technique to extract and measure an expected signal in a data set that is affected by a noisy background. We refer to Maturi et al. (2005) and Bellagamba et al. (2011) for a more complete introduction about Optimal Filters and their application to cluster detection in photometric surveys, respectively. In this section, we review the relevant quantities for the discussion of the specific implementation of the algorithm in AMICO.

2.1 General principles

The basic assumption of Optimal Filtering is that we can describe a set of data D as the sum of a signal component described by a model M with an unknown normalization A and a noise component N. In the case of a photometric catalogue, the data are represented by the galaxy density $D(\boldsymbol{\theta}, \boldsymbol{m}, z)$ as a function of angular position in the sky $\boldsymbol{\theta}$, of magnitudes $\boldsymbol{m}$ and of redshift z. We can write it as

$$D(\boldsymbol{\theta}, \boldsymbol{m}, z) = A(\boldsymbol{\theta}_c, z_c) \times M_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z) + N(\boldsymbol{m}, z). \quad (1)$$

The model $M_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z)$ is the expected galaxy distribution of a cluster centred in $(\boldsymbol{\theta}_c, z_c)$, the amplitude $A(\boldsymbol{\theta}_c, z_c)$ is the normalization of the cluster galaxy distribution, and the noise $N(\boldsymbol{m}, z)$ is the field galaxy distribution. In this formalism, the array $\boldsymbol{m}$ may include not only photometric magnitudes but also any other observable available for the galaxies in the catalogue, such as morphological type or ellipticity. Here, we will assume the distribution of the field galaxies to be completely random with a uniform mean, i.e. the power spectrum of the large-scale structure is neglected. An improvement over this simplistic assumption will be described in Section 2.6.

Following the general theory of Optimal Filtering, the optimal linear estimate of the amplitude A of the signal component in the position $(\boldsymbol{\theta}_c, z_c)$ is obtained by filtering the data D with a suitable function Ψ_c

$$A(\boldsymbol{\theta}_c, z_c) = \alpha^{-1}(z_c) \int \Psi_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z) \times D(\boldsymbol{\theta}, \boldsymbol{m}, z)\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z - B(z_c), \quad (2)$$

where α is a normalization constant given by

$$\alpha(z_c) = \int \Psi_c^2(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z) N(\boldsymbol{m}, z_c)\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z, \quad (3)$$

where n is the number of dimensions of the space describing the cluster properties $\boldsymbol{m}$, and B is the background component to be subtracted. Under the assumption that the noise is uniform and is produced by random Poissonian counts of galaxies, the Optimal Filter is given by the ratio of the model M_c over the data noise N:

$$\Psi_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z) = M_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}, z)/N(\boldsymbol{m}, z_c). \quad (4)$$

It is possible to demonstrate that the estimate of equation (2) is optimal in the sense that it is unbiased and with the minimum possible variance.

2.2 Applying the filter

Dealing with a catalogue of galaxies, it is convenient to rewrite equation (2) as

$$A(\boldsymbol{\theta}_c, z_c) = \alpha^{-1}(z_c)\, S(\boldsymbol{\theta}_c, z_c) - B(z_c), \quad (5)$$

where S, defined as

$$S(\boldsymbol{\theta}_c, z_c) = \sum_{i=1}^{N_{\rm gal}} \frac{M_c(\boldsymbol{\theta}_i - \boldsymbol{\theta}_c, \boldsymbol{m}_i)\, p_i(z_c)}{N(\boldsymbol{m}_i, z_c)}, \quad (6)$$

is the analogue of the integral of equation (2) for discrete data, and the subscript i runs over the galaxy catalogue. In the model M_c of a cluster at redshift z_c, we are now omitting the redshift distribution of the members, in favour of the photometric redshift distribution $p_i(z)$ for each galaxy. In fact, the intrinsic dispersion in redshift of the cluster members is negligible in comparison with the measurement uncertainty. The units of the model $M_c(r, \boldsymbol{m})$ are then $\mathrm{mag}^{-n}\,\mathrm{deg}^{-2}$, the noise $N(\boldsymbol{m}, z)$ instead has units $\mathrm{mag}^{-n}\,\mathrm{deg}^{-2}\,z^{-1}$, as the data $D(\boldsymbol{\theta}, \boldsymbol{m}, z)$.

The normalization α ensures that the amplitude A is the measure of the cluster signal in units of the model M_c. Equation (3) then becomes

$$\alpha(z_c) = \int \frac{M_c^2(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m})\, q^2(z_c, z)}{N(\boldsymbol{m}, z_c)}\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z. \quad (7)$$

In the previous equation, we introduced $q(z_c, z)$, which is the typical redshift probability distribution for a galaxy which lies at redshift z_c. If no a priori information is available, this distribution can be computed from the data as

$$q(z_c, z) = \left(\sum_{i=1}^{N_{\rm gal}} p_i(z_c) \right)^{-1} \sum_{i=1}^{N_{\rm gal}} p_i(z - z_c + z_{\rm peak}, i) p_i(z_c), \quad (8)$$

where $z_{{\rm peak},i}$ is the most probable redshift for the ith galaxy. Equation (8) is a sort of weighted mean $p(z)$ for galaxies whose true redshift is z_c computed in this way: each galaxy probability distribution is weighted with its value at redshift z_c and then is shifted so that it peaks at z_c. Instead of a generic $q(z_c, z)$ for all objects, it is possible to compute equation (8) for different classes of objects (e.g. as a function of magnitudes) and take this into account in the integral of equation (7) and in the following.

The same statistics on the redshift distribution can be used to calculate an analytical estimate of the background component to be subtracted, which is

$$B(z_c) = \alpha^{-1}(z_c)\beta(z_c), \quad (9)$$

where $\beta(z_c)$ is the expectation value of $S(\boldsymbol{\theta}_c, z_c)$ if the galaxy distribution corresponds to the field component only:

$$\beta(z_c) = \int \frac{M_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m})\, q(z_c, z)}{N(\boldsymbol{m}, z_c)} N(\boldsymbol{m}, z_c)\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z = \int M_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m})\, q(z_c, z)\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z. \quad (10)$$

The subtraction of $B(z_c)$ in equation (5) ensures that $A = 0$ when the galaxy distribution around a given position corresponds to the field one, namely $D(\theta_c, z_c) = N(\theta_c, z_c)$. By construction, $\beta(z_c)$ is also the total number of galaxies in the cluster model at redshift z_c.

It is important to stress that the treatment of photometric redshifts described here is completely generic, and does not assume any specific shape or accuracy of the p(z) distribution. This is different with respect to other MF algorithms, such as Dong et al. (2008) and Milkeraitis et al. (2010). Even in the absence of any redshift information one can apply the formalism described here by simply assuming a flat probability distribution in the considered redshift range.

2.3 Variance of the amplitude estimate

Following the assumptions described in the previous sections, one can derive an analytical estimate of the uncertainty of $A(\boldsymbol{\theta}_c, z_c)$. This is given by

$$\sigma_A^2(\boldsymbol{\theta}_c, z_c) = \alpha^{-1}(z_c) + A(\boldsymbol{\theta}_c, z_c)\, \frac{\gamma(z_c)}{\alpha^2(z_c)}, \quad (11)$$

where the first component derives from the background fluctuations, the second one accounts for the shot-noise contribution of the actual structures present in the data as measured with the filter. Here, the

quantity γ is given by

$$\gamma(z_c) = \int \frac{M_c^3(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m})\, q^3(z_c, z)}{N^2(\boldsymbol{m}, z_c)}\, \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z. \quad (12)$$

By construction, equation (11) is meaningless when A is negative, but this is not important as regions with negative A have a lower-than-average galaxy density, and so are negligible in the cluster detection.

2.4 Likelihood

Given the description of the data D as the sum of a signal component M with normalization A and a noise component N as in equation (1), their likelihood is

$$\mathcal{L}(\boldsymbol{\theta}_c, z_c) = -\int \frac{\{D - [AM_c(\boldsymbol{\theta} - \boldsymbol{\theta}_c, \boldsymbol{m}) + N(\boldsymbol{m}, z_c)]\}^2}{N(\boldsymbol{m}, z_c)} \mathrm{d}^2\theta\, \mathrm{d}^n m\, \mathrm{d}z. \quad (13)$$

Combining the previous equation with the amplitude estimate given by equation (5), one can show that

$$\mathcal{L}(\boldsymbol{\theta}_c, z_c) = \mathcal{L}_0 + A^2(\boldsymbol{\theta}_c, z_c)\, \alpha(z_c), \quad (14)$$

where $\mathcal{L}_0$ is a constant that does not depend on the position. Differently from AMF (Kepner et al. 1999; Dong et al. 2008), we do not consider a Poissonian likelihood, but use instead a Gaussian likelihood, coherently with the construction of the Optimal Filter.

At a given redshift, maxima of A are also maxima of the likelihood. Intuitively, this depends on the fact that, the bigger is the cluster, the bigger is the improvement in the likelihood when the cluster model is centred in that position. However, as noted in Bellagamba et al. (2011), the likelihood alone does not distinguish between overdensities ($A > 0$), which we are interested in, and underdensities with negative A. In the following, to perform the detection we will combine this information with the S/N A/σ_A, which is meaningful only when A is positive.

To summarize, following equations (5), (11) and (14), we can calculate an estimate of the amplitude of the candidate cluster, its uncertainty and the likelihood on a 3D grid of positions in the sky and redshift. We can then derive cluster detections from this map by selecting $(\boldsymbol{\theta}_c, z_c)$ locations with large S/N values and large likelihood. We will detail our procedure to extract detections from the map in Section 3.

We note that, differently from other works (Dong et al. 2008; Ascaso et al. 2012) that centre their cluster candidates on galaxy positions, we use a regular grid in angles and redshift. This choice is helpful for the optimization of the extraction of cluster candidates presented in Section 3. If needed, it is always possible to select the most likely central galaxy a posteriori.

2.5 Masked areas

It is common in galaxy surveys to have field regions with missing data, mainly due to foreground sources (e.g. stars). This is usually handled with the definition of suitable masks. In our formalism, we can account for masked regions by computing the constants α, β and γ as a function of $\boldsymbol{x}_c$ (and not just z_c), performing the integrals in equations (7), (10) and (12) only on the available (unmasked) area around the centre. The derived values of A are then comparable, because they are computed assuming the same model and the noise distribution, but considering the different available area for the amplitude estimate. By construction, if the galaxy distribution is the same, the S/N of an amplitude estimate derived from a partially masked area will be lower than the one obtained when the full area is completely available to observations.

2.6 Local background

In the previous sections, the background component $N(\boldsymbol{m}, z_c)$ has been considered to have a constant mean over the surveyed area. With this approximation, for a given redshift z_c, the contribute of the field population to the amplitude is random with mean equal to $B(z_c)$ and is subtracted as in equation (5). In reality, the large-scale structure of the Universe will produce correlations in the density of galaxies on scales which are larger than the typical cluster scale. In order to appropriately remove the background contribution to the amplitude estimate A, AMICO applies a local correction $f(\boldsymbol{\theta}, z_c)$ to $N(\boldsymbol{m}, z_c)$. For every redshift, this is computed by producing a map of the density of galaxies as a function of the angular coordinates. A k-σ clipping is applied to this map to remove obvious peaks (likely to be produced by clusters and not by the field distribution) and then a smoothing is performed over a scale which is significantly larger than the one of the filter. Finally, by dividing this map by its mean value, we obtain a correction term $f(\boldsymbol{\theta}, z_c)$ for each position.

It is easy to show that one can account for the local background correction a posteriori. In fact, after applying the filter to the galaxy population using the mean $N(m, z_c)$ (equation (6)), the relevant quantities for the corrected background density $f(\boldsymbol{\theta}, z_c)N(m, z_c)$ can be calculated in the following way:

$$A'(\boldsymbol{\theta}_c, z_c) = \frac{S(\boldsymbol{\theta}_c, z_c) - f(\boldsymbol{\theta}, z_c)\beta(z_c)}{\alpha(z_c)}\,, \quad (15)$$

$$\sigma^2_{A'}(\boldsymbol{\theta}_c, z_c) = \frac{f(\boldsymbol{\theta}, z_c)}{\alpha(z_c)} + A'(\boldsymbol{\theta}_c, z_c)\, \frac{\gamma(z_c)}{\alpha^2(z_c)}\,, \quad (16)$$

$$\mathcal{L}'(\boldsymbol{\theta}_c, z_c) = \mathcal{L}_0 + \frac{A^2(\boldsymbol{\theta}_c, z_c)\, \alpha(z_c)}{f(\boldsymbol{\theta}, z_c)}\,. \quad (17)$$

In this step, there are no approximations, the very same results would be obtained by accounting for the local correction from the beginning. We opted for this procedure for efficiency reasons.

2.7 Model

In the formalism described above, the model is an input parameter that describes the expected galaxy distribution in clusters as a function of redshift. In principle, one can construct it in several ways, for example by binning members of known clusters in magnitude and radial distance. For numerical stability and continuity, it is often preferred to use analytic expressions to describe the model. As a cluster galaxy distribution can be considered the sum of different components (e.g. central galaxy and satellites, red and blue populations), the model can be written in a generic form as

$$M_c(r, \boldsymbol{m}) = \sum_i \Psi_i(r)\Phi_i(\boldsymbol{m})\,, \quad (18)$$

where the index i runs over the considered galaxy populations and $\Psi_i(r)$ and $\Phi_i(\boldsymbol{m})$ are the radial and magnitude distributions for the ith component, respectively.

As a representative example of a practical choice for the model, we describe here the model used for the tests presented in Sections 4 and 5 of this paper. In this case, we consider separately the distribution of central and satellite galaxies, and we take into account only one observing band. The central galaxies are modelled with a Gaussian distribution in magnitude. In the radial distribution, they are described by a step distribution larger than zero only in the most

inner bin, as they lie at the centre of the haloes by construction. For the satellites' magnitude distribution, we use a Schechter function (Schechter 1976):

$$\Phi(m) = C \times 10^{-0.4(m-m_\star)(\alpha+1)} \exp\left(-10^{-0.4(m-m_\star)}\right), \tag{19}$$

where $m_\star$ is the typical magnitude, α controls the faint-end slope and C is the normalization. For the satellites' radial profile we use an NFW distribution (Navarro, Frenk & White 1997)

$$\Psi(r) = \frac{B}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2}, \tag{20}$$

where the scale radius is $r_s = R_{200}/c$, c is the concentration parameter and B is the normalization. The value of the parameters as a function of redshift is derived from mock catalogues as described in Section 4.2.

3 SELECTING THE CLUSTER CANDIDATES

The principles of Optimal Filtering, applied to the photometric galaxy survey data, lead to the construction of a 3D map of amplitude estimates $A(\boldsymbol{\theta}, z_c)$, as defined by equation (5). Each value of this map represents an estimate of the amplitude of a cluster signal, if a cluster is actually centred in that position. It is non-trivial to move from the amplitude map to the list of likely clusters, which is the desired output of a detection algorithm. In the following, we will outline how AMICO proceeds in this task. In particular, we will describe the *cleaning* procedure, which aims at removing the signal of the most significant objects from the map in the attempt to detect smaller objects that might be blended with them. This is one of the most important novelties introduced in AMICO with respect to traditional MF algorithms. We will see in Sections 4.4 and 4.5, a practical demonstration of its power.

After the 3D map of amplitude is built as described in Section 2.2, the algorithm proceeds iteratively. The total number of detections $N_{\rm det}$ depends on $(\mathrm{S/N})_{\rm min}$, the minimum S/N one wants to achieve. For j that runs from 1 to $N_{\rm det}$, the following steps are made:

(i) All the pixels with an S/N below the defined threshold $(\mathrm{S/N})_{\rm min}$ are discarded.

(ii) Among the remaining pixels, the one with the highest likelihood is selected. This identifies the jth cluster detection.

(iii) Some relevant properties of the detection are saved, such as centre position $(\boldsymbol{\theta}_j, z_j)$, amplitude $A_j \equiv A(\theta_j, z_j)$, S/N A_j/σ_{A_j} and a first estimate of the number of members N_j, given by $N_j = A_j \times \beta(z_j)$.

(iv) For each galaxy whose sky coordinates lies close to θ_j, a membership probability $P(i \in j)$ is computed as detailed in Section 3.1.

(v) The imprint of its potential members is removed from the map, as it will be detailed in Section 3.2.

Once step (v) is completed, all the values of amplitude, likelihood and S/N are updated and do not contain anymore the contribution of the jth detection (and of the previous ones). The process is repeated iteratively until after step (i) no pixel is left.

3.1 Membership assignment

We now describe the details regarding the step (iv) of our procedure. Given the jth detection at position $\boldsymbol{\theta}_j$ and redshift z_j with amplitude A_j, the probability $P(i \in j)$ of the ith galaxy of being a member of the jth detected object is defined as

$$P(i \in j) \equiv \frac{A_j M_j(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j)}{A_j M_j(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j) + N(\boldsymbol{m}_i, z_j)}, \tag{21}$$

i.e. the ratio between the galaxy density associated with the cluster component and the total one. This definition is strictly correct only if we consider one cluster at a time, while in reality several galaxies will have a positive probability to belong to more than one structure. In principle, equation (21) should be modified to

$$P(i \in j) = \frac{A_j M_j(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j)}{\sum_{k=1}^{N_{det}} A_k M_k(\boldsymbol{\theta}_i - \boldsymbol{\theta}_k, \boldsymbol{m}_i) p_i(z_k) + N(\boldsymbol{m}_i, z_j)}, \tag{22}$$

where the sum at the denominator runs over all the detected structures. However, even if formally correct, the application of this equation in the process outlined above is unfeasible. In fact, the probability must be computed when the jth detection is found and no information is available on the presence of other nearby clusters candidates that will be subsequently detected. Thus, we consider one cluster at a time in the total density, an assumption which is coherent with the measurement of A made via equation (5).

We take into account previous memberships assigned to the same galaxy by introducing the field probability $P_{f,i}$, which is defined as

$$P_{f,i} = 1 - \sum_{k=1}^{j-1} P(i \in k). \tag{23}$$

The field probability for a galaxy is 1 at the beginning of the process, and decreases while the galaxy is progressively attributed to new detected objects. Intuitively, it is a measure of the fraction of the galaxy that is still available for further associations. Then, equation (21) becomes

$$P(i \in j) = P_{f,i} \times \frac{A_j M_j(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j)}{A_j M_c(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j) + N(\boldsymbol{m}_i, z_j)}. \tag{24}$$

This modification also ensures that the sum of the membership probabilities of each galaxy is always ≤ 1.

3.2 Cleaning the detections

Once the probability of a galaxy to belong to a given detected object is calculated, in step (v) of the process summarized above, we can take this knowledge into account when determining the following detections. We do so by removing from all the pixels the contribution of possible cluster members weighted by $P(i \in j)$. This effectively removes the imprint of the detected cluster from all the pixels. In practice, equation (6) becomes

$$S(\boldsymbol{\theta}_j, z_j) = \sum_{i=1}^{N_{\rm gal}} P_{f,i} \frac{M_j(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{m}_i) p_i(z_j)}{N(\boldsymbol{m}_i, z_j)}. \tag{25}$$

From the computational point of view, the cleaning process forces the code to run through all the galaxies every time a new detection is found, and modify the value of the map of $S(\theta_j, z_j)$ in its surroundings. Moreover, following equations (11) and (14) also the values of σ_A and $\mathcal{L}$ must be updated accordingly. This makes the cleaning the computational *bottleneck* of AMICO.

An example of how the cleaning procedure modifies the initial amplitude map is shown in Fig. 1, from the analysis on the mock galaxy catalogues presented in Section 5.

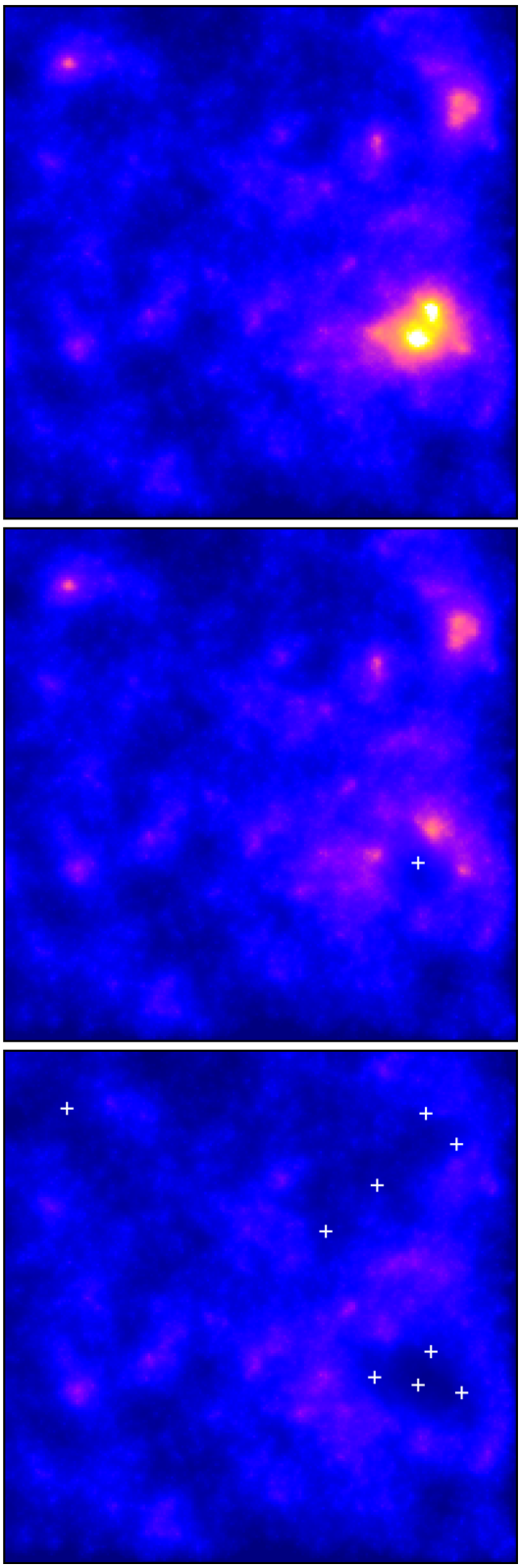

Figure 1. Amplitude map of $1 \times 1\,\mathrm{deg}^2$ at redshift $z = 0.33$ from one of the fields analysed in Section 5. Top panel: initial map produced by the algorithm. Central panel: map after the pixel with the highest S/N has been marked as a detection and its cleaning performed. Bottom panel: map after all the detections with S/N > 7 have been selected and cleaned. White crosses indicate the position of cluster detections removed from the map.

4 TESTS ON IDEAL MOCKS

In order to verify the performance of the algorithm outlined in the previous sections, we perform two sets of tests, which are distinguished by the kind of mock catalogues used. The first set of tests is run over so-called *ideal* mocks, where we place clusters in known positions over a background population with constant mean density. These tests are useful to show how the algorithm works in a completely controlled environment, where all the assumptions of the Optimal Filtering described in Section 2 are perfectly met. The second set of tests is run on *cosmological* mocks, where the galaxy distribution, derived from *N*-body numerical simulations and semi-analytic modelling, is more similar to the one we see in real observations. These tests show the algorithm at work in a more complex environment, where galaxy clusters and groups may have different shapes, luminosity functions, are embedded in the correlated large-scale structure with non-negligible chances of random alignments. All of this makes the identification of clusters more difficult, and the estimation of their properties more uncertain. This second set of tests will be described in Section 5.

4.1 Simulation

The starting point for our tests are the public mock catalogues described in Meneux et al. (2008). We use 50 catalogues, $2 \times 2\,\mathrm{deg}^2$ each, which were originally meant to mimic the VVDS survey (Le Fevre et al. 2003). They were obtained by applying the prescriptions of De Lucia & Blaizot (2007) to the dark-matter halo merging trees extracted from the Millennium simulation (Springel et al. 2005). The catalogues are complete in apparent magnitude up to $I = 25$ for redshift $z > 0.3$. In this section, we extract a cluster model and a noise distribution from these mocks and we use them to create new *ideal* mocks. The analysis of the mocks in their original form is discussed in the next section. The cosmology adopted in the simulation of the mocks is a flat ΛCDM (Λ cold dark matter) model, with $\Omega_m = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$ and $H_0 = 0.73$ km $\mathrm{s}^{-1}\,\mathrm{Mpc}^{-1}$.

4.2 Model and background

The catalogues provide us a halo identification made on the outputs of the simulation. Galaxies that lie at the centre of haloes are marked and for each halo its mass M_{200}[2] is provided. This allows us to build the model distribution $M_c(r, m)$ that describes the typical galaxy content of a cluster of mass $\sim 10^{14}\,\mathrm{M}_\odot\,h^{-1}$ at redshift z_c. To this end, we first select as cluster members the galaxies that are at distance $< 1.35 R_{200} \sim R_{100}$ from the central galaxy of a cluster with $13.8 < \log M_{200}/(\mathrm{M}_\odot\,h^{-1}) < 14.2$. Then, we build the radial distribution (in units of R_{200}) and the magnitude distribution for the cluster satellites in bins of redshift, weighing each galaxy by $w_i = [M_{200}/(10^{14}\,\mathrm{M}_\odot\,h^{-1})]^{-1}$. This weight is used to keep the desired normalization for the cluster model. Finally, in each redshift bin, we fit the derived average radial distribution with an NFW function (equation 20) and the average magnitude distribution with a Schechter function (equation 19).

As an example, we show the empirical distribution and the analytical fitting functions at $z = 0.4$ in Fig. 2. At all redshifts, the mean radial profile is consistent with a concentration $c = 3$, so we

[2] The mass of the haloes is provided in terms of M_{200}, which is the mass contained in a sphere with a density 200 times larger than the critical one. The radius of this sphere is R_{200}.

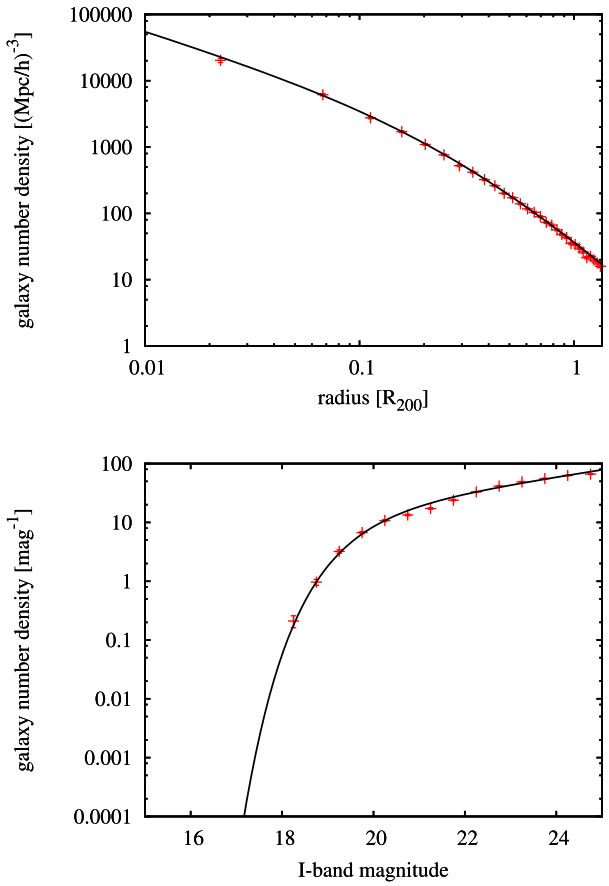

Figure 2. Empirical distributions and analytical fit for the cluster satellites at redshift $z = 0.4$. Top panel: radial profile. Bottom panel: magnitude distribution. The error bars derive from Poissonian statistics, but are often smaller than the point size. These distributions were used to create the idealized mock simulations.

use this value for the model. As for the other parameters (the total number of galaxies N, the typical magnitude $m_\star$, the faint-end slope α), we perform a fit over redshift to get a smoothly changing model. Central galaxies are added separately, following a Gaussian distribution in magnitude and adding one galaxy in the first bin of the radial profile, as they lie by definition exactly at the centre of the structures. This mimics the presence of a BCG in our model. For each redshift z_c, the desired $M_c(r, m)$ is then obtained by projecting the 3D radial distribution in 2D and multiplying it by the magnitude distribution.

The background distribution $N(m, z)$ is instead built from the galaxies that are not closer than $1.35R_{200}$ to any structure with $M_{200} > 0.5 \times 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. This choice is meant to be complementary to the one made in constructing the model $M_c(r, m)$, following the separation in two parts of the galaxy population that is an assumption of the algorithm. Anyway, $N(m, z)$ is very weakly affected by the specific values of the thresholds in mass and radius, as the vast majority of the galaxy population resides in much smaller haloes.

Given the properties of the catalogue, we decided to concentrate our analysis over the redshift range of $0.3 < z < 1$. In fact, at lower redshifts, the catalogue is not complete at the faint magnitudes, while at higher redshifts the galaxy density is so low that one cannot build properly the typical cluster distribution. In the remaining of the paper, we will restrict our analysis to this redshift interval. We note that, even if this work is not aimed at providing forecasts for any

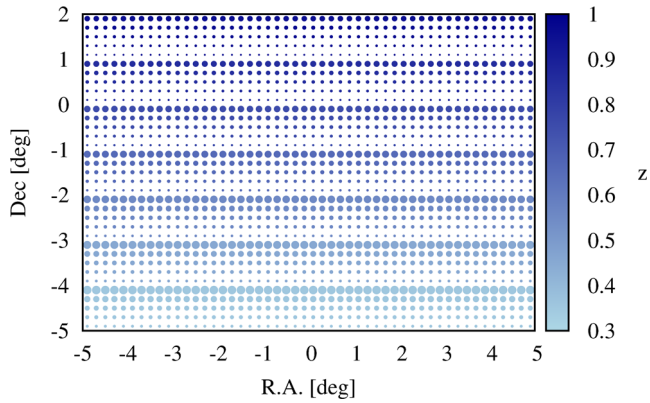

Figure 3. Distribution of clusters in the field described in Section 4.3. The colour indicates the input redshift, as coded by the bar on the right, the radius of the circle is proportional to the cubic root of the number of visible galaxies.

specific survey, this redshift range is common for cluster detection with optical data. AMICO can use the information coming from multiband observations, but in the tests presented here we use only the I band for simplicity.

4.3 Test on isolated clusters

Having built the cluster model and the background component in the redshift range of interest, we can prepare our *ideal* mock catalogues. As a first test case, we prepare a catalogue with structures disposed on a grid over a statistically uniform background. We consider seven redshift values, with z ranging from 0.35 to 0.95, separated by $\Delta z = 0.1$. For each redshift, we create 250 *Monte Carlo* realizations of our cluster model, 50 for each five different amplitude values $A_{\mathrm{true}} = (0.2, 0.5, 1.0, 2.0, 5.0)$. Clusters are distributed over a field with size $10 \times 7\,\mathrm{deg}^2$ on a grid with a distance 0.2 deg one from each other, as shown in Fig. 3. We underline that, following our assumptions, for a given redshift the expected galaxy distribution for clusters with any richness is the same, apart from a rescaling of the total number of galaxies. We add a homogeneous background component, following the density given by $N(m, z)$ and random positions over the field. Here and in the following, we applied a scatter to the redshift of the galaxies following a Gaussian distribution with $\sigma = 0.05(1+z)$, which is typical for wide-band photometric surveys.

Given a cluster model and a background distribution, and having fixed the photometric redshift uncertainty, we can use equations (5), (6) and (11) to estimate the expected S/N for a cluster of any amplitude at any redshift, and build the expected selection function for our cluster search. The result is shown in Fig. 4. This is a purely theoretical estimate that depends on our modellization of the cluster and of the background and allows us to make a first estimate of the kind of objects we can detect under certain survey conditions. In particular, for any redshift we can compute the minimum amplitude A_{min} required for the detection. In the following, we will compare this with the actual detection thresholds in our experiments on ideal and cosmological fields.

4.3.1 Completeness

Once the mock catalogue is produced, we run AMICO on it using the model and the background estimated as above. In this case, as the background distribution is following our assumption of uniformity over the field, we do not use the local estimation presented

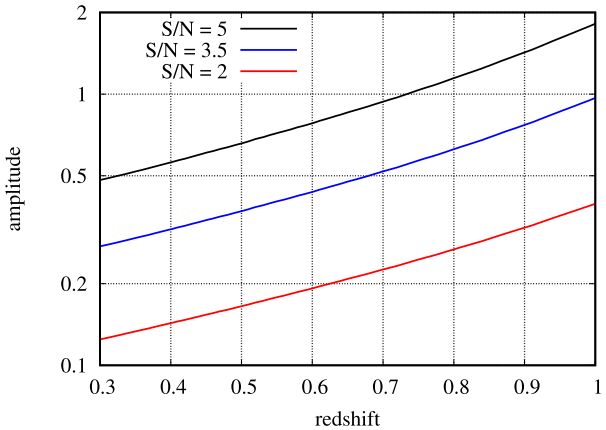

Figure 4. Theoretical minimum amplitude for detection of galaxy clusters as a function of redshift, for three different S/N thresholds.

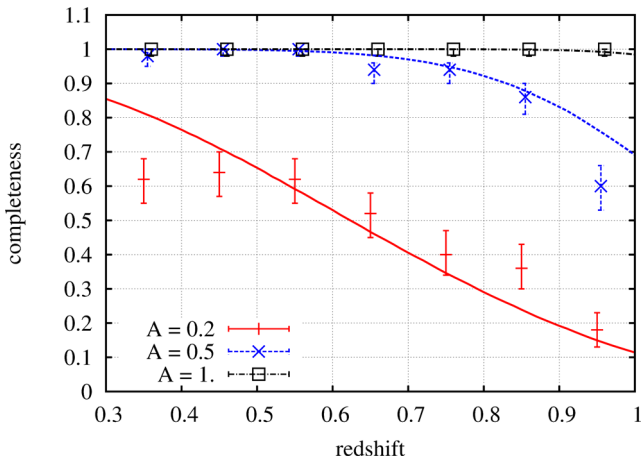

Figure 5. Points with error bars show the fraction of detected objects for different input amplitudes as a function of redshift. The lines indicate the theoretical detection probability. For clarity, points that refer to different input amplitude values are slightly offset along the *x*-axis.

in Section 2.6. The following quantities are read for each galaxy of the catalogue: angular position in the sky, magnitude in the I band, photometric redshift estimate and photometric redshift uncertainty. The algorithm produces a list of detections as described in Section 3. In order to be as complete as possible, we set an S/N threshold = 2 even if, as we will see, the amount of spurious detections due to noise fluctuations is not negligible for S/N < 3. Once we have a list of detections, we try to match them with the actual structures in the field, allowing for an angular separation equal to the radius R_{200} of each real structure and for a redshift separation equal to the 1σ uncertainty $0.05 \times (1 + z)$. When not explicitly mentioned, this *matching procedure* will be repeated for all the experiments we describe.

The resulting completeness for the three lower input values of amplitude (0.2, 0.5, 1.0) is shown in Fig. 5. For comparison, we show also the theoretical detection probability that we obtain from Fig. 4, considering a Gaussian scatter for the observed amplitude with mean equal to $A_{\rm true}$ and uncertainty σ_A calculated as in equation (11). The fraction of the resulting distribution of $A_{\rm obs}$ that is greater than $A_{\rm min}$ for the same redshift is the expected probability of detecting a structure with those properties. As we see, there is a good agreement between the theoretical prediction and the actual fraction of detected objects. For simplicity, we do not show the results for the structures

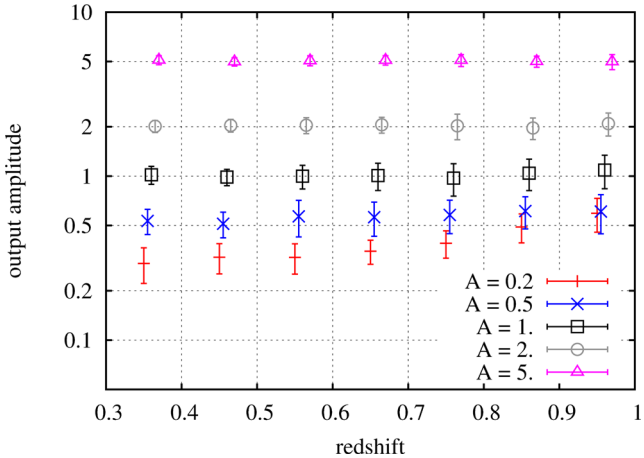

Figure 6. Mean and scatter of the measured amplitude as a function of the input redshift for structures with different input amplitudes. For clarity, points that refer to different input amplitude values are slightly offset along the *x*-axis.

with input amplitude equal to 2 and 5, as the completeness is equal to 1 both in theory and in practice all over the redshift range for these objects.

4.3.2 *Properties of detected structures*

For the detected objects, we can compare the measured amplitude $A_{\rm obs}$ with the input amplitude $A_{\rm true}$. The results are shown in Fig. 6. The estimated amplitude is unbiased with respect to the input value as long as the sample is complete; that is, we are sampling the whole distribution of $A_{\rm obs}$ for a given input $A_{\rm true}$. This is true at all redshifts for the three most massive bins and at $z < 0.6$ for clusters with $A_{\rm true} = 0.5$ (see Fig. 5). Instead, the sample of clusters with $A_{\rm true} = 0.2$ is never complete, because the theoretical amplitude of all objects lay below the detection line given by $A_{\rm true}$; therefore, only those objects whose $A_{\rm obs}$ is increased by the random noise enter the detection catalogue, and this translates into a positive bias in the mean $A_{\rm obs}$ for detected objects. The scatter in the measurement of A is increasing for higher redshift and lower $A_{\rm true}$, as expected given the S/N estimates.

The redshift of the detection can also be compared with the input one. As we see in Fig. 7, the redshift estimate is always consistent with the input value, considering the scatter. The error bars are shrinking for bigger structures, as the number of member galaxies increases and reduces the statistical uncertainty. We note that there is a slight tendency towards an underestimation of z for low redshifts and an overestimation for high redshift, which is mildly significant for the biggest structures. This is due to the redshift evolution of the cluster model $M_c(r, m)$, which may cause the likelihood for a structure at redshift z_c to peak at a slightly different redshift. As we will see in Section 5, this source of bias is negligible in a realistic environment, with correlated noise and increasing confusion due to structures that may be close to each other and aligned along the line of sight.

4.3.3 *Membership*

As described in Section 3.1, after a detection is made, AMICO runs through the galaxy catalogue and assigns to each galaxy a probability of being a member of the identified object. We can verify how accurate is this probabilistic association by comparing

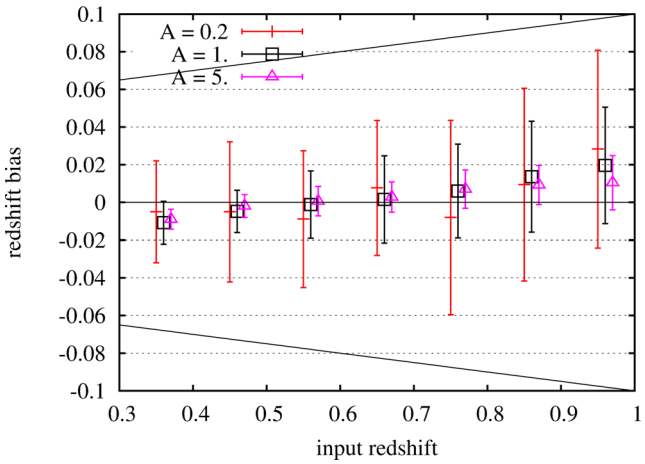

Figure 7. Mean and scatter of the redshift bias (output–input) as a function of the input redshift for structures with different input amplitudes. For clarity, the results for just three input amplitude values are shown and points that refer to different input amplitude values are slightly offset along the x-axis. The solid lines limit the region of 1σ uncertainty of the photometric redshift of individual galaxies.

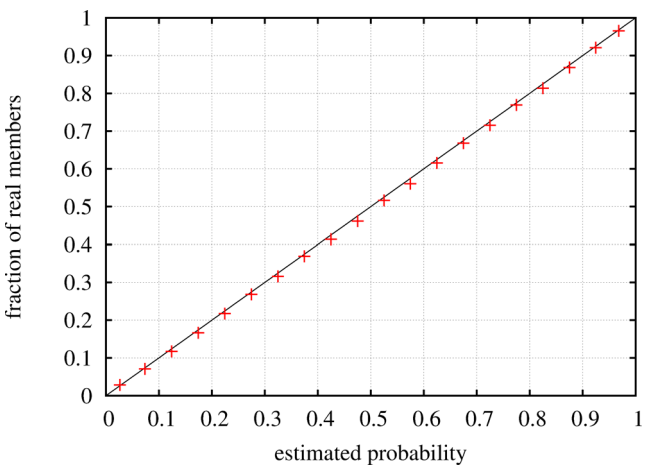

Figure 8. The mean membership probability as estimated by the algorithm versus the fraction of actual members in bins of estimated probability is shown by red pluses. The error bars are negligible because they are smaller than the point size. The 1-to-1 relation is shown for reference as black solid line.

it with the fraction of members that actually belong to the identified structure. The results are shown in Fig. 8. As we see there is an almost perfect agreement between the probability estimated by the code and the fraction of actual members. This is a very important result that verifies the effectiveness of the algorithm when all its assumptions are met. We will verify this result in a more realistic environment in Section 5.5.

4.3.4 Spurious detections

Up to now we focused only on detections that were matched to the objects we inserted in the catalogue. We now consider the fraction of detections that does not correspond to any structure, but instead appears just due to random fluctuations of the background distribution. Even if the Optimal Filtering formalism minimizes their presence, their appearance is expected because of the probabilistic Gaussian nature of the measured amplitude that has the well-defined

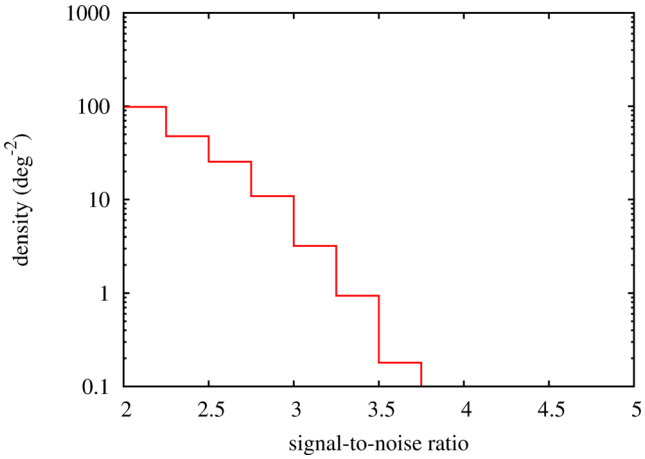

Figure 9. Number of spurious detections per square degree as a function of S/N.

variance given by equation (11). Their distribution as a function of the S/N is shown in Fig. 9. As we see, the density of spurious detections decreases steeply with the S/N. We can estimate a density $\sim$4.5 deg^{-2} at S/N $>$ 3 and $\sim$0.2 deg^{-2} at S/N $>$ 3.5, which is a strong threshold against detections created by purely random background distribution.

4.4 Test on angular deblending

In order to quantify the capability of AMICO to clean the galaxy catalogue from the detected structures (see Section 3.2) and allow for the identification of other neighbouring structures, we create a specific simulation with couples of clusters that are close to each other in angular distance (the same problem along the z-direction will be analysed in Section 4.5). The employed background distribution and cluster galaxy population are the same described in Section 4.3. For this test, all the structures are at redshift 0.5 and have amplitude equal to 1. We created couples of clusters with a relative distance that goes from 0.01 to 0.1 deg in steps of 0.01 deg. For each distance bin 25 couples were used.

As it can be seen from Fig. 4, the expected S/N for a single structure with the parameters we chose is larger than 5. For this reason, we expect that for every couple at least one of the two structures should be correctly identified, and this is verified. To verify how efficient is AMICO to disentangle the two components of the couples, we evaluate the completeness that here is defined as the fraction of couples that are correctly identified as two separate objects over the total number of pairs.

The results are displayed in Fig. 10 for different thresholds in S/N, along with the projected profile of the cluster model. For the case under investigation, the deblending is virtually impossible for clusters with a relative distance = 0.01 deg, while it becomes more and more likely for larger distances, and almost certain once the two structures are at a distance larger than 0.04 deg, which is close to their R_{200}. The completeness for different S/N thresholds highlights the fact that for close-by companions, the significance of the second detection is lowered with respect to the first one. This happens because, during the iterative detection procedure, the membership probability $P(i \in j)$ is assigned considering one detection at a time. Galaxies belonging to the two objects concur to increase the amplitude estimation of the first detection, and the membership probability is then assigned following equation (21) to the galaxies surrounding it, including the ones that belong to the other object.

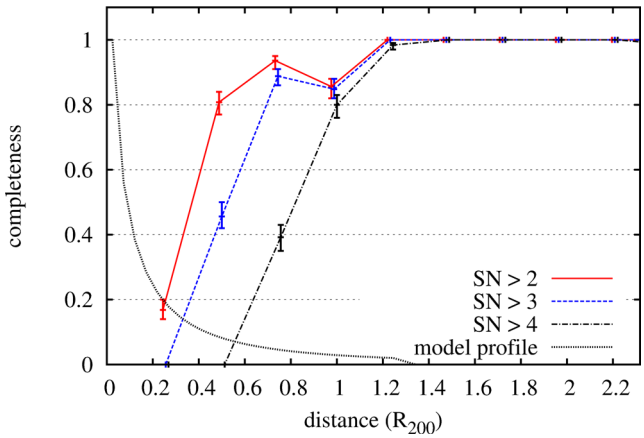

Figure 10. Points with error bars show the fraction of couples correctly identified as two separate objects as a function of the relative distance in degrees for different S/N thresholds (red: S/N ≥ 2, blue: S/N ≥ 3, black: S/N ≥ 4). For reference, the radial profile of the cluster is shown as a dashed line.

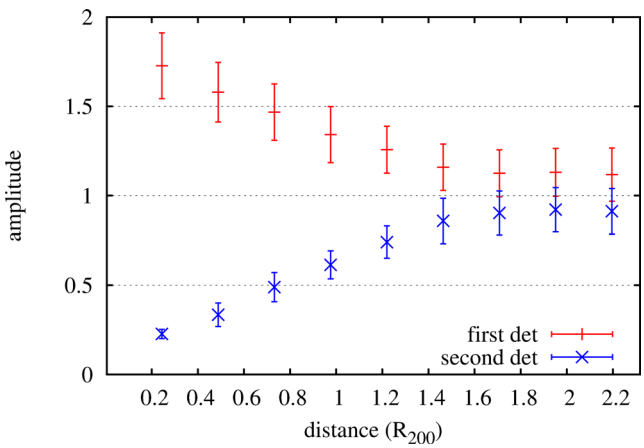

Figure 11. Mean and scatter of the estimated amplitude for the first (red) and the second (blue) detection of each couple as a function of the relative distance in degrees. Only couples for which both objects were detected are considered in this plot.

For this reason, at the stage when AMICO finds the second detection the value of $P_{f,i}$ of its members is already <1, causing an underestimate in the resulting $A_{\rm obs}$ (see equation 25). This can be seen also in Fig. 11, where the mean estimated amplitudes for the first and the second detections of the couple are shown. In the filtering procedure, the two structures interact with each other and, the closer they are, the more the first one 'steals' amplitude from the second one.

That said, it is notable that AMICO with its cleaning approach is able to identify objects whose galaxy distributions are significantly overlapping on the sky and whose relative distance in some cases is smaller than R_{200}. In our simulations, the two objects are exactly at the same redshift, but the same would apply to structures with relative line-of-sight distance lower than the typical photometric redshift uncertainty. We underline that the case we analysed is the most pessimistic from two points of view: the two structures have exactly the same redshift and the same input amplitude. Clearly, deblending of structures at different z would be helped by their different photometric redshift distributions. Moreover, if one of the two structures is smaller than the other, the bias induced on $A_{\rm obs}$

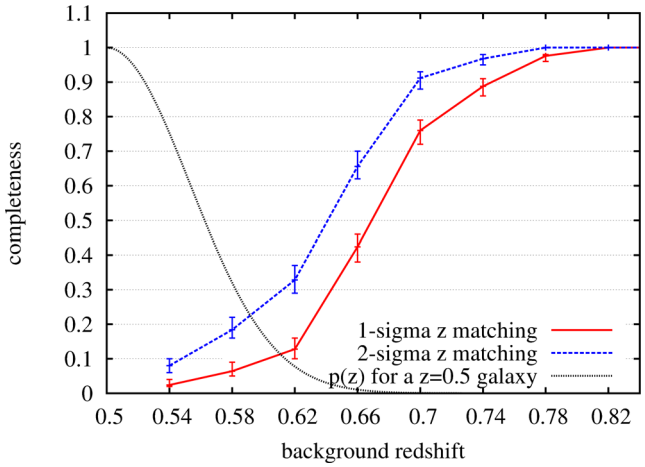

Figure 12. Points with error bars show the fraction of couples correctly identified as two separate objects as a function of the redshift of the most distant structure for two different matching criteria in redshift: 1σ in red and 2σ in blue. The black dashed line is the shape of the photometric redshift distribution for the object at $z = 0.5$.

of the first detected structure would be less significant than the one shown in Fig. 11, making more accurate the probability association of galaxies that is needed in the following step of the iteration.

4.5 Test on line-of-sight projections

Complementary to Section 4.4, here we aim to study the possibility of AMICO to disentangle different structures when they are aligned along the line of sight but at different redshifts. To this end, we create a different simulation where clusters are made in couples. For each pair, the closest (foreground) object lies at $z = 0.5$, while the second (background) one is at a redshift ranging from 0.54 to 0.9, in steps of 0.04. All the structures have an input amplitude equal to 1, but this corresponds to an expected S/N that is decreasing in redshift, as we can see from Fig. 4. As in Section 4.4, we are interested in the capability of the algorithm to detect both objects, as one of the two (in this case, the foreground one) is obviously identified.

The results as a function of the redshift of the background structure are shown as the red line in Fig. 12. We see that the fraction of detected structures goes above 50 per cent when the relative separation in redshift $\Delta z_{\rm pair}$ is larger than $0.16 \sim 2\sigma_z$, while for smaller distances the confusion induced by the photometric redshifts makes it impossible to distinguish the two objects. We can better understand this behaviour by looking at the top panel in Fig. 13, where the bias in the redshift determination for the foreground and the background structures is shown. Galaxies of the background structure tend to increase the signal, and thus the likelihood, of a detection at redshifts slightly higher that the true one of the foreground object, thus biasing high the measured redshift. In turn, the cleaning procedure for the foreground detection tends to remove galaxies of the other object which were scattered low in z. As a result, the redshift of the second detection is biased up as well. This phenomenon is especially relevant for $\sigma_z \leq \Delta z_{\rm pair} \leq 3\sigma_z$, when the galaxies of the structures interact in the amplitude measurement. In particular, it seems likely that for $0.58 \leq z_{\rm back} \leq 0.7$ some of the structures in the background have been detected but were not associated with a halo in the simulation because of the excessive redshift bias. In fact, as it is shown in Fig. 13, in this redshift range the upper limit of the scatter of the redshift bias is close to or larger than the limit in z

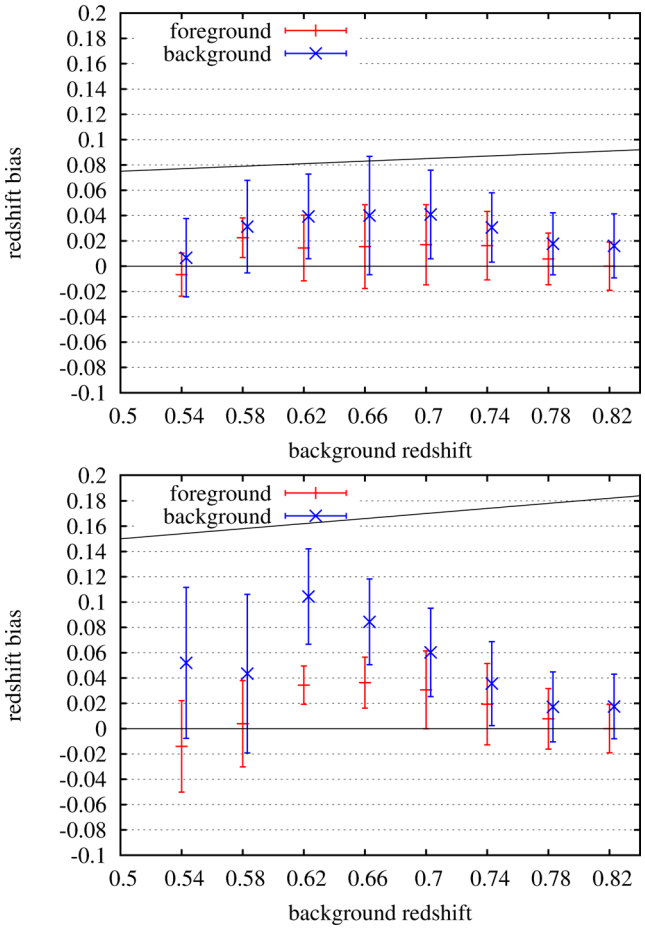

Figure 13. Mean and scatter of the bias in the estimated redshift for the first (red) and the second (blue) detections of each couple as a function of the redshift of the background structure. Top panel: 1σ z matching. Bottom panel: 2σ z matching. Only couples for which both objects were detected are considered in this plot. The upper solid line represents the limiting z bias for which the object is considered matched in the two cases.

for which the background object is considered matched, which is $0.05 \times (1 + z)$.

To test this hypothesis, we relaxed the redshift criterion for matching to 2σ, equivalent to $0.1 \times (1 + z)$, and the result for the completeness is the blue line in Fig. 12. We see there is an increase in the fraction of matched couples, and this is especially relevant for $z_{\rm back} = 0.62$ and $z_{\rm back} = 0.66$ whose completeness increases by ~20 per cent. The corresponding plot about redshift measurement is in the bottom panel of Fig. 13. We see the increase in the redshift bias especially for $z_{\rm back} = 0.62$ and $z_{\rm back} = 0.66$, due to the insertion in the catalogue of couples whose redshift is significantly biased high, confirming our hypothesis. Now, the distribution of measured redshifts is well distant from the matching limit, indicating that its exact location is not affecting anymore the matching results.

Again, we note that these results are relative to a very pessimistic case (two structures at equal amplitude perfectly aligned on the sky) and thus must be considered as a worst-case scenario when facing this problem. We note anyway that AMICO is able in a relevant fraction of cases to detect a structure that lies just behind another one, completely overlapping in the sky and with the two galaxy populations that are overlapping in redshift, with a $\Delta z \sim 2 \times \sigma_z$, where σ_z is the typical photometric redshift uncertainty of the galaxies.

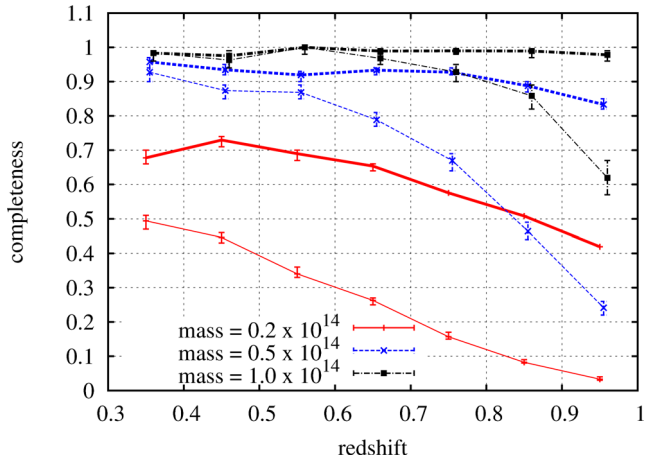

Figure 14. Fraction of detected structures (completeness) as a function of redshift for three different bins in logarithmic mass ($\Delta \log M = 0.1$) and two S/N thresholds: S/N > 2 in solid, S/N > 3.5 in dashed.

As a general remark, the cases presented here and in Section 4.4 quantify the possible biases in the determination of clusters position and amplitude when two or more structures are nearby, and the description of the data as the sum of a field and a cluster component fails. A simple check on the spatial proximity of the output detections of AMICO provides a list of possible cases where these small systematics may occur.

5 TESTS ON COSMOLOGICAL MOCKS

Now we turn to the analysis of the mock catalogues presented in Section 4.1, as they are built from the *N*-body cosmological simulation by applying a semi-analytical modelling of the galaxy evolution. To make the original catalogues more realistic, we introduced a scatter in the redshift of the galaxies to simulate the photometric redshift uncertainties expected from ground-based observations. As in the previous section, galaxies were scattered following a Gaussian distribution with rms $\sigma_z = 0.05 \times (1 + z)$. The 50 mock catalogues with a size equal to $2 \times 2\,{\rm deg}^2$ were analysed separately and then the matching with real haloes was performed following the same procedure described in Section 4.3.1. We initially considered in the matching all the haloes with mass M_{200} larger than $10^{12}\,{\rm M}_\odot\,h^{-1}$ as identified in the catalogue. In Section 5.3, we will discuss and modify this mass threshold for the halo catalogue. Position and redshift of the haloes are set equal to the ones of their central galaxy, as indicated in the catalogue. As in the previous section, the cluster search was performed in the redshift range of $0.3 \leq z \leq 1$. We used the same model $M_c(r, m)$ and background distribution $N(m, z)$ we described in Section 4.2. Differently from the analysis of the ideal mocks, in this case we ran AMICO applying the local background estimation (see Section 2.6) that helps taking into account the large-scale fluctuations of the field. We discuss its impact on the quality of the detections in Section 5.3.

5.1 Completeness

Once the matching is performed, we can quantify the fraction of real objects detected by AMICO per bin of redshift and mass. We plot the results in Fig. 14, where for simplicity we show only three mass bins, corresponding to masses, respectively, ~0.2, ~0.5, ~$1.0 \times 10^{14}\,{\rm M}_\odot\,h^{-1}$, each with a size $\Delta \log M = 0.1$. The completeness is shown for two different S/N cuts: S/N > 2 as we used

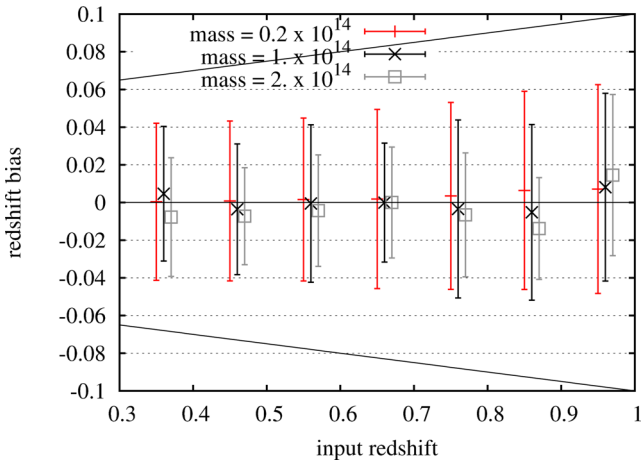

Figure 15. Mean and scatter of the redshift bias (output–input) as a function of the input redshift for structures with different masses. For clarity, just three mass bins are shown, each with $\Delta \log M = 0.1$, and points that refer to different mass bins are slightly offset along the x-axis. The solid lines limit the region of 1σ uncertainty of the photometric redshift.

in Section 4, and S/N > 3.5, which protects against pure random fluctuations of the background, as shown in Section 4.3.4. In general, as one would expect, the completeness is decreasing both with redshift and with mass.

Relating the measured completeness as a function of mass with the analytical estimates as a function of amplitude shown in Fig. 4 is not trivial, because of the mass–amplitude relation, which we will analyse in Section 5.3. The only sample for which we can directly compare results with the analytical expectations is the one with mass $\sim 1.0 \times 10^{14}\,\mathrm{M}_{\odot}\,h^{-1}$. Objects in this mass bin are detected with completeness close to 1 at S/N > 2 for the entire redshift range. The completeness at S/N > 3.5 of this sample is instead decreasing at high redshift, coherently with the theoretical S/N that is close to 3.5 at $z \geq 0.9$, as shown in Fig. 4.

5.2 Redshift estimation

We can perform in this more realistic environment the same analysis we performed in Section 4.3.2 about the redshift estimation. The results are shown in Fig. 15 for three different mass bins. Comparing the results with Fig. 7, we see that the scatter is not as much dependent on mass as it was on amplitude on the ideal mock. The small bias that was visible in the idealized situation is completely negligible in the realistic case. Both these differences depend on the presence of other sources of noise in these mock catalogues, such as the structures that are located along the line of sight to the detected objects, making the background much more complicated than our assumption of a random component with uniform mean. We see that the scatter is also mildly dependent on redshift, as expected since the scatter in the photometric redshifts is linearly dependent on $(1+z)$, and is of the order of $0.5\sigma_z$. This value may look large considering the number of cluster members that contribute to the redshift determination, but the precision is limited by the intervening galaxies belonging to the large-scale structure. The absence of redshift bias in this test depends also on the choice of using an unbiased Gaussian distribution for the redshifts of the galaxies in the mocks. In applications to real data, some bias in the cluster redshift might arise if photometric redshifts in the catalogue are biased.

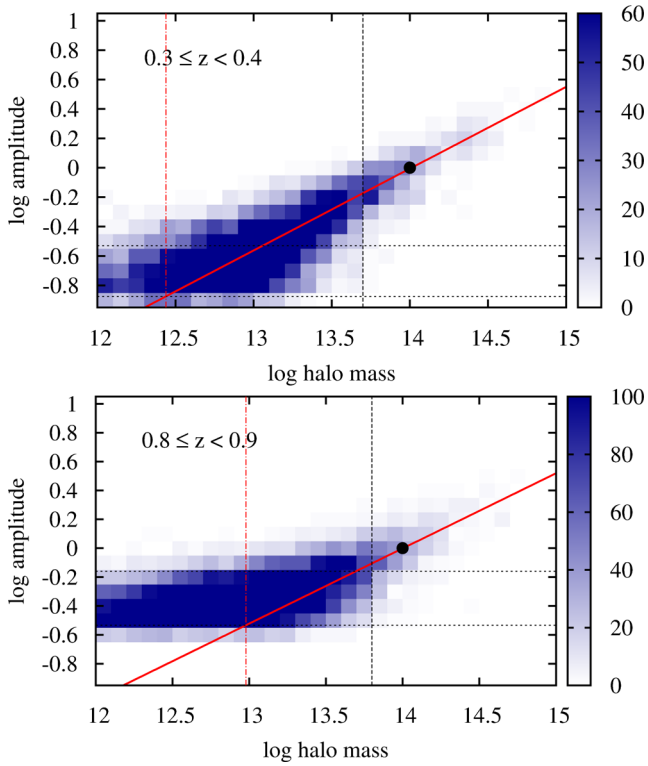

Figure 16. Density plot for the relation between halo mass and estimated amplitude for two redshift bins: $0.3 \leq z < 0.4$ in the top panel, $0.8 \leq z < 0.9$ in the bottom panel. The red line marks the best-fitting mass–amplitude relation. The dashed black vertical line indicates the 90 per cent completeness limit for that redshift bin. Only objects with mass larger than that threshold were considered in the fit. The two horizontal lines indicate the amplitude corresponding to an S/N equal to 2 (threshold for detection) and 3.5. The dashed red vertical line indicates the minimum mass for which the expected amplitude corresponds to an S/N above the detection threshold, according to the best-fitting mass–amplitude relation. The black dot indicates the original calibration of the model, such that amplitude equal to 1 corresponds to a mass of $10^{14}\,\mathrm{M}_{\odot}\,h^{-1}$.

5.3 Mass–amplitude relation

One of the most important properties of a detection method for galaxy clusters, especially for cosmological exploitation, is the relation between the halo mass and the observable. In the case of AMICO, the mass proxy is provided by the amplitude A of the detection (see equation 5). The density plots for the mass–amplitude relation in two different redshift bins are shown in Fig. 16, together with the best-fitting log-linear relation $\log A = a \times \log M + b$. To avoid a more complete treatment that takes into account, the incompleteness of the sample and the intrinsic scatter (see e.g. Andreon & Bergé 2012; Sereno & Ettori 2015), the best-fitting relation is computed considering only the mass range for which the completeness is larger than 90 per cent, to have an almost complete sample of measured amplitudes for a given bin of mass and redshift. The lower limit of this mass range goes from $\log M/(\mathrm{M}_{\odot}\,h^{-1}) = 13.6$ at low z to $\log M/(\mathrm{M}_{\odot}\,h^{-1}) = 13.9$ in the last redshift bin. As it can be seen in Table 1, the slope of the relation is between 0.5 and 0.6 at any redshift, somewhat flatter than what is usually found for other richness estimators in literature (see e.g. Rykoff et al. 2012; van Uitert et al. 2016; Sereno & Ettori 2017; Simet et al. 2017). This may depend on the fact that the amplitude is computed by AMICO with a fixed filter, which does not change according to the

Table 1. Best-fitting parameters for the relation $\log A = a \times \log M + b$ between mass M and amplitude A for different redshift bins. Between parentheses, we show the same parameters of the relation when AMICO is run without local background correction.

z range	Slope a	Intercept b	Scatter of $\log A$	Scatter of $\log M$
[0.3, 0.4[	0.556 ± 0.045 (0.599 ± 0.036)	-0.006 ± 0.011 (0.028 ± 0.010)	0.125 (0.156)	0.290 (0.310)
[0.4, 0.5[	0.580 ± 0.032 (0.592 ± 0.031)	-0.001 ± 0.009 (0.030 ± 0.009)	0.126 (0.150)	0.300 (0.326)
[0.5, 0.6[	0.563 ± 0.036 (0.568 ± 0.035)	0.000 ± 0.009 (0.030 ± 0.009)	0.140 (0.146)	0.313 (0.344)
[0.6, 0.7[	0.554 ± 0.039 (0.563 ± 0.031)	-0.009 ± 0.009 (0.019 ± 0.009)	0.135 (0.156)	0.298 (0.326)
[0.7, 0.8[	0.529 ± 0.046 (0.540 ± 0.045)	-0.011 ± 0.010 (0.022 ± 0.010)	0.144 (0.164)	0.332 (0.343)
[0.8, 0.9[	0.521 ± 0.065 (0.531 ± 0.049)	-0.001 ± 0.013 (0.033 ± 0.011)	0.151 (0.158)	0.366 (0.395)
[0.9, 1.0[	0.510 ± 0.092 (0.492 ± 0.068)	-0.008 ± 0.020 (0.033 ± 0.013)	0.137 (0.151)	0.378 (0.392)

physical size of the object. We expect that for objects less massive than $10^{14}\,\mathrm{M}_\odot\,h^{-1}$ (and thus typically smaller than the model) the correlated noise due to large-scale structure increases the measured amplitude. On the other hand, for more massive and larger objects we miss part of the galaxy population that exceeds the model radius. It is in principle possible to extend the filter to the observed size of the objects, employing the scale-adaptive version of it (Viola, Maturi & Bartelmann 2010).

We point out the very precise estimate of the amplitude for objects with mass $\sim 10^{14}\,\mathrm{M}_\odot\,h^{-1}$. In fact, as the model has been derived from the galaxy population of this sample of clusters, their expected amplitude is equal to one by construction. AMICO recovers perfectly this value, with bias b consistent with zero over the whole redshift range, with no significant difference with respect to the *ideal* case presented in Section 4.3.2. This proves that even when dealing with objects with different shapes, luminosity functions and embedded in the large-scale structure, AMICO provides an unbiased mass proxy, provided that the average cluster properties are known.

In Table 1, we also show the scatter of the amplitude at given mass $\sigma_{A|M}$, and the scatter of mass at given amplitude $\sigma_{M|A}$. The latter was estimated from structures detected at S/N $\geq$ 3.5. The shallow A–M relation makes $\sigma_{M|A}$ much larger than $\sigma_{A|M}$. However, both these values are somewhat dependent on the sample definition, because they decrease for larger amplitude and mass.

Finally, we report in Table 1 the parameters of the mass–amplitude relation that are obtained assuming a spatially uniform background (i.e. neglecting the local background estimate described in Section 2.6). By comparing parameters with and without local background correction, we see that the advantage of applying the $f(\theta, z_c)$ local correction is twofold: the intercept b of the relation is closer to zero and the scatter is reduced. This is consistent with the expectation that massive haloes are embedded in a higher background that, when neglected, biases high the amplitude estimate. It also proves the ability of our method to correctly capture these large-scale features, reducing their contribution to the A–M scatter.

5.4 Miscentering

For several applications (e.g. weak lensing mass measurements, galaxy density profiles), it is important to estimate accurately the centre of the detected structure. AMICO, differently from other methods, does not search for a central galaxy, but instead selects the centre as the position of highest likelihood, that at a given redshift corresponds to a maximum in amplitude as well (see equation 14). The identification of the central galaxy can then be done a posteriori according to the specific goal. We can measure the mean offset between the centre as defined by AMICO and the central galaxy as indicated in the catalogue. The results are shown in Fig. 17. The

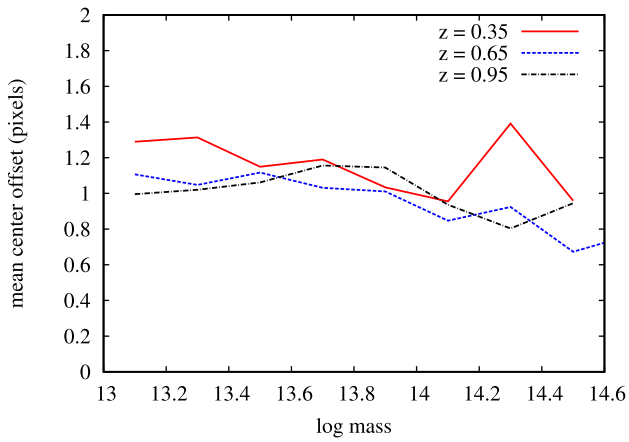

Figure 17. Mean offset between the centre of the halo and the centre of the detection, in pixels, as a function of halo mass. The pixel size is set to 0.005 deg that correspond to 63.7, 90.2 and 101.3 kpc h^{-1} at redshift 0.35, 0.65 and 0.95, respectively.

mean offset is typically of the order of the size of the amplitude map pixel (0.005 deg). In physical units, this translates in a value between 60 and 120 kpc h^{-1} that tends to increase with redshift. In principle, the resolution of the amplitude map can be increased at will, but there is a trade-off between precision and computing speed. Moreover, given that our centre depends on a filtering procedure, it is meaningless to give it with a precision smaller than the typical size over which the filter changes significantly.

We underline that in Fig. 17, we are measuring the distance between two quantities with a different definition: in the mock catalogues, the centre is identified with the BCG, while our estimator depends on the full galaxy distribution on the scale of the cluster. Even if one neglects resolution issues, it is not expected that the two correspond precisely for many of the objects. We note here that it is possible to select BCGs a posteriori from AMICO detections (see Radovich et al. 2017).

We also checked for systematics in the amplitude estimate with respect to the centring offset. There is no significant trend as a function of the offset in the mass proxy derived by AMICO. This is expected as, by construction, the algorithm selects the position of the maximum of A as the detection centre.

5.5 Membership

Having established the performance of the detection procedure, we now focus on the assignment of the membership probability of the galaxies, as we have done in Section 4.3.3 for the ideal case. Only members of detections matched to real objects were considered in the following analysis. In Fig. 18, we show the relation between

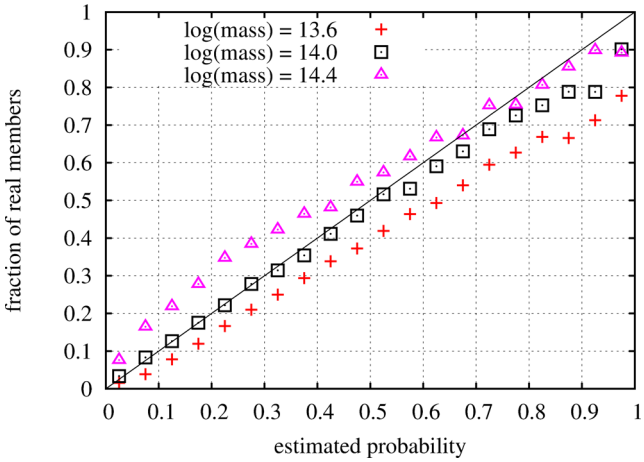

Figure 18. Mean membership probability as estimated by the algorithm versus the fraction of actual members. Red, black and green points refer to galaxy members of haloes with $\log M/(\mathrm{M}_\odot\, h^{-1})$ around 13.6, 14.0 and 14.4, respectively. The error bars are negligible because they are smaller than the point size. The 1-to-1 relation is shown for reference as a black solid line.

the estimated probability and the fraction of real members for three different bins of mass. As explained in Section 4.2, galaxies in the mock catalogues are defined as members when they are at a distance $\leq 1.35 R_{200}$ from the central galaxy. The plot in Fig. 18 shows that, even in a more realistic scenario, AMICO is able to provide a reliable membership assignment through about an order of magnitude in mass. For haloes with masses $\sim 10^{14}\,\mathrm{M}_\odot\, h^{-1}$, the relation is well calibrated with differences lower than 0.1.

The slight over(under)-estimation of the probability for smaller (larger) haloes depends on the mismatch between the model and the actual size of the objects. The flattening of the relation at high probabilities is mostly due to small miscenterings of the true halo position. To further improve the identification of the members of a cluster, one can a posteriori correct the membership probability, provided a reasonable estimate of the mass–amplitude relation is obtained, but this goes beyond the scope of this paper.

5.6 Spurious detections

In general, the definition of spurious detection has some ambiguity in this case, differently from the simple case analysed in Section 4.3.4. We showed in Section 5.3, that, as for any mass proxy based on photometric data, there is a significant scatter between mass and amplitude, and thus also objects of small mass are expected to enter the sample if we do not set a very high threshold in S/N or A_{obs}. On the other hand, due to the photometric redshift uncertainty, it is necessary to search for detection counterparts in a quite large window in z, inside which one will usually find some corresponding halo, if no significant mass threshold is set. This means that in this case there is not an obvious way to discriminate true and false detections, and that most of the information on the quality of the detection method is embedded in the mass-observable plots shown in Fig. 16.

For this reason, we decided to define the mass threshold for our sample considering the properties the mass–amplitude relation. Given the analytic S/N estimates as a function of amplitude and redshift shown in Fig. 4, for each redshift bin we derive the amplitude that corresponds to S/N = 2 (see the relation in Fig. 4) and define the value of $M_{\mathrm{thr}}(z)$ as the one derived from the A–M relation, without

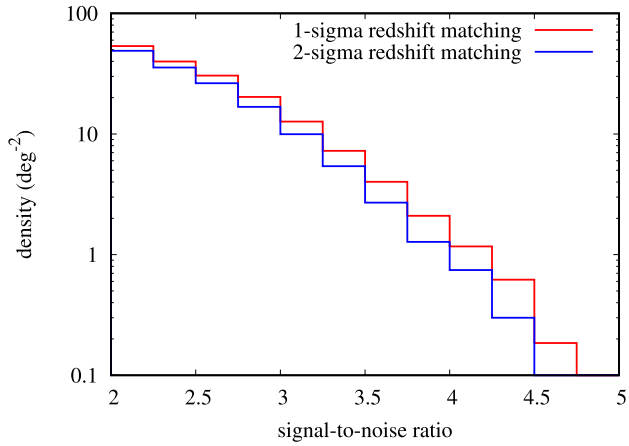

Figure 19. Number of spurious detections per square degree as a function of S/N. Red line: 1σ redshift matching. Blue line: 2σ redshift matching.

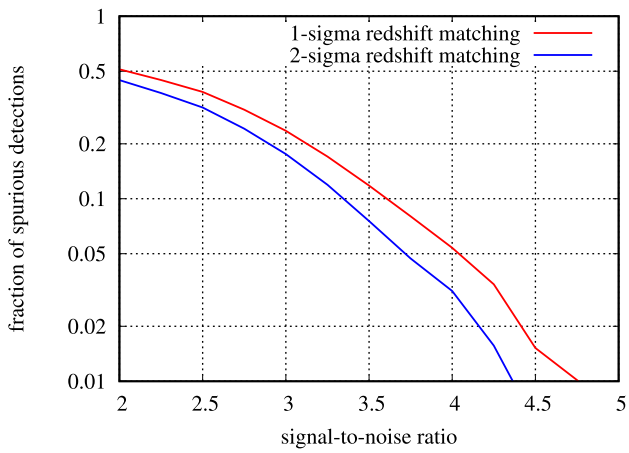

Figure 20. Fraction of spurious detections with a minimum S/N. Red line: 1σ redshift matching. Blue line: 2σ redshift matching.

considering the scatter. We then consider as valid counterparts only haloes that have mass bigger than $M_{\mathrm{thr}}(z)$. We note that this mass is actually quite low compared to what is usually defined a cluster and increases as a function of redshift, with $\log M_{\mathrm{thr}}/(\mathrm{M}_\odot\, h^{-1})$ that goes from 12.44 at $z = 0.35$ to 13.14 at $z = 0.95$, following the decrease in S/N for a fixed amplitude. All the detections without a counterpart with $\log M > M_{\mathrm{thr}}(z)$ are then considered as spurious.

We show in Fig. 19 the density per square degree of spurious detections as a function of their S/N, and in Fig. 20, the fraction of unmatched detections with an S/N larger than a given threshold. Both results are shown for the normal matching procedure (red line) and for the more relaxed criterion of 2σ in z. The S/N is confirmed to be a good indicator of the confidence in a detection, as the fraction of spurious detections is monotonically and steeply decreasing. Comparing Fig. 19 with Fig. 9, we see that the number of spurious detections at $2 < \mathrm{S/N} < 2.5$ is actually lower in this case than in the ideal case of pure random uncorrelated background. This happens because the cleaning procedure applied on high S/N structures removes most of the very low-significance detections. On the other hand, there is a small but not irrelevant number of detections considered as spurious at $3 < \mathrm{S/N} < 5$. The fraction of unmatched detections goes below 12 per cent (8 per cent) at S/N > 3.5 for the 1σ (2σ) redshift matching criterion and becomes negligible at S/N > 5. This value is consistent with what was obtained on COSMOS data

in Bellagamba et al. (2011) with a previous version of the algorithm that did not implement the deblending procedure and the local correction for local background variations.

We underline again that this definition of spurious detection is somewhat arbitrary. There is no reason why a halo of mass just above $M_{\rm thr}$ should be considered as a good counterpart while one just below this threshold should not, given the scatter in the A–M relation. Actually, following the relation between mass and amplitude derived in Section 5.3 and its scatter, one can estimate which fraction of objects of a certain mass and redshift is expected to enter the detection catalogue for a given S/N threshold. In our case, we found that this would lead to a large overestimation of the expected detections for objects of mass $M \lesssim M_{\rm thr}$. In practice, most of the very small objects are eliminated when bigger structures are found and cleaned, because they are physically connected or randomly aligned with them. As a consequence of this, we are actually detecting fewer objects of mass similar or lower than $M_{\rm thr}$ with respect to what one would expect combining the mass function and the scatter in the mass–amplitude relation.

6 SUMMARY AND CONCLUSIONS

In this work, we presented AMICO, a new implementation of the Optimal Filtering technique for cluster detection presented in Maturi et al. (2005) and applied to photometric data in Bellagamba et al. (2011). Differently from other approaches, Optimal Filtering is not based on the search of any specific feature of galaxy clusters, but instead aims at maximizing the S/N for the detection of objects that follow a given model in data affected by a given amount of noise. The model for clusters at redshift z can be considered an input parameter in this approach and can in principle be derived from previous observations or iteratively from the data. In this paper, we did not focus on the extraction of the cluster model, and we instead considered the workflow of AMICO once the model is known.

The first step in the detection procedure is the creation of a 3D map constructed through the application of the Optimal Filter to the data. This map contains in each pixel an estimate of the amplitude of a tentative cluster centred in that position in the sky and redshift, together with its S/N and likelihood. The main novelty we introduce with AMICO with respect to Bellagamba et al. (2011) is the method to derive the detection catalogue from the filtered map. Instead of just selecting peaks above a certain S/N in the map, we apply a procedure that aims at erasing iteratively the imprint of the identified structures. This helps in the detection of smaller structures partially overlapping with them in the sky. As a by-product, we also obtain for each galaxy a reliable estimate of its probability to be a member of the candidate clusters.

We tested the performance of the algorithm running it blindly on mock catalogues. We first used catalogues with clusters that follow our model and a constant random background to test the code in a controlled environment. In this way, we verified that AMICO is able to reach a completeness in agreement with theoretical expectations, and unbiased estimates of amplitude and redshift for the detected objects. All expectations were met. We then turned to specific tests on the capability of AMICO to disentangle structures that are close-by in the sky or aligned along the line of sight. In 50 per cent of the cases, AMICO proved to be able to distinguish two structures whose relative distance is $\sim 0.5 \times R_{200}$ in the sky or $\sim 2 \times \sigma_z$ in redshift, where σ_z is the typical photometric redshift uncertainty of the galaxies.

We finally turned our attention to the analysis of cosmological mocks, produced by semi-analytic modelling of galaxy evolution on a cosmological N-body simulation. In this case, the galaxy distribution is more similar to what observed in the Universe, with haloes of different masses embedded in the large-scale structure. This makes the detection procedure more difficult, because the background is not homogeneous and the galaxy distribution of each cluster may have peculiar properties. Despite the complexity of the data, AMICO achieves a completeness for haloes with mass $\sim 10^{14}\,{\rm M}_\odot\,h^{-1}$ close to 1 all over the considered redshift range ($0.3 < z < 1.0$), in line with the theoretical predictions and the results on the ideal mocks. The redshift of the detected haloes is recovered with no significant bias, and with a scatter which is of $\sim 0.5\sigma_z$, higher than in the ideal case.

To analyse in more depth the results of the detection procedure, we performed a fit of the mass–amplitude relation in different redshift bins, considering the mass range for which the sample is (almost) completely detected. The mass–amplitude relation has a (logarithmic) slope between 0.50 and 0.60. Most importantly, the bias for haloes with masses close to the calibration mass of $10^{14}\,{\rm M}_\odot\,h^{-1}$ (i.e. the intercept of the $\log A$–$\log M$ relation) is consistent with zero at all z. This shows that the measured amplitude respects the input calibration of the model. Also, the logarithmic scatter of the amplitude does not show any evolution with z and is always between 0.13 and 0.15. This indicates that the amplitude A is an unbiased redshift-independent mass proxy, provided that the model is calibrated with haloes of similar mass at each redshift. A key step in this result is the local estimate of the background population, which allows to subtract effectively the contribution of the large-scale structure correlated with the clusters.

Finally, we verified that the S/N computed by AMICO is a reliable indicator of the goodness of a detection: the fraction of false positives is $< 8-12$ per cent at S/N > 3.5 and practically zero at S/N > 5. Note that for this calculation we considered as spurious all the detections that do not have a counterpart in the halo catalogue with mass above $M_{\rm thr}(z)$, which corresponds to a theoretical S/N $= 2$. This sharp truncation in mass is motivated by our approach, but nevertheless it neglects all haloes with that can be, and actually are, detected because of the scatter in the mass–amplitude relation. Therefore, our estimate is conservative.

AMICO is an algorithm that builds on the well-established Optimal Filtering approach for cluster detection, solving in an effective way the problem of extracting detections from the map. The cleaning procedure allows the detection of groups and clusters that are close-by or aligned to larger structures, a particularly challenging task for the peak selection used traditionally in filtering methods (Kepner et al. 1999; Milkeraitis et al. 2010; Bellagamba et al. 2011). In some way, the cleaning procedure resembles the so-called *percolation* in redMaPPer (Rykoff et al. 2014), where a membership probability is attributed to the galaxies in a similar way to our equations (21) and (23). Similarly to redMaPPer, the final centre, redshift and richness of the detections are computed taking into account higher ranked close-by clusters. We point out, however, that the way in which detections are performed is different in the two algorithms: redMaPPer computes and ranks the candidate cluster centres on the original catalogue, while in AMICO each cluster candidate is detected only after those with larger likelihood have been erased. Moreover, redMaPPer removes from the list of possible centres the galaxies that have a probability of belonging to another structure larger than 50 per cent, while AMICO does not set any arbitrary threshold, as the S/N in each point takes naturally into account the available probability of the surrounding galaxies.

By choice, in this work, we ignored the problem of defining a correct cluster model for a given data set, which is for sure an

important part of the challenge of cluster detection, especially when the redshift limits of the surveys are pushed above $z = 1$, when less information about known structures is available. We are working on the capability of AMICO to extract the cluster model from a given data set, and we plan to show this feature in a future work.

Finally, we note that, thanks to its flexible structure, AMICO can, in principle, be applied to other astronomical problems, such as the detection of dwarf galaxies or globular clusters in star catalogues.

ACKNOWLEDGEMENTS

FB, MR and LM thank the support from the grants ASI n.I/023/12/0 'Attività relative alla fase B2/C per la missione Euclid' and PRIN MIUR 2015 'Cosmology and Fundamental Physics: Illuminating the Dark Universe with Euclid'. MM was supported by the SFB-Transregio TR33 'The Dark Universe'. We thank O. Cucciati and S. Bardelli for the useful discussions.

REFERENCES

Abell G. O., 1958, ApJS, 3, 211
Andreon S., Bergé J., 2012, A&A, 547, A117
Ascaso B., Wittman D., Benítez N., 2012, MNRAS, 420, 1167
Bellagamba F., Maturi M., Hamana T., Meneghetti M., Miyazaki S., Moscardini L., 2011, MNRAS, 413, 1145
Botzler C. S., Snigula J., Bender R., Hopp U., 2004, MNRAS, 349, 425
de Jong J. T. A. et al., 2013, The Messenger, 154, 44
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dong F., Pierpaoli E., Gunn J. E., Wechsler R. H., 2008, ApJ, 676, 868
Gladders M. D., Yee H. K. C., 2000, AJ, 120, 2148
Hansen S. M., McKay T. A., Wechsler R. H., Annis J., Sheldon E. S., Kimball A., 2005, ApJ, 633, 122
Kaiser N. et al., 2002, in Tyson J. A., Wolff S., eds, Proc. SPIE Conf. Ser. Vol. 4836, Survey and Other Telescope Technologies and Discoveries. SPIE, Bellingham, p. 154
Kepner J., Fan X., Bahcall N., Gunn J., Lupton R., Xu G., 1999, ApJ, 517, 78
Koester B. P. et al., 2007, ApJ, 660, 221
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Le Fevre O. et al., 2003, in Guhathakurta P., ed., Proc. SPIEConf. Ser. Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II. SPIE, Bellingham, p. 173
Licitra R., Mei S., Raichoor A., Erben T., Hildebrandt H., 2016, MNRAS, 455, 3020
LSST Science Collaboration et al., 2009, peprint (arXiv:0912.0201)
Lumsden S. L., Nichol R. C., Collins C. A., Guzzo L., 1992, MNRAS, 258, 1
Maturi M., Meneghetti M., Bartelmann M., Dolag K., Moscardini L., 2005, A&A, 442, 851
Meneux B. et al., 2008, A&A, 478, 299
Milkeraitis M., van Waerbeke L., Heymans C., Hildebrandt H., Dietrich J. P., Erben T., 2010, MNRAS, 406, 673
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Pace F., Maturi M., Meneghetti M., Bartelmann M., Moscardini L., Dolag K., 2007, A&A, 471, 731
Pace F., Maturi M., Bartelmann M., Cappelluti N., Dolag K., Meneghetti M., Moscardini L., 2008, A&A, 483, 389
Postman M., Lubin L. M., Gunn J. E., Oke J. B., Hoessel J. G., Schneider D. P., Christensen J. A., 1996, AJ, 111, 615
Radovich M. et al., 2017, A&A, 598, A107
Ramella M., Nonino M., Boschin W., Fadda D., 1999, in Giuricin G., Mezzetti M., Salucci P., eds, ASP Conf. Ser. Vol. 176, Observational Cosmology: The Development of Galaxy Systems. Astron. Soc. Pac., San Francisco, p. 108
Rykoff E. S. et al., 2012, ApJ, 746, 178
Rykoff E. S. et al., 2014, ApJ, 785, 104
Rykoff E. S. et al., 2016, ApJS, 224, 1
Schechter P., 1976, ApJ, 203, 297
Sereno M., Ettori S., 2015, MNRAS, 450, 3633
Sereno M., Ettori S., 2017, MNRAS, 468, 3322
Shapley H., 1933, Proc. Natl. Acad. Sci. USA, 19, 591
Shectman S. A., 1985, ApJS, 57, 77
Simet M., McClintock T., Mandelbaum R., Rozo E., Rykoff E., Sheldon E., Wechsler R. H., 2017, MNRAS, 466, 3103
Soares-Santos M. et al., 2011, ApJ, 727, 45
Springel V. et al., 2005, Nature, 435, 629
The Dark Energy Survey Collaboration, 2005, preprint (arXiv:astro-ph/0510346)
Trevese D., Castellano M., Fontana A., Giallongo E., 2007, A&A, 463, 853
van Uitert E., Gilbank D. G., Hoekstra H., Semboloni E., Gladders M. D., Yee H. K. C., 2016, A&A, 586, A43
Viola M., Maturi M., Bartelmann M., 2010, MNRAS, 403, 859
Wen Z. L., Han J. L., Liu F. S., 2012, ApJS, 199, 34
White M., Kochanek C. S., 2002, ApJ, 574, 24

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