Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **420,** 1861–1881 (2012) doi:10.1111/j.1365-2966.2011.19782.x

ORCA: The Overdense Red-sequence Cluster Algorithm

D. N. A. Murphy,[1⋆] J. E. Geach[1,2] and R. G. Bower[1]

[1]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE*
[2]*Department of Physics, McGill University, Ernest Rutherford Building, 3600 Rue University, Montréal, Québec H3A 2T8, Canada*

Accepted 2011 September 7. Received 2011 September 7; in original form 2010 August 8

ABSTRACT

We present a new cluster-detection algorithm designed for the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) survey but with generic application to any multiband data. The method makes no prior assumptions about the properties of clusters other than (i) the similarity in colour of cluster galaxies (the 'red sequence'); and (ii) an enhanced projected surface density. The detector has three main steps: (i) it identifies cluster members by photometrically filtering the input catalogue to isolate galaxies in colour–magnitude space; (ii) a Voronoi diagram identifies regions of high surface density; and (iii) galaxies are grouped into clusters with a Friends-of-Friends technique. Where multiple colours are available, we require systems to exhibit sequences in two colours. In this paper, we present the algorithm and demonstrate it on two data sets. The first is a 7-deg^2 sample of the deep Sloan Digital Sky Survey (SDSS) equatorial stripe (Stripe 82), from which we detect 97 clusters with $z \leq 0.6$. Benefitting from deeper data, we are 100 per cent complete in the `maxBCG` optically selected cluster catalogue (based on shallower single-epoch SDSS data) and find an additional 78 previously unidentified clusters. The second data set is a mock Medium Deep Survey Pan-STARRS catalogue, based on the Λ cold dark matter (ΛCDM) model and a semi-analytic galaxy formation recipe. Knowledge of galaxy–halo memberships in the mock catalogue allows for the quantification of algorithm performance. We detect 305 mock clusters in haloes with mass $>10^{13}\,h^{-1}\,\mathrm{M}_{\odot}$ at $z \lesssim 0.6$ and determine a spurious detection rate of <1 per cent, consistent with tests on the Stripe 82 catalogue. The detector performs well in the recovery of model ΛCDM clusters. At the median redshift of the catalogue, the algorithm achieves >75 per cent completeness down to halo masses of $10^{13.4}\,h^{-1}\,\mathrm{M}_{\odot}$ and recovers >75 per cent of the total stellar mass of clusters in haloes down to $10^{13.8}\,h^{-1}\,\mathrm{M}_{\odot}$. A companion paper presents the complete cluster catalogue over the full 270-deg^2 Stripe 82 catalogue.

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

1 INTRODUCTION

Galaxy clusters are integral tools in our drive to test the Λ cold dark matter (ΛCDM) cosmological model and our understanding of galaxy formation. The evolution of the cluster population with redshift, for example, can impose important constraints on the matter density of the Universe (Carlberg et al. 1996; Evrard 1997; Schuecker et al. 2003) and the growth of primordial density fluctuations (Frenk et al. 1990; White, Efstathiou & Frenk 1993; Fedeli, Moscardini & Matarrese 2009). The deep potential wells of clusters offer a suite of laboratories within which detailed studies of gas–galaxy interactions are possible. There is evidence that clusters have been in place for a significant fraction of the star-forming history of the Universe, meaning they can provide a unique insight into how environmental effects shape the evolutionary path of galaxies.

The cluster mass budget is dominated by the presence of dark matter (∼85 per cent, for a comprehensive review, see Voit 2005), making them ideal sites for identifying strongly lensed background galaxies (Smail et al. 2007) and thus provide glimpses of the early star-forming Universe (Swinbank et al. 2010). Weak-lensing studies can determine the projected mass distribution of clusters (e.g. Sheldon et al. 2004) and in some cases the dark matter itself (Clowe et al. 2006). Hot intracluster gas also leaves an imprint on the cosmic microwave background (CMB) by way of the Sunyaev–Zel'dovich (hereinafter SZ, Sunyaev & Zel'dovich 1980; Carlstrom, Holder & Reese 2002) effect via the inverse Compton scattering of CMB photons. At the megaparsec scale, clusters act as high-mass lamp-posts between the filamentary connected structures tracing out the cosmic

⋆E-mail: david.murphy@durham.ac.uk

web (Pimbblet & Drinkwater 2004; Colberg, Krughoff & Connolly 2005; Murphy, Eke & Frenk 2011).

There is therefore great merit in producing a homogeneous cluster census of the Universe, and much effort has gone into producing comprehensive cluster surveys. Efforts to this end are broadly separated into two wavelength domains: the optical–near-IR (optical–NIR) and X-ray. We note in passing that cluster detection by SZ-decrement in the microwave is an emerging cluster survey technique that holds promise at high redshift (McInnes et al. 2009; Brodwin et al. 2010; Hincks et al. 2010; Vanderlinde et al. 2010).

X-ray detections exploit the hot intracluster gas accounting for the bulk of the cluster baryonic mass component (Cavaliere & Fusco-Femiano 1976; Allen, Schmidt & Fabian 2002). X-ray-selected cluster catalogues tend to be robust to projection effects, probe large volumes and produce a cluster sample with well-characterized masses. Cluster catalogues from large-area X-ray surveys (e.g. Ebeling et al. 1998) identify bright, massive clusters, with their deep potential wells establishing the high electron densities required for strong X-ray emission. Whilst a cursory glance in the X-ray unveils the presence of high-mass systems, to select those with lower masses, unresolved gas components, and distant or gas-poor clusters, one must look to alternative approaches.

There has been a half-century history of cluster identification in the optical regime. Early 'eyeball' surveys of photographic plates produced the earliest cluster catalogues (Abell 1958; Zwicky, Herzog & Wild 1961; Abell, Corwin & Olowin 1989) and allowed the first statistical study of the cluster population. When cluster and group samples were later constructed with the help of digitized photographic plates (such as the APM; Dalton et al. 1992) and galaxy spectra (Eke et al. 2004), the task of identification passed from human to machine. With the advent of wide-field multiband CCD imaging, assembly of vast galaxy samples has become the standard. For example, Sloan Digital Sky Survey (SDSS; York et al. 2000) optical imaging data have vastly increased both the volume and the detail of detected astronomical sources, to date generating five-band *ugriz* photometry for ∼230 million objects (Data Release 7, hereinafter DR7, Abazajian et al. 2009). Although one can estimate galaxy redshifts photometrically based on spectral energy distribution (SED) template fitting (Csabai et al. 2003), neural networks (Collister & Lahav 2004) or a combination of the two (Abazajian et al. 2009, section 4.6), photo-zs are prone to large uncertainties and are generally unsuitable for accurate 3D reconstructions of the galaxy distribution (although for recent approaches using the entire photometric redshift distribution, see Liu et al. 2008).

Armed with only the angular positions of galaxies, automated algorithms have been developed to identify clusters as projected overdensities in the plane of the sky (Lidman & Peterson 1996; Postman et al. 1996). These often come at the expense of model dependency and sensitivity to the boundaries and holes common in galaxy catalogues. More geometric approaches have made use of the Voronoi Tessellation (VT) to map the projected density distribution of galaxies. Using the Voronoi cell area as a proxy for the local galaxy density, VTs were first suggested as a non-parametric means of astrophysical source detection by Ebeling & Wiedenmann (1993), and later cluster detection in Ramella et al. (2001). Voronoi techniques have also been used in void detection (Ryden 1995; El-Ad, Piran & da Costa 1996) and the identification of large-scale structure (Icke & van de Weygaert 1991). However, these approaches tend to suffer from contamination arising from the inclusion of background and foreground field galaxies.

Gladders & Yee (2000) proposed a powerful method that picks out the near-ubiquitous signature of galaxy clusters from photometric surveys. Star formation rates of galaxies bound in the potential wells of clusters are suppressed when the cold gas supply is depleted by environmentally driven stripping or starvation processes (Balogh, Navarro & Morris 2000). The passively evolving stellar populations in these galaxies develop strong metal absorption lines bluewards of 4000 Å giving rise to a break, or step, in their spectra. In broad-band photometric filters, these cluster members appear nearly uniformly red between the bands that straddle the spectral break. Because cluster galaxies occupy a wide range of masses (luminosities), these characteristic colours produce a distinct ridgeline, or 'red sequence' (Bower, Lucey & Ellis 1992), in colour–magnitude space. The dichotomy between this quiescent population of predominantly E/S0 galaxies and the star-forming population of spiral-dominated field galaxies is observed as a bimodality of galaxy colours. With increasing redshift, the 4000-Å break moves redwards; the Gladders & Yee (2000) prescription for cluster detection exploits both the strong colour bimodality in the galaxy distribution and the colour–redshift relation to isolate clusters of galaxies over a range of epochs.

With a growing body of IR data (specifically, the IRACs onboard the *Spitzer Space Telescope*), efforts such as the Spitzer Adaptation of the Red-Sequence Cluster Survey (Wilson et al. 2009) have already turned to pushing red-sequence cluster searches beyond the optical/NIR regime. With evidence of cluster sequences in place up to $z \sim 1.5$ (Papovich et al. 2010; Hayashi et al. 2011) and perhaps as early as $z = 3$ (Kodama et al. 2007; Doherty et al. 2010), tracking the 4000-Å break farther redwards shows great potential in filling the $1.4 < z < 2.2$ cluster desert. These distant systems may potentially hold some crucial clues for our understanding of galaxy formation and evolution.

Future observational campaigns such as the Large Synoptic Survey Telescope (Ivezic et al. 2008) are set to push forward the frontiers of wide-area, deep multiband optical sky surveys. More immediately the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) Telescope 1[1] (PS1; Kaiser et al. 2002), the first of four 1.8-m telescopes, is currently imaging three-fourths of the sky with deep, well-characterized (Stubbs et al. 2010) five-band photometry. Algorithms capable of processing the petabyte-scale sky surveys of these next-generation facilities will be best placed to supply data products fully exploiting their advances. Cluster selection by red sequence is set to remain highly relevant to the construction of cluster catalogues using these forthcoming surveys.

One approach to cluster detection in these deeper data sets is through 'matched-filter' (MF; Postman et al. 1996) algorithms that distill the large body of collected cluster data into a likelihood function, recovering systems by maximizing the likelihood of survey data fitting the model. In particular, these filters may specify the cluster luminosity function, radial density distribution, behaviour of the red-sequence ridgeline and in some cases the presence of a central brightest cluster galaxy (BCG) (`maxBCG`; Koester et al. 2007b). MF algorithms often confer redshift and richness estimates as part of the detection procedure. The MF technique has been successful in extracting cluster signals from a diverse range of galaxy surveys, including the SDSS (Goto et al. 2002; Koester et al. 2007a) and Canada–France–Hawaii Telescope Legacy Survey (Gladders & Yee 2005; Thanjavur, Willis & Crampton 2009). The `maxBCG` SDSS cluster catalogue (Koester et al. 2007a) has facilitated a more detailed study of the cluster red sequence (Hao et al. 2009), which may in turn provide added refinements to future algorithms.

[1] http://pan-starrs.ifa.hawaii.edu

However, the advantage of MF algorithms can also be their drawback: such techniques will preferentially recover the clusters they are designed to match, but those not fitting the model are less likely to be identified. Many matched filter approaches also are based on uniform background galaxy distributions, and experience a degraded performance (Kim et al. 2002) under more realistic backgrounds.

Our cluster detection philosophy is designed to be distinct from, but entirely complementary to, the variety of matched filter algorithms available. This study relaxes theoretically and observationally motivated constraints, permitting a broader exploration of systems with projected overdensities. Specifically, we do not assume cluster red sequences occupy a particular position in colour–magnitude space, nor do we stipulate preferred distributions for the projected position of cluster members on the sky. Through this approach, we hope to provide both an independent catalogue of clusters and a means to refine our understanding of characteristic cluster properties. The lack of selection criteria in our algorithm permits a double-check of the detections, since we can ask if the identified system conforms to our expectations. As we shall later demonstrate (see Section 5 and Fig. A2), the prescription presented here may lead to improved recovery of certain systems and better agreement with X-ray cluster data. Moreover, because our proposed technique makes only two assumptions about cluster properties, it is sensitive to a wide range of clusters, including aspherical/asymmetric systems in the process of merging (Clowe et al. 2006) and fossil groups (Schirmer et al. 2010) with luminosity functions unlike a Schechter (1976) function.

In this paper, we present our detection prescription, which involves a blind scan of colour–magnitude space (to locate cluster sequences) and a VT technique (to estimate the galaxy surface density distribution). Requiring only two bands to detect spectral breaks, our approach provides a very efficient method of detecting clusters in wide-area CCD imaging of the sky. Whilst algorithms have in the past used VTs to find clusters, previous attempts either do not exploit the red sequence or instead use photometric redshift distribution functions that rely sensitively on the absolute calibration and number of photometric bands (van Breukelen & Clewley 2009; Soares-Santos et al. 2011). In this paper, we describe the algorithm and apply it to a 7-deg^2 sample of SDSS equatorial stripe, Stripe 82, data. A companion paper (Geach, Murphy & Bower 2011, hereinafter GMB11) presents the full Stripe 82 catalogue covering the full 270 deg^2.

The outline of this paper is as follows. In Section 2, we define the data used for the cluster search in the SDSS and mock catalogues. Section 3 describes the algorithm step by step. Section 4 describes the application and testing of the algorithm using real astronomical data, followed by a brief comparison with existing cluster catalogues in Section 5. We describe the detection of mock clusters in simulated data in Section 6, followed by performance tests on the simulated catalogues. In Section 7, we summarize our findings.

Throughout, we assume a ΛCDM cosmology with $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$ and $h = H_0/100\,{\rm km\,s^{-1}\,Mpc^{-1}}$. For SDSS data, we use the Sloan photometric system (Gunn et al. 1998) and 'model' magnitudes.

2 DATA

2.1 SDSS Stripe 82

We extract SDSS DR7 *griz* photometry for all sources with extinction-corrected (Schlegel, Finkbeiner & Davis 1998) *r*-band model magnitudes, $r \leq 24$, in the deep co-add stripe centred on the celestial equator ('Stripe 82') from the SDSS Catalog Archive Server.[2] To minimize stellar contamination, we select only galaxies where the offset between the *r*-band point spread function (PSF) and model magnitudes satisfies $|r_{\rm PSF} - r_{\rm model}| > 0.05$. We exclude bright ($r_{\rm model} < 14$) galaxies and spurious sources such as overly deblended galaxies and fragmented stellar haloes.

Although no spectroscopic or photometric redshift estimates are used in detections, we post-process the cluster catalogue to estimate the redshift of each system. Cluster galaxies are assigned spectroscopic redshifts by matching source positions in the SDSS DR7, WiggleZ DR1 (Drinkwater et al. 2010) and 2SLAQ (Cannon et al. 2006) catalogues to within 1 arcsec. Where spectroscopic redshift data is unavailable, we use SDSS DR7 photometric redshifts (see Abazajian et al. 2009, and references therein). To increase both the source catalogue redshift completeness and the redshift accuracy for galaxies with no spectra, we supplement these data with additional photometric redshifts. We select all galaxies later identified by ORCA (The Overdense Red-sequence Cluster Algorithm) in the GMB11 Stripe 82 catalogue and estimate their redshifts using the HYPERZ code[3] (Bolzonella, Miralles & Pelló 2000) with *ugriz* model magnitudes and errors. The SDSS Stripe 82 input catalogue contains 11 358 087 galaxies with Galactic extinction corrected (Schlegel et al. 1998) *griz* model magnitudes, over $-50^\circ < \alpha < 59^\circ$ and $\delta = \pm 1\overset{\circ}{.}25$. In this study, we concentrate on a 7-deg^2 subregion within this catalogue, centred at $(\alpha, \delta) = (355\overset{\circ}{.}52, 0^\circ)$, comprising 291 389 galaxies (magnitude cuts applied to these galaxies for cluster detection are discussed in Section 3.7.1). This sample, covering the same area as the mock survey described below, was considered a large enough observational data set with which to test the algorithm. GMB11 describe findings from the ORCA catalogue based on the full 270-deg^2 data set.

2.2 Mock Pan-STARRS Medium Deep Survey catalogue

Cai et al. (2009) discuss the assembly of a light-cone from the Millennium Simulation (Springel et al. 2005) with a 3° opening angle, equivalent to a single pointing of the PS1, and the area of a single Medium Deep Survey (MDS) tile. The Millennium Simulation provides the ΛCDM architecture into which galaxies are populated using the Bower et al. (2006) semi-analytic GALFORM model (Cole et al. 2000). This creates a data set with PS1 *grizy* photometry for 2346 468 galaxies down to a magnitude limit of $r < 27.5$ (equivalent to the expected 5σ depth for the PS1 MDS) and a median redshift of $z = 1.05$. The similarity of the PS1 bands to the SDSS photometric system allows us to apply the same magnitude limits as those set for the Stripe 82 data (Section 3.7.1).

3 THE METHOD

In this section, we first outline and then detail the main components of the ORCA cluster detector.

3.1 Algorithm outline

Here we describe the main steps of the ORCA algorithm. With photometry in several bands, we calculate galaxy colours in consecutive ($g - r$, $r - i$, etc.) band pairs.

[2] http://casjobs.sdss.org
[3] http://webast.ast.obs-mip.fr/hyperz

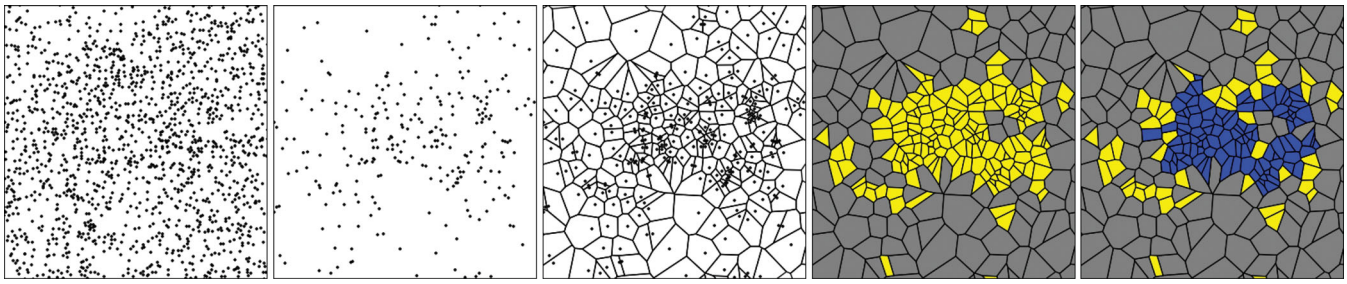

Figure 1. A depiction of the ORCA detector applied to a 9×9-arcmin^2 cut-out region of Stripe 82. Starting with all galaxies in the box (first panel from the left-hand side), a photometric selection (Section 3.2) isolates galaxies within a specific redshift range (second panel from the left-hand side); any clusters in this field will be evident as surface overdensities. In the third panel from the left-hand side, we compute the Voronoi diagram (Section 3.4) of the distribution to estimate the surface density of remaining galaxies. These are separated into overdense (yellow) and underdense (grey) cells in the fourth panel from the left-hand side, according to how likely they are to belong to a random distribution (Section 3.4). In the rightmost panel, we use a Friends-of-Friends percolation algorithm (Section 3.5) to connect overdense cells until the density of the whole system falls below a density threshold. Galaxies in the blue cells become members of a cluster if there are at least $N_{\rm min}$ linked members.

(1) We define a simple photometric selection using the colours and magnitudes of the sample. This selection could be simple, for example, a narrow slice(s) in colour–magnitude space(s), or a more complex selection function. This selection function can be modified in successive applications of the algorithm to blindly scan the full photometric space, and thus isolate red sequences across a range of redshifts (Gladders & Yee 2000, 2005).

(2) In each pass of the algorithm, we apply the photometric selection to the catalogue, thus greatly restricting the total number of galaxies under consideration. In the case of using two colours concurrently, this can be a very effective means of reducing foreground and background contamination of a putative cluster characterized by some red sequence.

(3) After the selection, we calculate the Voronoi diagram of the projected distribution of galaxies on the sky. The inverse of the area of each convex hull surrounding each galaxy can be used as an estimate of the local surface density.

(4) Galaxies residing in dense cells (satisfying some threshold criteria) can be connected together into conglomerations. If enough galaxies are joined together in this way, we define a cluster.

(5) In the blind scan, successive photometric cuts may select the same structures (since the adjustment of the selection is by design less than the typical width of a red sequence). Multiple detections of the same structure are identified and reduced to a single detection (we discuss how this was implemented in Section 3.6).

An illustrative overview of the above procedure can be seen in Fig. 1.

3.2 Photometric filtering

In large-scale imaging surveys, groups and clusters are apparent as overdensities in the projected distribution of galaxies. Cluster-detection methods reliant only on determining the projected galaxy density distribution are often plagued by two problems: (i) projection effects contaminating clusters with unassociated foreground and background galaxies; and (ii) the inclusion of spurious cluster detections arising from noisy data or chance projected overdensities.

To mitigate these problems, the contrast of genuine clusters can be enhanced by applying a photometric selection filter in colour–magnitude space, to isolate the red-sequence ridgeline. We parametrize our selection as a slice in colour–magnitude space, defined by a colour–magnitude normalization ($c_{\rm m20}$, the colour at 20th

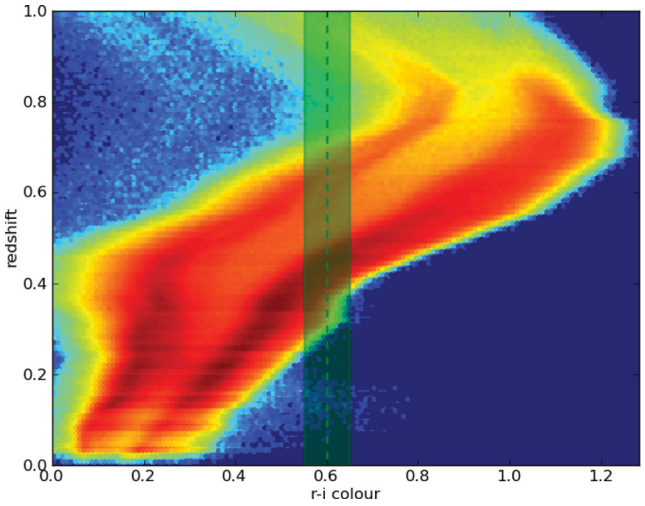

Figure 2. The redshift evolution of the observed-frame $r - i$ colour from a sample of mock galaxies. The colours indicate the density of galaxies at each point, with red being the highest. We are able to exploit this observed relation to isolate cluster galaxies within a specific redshift range by using a selection (such as the shaded strip in this figure) to select galaxies from a narrow colour range.

magnitude), slope $\beta(c_{\rm m20})$ and width $\sigma(c_{\rm m20})$. The expected evolution of red-sequence colours is constrained from simple stellar evolution models, meaning scans over an appropriate set of photometric selection filters allow for the isolation of clusters over a slew of redshifts. Fig. 2 shows the redshift evolution of galaxy colours in a sample of mock galaxies from Merson et al. (in preparation) and shows an additional advantage in using such filters. The two tracks visibly demonstrate the bimodality in galaxy colour that manifests itself as the 'red sequence' (lower track; Bower et al. 1992) and 'blue cloud' (upper track). By selecting galaxies within specific colour range Δc (as denoted by the green region in the figure), one may isolate red-sequence cluster galaxies within the redshift range Δz. Contaminants in this selection are bluer galaxies from higher redshifts. By simultaneously selecting galaxies from two photometric selections in different colours, one can eliminate degeneracies between colour tracks. We discuss this further in the following section.

The algorithm allows $\beta(c_{\rm m20})$ and $\sigma(c_{\rm m20})$ to adopt any values as the detector scans through colour–magnitude space. The simple

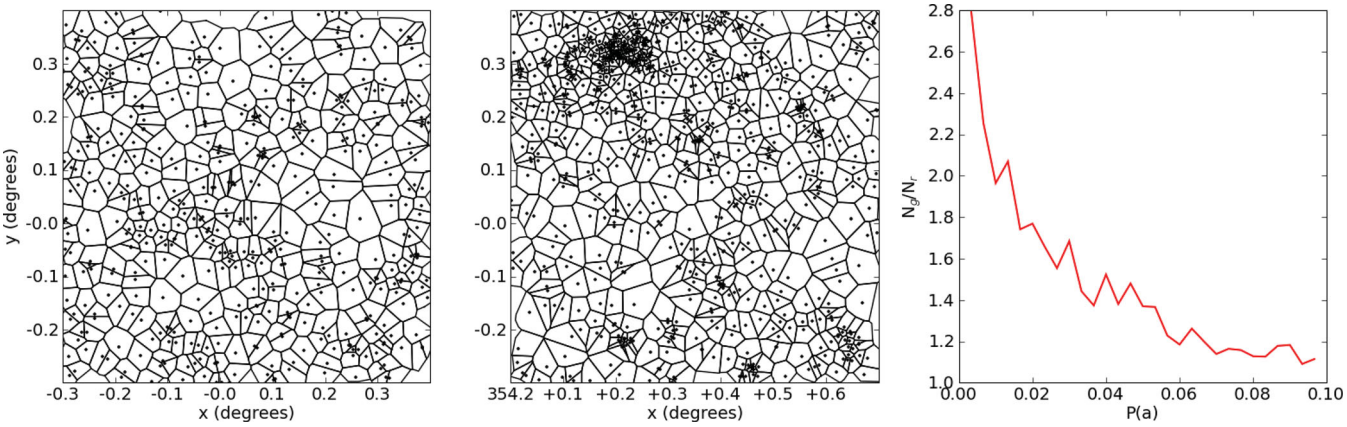

Figure 3. An illustration of the Voronoi technique described in Section 3.4. The left-hand panel is the Voronoi diagram of a random distribution of points. The middle panel is the equivalent diagram for galaxies in a field with the same mean density as the random field. The right-hand panel shows the ratio of galaxy cell counts to random cell counts for a range of values of the integral distribution of cell areas (equation 1 from Kiang 1966). There is a notable excess fraction of galaxy cells relative to random cells at low values of $P(a)$, permitting the use of a threshold to separate clustered galaxies from field galaxies.

prescription we adopt is that of a fixed slope and width with normalization. Although the observed-frame sequence slope is known to evolve with redshift (Gladders et al. 1998; Stanford, Eisenhardt & Dickinson 1998; Stott et al. 2009), our choice of photometric selection width encompasses a range of sequence gradients large enough to account for evolution as the algorithm searches to deeper redshifts. Analysis of mock clusters from the Millennium Simulation suggests this approach probes at least 2.5 (1.5) mag fainter (brighter) than the observed characteristic galaxy flux at the redshifts clusters are detected in this study. With measurements from a large ORCA cluster catalogue, future refinements to the algorithm may include a description of how the sequence slope varies with normalization $c_{\rm m20}$. The values adopted for β and σ are discussed in Section 3.7.

We scan through colour–magnitude space in a colour C_A from blue to red, placing down a series of M photometric selection filters $f(C_{A_1})$, $f(C_{A_2})$, ..., $f(C_{A_M})$ by increasing the normalization $c_{\rm m20}$ in small increments dc. The size of this increment, set in Section 3.7.1, allows adjacent filters to overlap, ensuring clusters close to the boundary of a filter are well sampled. Because each photometric selection isolates cluster galaxies (where they exist) from a specific redshift range, the detector can identify multiple clusters in the same line of sight. We determine the sensitivity of the algorithm to projection in Section 4.6.4.

3.3 Dual-colour photometric filtering

Although only one colour is necessary to detect clusters, Fig. 2 notes the colour–redshift degeneracy apparent in attempting to isolate a redshift regime from a single-colour selection. One can break the degeneracy and further reduce the field galaxy contamination by identifying the colour range cluster members have in a second colour C_B, and subsequently applying a series of joint photometric filters in both C_A and C_B. To establish the C_B colour range to scan, we take all cluster members from the preliminary detection (C_A only), detrend their sequence slopes and fit a Gaussian to the colour distribution. The C_B colour range ΔC_B is taken to be $\pm 1\sigma$ from the Gaussian mean.

If the Gaussian fit is poor, detection of a clear sequence in both C_B and C_A is less likely. In this case, ΔC_B is simply $\pm 1\sigma$ from the median of the C_B colour distribution. The algorithm then scans over this second colour range and attempts to detect the cluster in both colours.

A filter pair in C_A and C_B (hereinafter $\{C_A, C_B\}$) requires a detectable sequence in both colours, and amplifies the cluster signal by eliminating field galaxies in the C_A filter that fail to appear within the C_B filter. Any cluster in the final catalogue detected in C_A must therefore also have been detected in C_B. This improves the robustness of the algorithm and the reduction of contaminants from spurious detections. Because subfilters overlap in C_B colour–magnitude space, the same cluster may be detected in multiple filters. We apply the prescription described in Section 3.6 to identify and merge clusters that have been detected in more than one filter. The number of selection filters used to sample any colour range depends on the sampling interval dc set in Section 3.7.1.

3.4 Identifying overdensities with the VT

After increasing a cluster's detectability by suppressing field galaxies with photometric filters, the next step is to calculate the local surface density of each galaxy. Galaxies residing in common regions of enhanced density can then be grouped together into clusters. To quantify the surface density field, we divide the galaxies into Voronoi cells using QHULL[4] (Barber, Dobkin & Huhdanpaa 1996). The Voronoi diagram is a tessellation of convex hulls, or cells, with each galaxy occupying only one cell. All positions inside a given cell are closer to the cell's nucleus (the galaxy) than any other. Unlike many other detection techniques, the VT (for VT cluster detection, see Ebeling & Wiedenmann 1993; Ramella et al. 2001) does not smooth the data, is robust to cluster ellipticity (Plionis, Barrow & Frenk 1991) and can be applied to a variety of survey geometries. VTs do not suffer from spurious detections around survey boundaries and edges, and are thus well suited to analysing astronomical data with localized camera defects, excised bright stars and other sources of incompleteness. The left-hand and middle panels of Fig. 3, respectively, show the Voronoi diagrams for a random point distribution and galaxies with identical mean densities $\bar{\Sigma}$. Galaxies in more concentrated regions tend to have smaller cells.

We define the reciprocal of the galaxy cell area ($a_{\rm g}$) as an estimate of the galaxy's local surface density $\hat{\Sigma}_{\rm g}$. Searching for connected

[4] http://www.qhull.org

regions of high density identifies statistically significant structures. To determine if a galaxy resides in a high-density region of the survey, we evaluate the statistical significance of finding a cell of area a_g in a random field with mean cell area $\bar{a}_R$. We use the Kiang (1966) cumulative function for a Poissonian distribution of points:

$$P(a) = \int_0^a \mathrm{d}p = 1 - \mathrm{e}^{-4a}\left(\frac{32a^3}{3} + 8a^2 + 4a + 1\right), \quad (1)$$

where $a = (a_g/\bar{a}_R)$. The right-hand panel of Fig. 3 shows the distribution $P(a)$ for cells in an example galaxy field relative to a Poisson distribution of the same field size and number of points. Candidate cluster galaxies residing in overdense regions can be selected by cell areas statistically unlikely to arise in a random distribution. An excess of galaxy cells is apparent for low $P(a)$ compared to the random distribution. We identify all galaxies with $P(a_g) < P_{\mathrm{thresh}}$ in order to select a population of clustered galaxies. The choice of overdensity probability threshold is discussed in Section 3.7.2.

3.5 Connecting overdense regions to form clusters

Remaining galaxies belonging only to overdense cells are now grouped together to form clusters. We achieve this by applying a Friends-of-Friends algorithm to these cells. Rather than a distance criterion, we define a 'friend' as an adjacent Voronoi cell sharing at least one vertex. Potential clusters are seeded by ordering the cells with decreasing density, iterating through and connecting adjacent cells. These overdense regions grow by percolation until either no more adjacent overdense cells remain, or the mean cell density of the putative cluster

$$\bar{\Sigma}_{\mathrm{cells}} = N_{\mathrm{gal}} \sum_{i=1}^{N_{\mathrm{gal}}} \frac{1}{a_i} < \Sigma_{\mathrm{crit}}. \quad (2)$$

Groups of connected galaxies are classified as clusters if they have $N_{\mathrm{gal}} \geq N_{\mathrm{min}}$. The choice of the critical density threshold Σ_{crit} and N_{min} algorithm parameters is discussed in Section 3.7.1.

3.6 Producing a cluster catalogue

In Sections 3.2 and 3.3, we noted that adjacent photometric filters applied to the input catalogue overlap in colour–magnitude space. With this sampling strategy, the same cluster could be detected in multiple filters. Fig. 4 shows a sequence of VTs applied to the same area of the sky under photometric filters sensitive to different redshift ranges. Because colour-scans sample the colour range of a red sequence at a specific redshift, the cluster will be detected in multiple scans (with a peak contrast where the selection is most effective). In cases of clusters detected multiple times in different photometric filters, the 'best' cluster is identified and added to the final cluster catalogue.

For two candidates to be considered detections of the same system, they must have sufficiently similar spatial positions, red-sequence fits and cluster members. We quantify the similarity in cluster sequences using linear fits to the colour–magnitude relation (CMR) for the galaxies in each cluster detected. Sequence slopes can, in principle, adopt any value permitted by the width of the photometric filter (defined here as σ_f) it was selected in. We quantify the similarity between two sequences with the following criteria:

(i) *Sequence match 1* (ΔS_1). True if the sequence separation is $<0.5\sigma_f$ in colour for at least 25 per cent of the magnitude range $m_{\mathrm{BCG}} \leq m \leq m_{\mathrm{BCG}} + 5$.

(ii) *Sequence match 2* (ΔS_2). True if the sequence separation is $<\sigma_f$ in colour difference for at least 50 per cent of the range described in ΔS_1.

(iii) *Sequence match 3* (ΔS_3). True if the colour difference at 20th magnitude, (Δc_{m20}), between the two sequences is $<\sigma_f$.

(iv) *Sequence match 4* (ΔS_4). True if the clusters are detected in adjacent (overlapping) filters.

To define the similarity in cluster membership, spatial position and extent, we describe the *common-galaxy* fraction and *projection extent* for two clusters, CL_1 and CL_2:

(i) *Common galaxies* ($cg_{1,2}$). The fraction of galaxies in CL_1 that also belong to CL_2. Similarly, $cg_{2,1}$ is the fraction of CL_2 galaxies also appearing in CL_1. The BCG_{id} boolean notes when clusters share the same BCG.

(ii) *Projection extent* ($pe_{1,2}$). The fraction of galaxies in CL_1 that lie within the Voronoi cell boundaries of the CL_2 cluster. As with cg, $pe_{2,1}$ is the case for CL_2.

With these measures, five tests of 'cluster similarity' were devised (Table 1). A pair of clusters must pass *at least one* to be considered detections of the same system. Each of these tests accounts both for the spatial and for the colour characteristics of the clusters. Because no merging can proceed purely by colour similarity or spatial coincidence, this ensures the separation of associated but distinct systems, and clusters in projection. We balance these requirements with the need to prevent multiple instances of the same cluster appearing in the final catalogue. Where matches between two clusters exist, the thresholds in Table 1 make it likely the two systems will be merged.

To define the 'best' cluster from a list of candidates, we pick out the system with the largest *reduced flux* – the total flux (in

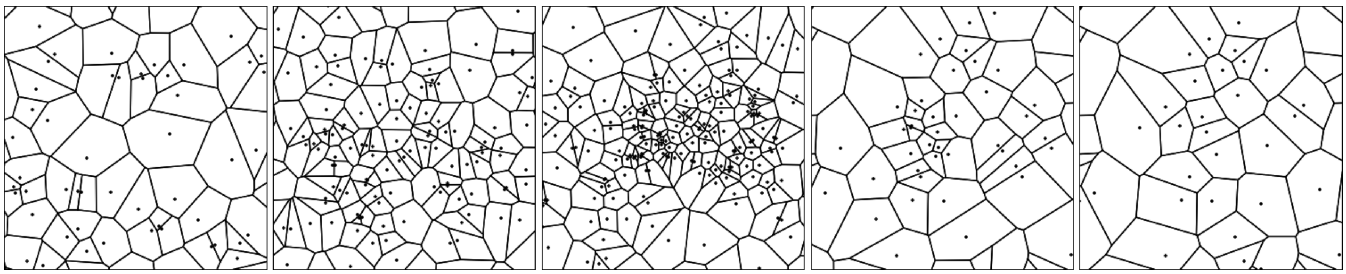

Figure 4. A sequence of Voronoi diagrams generated from galaxies in the same area of the sky, but selected from different photometric filters. A cluster signal is apparent for some filters, but is not apparent in others. This demonstrates the power of colour selection in isolating galaxies at specific redshifts. In cases where a cluster may be detected in more than one filter (such as the borderline detection in the second panel from the left-hand side), the algorithm must decide which cluster to select. This aspect of the detector is discussed in Section 3.6.

Table 1. The set of conditions used to consider whether two clusters are multiple detections of the same system. If any one of these conditions is satisfied, the algorithm picks the 'best' cluster of the two.

#	Constraint
1	$(cg_{1,2}$ OR $cg_{2,1}) \geq 0.5$
2	$(pe_{1,2}$ OR $pe_{2,1}) > 0$ AND ΔS_1
3	$\mathrm{BCG_{id}}$ AND ΔS_2
4	$(pe_{1,2}$ OR $pe_{2,1}) \geq 0.8$ AND ΔS_3
5	$(pe_{1,2}$ OR $pe_{2,1}) \geq 0.8$ AND ΔS_4

the detected band) of all but the three brightest cluster members. This prevents the selection of a cluster including one or two bright galaxies that may not be genuine members, but also makes the choice of best cluster largely independent of the BCG. Once the 'best' cluster is selected, the remaining candidates are discarded from the catalogue. However, to each cluster selected in this way, we attach a record of the candidate cluster galaxies that were not selected, forming an auxiliary catalogue of *associate cluster members*. In this way, we can keep track of galaxies the detector considered as members but did not include in the cluster. The degree of oversampling in colour space, and hence the number of multiple detections, depends on the sampling interval dc, relative to the width $\sigma(c_{\mathrm{m20}})$ of the filter. We set both of these parameters in Section 3.7.1.

3.7 Algorithm parameters

This section defines the values adopted for the algorithm parameters described in Sections 3.2–3.5.

3.7.1 Photometric filtering

In both mock and real data sets, we limit our search for clusters to three colours: $g - r$, $r - i$ and $i - z$. These are used to form joint selection filters combining two colours: $\{g - r, r - i\}$ and $\{r - i, i - z\}$.

Each photometric filter is described by a colour normalization c_{m20}, slope $\beta(c_{\mathrm{m20}})$ and width $\sigma(c_{\mathrm{m20}})$. For this study and that of GMB11, we demonstrate the detector with an unchanging filter slope and width. In order to set β and σ for each colour, 126 members of Abell 2631 (Abell et al. 1989) are visually identified in an i, r and g composite Stripe 82 image. At redshift $z = 0.278$ (Böhringer et al. 2000), this system is the richest Abell cluster in Stripe 82 and shows evidence of a clear sequence in all three colours used in this study.

A linear fit to the colour–magnitude sequence was applied to determine β for each colour. The filter widths were set using a method akin to that described in Gladders et al. (1998); we first remove the slope in each sequence and then exclude 3σ outliers. Starting at the line fitted to the cluster sequence, we increase the width in equal amounts above and below this line until we enclose 90 per cent of the remaining members. We define this as the filter width σ for that colour.

Fig. 5 shows the colour–magnitude sequence of the identified members in the three colours (top panels) compared to a field of

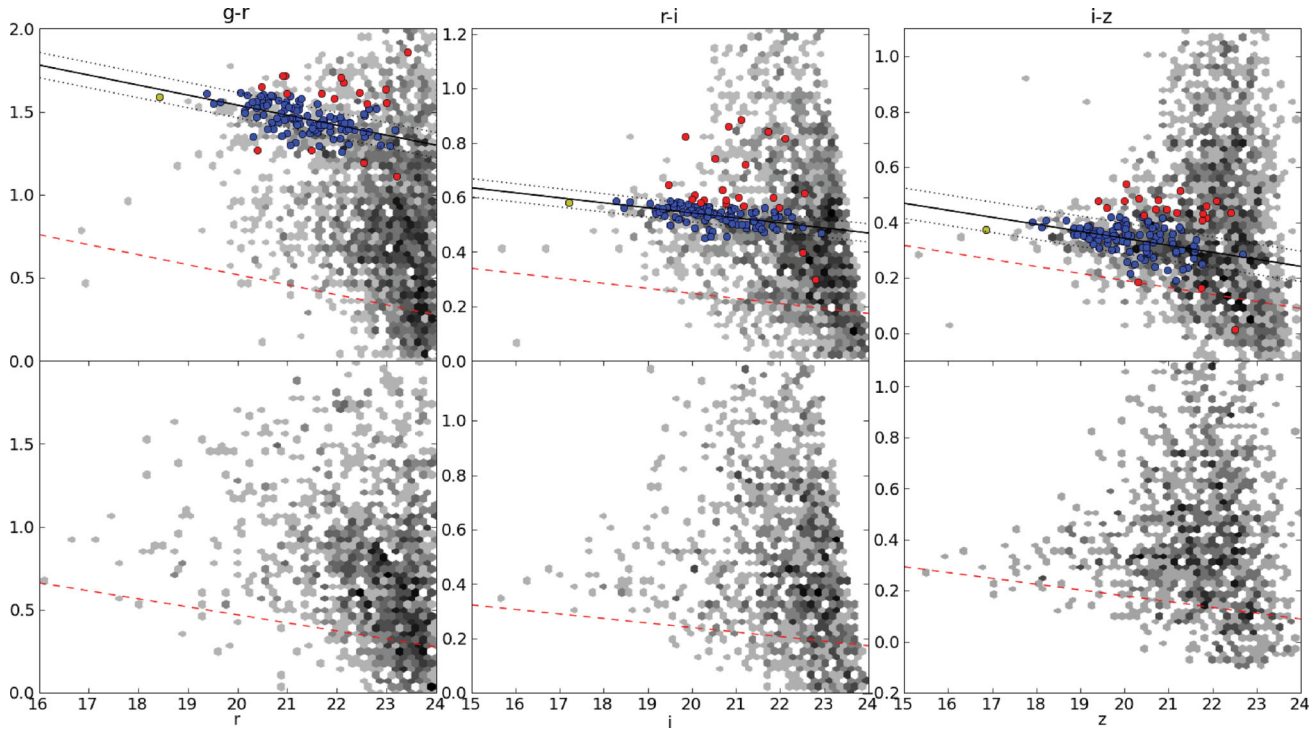

Figure 5. Top panels: colour–magnitude diagrams for the 126 Abell 2631 members selected in this study. The yellow dot notes the position of the cluster r-band BCG. The black lines denote photometric selection filter fits to the data and indicate the slope (β), normalization (solid line, c_{m20}) and width (dotted line, σ). The identified members are split into those inside (blue) and outside (red) the 3σ cut used to estimate the filter width. The grey data indicate all galaxies that were not identified as members of the cluster out to a radius of 7 arcmin from the cluster centre. The red dashed line in the $g - r$ colour indicates the blue limit imposed by the Virgo cluster, and the equivalent lines in $r - i$ and $i - z$ denote the lowest c_{m20} identified from cluster sequences in our search of the 7-deg^2 Stripe 82 survey. Bottom panels: the colour–magnitude diagrams for galaxies in a region of the same area located in a field environment.

Table 2. Filter parameters fitted from Abell 2631, the ranges searched and the number of filters in each colour. The blue limit in $g - r$ corresponds to an extrapolation of the Virgo CMR, whilst the others permit a full sweep of the available data. The emboldened figure is the largest filter width (σ_f) and is adopted for all colours.

Colour	Slope (β)	Width (σ_f)	Range	Filters
$g - r$	−0.048	**0.152**	0.47–2.00	39
$r - i$	−0.017	0.067	0.00–1.22	38
$i - z$	−0.023	0.110	−0.10–1.10	31

the same area with no cluster present (bottom panels). The blue (red) points identify members that were inside (outside) the 3σ cut used to identify outliers. The grey data correspond to galaxies that were within 7 arcmin of the cluster centre and not picked as cluster members. Table 2 lists the fitted filter parameters for each colour (corresponding to the black lines in Fig. 5) in addition to the colour range and number of filters used in our cluster search. Following our decision in Section 3.2 to use a fixed slope, we adopt the largest filter width (σ_f, 0.152) for all colours, and use this to define the input galaxy magnitude limit for each band. Magnitude limits are applied to reduce the number of input galaxies with high levels of photometric uncertainty. We set these as the faintest magnitude where the photometric uncertainties fall below $0.68\sigma_f$.

We set limits for each band based on a sample of 100 000 galaxies from Stripe 82. Fig. 6 shows the galaxy photometric error distribution for the r band, and from this, we set a magnitude limit of $r \leq 23.5$. This is slightly more conservative than the limit implied by the error distribution ($r \leq 23.8$) because we aim to include only sources with good photometry. The magnitude limits applied are 24.0, 23.5, 23.3 and 21.6 in the g, r, i and z bands, respectively, resulting in a source catalogue of 69 797 galaxies. With the added depth from Stripe 82 photometry, these limits permit an exploration of the red sequence to at least 2.5, 3 and 1.5 mag fainter than $M_\star$, respectively, for the r, i and z bands. As part of the algorithm design, we considered multiple searches through the data at different flux limits.

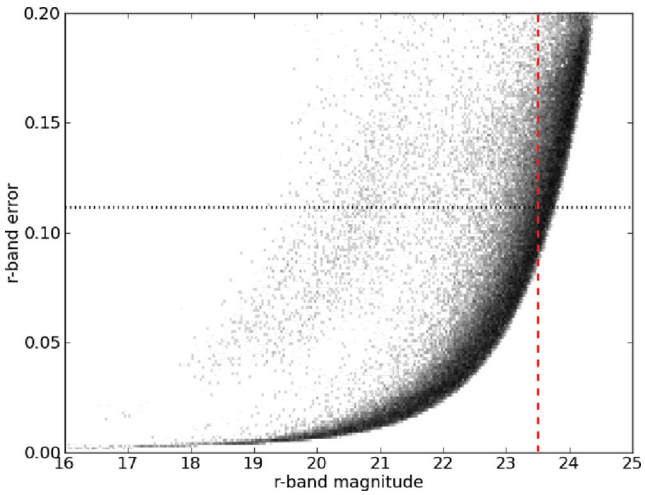

Figure 6. The SDSS model r-band photometric error in a sample of 100 000 Stripe 82 galaxies. These data are used to set a magnitude limit where at least 50 per cent (0.68σ, black horizontal dotted line) of the faintest galaxies remain in a colour slice of width $\sigma_f = 0.152$. Whilst the data suggest a limit of $r \leq 23.8$, we opt for a slightly more conservative $r \leq 23.5$ limit (red vertical dashed line).

Under this prescription, higher signal cluster sequences would be selected when re-detections of the same system were merged. In tests with the mock light-cone data analysed in Section 6, we found no significant advantage in this implementation, and instead kept our magnitude limits fixed.

The bluest filter pair we employ is $g - r$. To prevent the detection of spurious systems bluer than the $z = 0$ red sequence in this colour, we determine a blue limit by extrapolating the CMR for Coma (Smith et al. 2009) and Virgo (Rines & Geller 2008) to $r = 20$. The c_{m20} normalization for Coma (Virgo) was estimated as 0.6 (0.47); we use the latter as the bluest filter possible in the $g - r$ colour. We do not apply similar limits to the other colours, but the normalization below which no sequences were detected in $r - i$ and $i - z$ is described in Section 4.1. Fig. 5 shows these limits as the red dashed lines.

Finally, the detection algorithm uses photometric filters that overlap in colour–magnitude space, preventing clusters close to filter edges from being poorly sampled. A sampling interval in colour space of $dc = 0.04$ is chosen, corresponding to an overlap of approximately 75 per cent between adjacent filters based on σ_f, the filter width.

3.7.2 *VT and connection of overdense regions*

The initial identification of clusters in projected high-density regions and the subsequent percolation of their members depends, respectively, on the probability threshold P_{thresh} and the critical density Σ_{crit}. We parametrize the critical density Σ_{crit} as a scalar multiple of $\bar{\Sigma}$ such that both detection parameters have a mean density dependence. In the left-hand panels of Fig. 7, we note the effect a range of (P_{thresh}, Σ_{crit}) combinations have on the recovery of Abell 2631 within a box of scale 13.6 arcmin. By tracking the detector's assignment of Voronoi cells to clusters and fields, we compare members visually identified to the recovery of this cluster under different parameter combinations. The cells are colour-coded into four groups to differentiate detected and visually identified members. The grey cells show galaxies neither detected nor identified as cluster members. The green cells denote detected members that were also visually identified, orange for where the detector did not assign cluster membership despite our classification as such from the imaging, and finally red cells are detected members not visually identified as members. We stress the latter group in no way indicates the purity of the cluster, as we are both incomplete and subjective in our identification of genuine cluster members. However, this exercise does provide a useful indication of detector performance when compared to our visual impression of cluster membership.

The detection grids show re-detection is broadly insensitive to the range of parameters explored. At higher probability thresholds (increasing row number), the cluster expands to form a more extended structure. This growth is moderated by the introduction of a minimum cell density. We exclude $\Sigma_{crit} = 20\bar{\Sigma}$ as it removes a significant fraction of visually identified members on the periphery of the cluster. The middle ground between detecting a more compact system ($P_{thresh} = 0.005$) and potentially increasing the interloper fraction ($P_{thresh} = 0.015$) suggests the balance of detection completeness and cluster purity lies with $P_{thresh} = 0.01$. We note from Fig. 3 there are at minimum twice as many clustered cells as unclustered at $P(a) \leq 0.01$. Although $(0.01, 0\bar{\Sigma})$ and $(0.01, 10\bar{\Sigma})$ appear identical in their recovery of the cluster, we require a non-zero density constraint to filter out spurious low-amplitude systems and prevent large clusters from percolating into giant connected structures.

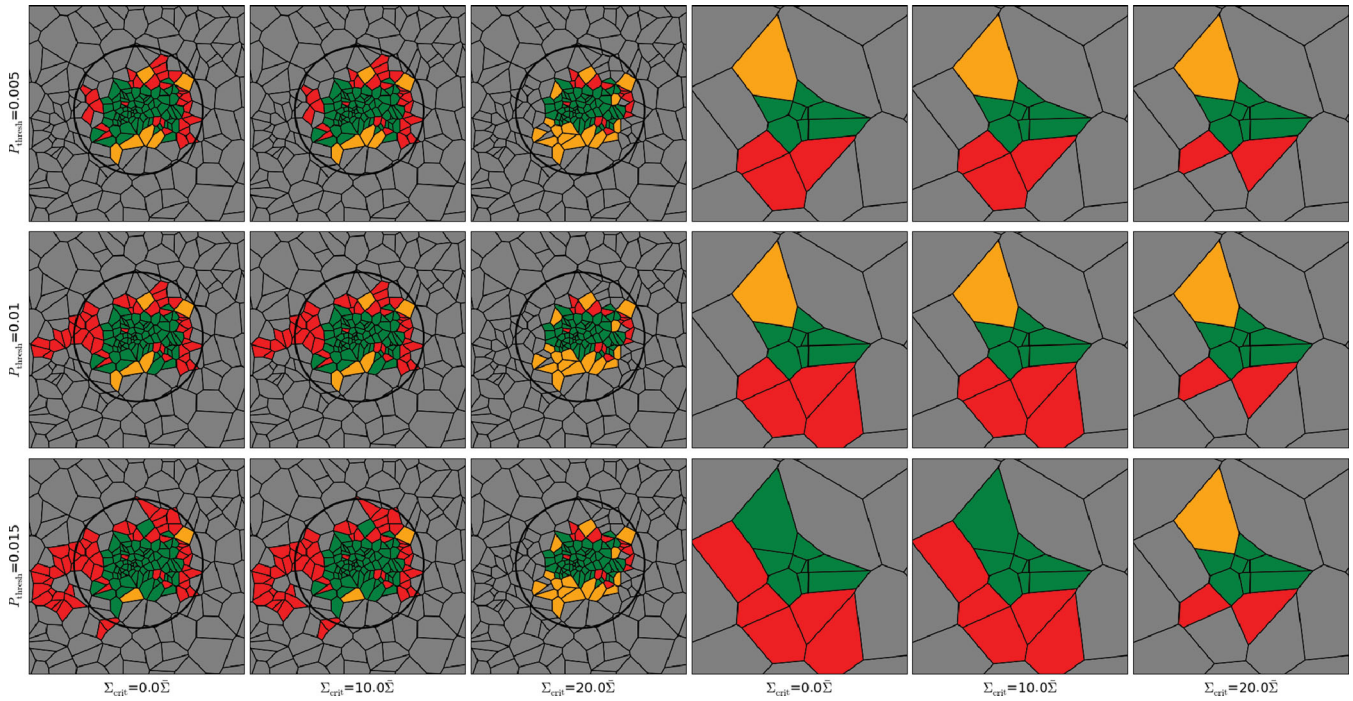

Figure 7. Effect of detection parameters on Abell 2631 (left-hand panels, box scale 13.6 × 13.6 arcmin^2) and a compact group (right-hand panels, box scale 3.5 × 3.5 arcmin^2). Colour key: the grey cells are cells with field galaxies and the green cells are galaxies identified by the algorithm that were also visually identified as members. The red cells are members assigned to the cluster by the detector but not visually identified as cluster members. The orange cells are galaxies that failed to be correctly identified by the algorithm as cluster galaxies, but were defined as such visually. The circle around Abell 2631 corresponds to a 1 h^{-1} Mpc radius at the cluster redshift.

We consequently adopt the parameter combination $(P_{\rm thresh}, \Sigma_{\rm crit}) = (0.01, 10\bar{\Sigma})$. To ensure these parameters are not biased to the detection of high-mass systems, we use 11 members of a visually identified compact group to perform a re-detection in the same parameter ranges. The right-hand panels in Fig. 7, with boxes of scale 3.5 arcmin, show the recovery of this group, and indicate group scale detection is robust to the range of parameters explored. The trade-off between completeness and purity is similarly evident here, with $(0.01, 10\bar{\Sigma})$ remaining a good compromise between the two.

In both cases (and more generally), there is a tendency to underestimate the total number of cluster members. This arises from an inevitable feature of Voronoi diagrams implying the algorithm is unlikely to recover all cluster members. The suppression of the field galaxy population with photometric filters causes an abrupt drop in galaxy surface density at the cluster boundary. Because the Voronoi cells of peripheral members have a limited number of field galaxies to constrain their boundaries, they adopt larger areas. Such cells may then be rejected as members because their areas are inconsistent with that population. Nevertheless, tests with mock catalogues allow us to quantify the impact this effect has on the cluster purity, as discussed later in Section 6.

Finally, we set the minimum membership of a cluster, $N_{\rm min}$, to five galaxies.

4 SDSS EQUATORIAL STRIPE 82 CLUSTER CATALOGUE

4.1 The catalogue

We applied the detector to a 7-deg^2 sample of Stripe 82, using the limits described in Section 2 and parameters described in Section 3.7. Here we describe the general characteristics of this catalogue, perform a series of tests on the data, and briefly compare our detections to existing optical and X-ray-detected clusters.

After applying the magnitude limits described in Section 3.7.1, a source catalogue of 69 797 galaxies are analysed by the algorithm. We find a total of 97 clusters, identifying a total of 1293 cluster galaxies (0.5 per cent of the original galaxy sample) and 813 associate cluster members (candidate cluster members that were not selected). Of these clusters, 34 per cent were detected in the $\{g-r, r-i\}$ combination and 66 per cent in the $\{r-i, i-z\}$ combination.

Although we define a blue limit for the $g-r$ CMR ($c_{\rm m20} > 0.47$), equivalent limits were not applied to the $r-i$ and $i-z$ colours. We can, however, place upper bounds on the blue limit in these colours by noting no clusters were detected below $r-i = 0.24$ and $i-z = 0.18$. Such limits serve to reduce the search time for future survey scans.

Table 3 shows an extract of the cluster catalogue. This 7-deg^2 sample of 97 Stripe 82 clusters is available online.[5] Each cluster is named according to the IAU convention, in the form MGB JHHMMSS+DDMM.m. We detail below the main features of both catalogues.

4.2 Cluster positions and redshifts (`cluster_z`, `cz_type`)

The `ra` and `dec` position quoted in the catalogue is the algorithm estimate of the centre of each cluster, based on the average positions of their members.

[5] http://orca.dur.ac.uk/

Table 3. A sample of the ORCA cluster catalogue generated in this study. Full details of the columns can be found in Sections 4.1–4.5. The first column contains the cluster name based on the IAU convention. The second and third columns note the J2000 estimated cluster positions in degrees. The fourth and fifth columns describe the cluster redshift and source data used to calculate the redshift. The sixth column notes how many members were found in the cluster, and we provide estimates for the cluster B_{gc} richness and sequence scatter in the seventh and eighth columns, respectively. The final two columns indicate the radius (in degrees) enclosing 80 per cent of the cluster members and the ratio of this value to the 20 per cent radius, a measure of cluster concentration.

Name	ra	dec	cluster_z	cz_type	N_{gal}	b_gc	scatter	θ_{80}	C
MGB J234017−00030.9	355.06912	−0.06455	0.245	c0s0w0q0d0b0p6h2	6	19 416	0.047	0.0001	1.700
MGB J233817+00190.0	354.56897	0.33309	0.208	c0s0w0q0d0b0p8h6	8	94 461	0.038	0.0003	3.667
MGB J234113−00000.4	355.30349	−0.00597	0.166	c0s0w0q0d0b0p6h2	6	182 181	0.018	0.0003	1.692
MGB J234400−00300.3	355.99952	−0.50461	0.181	c0s1w0q0d0b0p5h4	6	71 831	0.025	0.0001	1.750
MGB J234725+00190.7	356.85322	0.32867	0.201	c0s0w1q0d0b0p14h14	14	10 967	0.037	0.0004	2.545

Although we do not use any redshift data to generate our cluster catalogue, we provide redshift estimates for each system detected by the algorithm. These redshifts are weighted towards members with spectroscopic data, but two sets of photometric redshift data (HYPERZ and the DR7 photometric estimate) are used to provide each cluster galaxy with at least one redshift estimate. From the catalogue of 1293 cluster galaxies, 2.6 per cent have spectroscopic data (DR7 spectroscopic redshifts, WiggleZ and 2SLAQ), 93 per cent have DR7 photo-zs and 87 per cent have HYPERZ estimates. The HYPERZ estimates for cluster members were generated using only S0 and E SEDs, a Calzetti et al. (2000) reddening law and a two-stage convergence (over and above that performed by HYPERZ) to the redshift where a range identified in coarse redshift bins is resampled with a smaller bin width. Comparing these estimates to available spectroscopic redshifts, the measured error dispersions are higher in HYPERZ than in the DR7 pipeline (0.029 versus 0.016).

We calculate each cluster redshift by determining the weighted median redshift from the available member data. The weighting for a spectroscopic redshift, DR7 photo-z and HYPERZ redshift is 4, 2 and 1, respectively, the higher weighting for DR7 photo-zs reflecting the smaller error dispersion mentioned above. To gauge the accuracy of our redshift estimate, we note the calculated redshift of Abell 2631 is $z = 0.26$, some 0.02 lower than the value determined by Böhringer et al. (2000). The median cluster redshift of the whole catalogue is $z_{med} = 0.31$, and the maximum redshift is $z = 0.57$. Approximately, 25 per cent of the clusters have at least one member with a spectroscopically measured redshift.

Without access to spectroscopy, accurate photometric redshifts of red-sequence cluster galaxies are good measures of cluster redshifts. We quantify this in Fig. 8 by comparing the photometric and spectroscopic redshifts of cluster BCGs from a sample of the full GMB11 Stripe 82 cluster catalogue with spectroscopic redshifts. After removing a small systematic trend and 3σ outliers, the 1σ dispersion in $(z_s - z_p)/1 + z_s$ is 0.0157 (increasing to 0.0163 when ignoring the systematic error). This suggests BCG photometric redshifts are accurate estimates of the cluster redshift.

The `cz_type` property is a shorthand description of the available redshift data for each cluster, each letter defining a measurement type, followed by the number of that type. The letters denote data from the mo(c)k, DR7 (s)pectroscopic, (w)iggleZ, 2SLA(q), DR7 (p)hotometric and (h)YPERZ data sets, where mock is of course not used in these observational data.

4.3 Cluster richness (`b_gc`)

With access to cluster redshifts, we are able to calculate the B_{gc} optical cluster richness, a robust parameter known to correlate with cluster mass. We use the B_{gc} measure described in Yee & López-Cruz (1999):

$$B_{gc} = \frac{\rho_{bg} D(z_{cl})^{\gamma-3} A_{gc}}{I_\gamma \Phi(M_3, M_3 + 3, z_{cl})}, \quad (3)$$

where ρ_{bg} is the background surface density of all source catalogue galaxies (irrespective of their colour) inside a $0.5\,h^{-1}$ Mpc radius with luminosities between the third BCG (M_3) and 3 mag fainter. The integrated luminosity function, $\Phi(M_3, M_3 + 3, z_{cl})$, is measured over the same luminosity range. We evolve the $z = 0.1$ Blanton et al. (2003) SDSS r-band luminosity function ($\phi_\star = 1.49 \times 10^{-2}, M_\star = -20.44, \alpha = -1.05$) using the prescription described in Lin et al. (1999) that adds redshift-dependent terms to $\phi_\star$ and $M_\star$ with parameters $P = -1.06$ and $Q = 1.82$. D, the angular diameter distance, is derived from the cluster redshift z_{cl}. γ and I_γ, respectively, define the slope of the angular galaxy correlation function and the integration constant arising from deprojecting the

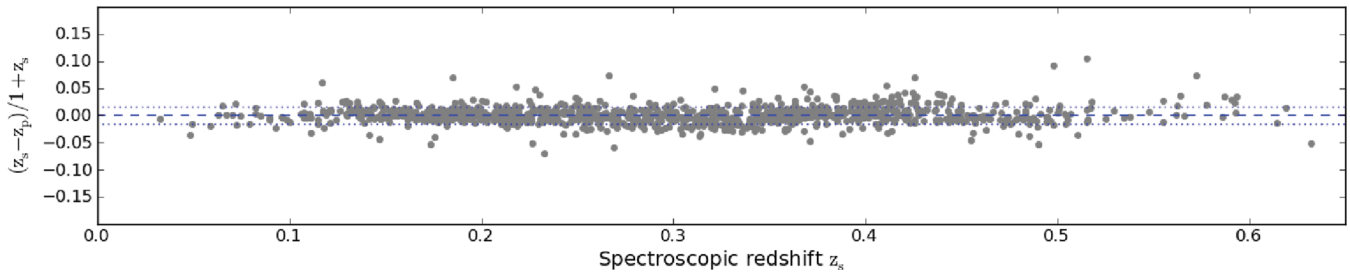

Figure 8. Comparison of photometric redshift accuracy $\delta_z(z_s) = (z_s - z_p)/1 + z_s$ for the cluster BCGs with spectroscopic redshifts. After outlier rejection (clipping galaxies with $|\delta_z| > 3\sigma_{\delta_z}$, or 0.4 per cent of the total sample) and removing the slight systematic photoredshift error, we find a 1σ scatter $\sigma_{\delta_z} = 0.0157$ (denoted by the dotted blue lines). This highlights the excellent redshift recovery using *ugriz* photometry alone. For a given cluster, we combine both the photometric and (where available) the spectroscopic redshifts of cluster members to derive a robust redshift estimate for the system as a whole.

cluster. We set these to $\gamma = 1.77$ and $I_\gamma = 3.78$. The correlation amplitude $A_{\rm gc}$ is defined as

$$A_{\rm gc} = \frac{N_{\rm net}}{N_{\rm bg}} \frac{(3-\gamma)}{2} \theta^{\gamma-1}, \quad (4)$$

where $N_{\rm net}$ is the background-corrected count of galaxies within the luminosity range described above, out to an angular separation θ that corresponds to $0.5\,h^{-1}$ Mpc at the cluster redshift. $N_{\rm bg}$ is the background galaxy count within this radius, estimated from the mean density of galaxies across the whole field. The full 270-deg^2 Stripe 82 catalogue provides additional definitions of cluster richness – we refer readers to GMB11 for the details of those measurements.

4.4 Cluster-sequence scatter (`scatter`)

To estimate the width of a detected cluster's sequence, we first make a fit to the slope of the sequence and remove the tilt. Using cluster members between $m_{\rm BCG} \leq m \leq m_{\rm BCG} + 3$, we estimate the sequence scatter by making a 2σ clip in the colour distribution.

The robustness of the red-sequence fit is sensitive to the number of members in the detection. Based on a bootstrap-resampling of the cluster sequences, we find the fitting procedure is robust in clusters with at least eight members. Below this, sequence scatter estimates are dominated by fitting uncertainty. For systems of at least 10 members, the characteristic error in the sequence scatter is 34 per cent, dropping to 19 per cent for clusters with up to 30 members and 8 per cent for those with at least 50 members. Future catalogues will provide improved estimates of the sequence-fitting error.

4.5 Projected scale (θ_{80}) and concentration (C)

For each cluster, a projected scale size θ_{80} is provided. This is calculated as the angular radius (in degrees) enclosing 80 per cent of cluster members from the centre.

A measure of the projected concentration (C) is determined by comparing the radius enclosing 80 per cent of the cluster members to the radius enclosing 20 per cent. High values of θ_{80}/θ_{20} indicate a centrally concentrated cluster.

4.6 Testing the algorithm

4.6.1 Cluster re-detection robustness

To determine how robust the detector is to catalogue incompleteness, we attempt re-detections of the Abell 2631 cluster after removing a random selection of members from the source data. Our sole constraint is that the cluster BCG remains in the source data. In the following analysis, we only consider the detected cluster closest to the original Abell 2631 position. Robustness is defined as the fraction of members detected in the new cluster from those remaining in the input catalogue. We use a test $g - r$ photometric filter that adopts a β_{g-r}, $c_{\rm m20}$ and σ_{g-r} best suited to the recovery of Abell 2631, selecting approximately 85 per cent (108) of the visually selected members. We experiment with removal fractions down to 95 per cent, corresponding to the largest fraction still retaining $N_{\rm min} = 5$ original cluster members in the sample.

50 random realizations of a depleted input catalogue are generated for each removal fraction, yielding a median recovery rate based on members that could have been added to the cluster. The solid blue line in Fig. 9 shows how increasing the removal fraction affects the fraction of cluster members recovered; error bars on this line represent 1σ uncertainties from the 50 re-detections in each bin. The recovery fraction when no galaxies have been ejected is ∼93 per cent of the 108 Abell 2631 members located inside the photometric filter. The other 7 per cent were rejected by the algorithm because either their Voronoi cells have insufficient densities to join the overdense collection of cells ($P_{\rm thresh}$, see Section 3.7.2) or their inclusion causes the percolating cluster to drop below the critical density ($\Sigma_{\rm crit}$).

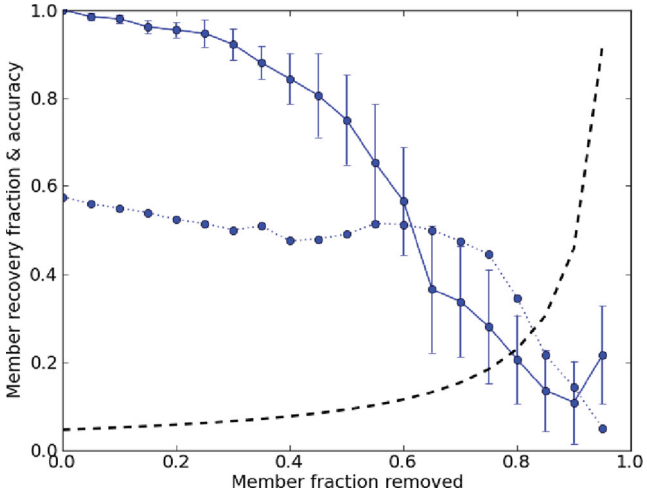

Figure 9. The recovery fraction (solid line) and recovery accuracy (dotted line). Some Abell 2631 cluster galaxies are randomly removed from the source catalogue, and the fraction subsequently identified in a re-detection of the cluster is the recovery fraction, with error bars of 1σ uncertainty calculated from 50 re-detections. The fraction of visually identified Abell 2631 galaxies making up the re-detected cluster defines the recovery accuracy. The fraction required to produce an $N_{\rm min} = 5$ member system is denoted by the black dashed line.

We take into account this intrinsic detection inefficiency, quoting yields from the cluster re-detection relative to the ∼93 per cent of members recovered where no additional galaxies are removed. Unsurprisingly, the fraction of detected members located in the cluster drops as more members are excised. However, over 75 per cent of the remaining members are re-detected even after half of the cluster is removed. Approaching larger removal fractions, the fragmentation of cluster members into spatially distinct groups hinders recovery of the complete set. The black dashed line in this plot corresponds to the minimum recovery fraction required to identify $N_{\rm min} = 5$ original members from the input data. The algorithm can robustly identify the original cluster down to an 80 per cent removal fraction, corresponding to 22 of the original 108 galaxies. Below this limit, an insufficient number of cluster members are recovered by the detector to identify a cluster associated with the halo.

For each ejection fraction, we also calculate the recovery accuracy: the fraction of visually identified Abell 2631 galaxies making up the re-detected cluster. The dotted blue line in Fig. 9 shows this parameter. The initial accuracy (no members are removed) is approximately 60 per cent, providing some estimate of our level of incompleteness when visually identifying cluster membership. As more members are removed, there is a gradual reduction in accuracy, implying replacement of these members with other galaxies becomes more commonplace. At large (>70 per cent) removal fractions, fragmentation acts to reduce the connectivity of cluster members, increasing the number of contaminant galaxies that share the photometric filter.

4.6.2 *Cluster displacement and edge effects*

A cluster detector should identify systems irrespective of the projected environment they are located in. Ideally then, recovery of identified members is achieved even if the system is moved to another position.

To determine the sensitivity of cluster identification to localized background fluctuations, we shift source data positions of known cluster members to a random location, keeping their spatial distribution intact. A buffer is created around the survey edge to ensure no cluster members are displaced outside the boundaries, then a re-detection of the cluster is attempted. The re-detection performance is quantified by the recovery efficiency – the fraction of original members in the new cluster, and the recovery accuracy remains as defined in the previous test.

Fig. 10 shows the recovery efficiency (solid blue line) and recovery accuracy (dotted blue line) for clusters spanning more than an order of magnitude in membership ($N_{\rm min} = 5$ to 174 galaxies). If there was a choice of cluster for a membership bin, we used the system with the smallest sequence scatter to determine the impact of displacement on the best candidate in that membership group. Each cluster was re-detected in the pair of selection filters it was originally identified in, meaning a re-detection with no displacement would yield a perfect recovery efficiency and *recovery accuracy* (both equal to unity). We perform 50 random displacements for each of the selected clusters, using their scatter to derive 1σ uncertainties from the mean. The black dashed line in Fig. 10 corresponds to the recovery fraction required to detect $N_{\rm min} = 5$ galaxies of the original system from each displaced cluster.

For the majority of cluster sizes, recovery accuracies are approximately constant at ∼90 per cent, meaning 10 per cent of the cluster members are background galaxies selected in the same photometric selection. Recovery efficiency data suggest the detector makes significant cluster re-detections for systems down to 10 members, but smaller groups are susceptible to higher levels of contamination and fragmentation. Our example case of Abell 2631 (at $\log_{10} N_{\rm gal} \sim 2.1$), with a recovery efficiency of 80 per cent, is approximately 13 per cent lower than the recovery fraction from robustness test calculated above. A recovery accuracy of ∼86 per cent is consistent with the detector swapping 13 per cent of the original members with background galaxies when the cluster is moved.

We next establish how survey edges bias the detection of systems at the boundaries. Using the same set of clusters, we repeat the above experiment, specifically placing systems close to the survey edges to quantify the impact of edge effects on group and cluster recovery. When moving each cluster, we ensure no members are outside the survey boundary. The average separation between the survey edge and the member farthest from the cluster centre is around 23 arcsec.

Galaxy cells at the boundary of a Voronoi diagram are unbounded, often resulting in very large cell areas. This may hamper the identification of low-membership clusters, where a member with a cell area exceeding the probability threshold may preclude the cluster from detection. Random positions are selected along any one of the four sides of the survey (allowing clusters to reside in a corner). In our source catalogue, the declination boundaries (at $\delta = \pm 1.^{\circ}25$) are set by the geometry of the stripe, whilst the right ascension boundaries are artificially defined. Distances between the cluster centroid and survey edge are large enough to include all members within the survey. The red line in Fig. 10 shows the recovery efficiency based again on 50 randomized displacements. This distribution is very similar to that of the displacement test above, suggesting edge effects do not hinder the recovery of clusters any more than the displacement of the members themselves. This is particularly significant at group scales, where the exclusion of one or two members could prevent the detection of the system.

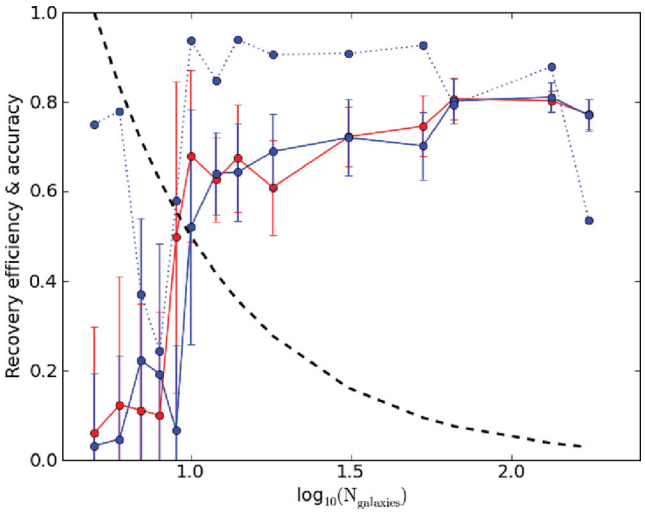

Figure 10. The algorithm's re-detection capability when a cluster has been moved to a random position. The recovery efficiency (solid blue line) is the fraction of original cluster galaxies found in the displaced cluster. The edge-effect recovery efficiency (red line) shows a similar test, instead moving the cluster to a random position near the survey boundary. Uncertainties in both lines are 1σ errors from 50 re-detections. The recovery accuracy (dotted blue line) is the ratio of input cluster members to the member count of the re-detected cluster. The black dashed line indicates the $N_{\rm min} = 5$ threshold required to secure a robust detection of the cluster's halo.

4.6.3 *False-positive detection rate*

We set the detector the task of attempting to detect spatially clustered systems with randomized colours. This establishes the importance of red sequences to cluster detection with this algorithm and provides an estimate of the false detection rate. We run the detector on the source catalogue in the same manner as before, having first shuffled the colours, so while cluster members still reside in high surface density regions, they no longer have red sequences. We identified two 'clusters' (with five and six members) in the 7-deg^2 survey, both located at the positions of original high-membership ORCA clusters. To ensure this calculation is uninfluenced by the size of the survey, we repeat this process on the full Stripe 82 data set ($-50^{\circ} < \alpha < 59^{\circ}$) covering 270 deg^2. The algorithm detects 15 'clusters' from these data, each consisting of five- or six-member groups. From this, we infer the number of spurious systems detected per 7 deg^2 is 0.39.

In a similar fashion, we next randomize galaxy positions while keeping the colours the same. This means cluster red sequences remain intact as the algorithm scans through colour–magnitude space, but points clustered in colour are no longer clustered in the sky. The algorithm detected four 'clusters' over the full 270-deg^2 Stripe 82 data set, implying an ∼0.1 per cent spurious cluster detection rate.

Both exercises suggest the detector cannot identify clusters without correlations in both colour and spatial position. Moreover, the probability of detecting systems based on random distributions of both colour and position is below 1 per cent.

4.6.4 *Projected cluster-pair resolution*

The ideal algorithm can identify two clusters with the same angular position on the sky, but at different radial distances. Using the $c_{\rm m20}$–z relation demonstrated in Fig. 2, one can, in principle, isolate superimposed systems by identifying them in different filters. Within a detection filter $f(C_A)$ of width $\sigma_{\rm f}$, two spatially coincident systems will be merged even if their sequences do not directly overlap. We overcome this limitation by splitting sequences in the next colour (C_B) through the application of joint filters (Section 3.3). The resolving power of the algorithm in projection is therefore limited by the merging of separate clusters that are mistaken as multiple detections in Section 3.6.

We test this effect with the same clusters as used in Section 4.6.2 by implanting a seven-member test cluster at the same spatial position and colour normalization $c_{\rm m20}$. We increase the test cluster C_B colour normalization by $\delta c_{\rm m20}$ and run the matching algorithm. This is repeated until the detector classifies the reddened test cluster as an independent system. The *resolving capability* of the algorithm can be parametrized as $\chi = \Delta c_{\rm m20}/\sigma$ – the minimum sequence colour separation between the two detected systems relative to the width of the filters they were identified in. Small values indicate a good resolution, and in all clusters tested against, we found $\chi < 0.5$. Moreover, for all but two membership bins ($N_{\rm gal} = 14$ and 18), the test cluster was resolved within $\chi < 0.25$. Whilst in our real astronomical data we observe some cluster pairs overlapping in projected space, these examples exhibit large separations in both colour space and redshift. For example, the two clusters MGB J234729−00080.4 and MGB J234733−00100.0 have redshifts of $z = 0.23$ and 0.53, respectively, and $\chi_{r-i} = 7.8$. Although our analysis here could benefit from a larger sample size, ORCA can distinguish between two separate systems even if their sequences lie in the same filter, subject to their colour separation being at least one-fourth the filter width. Below this level, their similarity in colour likely justifies classifying these systems as the same structure.

5 COMPARISON TO EXISTING CLUSTER DATA

The positions of detected clusters can be seen in Fig. 11, with the location of `maxBCG` clusters (Koester et al. 2007a) marked with the red circles, and the positions of known X-ray clusters marked with the blue squares. Clusters detected in the $\{g-r, r-i\}$ combination are shown as the blue filled cells and those detected in $\{r-i, i-z\}$ are shown as the red filled cells. In each case, the cluster BCG cells are yellow.

5.1 The `maxBCG` catalogue

The Koester et al. (2007a) `maxBCG` catalogue of 13 823 optically selected SDSS clusters uses the detection algorithm described in Koester et al. (2007b). This catalogue makes use of data from an earlier release of the SDSS, so was unable to take advantage of the added depth Stripe 82 offered this study. Because direct comparison of the two cluster selection functions is both non-trivial and unfair, we do not attempt a full analysis in this study. However, in the spirit of matching detections made here to those of the shallower data in the Koester et al. (2007a) catalogue, we include the positions of `maxBCG` clusters in Fig. 11 as a set of red circles. The centre of these circles is the location of the assigned BCG, whilst the radius corresponds to $1\,h^{-1}$ Mpc calculated from the published photometric redshift estimate of the cluster. We stress, however, that this does not necessarily correspond to the physical size of the cluster.

The survey area contains 22 `maxBCG` clusters. For ease of reference, salient details from that catalogue are reproduced in Table 4, along with a name of the form BCG JHHMMSS+DDMM.m. We attempt a simple match to the ORCA catalogue by looking for either common BCGs (and more generally a match to ORCA cluster members where BCGs are assigned differently) or statistically significant separations between ORCA centroids and `maxBCG` positions. We find a match to 18 of the 22 clusters; the four `maxBCG` clusters that do not have ORCA analogues are noted in Fig. 11 with the dashed circles and are apparent as two pairs with small angular separation.

We note the ORCA cluster (MGB J234341+00180.3) is situated between the western pair (BCG J234322+00190.6 and BCG J234403+00130.6). Optical-band imaging (Fig. A1 in Appendix A) shows evidence of early-type galaxies distributed in a filamentary chain, approximate comoving length $2\,h^{-1}$ Mpc, sampled by ORCA between the `maxBCG` detections.

The other pair (BCG J234106+00120.4 and BCG J234122+00190.0) may be part of an elongated structure sampled by both the four `maxBCG` entries in that area and also by the ORCA detector. Fig. A2 shows the ORCA cluster MGB J234105+00180.3. This cluster centre, situated between the two `maxBCG` clusters, matches the centroid of a ROSAT All Sky Survey (RASS) cluster to within 0.4 arcmin, with an uncertainty of ∼1 arcmin in the X-ray source.

Overall, we find very good agreement with the `maxBCG` catalogue of clusters, detecting 81 per cent of their entries in the survey region, rising to 100 per cent when taking into account how the different algorithms handle systems that by eye resemble filamentary structure.

5.2 X-ray-selected clusters

X-ray-selected clusters are a useful independent check on the population of clusters detected by optical cluster finders. We use cluster data from the ROSAT All Sky Survey-derived (RASS; Voges et al. 1999), NORAS (Böhringer et al. 2000) and BCS catalogues (for the latter, both main and extended catalogues; Ebeling et al. 1998, 2000), XCS (Romer et al. 2001; Mehrtens et al. 2011) and BLOX (Dietrich et al. 2007) from *XMM-Newton*, and CHaMP (Barkhouse et al. 2006) from *Chandra*. We combine these data sets, taking care to identify any duplicate detections, to form an X-ray catalogue consisting of 1463 unique clusters. From this catalogue, there are 58 X-ray clusters within the full 270-deg^2 footprint covered by Stripe 82, and two of these lie within the 7-deg^2 sample studied here. In future, we will provide a comparison of these X-ray data to an optical cluster catalogue covering a larger area.

The blue squares in Fig. 11 show the positions of the two clusters in the region we study here. The westernmost X-ray cluster, RXC J2337.6+0016 (also detected in the flux-limited brightest cluster sample, Ebeling et al. 1998), is the X-ray counterpart to ACO2631 (Abell et al. 1989) and has a redshift of 0.2780 (Crawford et al. 1995). The X-ray position coincides with the ORCA detection of this system (MGB J233740+00160.2; $z = 0.2571$) at a separation ($\Delta\theta$, Δz) of (0.′1, 0.021). The easternmost X-ray cluster (RXC J2341.1+0018) with a redshift of $z = 0.2766$ (Katgert et al. 1998, misidentified as ACO2644) was originally optically identified in Goto et al. (2002) and Lopes et al. (2004), and is in close proximity to MGB J234105+00180.3 ($z = 0.2588$), with ($\Delta\theta$, Δz) = (0.′4, 0.018). The latter match also appears to straddle two `maxBCG`

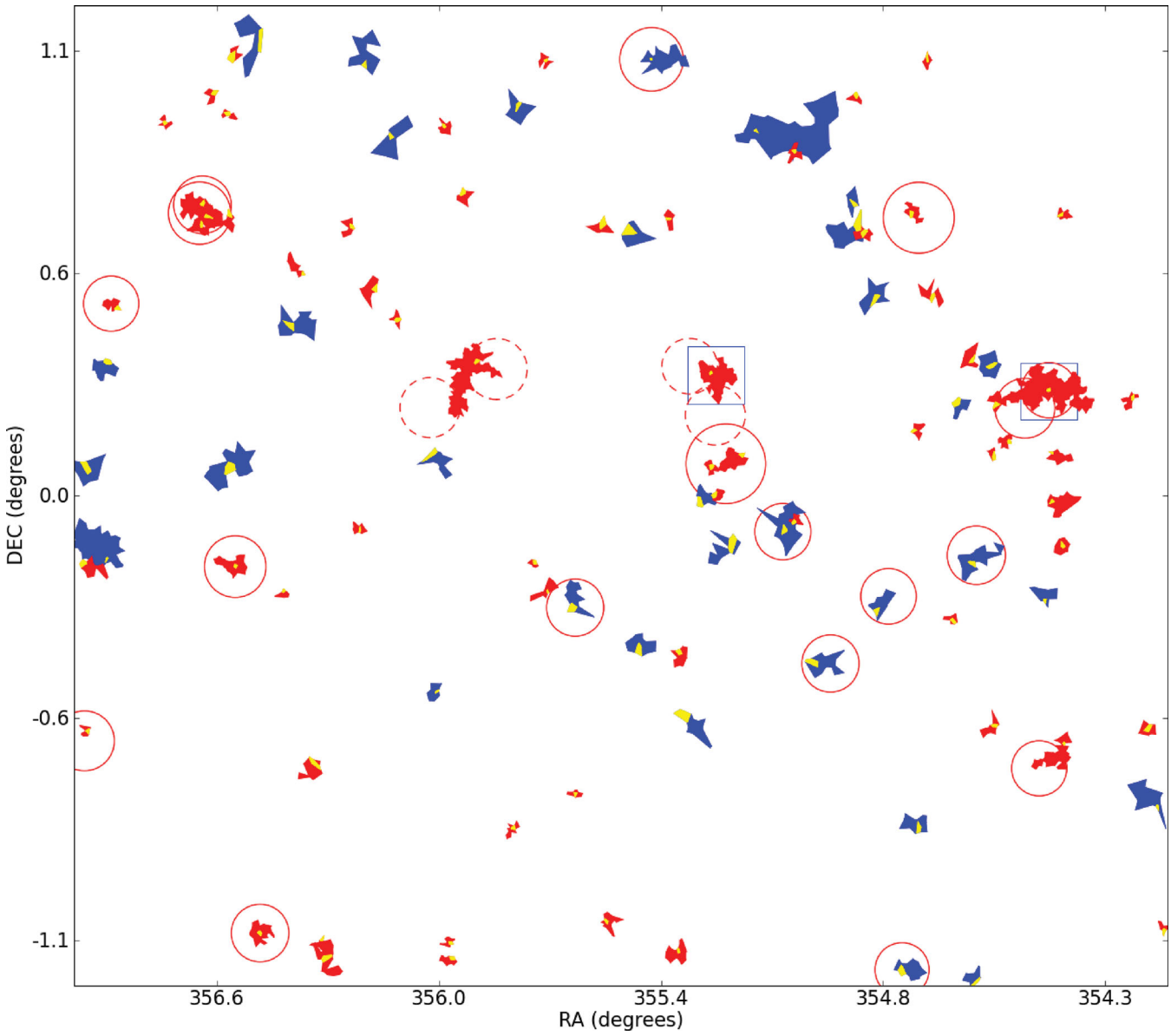

Figure 11. Clusters detected in the Stripe 82 field. The coloured cells represent clusters detected in different colour pairs. The blue cells correspond to clusters detected in $\{g-r, r-i\}$ filter pairs and the red cells correspond to clusters detected in $\{r-i, i-z\}$ filter pairs. The yellow cells indicate the BCG position of each cluster. The red circles indicate the positions of `maxBCG` clusters, based on data shallower than that used in this study. Circle radii correspond to $1\,h^{-1}$ Mpc, based on the `maxBCG` photometric redshift estimate of the cluster. The dashed red circles indicate the four `maxBCG` clusters discussed in Section 5.1 that also feature *gri*-colour imaging in Figs A1 and A2. The blue squares note the position of RASS X-ray sources, with half-lengths corresponding to $1\,h^{-1}$ Mpc.

clusters in the same region as the potentially elongated structure discussed in Section 5.1.

6 PS1 MOCK CLUSTER CATALOGUE

6.1 Simulations

In this section, we describe the application of ORCA to a mock PS1 light-cone. Theoretical simulations allow one the luxury of comparing clusters detected by the algorithm (ORCA clusters) to the galaxy membership of dark matter haloes (hereinafter ΛCDM clusters). Simulated galaxies are allocated to dark matter haloes using the Bower et al. (2006) semi-analytic model. This approach makes the assumption a satellite galaxy is stripped of hot gas immediately following accretion on to a large halo. Star formation is halted after the cold gas reservoir is depleted, and the galaxy joins the red sequence. Coupled with active galactic nucleus feedback, this prescription reproduces the observed bimodality in galaxy colours. However, a known flaw, the rate of gas depletion, results in redder than observed satellite galaxies. Recent treatments of ram-pressure stripping (e.g. McCarthy et al. 2008) hope to improve the understanding of the transition to early-type galaxies with improved semi-analytic models (Font et al. 2008; Benson & Bower 2010).

Although mock surveys are inaccurate realizations of the Universe (see Hilbert & White 2010, for an example in a cluster detection context), they can nevertheless serve as self-consistent tests of

Table 4. An extract from the Koester et al. (2007a) catalogue noting the 22 `maxBCG` clusters within the limits of this SDSS sample field. The cluster name follows the IAU JHHMMSS+DDMM.m format. The `ra` and `dec` are J2000 and measured in degrees. $z_{\rm photo}$ and $z_{\rm spec}$ are the estimated photometric and spectroscopic redshifts of the clusters, respectively. $N_{\rm gal}$ is the number of members in the cluster and $N_{\rm gal}^{\rm R200}$ is the scaled richness.

Cluster name	`ra`	`dec`	$z_{\rm photo}$	$z_{\rm spec}$	$N_{\rm gal}$	$N_{\rm gal}^{\rm R200}$
BCG J233740+00160.3	354.41553	0.27138	0.286	0.277	59	88
BCG J234624+00440.0	356.59955	0.74943	0.273	0.275	25	26
BCG J233746-00420.2	354.44067	−0.70310	0.286	0.287	20	17
BCG J234100+00040.9	355.24905	0.08161	0.194	0.185	23	23
BCG J233955−00250.0	354.97916	−0.43282	0.275	0.277	17	15
BCG J234548−01070.7	356.45068	−1.12775	0.273	–	18	18
BCG J234604−00100.0	356.51477	−0.18283	0.254	–	22	22
BCG J234322+00190.6	355.84039	0.32587	0.257	0.267	38	60
BCG J234146+01070.5	355.44077	1.12444	0.246	0.251	15	11
BCG J233919−00150.6	354.82941	−0.25941	0.284	–	14	11
BCG J234024−00050.6	355.10205	−0.09300	0.281	–	17	13
BCG J234720+00290.7	356.83487	0.49456	0.286	0.275	12	10
BCG J233900+00420.0	354.75143	0.71610	0.219	0.183	14	11
BCG J234122+00190.0	355.34253	0.33330	0.284	0.278	22	22
BCG J233911−01130.3	354.79459	−1.22236	0.292	–	14	10
BCG J234626+00430.7	356.60690	0.72794	0.251	–	25	29
BCG J234403+00130.6	356.01273	0.22646	0.262	–	16	11
BCG J234233−00170.3	355.63776	−0.28873	0.275	–	16	14
BCG J233755+00130.5	354.47760	0.22478	0.262	0.278	37	61
BCG J233825−00090.2	354.60291	−0.15397	0.270	–	14	11
BCG J234737−00370.9	356.90375	−0.63221	0.262	–	14	11
BCG J234106+00120.4	355.27640	0.20707	0.262	–	15	10

the detector. We emphasize, however, there is little merit in comparing mock cluster detections with those in survey data until models can reproduce the observed group and cluster galaxy population with more fidelity.

To compare ORCA detections to the model, we construct ΛCDM clusters with the aid of halo memberships and full 3D galaxy data. In each ΛCDM cluster, we calculate the approximate centre from cluster member positions. Outlier galaxies are identified by rejecting 3σ deviations from a bootstrap-estimated median galaxy–centroid distance. Following outlier ejection, we find the resultant cluster sizes agree well with the virial radii of the host haloes. We set a minimum cluster mass limit by selecting ΛCDM clusters residing in haloes with $M_{\rm H} \geq 10^{13}\,h^{-1}\,{\rm M}_{\odot}$.

6.2 Mock reference cluster

We select a 'reference cluster' from a set of ΛCDM-based detections generated from a preliminary scan of the simulation. The chosen cluster allows us to set the slope and width of the photometric filters in our search through the mock data. Candidate training clusters were identified from a redshift range bracketing Abell 2631 ($z = 0.278$), with similar memberships and a clear sequence in all colours. We selected the richest of these candidates, featuring 130 members and a redshift of $z = 0.3$. By applying the same fitting techniques as those described in Section 3.7.1, we set the filter parameters listed in Table 5 and apply the same colour ranges as those used on the SDSS. The fitted gradients are steeper in $g - r$ and $r - i$ than those used for the SDSS, and the filter widths are smaller. These values were nevertheless consistent with the other candidate reference clusters identified in the mock. As before, we use the most conservative width ($g - r$, 0.13) for filters in each colour.

Table 5. Filter parameters fitted from the mock reference cluster (by analogy with those derived from Abell 2631) along with colour ranges searched by the detector (the same as those used in the Stripe 82 data).

Colour	Slope (β)	Width ($\sigma_{\rm f}$)	Range	Filters
$g - r$	−0.070	**0.130**	0.47−2.00	39
$r - i$	−0.032	0.064	0.00−1.22	38
$i - z$	−0.012	0.035	−0.10−1.10	31

6.3 Producing ΛCDM and mock ORCA cluster catalogues

Except for the revised parameters listed in Table 5, the detector ran as described in Section 3, and applied magnitude limits created a source catalogue of 80 536 mock galaxies. Because the algorithm relies on the detection of colour–magnitude ridgelines, we do not want to include ΛCDM clusters without detectable sequences. We therefore construct the ΛCDM cluster list from galaxies selected in the same photometric filters as used by the detector, meaning ΛCDM clusters may also be detected multiple times. We group together ΛCDM clusters with common halo identifiers, but as before, selected the highest *reduced flux* candidate as the 'best' ΛCDM cluster.

We found a total of 305 ORCA clusters with $M_{\rm H} \geq 10^{13}\,h^{-1}\,{\rm M}_{\odot}$; at $M_{\rm H} \geq 10^{14}\,h^{-1}\,{\rm M}_{\odot}$, the counts are more equal. Although the majority of clusters identified are at $z \sim 0.3$, the tests we describe in the following section will highlight how well ORCA performs over this entire parameter space. Fig. 12 shows a simple comparison of the two catalogues by plotting both sets of clusters residing in haloes $M_{\rm H} \geq 10^{13.5}\,h^{-1}\,{\rm M}_{\odot}$ out to $z = 0.6$ (the highest cluster redshift in the SDSS cluster catalogue). The grey circle centres denote the position,

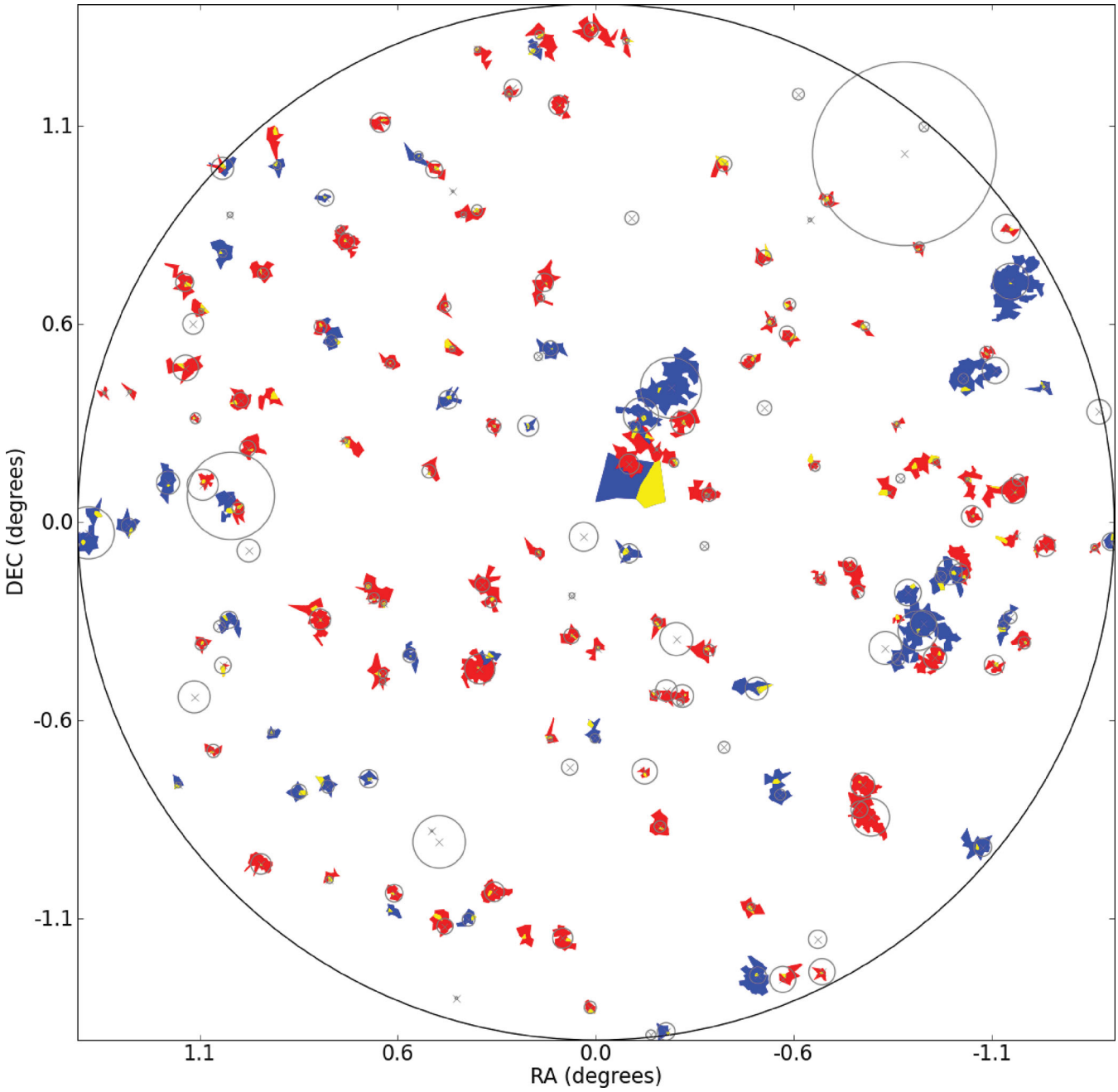

Figure 12. Clusters in haloes of mass $\geq 10^{13.5}\,h^{-1}\,\mathrm{M}_{\odot}$ from the mock ORCA cluster catalogue (cells) and the ΛCDM catalogue (circles). The cell colours correspond to clusters detected in different colour pairs. The blue cells are clusters detected in the $\{g-r, r-i\}$ filter pairs and the red cells are clusters detected in $\{r-i, i-z\}$. The yellow cells indicate the BCG of each cluster. The crosses denote the ΛCDM cluster centre, and circle radii indicating the angular distance between the centre and most distant member.

and their radii the maximum member–cluster centre distance of ΛCDM clusters. The blue and red cells represent ORCA clusters detected in $\{g-r, r-i\}$ and $\{r-i, i-z\}$, respectively.

6.4 Performance of the algorithm

To determine how well the detector recovers and characterizes the mock clusters, we illustrate here three simple tests to quantify the detection performance.

6.4.1 *Completeness*

We define completeness as the number of detected haloes as a function of halo mass and redshift. A halo is detected if at least N_{min} galaxies are identified, even if they are shared between multiple ORCA clusters (e.g. fragmenting a halo when the algorithm attempts to identify substructure). We compare this number to ΛCDM cluster counts (by definition unfragmented), with at least N_{min} members.

The fraction of detected ΛCDM clusters can be seen in Fig. 13, where we produce a grid of cells with sampling intervals of 0.05 in redshift and 0.2 in $\log_{10}$ halo mass. Because in some cases only a

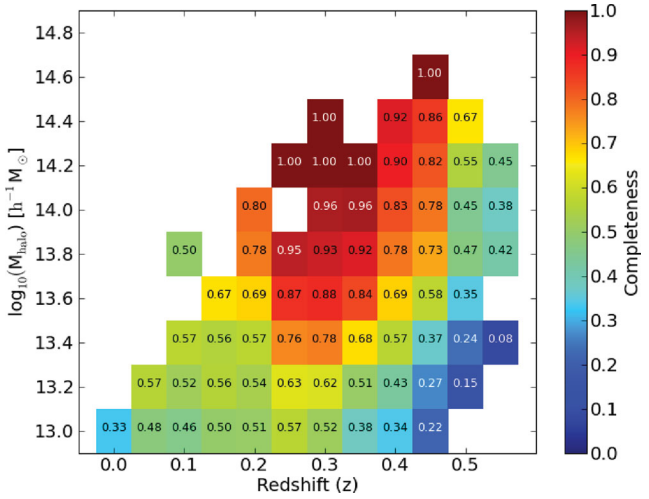

Figure 13. Completeness of mock ΛCDM clusters. The fraction of correctly detected clusters from the ORCA catalogue as a function of halo mass and redshift. The white regions indicate where there were no ΛCDM clusters in that bin.

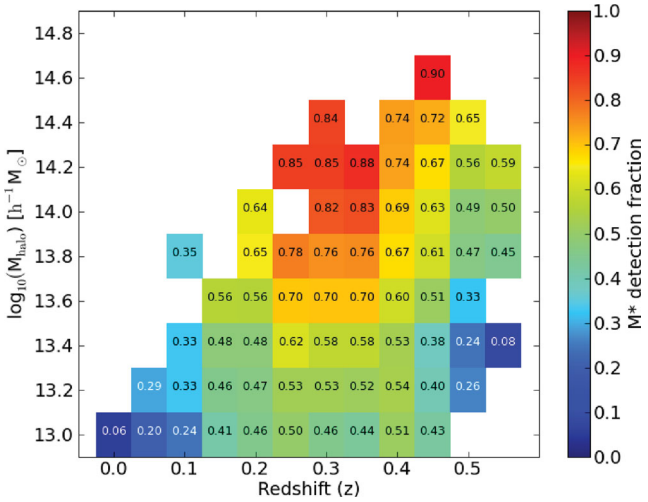

Figure 14. Stellar mass accuracy. The fraction of recovered stellar mass in mock clusters as a function of halo mass and redshift.

few detections occupy each cell, some regions will suffer from shot noise. We smooth the data using a 3 × 3 grid so the completeness for a given cell is the mean completeness over this region. The empty regions in Fig. 13 therefore indicate where either no ΛCDM clusters exist or too few clusters are found to reliably calculate the completeness (we set a threshold of at least five clusters detected over the 3 × 3 grid). Between $0.1 \leq z \leq 0.4$, the detector attains at least 68 per cent completeness for halo masses above $10^{13.6}\,h^{-1}\,\mathrm{M}_{\odot}$, and is over 90 per cent complete in halo masses exceeding $10^{14.3}\,h^{-1}\,\mathrm{M}_{\odot}$. This compares favourably with the `maxBCG` algorithm applied to mock simulations, where Koester et al. (2007b) report >90 per cent completeness between $0.1 \leq z \leq 0.3$ for $M_{\mathrm{H}} \geq 10^{14.3}\,h^{-1}\,\mathrm{M}_{\odot}$ with clusters containing at least 10 members (cf. $N_{\mathrm{min}} = 5$ in this study). Applying the completeness definition and the same selection criteria as that study, the ORCA detector is >90 per cent complete down to a halo mass of $10^{13.8}\,h^{-1}\,\mathrm{M}_{\odot}$. These results also compare well to the VT completeness of the `2TecX` (van Breukelen & Clewley 2009) algorithm, either matching or exceeding their stated completeness for $M_{\mathrm{H}} = 10^{13.7}$ and $10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$ up to our redshift limit.

At higher redshifts, there is a decline in completeness where there are only a few members brighter than the magnitude limit, reducing the algorithm sensitivity to distant clusters. This effect is more apparent among the lower mass haloes. At high redshift ($z > 0.4$) and low mass ($M_{\mathrm{H}} \leq 10^{13.3}\,h^{-1}\,\mathrm{M}_{\odot}$), there are 12 ΛCDM clusters, but the detector identifies only two of these. We also note a local incompleteness at $z \leq 0.08$. Arising from our choice of probability threshold (P_{thresh}), too few overdense cells are selected in filters featuring low signal-to-noise ratio clusters. The photometric filters best suited to detecting local, relatively blue clusters have galaxy populations dominated by the blue cloud component of the CMR. Successful detections in this crowded field are compounded by the larger scale-size of more local clusters such as the local ($z = 0.03$) seven-member group at the north-western boundary of the catalogue in Fig. 12. Under these circumstances, it becomes unlikely cluster Voronoi cells share common vertices, restricting potential membership links between them.

We classify spurious detections in the mock cluster catalogue as those clusters where each member belongs to a different halo. Of the 305 ORCA clusters, only two fit this description, suggesting a spurious detection rate (0.7 per cent) consistent with tests performed in Section 4.6.3.

6.4.2 *Stellar mass accuracy*

Stellar mass accuracy is the stellar mass of an ORCA cluster relative to that of the ΛCDM cluster belonging to the same halo. Because the algorithm may split the halo galaxies into multiple clusters, we combine the mass of all ORCA clusters sharing the same halo. In ΛCDM clusters with up to ∼12 members (approximately 75 per cent of the catalogue), over half of the total cluster stellar mass comes from the two most massive galaxies. Efficient detection of these galaxies is therefore essential in gaining accurate estimates of cluster stellar masses. The stellar mass accuracy for each ΛCDM cluster is $A_* = M_*^{\mathrm{cl}}/M_*^{\mathrm{true}}$, where M_*^{cl} is the stellar mass of all ORCA cluster members registered to the ΛCDM cluster's halo. We apply the same gridding technique as discussed in the previous section, requiring at least five clusters in a grid to define a reliable A_*. As Fig. 14 shows, between $0.1 \leq z \leq 0.4$, the algorithm recovers over half of the cluster stellar mass for systems with halo masses of at least $10^{13.4}\,h^{-1}\,\mathrm{M}_{\odot}$. This recovery fraction improves with increasing mass, reaching 90 per cent in some cases. Both local and distant clusters suffer from lower stellar mass estimates. For the former, higher levels of halo fragmentation (one halo being assigned to many ORCA clusters) result in galaxies lost to nearby systems with densities or memberships too low to qualify as clusters. Those systems with redshifts $z > 0.5$ tend to be unfragmented but contain fewer members, causing an underestimation of cluster stellar mass. The stellar mass accuracy at the median redshift of the survey ($z = 0.33$) remains above 50 per cent down to halo masses of $10^{13.2}\,h^{-1}\,\mathrm{M}_{\odot}$, and above 75 per cent from masses of $10^{13.8}\,h^{-1}\,\mathrm{M}_{\odot}$, suggesting the detector performs well in estimating the true cluster stellar mass content.

6.4.3 *Purity*

As discussed in Section 6.4.1, a halo is detected by the algorithm if it finds at least N_{min} members that have been allocated to ORCA clusters. For a cluster with seven members, the distinction between a cluster containing five halo galaxies and two interlopers, and one containing seven halo galaxies provides a measure of cluster purity.

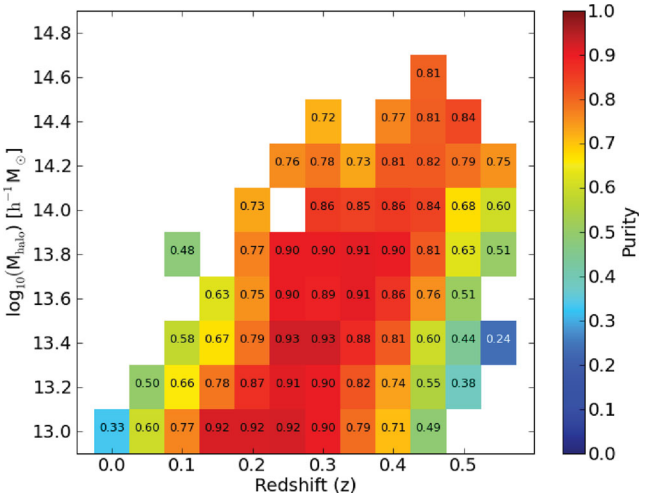

Figure 15. The purity of ΛCDM clusters detected by the ORCA algorithm. Low values indicate where clusters have included a large number of contaminating galaxies not belonging to the halo.

We define purity as the fraction of galaxies ORCA assigned to the cluster that are members additionally belonging to the host halo. This description is in line with the purity described by Koester et al. (2007b). However, we decide not to adopt a threshold above which a cluster is considered pure, instead directly assigning each cluster a purity fraction. Fig. 15 shows the purity of ORCA clusters with varying redshift and halo mass, the gridding method here being the same scheme as introduced in Section 6.4.1. ORCA clusters are at least 70 per cent pure at the median redshift of the survey over all halo masses. The purity appears to drop at higher redshifts, attributed to faint but genuine cluster members being replaced by brighter contaminants that lie on the cluster sequence. Relative to the completeness and stellar mass estimates, cluster purity is not as sensitive to halo mass. This is most likely a consequence of the membership incompleteness discussed in Section 3.7.2. Because peripheral members are less likely to be in Voronoi cells tagged as statistically significant, the inclusion of interlopers at cluster edges is reduced. As in the previous section, increased halo fragmentation drives the local drop in purity, serving to increase the contamination fraction by distributing the halo galaxies among local clusters and systems failing to achieve cluster status.

7 SUMMARY

We present and demonstrate a new cluster detection algorithm based on red-sequence cluster searches, the detection of overdensities using VTs, and connecting galaxies into clusters with a Friends-of-Friends algorithm. With this approach, we make only two assumptions about the systems we are looking for – they have detectable red sequences and are overdensities in the projected plane of the sky.

We calibrate the photometric selection filters to a rich Abell cluster found in SDSS data, and find that recovery of members from both this large cluster and a small group is largely insensitive to the choice of two algorithm parameters controlling the behaviour of the algorithm. When applying the algorithm to a sample of SDSS Stripe 82 galaxies with four bands, we find 97 clusters. Based on spectroscopic and photometric redshifts, we estimate these clusters are detected out to $z = 0.6$ and the catalogue has a median redshift of $z = 0.31$. We perform false-positive tests suggesting the spurious detection frequency is below 1 per cent. Tests on the catalogue suggest the detector is robust to sparsely sampled cluster fields and is not overly sensitive to survey edges. In comparing our data to existing optical and X-ray clusters, we find good agreement with the `maxBCG` and RASS catalogues in the same region.

We go on to test the performance of the detector with a mock survey generated from a semi-analytic galaxy formation model. In comparing the ORCA cluster detections with those generated from halo membership data, we make a quantitative assessment of the detector performance. The algorithm identifies 305 clusters, whilst the simulation produces 414 down to a halo mass of $10^{13}\,h^{-1}\,\mathrm{M}_{\odot}$. At the median redshift of the catalogues (both $z = 0.33$), we find ORCA is 75 per cent complete down to a cluster halo mass of $10^{13.4}\,h^{-1}\,\mathrm{M}_{\odot}$ and is able to recover approximately 75 per cent of the total stellar mass for clusters in haloes of at least $10^{13.8}\,h^{-1}\,\mathrm{M}_{\odot}$.

We have demonstrated this algorithm is capable of identifying clusters in both real and simulated data with minimal assumptions as to the nature of clusters. In combining comprehensive colour scans to search for cluster red sequences with Voronoi diagrams to estimate surface densities, we avoid making model-dependent decisions about what a cluster is. Cluster redshifts arise as a consequence, not condition, of our detection, affording additional freedom from model SEDs and the uncertainties inherent in photometric redshift data spanning the depths, fluxes and areas set to be commonplace in next-generation galaxy catalogues. This detector can be used in any survey where there are at least two photometric bands, but is most powerful when applied to multicolour surveys such as the forthcoming Pan-STARRS surveys. The scope for cluster detection with ORCA is not limited solely to the optical regime. Preliminary tests with optical–IR band-merged catalogues show great promise, requiring minimal adaptation to facilitate the detection of the 4000-Å break into the IR bands and beyond $z = 1$.

ACKNOWLEDGMENTS

We thank the referee for useful comments which improved the clarity of this paper. DNAM acknowledges an STFC PhD studentship, JEG and RGB thank the UK Science and Technology Facilities Council and the Natural Sciences and Engineering Research Council of Canada for financial support. The authors thank Alastair Edge, John Lucey, Kathy Romer, Ian Smail and John Stott for useful discussions, and Carlton Baugh, Yan-Chuan Cai and Shaun Cole for access to the mock MDS light-cone data.

Calculations in portions of this work were performed on the ICC Cosmology Machine, which is part of the DiRAC Facility jointly funded by the STFC, Large Facilities Capital Fund of BIS, and Durham University.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, the Astrophysical Institute Potsdam, the University of Basel, the University of Cambridge, the Case Western Reserve University, the University of Chicago, the Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the

Chinese Academy of Sciences (LAMOST), the Los Alamos National Laboratory, the Max Planck Institute for Astronomy (MPIA), the Max Planck Institute for Astrophysics (MPA), the New Mexico State University, the Ohio State University, the University of Pittsburgh, the University of Portsmouth, the Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abell G. O., 1958, ApJS, 3, 211
Abell G. O., Corwin H. G., Jr, Olowin R. P., 1989, ApJS, 70, 1
Allen S. W., Schmidt R. W., Fabian A. C., 2002, MNRAS, 334, L11
Balogh M. L., Navarro J. F., Morris S. L., 2000, ApJ, 540, 113
Barber C. B., Dobkin D. P., Huhdanpaa H., 1996, ACM Trans. Math. Softw., 22, 469
Barkhouse W. A. et al., 2006, ApJ, 645, 955
Benson A. J., Bower R., 2010, MNRAS, 405, 1573
Blanton M. R. et al., 2003, ApJ, 592, 819
Böhringer H. et al., 2000, ApJS, 129, 435
Bolzonella M., Miralles J., Pelló R., 2000, A&A, 363, 476
Bower R. G., Lucey J. R., Ellis R. S., 1992, MNRAS, 254, 601
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Brodwin M. et al., 2010, ApJ, 721, 90
Cai Y., Angulo R. E., Baugh C. M., Cole S., Frenk C. S., Jenkins A., 2009, MNRAS, 395, 1185
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cannon R. et al., 2006, MNRAS, 372, 425
Carlberg R. G., Yee H. K. C., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C. J., 1996, ApJ, 462, 32
Carlstrom J. E., Holder G. P., Reese E. D., 2002, ARA&A, 40, 643
Cavaliere A., Fusco-Femiano R., 1976, A&A, 49, 137
Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, ApJ, 648, L109
Colberg J. M., Krughoff K. S., Connolly A. J., 2005, MNRAS, 359, 272
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Collister A. A., Lahav O., 2004, PASP, 116, 345
Crawford C. S., Edge A. C., Fabian A. C., Allen S. W., Bohringer H., Ebeling H., McMahon R. G., Voges W., 1995, MNRAS, 274, 75
Csabai I. et al., 2003, AJ, 125, 580
Dalton G. B., Efstathiou G., Maddox S. J., Sutherland W. J., 1992, ApJ, 390, L1
Dietrich J. P., Erben T., Lamer G., Schneider P., Schwope A., Hartlap J., Maturi M., 2007, A&A, 470, 821
Doherty M. et al., 2010, A&A, 509, A83
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Ebeling H., Wiedenmann G., 1993, Phys. Rev. E, 47, 704
Ebeling H., Edge A. C., Bohringer H., Allen S. W., Crawford C. S., Fabian A. C., Voges W., Huchra J. P., 1998, MNRAS, 301, 881
Ebeling H., Edge A. C., Allen S. W., Crawford C. S., Fabian A. C., Huchra J. P., 2000, MNRAS, 318, 333
Eke V. R. et al., 2004, MNRAS, 348, 866
El-Ad H., Piran T., da Costa L. N., 1996, ApJ, 462, L13
Evrard A. E., 1997, MNRAS, 292, 289
Fedeli C., Moscardini L., Matarrese S., 2009, MNRAS, 397, 1125
Font A. S. et al., 2008, MNRAS, 389, 1619
Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1990, ApJ, 351, 10
Geach J. E., Murphy D. N. A., Bower R. G., 2011, MNRAS, 413, 3059 (GMB11)
Gladders M. D., Yee H. K. C., 2000, AJ, 120, 2148
Gladders M. D., Yee H. K. C., 2005, ApJS, 157, 1
Gladders M. D., Lopez-Cruz O., Yee H. K. C., Kodama T., 1998, ApJ, 501, 571
Goto T. et al., 2002, AJ, 123, 1807
Gunn J. E. et al., 1998, AJ, 116, 3040
Hao J., Koester B. P., Mckay T. A., Rykoff E. S., Rozo E., Evrard A., Annis J., Becker M., 2009, ApJ, 702, 745
Hayashi M., Kodama T., Koyama Y., Tadaki K. I., Tanaka I., 2011, MNRAS, 415, 2670
Hilbert S., White S. D. M., 2010, MNRAS, 404, 486
Hincks A. D. et al., 2010, ApJS, 191, 423
Icke V., van de Weygaert R., 1991, QJRAS, 32, 85
Ivezic Z. et al. (LSST Collaboration), 2008, preprint (arXiv:0805.2366)
Kaiser N. et al., 2002, in Tyson J. A., Wolff S., eds, SPIE Conf. Ser. Vol. 4836, Survey and Other Telescope Technologies and Discoveries. SPIE, Bellingham, p. 154
Katgert P., Mazure A., den Hartog R., Adami C., Biviano A., Perea J., 1998, A&AS, 129, 399
Kiang T., 1966, Z. Astrophys., 64, 433
Kim R. S. J. et al., 2002, AJ, 123, 20
Kodama T., Tanaka I., Kajisawa M., Kurk J., Venemans B., De Breuck C., Vernet J., Lidman C., 2007, MNRAS, 377, 1717
Koester B. P. et al., 2007a, ApJ, 660, 239
Koester B. P. et al., 2007b, ApJ, 660, 221
Lidman C. E., Peterson B. A., 1996, AJ, 112, 2454
Lin H., Yee H. K. C., Carlberg R. G., Morris S. L., Sawicki M., Patton D. R., Wirth G., Shepherd C. W., 1999, ApJ, 518, 533
Liu H. B., Hsieh B. C., Ho P. T. P., Lin L., Yan R., 2008, ApJ, 681, 1046
Lopes P. A. A., de Carvalho R. R., Gal R. R., Djorgovski S. G., Odewahn S. C., Mahabal A. A., Brunner R. J., 2004, AJ, 128, 1017
McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
McInnes R. N., Menanteau F., Heavens A. F., Hughes J. P., Jimenez R., Massey R., Simon P., Taylor A., 2009, MNRAS, 399, L84
Mehrtens N. et al., 2011, preprint (arxiv e-prints)
Murphy D. N. A., Eke V. R., Frenk C. S., 2011, MNRAS, 413, 2288
Papovich C. et al., 2010, ApJ, 716, 1503
Pimbblet K. A., Drinkwater M. J., 2004, MNRAS, 347, 137
Plionis M., Barrow J. D., Frenk C. S., 1991, MNRAS, 249, 662
Postman M., Lubin L. M., Gunn J. E., Oke J. B., Hoessel J. G., Schneider D. P., Christensen J. A., 1996, AJ, 111, 615
Ramella M., Boschin W., Fadda D., Nonino M., 2001, A&A, 368, 776
Rines K., Geller M. J., 2008, AJ, 135, 1837
Romer A. K., Viana P. T. P., Liddle A. R., Mann R. G., 2001, ApJ, 547, 594
Ryden B. S., 1995, ApJ, 452, 25
Schechter P., 1976, ApJ, 203, 297
Schirmer M., Suyu S., Schrabback T., Hildebrandt H., Erben T., Halkola A., 2010, A&A, 514, A60
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Schuecker P., Caldwell R. R., Böhringer H., Collins C. A., Guzzo L., Weinberg N. N., 2003, A&A, 402, 53
Sheldon E. S. et al., 2004, AJ, 127, 2544
Smail I. et al., 2007, ApJ, 654, L33
Smith R. J., Lucey J. R., Hudson M. J., Allanson S. P., Bridges T. J., Hornschemeier A. E., Marzke R. O., Miller N. A., 2009, MNRAS, 392, 1265
Soares-Santos M. et al., 2011, ApJ, 727, 45
Springel V. et al., 2005, Nat, 435, 629
Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461
Stott J. P., Pimbblet K. A., Edge A. C., Smith G. P., Wardlow J. L., 2009, MNRAS, 394, 2098
Stubbs C. W., Doherty P., Cramer C., Narayan G., Brown Y. J., Lykke K. R., Woodward J. T., Tonry J. L., 2010, ApJS, 191, 376
Sunyaev R. A., Zel'dovich I. B., 1980, ARA&A, 18, 537
Swinbank A. M. et al., 2010, Nat, 464, 733
Thanjavur K., Willis J., Crampton D., 2009, ApJ, 706, 571
van Breukelen C., Clewley L., 2009, MNRAS, 395, 1845
Vanderlinde K. et al., 2010, ApJ, 722, 1180
Voges W. et al., 1999, A&A, 349, 389

Voit G. M., 2005, Rev. Mod. Phys., 77, 207
White S. D. M., Efstathiou G., Frenk C. S., 1993, MNRAS, 262, 1023
Wilson G. et al., 2009, ApJ, 698, 1943
Yee H. K. C., López-Cruz O., 1999, AJ, 117, 1985
York D. G. et al., 2000, AJ, 120, 1579
Zwicky F., Herzog E., Wild P., 1961, Catalogue of Galaxies and of Clusters of Galaxies, Vol. I. California Institute of Technology (CIT), Pasadena

APPENDIX A: CLUSTER IMAGES

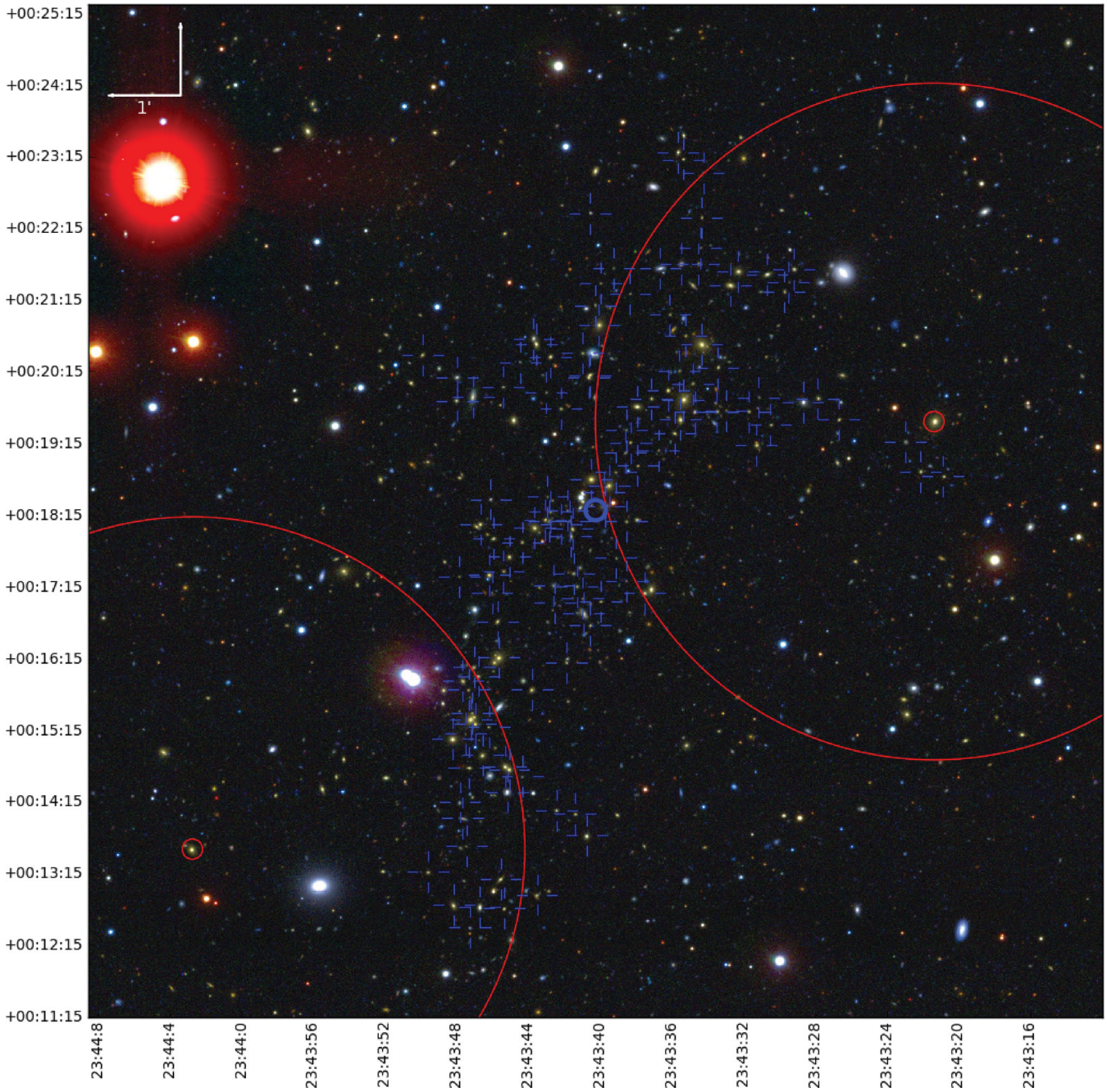

Figure A1. Stripe 82 cluster MGB J234341+00180.3 is an extended system detected between two `maxBCG` clusters (BCG J234322+00190.6 and BCG J234403+00130.6). For clarity, we have not plotted the Voronoi grid, but the cluster members are marked with blue cross-hairs. The `maxBCG` clusters are shown in red, with the central positions noted by the two smaller circles, and the larger circles corresponding to radii of $1\,h^{-1}$ Mpc based on the photometrically estimated cluster redshift from Koester et al. (2007a).

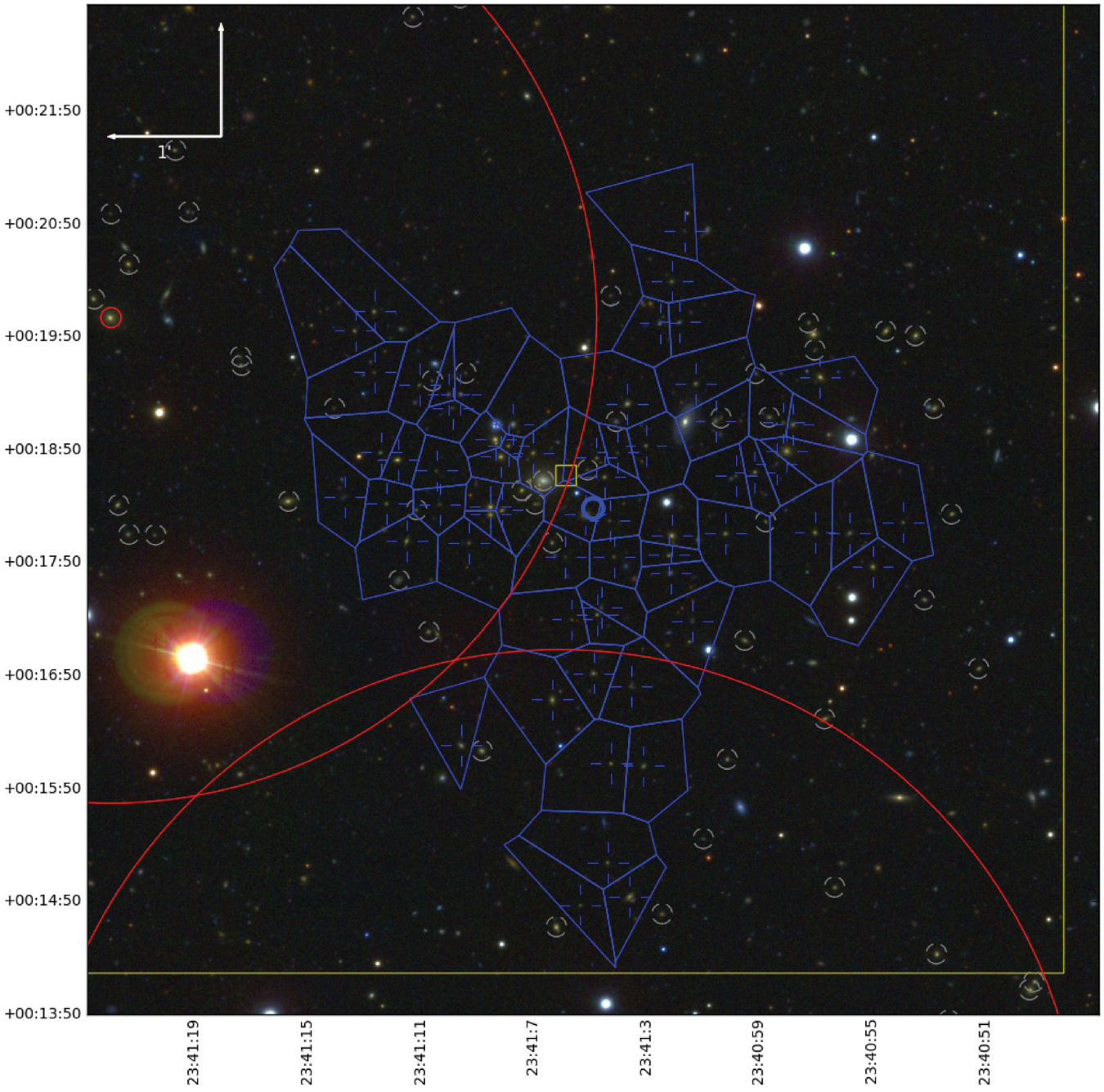

Figure A2. Stripe 82 cluster MGB J234105+00180.3: an ORCA detection between two maxBCG clusters and on top of an X-ray cluster position. Members and their Voronoi cells are marked in blue, the thick circle indicating the estimated cluster centre. The grey dashed circles are *associate cluster members* arising from multiple detections of this cluster (Section 3.6). The red data indicate the location of the `maxBCG` clusters BCG J234122+00190.0 and BCG J234106+00120.4, with the larger circles indicating a $1\,h^{-1}$ Mpc radius and the smaller circles indicating the BCG positions. The yellow data indicate the NORAS X-ray cluster RXC J2341.1+0018; the half-length of the large square corresponds to $1\,h^{-1}$ Mpc based on the cluster redshift, the small square noting the X-ray position, uncertain to approximately 1 arcmin. The X-ray-ORCA centroid separation is approximately 0.4 arcmin.

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