
SIDE A
TEN BILLION YEARS OF MASSIVE GALAXIES
EDWARD N. TAYLOR
I. Introduction and Summary
II. The MUSYC NIR-Selected Catalog of the ECDFS
III. The Rise of Red Galaxies
IV. On the Dearth of Compact Galaxies in the Local Universe
V. On the Masses of Galaxies in the Local Universe
VI. Nederlandse Samenvattig
recorded in 2005-2009
Upstairs Studio, Fitzroy
FZY LP 1001
STEREO - 33 1/3

# 10 Billion Years of Massive Galaxies

# 10 Billion Years of Massive Galaxies

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. dr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 15 december 2009
te klokke 17.15 uur

door

Edward Nairne Cunningham Taylor

geboren te Baltimore (VS)
in 1980

Promotiecommissie

| | |
|---|---|
| Promotores: | Prof. dr. M. Franx<br>Prof. dr. P. G. van Dokkum (Yale University) |
| Overige leden: | Prof. dr. E. F. Bell (University of Michigan)<br>Prof. dr. K. H. Kuijken<br>Prof. dr. G. K. Miley<br>Prof. dr. H. Röttgering<br>Prof. dr. S. C. Trager (Rijksuniversiteit Groningen) |

1. The quenching of star formation within the massive galaxy population (as opposed to individual galaxies) is a long and gradual process and is not a simple function of stellar mass (Chapter III).

2. The massive, 'passive' galaxies seen at $z \gtrsim 2$ have not finished their evolution: they must undergo significant structural evolution, without substantially increasing their stellar mass. Major mergers cannot be the sole driver of this evolution (Chapter IV).

3. Stellar mass-to-light ratios derived on the basis of a single rest-frame optical color are just as good — and in some circumstances better — than those derived from a full SED (Chapters III and V).

4. In photometric lookback studies, stellar mass is a more robustly measured quantity than absolute magnitude (Chapter III).

5. Galaxies are not homologous and should not be treated as such (Chapter V).

6. More than better data, future photometric lookback surveys will require better analysis to improve on current measurements of the galaxy stellar mass function and its evolution.

7. By providing a large sample of massive galaxies at $0.1 \lesssim z \lesssim 0.4$, the Galaxy And Mass Assembly (GAMA) survey will not only yield much stronger constraints on the recent evolution of the general galaxy population, it will also significantly reduce the uncertainties on past and future studies that focus on higher redshifts.

8. It is remarkable that the vast majority of professional astronomers have had little to no formal training in the four key aspects of their day-to-day work: programming, writing, teaching, and management. Providing this kind of training offers an easy, practical way of increasing the productivity of both individual researchers and scientific collaborations.

9. Language encourages us to ignore one of the most profound conclusions arrived at via science or philosophy — that of a fundamental unity of all things.

10. As an economically and physically viable means of mitigating the worst effects of climate change, geoengineering (including $CO_2$ removal and solar radiation management) deserves serious study and debate.

**for my mum and dad — i think you're cool, too.**

Here's to Cisco and Sonny and Leadbelly, too,
and all the good people that travelled with you;
here's to the hearts and the hands of the men
that come with the dust, and are gone with the wind.

— Bob Dylan, *Song to Woody* (1962)

# Table of Contents

# List of Figures

# Chapter I

## Introduction and Summary

### 1 Statement of the Problem

Galaxy formation and evolution is the mother of all astronomical problems. It takes as its starting point the description of the basic contents of the universe from cosmology, and seeks ultimately to explain the formation of stars and planets within galaxies, including the creation of the basic elements necessary for life. In this way, it links the two big questions — how did the universe begin; and what are the origins of life?

In its scale and complexity, galaxy formation and evolution is a diabolical problem. A complete, from–first–principles description involves physical processes that occur on scales ranging from many megaparsecs down to hundreds of kilometers, and indeed down to the scales of molecules, nuclei, and electrons. There is broad consensus on a general, qualitative picture of how galaxies form, and what physical processes are likely to be the most important in governing their evolution. Driven by gravitational pressure, minute fluctuations in the post-inflation, dark–plus–baryonic density field grow into a froth of filamentary structures and voids. As overdensities become gravitationally unstable, they collapse and condense to form individual, virialized halos — the cradles for future galaxies. In the course of collapse, thermal pressure separates the baryonic gas from the dark matter. Gas accretion onto and into the dark halo is then regulated by the cooling efficiency of the gas. As the gas density increases, so too does the cooling rate. At a certain point, too cold and dense for thermal pressure to withstand its self-gravity, the gas now begins to fragment and collapse. Once the gas density in the centers of smallest fragments becomes high enough to ignite fusion, a star is born. Somewhere along the way, a supermassive black hole develops in the center of each halo. Both the stars and the black hole then drive kinetic and/or energetic 'feedback' processes, which can incite, disrupt, or prevent further gas accretion and star formation. If the feedback is strong enough, galaxies can also affect their surrounding environment, and thus other galaxies. Further, as cosmic structure formation continues around them, galaxies can also grow through successive mergers. Tidal interactions between galaxies can also be important in triggering star formation, or in stripping gas out of galaxies. But virtually none of these processes are well understood individually, much less how they relate to one another.

The most successful approach to the problem has been through the use of 'semi-analytic' models (see Baugh, 2006, for an introductory overview; it should also be noted that with continued advances in computational power, cosmological hydrodynamical simulations are becoming practicable; see, e.g., Naab et al., 2006, Schaye et al., 2009). The basic idea behind semi-analytic modeling is to take the results of large-volume $N$-body simulations of structure formation within the framework of $\Lambda$CDM cosmology (see the review of Springel, Frenk & White, 2006), and then combine them with empirical, analytic 'laws' that describe the halo- and galaxy-scale physical processes that influence the baryons. In this way, semi-analytic models generate 'predictions' for the global properties of individual galaxies. Where the results differ from the observed universe, the assumptions underpinning the model are then refined to reduce the discrepancy. Given the large uncertainties and many approximations in these models, theorists have no shortage of knobs and levers with which to fine-tune their results.

Pedagogically, the virtue of this approach is that, by identifying precisely how changing the model assumptions affects the outcome, it provides a means of probing the relative importance of different physical processes at different scales or stages of evolution. But these models are not, strictly speaking, predictive. Instead, they focus on *post hoc* modifications to the assumptions underpinning the models in order to obtain consistency. Rather than a complete physical explanation of the process of galaxy evolution, the models thus aim for a consistent description of the evolving properties of the galaxy population (see, e.g., Croton et al., 2006; Bower, McCarthy & Benson, 2008; Somerville et al., 2008). The 'best' models are those that simultaneously reproduce the largest number of qualitatively different aspects of the observed galaxy population with the smallest amount of tuning.

The field of galaxy formation and evolution is thus largely observation driven, and likely to remain so for quite a while. In this context, the goal of this thesis is to provide new observational constraints on the evolution of galaxies with which to challenge these kinds of models. In particular, *this thesis is focused on the evolution of massive galaxies — in terms of their number, star formation activity, and structure — over 10 Gyr of cosmic history.*

## 2 Technical Background — The basic requirements of a modern lookback survey

### Lookback Surveys and Observational Cosmology

Because the speed of light is finite, it takes time for light to travel from one place to another. Thus, when you look over great distances, you see parts of the universe that are actually younger than here and now. Observing galaxies over a range of distances thus provides a kind of time-ordered series of momentary glimpses into the lives of individual galaxies at different points in the history of the universe. *Through comparisons between the statistical properties of the galaxy population across a range of distances, 'lookback' surveys provide a means of directly observing the evolution of galaxies over cosmic time.*

As in most of astronomy, the technical crux of any lookback survey is determining the distances to individual sources. In our expanding universe, the cosmic expansion history is imprinted on every photon: as photons stream through space they are caught up in the expansion and redshifted. The farther and longer they travel, the greater the effect. As a distance indicator, redshifts by themselves provide only a relative distance measurement; the precise relation between redshift and distance (or, equally, lookback time) depends on the cosmic expansion history.

The major advances in observational cosmology made between 1998 and 2003 thus revolutionized the field of lookback survey science. This revolution began in the late 1990s with the first solid evidence for dark energy, which came from supernova 'standard candle' measurements (e.g., Riess et al., 1998; Perlmutter et al., 1999), and culminated with the announcement of the first results from a small microwave satellite called the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al., 2003). In combination with these kinds of supernovae results, the Hubble key project measurement of the Hubble constant (Freedman et al., 2001), and clustering measurements based on nearby galaxies (e.g., Verde et al., 2002) and the Lyman-$\alpha$ forest (e.g., Croft et al., 2002), the WMAP data allowed the determination of the basic cosmological parameters to within $\lesssim 5$ % (Spergel et al. 2003; Verde et al. 2003; see also Spergel et al. 2005; Dunkley et al. 2009; Komatsu et al. 2009). Taken together, these results established the 'concordance cosmology' as a standard model.

Prior to this, the interpretation of lookback surveys was hampered by a degeneracy between evolutionary and cosmological effects; the emergence of the concordance cosmology broke this degeneracy. 'Precision' cosmology is thus a key enabling factor for quantitative studies of the evolution galaxies in terms of their global physical properties. It provides the crucial information needed to translate from observed quantities to intrinsic ones: from fluxes to absolute luminosities, from apparent to physical sizes, and from number counts to comoving densities.

### Spectroscopic Galaxy Redshift Surveys at High and Low Redshift

At least for large distances, the most robust and reliable distance indicator is a spectroscopic measure of redshift. While the advent of relatively sensitive CCD detectors made it possible, for the first time, to obtain spectra for relatively faint, distant galaxies, it was only with the multiplexing power of multiobject spectrographs that spectroscopic lookback surveys became practicable. By the early 1990s, based on samples of tens to hundreds of galaxies, the first strong evidence had begun to emerge for significant differences between 'high redshift' galaxies ($z \lesssim 0.5$, with a mean redshift of 0.1) and those in the local universe — that is, the first signs of evolution (see Koo & Kron, 1992, for a review of these results). In the mid 1990s, the sensitivity of newly commissioned 10 m-class telescopes pushed the frontiers of high redshift science. In 1996, for example, Cowie et al. used the LRIS spectrograph on Keck to assemble a sample of nearly 400 galaxies in the range $0.2 \lesssim z \lesssim 1.7$. Over the following five years, the DEEP survey (Vogt et al., 2005; Weiner et al., 2005) obtained 658 redshifts for galaxies

with a median redshift of 0.65. 2002 marked a second generational change and a watershed for spectroscopic lookback surveys. In this year, two major surveys began, both using a new generation of multiobject spectrographs mounted on 10 m-class telescopes. Using DEIMOS on Keck, the aim of DEEP-2 was to collect more than 50000 $z \gtrsim 0.7$ galaxies (Davis et al., 2003; Faber et al., in preparation). They collected 8000 spectra in their first year of operation. The goal of the VIMOS VLT Deep Survey (VVDS; Le Fèvre et al., 2005) was to collect on the order of 135000 redshifts; they obtained 10000 in their first year.

Since evolution can only be inferred from statistical differences between the high– and low–redshift galaxy populations, *local galaxy surveys provide the crucial 'control' sample with which to compare higher redshift results.* Just as multiobject spectrographs led to an explosion in our knowledge of the high-redshift universe, our knowledge and understanding of the local universe has been revolutionized by a succession of ambitious spectroscopic surveys. The Las Campanas Redshift Survey (LCRS; Shectman et al., 1996) collected more than 26000 redshifts between 1988 and 1994. Between 1995 and 2002, the Two Degree Field Galaxy Redshift Survey (2dF GRS; Colless et al., 2001, 2003) collected approximately 220000 galaxy redshifts. Using a dedicated telescope, the Sloan Digital Sky Survey (SDSS; York et al., 2000; Strauss et al., 2002) began in 2000, aiming to collect 1000000 redshifts over five years. The survey has since been extended, and is ongoing.

As an aside, it is worth reflecting on the pace of these developments. In their 1991 review of 'Redshift Surveys of Galaxies', Giovanelli & Haynes remark that '[b]y any standards of human activity, the redshift industry is among the most successful, as it can boast a sustained growth rate in excess of 10 % per year over its whole 80-year history, and has the potential to maintain its growth for the foreseeable future.' It 1980, the combined total of spectroscopic redshifts, including multiple determinations for individual objects, was on the order of 8000 (Palumbo, Tanzella-Nitti, & Vettolani, 1983). Had we sustained 10 % year-on-year growth, we would expect to now have something like 145000 redshifts. This number is comparable to the average *annual* output of SDSS.[1] The impact of multiobject spectrographs on observational cosmology cannot be understated.

### Photometric Redshifts

The generic problem with spectroscopic surveys remains that, in comparison to photometric imaging, they are observationally expensive. Moreover, with current technologies, spectroscopy is limited to objects brighter than $I \sim 24$ and $K \sim 20$ (see, e.g., Fernández-Soto et al., 2001; Cimatti et al., 2002; Kriek et al., 2006, 2008a). This limitation became especially important with the launch of the *Hubble Space Telescope* (*HST*), and in particular following the Hubble Deep Field (HDF) project (Williams et al., 1996). Given that spectroscopy was impossible, how best to those galaxies at the edge of the observable universe?

[1] Between them, the SDSS and the AAO are responsible for approximately 85 % of all redshifts ever collected (Driver et al., 2009).

As an alternative to spectroscopy, it is possible to derive an approximate redshift the photometric spectral energy distribution (SED). Rather than emission or absorption lines, these photometric redshift techniques rely on broad spectral shape and gross features like the Balmer/4000 Å or Lyman breaks to constrain a galaxy's redshift (see, e.g., Connolly et al., 1995; Steidel, Pettini & Hamilton, 1995). The idea of photometric redshifts was presented by Baum (1962), who used the technique to measure distances to several clusters in the range $0.1 \lesssim z \lesssim 0.5$.[2] Baum's measurements were based on photometry in nine broad bands, including two near infrared filters, coadded for 2–4 galaxies per cluster. In this way, he found a redshift of $0.44 \pm 0.03$ for the cluster 3C295; the modern value is 0.46. By 1985, using stellar population synthesis models to fit optical SEDs, Koo had achieved a photometric redshift accuracy of $\Delta z \sim 0.04$ for $z \lesssim 0.5$ and $\Delta z \sim 0.06$ for $z \sim 0.6$, thereby extending the idea of photometric redshifts to the general galaxy population (see also, e.g., Loh & Spillar, 1986). In this way, photometric redshifts — or "phot-$z$s" — provide a kind of "poor person's redshift machine" (Koo, 1985).

At least in principle, photometric redshifts also offer a means of probing those very faint galaxies for which spectroscopy is impractical. It is important to note, however, spectroscopic redshifts are still needed for a representative subsample to test the validity of the photometric determinations (see, e.g., Connolly et al., 1995; Brammer et al., 2008, but see also Quadri & Williams, 2009, who present an empirical test of photometric redshift accuracy using galaxy pairs.). This fact explains why photometric redshifts were not widely used until *after* large spectroscopic surveys were completed. Further, even with a good "spec-$z$" comparison sample, no guarantees can be made on the reliability of any individual galaxy's phot-$z$. Photometric redshift surveys can thus extend and complement spectroscopic surveys, but they do not eliminate the need for spec-$z$s for a large and representative sample (see, e.g., Fernández-Soto et al., 2001).

*The use of photometric redshifts makes it possible to analyze vastly larger samples of galaxies, albeit with significant uncertainties.* Phot-$z$s thus became the default means of analyzing the HDF data in particular (see, e.g., Subbaroo et al., 1996; Gwyn & Hartwick, 1996; Sawicki, Lin & Yee, 1997; Connolly et al., 1997), and high-redshift galaxies in general. The successes in the HDF (see, e.g., Hogg et al., 1999; Fernández-Soto et al., 2001) spurred a number of efforts to develop and apply new photometric redshift estimation techniques (e.g., Csabai et al., 2000; Bolzonella, Miralles & Pello, 2000; Benítez, 2000; Firth, Lahov & Somerville, 2003; Vanzella et al., 2004). But there is one project that deserves special mention here: by combining optical photometry in 5 broad- and 12 medium-band (roughly equivalent to $R \sim 10$ spectroscopy), the COMBO-17 survey achieved a photometric redshift accuracy of $\Delta z/(1+z) \sim 0.02$ for 25000 galaxies $z \lesssim 1$ galaxies (Wolf et al., 2003, 2004). In terms of the successful application of photometric redshift techniques, this survey probably contributed more than any other to the acceptance and adoption of photometric redshift techniques as a legitimate means of analyzing large, representative samples of high-redshift galaxies.

[2]See also Stebbins & Whitford 1948, who, pursuing an observing program devised by Hubble and Baade, derived "a relation between color index and red shift for extragalactic nebulae".

### Stellar Mass Estimation

Photometric redshift techniques require a set of template spectra with which to fit the observed SEDs of individual galaxies. By using empirical templates (*i.e.*, spectra of galaxies of known spectral type), the best fit template can then be used to provide an approximate spectral classification. The alternative is to use synthetic spectra based on stellar population models (e.g., Le Borgne & Rocca-Volmerange, 2002; Bruzual & Charlot, 2003; Maraston, 2005). The advantage to this approach is that each template can be associated with a stellar mass-to-light ratio.

*Using a large library of synthetic stellar population spectral templates that span a wide enough range of possible star formation histories, it is thus possible to derive stellar mass estimates from galaxy SEDs* — and, equally, star formation rates, mean stellar ages and the other essential stellar population parameters of interest for quantifying the evolution of the galaxy population (Tinsley & Gunn, 1976; Tinsley, 1978). While these techniques are conceptually related to photometric redshifts, they are obviously not specific to photometric surveys. As an important example, Kauffmann et al. (2003a,b) used spectral diagnostics to derive stellar mass estimates for SDSS galaxies (see also, e.g. Heavens, Jiminez & Lahov, 2000; Panter, Heavens & Jiminez, 2003).

The accuracy of these techniques is limited by (at least) three important factors. Firstly, they depend on accurate stellar population models, including a broad enough range of (parametric) star formation histories to describe the full diversity of real galaxies (see, e.g., Charlot, Worthey & Bressan, 1996; Maraston et al., 2006; Kannappan & Gawiser, 2007; Conroy, Gunn & White, 2009). Secondly, and related to this point, there is the issue of the stellar initial mass function (IMF). The shape of the IMF — and how or whether it varies with, e.g., star formation rates, environment, and redshift — is still a major unknown (see, e.g., Salpeter, 1955; Kroupa, 2002; Hopkins & Beacom, 2006; van Dokkum, 2008; Cerviño & Valls-Gabaud, 2008). Finally, the accuracy of stellar mass estimates is fundamentally limited by generic degeneracies between the observable properties of different stellar populations with different mass-to-light ratios (see, e.g. Rix & Rieke, 1993; Brinchmann & Ellis, 2000; Bell & de Jong, 2001; Gallazzi & Bell, 2009).

### The Critical Importance of Near Infrared (NIR) data

Most of the broad spectral features on which modern SED-fitting algorithms rely fall in the restframe optical (Connolly et al., 1995). For $z \gtrsim 1$, these features are redshifted beyond the observers' optical window and into the near infrared (NIR). For this reason, NIR imaging is a critical requirement for studies of the $z \gtrsim 1$ galaxy population: *deep NIR observations are the key to opening the door to the $z \gtrsim 1$ universe* (see, e.g., Hogg et al., 1997; Rudnick et al., 2001; Labbé et al., 2003; Förster-Schreiber et al., 2006).

There is a second reason why having NIR data is crucial. A galaxy's restframe ultraviolet emission is dominated by hot, bright, young stars. This means that optically selected galaxy samples at $z \gg 1$ become progressively more biased towards star forming galaxies. By selecting in the observed NIR, galaxies are se-

lected on their restframe optical emission, which is dominated by the longer lived, main sequence stars that typically constitute the bulk of a galaxy's mass. That is, NIR-selected samples of $z \gtrsim 1$ galaxies offer a practical means of constructing mass-limited samples of distant galaxies, and so to get a representative census of the massive galaxy population. (this point is discussed further below; see also, e.g., Adelberger & Steidel, 2000; Cimatti et al., 2002; Labbé et al., 2003; van Dokkum et al., 2006).

### The Essential Ingredients

Hence it was my good fortune that by the time I began work on this thesis in 2004, a series of technological, technical, and conceptual innovations had led to a true Kuhn-ian revolution in the science of galaxy formation and evolution (Kuhn, 1962). These included: the explosion of high– and low–redshift spectroscopic redshift surveys afforded by massively-multiplexing multiobject spectrographs; the increased sensitivity and resolution of 10 m class telescopes and space based observatories; the maturation and acceptance of techniques for estimating both photometric redshifts and stellar mass-to-light ratios; the advent of wide-field NIR imagers; and the establishment of the concordance cosmology as a standard model for cosmology. Together, these developments provided all the necessary requirements for detailed, quantitative studies of the general galaxy population in terms of their luminosities, stellar masses, star formation, sizes, structures, and morphologies.

Further, using these kinds of surveys, both the cosmic star formation history (Lilly et al., 1996; Madau et al., 1996; Hopkins, 2004, and references therein; Bowens et al. 2007, Bouwens et al. 2009) and the buildup of stellar mass (e.g., Fontana et al., 2004, Drory et al., 2005, Fontana et al., 2006, Arnouts et al., 2007, Pozzetti et al., 2007, Pérez-González et al., 2008, Marchesini et al., 2009; see also Trentham, Wilkins & Hopkins 2008) have now been constrained out to $z \sim 5$; that is, over approximately 90 % of cosmic history.

## 3 Galaxy Formation and Evolution – What we have learned

### Round Ones and Flat Ones; Red Ones and Blue Ones; Old Ones and Young Ones — A dichotomy among z ∼ 0 galaxies

Even before galaxies came to be known as 'galaxies', it was recognized that 'extragalactic nebulae' fell into two broad classes (Hubble, 1926). This classification was originally made on the basis of structure: the distinction was between 'late type' galaxies, which showed conspicuous spiral arm structures and/or a bright nuclear region, and smooth, featureless elliptical 'early type' galaxies. (At lower masses, irregular and peculiar galaxies are an important third class; in what follows, I will ignore these galaxies.) The two classes of galaxies have since been shown to have rather different properties. First, in general, elliptical galaxies have redder colors than spirals (Strateva et al., 2001; Blanton et al., 2003; Driver et al., 2006). This reflects the fact that early type galaxies tend to be dominated by relatively old stellar populations, whereas late type galaxies tend to be actively forming new stars (Kauffmann et al., 2003b; Brinchmann et al., 2004; Wyder et

al., 2007). Early type galaxies tend to lie preferentially in higher density environments (Blanton et al., 2005; Baldry et al., 2006; van der Wel et al., 2008). Further, the most massive and/or luminous galaxies tend to be early type (Strateva et al., 2001; Kauffmann et al., 2003b).

The emergent picture, then, is of a population of 'developed' massive, quiescent, centrally-concentrated, and old elliptical or spheroidal galaxies that are found in more dense environments, as distinct from the less massive, star forming, 'developing' disk-dominated population that dominates in the field and in small groups (see, e.g., Ellis et al., 2005; Conselice, 2006). That is, the two different classes of galaxies appear to correspond to distinct evolutionary states. *Elucidating the nature of and physical basis for the difference between developed and developing galaxies is a major challenge for cosmological models of galaxy formation and evolution.*

### The Central Importance of Stellar Mass

Within each of these populations, however, galaxies are remarkably well behaved. Although the basic, global properties of individual galaxies — for example, luminosity, mass, size, local density, star formation rate, mean stellar age, metallicity, and gas content — vary by orders of magnitudes, there exist very tight and well-defined correlations between essentially all of these properties for each of these two classes (see, e.g., Minkowski, 1962; Faber & Jackson, 1976; Tully & Fisher, 1977; Sandage & Visvanathan, 1978; Dressler, 1980; Djorgovsky & Davis, 1987; Dressler et al., 1987; Magorrian, 1998). Presumably, key information about the physical processes governing galaxies' formation and evolutionary histories are encoded in the slope of, and scatter around, these relations.

One of the most important insights gleaned from the SDSS has been to confirm the idea that most, if not all, of these relations can be understood as being primarily a sequence in mass (e.g. Kauffmann et al., 2003b; Shen et al., 2003; Blanton et al., 2005; Baldry et al., 2006; Gallazzi et al., 2006). Given a galaxy's stellar mass, it is thus possible to predict a wide variety of global properties with a remarkable degree of accuracy. Moreover, Kauffmann et al. (2003b) have shown that the distinction between developing and developed galaxies coincides with an apparent 'transition mass' of $\sim 3 \times 10^{10}$ $M_{\odot}$; above this limit, most galaxies are quiescent, early type galaxies (see also, e.g., Blanton et al., 2005). In this sense, stellar mass appears to be a fundamental parameter in determining — or at least describing — a galaxy's current state of evolution. (Although see, e.g., Kauffmann et al., 2006; Franx et al., 2008; Graves, Faber & Schiavon, 2009, who argue that stellar surface density or velocity dispersion may be a more fundamental parameter in describing this transition.)

*By observing changes in the scaling relations between stellar mass and other global galaxy properties, we can therefore hope to learn something about the processes that shape the lives of galaxies.* In particular, we would like to determine when and why these relations first come about, as well as when and how galaxies make the transition from developing to developed.

**Star Formation Quenching and the Transition From Blue to Red**

In a landmark study based on the COMBO-17 photometric redshift survey, Bell et al. (2004b) showed that the distinct color–magnitude relations for red and blue galaxies[3] are already in place at $z \sim 1$ (see also, e.g. Im et al., 2002; Tanaka et al., 2005; Weiner et al., 2005; Willmer et al., 2006; De Lucia et al., 2007). As in the local universe, the 'red sequence' is dominated by largely quiescent, structurally early type galaxies (Bell et al., 2004a; Holden et al., 2008). The morphology–density relation seen locally is also in place by this time (van der Wel et al., 2007), as is the 'fundamental plane' relation (a relation between dynamical mass and surface brightness) for early type galaxies (Treu et al., 2002; van der Wel et al., 2004; di Serego Alighieri et al., 2005). Further, both the color–magnitude relation for red galaxies and the fundamental plane for early type galaxies evolve in a manner that is consistent with passive fading of an old ($z \gtrsim 2$) stellar population (see, e.g., Bell et al., 2004b; van der Wel et al., 2004; Cimatti, Daddi & Renzini, 2006).

That is, the galaxy population at $z \sim 1$ appears to be qualitatively similar to that of the present day. But this is not to say that nothing has changed. The total mass density of red sequence galaxies has roughly doubled between $z \sim 1$ and the present (Borch et al., 2006; Faber et al., 2007; Brown et al., 2008). At the same time, that total mass density of blue galaxies, which are actively star forming, remains more or less constant (Borch et al., 2006; Arnouts et al., 2007; Bell et al., 2007). We are thus faced with a rather curious situation: the number of passive galaxies grows continually over time, while the combined stellar mass of actively star forming galaxies remains unchanged.

These results have been accommodated within the ΛCDM paradigm through the postulation of a 'quenching' mechanism, which acts to disrupt star formation in massive galaxies, thereby inciting a transition from blue to red. While the physical basis for this *ad hoc* inclusion to cosmological models of galaxy evolution is not understood, a number of candidates have been proposed, including the prevention of gas accretion onto high mass halos by shock heating (e.g., Dekel & Birnboim, 2006; Cattaneo et al, 2006, 2008; van den Bosch et al., 2008) and energetic or kinetic feedback from active galactic nuclei (e.g. Croton et al., 2006; Bower et al., 2006; Menci et al, 2006; Somerville et al., 2008; Bower, McCarthy & Benson, 2008), possibly triggered by a merger event. This quenching mechanism is also required to get the right number of massive galaxies at high redshift (see, e.g., Cattaneo et al, 2006; Menci et al, 2006)

*Quantifying the evolution in the red/blue fraction among massive galaxies thus provides basic observational constraints on the mechanism whereby star formation is quenched.* In so doing, it offers a potential means of constraining the relative

[3]Here and in what follows, 'red' is often used as a proxy for 'quiescent' and/or 'early-type'. There is considerable — but not total — overlap between galaxy samples selected by morphology (*i.e.*, early-/late-type), photometry (*i.e.*, red/blue), and spectroscopy (*i.e.*, quiescent/star forming). In this sense, it reasonable, but not strictly accurate, to use the terms red, quiescent, and early type (or blue, star forming, and late type) as if they were interchangeable. However, it is still unclear which which of these distinctions is/are the most 'fundamental' (see, e.g. Blanton et al., 2005).

importance of AGN feedback, mergers, and gas accretion in regulating the star formation process in particular, and in galaxy evolution in general.

**The Situation at High Redshift**

Without the luxury of spectroscopic redshifts for representative samples of high redshift galaxies, the pioneering studies of the high redshift universe relied on samples selected on the basis of observed colors.[4] In particular, the Lyman-break selection criterion proposed by Steidel et al. (1993, 1996, 1999, 2003) was shown to be an efficient means of selecting galaxies at $z \gtrsim 3$. This selection works by isolating the sharp spectral break caused by absorption of UV photons between Lyman-$\alpha$ (1216 Å) and the Lyman limit (912 Å) by hydrogen atoms in the IGM. Their small sizes (Steidel, Pettini & Hamilton, 1995; Giavalisco, Steidel & Macchetto, 1996), relatively strong clustering (Adelberger et al., 1998; Giavalisco et al., 1998; Giavalisco & Dickinson, 2001), moderate star formation rates (see, e.g. Adelberger & Steidel, 2000; Pettini et al., 2001; Papovich, Dickinson & Ferguson, 2001), and relatively low masses (Sawicki & Yee, 1998; Shapley et al., 2001) were all consistent with the idea that these Lyman Break Galaxies (LBGs) being 'primordial' massive galaxies (see, e.g., Giavalisco & Dickinson, 2001; Giavalisco, 2002).

While this color selection was deliberately targeted towards high-redshift galaxies, by selecting on the basis of a restframe UV feature, it was also implicitly limited to star forming galaxies with little or no dust obscuration (see, e.g. Adelberger & Steidel, 2000). Using a NIR color selection criterion, which is based on the restframe optical Balmer and 4000 Å breaks in the spectra of old stellar populations, Franx et al. (2003) identified a nearly completely disjoint population of Distant Red Galaxies (DRGs). These galaxies were quickly spectroscopically confirmed to lie mostly at $z \gtrsim 2.3$ van Dokkum et al. (2003, 2004), and to be more massive, older, and dustier than the LBGs (van Dokkum et al., 2004; Förster-Schreiber et al., 2004; Labbé et al., 2005).

The discovery of the DRG population had two important implications. First, the fact that many DRGs were found to have genuinely old stellar populations significantly pushed back the epoch of formation for massive galaxies. It had been shown that a decent fraction of Extremely Red Objects (EROs, selected on the basis of their optical–minus–NIR colors; see McCarthy 2004 and references therein) were evolved and quiescent galaxies at $z \lesssim 1.3$, and quiescent galaxies at $1.4 \lesssim z \lesssim 2$ had been found in spectroscopic surveys (e.g. Cimatti et al., 2004; Glazebrook et al., 2004; McCarthy et al., 2004; Daddi et al., 2004, 2005). But the DRGs provided evidence for the emergence of significant numbers of massive, evolved galaxies in the first 2–3 Gyr of the history of the universe.

[4] Color selection techniques had in fact been an important part of moderate- and high-redshift spectroscopic studies since the mid 1990s. It was quickly realized that purely flux-limited samples were an inefficient means of isolating distant galaxies: most faint galaxies are nearby, low luminosity galaxies (see, e.g. Lilly et al., 1996; Ellis et al., 1996; Cowie et al., 1996). In the context of the earlier discussion of photometric versus spectroscopic redshifts, as well as in what follows, these kinds of color selection techniques can be thought of as an extremely crude photometric redshift. The same can be said of the 'red sequence cluster method' of identifying massive, red sequence galaxies in clusters (Gladders & Yee, 2000, 2005).

Secondly, particularly given that DRGs were found to exist in numbers comparable to the LBGs, these results also showed that the use of the Lyman break technique in particular, and color-selected samples in general, could yield a badly biased and significantly incomplete view of the high redshift universe (Labbé et al., 2003, 2005; Reddy et al., 2005; van Dokkum et al., 2006). *To get a fairer picture of the high redshift universe, what was needed was a complete sample of galaxies selected by stellar mass*; van Dokkum et al. (2006) showed that this could be done efficiently on the basis of photometric redshifts, using NIR-selected galaxy samples.

**Old Galaxies in the Young Universe**

By obtaining very deep, restframe optical spectra of a mass-limited sample of $z \sim 2.3$ galaxies, Kriek et al. (2006, 2008a) made a significant advance on previous photometric and spectroscopic studies of high redshift galaxies. These galaxies were $K$-selected, and were selected to have photometric redshifts greater than 2 (11 out of 36 of these galaxies were selected from an early version of the data presented in Chapter II). On the basis of their spectra, roughly half of these galaxies were shown unambiguously to have evolved stellar populations, and little or no ongoing star formation. Further, there was the tantalizing suggestion that the passive galaxies may already follow a red sequence (Kriek et al., 2008b; see also the recent results by Williams et al., 2009; Brammer et al., 2009). In other words, *a significant number of the massive galaxies at $z \sim 2.3$ have stellar populations that are consistent with their being "fully formed" massive galaxies.* (At least some) massive galaxies form or assemble their stars very early on in the history of the universe — and in a very short time.

**The Continued Evolution of Quiescent Galaxies**

Within the paradigm of hierarchical structure formation, the most massive objects are expected to be both young *and* old. They are old in the sense that they are expected to form preferentially from the highest overdensities, which are the first to collapse. On the other hand, since large structures form through successive mergers between smaller progenitors, the most massive galaxies are expected to have assembled only relatively recently, and in this sense are quite young. There is thus a crucial distinction to be made between a galaxy's mean stellar age, and its formation age; *i.e.*, the time since it first assumed its present form. With the introduction of a quenching mechanism, the models thus 'predict' that these massive galaxies will continue to grow through accretion and/or minor and major mergers, even after their star formation has effectively ceased (see, e.g., De Lucia et al., 2006).

With this in mind, there is (at least) one important difference between the $z \sim 2.3$ galaxies and local galaxies of the same stellar mass: the high redshift galaxies are much smaller. Using a combination of *HST* and Keck laser guide-star assisted adaptive optics imaging, van Dokkum (2008) measured sizes for 9 out of the 11 quiescent galaxies in the Kriek et al. (2006) sample. They found sizes on the order of 3–10 times smaller than typical galaxies of the same mass in the local universe. These galaxies have physical surface densities (measured within the

central kiloparsec) that are 2–3 times higher than their local counterparts (Bezanson et al., 2009). As further confirmation of the remarkable compactness of these galaxies, van Dokkum, Kriek & Franx (2009) measured a velocity dispersion of $510^{+165}_{-95}$ km/s for one of these galaxies (see also the $z \sim 1.6$ results of Glazebrook, 2009). This work confirmed and consolidated the work of a number of authors, including Daddi et al. (2005), Trujillo et al. (2006), Trujillo et al. (2007), Zirm et al. (2007), and Toft et al. (2007). Similarly compact galaxies have since been found at $1 < z < 2$ by Cimatti et al. (2008) and Damjanov et al. (2009; see also Longhetti et al., 2007; Saracco et al., 2009), as well as at $1.7 < z < 3.0$ by Buitrago et al. (2008).

The observation that massive galaxies at high redshift are so much smaller than their local counterparts implies that each of these massive galaxies have to significantly grow in size in order to match the properties of galaxies found in the local universe. Theoretical candidates for the physical processes that drive this strong size evolution include a combination of accretion and mergers (van der Wel et al., 2009; Hopkins et al., 2009), and kinetic feedback from AGN (Fan et al., 2008). In this way, *confirming and quantifying the $z \lesssim 2.3$ size evolution of massive galaxies has the potential to provide constraints on the recent merger histories of massive galaxies, and thus the relative importance of mergers in galaxy growth.*

## 4 This Thesis

With all of the above as background, the three key questions addressed in this thesis are:

- When are massive galaxies formed?
- When is star formation quenched in massive galaxies?
- What happens to these galaxies after their star formation has ended?

A recurring theme throughout this work is the importance of systematic errors. Those who know me will be aware of my predilection for asking, "What could possibly go wrong?"; this is a question that appears more than once in what follows. While not nearly as glamorous or inspiring as being able to conclude that, say, at least half of all massive galaxies were formed in the last 7 Gyr, the results connected to error analysis are at least as important as the more 'astronomical' results — if not more so. It is only through detailed error analysis that we can learn exactly how well we know what we think we know.

### 4.1 The Rise of Red Galaxies

The first two of these key questions are addressed in Chapters II and III. This work is based on a NIR-selected catalog of the Extended Chandra Deep Field South (ECDFS), based on photometry in ten broadband filters compiled as part of the MUtiwavelength Survey by Yale–Chile (MUSYC; Gawiser et al., 2006). This catalog is based on publicly available optical imaging obtained from a number of different sources (Hildebrandt et al., 2006, and references therein), supplemented

by original optical and NIR data taken by the MUSYC team. The ECDFS is one of the premier sites for deep field galaxy evolution studies, with observations spanning the UV to the radio. The additional of NIR data fills a crucial gap in the wavelength coverage of this important field. Particularly in concert with the many existing and upcoming survey projects targeting the ECDFS (see references given in Chapter II), the MUSYC NIR-selected catalog provides an outstanding laboratory for $z \gtrsim 1$ galaxy studies.

**Chapter II** is devoted to the integration of these different datasets into a mutually consistent whole. The resultant catalog comprises over 10000 reliable detections above our nominal selection and completeness limit of $K^{\rm AB} = 22$ over an effective survey area of 818 square arcmin, including nearly 9000 high-redshift galaxies, and an approximately complete sample of nearly 1300 $M_* > 10^{11}$ $\rm M_\odot$ galaxies at $z_{\rm phot} < 1.8$. In order to maximize the legacy value of the data, the data calibrations have been extensively tested through both internal consistency checks and external comparisons to existing surveys. In particular, we found a major calibration error in the COMBO-17 data of the ECDFS, which has since been corrected (Wolf et al., 2008)

The MUSYC ECDFS catalogs have been made freely and publicly available, including the reduced images, ten band photometry, a comprehensive compilation of spectroscopic redshifts from literature sources, state of the art photometric redshift determinations, and restframe photometry. This restframe photometry has been derived using an IDL implementation of an algorithm described by Rudnick et al. (2003), which I developed and tested. This utility, dubbed InterRest, has also been made freely available. The imaging data have since been incorporated into the FIREWORKS (Wuyts et al., 2008) catalog of the GOODS region, and form the backbone of the *Spitzer* Infrared MUSYC Public LEgacy (SIMPLE) survey (Damen et al., 2009), which adds extremely deep *Spitzer Space Telescope* data for the full field.

In **Chapter III**, the MUSYC data are used to construct the color–magnitude and color–stellar mass diagrams for $z \lesssim 2$. A red sequence is detected out to at least $z \sim 1.2$, but beyond this point, the NIR data are not deep enough to distinguish distinct red and blue populations. The $z \lesssim 1.2$ color evolution of the red sequence is consistent with the passive fading of old stars, with no evidence of evolution in the scatter around the color-magnitude or color-mass relations (see also Ruhland et al., 2009). But this is not to say that the red sequence does not evolve as a population: the number density of red galaxies grows by a factor of at least 5 between $z \approx 2$ and the present, and by a factor of at least 2 after $z \approx 1$. In contrast, the total number density of massive galaxies is approximately constant over $0 < z < 1$.

Our results link the $z \lesssim 1$ results of, e.g., Bell et al. (2004b) and Faber et al. (2007) to the $z \sim 2.3$ results of Kriek et al. (2006, 2008a,b). Bridging this $1 \lesssim z \lesssim 2$ gap is of particular interest since the cosmic star formation rate drops by an order of magnitude in this interval (see, e.g., Hopkins, 2004; Nagmine et al.,

2006; Panter et al., 2007; Tresse et al., 2007; Pérez-Gonzaléz et al., 2008). This is also the era in which massive galaxies first emerge in large numbers (Juneau et al., 2005; Borch et al., 2006; Fontana et al., 2006; Pozzetti et al., 2007). These results are also complementary to studies that consider the mean star formation rate as a function of stellar mass (e.g., Juneau et al., 2005; Zheng et al., 2007; Damen et al., 2009a).

Since all passive galaxies are red, but not all red galaxies are passive, the results presented in Chapter III can be used to place an upper limit on the number of massive galaxies that have had their star formation effectively quenched. We therefore conclude that *at most 20 % of all local massive, red sequence galaxies had finished their star formation by $z \sim 2$, and that at least 50 % stopped forming stars only after $z \sim 1$.* Whatever the mechanism that is responsible for quenching star formation in massive galaxies, this is when it operates.

This work was the first of its kind to include a detailed investigation of the systematic uncertainties associated with these kinds of measurements. By systematically varying individual aspects of the experimental design, we were able to directly quantify the relative importance of a number of effects, including data calibration, photometric methods, photometric redshift uncertainties, errors in stellar mass estimates, and field-to-field variance. This allows us to identify the most important sources of systematic error or uncertainty; these are, in order: 1.) systematic differences in the analysis of the high-redshift galaxies and the $z \sim 0$ comparison sample; 2.) details of the photometric redshift calculation; and 3.) the basic photometric calibration of the data. Each of these sources of systematic uncertainty outweighs the statistical uncertainties, including those due to field–to–field variance. We also show that, for example, the choice of templates for the photometric redshift calculation, random photometric redshift errors, and systematic errors in the stellar mass estimates are not dominant sources of uncertainty.

By identifying and focusing on the most important sources of error and uncertainty, we have minimized our vulnerability to these effects, and so provide a more robust result than previous studies. Bearing this in mind, Chapter III also includes a complementary analysis that, while model dependent, is largely immune to the three most important sources of systematic errors. These results agree remarkably well with those from our more sophisticated analysis based on photometric redshifts. To the extent that these two analyses are consistent, this gives some confidence that we may in fact have the important systematic effects 'under control'.

Quantifying the systematic uncertainties associated with each different aspect of our experimental design is not only useful in interpreting our specific results, but also as a guide for the design of future surveys and experiments. In terms of future work, it is highly significant that, even for this relatively modest-sized field, systematic uncertainties outweigh the statistical errors, even after accounting for the effects of field-to-field variance (see also Marchesini et al., 2009). This implies that *future surveys will require better analysis as much as better data in order to improve on the results given in Chapter III.*

### 4.2 (No) Compact Galaxies in the Local Universe

**Chapter IV** addresses the third key question: what happens to galaxies after they finish their star formation? Specifically, we test the claim that the passive galaxies observed at $z \sim 2.3$ have to undergo significant structural evolution in order to match the $z \sim 0$ galaxy population by using data from the SDSS to look for local red sequence galaxies with comparable sizes and masses. Even more specifically, Chapter IV considers the possibility that such massive, compact galaxies might be missing from the SDSS catalog due to selection effects.

As part of the SDSS spectroscopic target selection algorithm, there are two selection criteria that exclude high surface brightness objects; these are intended to ensure against saturation and cross-talk in the spectroscopic detectors (Strauss et al., 2002). We show that this makes incompleteness a concern for bright and compact galaxies at *low* redshifts: even if red sequence galaxies with the sizes and masses seen at $z \sim 2.3$ were to exist $z \lesssim 0.05$, they would be not be selected as SDSS spectroscopic targets. For this reason, we look for massive, red sequence, compact galaxy candidates in the range $0.066 < z < 0.10$, where SDSS should be $\gtrsim 50$ % complete, and still be able to adequate resolve such compact galaxies.

If the $z \sim 2.3$ galaxies were not to evolve in either size or number density, we would expect to have found on the order of $\sim 6500$ such galaxies within this sample. Instead, after discarding those galaxies with obvious reasons to distrust their size and/or mass measurements, and after corroborating the size and mass measurements of the remainder of the sample based on their velocity dispersions, we find no (0) galaxies that are consistent with being passively evolved versions of the $z \sim 2.3$ galaxies. Chapter IV also includes a search for massive, compact galaxies in the SDSS photometric sample, for which selection effects should not be a concern; again, we find no plausible candidates. Massive, compact, red sequence galaxies are not just missing from the SDSS catalogs, they are simply not there to be found in the local universe.

This confirms the conclusions of van Dokkum (2008): *massive galaxies must undergo significant structural evolution after $z \sim 2$, even after their star formation has effectively ended.* The mechanism for this size evolution is not clear. However, using a simple statistical argument, we suggest that the fact that each and every one of the $z \sim 2.3$ galaxies must evolve in size implies this growth cannot be explained by a highly stochastic mechanism like major mergers.

### 4.3 Estimating Galaxies' Masses

The connection between observations and theory of galaxy evolution hinges on our ability to make the link between galaxies' observed SEDs and their intrinsic stellar populations (or, conversely, to predict SEDs for galaxies in the models given their star formation histories). As described above, these estimates are plagued by a number of different kinds of random and systematic uncertainties. The primary goal of **Chapter V** is to test our ability to derive robust stellar mass estimates from five band optical SEDs for galaxies in the local universe. Specifically, we compare stellar mass estimates to estimates of total mass based on galaxy dy-

namics, derived using the latest generation of SED-fit stellar mass-to-light ratios and the most robust size and flux measurements available for galaxies in the SDSS.

We find very good correspondence between stellar and dynamical mass estimates, but only provided that we account for non-homology (*i.e.* structural differences among galaxies) when deriving the dynamical mass estimates. In particular, we find no statistically significant trends in the stellar-to-dynamical mass ratio as a function of spectroscopic stellar population indicators, or as a function of derived stellar population parameters. Nor do we find any significant trends as a function of direct observables (e.g., apparent magnitude or size), which would indicate a bias in these measurements. Further, we find no appreciable difference between the relation between stellar– and dynamical–mass estimates for galaxies in different states of activity. Because both of these mass estimates are model-dependent, neither one offers a truly solid basis for comparison; the very good consistency thus provides very strong circumstantial evidence — but not proof beyond a reasonable doubt — that there are no significant biases in either mass estimate, including the models used to derive them. With this caveat, we conclude that, at 99 % confidence, *across a broad range of stellar populations, the systematic, differential errors in stellar mass estimates based on five band optical photometry are less than 0.12 dex (40 %).*

We also find a rather mild mass-dependence for the stellar-to-dynamical mass ratio. This same trend is seen for subsamples of galaxies with the same structures; this observation is thus independent of the model used to account for structure and dynamical non-homologies. This implies a relatively small variation in the dark–to–baryonic mass ratios of galaxies as a function of mass, at least among the most massive galaxies. Moreover, at fixed mass, the observed scatter in stellar-to-dynamical mass ratios is small; the intrinsic variations in the dark-to-baryonic mass ratios among galaxies of the same mass may be as small as 0.04 dex ($\sim 10$ %).

## 5 Outlook

While our knowledge and understanding of the galaxy population and its evolution has exploded since the early 1990s, the past five or ten years has seen the establishment of galaxy formation and evolution as a mature science. One very important sign of this is the growing awareness of the importance of systematic errors. This awareness extends to interpretation of the theoretical models (see, e.g. Schaye & Dalla Vecchia, 2008; Wiersma, Schaye & Smith, 2009; Booth & Schaye, 2009). On the observers' side, this awareness is particularly apparent in connection with stellar mass estimates (see, e.g., van der Wel et al., 2006; Kannappan & Gawiser, 2007; Wuyts et al., 2007; Kriek et al., 2008a; Wuyts et al., 2009; Marchesini et al., 2009; Conroy, Gunn & White, 2009, Chapters III and V). Continued refinement of stellar evolution models will help to reduce the uncertainties in high redshift science. (Although there remains the question as to just how universal the stellar IMF is; the IMF is the new cosmology.)

In this sense, the upcoming NIR multiobject spectrographs will be extremely valuable. By providing spectroscopic redshifts for large numbers of $z \gtrsim 1$ galaxies,

they will allow significantly more robust redshift and stellar mass determinations for galaxies at $1 \lesssim z \lesssim 2$. The results of the NEWFIRM medium band photometric redshift survey (van Dokkum et al., 2009; Brammer et al., 2009) are also likely to be important in this regard. However, given the maturity of SED-fitting techniques, these results seem more likely to extend and refine our present knowledge than they are to overturn it. (Here, it is relevant that the results presented in Chapter III are dominated by systematic errors in the data analysis, rather than random uncertainties associated with statistical or measurement errors, and in particular that stellar mass estimates are not a dominant source of uncertainty.) In this sense, NIR spectrographs seem less likely to revolutionize our understanding of the $z \gtrsim 1.5$ universe in the same way as our understanding of the $0 \lesssim z \lesssim 1.5$ universe has been by optical spectra. That said, obtaining dynamical estimates of the total masses of $z \lesssim 2$ galaxies will provide important new consistency checks on stellar mass estimates, and in particular constraints on any redshift dependency in the IMF.

There are a number of instruments and observatories planned for the next decade that promise to provide qualitatively new information on the properties of high redshift galaxies. With ALMA, it will be possible to probe the molecular gas content of high redshift galaxies; ASKAP, MEERCAT, and the SKA will do the same for atomic gas. These telescopes will thus provide a means of exploring evolution in the process of star formation. LOFAR will map massively star forming galaxies and AGN out to $z \gtrsim 2$ As well as being able to push further down the luminosity function at lower redshifts, the next generation of space telescopes, including the JWST, should open a window to the $z \gtrsim 5$ universe.

In the meantime, however, the challenge remains to disentangle the relative importance AGN feedback, secular evolutionary processes, and environmental effects in the evolution of massive galaxies, and in particular which of these is/are responsible for the quenching of star formation and the structural transformation from disk to elliptical. Given the recent explosion of survey projects, in terms of both number and scope, the prospects for making progress on these questions with the data presently available are excellent. In particular, there have been several recent works that argue that environmental effects play a far less important role triggering star formation than previously thought (e.g., van der Wel et al., 2008; van den Bosch et al., 2009; Pasquali et al., 2009). There are also indications that only a small fraction of the total cosmic star formation can be directly linked to merger events (e.g., Noeske et al., 2009; Robaina et al., 2009). Further exploration of these results may be very revealing.

The most immediately obvious way of determining which processes are the most relevant in driving the transition from blue to red and from late to early type morphologies is still through observing the evolution in galaxy demographics. Two key questions are: the extent to which the quenching of star formation and the morphological transition is coupled, and how the AGN activity of the red/blue early-/late-type populations compare. If it can be shown, for example, that the quenching of star formation happens before the morphological transition (or vice versa), that would imply that these two transitions are the products of

distinct physical processes (see, e.g. Pozzetti et al., 2009). Another possibility is that, for example, if the peak of AGN activity was shown to occur only *after* the quenching of star formation, this might rule out AGN feedback as the primary quenching mechanism (in this context, see the recent results by Brown et al., 2009; Wild et al., 2009).

In all of this, morphology remains a relatively unexplored dimension of parameter space. Each of the major processes have rather different morphological signatures: for example, secular star formation in galaxies is clumpy, but largely symmetric; mergers and interactions induce strong asymmetries. In this sense, a comprehensive morphological census of the high- and low-redshift galaxy populations may also to qualitatively new constraints on the relative importance of the different processes thought to underpin the evolution of galaxies. This would require the development of new, non-parameteric measures of morphology (see, e.g., Conselice, 2003; Lotz, Primack & Madauo, 2004; Kelly & McKay, 2005; Abraham et al., 2007; Heurtas-Company et al., 2008).

These questions will only be resolved using very large, representative samples of galaxies at $z \gg 0$. The SDSS has greatly expanded and concretized our knowledge and understanding of the local galaxy population: what is needed is a similar sized survey with redshift resolution. By obtaining both optical spectra and subarcsec-resolution optical and NIR imaging for 250000 $z \lesssim 0.5$ galaxies, the GAMA survey Driver et al. (2009) may go a long way towards addressing these issues. Ultimately, however, what we would really like is an SDSS-sized sample of galaxies at $z \sim 1$ or even $z \sim 2$, with diffraction-limited NIR imaging and spectra obtained using a 30 m-class telescope.

## References


Abraham R G, Nair P, McCarthy P J, Glazebrook K et al., 2007, ApJ 669,184

Adelberger K L, Steidel C C, Giavalisco M, Dickinson M, Pettini M, Kellogg M, 1998, ApJ 505, 18

Adelberger K L & Steidel C C, 2000, ApJ 544, 218

Arnouts S et al., 2007, A&A 476, 137

Baldry I K, Glazebrook K, Brinkmann J, Ivezić Ž, Lupton R H, Nicol R C, Szalay A S, 2004, ApJ 600, 694

Baldry I K, Balough M L, Bower R G, Glazebrook K, Nichol R C, Bamford S P, Budavari T, 2006, MNRAS 373, 469

Baugh C M, 2006, RPPh 69, 3101

Baum W A, 1962, in *Problems of Extragalactic Research* (IAU Symp. 15), G C McVittie, ed. (Macmillan, New York)

Bell E F, de Jong R S, 2001, ApJ 550, 212

Bell E F, McIntosh D H, Katz N, Weinberg M D, 2003, ApJS 149, 289

Bell E F et al., 2004, ApJ 600, L11

Bell E F et al., 2004, ApJ 608, 752

Bell E F, Zheng X Z, Papovich C, Borch A, Wolf C, Meisenheimer K, 2007, ApJ 663, 834

Benítez N, 2000, ApJ 536, 571

Bennett C L et al., 2003 ApJ 583, 1

Bezanson R, van Dokkum P G, Tal T, Marchesini D, Kriek M, Franx M, Coppi P, 2009, ApJ 697, 1290

Blanton M R et al., 2003, ApJ 594, 186

Blanton M R, Eisenstein D, Hogg D W, Schlegel B K, Brinchmann J, 2005, ApJ 629, 143

Bolzonella M, Miralles J-M, Pelló R, 2000, A&A 363, 476

Booth C M, Schaye J, 2009, MNRAS 398, 53

Borch A, Meisenheimer K, Bell E F, Rix H-W, Wolf C, Dye S, Kleinheinrich M, Kovacs Z, Wisotzki L, 2006, A&A 453, 869

Bouwens R J, Illingworth G D, Franx M, Ford H, 2007, 670, 928

Bouwens R J, Illingworth G D, Franx M et al., 2009, ApJ 705, 936

Bower R G, Benson A J, Malbon R, Helly J C, Frenk C S, Baugh C M, Cole S, Lacey C G, 2008, MNRAS 370, 645

Bower R G, McCarthy I G, Benson A J, 2008, MNRAS 390, 1399

Brammer G B, van Dokkum P G, Coppi P, 2008, ApJ 686, 1503

Brammer G, Whitaker K E, van Dokkum P G et al., 2009, ApJ (accepted; arXiv:0910.227)

Brinchmann J & Ellis R S, 2000, ApJ 536, L77

Brinchmann J, Charlot S, White S D M, Tremonti C, Kauffmann G, Heckman T, 2004, MNRAS 351, 1151

Brown M J I, Zheng Z, White M, Dey A, Jannuzi B T, Benson A J, Brand K, Brodwin M, Croton D J

Brown M J I, Moustakas J, Caldwell N, Palamara D et al., 2009, ApJ 703, 150

Bruzual G, Charlot S, 2003, MNRAS 344, 1000

Buitrago F, Trujillo I, Barro G, Gallego J, Zamorano J, Conselice C J, 2008, ApJL 687, 61

Cattaneo A, Dekel A, Devriendt J, Guiderdoni B, Blaizot J, 2006, 370, 1651

Cattaneo A, Dekel A, Faber S M, Guiderdoni B, 2008, MNRAS 389, 567

Cerviño M & Valls-Gabaud D, 2008, MNRAS 338, 481

Charlot S, Worthey G & Bressan A, 1996, ApJ 457, 625

Cimatti A, Mignoli M, Daddi E, Pozzetti L, Fontana A, Saracco P, Poli F, Renzini A, Zamorani G, Broadhurst T, Cristiani S, D'Odorico S, Giallongo E, Gilmozzi R, Menci N, 2002, A& A 392, 395

Cimatti A, Daddi E, Renzini A et al., 2004, *Nature* 430, 184

Cimatti A, Daddi E, Renzini A, 2006, A& A 453, L29

Cimatti A, Cassata P, Pozzetti L, Kurk J, Mignoli M, Renzini A, Daddi E et al., 2008, A&A 482, 21

Cole S et al., 2001, MNRAS 326, 255

Conroy C, Gunn J E, White M, 2009, ApJ 699, 486

Conselice C J, 2003, ApJSS 147, 1

Conselice C J, 2006, MNRAS 373, 1389

Csabai I, Connolly A, Szalay A S, Budavári T, 2000, AJ 119, 69

Colless M, Dalton G, Maddox S, Sutherland W, Norberg P et al., 2001, MNRAS 328, 1039

Colless M, Peterson B A, Jackson C, Peacock J A, Cole S, Norberg P et al., 2003, arXiv:astro-ph/0306581

Conroy C, Gunn J E, White M, 2009, ApJ 699, 486

Connolly A J, Csabai I, Szalay A S, Koo D C, Kron R G, Munn J A, 1995, AJ 110, 2655

Connolly A J, Szalay A S, Dickinson M,Subbarao M U, Brunner R J, 1997, ApJ 486, L11

Cowie L L, Songaila A, Hu E M, Cohen J G, 1996, AJ 112, 839

Croft R A C, Weinberg D H, Bolte M, Burles S, Hernquist L, Katz N, Kirkman D, Tytler D, 2002, ApJ 581, 20

Croton D J, Springel V, White S D M, De Lucia G, Frenk C S, Gao L, Jenkins A, Kauffmann G, Navarro J F, Yoshida N, 2006, MNRAS 365, 11

Daddi E, Renzini A, Pirzkal N, Cimatti A, Malhotra S, Stiavelli M, Xu C, Pasquali A, Rhoads J E et al., 2004, ApJ 617, 746

Daddi E, Renzini A, Cimati A, Malhotra S, Stiavelli M, Xu C, Pasquali A, Rhoads J E, Brusa M, di Serego Alighieri S, Ferguson H C, Koekemoer A, Moustakas L A, Panagia N, Windhorst R A, 2005, ApJ 626, 680

Damen M, Labbé I, Franx M, van Dokkum P G, Taylor E N, Gawiser E, 2009, ApJ 690, 937

Damen M et al., 2009, ApJ (submitted)

Damjanov I, McCarthy P J, Abraham R G, Glazebrook K, Yan H et al., 2009, ApJ 695, 101

Davis M, Faber S M, Newman J, Phillips A C, Ellis R S, Steidel C C et al., 2003, Proc. SPIE 4834, 161

Dekel A, Birnboim Y, 2006, MNRAS 368, 2

De Lucia G, Springel V, White S D M, Croton D, Kauffmann G, 2006, MNRAS 366, 2, 499

De Lucia G, Poggianti B et al., 2007, MNRAS 374, 809

di Serego Alighieri S, Vernet J, Cimatti A, Lanzoni B et al., 2005, A&A 442, 125

Djorgovsky S & Davis M, 1987, ApJ 313, 59

Dressler A, 1980, ApJ 313, 42

Dressler A, Lynden-Bell D, Burstein D, Davies R L, Faber S M, Terlevich R, Wegner G, 1987, ApJ 313, 42

Driver S P, Allen P D, Graham A W, Cameron E, Liske J, Ellis S C, Cross N J G, De Propris R, Phillipps S, Couch W J, 2006, MNRAS 368, 414

Driver S P, Norberg P, Baldry I K, Bamford S P, Hopkins A M, Liske J, Loveday J, Peacock J A, 2009, A&G 50, 5.12

Drory N, Salvato M, Gabasch A, Bender R, Hopp U, Feulner G, Pannella M, 2005, ApJ 619, L131

Dunkley J, Komatsu E, Nolta M R, Spergel D N, Larson D, Hinshaw G, Page L et al., 2009, ApJS 180, 306

Ellis R S, Colless M, Broadhurst T, Heyl J, Glazebrook K, 1996, MNRAS 280, 235

Ellis S C, Driver S P, Allen S D, Liske J, Bland-Hawthorn J, De Propris R, 2005, MNRAS 363, 1257

Faber S M & Jackson R E, 1976, ApJ 204, 668

Faber S M et al., 2007, ApJ 665, 265

Fan L, Lapi A, De Zotti G, Danese L, 2009, ApJ 689, L101

Fernández-Soto A, Lanzetta K M, Chen H-W, Pascarelle S M, Yahata N, 2001, ApJS 135, 41

Firth A E, Lahov O, Somerville R S, 2003, MNRAS 339, 1195

Fontana A et al., 2004, A&A 424, 23

Fontana A et al., 2006, A&A 459, 745

Förster-Schreiber N M, van Dokkum P G, Franx M, Labbé I, Rudnick G et al., 2004, ApJ 616, 41

Förster-Schreiber N M, Franx M, Labbé I, Rudnick G, van Dokkum P G et al., 2006,

AJ 131, 1891
Franx M, Labbé I, Rudnick G, van Dokkum P G et al., 2003, ApJ 587, L79
Franx M, van Dokkum P G, Fórster-Schreiber N M, Wuyts S, Labbé I, Toft S, 2008, ApJ 688, 770
Freedman W L, Madore B F, Gibson B K, Ferrarese L, Kelson D D, Sakai S, Mould J R, Kennicutt R C et al., 2001, ApJ 553, 47
Gallazzi A, Charlot S, Brinchmann J, White S D M, 2006, MNRAS 370, 1106
Gallazzi A & Bell E F, 2009, ApJS (accepted; arXiv:0910.1591)
Gawiser E et al., 2006a, ApJS 162, 1
Giavalisco M, Steidel C C, Macchetto F D, 1996, 470, 189
Giavalisco M, Steidel C C, Adelberger K L, Dickinson M E, Pettini M, Kellog M, 1998, ApJ 503, 543
Giavalisco M & Dickinson M, 2001, ApJ 550, 177
Giavalisco M, 2002, ARA&A 40, 579
Giovanelli R & Haynes M R, 1991, ARA&A 29, 499
Gladders M D & Yee H K C, 2000, AJ 120, 2148
Gladders M D & Yee H K C, 2005, ApJSS 157, 1
Graves G J, Faber S M, Schiavon R P, 2009, ApJ 698, 1950
Glazebrook K, Abraham R G, McCarthy P J et al, 2004, *Nature* 430, 181
Glazebrook K, 2009, *Nature* 460, 7256
Gwyn S D J & Hartwick F D A, 1996, ApJ 468, L77
Heavens A, Jiminez R & Lahov O, 2000, MNRAS 317, 965
Hildebrandt H et al., 2006, A&A 452, 1121
Hogg D W, Neugebauer G, Armus L, Matthews K, Pahre M A, Soier B T, Wienberger A J, 1997, AJ 113, 474
Hogg D W, Cohen J G, Blandford R, Gwyn S D J, Hartwick F D A, Mobasher B, Mazzei P, Sawicki M, Lin H, Yee H K C et al., 1999, AJ, 115, 1418
Holden B P, Illingworth G D, Franx M et al., 2008, ApJ 670, 190
Hopkins A M, 2004, ApJ 615, 209
Hopkins A M & Beacom J F, 2006, ApJ 651, 142
Hopkins P F, Bundy K, Murray N, Quataert E, Lauer T, Ma C-P, 2009, MNRAS 398, 898
Hubble E P, 1926, ApJ 64, 321
Huertas-Company M, Rouan D, Tasca L, Soucail G, Le Fèvre O, 2008, A&A 478, 971
Im M et al., 2002, ApJ 571, 136
Juneau S et al., 2005, ApJ 619, L135
Kannappan S J & Gawiser E, 2007, ApJ 657, L5
Kauffmann G, Heckman T M, White S D M et al., 2003, MNRAS 341, 33
Kauffmann G, Heckman T M, White S D M et al., 2003, MNRAS 341, 54
Kauffmann G, Heckman T M, De Lucia G, Brinchmann J, Charlot S, Tremonti C, White S D M, Brinkmann J, 2006, MNRAS 367, 1394
Kelly B C, McKay T A, 2005, AJ 129, 1287
Kriek M, van Dokkum P G, Franx M, Quadri R, Gawiser E, Herrera D Illingworth G D, Labbé I, Lira P, Marchesini D, Rix H-W, Rudnick G, Taylor E N, Urry M C, Wuyts S, 2008, ApJL 649, 71
Kriek M, van Dokkum P G, Franx M, Illingworth G D, Marchesini D, Quadri R, Rudnick G, Taylor E N, Förster-Schreiber N M, Gawiser E, Labbé I, Lira P, Wuyts S, 2008, ApJ 677, 219
Kriek M, van der Wel A, van Dokkum P G, Franx M, Illingworth G D, 2008, ApJ 682, 896
Kroupa P, 2002, in *Modes of Star Formation and the Origin of Field Populations*(Grebel & Brandner, ed.s), ASP Conf. Proc., 285, 86
Komatsu E, Dunkley J, Nolta M R et al., 2009, ApJS 180, 330
Koo D C, 1985, AJ 90, 418
Koo D C & Kron R J, 1992, ARA&A 30, 613
Kuhn T S, *The Structure of Scientific Revolutions*, 1962 (University of Chicago Press, Chicago; 3rd ed., 1996)
Labbé I, Franx M, Rudnick G, Förster-Schreiber N M et al., 2003, AJ 125, 1107
Labbé I, Huanf G, Franx M, Rudnick G et al., 2005, ApJ 624, L81
Le Borgne D & Rocca-Volmerange B, 2002, A&A 386, 446
Le Fèvre O, Vettolani G, Garilli B, Tresse L et al., 2005, A&A 439, 845
Lilly S, Le Fèvre O, Hammer F, Crampton D, 1996, ApJ 460, L1
Loh E D & Spillar E J, 1986, ApJ 303, 154
Longhetti M, Saracco P, Severgnini P, Della Caca R, Mannucci F, Bender R, Drory N, Feulner G, Hopp U, 2007, MNRAS 374, 614
Lotz J, Primack J, Madau P, 2004, AJ 128, 163
Madau P, Ferguson H C, Dickinson M E, Giavalisco M, Steidel C C, Fruchter A, 1996, MNRAS 283, 1388
Magorrian J, Tremaine S, Richstone D et al., 1998, AJ 115, 2285
Maraston C, 2005, MNRAS 362, 799
Maraston C Daddi E, Renzini A, Cimatti A, Dickinson M, Papovich C, Pasquali A, Pirzkal N, 2006, ApJ 652, 85
Marchesini D, van Dokkum P G, Förster-Schreiber N M, Franx M, Labbé I, Wuyts S, 2009, ApJ 701, 1765
McCarthy P J, 2004, ARA&A 42, 477
McCarthy P J, Le Borgne D, Crampton D, Cen H-W et al., 2004, ApJ 614, L9
Menci N, Fontana A, Giallongo E, Salimbeni S, 2005, ApJ 632, 49
Menci N, Fontana A, Giallongo E, Grazian A, Salimbeni S, 2006, ApJ 647, 753

Minkowski R, 1962, in *Problems of Extragalactic Research* (IAU Symp. 15), G C McVittie, ed. (Macmillan, New York)

Naab T, Johansson P H, Ostriker J H, Efstathiou G, 2007, ApJ 658, 710

Nagamine K, Ostriker J P, Fukugita M, Cen R, 2006, ApJ 653, 881

Noeske K G, Weiner B J, Faber S M, Papovich C, Koo D C, Somerville R S et al., 2007, 660, L43

Palumbo G G C, Tanzella-Nitti G & Vettolani G, 1983, *Catalogue of radial velocities of galaxies* (Gordon and Breach Science Publishers; New York)

Panter B, Heavens A F, Jimenez R, 2003, MNRAS 343, 1145

Panter B, Jimenez R, Heavens A F, Charlot S, 2007, MNRAS 378, 1550

Papovich C, Dickinson M, Ferguson H C, 2001, ApJ 559, 620

Pasquali A, Mo H J, Yang X, Somerville R, 2009, MNRAS 394, 38

Perlmutter S et al., 1999, ApJ 517, 565

Pérez-Gonzaléz P G et al., 2008, ApJ 675, 234

Pettini M, Shapley A E, Steidel C C et al., 2001, ApJ 554, 981

Pozzetti L et al., 2007, A&A, 474, 447

Pozzetti L et al., 2009, A&A (submitted; arXiv:0907.5416)

Quadri R & Williams R J, 2009, ApJ (submitted; arXiv:0910.2704)

Reddy N, Erb D K, Steidel C C, Shapley A E, Adelberger K L, Pettini M, ApJ 633, 748

Riess A G, Fillipenko A V et al., 1998, AJ 116, 1009

Rix H-W & Rieke M J, 1993, ApJ 418, 123

Robaina A R, Bell E F, Skelton R E, McIntosh D H, Somerville R S, Zheng X, Rix H-W, 2009, ApJ 704, 324

Rudnick G, Franx M, Rix W-H, Moorwood A, Kuijken K, van Starkenburg L, van der Werf P, Röttgering H, van Dokkum P G, Labbé I, 2001, AJ 122, 2205

Rudnick G, Rix H-W, Franx M, Labbé I et al., 2003, ApJ 599, 847

Ruhland C, Bell E F, Häußler B, Taylor E N, Barden M, McIntosh D, 2009, ApJ 1058, 1069

Salpeter E E, 1955, ApJ 121, 161

Sandage A & Visvanathan N, 1978, AJ 225, 742

Saracco P, Longhetti M, Andreon S, 2009, MNRAS 392, 718

Sawicki M J, Lin H & Yee H Y K, 1997, AJ 113, 1

Sawicki M J, & Yee H Y K, 1998, AJ 115, 1329

Schaye J & Dalla Vecchia C, 2008, MNRAS 383, 1210

Schaye J, Dalla Vecchia C, Booth C M, Wiersma R P C, Theuns T, Haas M R et al, 2009, MNRAS (submitted; arXiv:0909.5196)

Shapley A E, Steidel C C, Adelberger K L, Dickinson M, Giavalisco M, Pettini M, 2001, ApJ 562, 95

Shectman S A, Landy S D, Oemler A, Tucker D L, Lin H, Kirshner R P, Schecter P L, 1996, ApJ 470, 172

Shen S, Mo H J, White S D M, Blanton M R, Kauffmann G, Voges W, Brinkmann J, Csabai J, 2003, MNRAS 343, 978

Somerville R S, Hopkins P F, Cox T J, Robertson B E, Hernquist L, 2008, MNRAS 391, 481

Spergel D N, Verde L, Peiris H V, Komatsu E, Nolta M R et al., 2003, ApJS 148, 175

Spergel D N, Bean R, Doré O, Nolta M R et al., 2007, ApJS 170, 377

Springel V, Frenk C S and White S D M, 2006, *Nature* 440, 1137

Stebbins J & Whitford A E, 1948, AJ 53, 204

Steidel C C & Hamilton D, 1993, AJ 105, 2017

Steidel C C, Pettini M & Hamilton D, 1995, AJ 110, 2519

Steidel C C, Adelberger K L, Giavalisco M, Dickinson M, Pettini M, 1996, ApJ 462, L17

Steidel C C, Giavalisco M, Pettini M, Dickinson M, Adelberger K L, 1999, ApJ 519, 1

Steidel C C, Adelberger K L, Shapley A E, Pettini M, Dickinson M, Giavalisco M, 2003, ApJ 592, 728

Strateva I et al., 2001, AJ, 122, 1861

Strauss M A, Weinberg S H, Lupton R H, Narayanan V K et al., 2002, AJ, 124, 1810

Subbaroo M U, Connolly A J, Szalay A S, Koo D C, 1996, AJ 112, 929

Tanaka M, Kodama M, Arimoto N, Okamura S, Umetsu K, Simasaku K, Tanaka I, Yamada T, 2005, MNRAS 363, 268

Tinsley B M & Gunn J E, 1976, ApJ 203, 52

Tinsley B M, 1978, ApJ 222, 14

Toft S, van Dokkum P G, Franx M, et al., 2007, ApJ 671, 285

Tresse L et al., A&A 472, 403

Treu T, Stiavelli M, Casertano S, Møller P, Bertin G, 2002, ApJ 564, L13

Trujillo I, Fórster-Schreiber N M, Rudnick G, et al., 2006, ApJ 650, 18

Trujillo I, Conselice C, Bundy K, Cooper M C, Eisenhardt P, Ellis R C, 2007, MNRAS 382, 109

Tully R B & Fisher J R, 1977, A&A 54, 66

van den Bosch F C, Aquino D, Yang X, Pasquali A, McIntosh D H, Weinmann S M, Kang X, 2008, MNRAS 387, 79

van den Bosch F C, Yang X, Mo H J, Weinmann S, McIntosh D H, Aqunio D, 2009, MNRAS (submitted; arXiv:0805.0002)

van der Wel A, Franx M, van Dokkum P G, Rix H-W, 2004, ApJ 601, L5

van der Wel A, Franx M, Wuyts S, Huang J, Rix H-W, Illingworth G, 2006, ApJ 652, 97

van der Wel A, Holden B P, Franx M et al., 2007, ApJ 670, 206

van der Wel A, 2008, ApJ 675, L13

van der Wel A, Bell E F, van den Bosch F C,

Gallazzi A, Rix H-W, 2009, ApJ 698, 1232

van Dokkum P G, Förster-Schreiber N M, Franx M et al., 2003, ApJ 587, L83

van Dokkum P G, Franx M, Förster-Schreiber N M, Illingworth G et al., 2004, ApJ 611, 703

van Dokkum P G et al., 2006, ApJ 638, 59

van Dokkum P G, 2007, ApJ 674, 29

van Dokkum P G, Labbé I, Marchesini D, Quadri R, Brammer G et al., 2009, PASP 121, 2

van Dokkum P G, Kriek M, Franx M, 2009, *Nature* 460, 717

Vanzella E, Cristiani S, Fontana A, Nonino M et al., 2004, A&A 423 761

Verde L, Heavens A F, Percival W J, Matarrese S et al., 2002, MNRAS 335, 432

Verde L, Peiris H V, Spergel D N, Nolta M R et al., 2003, ApJS 148, 195

Vogt N P, Koo D C, Phillips A C, Wu K, Faber S M, Willmer C N A, Simard L, Weiner B J et al., 2005, ApJS 159, 41

Weiner B J, Phillips A C, Faber S M, Willmer C N A, Vogt N P, Simard L, Gebhardt K, Im M, Koo D C, Sarajedini V L, Wu K L, 2005, ApJ 620, 595

Wiersma R P C, Schaye J, Smith B D, 2009, MNRAS 393, 99

Wild V, Walcher C J, Johansson H, Tresse L et al., 2009, MNRAS 395, 144

Wilkins S M, Trentham N, Hopkins A M, 2008, MNRAS 385, 687

Williams R E et al., 1996, AJ 112, 1335

Williams R J, Quadri R F, Franx M, van Dokkum P G, Labbé I, 2009, ApJ 691, 1879

Willmer C N A et al., 2006, ApJ, 647, 853

Wolf C, Meisenheimer K, Rix H-W, Borch A, Dye S, Kleinheinrich M, 2003, A&A 401, 73

Wolf C, Meisenheimer K, Kleinheinrich M et al., 2004, A&A 421, 913

Wuyts S, Labbé I, Franx M, Rudnick G, van Dokkum P G et al., 2007, ApJ 655, 51

Wuyts S, Labbé I, Förster Schreiber N M, Franx M, Rudnick G, Brammer G B, van Dokkum P G, 2008, ApJ, 682, 985

Wuyts S, Franx M, Cox T J, Hernquist L, Hopkins P F, Robertson B E, van Dokkum P G, 2009, ApJ 696, 348

Wyder T K et al., 2007, ApJSS 173, 293

Wolf C, Hildebrandt H, Taylor E N, Meisenheimer K, 2008, A&A 942, 933

York D G et al., 2000, AJ 120, 2131

Zheng X Z, Bell E F, Papovich C, Wolf C, Meisenheimer K, Rix H-W, Rieke G H, Somerville R, 2007, ApJ 661, 41

Zirm A W, van der Wel A, Franx M et al., 2007, ApJ 656, 66

# Chapter II

## A Public, $K$-Selected, Optical–to–Near-Infrared Catalog of the Extended Chandra Deep Field South from the Multiwavelength Survey by Yale–Chile

We present a new, $K$-selected, optical–to–near infrared photometric catalog of the Extended Chandra Deep Field South (ECDFS), making it publicly available to the astronomical community. The dataset is founded on publicly available imaging, supplemented by original $z'JK$ imaging data collected as part of the MUltiwavelength Survey by Yale–Chile (MUSYC). The final photometric catalog consists of photometry derived from $UU_{38}BVRIz'JK$ imaging covering the full $\frac{1}{2} \times \frac{1}{2}$ $\square^\circ$ of the ECDFS, plus $H$ band photometry for approximately 80 % of the field. The $5\sigma$ flux limit for point sources is $K_{\rm tot}^{\rm (AB)} = 22.0$. This is also the nominal completeness and reliability limit of the catalog: the empirical completeness for $21.75 < K < 22.00$ is $\gtrsim 85$ %. We have verified the quality of the catalog through both internal consistency checks, and through external comparisons to other existing and publicly available catalogs. As well as the photometric catalog, we also present catalogs of photometric redshifts and restframe photometry derived from the ten-band photometry. We have collected robust spectroscopic redshift determinations from published sources for 1966 galaxies in the catalog. Based on these sources, we have achieved a $(1\sigma)$ photometric redshift accuracy of $\Delta z/(1+z) = 0.036$, with an outlier fraction of 7.8 %; most of these outliers are X-ray sources. Finally, we describe and release a utility for interpolating restframe photometry from observed spectral energy distributions (SEDs), dubbed InterRest. Particularly in concert with the wealth of already publicly available data in the ECDFS, this new MUSYC catalog provides an excellent resource for studying the changing properties of the massive galaxy population at $z \lesssim 2$.

Taylor E N, Franx M, van Dokkum P G, Quadri R F, Gawiser E, Bell E F, Barrientos L F, Blanc G A, Castander F J, Damen M, Gonzalez-Perez V, Hall P B, Herrera D, Hildebrandt H, Kriek M, Labbé I, Lira P, Maza J, Rudnick G, Triester E, Urry C M, Willis J P, Wuyts S,
*The Astrophysical Journal Supplement Series*, **183**, 295–319 (2009)
(submitted September 2008; published August 2009)

## 1 Introduction

Over the past decade, multiband deep field imaging surveys have provided new opportunities to directly observe the changing properties of the general (or 'field') galaxy population with lookback time. By quantifying the star formation, stellar mass, and morphological evolution among galaxies, these new datasets have led to new and fundamental insights into the physical processes that govern the formation and evolution of galaxies. These advances have been made possible not only by the advent of a new generation of space-based and 8 m class telescopes, but also the maturation of techniques for estimating redshifts and intrinsic properties like stellar masses from observed spectral energy distributions (SEDs). These two developments have made it possible not only to go deeper — pushing to higher redshifts and probing further down the luminosity function — but also to consider many more galaxies per unit observing time. This has made possible the construction of large, representative, and statistically significant samples of galaxies spanning a large proportion of cosmic time.

The Chandra Deep Field South (CDFS; Giacconi et al., 2002) is one of the premier sites for deep field cosmological surveys (see **Figure 1)**. It is one of the most intensely studied regions of the sky, with observations stretching from the X-ray to the radio, including ultraviolet, optical, infrared, and submillimeter imaging, from space-based as well as the largest terrestrial observatories. It has also become traditional for surveys targeting the CDFS to make their data publicly available. As a direct result of this commitment to collaboration within the astronomical community, the wealth of data available — in terms of both volume and quality — provide an exceptional opportunity to quantify the evolution of the galaxy population out to high redshift.

With this goal in mind, the key to gaining access to the $z \gtrsim 1$ universe is near infrared (NIR) data. Most of the broad spectral features (e.g. the Balmer and 4000 Å breaks) on which modern SED-fitting algorithms rely are in the restframe optical; for $z \gtrsim 1$, these features are redshifted beyond the observer's optical window and into the NIR. For this reason, we have combined existing imaging of the Extended Chandra Deep Field South (ECDFS; see Figure 1) with new optical and NIR data taken as part of the MUltiwavelength Survey by Yale–Chile (MUSYC).

The primary objective of MUSYC is to obtain deep optical imaging and spectroscopy of four $\frac{1}{2} \times \frac{1}{2}$ $\square^\circ$ southern fields, providing parent catalogs for followup with ALMA. Coupled with the optical ($UBVRIz'$) imaging program (Gawiser et al., 2006), there are two NIR components to the MUSYC project: a deep component ($K < 23.5$; Quadri et al., 2007), targeting four $10 \times 10$ $\square''$ regions within the MUSYC fields, and a wide component ($K < 22$; Blanc et al., 2008, this Chapter) covering three of the $\frac{1}{2} \times \frac{1}{2}$ $\square^\circ$ MUSYC fields in their entirety. These data are intended to allow, for example, the restframe-UV selection of galaxies at $z \gtrsim 3$ using the Lyman break technique (e.g., Steidel et al., 1996), the restframe-optical selection of galaxies at $z \gtrsim 2$ using the Distant Red Galaxy (DRG) criterion (Franx et al., 2003), and the color-selection of $z \gtrsim 1.4$ galaxies using the $BzK$ criterion (Daddi et al., 2004).

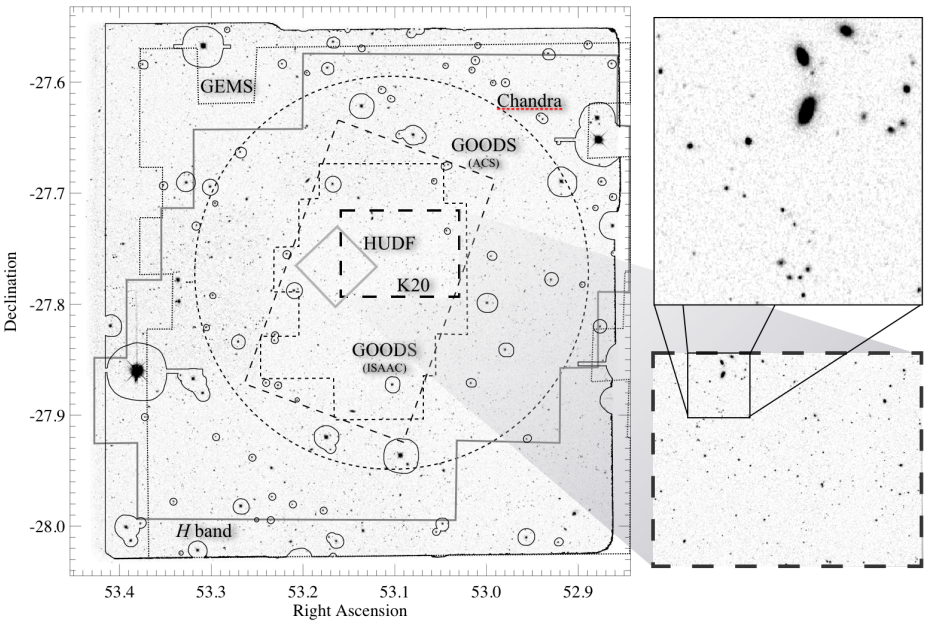


**Figure 1.** — MUSYC in the ECDFS. — The greyscale image shows the new $K$ band data. The solid black contour shows the area with useful photometry in all of $UU_{38}BVRIz'JK$ in the MUSYC catalog. (Areas badly affected by bright stars in the $z'$ band have been masked.) The catalog also includes $H$ band photometry for $\sim 80\%$ of the field (*solid grey contour*). For comparison, we also show the area covered by several other important (E)CDFS surveys (in order of field size, from largest to smallest): GEMS (*dotted lines*; Rix et al., 2004), the original Chandra CDFS (*short-dashed circle*; Giacconi et al., 2002), the GOODS (Dickinson et al., 2002) HST ACS optical (*light long-dashed rectangle*) and ISAAC NIR (*short-dashed region*) imaging, the K20 survey (*heavy long-dashed rectangle*; Cimatti et al., 2002), and the HUDF (*grey solid diamond*; Beckwith et al., 2006). The FIREWORKS catalog (Wuyts et al., 2008) combines the GOODS ACS and ISAAC data with the $UU_{38}BRVIz'$ data described in this Chapter for the central GOODS ISAAC region. SIMPLE (Damen et al., 2009) will add very deep *Spitzer Space Telescope* IRAC imaging to the whole region shown here. A medium band NIR survey is also underway using the NEWFIRM instrument (van Dokkum et al., 2009). At right, we show a detail of the K20 survey area (below), and futher detail of an approximately $2'' \times 2''$ area (above).

In the ECDFS, the broadband imaging data have been supplemented by a narrow-band imaging survey, targeting Ly-$\alpha$ emitters at $z = 3.1$ (Gawiser et al., 2006b; Gronwall et al., 2007), and a spectroscopic survey (Treister et al., 2008) targeting X-ray sources from the 250 ks ECDFS X-ray catalog (Lehmer et al., 2005; Virani et al., 2006). Further, the Spitzer IRAC/MUSYC Public Legacy in the ECDFS (SIMPLE; M Damen et al., 2009) project has obtained very deep IRAC imaging across the full ECDFS. There is also a deep medium band optical survey underway (Cardamone et al., in preparation), and a planned medium band NIR survey (van Dokkum et al., 2009).

This Chapter describes the MUSYC wide NIR-selected catalog of the ECDFS (which we will from now on refer to as 'the' MUSYC ECDFS catalog, despite the

existence of several separate MUSYC catalogs, as described above), and makes it publicly available to the astronomical community. A primary scientific goal of the wide NIR component of the survey is to obtain statistically significant samples of massive galaxies at $z \lesssim 2$. In Chapter III, we will use this dataset to quantify the $z \lesssim 2$ color and number density evolution of massive galaxies in general, and in the relative number of red sequence galaxies in particular.

The MUSYC ECDFS dataset is founded on existing and publicly available imaging, supplemented by original optical ($z'$) and NIR ($JK$) imaging. Apart from the $JK$ imaging, all of these data have been described elsewhere. Accordingly, the data reduction and calibration of the new $JK$ imaging is a prime focus of this Chapter. However, when it comes to constructing panchromatic catalogs with legacy value from existing datasets, the whole is truly more than the sum of parts: ensuring both absolute and relative calibration accuracy is paramount. We have invested substantial time and effort into checking all aspects of our data and catalog, using both simulated datasets, and through comparison to some of the many other existing (E)CDFS catalogs.

The structure of this Chapter is as follows: we describe the acquisition and basic reduction of the MUSYC ECDFS broadband imaging dataset in Section 2. The processes used to combine these data into a mutually consistent whole are described in Section 3. In Section 4, we describe the construction of the photometric catalog itself, including checks on the completeness and reliability, and on our ability to recover total fluxes. We present external checks on the astrometric and photometric calibration in Section 5. After a simple comparison of our catalog to other NIR-selected catalogs in Section 6, we describe our basic analysis of the multiband photometry in Section 7, including star/galaxy separation, and the derivation of photometric redshifts, as well as the tests we have performed to validate our analysis. In Section 8, we introduce InterRest; a new utility for interpolating restframe fluxes. This utility is also being made public. Additionally, in Appendix A, we describe a compilation of 2213 robust spectroscopic redshift determinations for objects in the MUSYC ECDFS catalog.

Throughout this work, all magnitudes are expressed in the AB system; the only exception to this is Section 5.2, where it will be convenient to adopt the Vega system. Where necessary, we assume the concordance cosmology; *viz.* $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_0 = 1.0$, and $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$. When discussing photometric redshifts, we will characterise random errors in terms of the NMAD[1] of $\Delta z/(1+z)$; we will abbreviate this quantity using the symbol $\sigma_z$.

## 2 Data

This section describes the acquisition of the imaging data comprising the MUSYC ECDFS dataset; the vital statistics of these data are given in **Table 2**. Of these data, only the $z'JK$ are original; the WFI $UU_{38}BVRI$ imaging has been reduced

[1]Here, NMAD is an abbreviation for the normalized median absolute deviation, and is defined as $1.48 \times \mathrm{med}[x - \mathrm{med}(x)]$; the normalization factor of 1.48 ensures that the NMAD of a Gaussian distribution is equal to its standard deviation.

| (1) Band | (2) $\lambda_0$ [Å] | (3) $\Delta\lambda$ [Å] | (4) $m_{\rm Vega}^{\rm (AB)}$ | (5) Int. Time [hr] | (6) Area [$\square''$] | (7) Eff. Seeing | (8) Noise Limit ($5\sigma$) | (9) $N_{\rm cov}$ | (10) $f_{5\sigma}$ | (11) $N_{\rm main}$ | (12) $N_{\rm gals}$ | (13) $N_{\rm stars}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $U$ | 3505 | 625 | +1.01 | 21.91 | 975 | $1\farcs07$ | 26.5 | 15136 | 0.631 | 6213 | 6424 | 576 |
| $U_{38}$ | 3655 | 360 | +0.82 | 13.75 | 947 | $1\farcs01$ | 26.0 | 14280 | 0.554 | 5505 | 5715 | 504 |
| $B$ | 4605 | 915 | −0.12 | 19.29 | 1012 | $1\farcs03$ | 26.9 | 15153 | 0.852 | 8223 | 8322 | 880 |
| $V$ | 5383 | 895 | −0.01 | 29.06 | 1022 | $0\farcs95$ | 26.6 | 15154 | 0.863 | 8370 | 8463 | 891 |
| $R$ | 6520 | 1600 | +0.19 | 24.35 | 1017 | $0\farcs88$ | 26.3 | 15148 | 0.894 | 8647 | 8758 | 897 |
| $I$ | 8642 | 1500 | +0.51 | 9.60 | 977 | $0\farcs98$ | 24.8 | 15128 | 0.826 | 8456 | 8545 | 897 |
| $z'$ | 9035 | 995 | +0.54 | 1.30 | 996 | $1\farcs13$ | 24.0 | 13972 | 0.751 | 8043 | 8000 | 897 |
| $J$ | 12461 | 1620 | +0.93 | 1.33 | 906 | $\leq 1\farcs49$ | 23.1 | 14580 | 0.683 | 7894 | 7859 | 896 |
| $H$ | 16534 | 2960 | +1.40 | 1.00 | 560 | $\leq 1\farcs22$ | 23.1 | 10518 | 0.579 | 7005 | 6313 | 692 |
| $K$ | 21323 | 3310 | +1.83 | 1.00 | 906 | $\leq 1\farcs05$ | 22.4 | 14355 | 0.695 | 8782 | 8911 | 897 |

**Table 1.** — Summary of the data comprising the MUSYC ECDFS catalog. — For each band (Col. 1) that has gone into the MUSYC ECDFS catalog, we give the effective wavelength (Col. 2), the filter FWHM (Col. 3), and the apparent magnitude of Vega, in the AB system (*i.e.* the conversion factor between the AB and Vega magnitude systems, Col. 4). We also give the mean integration time (Col. 5) for each image, the effective imaging area (Col. 6, defined as the region receiving more than 75 % of the nominal integration time ), and the final effective seeing (FWHM, Col. 7). The $5\sigma$ limiting depths given in Col. (8) are as measured in $2\farcs5$ diameter apertures on the $1\farcs5$ FWHM PSF-matched images (see Section 3.2); for a point source, these can be translated to total magnitudes by subtracting 0.45 mag. Note that, whereas the optical data are taken in single pointings, the final NIR images are mosaics of many pointings. Note that the central $\sim 10 \times 10$ $\square''$ of the field received an extra three hr integration time in the $H$ band; these data are approximately 0.3 mag deeper than the figure quoted above. Col. (9) gives the number of $K$ detections that have useful coverage (*i.e.* an effective weight, $w$, of 0.6 or greater) in each band ; Col. (10) gives the fraction of those objects that have $> 5\sigma$ detections. Both of these columns refer to the full catalog. Col. (11) gives the number of objects in the main science sample ($K_{\rm tot} < 22$, $K$ S:N $> 5$, $w_B > 0.6$, $w_{z'} > 0.6$, $w_K > 0.75$); Col.s (12) and (13) give the numbers of stars and galaxies separately (see Section 7.1).

and described by Hildebrandt et al. (2006), and the SofI $H$ band data by Moy et al. (2003). Further, the original $z'$ data have been reduced as per Gawiser et al. (2006) for the MUSYC optical ($BVR$-selected) catalog. We have therefore split this section between a summary of the data that are described elsewhere (Section 2.1), and a description of the new ISPI $JK$ imaging (Section 2.2). Note that what we refer to as the $K$ band is really a '$K$ short' filter; we have dropped the subscript for convenience. For a complete description of the other datasets, the reader is referred to the works cited above.

## 2.1 Previously Described Data

### *2.1.1 The WFI Data — $UU_{38}BVRI$ Imaging from the ESO Archive*

Hildebrandt et al. (2006) have collected all (up until December 2005) archival $UU_{38}BRVI$[2] imaging data taken using the Wide Field Imager (WFI, $0\farcs238$ pix$^{-1}$; Baade et al., 1998, 1999) on the ESO MPG 2.2 m telescope for the four fields that make up the ESO Deep Public Survey (DPS; Arnouts et al., 2001). In addition the original DPS ECDFS data (DPS field $2c$), this combined dataset includes WFI commissioning data, the broadband data from the COMBO-17 survey (Wolf et al., 2004), and observations from seven other observing programs. Hildebrandt et al. (2006) have pooled and rereduced these data using the automated THELI pipeline described by Erben et al. (2005) under the moniker GaBoDS (Garching Bonn Deep Survey). The final products are publicly available through the ESO Science Archive Facility.[3] The final image quality of these images is $0\farcs9$–$1\farcs1$ (FWHM). Hildebrandt et al. (2006) estimate that their basic calibration is accurate to better than $\sim 0.05$ mag in absolute terms, and that, based on color–color diagrams for stars, the relative or cross-calibration between bands is accurate to $\lesssim 0.1$ mag for all images.

### *2.1.2 The Mosaic-II data — Original $z'$ Imaging*

We have supplemented the WFI optical data with original $z'$ band imaging taken using the Mosaic-II camera ($0\farcs267$ pix$^{-1}$; Muller et al., 1998) on the CTIO 4m Blanco telescope. The data acquisition strategy is the same as for the optical data in other MUSYC fields (Gawiser et al., 2006); the ECDFS data were taken in January 2005. The final integration time was 78 min, with an effective seeing of $1\farcs1$ (FWHM), although we note that the point-spread function (PSF) does have broad, non-Gaussian 'wings'. The estimated uncertainty in the photometric calibration is $< 0.03$ mag (Gawiser et al., 2006).

[2]Two separate WFI $U$ filters have been used. The first, ESO#877, which we refer to as the $U$ filter, is slightly broader than a Broadhurst $U$ filter. This filter is known to have a red leak beyond 8000 Å. The second filter, ESO#841, which we refer to as $U_{38}$, is something like a narrow Johnson $U$ filter. There is, unfortunately no clear convention for how to refer to these filters; for instance, Arnouts et al. (2001) refer to what we call the $U$ and $U_{38}$ filters as $U'$ and $U$, respectively.

[3]http://archive.eso.org/cms/eso-data/data-packages/gabods-data-release-version-1.1-1/

### *2.1.3 The SofI Data — H Imaging Supporting the ESO DPS*

We include the $H$ band data described by Moy et al. (2003), which was taken to complement the original DPS WFI optical data and SofI NIR data (Vandame et al., 2001; Olsen et al., 2006). This dataset, consisting of 32 separate $4\farcm9 \times 4\farcm9$ pointings, covers approximately 80 % of the ECDFS , and were obtaining using SofI ($0\farcs288$ pix$^{-1}$; Moorwood et al., 1998) on the ESO NTT 3.6 m telescope. The data were taken as a series of dithered (or 'jittered') 1 min integrations, totaling 60 min per pointing; the central four fields received an extra 3 hs integration time. We received these data (Pauline Barmby, private communication) reduced as described by Moy et al. (2003); *i.e.*, as 32 separate, unmosaicked fields. The effective seeing in each pointing varies from $0\farcs4$ to $0\farcs8$ (FWHM). Moy et al. (2003) found that their photometric zeropoint solution varied by $\leq 0.04$ mag over the course of a night; they offer this as an upper limit on possible calibration errors. Further, in comparison to the Los Campanas Infrared Survey (LCIRS; Chen et al., 2002), and the v0.5 (April 2002) release of the GOODS ISAAC photometry, Moy et al. (2003) found their calibration to be 0.065 mag brighter, and 0.014 mag fainter, respectively.

## 2.2 The ISPI Data — Original $JK$ Imaging

The new MUSYC NIR imaging consists of two mosaics in the $J$ and $K$ bands, each made up of $3 \times 3$ pointings, and covering approximately 950 $\Box'$. The data were obtained using the Infrared SidePort Imager (ISPI; Probst et al. 2003; Van der Bliek et al. 2004) on the CTIO Blanco 4m telescope. ISPI uses a $2048 \times 2048$ pix HgCdTe HAWAII-2 detector, which covers approximately $10\farcm5 \times 10\farcm5$ at a resolution of $\approx 0\farcs3$ pix$^{-1}$. The aim was to obtain uniform $J$ and $K$ coverage of the full $\frac{1}{2} \times \frac{1}{2}$ $\Box^\circ$ of the ECDFS to $\sim$ 80 min and $\sim$ 60 min, respectively; our target ($5\sigma$, point source) limiting magnitudes were $J \approx 22.5$ and $K \approx 22$.

The data were taken over the course of 15 nights, in 4 separate observing runs between January 2003 and February 2004. In order to account for the bright and variable NIR sky ($\sim$ 10000 times brighter than a typical astronomical source of interest, and variable on many-minute timescales), the data were taken as a series of short, dithered integrations. A non-regular, semi-random dither pattern within a $45''$ box was used for all but three sub-fields; these three earliest pointings were dithered in regular, $\sim 10''$ steps. An integration of $4 \times 15$s (*i.e.*, 4 individual integrations of 15 sec, coadded) was taken at each dither position in $K$; in $J$, integrations were typically $1 \times 100$s.

Conditions varied considerably over the observing campaign, with seeing ranging from $\lesssim 0\farcs7$ to $\gtrsim 1\farcs5$ (FWHM). All nine $K$ band pointings were observed under good conditions ( $\lesssim 1\farcs0$ FWHM). However, observing condititions were particularly bad for two of the nine $J$ pointings; the final effective seeing of both the south and southwest pointings are nearer to $1\farcs5$ (FWHM).

For each of the subfields comprising the MUSYC ISPI coverage of the ECDFS, the data reduction pipeline is essentially the same as for the other MUSYC NIR imaging, as described by Quadri et al. (2007) and Blanc et al. (2008), following

the same basic strategy as, e.g., Labbé et al. (2003). The data reduction itself was performed using a modified version of the IRAF package `xdimsum`.[4]

### 2.2.1 Dark Current and Flat Field Correction

The ISPI detector has a non-negligible dark current. To account for this, nightly 'dark flats' were constructed by mean combining (typically) 10 to 20 dark integrations with the appropriate integration times; these 'dark flats' are then subtracted from each science image. These dark flats show consistent structure from night to night, but vary somewhat in their actual levels. Note that this correction is done before flat-fielding and/or sky subtraction (see also Blanc et al., 2008).

Flat field and gain/bias corrections (*i.e.*, spatial variations in detector sensitivity due to detector response, optic throughput, etc.) were done using dome-flats, which were constructed either nightly or bi-nightly. These flats were constructed by taking a number of integrations with or without a lamp lighting the dome screen. Each flatfield was constructed using approximately ten 'lamp on' and 'lamp off' images, mean combined. In order to remove background emission from the 'lamp on' image, we subtract away the 'lamp off' image, to leave only the light reflected by the dome screen (see also Quadri et al., 2007). These flats are very stable from night to night, with some variation between different observing runs.

### 2.2.2 Sky Subtraction and Image Combination

Because the NIR sky is bright, non-uniform, and variable, a separate sky or background image must be subtracted from each individual science image. The basic `xdimsum` package does this in a two-pass procedure. In the first pass, a background map is constructed for each individual science image by median combining a sequence of (typically) eight dithered but temporally continguous science integrations: typically the four science images taken immediately before and after the image in question. In the construction of this background image, a 'sigma clipping' algorithm is used to identify cosmic rays and/or bad pixels, which are then masked out. The resultant background image (which at this stage may be biased by the presence of any astronomical sources) is then subtracted from the science image to leave only astronomical signal. The sky subtracted images are then shifted to a common reference frame using the positions of stars to refine the geometric solution (undoing the dither) and then mean combined, again masking bad pixels/cosmic rays. This combined image is used to identify astronomical sources, using a simple thresholding algorithm. The entire process is repeated in a second 'mask pass', with the difference that this time astronomical sources are also masked when the background map is constructed.

Following Quadri et al. (2007), we have made several modifications to the basic `xdimsum` algorithm in order to improve the final image quality. We have

[4]IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation. The `xdimsum` package is available from http://iraf.noao.edu/iraf/ftp/iraf/extern-v212/xdimsum020806.

constructed an initial bad pixel mask using the flat-field images. Further, each individual science image is inspected by eye, and any 'problem' integrations (especially those showing telescope tracking problems or bad background subtraction) are discarded; artifacts such as satellite trails and reflected light from bright stars are masked by hand. These masks are used in both the first pass and mask pass.

Persistence is a problem for the ISPI detector: as a product of detector memory, 'echoes' of particularly bright objects linger for up to eight integrations. For this reason, we have also modified `xdimsum` to create separate masks for such artifacts; these masks are used in the mask pass. Note that for the three subfields (including the eastern $K$ pointing) observed using a regular, stepped dither pattern, this leads to holes in the coverage near bright objects: the 'echoes' fall repeatedly at certain positions relative to the source, corresponding to the regular steps of the dither pattern. At worst, coverage in these holes is $\sim 25\%$ of the nominal value.

Even after sky-subtraction, large-scale variations in the background were apparent; these patterns were different and distinct for each of the four quadrants of the images, corresponding to ISPI's four amplifiers. To remove these patterns, we have fit a 5th-order Legendre polynomial to each quadrant separately, using 'sigma clipping' to reduce the contribution of astronomical sources, and then simply subtracted this away (see also Blanc et al., 2008). This subtraction is done immediately after `xdimsum`'s normal sky-subtraction.

In the final image combination stage, we adopt a weighting scheme designed to optimize signal–to–noise for point sources (see, e.g., Gawiser et al., 2006; Quadri et al., 2007). At the end of this process, `xdimsum` outputs a combined science image. Additionally, `xdimsum` outputs an exposure or weight map, and a map of the rms in coadded pixels. Note that while this rms map is not accurate in an absolute sense, it does do an adequate job of mapping the spatial variation in the noise; see Section 4.6 below.

### *2.2.3 Additional Background Subtraction*

The sky subtraction done by `xdimsum` is imperfect; a number of large scale optical artifacts (particularly reflections from bright stars and 'holes' around very bright objects) remain in the $K$ images as output by `xdimsum`. Using these images, in the object detection/extraction phase (see Section 4 below), we were unable to find a combination of SExtractor background estimation parameters (*viz.* BACK_SIZE and BACK_FILTERSIZE) that was fine enough to map these and other variations in the background but still coarse enough to avoid being influenced by the biggest and brightest sources. This led to significant incompleteness where the background was low, and many spurious sources where it was high. We were therefore forced to perform our own background subtraction, above and beyond that done by `xdimsum`.

This basic idea was to use SExtractor 'segmentation maps' associated with the optical ($BVR$[5]) and NIR ($K$) detection images to mask real sources. Note

[5]Here, by $BVR$, we are referring to the combined $B+V+R$ optical stack used for detection in the construction of the MUSYC optically-selected catalog of the ECDFS (Gawiser et al., 2006b).

that the much deeper $BVR$ stack includes many faint sources lying below the $K$ detection limit. To avoid the contributions of low surface brightness galaxy 'wings', we convolved the combined ($BVR$+$K$) segmentation maps with a 15 pix ($4''$) boxcar filter to generate a 'clear sky' mask. Using this mask to block flux from astronomical sources, we convolved the science image with a 100 pix ($26\farcs7$) FWHM Gaussian kernel to generate a new background map; this was then subtracted from the `xdimsum`-generated science image.

Note that the background subtraction discussed above is important only in terms of object *detection*; background subtraction for *photometry* is discussed in Section 4.3 below. While this additional background subtraction step results in a considerably flatter background across the detection image, it does not significantly or systematically alter the measured fluxes of most individual sources.

#### *2.2.4 Photometric Calibration*

Because not all pointings were observed under photometric conditions, we have secondarily calibrated each NIR pointing separately with reference to the Two Micron All Sky Survey (2MASS; Cutri et al., 2003; Skrutskie et al., 2006) Point Source Catalog.[6] Taking steps to exclude saturated, crowded, and extended sources, we matched ISPI magnitudes measured in $16''$ diameter apertures to the 2MASS catalog 'default' magnitude (a $4''$ aperture flux, corrected to total assuming a point-source profile). For each subfield, the formal errors on these zeropoint determinations are at the level of 1–2 percent. The uncertainty is dominated by the 2MASS measurement errors, and are highest for the central pointing where there are only 6–8 useful 2MASS–detected point sources. For comparison, the formal 2MASS estimates for the level of systematic calibration errors is $\lesssim 0.02$ mag.

## 3 Data Combination and Cross-Calibration

This section is devoted to the combination and cross-calibration of the distinct datasets described in the previous section into a mutually consistent whole. In Section 3.1, we describe the astrometric cross-calibration of each of the ten images, including the mosaicking of the NIR data. We describe and validate our procedure for PSF-matching each band in Section 3.2.

### 3.1 Astrometric Calibration and Mosaicking

To facilitate multiband photometry, each of the final science images is transformed to a common astrometric reference frame: a north-up tangential plane projection, with a scale of $0\farcs267$ pix$^{-1}$. This chosen reference frame corresponds to the stacked $BVR$ image used as the detection image for the optically-selected MUSYC ECDFS catalog (see Gawiser et al., 2006,b), based on an early reduction of the WFI data.

Whereas WFI and Mosaic-II are both able to cover the entire ECDFS in a single pointing, the SofI and ISPI coverage consists of 32 and 9 subfields, respec-

[6] Available electronically via GATOR: http://irsa.ipac.caltech.edu/applications/Gator/.

tively. For these bands, each individual subfield was astrometrically matched to the $BVR$ reference image using standard IRAF/PyRAF tasks. For the ISPI data, each subfield is then combined, weighted by S:N on a per pixel basis, in order to create the final mosaicked science image. (Note that individual subfields are also 'PSF-matched' before mosaicking — see Section 3.2 below.)

One severe complication in this process is that exposure/weight maps were not available for the SofI imaging. We have worked around this problem by constructing mock exposure maps based on estimates of the per pixel rms in each science image. Specifically, we calculate the biweight scatter in rows and columns: $\sigma_B(x)$ and $\sigma_B(y)$. The effective weight for the pixel ($x$, $y$) is then estimated as $[\sigma_B(x)\sigma_B(y)]^{-2}$. The map for each subfield is normalized so that the median weight is 1 for those pointings that received 1 hr integration, and 4 for the four central pointings.

In line with Quadri et al. (2007), we found it necessary to fit a high order surface (*viz.*, a 6th-order Legendre polynomial, including $x$ and $y$ cross terms) to account for the distortions in the ISPI focal plane. For the SofI data, a 2nd-order surface was sufficient, although we did find it necessary to revise the initial astrometric calibration by Moy et al. (2003).

As an indication of the relative astrometric accuracy across the whole dataset, **Figure 2** illustrates the difference between the positions of all $K < 22$ sources measured from the K band, and those measured in each of the $Rz'JH$ bands (observed using WFI, Mosaic-II, ISPI, and SofI, respectively). Systematic 'shears' between bands are typically much less than a pixel. Comparing positions measured from the registered $R$ and $K$ band images, averaged across the entire field, the mean positional offset is $0\farcs15$ (0.56 pix). Looking only at the $x/y$ offsets, we find the biweight mean and variance to be $0\farcs03$ (0.11 pix) and $0\farcs3$ (1.1 pix), respectively.

### 3.2 PSF Matching

The basic challenge of multiband photometry is accounting for different seeing in different bands, in order to ensure that the same fraction of light is counted in each band for each object. We have done this by matching the PSFs in each separate pointing to that with the broadest PSF. Of all images, the southwestern $J$ pointing has the broadest PSF: $1\farcs5$ (FWHM). This sets the limiting seeing for the multiband SED photometry. Among the $K$ pointings, however, the worst seeing is $1\farcs0$ (FWHM); this sets the limiting seeing for object detection, and the measurement of total $K$ magnitudes (see Sections 4.1 and 4.3 below). We have therefore created eleven separate science images: one $1\farcs5$ FWHM image for each of the ten bands to use for SED photometry, plus a $1\farcs0$ FWHM $K$ image for object detection and the measurement of total $K$ fluxes.

The PSF-matching procedure is as follows: for each pointing, we take a list of SED-classified stars from the COMBO-17 catalog; these objects are then used to construct an empirical model of the PSF in that image, using an iterative scheme to discard low S:N, extended, or confused sources. Our results do not

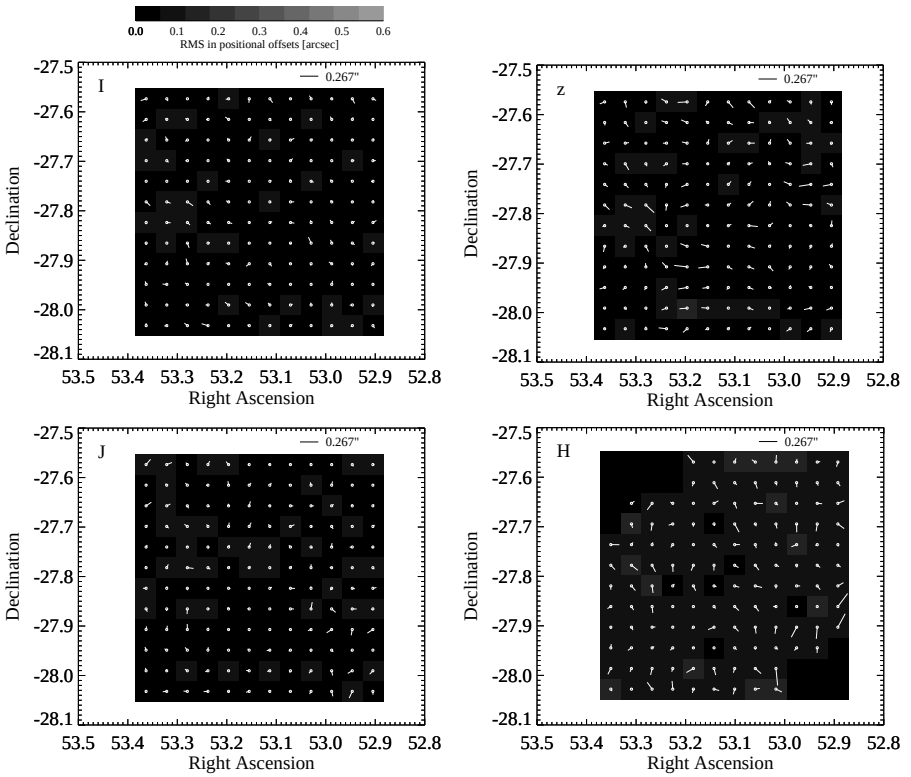


**Figure 2.** — Astrometric registration of the (from top to bottom) $Iz'JH$ images (obtaining using WFI, Mosaic-II, ISPI, and SofI, respectively), relative to the $K$ detection image. — In each panel, vectors give the biweight mean positional offset between the two images in $2\farcm5 \times 2\farcm5$ cells, based on all $K < 22$ sources; the greyscale gives the biweight variance. Systematic astrometric shears in individual images images are typically much less than a pixel.

change if we begin with $Bz'K$ selected stars, or GEMS point sources. We then use the IRAF/PyRAF task `lucy` (an implementation of the Lucy-Richardson deconvolution algorithm, and part of the STSDAS package[7]) to determine the convolution kernel required to 'degrade' each subfield to the target effective seeing. Finally, the convolution is done using standard tasks. Note that each of the NIR subfields is treated individually, prior to mosaicking.

In order to quantify the random and systematic errors resulting from imperfect PSF matching, **Figure 3** shows the relative difference between the curves of growth of individual point sources across 9 of our 10 bands, after matching to the target $1\farcs5$ FWHM PSF. In this Figure, we compare the growth curves of many bright, unsaturated, isolated point sources as a function of aperture diameter; specifically, we plot the relative difference between the normalized growth curves

[7]STSDAS is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

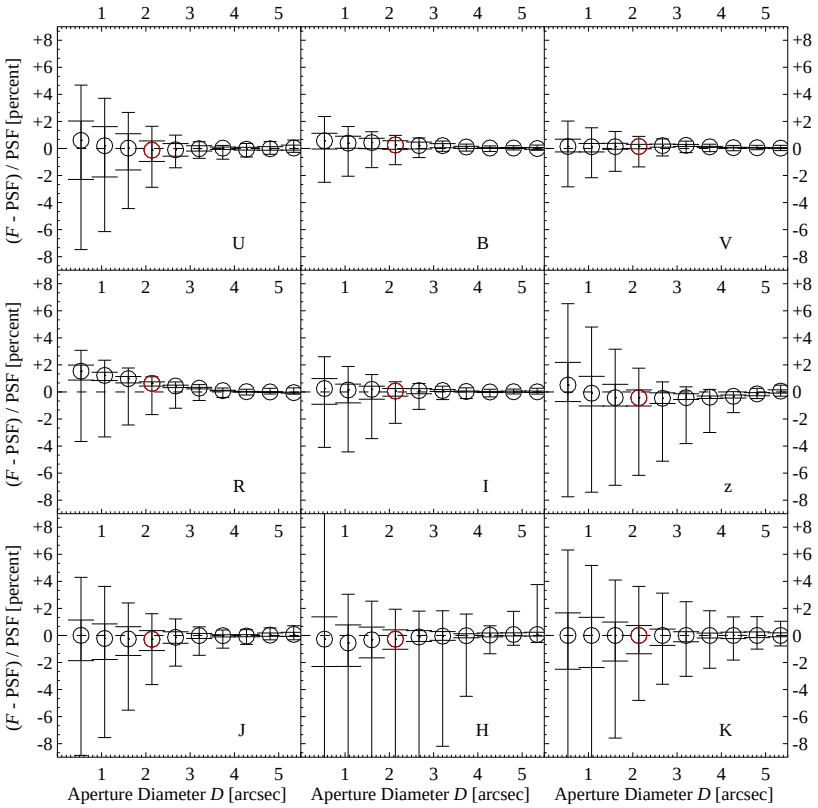


**Figure 3.** — Relative deviations in the curves of growth for point-sources in each of nine bands, from four different instruments, after PSF matching (1.″5 FWHM). — Each panel shows the relative differences between the normalized growth curves of bright, unsaturated, isolated point sources, plotted as a function of aperture diameter. Circles show the median of all growth curves in each band; large and small error bars show the 33/67 and 5/95 percentiles, respectively. The growth curves in different bands are all normalized with respect to the $K$ band median; the systematic errors in the $K$ panel are thus zero by construction. For our smallest apertures (2.″5), systematic offsets due to imperfect PSF matching are at worst 0.006 mag; random errors, due to, for example, spatial variation of the PSF, are $\lesssim 0.03$ mag.

in each band, compared to the median $K$ band growth curve. Within each panel, the circles represent the median growth curve in each band (zero for the $K$ band by construction), and the large and small error bars represent the 33/67 and 5/95 percentiles, respectively.

After PSF matching, there are signs of spatial variations in the FWHM of the $J$ and $K$ PSFs at the few percent level, particularly toward the edges of each pointing. But since the scatter in these plots represents both real spatial deviations in the PSF, as well as normalization errors, the results in Figure 3 can thus can be taken as an upper limit on the random PSF-related photometric errors. Looking at the $z'$-band panel, it is possible that the broad $z'$ band PSF wings are impor-

tant at the $\lesssim 0.005$ mag level for $2\farcs5$–$5\farcs0$. Note that the smallest apertures we use are $2\farcs5$ in diameter — for these apertures, random errors due to imperfect PSF matching are typically $\lesssim 0.03$ mag, and systematic errors are at worst 0.006 mag.

## 4 Detection, Completeness, Photometry, and Photometric Errors

In this section, we describe our scheme for building our multiband catalog of the ECDFS; a summary of the contents of the final photometric catalog is given in **Table 2**. We rely on SExtractor (Bertin & Arnouts, 1996) for both source detection and photometry; in Section 4.1 we describe our use of SExtractor, and we quantify catalog completeness and reliability in Section 4.2. There are two separate components to the reported photometry for each object: the total $K$ flux, which is discussed in Sections 4.3 and 4.4, and the 10-band SED, which is discussed in Section 4.5. Finally, in Section 4.6, we describe the process by which we have quantified the photometric measurement uncertainties.

### 4.1 Detection

Source detection and photometry for each band was performed using SExtractor in dual image mode; that is, one image is used for detection, and photometry is done on a second 'measurement' image. In all cases, the $1\farcs0$ FWHM $K$ band mosaic (see Section 3.2) was used as the detection image; since flexible apertures are always derived from the detection image, this assures that the same apertures are used for all measurements in all bands.

As a standard part of the SExtractor algorithm, the detection image is convolved with a 'filter' function that approximates the PSF; we use a 4 pix ($\sim 1\farcs0$) FWHM Gaussian filter. We adopt an absolute detection threshold equivalent to 23.50 mag / $\square''$ in the filtered image, requiring 5 or more contiguous pixels for a detection. Since we have performed our own background subtraction for the NIR images (see Section 2.2.3), we do not allow SExtractor to perform any additional background subtraction in the detection phase. For object deblending, we set the parameters `DEBLEND_NTHRESH` and `DEBLEND_MINCONT` to 64 and 0.001, respectively. These settings have been chosen by comparing the deblended segmentation map for the $K$ detection image to the optical $BVR$ detection stack, which has a considerably smaller PSF.

Near the edges of the observed region, where coverage is low, we get a large number of spurious sources. We have therefore gone through the catalog produced by SExtractor, and culled all objects where the $K$ effective weight, $w_K$, is less than 0.2 (equivalent to $\lesssim 12$ min per pointing). This makes the effective imaging area 953 $\square''$. Further, we find that a large number of spurious sources are detected where there are 'holes' in the coverage map (a product of the regular dither pattern used for the earliest eastern and northeastern tiles; see Section 3.1.) To avoid these spurious detections, for scientific analysis we will consider only those

| Column No. | Column Title | Description |
|---|---|---|
| 1 | `id` | Object identifier, beginning from 1 |
| 2, 3 | `ra`, `dec` | Right ascension and declination (J2000), expressed in decimal degrees |
| 4 | `field` | An internal MUSYC field identifier (ECDFS=8) |
| 5, 6 | `x`, `y` | Center of light position, expressed in pixels |
| 7 | `ap_col` | Effective diameter (*i.e.* $\sqrt{4\pi A}$, where $A$ is the aperture area) in arcsec; we use the larger of SExtractor's `ISO` aperture and a $2\farcs5$ diameter aperture to measure colors (see Section 4.5) |
| 8–27 | `U_colf`, `U_colfe`, etc. | Observed flux[a], with the associated measurement uncertainty, in each of the $UU_{38}BVRIz'JHK$ bands, as measured in the 'color' aperture |
| 28 | `ap_tot` | Effective diameter of the `AUTO` aperture, on which the $K$ flux measurement is based |
| 29, 30 | `K_totf`, `K_totfe` | Total $K$ flux[a], based on SExtractor's `AUTO` measurement, with corrections applied for missed flux and background over-subtraction (see Section 4.3) and the associated measurement uncertainty (accounting for correlated noise, random background subtraction errors, spatial variations in the noise, Poisson shot noise, etc.; see Section 4.6) |
| 31, 32 | `K_4arcsecf`, `K_4arcsecfe` | $K$ flux[a], as measured in a $4''$ aperture, with the associated measurement uncertainty |
| 33, 34 | `K_autof`, `K_autofe` | $K$ flux[a] within the `AUTO` aperture, with the associated measurement uncertainty |
| 35–37 | `Kr50`, `Keps`, `Kposang` | Morphological parameters from SExtractor, measured from the $1\farcs0$ FWHM $K$ image; *viz.*, the halflight radius (where the 'total' light here is the `AUTO` flux), ellipticity, and position angle |
| 38–47 | `Uw`, etc. | Relative weight in each of the $UU_{38}BVRIz'JHK$ bands[b] |
| 48 | `id_sex` | The original SExtractor ID,[c] for use with the SExtractor-generated segmentation map |
| 49, 50 | `f_deblend1`, `f_deblend2` | Deblending flags from SExtractor, indicating whether an object has been deblended, and whether that object's photometry is significantly affected by a near neighbor, respectively |
| 51 | `star_flag` | A flag indicating whether an object's $Bz'K$ color suggests it is a star (see Section 7.1) |
| 52–54 | `z_spec`, `qf_spec`, `spec_class` | Spectroscopic redshift determination, if available, along with the associated quality flag and spectral classification, if given (see Appendix A) |
| 55, 56 | `source`, `nsources` | A code indicating the source of the spectroscopic redshift, and the number of agreeing determinations |
| 57, 58 | `qz_spec`, `spec_flag` | A figure of merit, derived from the MUSYC photometry, for the spectroscopic redshift determination, and a binary flag indicating whether the spectroscopic redshift is considered 'secure' |

**Table 2.** — Summary of the contents of the photometric catalog. — *Notes:* [a] All fluxes are given in such a way that they can be transformed to AB magnitudes using a zeropoint of 25; in other words, fluxes are given in units of 0.363 mJy. [b] For all but the $z'$ and $H$ bands, this is essentially the integration time, normalized by the nominal values given in Table 2. For the $H$ band, this value is derived from the mock exposure map described in Section 3.1; for the $z'$ band, this is a binary flag indicating whether the $z'$ photometry is significantly affected by light from a nearby bright star. [c] Recall that we have excised all detections with an effective exposure time of less than 12 min from the catalog output by SExtractor.

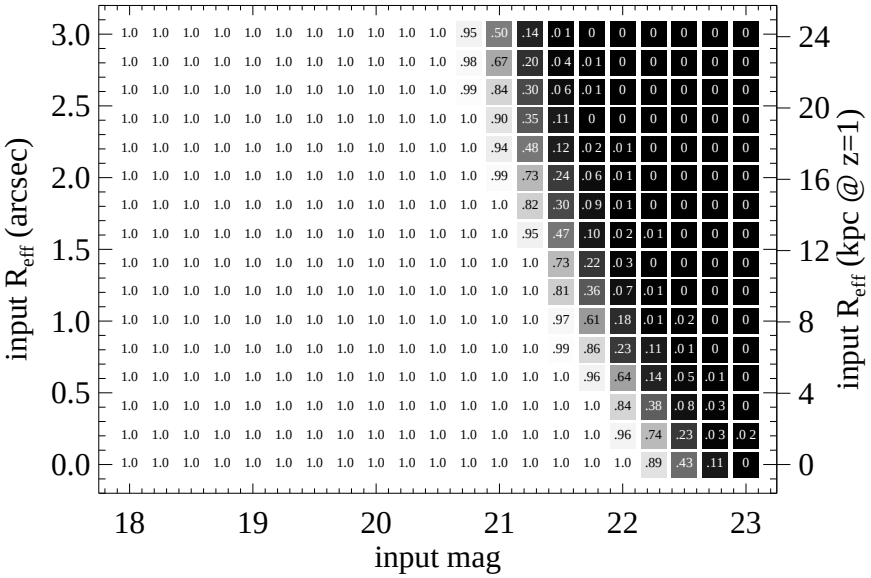


**Figure 4.** — Completeness for synthetic $R^{1/4}$-law sources added to the 1.″0 FWHM $K$ detection mosaic. — At each point, we give the fraction of synthetic sources (characterized by their total $K$ magnitude and effective or halflight radius, $R_{\rm e}$) that we are able to recover, using identical processes as for 'live' detection. Each synthetic source has been isolated from any other real or synthetic source by at least 50 pix to prevent confusion. While we are 100 % complete for point sources (*i.e.*, $R_{\rm e} = 0$) at our nominal limit of $K = 22$, the completeness drops rather rapidly for larger radii/lower surface brightnesses. At fixed $K$ and $R_{\rm e}$, both surface brightness and completeness are strong functions of the profile shape; we present these results for $R^{1/4}$ sources as loose lower limits on the true completeness.

detections with an $w_k > 0.75$ (equivalent to $\sim 45$ min per pointing) or greater.[8] This selection reduces the effective area of the catalog to 887 $\Box''$.

### 4.2 Completeness and Reliability

In order to estimate the catalog completeness, we have added a very large number of simulated sources into the 1.″0 FWHM detection image, and checked which are recovered by SExtractor, using the same settings as for 'live' detection. The completeness is then just the fraction of inputed sources which are recovered, as a function of source size and brightness. We adopted a de Vaucouleurs ($R^{1/4}$-law) profile for all simulated sources, each with a halflight radius, $R_{\rm e}$, between 0″ (*i.e.* a point source) and 3″, an ellipticity of 0.6, and total apparent $K$ magnitude in the range 18–23 mag. We truncate each object's profile at 8 $R_{\rm e}$. No more than 750 artificial galaxies were added at any given time, corresponding to 3–5 % increase

[8] In other words, the catalog is based on the area that received the equivalent of $\gtrsim 12$ min integration, but our scientific analysis is restricted to those objects that received $\gtrsim 45$ min integration. While objects with $0.2 < w_K < 0.75$ are given in the catalog, we do not include them in our main science sample, because of the poorer completeness and reliability among these objects.

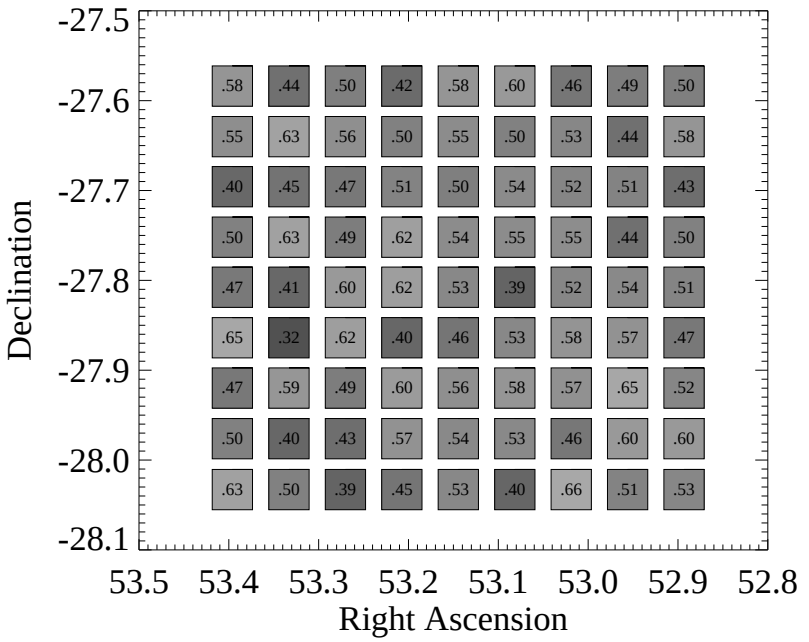


**Figure 5.** — Spatial variation in the completeness of the MUSYC ECDFS catalog. — Completeness for synthetic $K = 22.4$ point sources randomly added to the K detection image; our nominal completeness for these objects is $\approx 50$ %. Completeness is slightly lower around the very bright star towards the eastern edge of the field, but is otherwise reasonably uniform.

in the number of detected sources. Simulated sources were placed at least $13\farcs35$ (50 pix) away from any other detected or simulated source; these completeness estimates therefore do not account for confusion.

The results of this exercise are shown in **Figure 4**, which plots the completeness as a function of size and brightness. For point sources, we are 50 %, 90 %, and 95 % complete for K = 22.4, 22.2, and 22.1 mag, respectively. At a fixed total magnitude, the completeness drops for larger, low surface brightness objects. At $K = 22$, the nominal completeness limit of the catalog, we are in fact only 84 % complete for $R_{\rm e} = 0\farcs4$, assuming an $R^{1/4}$ profile. Note that we detect quite a few objects that 'really' lie below our formal (surface brightness) detection limit: just as noise troughs can 'hide' galaxies, noise peaks can help push objects that would not otherwise be detected over the detection threshold. (See also Section 4.3.)

Note that the above test deliberately avoids incompleteness due to source confusion. If we repeat the above test without avoiding known sources, we find that where completeness is low, confusion actually increases the completeness by a factor of a few, with faint sources hiding in the skirts of brighter ones (see also Berta et al., 2006). However, the flux measurements for these objects are naturally dominated by their neighbours; in this sense, it is arguable as to whether the synthetic object is actually being 'detected'. Where completeness is high

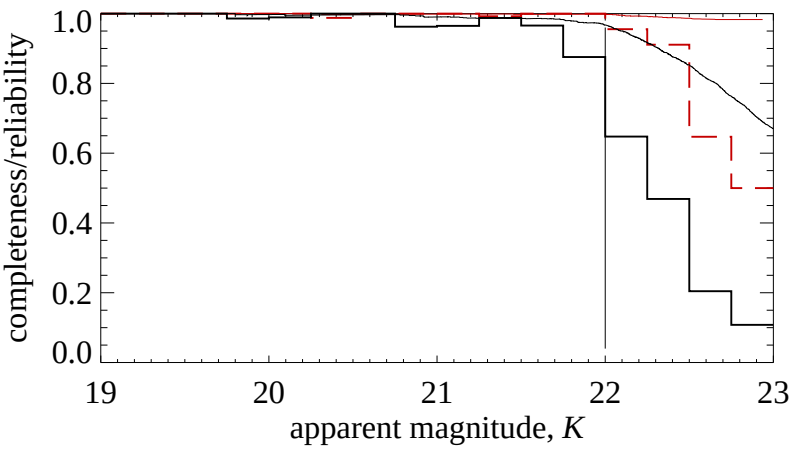


**Figure 6.** — Completeness and reliability of the MUSYC ECDFS catalog in comparison to the much deeper FIREWORKS catalog (Wuyts et al., 2008) of the GOODS-CDFS data. — We show the fraction of FIREWORKS sources that are detected by MUSYC (*i.e.*, the completeness of the MUSYC catalog; *solid black histograms*), and the fraction of MUSYC sources that are confirmed by FIREWORKS (*i.e.*, the reliability of the MUSYC catalog; *dashed red histograms*), in bins of total apparent $K$ magnitude. For the $21.75 < K < 22.00$ bin, we are more than 85 % complete, with essentially all detections confirmed by the GOODS data. Cumulatively, to $K < 22$, the MUSYC catalog is 97 % complete (*black curve*), and better than 99 % reliable (*red curve*).

($K \lesssim 20.5$), confusion reduces completeness by a few percent, but again, the exact amount is sensitive to the position and flux agreement required to define a successful detection. These tests suggest that $\lesssim 2$ % of sources are affected by confusion due to chance alignments with foreground/background galaxies (*cf.* physical associations). For comparison, based on the SExtractor segmentation map, $K$-detected objects cover 2.34 % of the field.

We have also done a similar test to investigate any variations in completeness across the field. We placed 5000 point sources with K = 22.4 — our 50 % completeness limit for point sources — across the field, each isolated by at least $26\farcs7$ (100 pix). The results are shown in **Figure 5**. Although it is perhaps slightly lower for the slightly shallower east and northeast pointings, the completeness is indeed quite uniform across the full field.

Finally, we can obtain empirical measures of both completeness and reliability by comparing our catalog to the much deeper $K$-selected FIREWORKS catalog of the GOODS-CDFS region (Wuyts et al., 2008) The results of this exercise are shown in **Figure 6**. Here, the completeness is just the fraction of FIREWORKS sources which are found in the MUSYC catalog; similarly, the reliability is the fraction of MUSYC sources which also appear in the FIREWORKS catalog. For the $21.8 < K < 22.0$ bin, the MUSYC catalog is 87.5 % complete, and 97 % reliable. For $K < 22$, the overall completeness and reliability are 97 % and 99 %, respectively.

Since the GOODS-ISAAC data are so much deeper, the high completeness at K $\sim$ 22 implies that K $\sim$ 22, $R_e \gtrsim 0\farcs5$ objects make up at most a small fraction of the FIREWORKS catalog. This might imply that our catalog is primarily flux, rather than surface brightness, limited. It must also be remembered, however, that the main motivation for large area surveys like MUSYC is to find the rare objects that may be missed in smaller area surveys like GOODS.

### 4.3 Total Fluxes — Method

We measure total fluxes in the $1\farcs0$ FWHM $K$ band mosaic, using SExtractor's `AUTO` measurement, which uses a flexible elliptical aperture whose size ultimately depends on the distribution of light in 'detection' pixels (*i.e.*, an isophotal region). We do specify a minimum `AUTO` aperture size (using the parameter `PHOT_AUTOAPERS`) of $2\farcs5$, although in practice this limit is almost never reached for sources with $K < 22$. The $2\farcs5$ limit has been chosen to be small enough to ensure high S:N for faint point sources, while still avoiding any significant aperture matching effects (see both Sections 3.2 and 4.6). We apply two corrections to the `AUTO` flux to obtain better estimates of galaxies' total fluxes; these are described below. We will then quantify the effect and importance of these corrections in the following section.

Even for a point source, any aperture that is comparable in size to the PSF will miss a non-negligible amount of flux (see, e.g., Bertin & Arnouts, 1996; Fasano, Filippi & Bertola, 1998; Cimatti et al., 2002; Labbé et al., 2003; Brown et al., 2007). Brown et al. (2007) have shown that fraction of light missed by the `AUTO` aperture correlates strongly with total magnitude; this is simply due to the fact that the `AUTO` aperture size correlates strongly with total brightness. Labbé et al. (2003) find that up to 0.7 mag can be missed for some objects, and Brown et al. (2007) suggest that the systematic effect at the faint end is $\sim$ 0.2 mag.

It is therefore both appropriate and important to apply a correction for missed flux laying outside the 'total' aperture. Following Labbé et al. (2003), we do this treating every object as if it were a point source: using empirical models of the PSF constructed as per Section 3.2, we determine the fraction of light that falls outside each aperture as a function of its size and ellipticity, and scale SExtractor's `FLUX_AUTO` measurement accordingly. Since no object can have a growth curve which is steeper than a point source, this is a minimal correction: it leads to a lower limit on the total flux.

Further, we find that SExtractor's background estimation algorithms systematically overestimate the background level, which also produces a bias towards lower fluxes. Because SExtractor does not allow the user to turn off background subtraction when doing photometry (*cf.* detection), we are forced to undo SExtractor's background subtraction manually for the final catalog, using the output `BACKGROUND` values, and the area of the `AUTO` aperture. We have done this only for the total $K$ fluxes; since we have performed our own background subtraction for the NIR images (as described in Section 2.2.3), undoing SExtractor's background subtraction is equivalent to trusting our own determination. Note that,

for the SED fluxes, we still rely on SExtractor's `LOCAL` background subtraction algorithm, with `PHOTO_THICK` set to 48.

### 4.4 Total Fluxes — Validation

Our overarching concern here is the correspondence between our measured fluxes and the true total fluxes of real sources. We have tested our total flux measurements by checking our ability to recover the known fluxes of large numbers synthetic sources, inserted into the 1.″0 FWHM $K$ science image as in Section 4.2. The results of these tests are shown in **Figure 7**. In this Figure, we compare the performance of SExtractor's `AUTO` measurement before (*upper panels*) and after (*lower panels*) our corrections for missed flux and background over-subtraction are applied. In each case, the contours show the systematic (*left panels*) and random (*right panels*) errors in the recovered magnitude. The red lines show the approximate 90 %, 50 %, and 10 % completeness limits for $R^{1/4}$-law sources, as derived in Section 4.2.

Further, in order to gauge the way these measurements are affected by noise, we have performed several variations of this test. In each test we add the synthetic sources either to a noiseless image, or to the actual 1.″0 FWHM $K$ mosaic; we have trialled the four possible permutations of using the noiseless or real image for detection or measurement. We briefly summarize the results of these tests below.

The reader wishing to avoid such a technical discussion of SExtractor's photometry algorithms may wish to skip to Section 4.5 after noting that, comparing the upper and lower panels of Figure 7, the effect of our two corrections to the `AUTO` measurement is to reduce the systematic underestimate of total fluxes by $\gtrsim 0.10$ mag. For point sources, the total flux is recovered to within 0.02 mag for $K < 22$.

#### *4.4.1 Missed Flux and Aperture Size Effects*

In order to determine the bias inherent in the `AUTO` algorithm, we have checked our ability to recover the fluxes of synthetic sources placed in a noiseless image, using this image for both detection and measurement. For point sources, the photometric bias inherent in the `AUTO` algorithm is $\lesssim 0.05$ mag for $K < 20.5$, but rises to 0.10 mag for $K = 22$. It is also a strong function of $R_{\rm e}$: at $K = 21.5$, the `AUTO` aperture misses 0.12 mag for $R_{\rm e} = 0.''4$, and more than 0.25 mag for $R_{\rm e} = 1.''0$. Applying our 'point source' correction for missed flux reduces this bias to $< 0.02$ mag for all $K < 22$ point sources; and, at $K = 21.5$, to 0.08 and 0.21 mag for $R_{\rm e} = 0.''4$ and $1.''0$, respectively.

The above numbers indicate the bias inherent in the `AUTO` algorithm, even for infinite signal–to–noise; considering synthetic sources introduced into the real $K$ science image, we find that noise exacerbates the problem. For point sources, the mean offset between the uncorrected `AUTO` and total fluxes are $\lesssim 0.05$ mag for $K < 20.0$, 0.10 mag for $K = 21.5$ and 0.17 mag for $K = 22.0$. For $K = 21.5$, the systematic offset is 0.16 mag for $R_{\rm e} = 0.''4$ and 0.50 mag for $R_{\rm e} = 1.''0$. For $K = 22$, the average 'point source' correction for missed flux goes from 0.05 mag for true point sources up to 0.10 mag for $R_{\rm e} = 1.''0$, and 0.15 mag for $R_{\rm e} = 1.''5$.

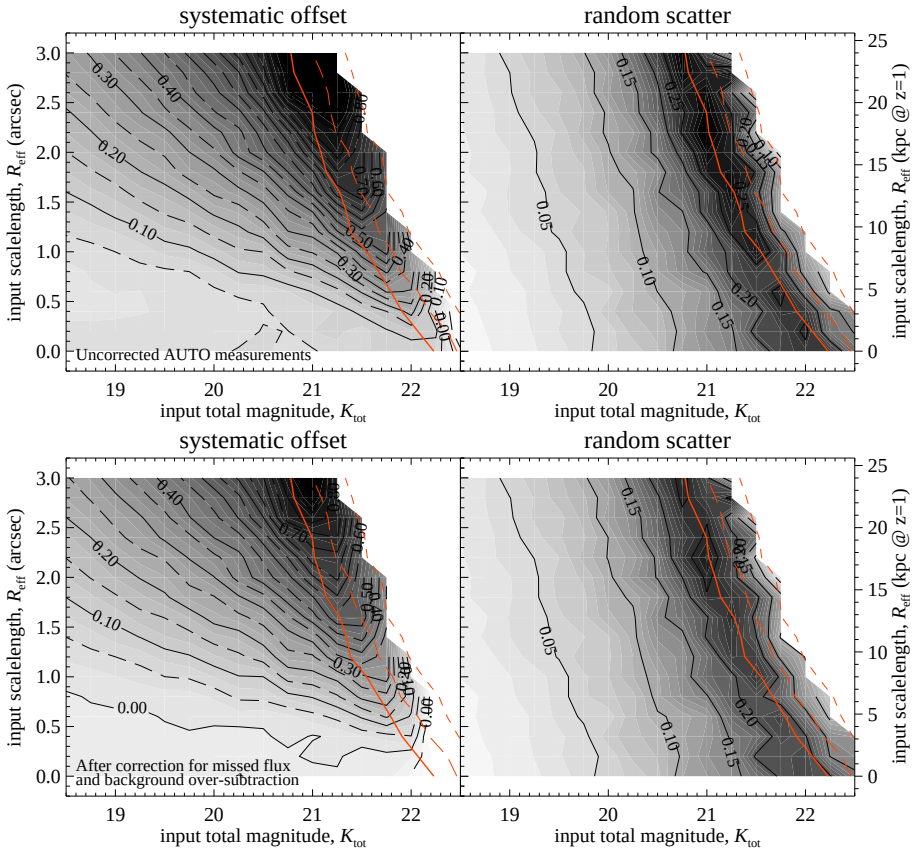


**Figure 7.** — Validating our total flux measurements. — The systematic (*left panels*) and random (*right panels*) errors in the recovered fluxes of synthetic $R^{1/4}$-law sources introduced into the $1\farcs0$ FWHM $K$ science image, based on SExtractor's `AUTO` aperture, before (*upper panels*) and after (*lower panels*) applying corrections for missed flux and background oversubtraction. The red lines in each panel show the approximate 90 % (*solid*), 50 % (*long dashed*), and 10 % (*short dashed*) completeness limits for $R^{1/4}$ law sources, as in Figure 4. As in Figure 4, the results shown in this Figure are presented as upper limits on the systematic errors; both the random and systematic errors are significantly less assuming exponential profiles. In order to account for flux laying beyond the `AUTO` aperture, we correct the flux of each source as if it were a point source; this is thus a minimal correction. This correction reduces the systematic error in total fluxes by $\gtrsim 0.1$ mag for an $R^{1/4}$ profile, and from $\lesssim 0.10$ mag to $\lesssim 0.01$ mag for point sources. For the faintest sources, *the missed flux correction also reduces the random error by as much as* 0.05 *mag*. We also find that SExtractor's LOCAL background subtraction algorithm tends to overestimate and oversubtract the background, leading to a systematic offset at the level of 0.03 mag. With both of these corrections, the systematic errors in total fluxes for point sources are reduced to $\lesssim 2$ %

After applying our correction for missed flux, the photometric offset is reduced to $< 0.03$ mag for all point sources; at $K = 21.5$, the numbers for $R_e = 0\farcs4$ and $R_e = 1\farcs0$ become 0.10 mag, and 0.35 mag, respectively.

As an aside, we have also looked at how noise *in the detection image* affects the `AUTO` measurement, by using the real image (with synthetic sources added), for detection, and using a noiseless image for measurement. The effect of noise in the detection image is to induce scatter in the isophotal area, and so the `AUTO` aperture size, at a fixed $R_e$ and $K$. By eliminating the first order effects due to aperture size, applying a correction for missed flux thus reduces the random scatter in the recovered fluxes of low surface brightness sources: the random scatter in recovered fluxes is reduced by $\sim 0.05$ mag for all $K \lesssim 21$ sources. This can also be seen in Figure 7.

As in Section 4.2, note that the numbers given above all apply to galaxies with an $R^{1/4}$ profile, and so should be treated as approximate upper limits on the random and systematic errors. We have performed the same test assuming exponential profiles: the systematic error in the recovered flux is less than 0.03 mag for all $K < 22$ and $R_e < 0\farcs6$.

### *4.4.2 Background Oversubtraction*

Even after correcting for missed flux, and even for point sources, SExtractor's photometry systematically underestimates the total fluxes of synthetic sources. At least part of this lingering offset is a product of the LOCAL background subtraction algorithm. This algorithm uses a 'rectangular annulus' with a user-specified thickness, surrounding the quasi-isophotal detection region. Any flux from the source lying beyond this 'aperture' (which may well be smaller than the `AUTO` aperture!) will therefore bias the background estimate upwards, leading to oversubtraction, and so a systematic underestimate of the total flux.

If we undo SExtractor's background subtraction[9], then the photometric offset for point sources is reduced to $< 0.02$ mag for all $K < 22$. The size of this correction is only weakly dependent on source size and flux, varying from $\gtrsim -0.025$ mag for ($K$, $R_e$) = (19, $0\farcs4$) to -0.038 for ($K$, $R_e$) = (22, $0\farcs4$).

## 4.5 Multiband SEDs

In order to maximize signal–to–noise for the faintest objects, instead of measuring total fluxes in all bands, we construct multicolor SEDs based on smaller, 'color' apertures; we then use the $K$ band total flux to normalize each SED.

The 'color' photometry is measured from $1\farcs5$ FWHM PSF-matched images (see Section 3.2), again using the $1\farcs0$ FWHM $K$ mosaic as the detection image. Specifically, we use SExtractor's `MAG_ISO`, again enforcing a minimum aperture size of $2\farcs5$ diameter. This limit is reached by essentially all objects with $K > 21.5$,

[9] Again, note that SExtractor does not allow the user to turn off background subtraction for photometry. In practice, we have undone SExtractor's background subtraction using the output BACKGROUND value, multiplied by the area of the `AUTO` aperture. The `AUTO` aperture area is given by `KRON_RADIUS`$^2 \times \pi \times$ `A_IMAGE` $\times$ `B_IMAGE`. Note, too, that we apply this correction *before* the missed flux correction discussed in Section 4.4.1.

and essentially none with $K < 20.5$. Note that, even though the `ISO` aperture is defined from 1.″0 FWHM $K$ mosaic, (after SExtractor's internal filtering; see Section 4.1), all 'color' measurements are made using matched apertures on 1.″5 FWHM PSF-matched images.

In order to test our sensitivity to color gradients, we have verified that $R_{\rm tot} = (R-K)_{\rm col} + K_{\rm tot}$, where $R_{\rm tot}$ comes from using the $R$ band image in place of the $K$ band image for detection and total flux measurement. Particularly for the brightest and biggest (and so, presumably, the nearest) galaxies, the use of the `ISO` aperture (*cf.* a fixed aperture) is crucial in ensuring that this is indeed true.

### 4.6 Photometric Errors

Following, for example, Labbé et al. (2003), Gawiser et al. (2006), and Quadri et al. (2007), we empirically determine the photometric measurement uncertainties by placing large numbers of apertures on empty or blank regions in our measurement images. The principal advantage of this approach is that it correctly accounts for pixel–pixel correlations introduced in various stages of the data reduction process (including interpolation during astrometric correction and convolution during PSF matching).

For the 'color' apertures, we have placed 2.″5–8″ apertures on $10^4$ independent (*i.e.*, non-overlapping) 'empty' regions, where 'empty' is defined using the combined optical ($BVR$) and NIR ($K$) segmentation maps.. With this information, we can build curves of $\sigma(A)$ for each band, where $\sigma$ is the measurement uncertainty in an aperture with area $A$. Similarly, for the 'total' apertures, which are somewhat larger, we have placed 2.″5–12″ independent apertures at 3500 'empty' locations on the 1.″0 FWHM $K$ detection mosaic, using only the NIR segmentation map to define 'empty'. Note that since the 'empty aperture' photometry is done using SExtractor in the same manner as for our final photometry, the errors so derived also account for random uncertainties due to, for example, errors in background estimation, etc.

There is one additional layer of complexity for the ISPI bands: in order to track the spatial variations in the 'background' rms, both within and between subfields, we use the rms maps produced during mosaicking by `xdimsum` (see Section 2.2.2). While these maps are not accurate in an absolute sense, they do adequately map the shape of rms variations across each subfield. We have therefore normalized these maps by the rms flux in empty 2.″5 apertures, and then combined them to construct a (re)normalized 'rms map' for the full $30' \times 30'$ field. Then, in practice, the photometric uncertainty for a given object is estimated by taking the median pixel value within the SExtractor segmentation region associated with that object, corrected up from 2.″5 to the appropriate aperture size using the $\sigma(A)$ curves described above.

In **Figure 8**, we validate these error estimates by showing the 'empty aperture' fluxes, $F$, as measured in 2.″5 diameter apertures, as a function of the photometric error, $\Delta F$, estimated as above. The line with error bars shows the mean and variance of the 'empty aperture' fluxes in bins of $\Delta F$; in other words, the error

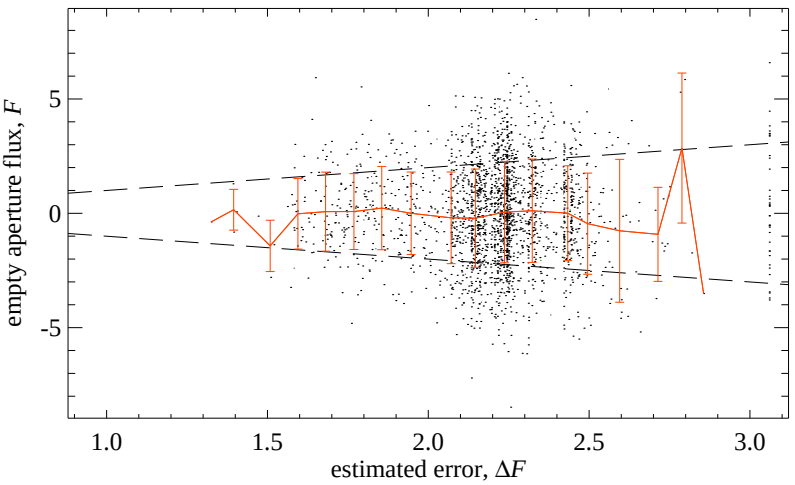


**Figure 8.** — Validating our flux error estimates. — For the $JK$ imaging, we use the rms maps output by `xdimsum` to account for spatial variations in the background noise level. Although these maps are not accurate in an absolute sense, they do adequately map the relative variations in the noise; accordingly, we have normalized these rms maps using the scatter in 2.″5 diameter apertures placed on empty regions of the science image. The points in this Figure show the integrated flux in each of 3000 of these apertures, $F$, as a function of the estimated error, $\Delta F$, derived using these rms maps. The red error bars show the rms in $F$, binned by $\Delta F$. The observed rms in empty apertures agrees extremely well with the estimated errors based on the normalized `xdimsum` rms map.

bars show the *actual* error, plotted as a function of the estimated error. The agreement between the photometric errors estimated using the rms map, and the variance in 'empty aperture' fluxes is excellent. This is more than just a consistency check: while the rms maps have been normalised to match the variance in empty aperture fluxes on average, the fact that the observed scatter scales so well with the predicted error demonstrates that the rms map does a good job of reproducing the spatial variations in the noise.

For a Gaussian profile (*i.e.*, a point source), and in the case of uncorrelated noise, an aperture with a diameter 1.35 times the FWHM gives the optimal S:N (Gawiser et al., 2006). Based on the 'empty aperture' analysis described in Section 4.6, the 2.″5 aperture size is slightly larger than optimal for a point source in the $J$ (1.″5 FWHM) image. For the 1.″0 FWHM $K$ detection image, the optimal aperture diameter for a point source is 1.″33; the S:N in a 2.″5 diameter aperture is $\approx 25$ % lower. Using slightly larger apertures presumably increases S:N for slightly extended sources, as well as reducing sensitivity to systematic effects due to various classes of aperture effects (e.g., imperfect astrometric and PSF matching, etc.).

Within a 2.″5 diameter aperture, the formal $5\sigma$ limits in the $K$ band are 22.25 mag at an effective weight of 0.75, and 22.50 mag at an effective weight of 1.0.

Averaged across the image, the $5\sigma$ limit is 22.42 mag; the limits for all bands are given in Table 2. For a point source, these limits can be translated to total fluxes by simply subtracting 0.45 mag.

# 5 Additional Checks on the MUSYC Calibration

## 5.1 Checks on the Astrometric Calibration

In order to test the astrometric calibration of the MUSYC ECDFS imaging, we have compared the cataloged position of sources from the $K$-selected catalog with those from version 3.3 of the Yale/San Juan Southern Proper Motion (SPM) catalog (Girard et al., 2004). This catalog is based on observations made using the 51 cm double astrograph of Cesco Observatory in El Leoncito, Argentina. For $V < 17$, the positional accuracy of the catalog is $0\farcs04$–$0\farcs06$.

In **Figure 9**, we show an astrometric comparison for 113 objects common to the SPM and MUSYC catalogs; these objects are plotted as black circles. For this comparison we have selected objects with $14 < V < 18$ and proper motions of less than 20 mas/yr. All these objects have $14 < K < 18$; the median has $K = 16$ mag.

The systematic offset between SPM– and MUSYC–measured positions, averaged across the entire field, is $0\farcs079$ in right ascension and $0\farcs222$ in declination; that is, a mean offset of $0\farcs235$ (0.88 pix), 20° east of north. For these sources, the random error in the MUSYC positions is $0\farcs30$ and $0\farcs27$ in $x$ and $y$, respectively.

We have performed the same comparison for the 2MASS sources that were used in the photometric calibration of the $K$ images; these objects are shown in Figure 9 as the grey crosses. The median $K$ magnitude of these objects is 14.75 mag, considerably brighter than the SPM sources used above. In comparison to the 2MASS catalog, which has astrometric accuracy of $\lesssim 0\farcs1$ for $K < 14$, we find a slightly larger random offset: ($0\farcs22$, $0\farcs39$) in ($\alpha$, $\delta$). For these sources, the random error in ($\alpha$, $\delta$) is ($0\farcs22$, $0\farcs19$).

In the lower part of Figure 9, we plot the positional offsets as a function of position accross the field. In these panels, the solid grey line shows the median-filtered relation between SPM– and MUSYC–measured positions. There appears to be a slight astrometric shear in the RA direction at the $\lesssim 0\farcs3$ level from the east to the west edge of the $K$ mosaic. Otherwise, however, the offsets are consistent with the direct shift of $0\farcs235$ derived above.

## 5.2 Checks on the Photometric Calibration

### *5.2.1 Comparison with FIREWORKS*

In order to test our photometric calibration, we have compared our catalog to the FIREWORKS catalog (Wuyts et al., 2008) of the GOODS-CDFS region (the central $\sim 150\ \square''$ of our field), which includes *HST*-ACS optical imaging, and significantly deeper NIR imaging taken using ISAAC on the VLT. Since the FIREWORKS catalog uses different filters, we are forced to use stellar colors to make this comparison. The results of this comparison are shown in **Figure 10** (p. 50). Each panel in this Figure shows the color–color diagram for stars in terms of

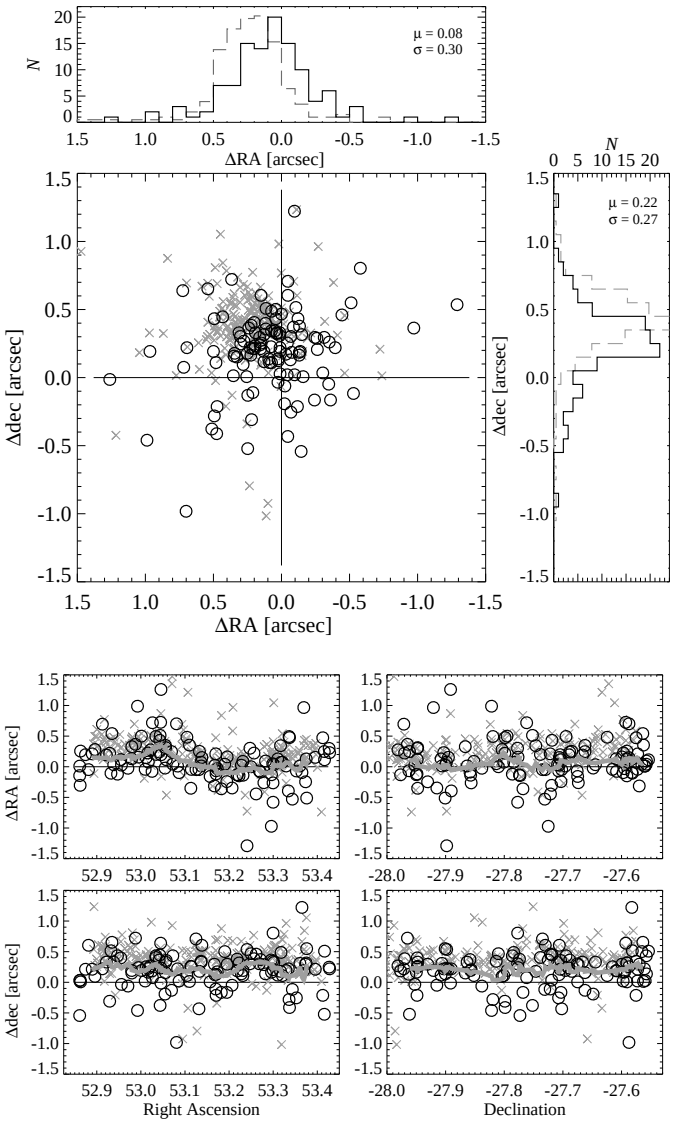


**Figure 9.** — Validating the MUSYC ECDFS astrometric calibration. — In the upper part of this Figure, we show the field-averaged astrometric comparison between the MUSYC ECDFS catalog, and the Yale/San Juan Southern Proper Motion (SPM) catalog v3.3 (*circles*), as well 2MASS (*crosses*). In the flanking panels, the solid (dashed) histograms show the distribution of RA/dec offsets with respect to the SPM (2MASS) catalogs; we also give the mean and rms offset between MUSYC and SPM catalog positions. In the lower part of this Figure, we show astrometric offsets as a function of position; in these panels, the solid grey lines show the median-filtered relation derived from the SPM points. In comparison to the SPM catalog, the MUSYC astrometry is offset by $0\farcs23$ (0.87 pix); there is also evidence of an astrometric shear of $\lesssim 0\farcs3$ (1.1 pix) in the RA direction across the full field.

their FIREWORKS ($V_{606W} - I_{775W}$) color, and a MUSYC–minus–FIREWORKS 'color'. In each panel, the circles with error bars show the data; the error bars refer only to errors in the MUSYC photometry.

We have used spectra for luminosity class V stars from the BPGS stellar spectral atlas (Gunn & Stryker, 1983) to generate predictions for where the stellar sequence should lie in these diagrams. These predictions are the solid red lines in each panel; the small blue stars show the predicted photometry for individual BPGS stars. Our results do not change if we use the Pickles (1998) stellar atlas. Note that, for the purposes of this comparison, we have converted to the Vega magnitude system, so that the stellar sequence necessarily passes through the point (0, 0).

We calculate the photometric offset in each band as the S:N–weighted mean difference between the observed stellar photometry and the predicted stellar sequence. These values are given in each panel; the dashed red line is just the predicted stellar sequence offset by this amount.

Particularly for the NIR data, the absolute calibrations of the MUSYC and FIREWORKS data agree very well: typically to better than 0.03 mag. In terms of the relative calibration across different bands, we see a discrepancy between the $I$ and $z'$ band calibrations of $\Delta(I - z') = 0.05$ mag, as well as a discrepancy between the $U_{38}$ and $B$ bands at the level of $\Delta(U_{38} - B) = -0.07$ mag.

### *5.2.2 Comparison with COMBO-17*

Although the COMBO-17 broadband $U_{38}BRVI$ imaging is a subset of the raw data used to produce the MUSYC imaging, the data reduction and analysis strategies used by each team are very different. In particular, rather than a single measurement from a coadded image, the COMBO-17 flux measurements are based on the coadding of many distinct measurements from the individual integrations, and SED or 'color' measurements are made using adaptive, weighted 'apertures', rather than traditional (top-hat) apertures. Direct, object–by–object comparison between the two catalogs thus offers the chance to test both the photometric calibration, and the methods used for obtaining photometry.

The results of this comparison are shown in the middle panels of Figure 11; these panels show the difference in the MUSYC and COMBO-17 cataloged $U_{38}BVRI$ fluxes, plotted as a function of total $R$ magnitude in the COMBO-17 catalog, $R_{\rm C17}$. The comparison is between total fluxes: *i.e.*, $I_{\rm C17} = R_{\rm tot,C17} + (I - R)_{\rm C17}$; $I_{\rm MUS} = K_{\rm tot,MUS} + (I - K)_{\rm MUS}$. As in the previous section, we have also transformed our data to the Vega magnitude system. For the purposes of this comparison, we distinguish between stars (*red stars*) and galaxies (*black points*), on the basis of the COMBO-17 SED classification; the results do not change significantly using $Bz'K$ selected stars or GEMS point sources. We have used those stars with $R < 21$ to identify differences in the two surveys' calibrations; these offsets are given in each panel, and shown as the dotted black lines.

There are significant differences between the MUSYC and original COMBO-17 calibrations. These are due to calibration errors in the COMBO-17 catalog (Wolf

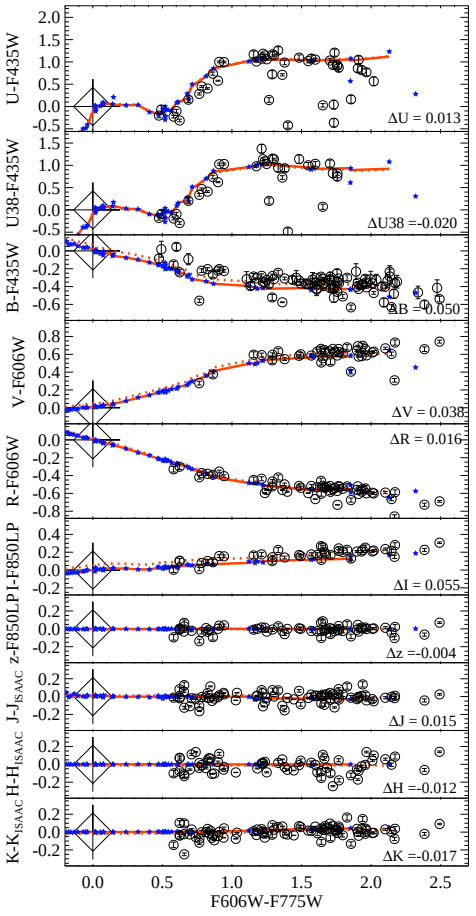


**Figure 10.** — Photometric comparison between GOODS and MUSYC in the CDFS, based on the FIREWORKS catalog (Wuyts et al., 2008) of the GOODS data. — In effect, these panels show the empirical color–transforms for stars between the GOODS (ACS/ISAAC) and MUSYC (WFI/Mosaic-II/SofI/ ISPI) filters, plotted as a function of ACS color. Note that for this comparison, we have adopted the Vega magnitude system, so that the stellar sequence must necessarily pass through the point (0, 0). In each panel, the open circles with error bars represent the data; the errors shown here pertain only to the MUSYC photometry. The closed blue stars show predicted stellar photometry based on the BPGS (Gunn & Stryker, 1983) stellar spectral atlas, convolved with the known filter curves; the solid red line shows a prediction for the stellar sequence in these color–color diagrams, obtained by median filtering the BPGS points. For each filter, we derive a photometric offset by taking the mean difference, weighted by S:N in the MUSYC catalog, between the observed and predicted location of the stellar sequence in the $y$ direction. These values are given in each panel (the '$\Delta$' having the sense of MUSYC–minus–GOODS); the dotted lines show the predicted stellar sequence offset by this amount. Particularly for the reddest bands, this comparison validates the MUSYC photometric calibration at the few percent level.

et al., 2008). The original COMBO-17 calibration was based on spectrophotometric observations of two stars, each of which suggested different calibrations; in the end, the wrong star was chosen.[10] Partially motivated by the comparison in Figure 11, Wolf et al. (2008) have since revised the basic calibration of the COMBO-17 ECDFS data using the other spectrophometric star, shifting the $U_{38}BVRI$ calibration by –0.143, +0.040, +0.003, –0.054, and –0.123 mag, respectively.

We note that these rather large calibration errors do not have a huge effect on the COMBO-17 redshift determinations (Wolf et al., 2008; Hildebrandt, Wolf & Benítez, 2008, Chapter III;). This is because the medium bands, which are key to measuring break strengths and so choosing the redshift, are calibrated with respect to the nearest broad band. However, we show in Chapter III that the effect on derived quantities like restframe colors and stellar masses is large.

[10]Note that these calibration issues affect only the ECDFS, and not the other three COMBO-17 fields, where multiple calibration stars give consistent results (Wolf et al., 2008).

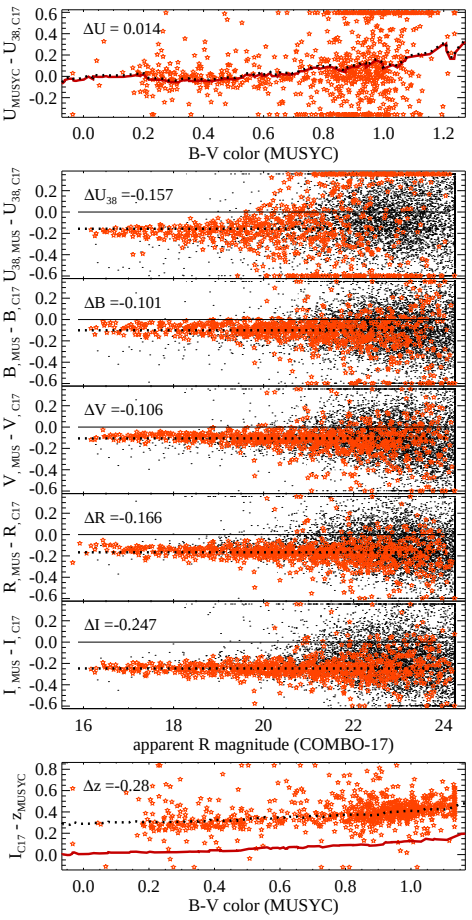


**Figure 11.** — Photometric comparison between COMBO-17 and MUSYC in the ECDFS. — Each panel shows the difference between the original COMBO-17 and MUSYC photometry; the red stars represent the observed photometry for stars, selected using COMBO-17's SED classification. In the central panels, which show direct object–by–object comparisons for those filters common to both MUSYC and COMBO-17, the black points show the same information for galaxies, plotted as a function of apparent $R$ magnitude in the COMBO-17 catalog. The top and bottom panels compare the MUSYC $U$ and $z'$ bands to the COMBO-17 $U_{38}$ and $I$ bands, as a function of the MUSYC $(B-V)$ color (in the Vega system), and based on synthetic photometry for main sequence stars. The median MUSYC–minus–COMBO-17 photometric offset for each band is given in each panel; the dotted lines in each panel show the expected location of the stellar locus offset by this amount. Note that while the COMBO-17 raw data are a subset of the MUSYC raw data, the data reduction and analysis pipelines are completely independent. There are significant differences between the COMBO-17 and MUSYC photometry, due at least in part to photometric calibration errors in the original COMBO-17 catalog (Wolf et al., 2008). Even after recalibrating the COMBO-17 photometry following (Wolf et al., 2008), however, significant differences remain: for $U_{38}BVRI$, the offsets are –0.014, –0.141, –0.109, –0.112, and –0.124 mag, respectively.

After recalibration using the other spectrophotometric standard, the MUSYC and COMBO-17 stellar colors agree at the level of a few hundredths of a magnitude for $BVRI$; for $U_{38}$ a discrepancy remains at the 0.1 mag level. Moreover, a discrepancy in the overall calibration remains, such that stars are 0.1 mag brighter in the MUSYC catalog. Our correction for missed flux accounts for 0.03 mag of this offset; the source of the remaining 0.07 mag offset has not been identified.

Secondly, notice that there are apparently different offsets for galaxies and stars: even after matching the two surveys' calibrations *for stars* using Figure 11, *galaxies are still fainter and bluer* in the COMBO-17 catalog than they are in ours. Quantitatively, the $U_{38}BVRI$ galaxy-minus-star offsets are 0.102, 0.020, 0.010, 0.067, and 0.088 mag, respectively. Further, excepting the $U_{38}$ band, the random scatter between the COMBO-17 and MUSYC galaxy photometry is 2–3 times greater than that for stars. It is difficult to say what might produce this

effect, but the effect persists even when we use our $R$ band image for detection and measurement; that is, this is not a product of our measuring total fluxes in $K$ rather than $R$. We do not believe that the combination of COMBO-17's smaller effective apertures and galaxy color gradients can fully account for these effects. For $R \gtrsim 21$, the effective diameter of the `ISO` aperture is almost always smaller than $2\farcs5$; for these objects the MUSYC photometry effectively uses fixed apertures. While the agreement between star and galaxy colors is noticeably better for $R \lesssim 21$ using fixed $2\farcs5$ apertures to construct SEDs, it does not have a significant effect for $R \gtrsim 21$, where the problem is greatest.

While we cannot directly compare our $U$ and $z'$ photometry to COMBO-17, it is still possible to use stellar colors to check these bands, as we have done for the FIREWORKS catalog. This is shown in the top and bottom panels of Figure 11. For the $z'$ band, this analysis suggests a possible discrepancy between the MUSYC $I$ and $z'$ band calibrations of $\Delta(I-z')_{\rm MUS} = 0.03$ mag. For the $U$ band, however, it suggests a discrepancy of $\Delta(U-U_{38})_{\rm MUS} \sim 0.15$ mag. While we have been unable to identify the cause of this offset, we note both that the shape of the observed and predicted stellar sequences do not obviously agree as well for the $U$ band as for the $z'$, and also that the results of both Sections 5.2.1 and 5.2.3 do not support the notion of an offset of this size. We do not believe that this indicates an inconsistency in the calibrations of the $U$ and $U_{38}$ bands.

### *5.2.3 Refining the Photometric Cross-Calibration using Stellar SEDs*

In the construction of SEDs covering a broad wavelength range, the relative or cross-calibration across all bands is at least as important as the absolute calibration of each individual band. As a trivial example, if the zeropoints of two adjacent bands are out by a few percent, but in opposite senses, this can easily introduce systematic offsets in color on the order of 0.1 mag; the worry is then that these apparent 'breaks' might seriously affect photometric redshift determinations. This is a particular concern in the case of the MUSYC ECDFS dataset, which incorporates data from four different instruments, each reduced and calibrated using quite different strategies.

We have therefore taken steps to improve the photometric cross-calibration of the MUSYC ECDFS data. The essential idea here is to take a set of objects whose SEDs are known *a priori* (at least in a statistical sense) and to ensure agreement between the observed and expected SEDs. Stars are, in fact, ideal for this purpose, since they form a narrow 'stellar sequence' when plotted in color–color space. At least in theory, and modulo the effects of, e.g., metallicity, a star's (*cf.* a galaxy's) full SED can be predicted on the basis of a single color.

Our method is as follows. We begin with a set of more than 1000 objects with unambiguous `Star` classifications in the COMBO-17 catalog, of which nearly 600 have photometric S:N $\gtrsim$ 10 in $K$, and are unsaturated in all MUSYC bands. Again, our results do not change if we use $Bz'K$ selected stars or GEMS point sources. Using EAZY (see Section 7.2 for a description), we fit the objects' photometry with luminosity class V stellar spectra from the BPGS stellar spectral

| Band | Photometric Offset with respect to | | |
|---|---|---|---|
| | FIREWORKS | COMBO-17 | Stellar SEDs |
| (1) | (2) | (3) | (4) |
| $U$ | +0.013 | +0.02 | −0.004 |
| $U_{38}$ | −0.020 | −0.15 | −0.051 |
| $B$ | +0.050 | −0.09 | −0.017 |
| $V$ | +0.038 | −0.09 | −0.006 |
| $R$ | +0.016 | −0.15 | +0.017 |
| $I$ | +0.055 | −0.23 | +0.023 |
| $z'$ | −0.004 | −0.27 | −0.011 |
| $J$ | +0.015 | — | +0.032 |
| $H$ | −0.012 | — | −0.032 |
| $K$ | −0.017 | — | — |

**Table 3.** — Checks on the photometric calibration. — This Table summarizes the results of Section 5.2. For each band (Col. 1), we give: the photometric offset between MUSYC and the FIREWORKS (Wuyts et al., 2008) catalogs of the GOODS ACS and ISAAC imaging data (Col. 2); the photometric offset between the MUSYC and COMBO-17 (Wolf et al., 2004) optical imaging data (Col. 3); the residuals from fitting stellar SEDs from the MUSYC catalog using main sequence stellar spectra from the BPGS atlas (Col. 4).

atlas as a template set, and the redshift fixed to zero. Note that, by default, EAZY includes a 0.05 mag systematic error on each SED point, added in quadrature with the measurement uncertainty.

Using the output $\chi^2$ to discard objects whose SEDs are not consistent with being a main sequence star, we can then interpret the median residual between the observed and best-fit photometry as being the product of calibration errors, and so refine the photometric calibration of each band to ensure consistency across all bands. Specifically, given the photometric errors, we use $\chi^2$ minimization to determine the zeropoint revision.

The zeropoint revisions derived in this way are small; $\lesssim 0.05$ mag in all cases. The exact revisions are given in **Table 3**, which also serves as a summary of the results presented in this Section.. Across the WFI data, there appears to be an offset that is roughly monotonic between the $U_{38}$ and $I$ bands, where the offset in $U_{38} - I$ is $-0.075$ mag; *cf.* $-0.074$ mag from the comparison to the FIREWORKS catalog. Similarly, there is an apparent inconsistency between the $I$ and $z'$ calibrations, such that the offset in $(I - z')$ is 0.03 mag; *cf.* 0.05 mag from the comparison to FIREWORKS.

The crux of this method is that whatever zeropoint discrepancies exist do not affect the choice of the best fit template in a systematic way. For example, a large offset in the $U$ bands or a wavelength–dependent offset might lead to stars being fit with systematically bluer or redder template spectra, so biasing the derived photometric offsets. In this sense, it is reassuring that the derived offsets are small, and comparable to the quoted uncertainties on the photometric calibration. Further, we note that we get very similar results if we increase the systematic uncertainty used by EAZY to 0.10 mag.

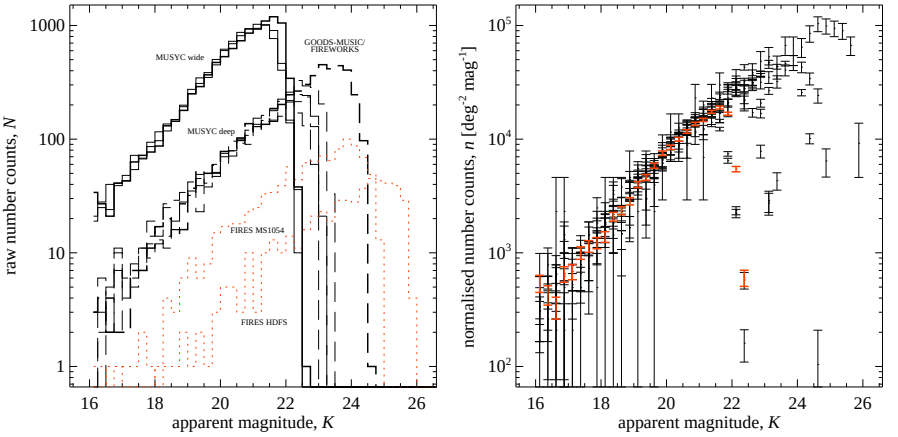


**Figure 12.** — $K$ band apparent magnitude number counts, comparing the MUSYC ECDFS catalog to other $K$-selected catalogs. — *Left Panel:* the raw numbers of detected sources, in bins of $K$ magnitude for the ECDFS (*heavy solid histogram*), in comparison to: the other MUSYC wide fields (*light solid histograms*; Blanc et al., 2008); the MUSYC deep fields (*light dashed histograms*; Quadri et al., 2007); the FIREWORKS catalog of the GOODS-CDFS data (*heavy dashed histogram*; Wuyts et al., 2008); and the two FIRES fields (*red dashed histograms*; Labbé et al., 2003; Förster-Schreiber et al., 2006). *Right panel:* The normalized number counts for the same collection of datasets; the MUSYC ECDFS data are highlighted (*heavy red points*). At a fixed $K$ magnitude, while the GOODS region of the ECDFS has approximately 80 % as many sources as are found in the ECDFS as a whole, in comparison to the other MUSYC wide fields, the ECDFS is underdense at the level of $\sim 5$ %.

Given the agreement between the results of the external comparison to FIREWORKS and those from the internal consistency check on stellar colors, we have chosen to adopt the zeropoint revisions suggested by this stellar colors exercise. With these revisions, we believe that our photometric calibration is accurate, in both an absolute and a relative sense, to the level of a few hundredths of a magnitude.

## 6 Number Counts

As a very basic comparison between our catalog and other $K$-selected catalogs, **Figure 12** shows the number of detected galaxies as a function of total apparent $K$ magnitude. Note that all the catalogs shown apply a similar correction for flux missed by SExtractor's `AUTO` measurement. The left panel of this figure shows the raw number counts; the right shows the number counts normalized by area. In both panels, it can be seen that our number counts drop off for $K \gtrsim 22$; our catalog is nearly, but not totally, complete for $K = 22$.

The overall agreement between these different catalogs is very good. Assuming that the calibration of all catalogs is solid, the left panel of Figure 12 would indicate that the ECDFS is slightly underdense — at the level of 4–6 % for

$17.5 < K < 21.5$. — in comparison to the two other MUSYC wide NIR selected catalogs (Blanc et al., 2008). Conversely, the ECDFS number counts could be matched to the other two wide catalogs by adjusting the ECDFS $K$ photometric calibration by $-0.06$ or $-0.09$ mag.

In comparison to the number counts from the FIREWORKS catalog of the GOODS CDFS region, the GOODS region contains approximately 18 % fewer sources per unit area than the ECDFS as a whole. Even after matching the MUSYC ECDFS $K$ calibration to the FIREWORKS catalog (see Section 5.2.1), the GOODS region remains underdense by 16 % in comparison to the ECDFS.

# 7 Photometric Redshifts

## 7.1 Star/Galaxy Separation

We separate stars and galaxies from within the MUSYC ECDFS catalog on the basis of their $Bz'K$ colors. The $Bz'K$ diagram is known as a means of selecting moderate redshift ($z \gtrsim 1.4$) galaxies (Daddi et al., 2004), but can also be used as a efficient means of distinguishing stars from galaxies (see, e.g., Grazian et al., 2006; Blanc et al., 2008). In **Figure 13**, we evaluate the performance of this criterion in comparison to the stellar SED classification from COMBO-17 (Wolf et al., 2004), as well as to a catalog of point sources from GEMS (Häussler et al., 2007).

Both panels of Figure 13 show the $Bz'K$ diagram for the MUSYC ECDFS catalog (*black points*); the $Bz'K$ stellar selection line:

$$(z' - K) \leq 0.3\ (B - z') - 0.5\ , \tag{1}$$

is shown as the dashed line. In total, from the main $K < 22$ sample, 755 sources are selected as stars on the basis of their $Bz'K$ colors. The left-hand panel of Figure 13 shows where $Bz'K$ star selection agrees with other indicators; the right-hand panel shows where there is disagreement. For instance, on the left, the star-shaped symbols show objects that are classified as 'stars' by COMBO-17; on the right, they represent those $Bz'K$-selected 'stars' which are *not* classified as such by COMBO-17. Similarly, the circles refer to point sources in the GEMS catalog. In both panels, objects that have been spectrally identified as stars are highlighted in red. In either panel, the stellar sequence is immediately obvious and, for a given $(B - z')$ color, can be seen to be separated from the galaxy population by at least a few tenths of a magnitude in $(z' - K)$.

Looking at the left panel, there is near complete overlap between COMBO-17's star classification and $Bz'K$ selection: only a very few COMBO-17 'stars' lie above the $Bz'K$ selection line. There are a few dozen GEMS point sources found above the $Bz'K$ selection line. In the MUSYC and GEMS optical images, some are clearly non-circular, and only a few show diffraction spikes; these appear to be compact, un– or barely–resolved galaxies. Note, too, that this region of the $Bz'K$ diagram is sparsely populated by X-ray sources (*i.e.* QSOs; Daddi et al., 2004; Grazian et al., 2006).

There are also a handful of objects that are spectroscopically identified as stars, which also fall above the $Bz'K$ star selection line. With one exception,

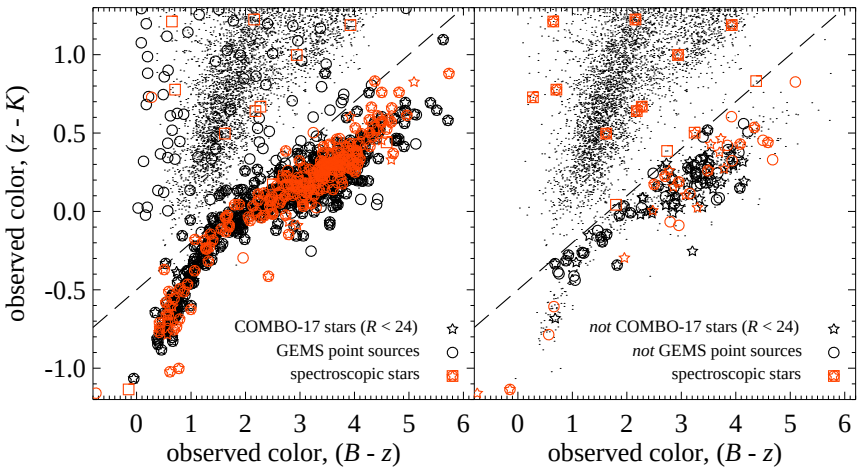


**Figure 13.** — Stellar identification using $Bz'K$ colors. — In each panel, we show the $Bz'K$ diagram for sources in the MUSYC ECDFS catalog (black points), and compare our $Bz'K$ star selection (*dashed line*) to other complementary stellar classifications: *viz.* SED-classified 'stars' from the COMBO-17 survey (*open stars*; Wolf et al., 2004), GEMS point sources (*open circles*; Häussler et al., 2007), and spectrally classified stars (*open squares/red symbols*). The left panel shows the agreement between $Bz'K$ selection and these other indicators; in the right panel we show where $Bz'K$-selection does not agree with the other indicators. So, for example, circles in the left panel show all GEMS point sources, whereas in the right panel they show those $Bz'K$-selected 'stars' that are *not* GEMS point sources. In the either panel, the stellar sequence in $Bz'K$ color space can be seen to be isolated by $\gtrsim 0.1$ mag in $(z' - K)$ from deep field galaxies. This includes QSOs, which can be seen in the left panel as GEMS point sources scattered throughout the galaxy population. Although there are a handful of spectrally classified stars lying well outside the $Bz'K$ stellar selection region (open squares in the left panel), these objects are neither COMBO-17 'stars' nor GEMS point sources (stars and circles in the right panel); *i.e.* the spectral classification is wrong. Of the $Bz'K$-selected stars which are not GEMS point sources (circles in the right panel), roughly half are faint stars superposed over a diffuse background galaxy, and roughly half are faint galaxies whose photometry is significantly affected by a bright, nearby star.

however, these objects are not GEMS point sources (squares in the left panel; circles in the right); neither are they classified as stars by COMBO-17 (squares in the left panel; stars in the right). These are, therefore, probably erroneous spectral classifications. There are no spectroscopic galaxies that lie in the stellar region of the $Bz'K$ diagram.

Turning now to the right panel, there are 66 $Bz'K$-selected 'stars' which do not appear in the GEMS point source catalog. A handful of these simply did not receive GEMS coverage. Of the rest, visual inspection shows these sources to be, in roughly equal proportions, faint stars superposed over a faint, background disk galaxy, or faint galaxies whose photometry is significantly affected by a nearby bright star. There are also 76 $Bz'K$-selected 'stars' which are not classified as such in the COMBO-17 catalog. In $(J - K)$–$K$ color–magnitude space, these objects almost all have $(J - K) < 0$ and $K < 21$; this would suggest that these

are faint stars misclassified by COMBO-17.

### 7.2 Photometric Redshifts — Method

The basic idea behind photometric redshift estimation is to use the observed SED to determine the probability of an object's having a particular spectral type, $T$ (drawn or constructed from a library of template spectra), and being at a particular redshift, $z_{\rm phot}$: *i.e.* $p(z_{\rm phot},\ T|{\rm SED})$. We have derived photometric redshifts for every object in the catalog using a new photometric redshift code called EAZY (Easy and Accurate $z_{\rm phot}$s from Yale; for a more detailed and complete discussion, see Brammer et al., 2008). EAZY combines many features of other commonly used photometric redshift codes like a Bayesian luminosity prior (e.g. BPZ; Benítez, 2000) and template combination (Rudnick et al., 2001, 2003) with a simple user interface based on the popular hyperz code (Bolzonella, Miralles & Pello, 2000). Novel features include the inclusion of a 'template error function'; a restframe wavelength dependent systematic error, which down-weights those parts of the spectrum like the restframe UV, where galaxies show significant scatter in color–color space. The user is offered full control over whether and how each these features are employed.

Another key difference is that objects are assigned redshifts by taking a probability weighted integral over the full redshift grid (*i.e.* marginalizing over the posterior redshift probability distribution), rather than, for example, choosing the single most likely redshift. (Although again the user is given the choice of which estimator to use.) EAZY also outputs 68/95/99 % confidence intervals, as derived from the typically asymmetric $p(z)$. EAZY thus outputs meaningful and reliable photometric redshift errors, including the effects of 'template mismatch'; *i.e.*, degeneracies between the redshift solution and the spectral type. By Monte Carlo'ing our catalog (*i.e.*, reanalyzing many Monte Carlo realizations of our photometry, each perturbed according to the photometric errors), we have verified that the EAZY $p(z)$ does in fact provide a good description of the redshift uncertainties due to photometric errors.

We have adopted EAZY's default parameter set for our redshift calculations.[11] That is, we use a library of six template spectra, allowing non-negative linear combinations between these basis templates, and including an apparent $K$ magnitude prior, $p(z|K)$, and using the default EAZY template error function. We also require photometry in at least 5 bands to compute a photometric redshift, although in practice this requirement is less restrictive than the weight cuts we apply in defining our science sample (see Section 9), and so has no effect

Both the base template set and the $K$ prior have been derived by Brammer et al. (2008) using synthetic photometry from the semi-analytic model of De Lucia & Blaizot (2007), which is in turn based on the Millenium Simulation (Springel et al., 2005). The motivation for this approach is to approximately account for

[11] In Chapter III, we present a number of variations on the photometric redshift computation described here; the redshifts described here correspond to the 'recalibration using stellar SEDs' test in Table 3 of Chapter III.

| Column No. | Column Title | Description |
|---|---|---|
| 1 | `id` | Object identifier, beginning from 1, as in the photometric catalog |
| 2 | `z_spec` | Spectroscopic redshift determination, where available, as given in the photometric catalog |
| 3, 4 | `z_a`, `chi_a` | Maximum likelihood redshift, allowing non-negative combinations of all six of the default EAZY templates, and the $\chi^2$ value associated with each fit |
| 5, 6 | `z_p`, `chi_p` | As above, but with the inclusion of a $K$ luminosity prior |
| 7, 8 | `z_m1`, `z_m2` | Probability-weighted mean redshift, without and with the inclusion of a $K$ luminosity prior, respectively; we recommend the use of the `z_m2` redshift estimator |
| 9–14 | `l68`, `u68`, etc. | Lower and upper limits on the redshift at 68, 95, and 99 % confidence, as computed from the same posterior probability distribution used to calculate `z_m2` |
| 15 | `odds` | The fraction of the total integrated probability within $\pm 0.2$ of `z_m2` |
| 16 | `qz` | The $Q_z$ figure of merit proposed by Brammer et al. (2008), calculated for `z_m2` |
| 17 | `nfilt` | The number of photometric points used to calculate all of the above |

**Table 4.** — Summary of the contents of the photometric redshift catalog.

the full diversity in $0 < z \lesssim 4$ galaxies' SEDs due to differences in their individual star formation and assembly histories. The $K$ prior is constructed directly from the De Lucia & Blaizot (2007) simulation.

In order to derive the base template set, Brammer et al. (2008) have applied the non-negative matrix factorization (NMF) algorithm of Blanton & Roweis (2007) to this synthetic catalog. In essence, this algorithm takes a large template library and distills from it a reduced set of basis templates that best describe the full range of 'observed' photometry. For this purpose, Brammer et al. (2008) have used the template library used by Grazian et al. (2006) to generate photometric redshifts for the GOODS-MUSIC catalog. This library consists of $\sim 3000$ Pégase synthetic spectra with a variety of dust obscurations, star formation histories, and ages. In additional to the five base templates output by the NMF algorithm, Brammer et al. (2008) also include one young, dusty template ($t = 50$ Myr; $A_V = 2.75$), to compensate for the lack of dusty galaxies in the De Lucia & Blaizot (2007) similuation.

Grazian et al. (2006), using their full template library, achieved a photometric redshift accuracy of $\sigma_z = 0.045$ for their GOODS-MUSIC catalog of the GOODS ACS-ISAAC-IRAC data. For the same data, and using the default setup described above, the EAZY photometric accuracy is $\sigma_z = 0.036$. This represents the current state of the art for photometric redshift calculations based on broadband photometry.

**Table 4** gives a summary of the information contained within the photometric redshift catalog. Note that when computing photometric redshifts, we only use photometry with an effective weight of 0.6 or greater. In addition to the basic

EAZY output, we have included two additional pieces of information. The first is simply a binary flag indicating whether or not each object is classified as a star on the basis of its $Bz'K$ colors. The second is the figure of merit proposed by Brammer et al. (2008):

$$Q_z(z_{\rm phot}) = \frac{\chi^2}{N_{\rm filt} - 3} \, \frac{z_{\rm up}^{99} - z_{\rm lo}^{99}}{p_{\Delta z=0.2}} \, . \tag{2}$$

This quantity combines the $\chi^2$ of the fit at the nominal redshift, the number of photometric points used in the fit, $N_{\rm filt}$, the width of the 99 % confidence interval, $(z_{\rm up}^{99} - z_{\rm lo}^{99})$, and the fractional probability that the redshift lies within $\pm 0.2$ of the nominal value, $p_{\Delta z=0.2}$; all of these quantities are output by EAZY by default. Brammer et al. (2008) have shown that a cut of $Q_z > 2$–3 can remove a large fraction of photometric redshift outliers.

### 7.3 Photometric Redshifts — Validation

In Appendix A, we describe both the spectroscopic redshift determinations that we have compiled for objects in the ECDFS, and show the $z_{\rm phot}$–$z_{\rm spec}$ agreement for individual $z_{\rm spec}$ samples. For all 'secure' redshift determinations, the random and systematic photometric redshift error is $\sigma_z = 0.036$ and med$[\Delta z/(1+z)] = -0.025$. In comparison to spectroscopic redshifts from the K20 survey, which is highly spectrally complete in the magnitude regime in which we are operating, the random error is $\sigma_z = 0.033$, with an outlier fraction of less than 5 %. (Here, we define the outlier fraction as the relative number of sources for which $\Delta z/(1+z) > 5\sigma_z$.) We also draw particular attention to the excellent agreement between our photometric redshifts and the spectroscopic determinations for the sample of van der Wel et al. (2005), which is a sample of 28 early type, red sequence galaxies at $z \sim 1$; we find $\sigma_z = 0.022$, with no outliers, and essentially no systematic offset. For comparison, the overall photometric redshift accuracy of the COMBO-17 survey for our $z_{\rm spec}$ comparison sample, but limited to $z_{\rm spec} < 1$, is $\sigma_z = 0.020$.

However, we also show in Appendix A that none of the available spectroscopic samples is particularly representative of the MUSYC ECDFS sample. In particular, in almost all cases there is a correlation between redshift security and $(J-K)$ color, such that redshift determinations for blue galaxies tend to be more secure, and so these galaxies are over-represented among MUSYC ECDFS galaxies. Even the K20 sample, which is 92 % complete for $K^{\rm (Vega)} < 20$, does not probe the reddest galaxies in our sample, presumably because they are too rare to be found in that survey's rather small area. There is, therefore, the very real danger that looking only at the $z_{\rm spec}$–$z_{\rm phot}$ agreement provides a false sense of security (see also Brammer et al., 2008), since there are comparatively few $z_{\rm spec}$s available for the faintest and reddest galaxies in the catalog — especially given that these are the main objects of interest.

For this reason, we have compared our photometric redshifts to those from COMBO-17 (Wolf et al., 2004) and GOODS-MUSIC (Grazian et al., 2006); the results of this comparison are shown in **Figure 14**. While these comparisons

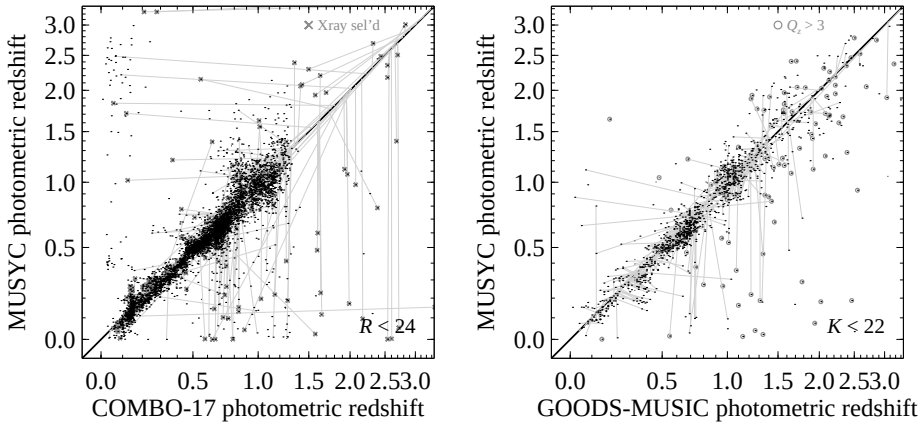


**Figure 14.** — Validating the MUSYC ECDFS photometric redshifts. — Each panel shows an object–by–object comparison between the MUSYC photometric redshift, and that from COMBO-17 (*left panel*; Wolf et al., 2004), and the GOODS-MUSIC catalog of the GOODS-CDFS data (*right panel*; Grazian et al., 2006). In order to discriminate between the two $z_{\rm phot}$s where there is disagreement, for those galaxies for which a robust spectroscopic redshift determination is available (see Appendix A), the red lines connect each point in ($z_{\rm phot}$, $z_{\rm phot}$) point to the point ($z_{\rm spec}$, $z_{\rm spec}$); vertical lines thus indicate catastrophic errors in the MUSYC $z_{\rm phot}$s, where horizontal lines show catastrophic failures in the COMBO-17/GOODS-MUSIC $z_{\rm phot}$s. COMBO-17 suffers from a few different classes of systematic effects, owing principally to the lack of NIR data. Note, however, that very few spectroscopic redshifts are available for these objects — *these effects would not be noticeable in a* $z_{\rm spec}$–$z_{\rm phot}$ *diagram* (see also Brammer et al., 2008). In this panel, X-ray selected sources are marked with a cross; within this $R < 24$ sample, X-ray-selected sources are roughly three times as likely to have $|\Delta z|/(1 + z_{\rm spec}) > 0.1$ (see also Appendix A). In comparison to GOODS-MUSIC, the MUSYC $z_{\rm phot}$s have a slightly greater number of catastrophic outliers, such that the MUSYC $z_{\rm phot}$ is far too low; again, many of these objects are X-ray sources. In this panel, objects with poorly constrained photometric redshifts (*viz.*, $Q_z > 3$) are marked with a circle; these objects are roughly twice as likely to have $|(z_{\rm MUS} - z_{\rm GDS})|/(1 + z_{\rm GDS}) > 0.2$. The overall agreement between the two redshift determinations is really very good, especially moving towards the 'redshift desert' at $z_{\rm phot} \gtrsim 1.5$, where few $z_{\rm spec}$s are available.

are extremely useful for identifying systematic differences between different $z_{\rm phot}$ solutions, without spectroscopic redshifts as a referent, they cannot be used to decide which is 'better' in the case of a disagreement. To this end, the red lines in each panel of this Figure show the spectroscopic redshifts (where available) by connecting the ($z_{\rm phot}$, $z_{\rm phot}$) point to the point ($z_{\rm spec}$, $z_{\rm spec}$). In each panel, vertical lines thus indicate where the COMBO-17 or GOODS-MUSIC $z_{\rm phot}$ is 'right', while the MUSYC $z_{\rm phot}$ is 'wrong'; conversely, horizontal lines show where the MUSYC $z_{\rm phot}$ is 'better' than that from COMBO-17 or GOODS-MUSIC. Note that for the comparison to COMBO-17, we restrict our attention to those galaxies with $R < 24$, since this is the reliability limit of the COMBO-17 catalog.

Owing to its medium-band photometry, the COMBO-17 redshifts should be

significantly better than our own for $z \lesssim 1$, but without NIR photometry, the redshifts of $z \gtrsim 1$ galaxies are poorly constrained. The agreement between the COMBO-17 photometric redshifts and our own (Figure 14; left panel), the agreement is indeed very good for $z_{\rm phot} < 0.8$. For $R < 24$ and $z_{\rm C17} < 1.0$, the random scatter between the COMBO-17 and MUSYC photometric redshifts is $\sigma_z = 0.034$; separately, for $R < 24$ and $z_{\rm spec} < 1$, the photometric redshift error is $\sigma_z = 0.030$ for MUSYC, and 0.020 for COMBO-17.

There are, however, several important differences between the MUSYC and COMBO-17 redshifts. First, note the effect of the $z_{\rm C17} < 1.4$ grid used by COMBO-17; coupled with their method of assigning redshifts (*viz.*, marginalizing over the redshift probability distribution), this means that galaxies are essentially never given $z_{\rm C17} \gtrsim 1.3$.

The exceptions to this rule are those objects that COMBO-17 has classified as QSOs on the basis of their optical SEDs; where MUSYC tends to place these objects at $z_{\rm MUS} \lesssim 1$, the COMBO-17 redshifts are very good. (Note that we have made no attempt to explicitly accommodate active galactic nucelii (AGNs) or QSOs in our photometric redshift calcuation.) In the left panel of Figure 14, we mark X-ray selected galaxies from the Szokoly et al. (2004) and Treister et al. (2008) catalogs with a cross. For this $R < 24$ sample, X-ray selected sources are roughly three times as likely to be outliers (here, we define outliers as those objects with $|\Delta z|/(1 + z_{\rm spec}) > 0.1$): the outlier fraction for X-ray sources is 35% (75/217), compared to 11% (164/1438) overall. Said another way, roughly half of all ($R < 24$) outliers are X-ray sources.

Secondly, there are two populations of objects with $z_{\rm C17} \lesssim 0.2$ that are placed by MUSYC at either $z_{\rm MUS} \sim 0.4$ or $z_{\rm MUS} \gtrsim 1.4$. From this first population, no $z_{\rm spec}$s are available; for the second, the handful of available $z_{\rm spec}$s confirm that these galaxies are at $z \gtrsim 1.4$. On the other hand, for the diffuse cloud of galaxies given $z_{\rm MUS} \lesssim 0.5$ and $0.5 \lesssim z_{\rm C17} \lesssim 1.0$, the $z_{\rm spec}$s support the COMBO-17 determinations.

Thirdly, while objects given $0.8 \lesssim z_{\rm phot} \lesssim 1.2$ in one catalog generally lie in the same redshift interval in the other, there is only a very weak correlation between the redshifts within this interval: the implication here is that objects with $z_{\rm spec} \gtrsim 0.8$ are assigned $0.8 \lesssim z_{\rm C17} \lesssim 1.2$ more or less at random on the basis of optical data alone. In other words, while the COMBO-17 $z_{\rm phot}$–$z_{\rm spec}$ agreement is excellent for $z_{\rm spec} \lesssim 0.8$, a $0.8 \lesssim z_{\rm C17} \lesssim 1.0$ selected sample may suffer significant contamination from $z_{\rm spec} \gtrsim 1$ galaxies with poorly constrained redshifts.

Looking now at the comparison with the GOODS-MUSIC redshifts (Figure 14; right panel), it is clear that, while the random scatter between the two determinations is larger than for the previous comparison, at least for $z \lesssim 1$, there are no signs of major systematic discrepancies. The random scatter between the GOODS-MUSIC and MUSYC photometric redshifts is $\sigma_z = 0.065$; separately, for the same $z_{\rm spec}$ comparison sample, the random errors are $\sigma_z = 0.036$ for MUSYC, and 0.043 for GOODS-MUSIC. Both MUSYC and GOODS-MUSIC suffer from catastrophic failures, where $z_{\rm spec} \sim 0.7$ galaxies are given $z_{\rm phot} \sim 0.2$; although

this appears to be a greater problem for MUSYC. GOODS-MUSIC also seems to have some systematic issues for $z_{\rm phot} \approx 0.4$.

In this panel, we also mark with a circle those objects with $Q_z > 3$. Whereas roughly half (938/1787) of the objects plotted in this panel have robust $z_{\rm spec}$s, the fraction among those with $Q_z > 3$ is just 33% (242/735); again, this underscores the importance of having a representative spectroscopic comparison sample. Using the cut $|(z_{\rm MUS} - z_{\rm GDS})|/(1 + z_{\rm GDS}) > 0.2$ to quantify the level of disagreement between the GOODS-MUSIC and MUSYC redshifts, objects with $Q_z > 3$ are twice as likely to be outliers: the fraction is 60% (99/166) for $Q_z > 3$, compared to 33% (586/1787) overall. We note that the fraction of sources with $Q_z > 3$ increases from $\lesssim 5\%$ for $z_{\rm GDS} \lesssim 1.2$ to $\sim 15$ % for $1.2 \lesssim z_{\rm GDS} \lesssim 2.2$. For $z_{\rm GDS} > 2.5$, roughly half (9/21) of all galaxies have $Q_z > 3$. Similarly, X-ray-selected galaxies are more likely to be outliers: the outlier fraction for X-ray sources is 43% (16/37).

Again, we caution that, without spectra for a large, representative subsample of the objects common to these two catalogs, it is not possible to determine whether one catalog is truly 'better' than the other. Moreover, given the differences between the MUSYC and GOODS-MUSIC catalogs — particularly the inclusion of ACS and IRAC imaging in the GOODS-MUSIC catalog — it is not possible to say whether any differences in photometric redshifts are due to the photometric redshift algorithms or to differences in the data themselves. Even with these caveats, however, the broad agreement between the MUSYC and GOODS-MUSIC $z_{\rm phot}$s — and especially for $z_{\rm phot} \gtrsim 1$, where $z_{\rm spec}$s are increasingly hard to come by — is certainly encouraging.

## 8 Interpolating Restframe Photometry — Introducing InterRest

Given an SED and a redshift, we have derived restframe photometry following the method described in Appendix C of Rudnick et al. (2003). This method is best understood as interpolating between two points in the observed SED to come up with a restframe flux. We have developed an IDL implementation of this algorithm for interpolating restframe photometry, dubbed InterRest. InterRest has been specifically designed to dovetail with EAZY: it accepts the same inputs and configuration files, and uses the same algorithms for integration, etc. We have made this utility freely available to the astronomical community.

The essential idea is to use a set of template spectra to construct a color–color relation for galaxies at a given redshift. Specifically, we relate a color in terms of two observed filters to another color in terms of an observed filter and the desired restframe filter. For example, in order to find the restframe $r$ flux of a galaxy at $z = 1.2$ ($\lambda_{\rm em} = 6220$ Å; $\lambda_{\rm ob} = 13700$ Å), we would relate the $(z' - J)$ color to the $(r_{z=1.2} - J)$ color; the $r_{z=1.2}$ flux then immediately follows.

This process is illustrated in **Figure 15**, with one crucial difference: whereas normally, in order to interpolate a restframe flux, we would relate an observed–minus–observed color to a restframe–minus–observed color, in this example we are concerned with using the observed $(V - I)$ color to *predict* the observed $(R - I)$

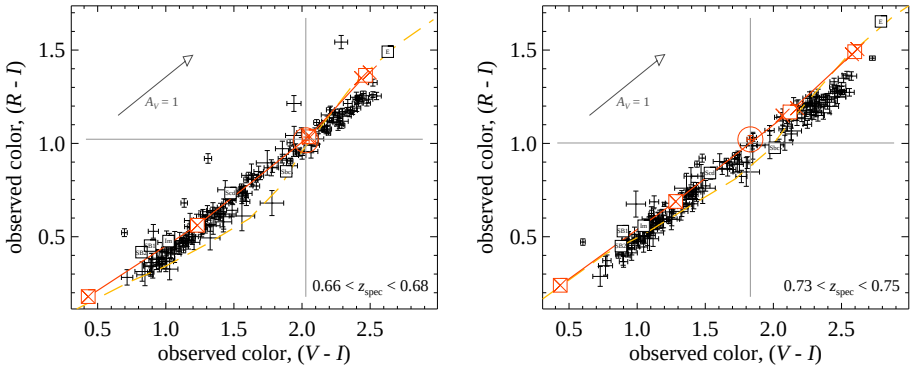


**Figure 15.** — Illustrating the InterRest algorithm for interpolating restframe fluxes. — Note that under normal circumstances, in order to interpolate a restframe flux, we would relate an observed–minus–observed color to a restframe–minus–observed color; in this case we are using the $(V-I)$ color to *predict* the flux in the observers' $R$ band. By comparing the interpolated and observed $R$ band fluxes, we will then be able to validate the algorithm (Figure 16). The algorithm works as follows: using a set of template spectra (*red crosses*), we construct a (redshift dependent) color–color relation for galaxies (*red line*); once the galaxy color–color relation has been defined, it is possible to read off the 'unknown' color (in this case, $R-I$) of any object, given its known, observed color (in this case, $V-I$). In both panels, the points with error bars show galaxies in a narrow spectroscopic redshift range, with colors measured to better than 0.05 mag; these galaxies are used in Figure 16 to validate our restframe color determinations. (See Section 8 for further discussion and explanation.)

color, and so the observed $R$ flux. In this way, we will be able to test the accuracy of the algorithm, through comparison between the predicted and observed $R$ fluxes. Even so, the example still serves to illustrate the idea behind the algorithm.

In each panel of Figure 15, the points show the observed $VRI$ colors of galaxies with spectroscopic redshifts in a narrow interval; we have selected the two most prominent $z_{\rm spec}$ spikes in the catalog, and restrict our attention to galaxies with $VRI$ colors measured to better than 0.05 mag. The red crosses in each panel show the synthetic $VRI$ colors for the default EAZY/InterRest template set, which we use to construct an approximate color–color relation for galaxies at each redshift. In both panels, the default EAZY/InterRest template spectra can be seen to do a reasonable job of describing the true color–color relation for galaxies at each of the two redshifts in question.

Now, for any individual galaxy (red point, circled; chosen at random), using the $(V-I)$ color, it is possible to read off the $(R-I)$ color (grey lines) from the synthetic color–color relation. Again, under normal circumstances, we would be relating an observed–minus–observed color to an restframe–minus–observed color; our interest here is in validating the performance of the algorithm.

As a single algorithmic detail, it is possible that the known–known colors (*i.e.*, $(V-I)$ in the above example) of two templates are very close, but for quite different known–unknown colors (*i.e.*, $(R-I)$ above): in this case, small changes in color or redshift can produce very large changes in the final result. To avoid this

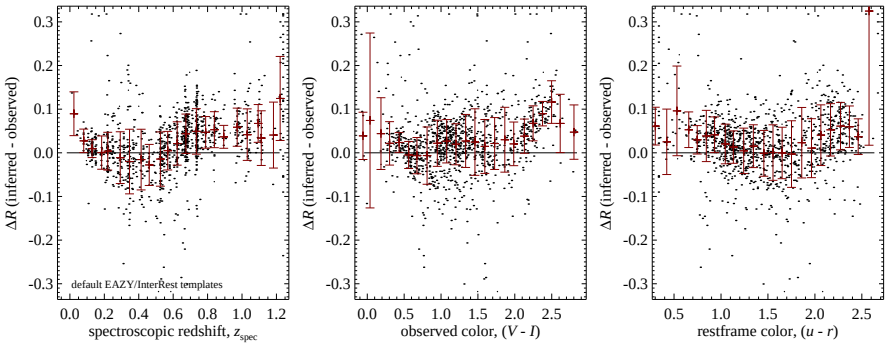


**Figure 16.** — Validating the InterRest algorithm for interpolating restframe photometry. — In each panel, we show the difference, $\Delta R$, between the $R$ band magnitude inferred from the $(V - I)$ color as in Figure 15 and that directly observed, plotted as a function of (*left to right*) spectroscopic redshift, observed color, and restframe color. The points in each panel show galaxies with robust spectroscopic redshift determinations, and $VRI$ colors measured to better than 0.05 mag; the red points with error bars show the biweight mean and scatter in $\Delta R$ as a function . The random error in the interpolated $R$ band magnitude is typically $\lesssim 0.05$ mag; comparable to the observational uncertainties themselves. Systematic uncertainties, as functions of both redshift (*i.e.* restframe wavelength) and restframe color are at the level of $\lesssim 0.05$ mag. Note, however, that the (logarithmic) wavelength interval between $V$ and $I$ is roughly twice as large as we would normally use to derive restframe photometry for real galaxies. We therefore present these numbers as upper limits on the true errors; we expect the true errors to be smaller by a factor of 2–4. That, at least for this dataset, we estimate the random and systematic errors on the interpolated restframe optical photometry to be $\lesssim 0.2$ mag.

situation, where the known–known colors are too close, we simply replace these points with their mean (in magnitude space). This can be seen in Figure 15, where the crosses show the points for the individual template spectra, and the squares show the points used to construct the color–color relation. Algorithmically, we define 'too close' as two points being separated by less than 5 % of the range (in mag) spanned by all template spectra.

In **Figure 16**, we show the differences between the $R$ fluxes interpolated as described above, and the observed $R$ fluxes in the MUSYC catalog, plotted as a function of (*left to right*), spectroscopic redshift, observed color, and restframe color. These plots are based on the $z_{\rm spec}$ compilation used in Figure 14, and described in Appendix A, but limited to those galaxies with $VRI$ colors measured to better than 0.05 mag. The black points are for individual galaxies; the red error bars show the mean error and random scatter in bins.

Both the random scatter and the systematic offset between the observed and interpolated $R$ fluxes are at the level of 0.05 mag. There are clear systematics with redshift (*i.e.* restframe wavelength), which appear to be related to the 4000 Å break. There also appears to be a problem at the level of 0.05 mag for the reddest galaxies ($u - r \gtrsim 2$). The random error in the interpolated $R$ fluxes is typically $\sim 0.05$ mag. This is comparable to the uncertainties in the photometry itself, but probably at least partially reflects the intrinsic width of the galaxy

| Column No. | Column Title | Description |
|---|---|---|
| 1 | `id` | Object identifier, beginning from 1, as in the photometric catalog |
| 2 | `redshift` | Assumed redshift; we use either the `z_m2` value output by EAZY, or the spectroscopic redshift, where available. |
| 3–17 | `RF_F1`, etc. | Restframe photometry for Bessel $UBVRI$ filters[a] |
| 18–32 | `RF_F6`, etc. | Restframe photometry for Johnson–Cousins $UBVRI$ filters |
| 33–47 | `RF_F11`, etc. | Restframe photometry for Gunn $ugriz$ filters |
| 48–54 | `RF_F16`, etc. | Restframe photometry for GALEX NUV and FUV filters |
| 55 | `distmod` | The distance modulus implied by redshift, assuming a given cosmology[b] |

**Table 5.** — Summary of the contents of the restframe photometry catalogs. — *Notes:* [a] For each object and filter, InterRest outputs two flags: `extrapN` (where `N` refers to the restframe filter number), which indicates where it has extrapolated beyond the observed SED, and `widegapN`, which indicates where it has not used neighboring filters due to, for example, missing or negative photometry. [b] Note that the fluxes output by InterRest are *observed fluxes* through *restframe filters*; that is, they have the same units as the observed, input photometry. The user must therefore perform the conversion to apparent and restframe magnitudes using the appropriate zeropoint and distance modulus.

color–color(redshift) sequence; if so, this represents a fundamental limit on the accuracy of the algorithm.

Note that whereas we would typically use two neighboring filters to interpolate restframe photometry, the wavelength span here is roughly twice as large; we therefore expect the true systematic errors in restframe fluxes (*cf.* colors) to be 2–4 times smaller than in the above example; *i.e.* at the level of 0.01–0.02 mag.

As a final aside, we note that we acheive comparable accuracies using the E, Scd, Sbc, and Im templates from Coleman, Wu & Weedman (1980), supplemented with a starburst template from Kinney et al. (1996). These templates are plotted in Figure 15 (black squares, labeled) for comparison to the default EAZY/InterRest templates. We have also tried using Bruzual & Charlot (2003) synthetic spectra, assuming Single Stellar Populations (SSPs; $\log t$ = 6.5, 7.0, ..., 10.0, 10.3 Gyr) with a Salpeter IMF and solar metallicity, and no dust extinction (shown by the dashed yellow line in Figure 15). Using BC03 spectra, we find serious systematic errors — on the level of up to 0.2 mag — both as a function of redshift, and of restframe color; this is true whether we assume a SSP or exponentially declining star formation history. These models do not reproduce the observed colors of real galaxies, and so are unsuitable for this purpose. Similarly, using the Blanton & Roweis (2007) template set, which are derived from a library of BC03 spectra with a wide range of ages and metallicities using the NMF algorithm, we find peak–to–peak systematic errors at the $\sim$ 0.1 mag level; the random errors are also at the 0.1 mag level.

In **Table 5**, we summarize the contents of the restframe photometry catalogs that we are releasing: note that we provide two separate catalogs based on photometric and spectroscopic redshift determinations, respectively.

## 9 Summary

We have described a new $K$-selected catalog of the ECDFS based on existing optical and NIR data, supplemented by original $z'JK$ imaging taken as part of the MUSYC project. The final $UU_{38}BVRIz'JHK$ photometric catalog (Section 4; Table 2) covers $\sim$ 900 $\square''$ to a ($5\sigma$, point source) limiting magnitude of $K = 22.0$ mag; note, however, that $H$ band data are available for only 80 % of the field. Included in the photometric catalog are a spectroscopic redshifts for 2914 unique objects, collected from the literature (Appendix A). In addition, we are also making available a photometric redshift catalog, derived from the MUSYC ECDFS photometry using EAZY (Section 7; Table 4), as well as catalogs of interpolated restframe photometry generated using InterRest (Section 8; Table 5).

The data described in this Chapter will form an important part of three ongoing NIR survey projects. The $K$ imaging is key for analysing the SIMPLE IRAC data (Damen et al., 2009). The broadband imaging provides the backbone for an optical medium-band survey, which will add 18 additional bands (Cardamone et al., in preparation). There is also a NEWFIRM medium band NIR survey planned, which will allow much greater photometric redshift accuracy for $z \gtrsim 1$ (van Dokkum et al., 2009). In an attempt to maximize the legacy value of our catalogs, we have invested significant time and effort in validating the absolute and relative calibration of the imaging data, as well as our analysis techniques. We summarize the results of these checks below.

*Astrometry* — The relative astrometric calibration of each band has been validated to $0\farcs15$ (0.56 pix). The absolute astrometry is accurate to $0\farcs3$, with a slight shear across the field at the level of $\lesssim 0\farcs3$ (Section 5.1).

*Completeness* — We have quantified the completeness of the catalog for sources with an $R^{1/4}$ profile in Figure 4; we present these values as lower limits on the completeness. While the catalog is formally surface brightness limited, a comparison to much deeper NIR imaging over the GOODS area of the field suggests that the catalog is more nearly flux limited. This comparison suggests that for $K = 22$, the catalog is $\sim$ 85–90 complete, and $\gtrsim$ 95 % reliable (Section 4.2).

*Photometric calibration* — While there are significant differences between the photometry in the COMBO-17 and MUSYC catalogs of the ECDFS (Section 5.2.2), a comparison between the MUSYC and GOODS photometry in the region of overlap validates the MUSYC photometry to $\lesssim 0.05$ mag (Section 5.2.1). We have refined the basic photometric calibration using the observed SEDs of main sequence stars (Section 5.2.3); we estimate that after this recalibration, the photometric cross-calibration is accurate to $\lesssim 0.02$ mag.

*Photometry* — Random and systematic photometric errors due to various aperture effects (including astrometric errors and imperfect PSF matching) are limited to $\lesssim 0.03$ mag and $\lesssim 0.006$ mag, respectively (Figure 3). We have applied corrections to SExtractor's `AUTO` flux measurements to account for missed flux and background oversubtraction; for synthetic $R^{1/4}$-law sources, these corrections typically reduce the offset between the known and recovered total fluxes by 0.05–0.10 mag (Section 4.3). We have also demonstrated that the photometric errors given

in the catalog accurately trace variations in the background rms in the NIR images (Figure 8).

*Spectroscopic Redshifts* — We have collected and collated 5374 spectroscopic redshift determinations from literature sources, of which 3815 are matched to 2914 unique sources in our catalog (Appendix A). Of these, 2213 redshifts are deigned 'secure', including 247 stars, and 1966 $z \gg 0$ galaxies.

*Photometric Redshifts* — There are some systematic discrepancies between the COMBO-17 and MUSYC photometric redshift determinations in the ECDFS, owing to the lack of NIR data in the COMBO-17 catalog; where available, spectroscopic redshifts validate the MUSYC values. The agreement between the MUSYC and GOODS-MUSIC photometric redshifts is very good, however there are a significant number of catastrophic errors in both redshift catalogs (Figure 14). In comparison to spectroscopic redshifts from the K20 survey (Cimatti et al., 2002; Mignoli et al., 2005), the random photometric redshift error is $\sigma_z = 0.033$, with an outlier fraction of 4.7 %; the outlier fraction is significantly higher for X-ray–selected spectroscopic redshift catalogs (Appendix A).

*Restframe Colors* — We have interpolated restframe photometry for the galaxies in our catalog using an IDL utility called InterRest (Section 8); we also make this utility publicly available. Estimated systematic errors in these interpolated restframe fluxes, as functions both of restframe wavelength and of galaxy color, are estimated to be $\lesssim 0.02$ mag (Figure 16). Random errors inherent to the algorithm are at a similar level.

*Recommendations for use* — Beyond the corrections to total magnitudes and zeropoints described above, the data given in the catalog have not been edited in any way; a modicum of care is therefore required when using the catalogs. To this end, there are several simple selections that we recommend. First, it is important to enforce a minimum weight criterion to ensure useful coverage in each band: we recommend using only those points with a relative weight of 0.6 or greater. This is particularly important for the $H$ band, where useful coverage is only available for $\sim 80\%$ of the field. Also, recall that exposure maps are not available for the $z'$ band; instead, the `zw` column contains a binary flag indicated whether or not the data are significantly affected by scattered light from bright stars. Further, in order to protect against false detections, we also recommend using only those sources with `Kw` $> 0.75$. To ensure against extremely poorly constrained photometric redshift solutions, we also recommend that users restrict their sample to those objects with S:N $> 5$ in the $K$-band color aperture; this cut is very efficient at eliminating those few objects that are detected only in the $K$-band. Lastly, we recommend use of the `star_flag` for identifying stars, and the `spec_flag` parameter for identifying those objects with robust spectroscopic redshifts.

The primary scientific motivation for the $K$-selected catalog that we have presented here is to characterise the properties of massive galaxies at $z \lesssim 2$, including their evolution. In Chapter III we demonstrate that this catalog is approximately complete (volume limited) for $M_* \gtrsim 10^{11}$ $M_\odot$ and $z_{\rm phot} \lesssim 1.8$, and use this catalog to quantify the $z \lesssim 2$ evolution in number density and color of

massive galaxies in general, and of red sequence galaxies in particular.

In this context, the MUSYC ECDFS dataset provides a valuable complement to existing optical surveys in the ECDFS targeting the $z \lesssim 1$; e.g. the COMBO-17 (Wolf et al., 2004) and GEMS projects (Rix et al., 2004). Further, the $z \lesssim 2$ comoving volume contained with the ECDFS field is approximately three times greater than that at $z \lesssim 3.5$ within the GOODS region in the CDFS. By allowing better sampling of rare objects, including the most massive galaxies at moderate–to high–redshifts, the MUSYC ECDFS catalog thus also complements the much deeper GOODS-CDFS data. Taken together, these combined datasets form an outstanding laboratory to study the basic properties of galaxies over nearly 90 % of the history of the universe.

**Acknowledgements.** This work was supported through grants by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), the Leids Kerkhoven-Bosscha Fonds (LKBF), and National Science Foundation (NSF) CAREER grant AST 04-49678. We thank the referee, Stefano Berta, for a close and thorough reading of the manuscript, which helped to clarify a number of points. We also wish to thank the organizers and participants of the several workshops hosted by the Lorentz Center, where many aspects of this work were developed and refined. SW gratefully acknowledges support from the W M Keck Foundation.

## Appendix

## A A Compilation of Public Spectroscopic Redshift Determinations for the MUSYC ECDFS Catalog

The ECDFS has been targeted by a number of large spectroscopic redshift campaigns, including: optical spectroscopy of the original CDFS X-ray catalog by Szokoly et al. (2004), the K20 survey (Cimatti et al., 2002; Mignoli et al., 2005), the VVDS (Le Fèvre et al., 2004), the GOODS FORS2 (Vanzella et al., 2005, 2006, 2008) and VIMOS (Popesso et al., 2008) campaigns, the IMAGES survey (Ravikumar et al., 2007), a MUSYC program targeting X-ray sources in the full ECDFS (Treister et al., 2008), and a VIMOS campaign by S Koposov (in preparation). A summary of the spectroscopic redshift resources we have used is given in **Table A.1**. Altogether, we have collected 5374 separate spectroscopic redshift determinations, of which 3815 are matched to 2914 unique objects in our catalog.

In cases where multiple spectroscopic redshift determinations/identifications are available for individual objects, our guiding principles for selecting a redshift were as follows. First, we adopt the most common redshift determination (where $\Delta z < 0.01$ is taken as agreement, and we do not consider repeat observations by the same team as an independent measurement). 574 objects in the catalog have multiple, consistent redshift determinations. Where there is no consensus, we discriminate between redshift solutions on the basis of the $Q_z$ figure of merit developed by Brammer et al. (2008), evaluated for the spectroscopic redshift. An exception to this rule is for redshifts from the X-ray selected catalogs, which do occasionally have extremely high values of $Q_z(z_{\rm spec})$, even when confirmed

by other secure determinations from other catalogs. Where $Q_z(z_{\rm spec})$ does not clearly discriminate between the possible solutions, we fall back onto the quality flags given by the different spectroscopic surveys. Note that for this purpose, we do not consider the VVDS '2' flag as 'secure'. Similarly, we give preference to the results of smaller studies, which presumably have devoted greater care on a per object basis. Reassuringly, in almost all cases, these criteria reinforce one another. Finally, we choose to adopt redshifts from sources that provide classification information where available; this means that we tend not to adopt redshifts from, for example, the VVDS catalog where other determinations are available. In this context, we consider X-ray selection as an additional piece of classification information; accordingly, we adopt redshifts from the Szokoly et al. (2004) and Treister et al. (2008) catalogs where available.

In this way, we have constructed a compendium of spectroscopic redshift determinations for 2914 unique objects in the MUSYC ECDFS catalog, including 283 spectrally-classified stars. Although all of these determinations are given in the catalog, we restrict our attention to those deigned 'secure', either by virtue of their quality flags, or through agreement between multiple sources. This leaves 2213 robust spectroscopic redshifts for objects in the MUSYC catalog; 1966 of these objects are identified as $z \gg 0$ galaxies.

**Figure A.1** shows the $z_{\rm phot}$–$z_{\rm spec}$ diagram, broken up by the $z_{\rm spec}$ source catalog, and quality flag (where possible). Within each panel, we give the NMAD and median offset in $\Delta z/(1+z)$; these values are also given in Table A.1; the grey region indicates the $3\sigma_z$ errors around the $z_{\rm phot} = z_{\rm spec}$ line. Above each $z_{\rm phot}$–$z_{\rm spec}$ diagram, we also show the distribution of each $z_{\rm spec}$ sample in observed $(J-K)$–$K$ color–magnitude space, in comparison to the full MUSYC catalog.

For 'secure' redshift determinations, the $z_{\rm phot}$–$z_{\rm spec}$ agreement is really quite good: the typical random scatter is $\sigma_z \lesssim 0.040$. Particularly for $z_{\rm spec} \lesssim 1$, we do appear to slightly underestimate galaxies' redshifts; typical systematic errors are $\Delta z/(1+z) \sim -0.025$. For the Szokoly et al. (2004) catalog, the random scatter in $z_{\rm phot}$ determinations is still quite good, but for the MUSYC spectroscopic redshift program (Treister et al., 2008), which targets brighter X-ray sources, the $z_{\rm phot}$–$z_{\rm spec}$ agreement is poor.

Further, while the outlier fraction is generally at the level of a few percent, catastrophic redshift failures appear to be a significant problem for X-ray selected sources. (Recall that we make no attempt to explicitly incorporate AGNs or QSOs in our photometric redshift calculation.) Among X-ray-selected sources, the fraction of galaxies with $|\Delta z|/(1+z) > 0.15$ is 30% (82/271); for the full gamut of robust spectroscopic redshifts, the fraction is 9% (178/1966). Said another way, 46 % (82/178) of all outliers are X-ray sources.

We draw particular attention to the comparison with the results from K20, which is highly spectrally complete in the magnitude range that we are operating in. In comparison to the K20 redshifts, we have achieved a photometric redshift accuracy of $\sigma_z = 0.034$. We also draw attention to the sample of van der Wel et al. (2005), which consists of 28 early type, red sequence galaxies at $z \sim 1$, for which we have achieved a photometric redshift accuracy of $\sigma_z = 0.022$; in fact,

| Reference(s) | Source Code | Internal Qual. Flag | No. Galaxies | No. Adopted | Median $\Delta z/(1+z)$ | NMAD $\Delta z/(1+z)$ | Outlier Fraction |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Cimatti et al. (2002); | K20 | 1 | 267 | 232 | -0.025 | 0.033 | 0.047 |
| Mignoli et al. (2005) | | 0 | 14 | 2 | -0.012 | 0.069 | 0.182 |
| Szokoly et al. (2004) | Xray | $\geq 2.0$ | 114 | 114 | -0.024 | 0.037 | 0.135 |
| | | $< 2.0$ | 17 | 4 | 0.045 | 0.146 | 0.133 |
| Le Fèvre et al. (2004) | VVDS | 4 | 172 | 131 | -0.030 | 0.027 | 0.027 |
| | | 3 | 347 | 267 | -0.030 | 0.032 | 0.035 |
| | | 2 | 342 | 19 | -0.022 | 0.058 | 0.080 |
| | | 1 | 82 | 1 | -0.003 | 0.127 | 0.017 |
| | | 9 | 49 | 1 | 0.016 | 0.199 | 0.036 |
| Vanzella et al. (2005, 2006, 2008) | GDS-F | A | 306 | 226 | -0.023 | 0.044 | 0.034 |
| | | B | 77 | 14 | -0.029 | 0.080 | 0.054 |
| | | C | 52 | 4 | 0.025 | 0.106 | 0.079 |
| Popesso et al. (2008) | GDS-V | A | 289 | 197 | -0.036 | 0.030 | 0.048 |
| | | B | 59 | 3 | -0.026 | 0.081 | 0.087 |
| | | C | 48 | 1 | -0.008 | 0.144 | 0.051 |
| Ravikumar et al. (2007) | IMAGES | 1 | 267 | 219 | -0.032 | 0.030 | 0.067 |
| | | 2 | 168 | 24 | -0.025 | 0.046 | 0.056 |
| | | 3 | 51 | 7 | -0.012 | 0.095 | 0.000 |
| Treister et al. (2008) | MUS-I | N/A | 165 | 120 | 0.001 | 0.112 | 0.125 |
| | MUS-V | N/A | 34 | 33 | 0.011 | 0.295 | 0.000 |
| S Koposov (in preparation) | Kopsv | N/A | 455 | 283 | -0.034 | 0.025 | 0.043 |
| Croom et al. (2001) | KX | N/A | 17 | 5 | -0.016 | 0.029 | 0.353 |
| Strolger et al. (2004) | SNe | N/A | 9 | 2 | — | — | — |
| van der Wel et al. (2004, 2005) | vdWel | N/A | 28 | 26 | -0.007 | 0.022 | 0.000 |
| Daddi et al. (2005) | Daddi | N/A | 5 | 5 | — | — | — |
| Doherty et al. (2005) | LCIRS | 1–3 | 14 | 10 | 0.003 | 0.050 | 0.071 |
| Kriek et al. (2006) | Kriek | N/A | 12 | 12 | 0.056 | 0.134 | 0.000 |
| Total | | | 2863 | 1966 | -0.029 | 0.036 | 0.078 |

**Table A.1.** — Summary of the spectroscopic redshifts compiled for the MUSYC ECDFS catalog. — For each spectroscopic redshift sample we have used, we give both the redshift source catalog (Col. 1) and the identifier used in the MUSYC catalog (Col. 2); further, we have broken up each sample by the internal quality flag (Col. 3), where available. For each (sub)sample, we give the number of galaxies matched to the MUSYC ECDFS catalog (Col. 4), and the number of galaxy redshifts adopted in the final catalog (Col. 5). We also give the systematic (Col. 6) and random (Col. 7) photometric redshift error, computed as the median and NMAD of $\Delta z/(1+z)$, and the outlier fraction (Col. 8), defined as the fraction of galaxies with $\Delta z/(1+z) > 0.1$; these quantities are all computed for galaxies in our main scientific sample (*i.e.* those galaxies counted in Col. 4 with useful coverage in optical and NIR bands, and with $K < 22$ and $K$ S:N $> 5$).

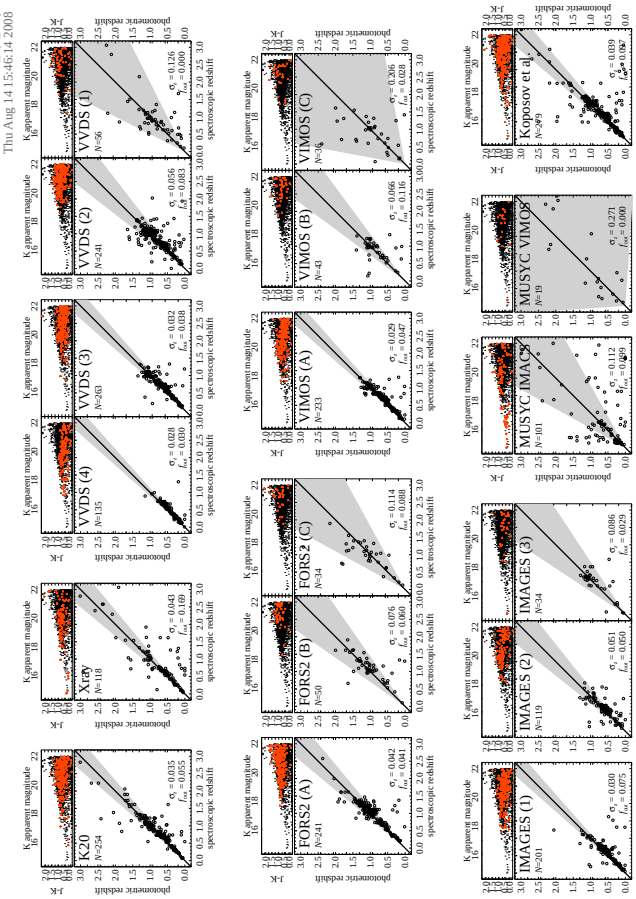


**Figure A.1.** — The $z_{\rm phot}$–$z_{\rm spec}$ diagram for individual spectroscopic redshift sources and quality flags. — We show the agreement between the MUSYC photometric redshift and many literature spectroscopic redshift catalogs, broken up by internal quality flag where available. Within each panel, we give the random photometric redshift error, $\sigma_z$, as well as the outlier fraction, which is defined as the fraction of objects with $\Delta z/(1+z) > 5\sigma_z$. The grey regions in each panel indicate the $3\sigma_z$ regions. Above each $z_{\rm phot}$–$z_{\rm spec}$ diagram, we highlight the particular $z_{\rm spec}$ sample in $(J-K)$–$K$ color–magnitude space. It is harder to obtain robust $z_{\rm spec}$ determinations for redder galaxies; these galaxies are therefore underrepresented in all $z_{\rm spec}$ samples. For this reason, we have validated our photometric redshift determinations through comparison with those from COMBO-17 and GOODS-MUSIC (see Figure 14).

this is the sample for which we have the best photometric redshift agreement.

The crucial point to be made from Figure A.1, however, is that since most of the different $z_{\rm spec}$ samples that are available in the ECDFS are not NIR-selected, they are not generally representative of the sources in our photometric catalog. For this reason, we validate our photometric redshift determinations in Section 7.3 through comparison with the COMBO-17 and GOODS-MUSIC photometric redshifts.

## References


Arnouts S, Vandame B, Benoist C, Groenewegen M A T, da Costa L, Schirmer M, Mignani R P, Slijkhuis R, Hatziminaoglou, Hook R, Madejsky R, Rité C, Wicenec A, 2001, A&A 379, 740

Baade D et al., 1998, Messenger 93, 13

Baade D et al., 1999, Messenger 95, 15

Beckwith S V W et al., 2006, AJ 132, 1729

Benítez N, 2000, ApJ 536, 571

Berta S et al., 2006, A&A 451, 881

Bertin E & Arnouts S, 1996, A&A 117, 393

Blanton M R & Roweis S, 2007, AJ 133, 734

Blanc G A et al., 2008, ApJ 681, 1099

Bolzonella M, Miralles J-M, Pelló R, 2000, A&A 363, 476

Brammer G B, van Dokkum P G, Coppi P, 2008, ApJ 686, 1503

Brown M J I, Dey A, Jannuzi B T, Brand K, Benson A J, Brodwin M, Croton D J, Eisenhardt P R, 2007, ApJ 654, 858

Bruzual G, Charlot S, 2003, MNRAS 344, 1000

Chen H-W et al., 2002, ApJ 570, 50

Cimatti A, Mignoli M, Daddi E, Pozzetti L, Fontana A, Saracco P, Poli F, Renzini A, Zamorani G, Broadhurst T, Cristiani S, D'Odorico S, Giallongo E, Gilmozzi R, Menci N, 2002, A&A 392, 395

Coleman G D, Wu C C, Weedman D W, 1980, ApJS 43, 393

Croom S M, Warren S J, Glazebrook K, 2001, MNRAS 328, 150

Cutri R M, Skrutskie M F, van Dyke S, et al., 2003, 'Explanatory Supplement to the 2MASS All Sky Data Release and Extended Mission Products', NASA/IPAC Infrared Science Archive, http://www.ipac.caltech.edu/2mass/releases/allsky/doc/

Daddi E, Cimatti A, Renzini A, Fontana A, Mignoli M, Pozzetti L, Tozzi P, Zamorani G, 2004, ApJ 617, 746

Daddi E et al., 2005, ApJ 626, 680

Damen M et al., 2009, ApJ (submitted)

De Lucia G & Blaizot J, 2007, MNRAS 375, 2

Dickinson M E, Giavalisco M, et al., 2002, arXiv:astro-ph/0204213v1

Doherty M, Bunker A J, Ellis R S, McCarthy P J, 2005, MNRAS 361, 525

Erben T et al., 2005, AN 326, 432

Fasano G, Filippi M, Bertola F, 1998, A&AS 129, 583

Förster-Schreiber N M, Franx M, Labbé I, Rudnick G, van Dokkum P G, Illingworth G D, Kuijken K, Moorwood A F M, Rix, HW., Röttgering H, van der Werf P, 2006, AJ 131, 1891

Franx M et al., 2003, ApJ 587, L79

Gawiser E et al., 2006a, ApJS 162, 1

Gawiser E et al., 2006b, ApJ 642, L13

Giaconni R, 2002, ApJS 139, 369

Girard T M, Dinescu D L, van Altena W F, Platais I, Monet D G, López C E, AJ 127, 5

Grazian A, Fontana A, De Santis C, Nonino M, Salimbeni R, Giallongo E, Cristiani S, Gallozzi S, Vanzella E, 2006, A&A 449, 951

Gronwall C et al., 2007, ApJ, 667. 79

Gunn J E & Stryker L L, 1983, ApJS 52, 121

Häussler B, et al., 2007, ApJS 172, 615

Hildebrandt H et al., 2006, A&A 452, 1121

Hildebrandt H, Wolf C, Benítez N, 2008, A&A, 480, 703

Kinney A L, Calzetti D, Bohlin R C, McQuade K, Storchi-Bergmann T, Schmitt H R, 1996, ApJ 467, 38

Kriek M, et al., 2006, ApJ, 649, L71

Labbé I, Franx M, Förster Schreiber N M, Rix H-W, Moorwood A, van Dokkum P G, van der Werf P, Röttgering H, van Starkenburg L, van der Wel A, Kuijken K, Daddi E, 2003, AJ 125, 1107

Lawrence A, et a., 2007, MNRAS 379, 1599

Le Fèvre O et al., 2004, A&A 428, 1043

Lehmer B D et al., 2005, ApJS 161, 21

McCarthy P J, et al., 2001, ApJ 560, L131

Mignoli M, Cimatti A, Zamorani G, Pozzetti L, Daddi E, Renzini A, Broadhurst T, Cristiani, D'Odorico S, Fontana A, Giallongo E, Gilmozzi R, Menci N, Saracco P, 2005, A&A 437, 883

Moy E et al., 2003, A&A 403, 493

Moorwood A, Cuby J-G, Lidman C, 1998, The Messenger 91, 9

Muller G P, Reed R, Armandroff T, Boroson T A, Jacoby G A, 1998, Proc. SPIE 3355, 577

Olsen L F et al., 2006, A&A 456, 881

Pickles A J, 1998, PASP 110, 863

Probst R G, Montane A, Warner M, Boccas M, Bonati M, Galvez R, Tighe R, Ashe M C, van der Bliek N S, Blum R D, 2003, APIE 4841,

411

Popesso P et al., 2009, A&A 494, 443

Quadri R et al., 2007, AJ 134, 3

Ravikumar C D et al., 2007, A&A 465, 1099

Rix H-W, et al., 2004, ApJS 152, 163

Rudnick G, Franx M, Moorwood A, Kuijken K, van Starkenburg L, van der Werf P, Röttgering H, van Dokkum P G, Labbé I, 2001, AJ 122, 2205

Rudnick G, Rix H-W, Franx M, Labbé I, Blanton M, Daddi E, Förster-Schreiber N M, Moorwood A, Röttgering H, Trujillo I, van der Wel A, van der Werf P, van Dokkum P G, van Starkenburg L, 2003, ApJ 599, 847

Skrutskie M F, Cutri R M, Stiening R, Weinberg M, Schneider S, Carpenter J M, et al., 2006, ApJ 131, 1163

Springel V et al., 2005, Nature 435, 629

Steidel C C, Giavalisco M, Pettini M, Dickinson M, Adelburger K, 1996, ApJ 462, L17

Strolger L-G, et al., 2004, ApJ 613, 200

Szokoly G P et al., 2004, ApJS 155, 271

Treister E, Virani S, Gawiser E, Urry C M, Lira P, Francke H, Blanc G A, Cardamone C N, Damen M, Taylor E N, Schawinski K, 2009, ApJ 693, 1713

Vandame B et al., 2001, astro-ph/0102300

van der Bliek N S, Norman D, Blum R D, Probst R G, Montane A, Galvez R, Warner M, Tighe R, Delgado F, Martinez M, 2004, SPIE 5492, 1582

van der Wel A, Franx M, van Dokkum P G, Rix H-W, 2004, ApJ 601, L5

van der Wel A, Franx M, van Dokkum P G, Rix H-W, Illingworth G D, Rosatti P, 2005, ApJ 631, 145

van Dokkum P G, Labbé I, Marchesini D, Quadri R, Brammer G et al., 2009, PASP 121, 2

Vanzella et al., 2005, A&A 434, 53

Vanzella et al., 2006, A&A 454, 423

Vanzella et al., 2008, A&A 478, 83

Warren S J, et al., 2007, MNRAS 375, 213

Virani S N, Treister E, Urry C N, Gawiser R, 2006, AJ 131, 2373

Wolf C et al., 2004, A&A 421, 913

Wolf C, Hildebrandt H, Taylor E N, Meisenheimer K, 2008, A&A 942, 933

Wuyts S, Labbé I, Förster Schreiber N M, Franx M, Rudnick G, Brammer G B, van Dokkum P G, 2008, ApJ, 682, 985

# Chapter III

## The Rise of Massive Red Galaxies: the Color–Magnitude and Color–Stellar Mass Diagrams for $z_{\rm phot} \lesssim 2$

We present the color–magnitude and color–stellar mass diagrams for galaxies with $z_{\rm phot} \lesssim 2$, based on a $K^{(\rm AB)} < 22$ catalog of the $\frac{1}{2} \times \frac{1}{2}$ $\Box^\circ$ Extended Chandra Deep Field South (ECDFS) from the MUltiwavelength Survey by Yale–Chile (MUSYC). Our main sample of 7840 galaxies contains 1297 $M_* > 10^{11}$ $\rm M_\odot$ galaxies in the range $0.2 < z_{\rm phot} < 1.8$. We show empirically that this catalog is approximately complete for $M_* > 10^{11}$ $\rm M_\odot$ galaxies for $z_{\rm phot} < 1.8$. For this mass-limited sample, we show that the locus of the red sequence color–stellar mass relation evolves as $\Delta(u-r) \propto (-0.44 \pm 0.02)\, z_{\rm phot}$ for $z_{\rm phot} \lesssim 1.2$. For $z_{\rm phot} \gtrsim 1.3$, however, we are no longer able to reliably distinguish red and blue subpopulations based on the observed optical color distribution; we show that this would require much deeper near infrared (NIR) data. At $1.5 < z_{\rm phot} < 1.8$, the comoving number density of $M_* > 10^{11}$ $\rm M_\odot$ galaxies is $\approx 50\%$ of the local value, with a red fraction of $\approx 33$ %. Making a parametric fit to the observed evolution, we find $n_{\rm tot}(z) \propto (1+z_{\rm phot})^{-0.52 \pm 0.12 (\pm 0.20)}$. We find stronger evolution in the red fraction: $f_{\rm red}(z) \propto (1+z_{\rm phot})^{-1.17 \pm 0.18 (\pm 0.21)}$. Through a series of sensitivity analyses, we show that the most important sources of systematic error are: 1. systematic differences in the analysis of the $z \approx 0$ and $z \gg 0$ samples; 2. systematic effects associated with details of the photometric redshift calculation; and 3. uncertainties in the photometric calibration. With this in mind, we show that our results based on photometric redshifts are consistent with a completely independent analysis which does not make use of any redshift information for individual galaxies. Our results suggest that, at most, 1/5 of local red sequence galaxies with $M_* > 10^{11}$ $\rm M_\odot$ were already in place at $z \sim 2$.

Taylor E N, Franx M, van Dokkum P G, Bell E F, Brammer G B, Rudnick G, Wuyts S, Gawiser E, Lira P, Urry C M, Rix H-W
*The Astrophysical Journal*, **694** 1171–1199 (2009)
(submitted July 2008, published March 2009)

## 1 Introduction

Observing the evolution of the massive galaxy population provides basic constraints on cosmological models of structure formation, and so helps to identify the physical processes that govern the formation and evolution of massive galaxies. In this context, the color–magnitude diagram (CMD) — astronomy's most basic diagnostic plot — has been particularly important and useful over the past five years. Physically, a galaxy's restframe color is determined by the (luminosity weighted) mean stellar age, modulo the mean stellar metallicity and extinction from dust in the ISM. The restframe optical brightness acts as a proxy for the total stellar mass, although the connection between the two has a similarly complicated dependence on star formation history, metallicity, and dust. The CMD thus offers two complementary means of characterizing the star formation history of individual galaxies, in terms of the amount and character of their starlight.

In the local universe, galaxies can be separated into two distinct but overlapping populations in color–magnitude space (see, e.g., Baldry et al., 2004): a relatively narrow and well-defined 'red sequence', as distinct from the more diffuse 'blue cloud', with each following its own color–magnitude relation (CMR). Red sequence galaxies dominate the bright galaxy population, and tend to have the more concentrated light distributions typical of morphologically early type galaxies (Strateva et al., 2001; Blanton et al., 2003; Driver et al., 2006; van der Wel, 2008). They typically have stellar masses greater than $10^{10.5}$ $M_{\odot}$ and are dominated by old stars, whereas blue cloud galaxies are typically less massive and continue to be actively star forming (Kauffmann et al., 2003a; Brinchmann et al., 2004; Wyder et al., 2007). Further, red sequence galaxies lie preferentially in higher density environments (Hogg et al., 2003; Blanton et al., 2005b; Baldry et al., 2006). The emergent picture is of a population of massive, quiescent, concentrated, and strongly clustered red sequence galaxies, as distinct from the typically less massive, disk-dominated, and star forming blue cloud population (Ellis et al., 2005). This Chapter focuses on the redshift evolution of the red sequence galaxy population.

Using high quality photometric redshifts from the COMBO-17 survey, Bell et al. (2004b) showed that a red galaxy sequence is already in place at $z_{\rm phot} \sim 1$ (see also, e.g., Im et al., 2002; Weiner et al., 2005; Willmer et al., 2006). Further, as in the local universe, the $z_{\rm phot} \approx 0.7$ red sequence is dominated by passive, morphologically early type galaxies (Bell et al., 2004a). The combined mass of red sequence galaxies at $z \sim 1$ is at least half of the present day value (Bell et al., 2004b; Faber et al., 2007; Brown et al., 2008). By contrast, the stellar mass density of actively star forming blue cloud galaxies remains more or less constant for $z \lesssim 1$ (Arnouts et al., 2007; Bell et al., 2007), even as the combined star formation rate drops by an order of magnitude over the same interval (Lilly et al., 1996; Madau et al., 1996; Hopkins, 2004). These results — a steadily growing number of passively evolving red galaxies, and a relatively constant number of actively star forming blue galaxies — have led to the idea of a quenching mechanism for star formation, operating to incite a transformation that moves active galaxies from the blue cloud onto the passive red sequence (Menci et al, 2005; Croton et al., 2006; Cattaneo et

al, 2006; Dekel & Birnboim, 2006; De Lucia et al., 2007).

Our specific goal in this Chapter is to quantify the $z_{\rm phot} \lesssim 2$ evolution of massive galaxies in general, and of red sequence galaxies in particular, in the color–magnitude and color–stellar mass planes. The $1 \lesssim z \lesssim 2$ interval is particularly interesting: whereas the $z \sim 1$ galaxy population appears qualitatively similar to the local universe (at least in terms of the existence and properties of red sequence galaxies) the situation at $z \gtrsim 2$ may be quite different. While massive, passive galaxies have been confirmed at $z \gtrsim 1.5$ (Daddi et al., 2005; McGrath et al., 2007) and even $z \gtrsim 2$ (Kriek et al., 2006), these galaxies do not appear to dominate the massive galaxy population as they do at $z \lesssim 1$. Indeed, it appears that the median massive galaxy at $z \sim 2$ has the infrared luminosity of a LIRG or ULIRG (Reddy et al., 2006). Moreover, whereas the number density of massive galaxies at $z \sim 1$ is $\gtrsim 50$ % of the local value (Juneau et al., 2005; Borch et al., 2006; Scarlata et al., 2007), at $z \gtrsim 2$ it is inferred to be $\lesssim 15$ % (Fontana et al., 2006; Arnouts et al., 2007; Pozzetti et al., 2007; Pérez-Gonzaléz et al., 2008). This marks the redshift interval $1 \lesssim z \lesssim 2$ as potentially being an era of transition in the universe, in which massive galaxies first begin both to appear in large numbers, and to take on the appearance of their local antecedents. This coincides with end of the period of peak star formation in the universe; while the cosmic star formation rate rises sharply for $z \lesssim 1$, it appears to plateau or even peak for $z \gtrsim 2$ (see, e.g., Hopkins, 2004; Nagmine et al., 2006; Panter et al., 2007; Tresse et al., 2007; Pérez-Gonzaléz et al., 2008). Whatever the mechanism that quenches star formation in massive galaxies may be, it is in operation at $1 < z < 2$.

The technical key to gaining access to the $1 \lesssim z \lesssim 2$ universe is deep near infrared (NIR) data (Connolly et al., 1997). Both spectroscopic and photometric redshift and stellar mass determinations rely primarily on broad spectral features which lie in the restframe optical; for $z \gtrsim 1$, these features are shifted beyond the observer's optical window and into the NIR. Moreover, the inclusion of NIR data makes it possible to construct stellar mass limited samples with high completeness (van Dokkum et al., 2006). Among the next generation of NIR-selected cosmological field galaxy surveys, the MUltiwavelength Survey by Yale–Chile (MUSYC; Gawiser et al. 2006) is among the first to become public. MUSYC has targeted four widely dispersed Southern fields, covering a total of one square degree in the $UBVRIz'$ bands. Coupled with this optical imaging program, MUSYC also has two NIR imaging campaigns: a wide ($K^{(\rm AB)} \lesssim 22$) component over three of the four fields Blanc et al. (2008, ; Chapter II), and a deeper ($K^{(\rm AB)} \lesssim 23.5$) component for four 10' × 10' fields (Quadri et al., 2007).

This present Chapter focuses on the Extended Chandra Deep Field South (ECDFS), one of the four $\frac{1}{2} \times \frac{1}{2}$ $\square^\circ$ MUSYC fields. Centered on the historical Chandra Deep Field-South ($\alpha = 03^{\rm h}32^{\rm m}28{\rm s}$, $\delta = -27^\circ 48' 30''$; J2000; Giacconi et al., 2001), this is one the best studied fields on the sky, with observations spanning the full electromagnetic spectrum from the X-ray to the radio. Notably, this field is also a part of the COMBO-17 survey (Wolf et al., 2004), and has received *Hubble Space Telescope* ACS coverage as part of the GEMS project (Rix et al.,

| Band | $\lambda_0$ [Å] | Instrument | Exp. Time [min.] | Area [□'] | FWHM | $5\sigma$ depth (AB) |
|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $U$ | 3560 | WFI | 1315 | 973 | $1\farcs1$ | 26.5 |
| $U_{38}$ | 3660 | WFI | 825 | 942 | $1\farcs0$ | 26.0 |
| $B$ | 4600 | WFI | 1157 | 1011 | $1\farcs1$ | 26.9 |
| $V$ | 5380 | WFI | 1743 | 1020 | $1\farcs0$ | 26.6 |
| $R$ | 6510 | WFI | 1461 | 1016 | $0\farcs9$ | 26.3 |
| $I$ | 8670 | WFI | 576 | 976 | $1\farcs0$ | 24.8 |
| $z'$ | 9070 | MOSAIC-II | 78 | 997 | $1\farcs1$ | 24.0 |
| $J$ | 12500 | ISPI | $\approx 80$ | 882 | $< 1\farcs5$ | 23.3 |
| $H$ | 16500 | SofI | $\approx 60$ | 650 | $< 0\farcs8$ | 23.0 |
| $K$ | 21300 | ISPI | $\approx 60$ | 887 | $< 1\farcs0$ | 22.5 |

**Table 1.** — Summary of the data comprising the MUSYC ECDFS catalog. — For each band, we give the filter identifier (Col. 1) and effective wavelength (Col. 2), as well as the detector name (Col. 3). For the NIR data, exposure times (Col. 4) are given per pointing; the effective seeing (Col. 6) is given for the pointing with the broadest PSF. The $5\sigma$ limiting depths (Col. 7) are as measured in $2\farcs5$ diameter apertures; the smallest apertures we use. The references for each set of imaging data are given in the main text (§2.1).

2004), as well as extremely deep *Spitzer Space Telescope* imaging from the SIMPLE (Damen et al., 2008) project. Further, the GOODS project (Dickinson et al., 2002) covers the 160 □' at the centre of this field, including supporting NIR data from the ISAAC instrument on the VLT (Grazian et al., 2006; Wuyts et al., 2008). Complementing these and other imaging surveys, a wealth of spectroscopic redshifts are available from large campaigns including the K20 survey (Cimatti et al., 2002), the VVDS project (Le Fèvre et al., 2004), the two GOODS spectroscopic campaigns (Vanzella et al., 2008; Popesso et al., 2009), and the IMAGES survey (Ravikumar et al., 2006), among others.

The plan for this Chapter is as follows. We begin in Section 2 by giving a brief overview of the data used in this Chapter — both the $z \gg 0$ MUSYC ECDFS dataset, and the $z \approx 0$ comparison sample from Blanton et al. (2005b). Next, in Section 3, we describe our basic methods for deriving redshifts and restframe properties for $z \gg 0$ galaxies; our analysis of the $z \approx 0$ comparison sample is described separately in Appendix A. In Section 4, we construct a stellar mass-limited sample of massive galaxies from our $K$-selected catalog.

Our basic results — the color–magnitude and color–stellar mass diagrams for $z_{\rm phot} < 2$ — are presented in Section 5. We then analyze three separate aspects of the data: evolution in the color distribution of the massive galaxy population in (Section 6); the color evolution of the red galaxy population as a whole (Section 7); and the $z_{\rm phot} \lesssim 2$ evolution in the absolute and relative numbers of massive red/blue galaxies (Section 8). Our final results are in conflict with those from COMBO-17 in the same field; in Appendix B, we show that this is a product of calibration errors in the COMBO-17 data, rather than differences in our analyses. In Section 9, we present a series of sensitivity analyses in which we repeat our analysis a number of times, while varying individual aspects of our experimental

design, and seeing how these changes affect our results; this tests thus enable us to identify and quantify the most important sources of systematic uncertainty in our main results.

Finally, in Section 10, we present a completely independent consistency check on our results: we measure the $z \lesssim 2$ evolution of the relative number of bright, red galaxies based only on directly observed quantities — that is, without deriving redshifts or stellar masses for individual galaxies. A summary of our results and conclusions is given in Section 11.

Throughout this work, magnitudes are expressed in the AB system; exceptions are explicitly marked. All masses have been derived assuming a 'diet Salpeter' IMF (Bell & de Jong, 2001), which is defined to be 0.15 dex less massive than a standard Salpeter (1955) IMF. In terms of cosmology, we have assumed $\Omega_\Lambda = 0.70$, $\Omega_m = 0.30$, and $H_0 = 70\ h_{70}$ km s$^{-1}$ Mpc$^{-1}$, where $h_{70} = 1$.

## 2 Data

### 2.1 An Overview of the MUSYC ECDFS Dataset

This work is based on a $K$-selected catalog of the ECDFS from the MUSYC wide NIR imaging program; these data are described and presented in Chapter II. We will refer hereafter to this dataset as 'the' MUSYC ECDFS catalog, although it should be distinguished from the optical ($B$+$V$+$R$)-selected catalog, as well as the narrow band (5000 Å)-selected photometric catalogs described by Gronwall et al. (2007), and the spectroscopic catalog described by Treister et al. (2008).

The vital statistics of the imaging data that have gone into the MUSYC ECDFS catalog are summarized in **Table 1**. Unlike the three other MUSYC fields, the ECDFS dataset was founded on existing, publicly available optical imaging: specifically, archival $UU_{38}BVRI$ WFI data, [1] including those taken as part of the COMBO-17 survey (Wolf et al., 2004) and ESO's Deep Public Survey (DPS; Arnouts et al., 2001), which have been rereduced as part of the GaBoDS project (Erben et al., 2005; Hildebrandt et al., 2006). We also include $H$ band imaging (P Barmby, 2006; private communication) from SofI on the 3.6m NTT, covering $\sim 80$ % of the field, taken to complement the ESO DPS data, reduced and described by Moy et al. (2003).

#### *2.1.1 Original Data Reduction and Calibration*

These existing data have been supplemented with original $z'$ band imaging from MOSAIC-II, reduced as per Gawiser et al. (2006), as well as $J$ and $K$ band imaging from ISPI; both instruments are mounted on the Blanco 4m telescope at CTIO. To cover the full $\frac{1}{2} \times \frac{1}{2}$ □° ECDFS in the $JK$ bands, we have constructed a mosaic of nine $\sim 10 \times 10$ □' subfields (the size of the ISPI field of view). The data reduction

[1] Two separate WFI $U$ filters have been used. The first, ESO#877, which we refer to as the $U$ filter, is slightly broader than a Broadhurst $U$ filter, and has ($\lambda_0$, $\Delta\lambda$) = (3400 Å, 732 Å). This filter is known to have a red leak beyond 8000 Å. The second filter, ESO#841, is something like a narrow Johnson $U$ filter, with ($\lambda_0$, $\Delta\lambda$) = (3637 Å, 383 Å), and which we refer to as $U_{38}$.

for the $JK$ imaging closely follows Quadri et al. (2007) and Blanc et al. (2008), and is described in detail in Chapter II, where we present the MUSYC ECDFS catalog.

In brief, to facilitate multiband photometry, each reduced image has been shifted to a common astrometric reference frame ($0\farcs267$ pix$^{-1}$). The relative astrometry has been verified to $0\farcs15$ (0.56 pix). To combat aperture effects (*i.e.* similar apertures capturing different fractions of light, due to variable seeing across different images), we have PSF-matched our images to the one with the worst effective seeing. Among the $K$ band pointings, the worst effective seeing is $1\farcs0$ FWHM; this sets our limits for detection and for total $K$ band flux measurements. Among the other bands, the worst seeing is $1\farcs5$ FWHM in the Eastern $J$ subfield; this sets the limit for our multicolor photometry. After PSF matching, systematic errors due to aperture effects are estimated to be $\lesssim 0.006$ mag for the smallest apertures we use.

In Section 5.2 of Chapter II, we have tested the photometric calibration through comparison with the COMBO-17 catalog of the ECDFS (Wolf et al., 2004), and with the FIREWORKS catalog of the GOODS-CDFS region (Wuyts et al., 2008). While there are significant differences between the COMBO-17 and MUSYC photometry, the comparison to FIREWORKS validates our photometry and photometric calibration to $\lesssim 0.02$ mag in most cases, particularly for the redder bands. Further, we have tested the relative calibration of all bands using the observed colors of stars; this test validates the photometric cross-calibration to $\lesssim 0.05$ mag. (See Section 9.1 for a discussion of how sensitive our main results are to photometric calibration errors.)

### *2.1.2 Photometry*

The photometry itself was done using SExtractor (Bertin & Arnouts, 1996) in dual image mode, using the $1\farcs0$ FWHM $K$ mosaic as the detection image. Note that, as we were unable to find a combination of SExtractor background estimation parameters (for the detection phase) that were fine enough to map real spatial variations in the background level, but still coarse enough to avoid being influenced by the biggest, brightest sources, we were forced to perform our own background subtraction for the NIR images. Total fluxes were measured from this $1\farcs0$ $K$ image, using SExtractor's `FLUX_AUTO`. In Chapter II, we that (in the photometry phase) SExtractor systematically overestimates the background flux level by $\sim 0.03$ mag; we have taken steps to correct for this effect. Following Labbé et al. (2003), we also apply a minimal correction to account for missed flux beyond the finite `AUTO` aperture, treating each object as though it were a point source. We quantify the impact these two corrections have on our final results in Sections 9.2.1 and 9.2.2.

Multicolor spectral energy distributions (SEDs) were constructed for each object using the larger of SExtractor's `ISO` aperture and a $2\farcs5$ diameter circular aperture, measured from the $1\farcs5$ FWHM $U$–$K$ images; we then normalize each object's SED using the total $K$ flux. This flexibility in aperture size is important to compromise between using apertures that are small enough to optimize signal–to–noise (the $2\farcs5$ diameter aperture is close to optimal in terms of S:N for

a point source in the 1.″5 FWHM $J$ band image), but also large enough to account for color gradients, which are important for the nearest, brightest objects. (See Section 9.2.3 for a discussion of how our results vary using only fixed aperture photometry to construct SEDs.)

Photometric errors (accounting for sky noise, imperfect background subtraction, etc., as well as the pixel–pixel correlations introduced at various stages in the reduction process) were derived empirically by placing large numbers of 'empty' apertures on each image (see also Labbé et al., 2003; Gawiser et al., 2006; Quadri et al., 2007). For the $J$ and $K$ bands, this was done for each subfield individually.

### *2.1.3 Completeness and Reliability*

We have assessed the completeness of the MUSYC catalog of the ECDFS in two ways (Chapter II). First, we have tested our ability to recover synthetic, $R^{1/4}$ profile sources of varying total flux and size introduced into empty regions of the data, using procedures identical to 'live' detection. This analysis suggests that, at $K = 22$, the MUSYC catalog should be 100 % complete for point sources, dropping to 64 % for $R_e = 0.''6$ (2.25 pix, or 4.8 kpc at $z = 1$) ellipticals. Secondly, we have compared our catalog to the much deeper FIREWORKS (Wuyts et al., 2008) catalog of the CDFS-GOODS region. In the region of overlap, for $21.8 < K \leq 22.0$ bin, $\gtrsim$ 85 % of FIREWORKS detections are found in the MUSYC catalog; all MUSYC detections in this bin are confirmed in the FIREWORKS catalog. Taken together, these two analyses suggest that, at our limiting magnitude of $K = 22$, the MUSYC catalog is primarily magnitude (*cf.* surface brightness) limited, $\gtrsim$ 85 % complete, and $\sim$ 100 % reliable.

### *2.1.4 Sample Selection*

In constructing our main galaxy sample, we have identified stars on the basis of the $(B-z')$–$(z'-K)$ color–color diagram. This selection performs extremely well in comparison with both COMBO-17's SED classification, and with GEMS point sources (Chapter II). We also make three further selections. First, to protect against false detections in regions with lower weights (e.g. the mosaic edges, and exposure 'holes' in the Eastern $K$ subfield), we require the effective weight in $K$ to be greater than 75 %, corresponding to an effective exposure time of 45 min or more. Secondly, we have masked out, by hand, regions around bright stars where the SEDs of faint objects may be heavily contaminated; this problem is most severe in the $z'$ band, where the PSF has broad wings. With these two selections, the effective area of the catalog becomes 818 □'. Thirdly, to protect against extremely poorly constrained redshift solutions, we will limit our analysis to those objects with S:N $> 5$ for the $K$ SED point. (See Section 9.3.1 for a discussion of how this selection impacts our results.)

Of the 16910 objects in the MUSYC ECDFS catalog, these selections produce 10430 reliable $K \leq 22$ detections, of which 9520 have reliable photometry in $UU_{38}BVRIz'JK$. Of these, 8790 cataloged objects have $K$ S:N $> 5$; 950 of these objects are excluded as stars, leaving 7840 galaxies in our main sample.

### 2.2 The $z \approx 0$ Comparison Sample

We will investigate the $z_{\rm phot} \lesssim 2$ evolution in the massive galaxy population by comparing the situation at $z \gg 0$ to that for $z \approx 0$ galaxies in the Sloan Digital Sky Survey (SDSS; York et al., 2000; Strauss et al., 2002); specifically, we use the 'low-z' sample from the New York University (NYU) Value Added Galaxy Catalog (VAGC) of the SDSS presented by Blanton et al. (2005b). The (Data Release 4) low-z catalog contains *ugriz* photometry for 49968 galaxies with $10 < D < 150$ Mpc ($z_{\rm spec} \lesssim 0.05$), covering an effective area of 6670 $\square^{\circ}$; 2513 of these galaxies have $M_* > 10^{11}$ $M_{\odot}$. Our analysis of these data closely follows that of the $z \gg 0$ sample, and is described separately in Appendix A.

## 3 Photometric Redshifts and Restframe Properties

### 3.1 Photometric Redshifts

The technical crux on which any photometric lookback study rests is the determination of redshifts from broadband SEDs. We have computed our photometric redshifts using EAZY (Brammer et al., 2008), a new, fully-featured, user-friendly, and publicly-available photometric redshift code. By default, the $z_{\rm phot}$ calculation is based on all ten bands, although we do require that the effective weight in any given band is greater than 0.6; in practice, this requirement only affects the $H$ band, where we do not have full coverage of the field. The characteristic filter response curves we use account for both atmospheric extinction and CCD response efficiency as a function of wavelength.

For our fiducial or 'default' analysis, we simply adopt the recommended default settings for EAZY: *viz.*, we adopt the EAZY default redshift and wavelength grids, template library, template combination method, template error function, $K$ luminosity prior, etc. (See Brammer et al., 2008, for a complete description.) Note that, by default, EAZY assigns each object a redshift by marginalizing over the full probability distribution rather than, say, through $\chi^2$ minimization. For this work, a key feature of EAZY is the control it offers over how the SED fitting is done: the user is able to specify whether and how the basic template spectra are combined, whether or not to include luminosity prior and/or a template error function, and how the output redshift is chosen. We will make use of EAZY's versatility in Section 9.4 to explore how these particular choices affect our results.

One of the unique aspects of the ECDFS is the high number of publicly available spectroscopic redshift determinations, which can be used to validate and/or calibrate our photometric redshifts. In Appendix A of Chapter II, we describe a compilation of spectroscopic redshifts for 2914 unique objects in our catalog, including 'robust' redshifts for 1656 galaxies in our main $K < 22$ sample[2]. These redshifts come from some of the many literature sources available in the ECDFS,

[2]In the end, this Chapter was finished and published *before* Chapter II. In particular, the full compilation of $z_{\rm spec}$s had not been completely finalized. Here, by 'robust', what is meant is only those $z_{\rm spec}$ determinations with high quality flags; it does not include those determinations that were declared 'robust' by virtue of independent confirmation by multiple authors. There are thus small differences in the values of $\sigma_z$ given in this Chapter and in Chapter II.

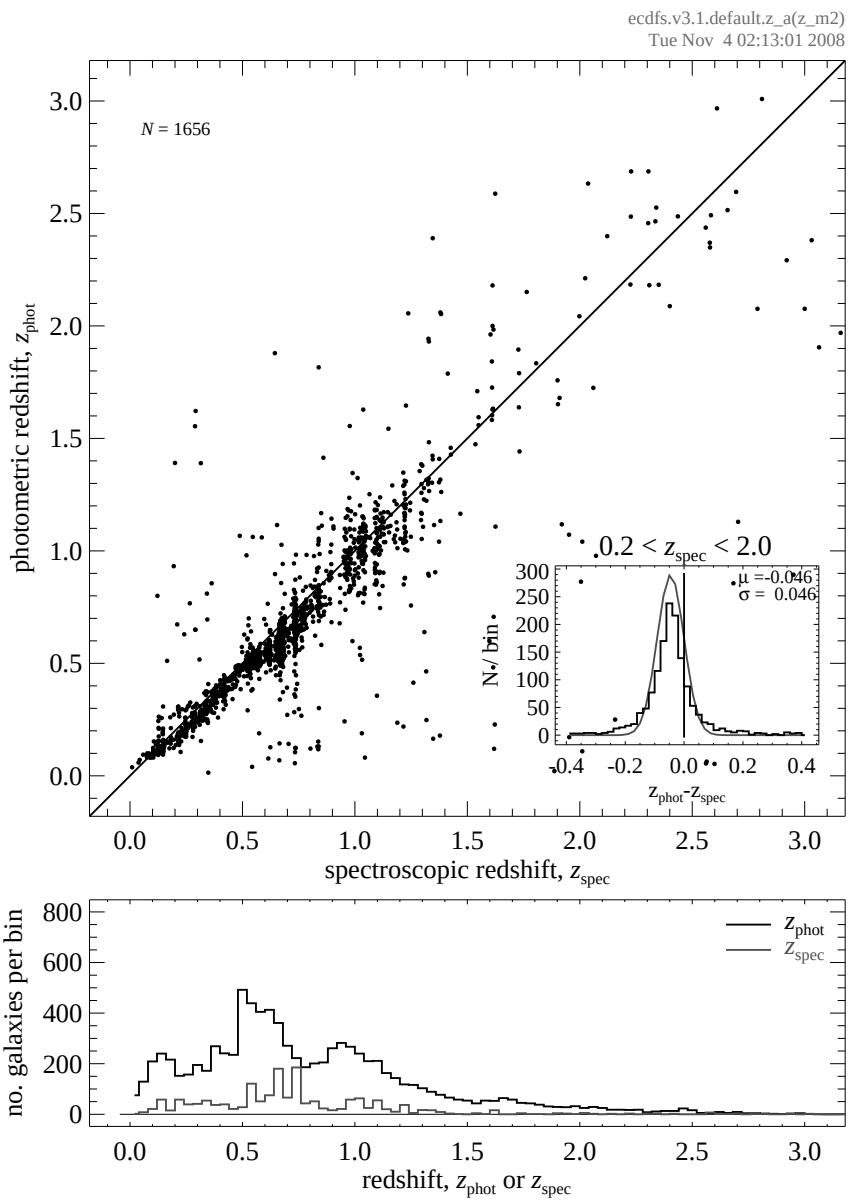


**Figure 1.** — Validating the MUSYC catalog photometric redshift determinations. — *Main panel*: the $z_{\rm spec}$–$z_{\rm phot}$ diagram for the 1656 galaxies from our main $K < 22$ sample, using a compendium of 'robust' spectroscopic redshifts from the literature (see Chapter II). *Inset*: the distribution of $\Delta z = (z_{\rm phot} - z_{\rm spec})$ for the same set of galaxies; the curve shows a Gaussian fit to this distribution, with parameters as given. *Lower panel*: the redshift distributions of the main and spec-$z$ samples; our photometric redshifts appear to mildly underestimate the redshift of the three overdensities at $0.5 < z_{\rm spec} < 0.8$. Quantitatively, we find the median and NMAD of $\Delta z/(1 + z_{\rm spec})$ for the full spec-$z$ sample to be $-0.029$ and 0.036, respectively (See also Figure 2); for $z_{\rm spec} > 1$, we find these numbers to be $-0.023$ and 0.060; for the K20 sample, which is 92 % complete for $K^{(\rm AB)} < 21.8$, these numbers are $-0.028$ and 0.033.

including those large surveys referred to in Section 1, as well as the X-ray selected spectroscopic redshift catalogs of Szokoly et al. (2004) and Treister et al. (2008), a new survey by S Koposov (in preparation), and a number of smaller projects. K20 is particularly useful in this regard, given its exceptionally high spectroscopic completeness, albeit over a very small area: 92 % of $K^{(\mathrm{Vega})} < 20$ sources over 52 □'.

The main panel of **Figure 1** shows our $z_{\mathrm{spec}}$–$z_{\mathrm{phot}}$ plot. We prefer to quantify the photometric redshift quality in terms of the normalized median absolute deviation (NMAD[3]) in $\Delta z/(1+z_{\mathrm{spec}})$, which we will abbreviate as $\sigma_z$; for this comparison sample, $\sigma_z = 0.035$. Further, the outlier fraction is acceptably small: 5.9 %. Comparing only to the 241 redshifts from K20, we find $\sigma_z = 0.033$; for the van der Wel et al. (2005) sample of 28 $z \sim 1$ early type galaxies the figure is 0.022. For $1 < z_{\mathrm{spec}} < 2$, we find $\sigma_z = 0.059$. For the 20 % (269/1297) of $0.2 < z_{\mathrm{phot}} < 1.8$ galaxies from our mass limited sample defined in Section 4 that have spectroscopic redshifts, we find $\sigma_z = 0.043$.

Based on their catalog of the GOODS ACS and ISAAC data, Grazian et al. (2006) have achieved a photometric redshift accuracy of $\langle \Delta z/(1+z_{\mathrm{spec}}) \rangle$ of 0.045. For comparison to Grazian et al. (2006), the inset panel shows the distribution of $\Delta z = (z_{\mathrm{phot}} - z_{\mathrm{spec}})$ for $0 < z_{\mathrm{spec}} < 2$; although offset by $-0.046$, the best fit Gaussian to the distribution has a width of 0.046, as opposed to 0.06 for Grazian et al. (2006). For an identical sample of 938 galaxies with $z_{\mathrm{spec}}$s, we find $\sigma_z = 0.043$ for the Grazian et al. (2006) $z_{\mathrm{phot}}$s and $\sigma_z = 0.035$ for ours. In other words, our photometric redshift determinations are at least as good as the best published for $K$-selected samples at high redshifts. We also note in passing that our $z_{\mathrm{phot}} \lesssim 1$ photometric redshifts agree very well with those from COMBO-17 (Wolf et al., 2004); a detailed comparison to both these catalogs is presented in Chapter II.

The lower panel of Figure 1 shows the redshift distributions for both our main galaxy sample (based on $z_{\mathrm{phot}}$), and the spectroscopic comparison sample (based on $z_{\mathrm{spec}}$). Note the presence of three prominent redshift spikes at $0.5 < z_{\mathrm{spec}} < 0.8$ (see also Vanzella et al., 2008); it appears that our redshift determinations may slightly underestimate the redshifts of these structures. The structures at $z \sim 1.0$, 1.1, 1.2, 1.3, and 1.4 (Vanzella et al., 2008) are also visible in the $z_{\mathrm{spec}}$ distribution, but are 'washed out' in the $z_{\mathrm{phot}}$ distribution.

### 3.2 Restframe Photometry and Stellar Masses

The many degeneracies between SED shape and the intrinsic properties of the underlying stellar population, which are actually a help when deriving photometric redshifts, make the estimation of such properties from SED fitting highly problematic. Systematic uncertainties associated with parameterizations of the assumed star formation history are at the level of 0.1 dex (Pozzetti et al., 2007), while uncertainties in the stellar population models themselves are generally accepted to be $\lesssim 0.3$ dex; this is comparable to the uncertainty associated with the choice of stellar IMF. For these reasons, we have opted for considerably simpler

[3]The NMAD is defined as $1.48 \times \mathrm{median}[|x - \mathrm{median}(x)|]$; the normalization factor of 1.48 ensures that the NMAD of a Gaussian distribution is equal to its standard deviation.

means of deriving restframe parameters.

Once the redshift is determined, we have interpolated restframe fluxes from the observed SED using a new utility dubbed InterRest, which is a slightly more sophisticated version of the algorithm described in Appendix C of Rudnick et al. (2003), and is described in detail in Chapter II. InterRest is designed to dovetail with EAZY, and is also freely available.[4] We estimate the systematic errors in our interpolated fluxes (*cf.* colors) to be less than 2 % (Chapter II).

We then use this interpolated restframe photometry to estimate galaxies' stellar masses using a prescription from Bell & de Jong (2001), which is a simple relation between restframe $(B-V)$ color and stellar mass–to–light ratio: $M_*/L_V$:

$$\log_{10}\ M_*/L_V = -.734\ +\ 1.404\ \times\ (B-V+0.084)\ , \tag{1}$$

assuming $M_{V,\odot} = 4.82$. (Here, the factor of 0.084 is to transform from the Vega magnitude system used by Bell & de Jong (2001) to the AB system used in this work.) This prescription assumes a 'diet Salpeter' IMF, which is defined to be 0.15 dex less massive than the standard Salpeter (1955) IMF, and is approximately 0.04 dex more massive than that of Kroupa (2001). Further, to prevent the most egregious overestimates of stellar masses, we limit $M_*/L_V \leq 10$ (see also Figure 4 of Borch et al. 2006). Although this limit affects just 1.2 % of our main sample, we found it to be important for getting the high-mass end of the $z \approx 0$ mass function right (Appendix A).

It is not immediately obvious whether using these color-derived $M/L$s is significantly better or worse than, say, from using stellar population synthesis to fit the whole observed SED. The prescription we use has been derived from SED-fit $M/L$s; the scatter around this relation is on the order of 0.1 dex. By comparison, the precision of SED-fit $M/L$ determinations is limited to 0.2 dex by degeneracies between, e.g., age, metallicity, and dust obscuration (see, e.g., Pozzetti et al., 2007; Conselice et al., 2007). Thus, the increase in the random error in $M_*$ due to the use of color-derived rather than SED-fit stellar masses is $\sim 10$ %.

In addition to being both simpler and more transparent, however, the use of color-derived $M/L$s has the major advantage of using the same restframe information for all galaxies, irrespective of redshift. This is especially important when it comes to the comparison between the high-z and low-z samples, where the available photometry samples quite different regions of the restframe spectrum. Since Equation 1 has ultimately been derived from SED-fit mass estimates, however, our color-derived mass estimates are still subject to the same systematic uncertainties. To the extent that such systematic effects are independent of color, redshift, etc., they can be accommodated within our results by simply scaling our limiting mass. On the other hand, if there is significant evolution in the color–$M/L$ relation with redshift, then there is the risk that the use of color-derived $M/L$s may introduce significant systematic errors with redshift. We investigate this issue further in Section 9.5, in which we also demonstrate that our results are essentially unchanged if we use conventional SED-fitting techniques to derive $M/L$s.

[4] Code and documentation can be found at: http://www.strw.leidenuniv.nl/∼ent/InterRest/.

### 3.3 The Propagation of Redshift Errors

A primary concern in this Chapter is the importance of systematic errors. To address this concern in the context of our photometric redshift determinations, we show in the top panels of **Figure 2** the $z_{\rm spec}$–$z_{\rm phot}$ agreement as a function of redshift, S:N in the $K$ 'color' aperture, and restframe color. In each panel, the points with error bars show the median and 15/85 percentiles in discrete bins.

Looking first at our photometric redshifts: the first panel of Figure 2, shows the photometric redshift error, $\Delta z/(1+z_{\rm spec})$, as a function of $z_{\rm spec}$. We see that there is a systematic offset between $z_{\rm phot}$ and $z_{\rm spec}$ for $z_{\rm spec} \lesssim 1$, such that our $z_{\rm phot}$s tend to be slightly too low (see also Figure 1); for $z \gtrsim 1$, this effect appears to be less. At least for $K$ S:N $\gtrsim 10$, random errors in the photometric redshifts do not appear to be a strong function of S:N.

There is a clear systematic effect as a function of restframe color. For galaxies redder than $(u-r) \approx 2$ (approximately the lower limit for $z \approx 0$ red sequence galaxies), the agreement between $z_{\rm phot}$ and $z_{\rm spec}$ is very good. For galaxies with $(u-r) \lesssim 2$, however, it seems that we systematically underestimate the true redshift by approximately $\Delta z \lesssim 0.02(1+z_{\rm spec})$. It is plausible that this is in fact the driver of the weak apparent systematic effect with redshift, coupled with there being a greater proportion of blue galaxies in the spectroscopic redshift sample at lower redshifts.

How do these errors in redshift estimation play out in the derivation of restframe properties? We address this question with reference to the lower panels of Figure 2, which illustrate how redshift uncertainties affect our derivation of three basic restframe quantities (*top to bottom*): absolute luminosity, restframe color, and stellar mass. In each panel, we plot the difference between the values derived adopting the spectroscopic or photometric redshift, as a function of (*left to right*) redshift, $K$ band S:N, and restframe color, as well as photometric redshift error.

Quantitatively, for our $z_{\rm spec}$ comparison sample, the random photometric redshift error of $\Delta z/(1+z_{\rm spec}) = 0.035$ translates into a 0.360 mag error in absolute magnitude, 0.134 mag error in restframe color, and a 0.107 dex error in stellar mass. (By contrast, the typical uncertainty in for a $K$ S:N = 10 galaxy is $\Delta K =$ 0.12–0.16 mag $\approx$ 0.05–0.07 dex.) Just as for the redshifts themselves, the clearest systematic effects in the derived quantities is with restframe color: there appear to be mild systematics with redshift for $z \lesssim 1$, and no clear trend with S:N, at least for S:N $> 10$.

It is easy to understand why redshift errors play a larger role in the derivation of magnitudes rather than colors. When calculating magnitudes, the primary importance of the redshift is a distance indicator. For the $z_{\rm spec}$ sample shown in Figure 2, the random scatter in $\Delta M_r$ due to distance errors alone (calculated by taking the difference in the distance modulus implied by $z_{\rm phot}$ versus that by $z_{\rm spec}$) is 0.28 mag; *i.e.* $\sim 75$ % of the scatter seen in Figure 2. On the other hand, colors are distance independent, and this element of uncertainty is cancelled out.

What is surprising is the relative insensitivity of our stellar mass estimates to redshift errors. Focusing on the right panels of Figure 2, it can be seen that where

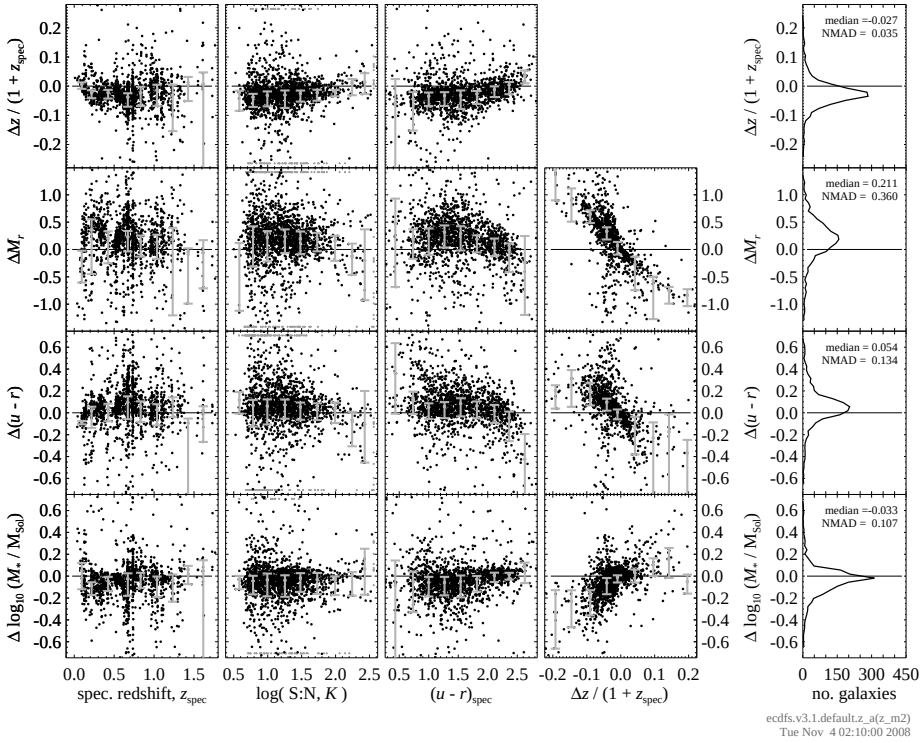


**Figure 2.** — Photometric redshift errors, and their effect on other derived quantities. — In each panel, the abscissa shows the difference in a derived quantity, derived assuming the spectroscopic or photometric redshift; in all cases, '$\Delta$' should be understood as the difference between the $z_{\rm phot}$– minus $z_{\rm spec}$–derived values. We show: (*top to bottom, in rows*) redshift, absolute magnitude, restframe color, and stellar mass, as a function of (*left to right, in columns*) redshift, observed signal–to–noise, restframe color, and photometric redshift error. The black points are for the spectroscopic sample shown in Figure 1; the red points show the median offset in bins; the error bars reflect the 15/85 percentiles. In each panel, points that fall outside of the plotted range are shown as small grey plusses. The separate panels at right show the distribution of $\Delta$s for all galaxies in the $z_{\rm spec}$ sample; the median and scatter (NMAD) in the difference between $z_{\rm phot}$– and $z_{\rm spec}$–derived quantities are as shown. In all cases, the clearest systematic effects are as a function of redshift, with some systematic errors for the reddest and bluest galaxies. The way that redshift errors propagate mean that the uncertainties *due to redshift errors* are much smaller for stellar masses than they are for either absolute magnitudes or restframe colors.

the photometric redshift underestimates the true redshift/distance, we will infer both too faint an absolute luminosity and too red a restframe color. When it comes to computing a stellar mass, however, these two effects operate in opposite directions: although the luminosity is underestimated, the too–red color leads to an overestimate of the stellar mass–to–light ratio. The two effects thus partially cancel one another, leaving stellar mass estimates relatively robust to redshift errors.

This means that *in a photometric redshift survey, the measurement uncertainty due to random photometric redshift errors is considerably less for stellar masses than it is for absolute magnitudes.* This conclusion remains unchanged using SED–fit stellar masses, rather than our favored color–derived ones. Conversely, we can say that (random) photometric redshift errors are not a dominant source of uncertainty in our stellar mass estimates. Indeed, as we have already noted, the size of these errors is comparable to the uncertainties in our total flux measurements.

## 4 Constructing a Stellar Mass Selected Sample

For moderate to high redshifts, NIR selection has the key advantage of probing the restframe optical light, which is a reasonable tracer of stellar mass. In this section, we empirically relate our observed flux detection/selection limit to an approximate completeness limit in terms of stellar mass and redshift.

To this end, we have taken galaxies with $K$ fluxes immediately below our detection limit from three significantly deeper $K$-selected catalogs; *viz.* the MUSYC deep NIR catalogs (Quadri et al., 2007), the FIREWORKS catalog (Wuyts et al., 2008), and the FIRES catalogs (Labbé et al., 2003; Förster-Schreiber et al., 2006). By taking objects from these catalogs that lie immediately below our detection threshold, and scaling their fluxes (and so stellar masses) to match our $K = 22$ limit, it is then possible to empirically determine the stellar mass–redshift relation for $K \approx 22$ galaxies. The upper envelope of points in ($M_*$, $z_{\rm phot}$) space thus represents the most massive galaxies at our observed flux limit, and so directly provides a redshift-dependent mass completeness limit.

This is illustrated in the left panel of **Figure 3**. In this panel, the large, open, colored symbols represent $22.0 < K < 22.5$ objects from the deeper catalogs, scaled up in flux to $K = 22$; *viz.* the MUSYC deep NIR catalogs (*blue crosses*), the FIREWORKS catalog (*yellow circles*), and the FIRES catalogs (*red squares*). Again, these points represent objects immediately at our detection limit; the upper envelope of these points therefore represents the most massive galaxies that might escape detection/selection in our analysis. This suggests that for $M_* > 10^{11}$ M$_\odot$, we are approximately complete for $z_{\rm phot} < 1.8$.

It is possible to do the same thing using the faintest detections from our own catalog, scaled down in flux to our selection limit. Specifically, we have taken galaxies with $21.5 < K < 22.0$ and scaled their fluxes (and masses) down to $K = 22$. For this test, we also restrict our attention to galaxies with well constrained redshifts, by requiring that the EAZY `odds` parameter be greater than 0.95.

These points are shown as the closed circles in the left panel of Figure 3. While the results of this 'internal' test are broadly consistent with the previous 'external'

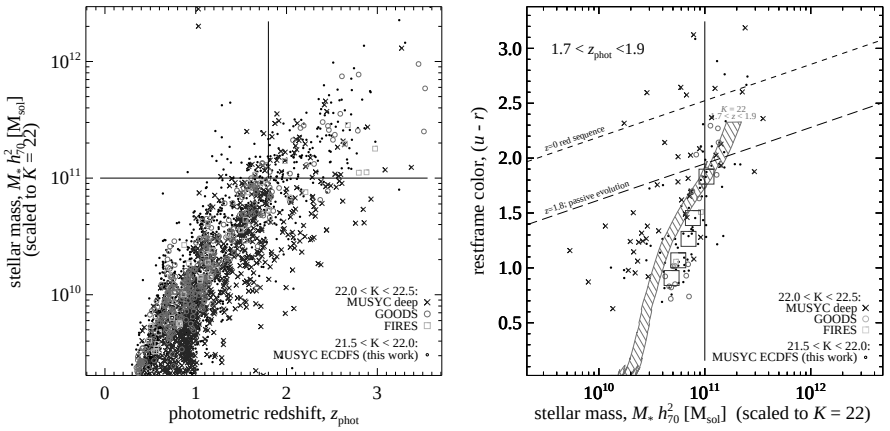


**Figure 3.** — Empirically determining our mass completeness limit as a function of redshift and color. — *Left panel:* The black points show stellar masses for MUSYC ECDFS galaxies with $21.5 < K < 22.0$, scaled down in flux to match our $K = 22$ detection limit, and plotted as a function of photometric redshift. The other symbols show stellar masses for $22.5 > K > 22.0$ galaxies, scaled up in flux to $K = 22$; these galaxies are drawn from the MUSYC deep fields (*blue crosses*), the FIREWORKS catalog (*yellow circles*), and the FIRES catalogs (*red squares*). Each sample has been analyzed in exactly the same manner. The upper envelope of these points effectively defines, as a function of redshift, the limiting stellar mass corresponding to our observed $K$ flux limit. For $M_* > 10^{11}$ $M_\odot$, we are nearly complete ($\gtrsim 90\%$) to $z_{\rm phot} = 1.8$. *Right panel:* the color–stellar mass diagram for $K \approx 22$ galaxies at $1.7 < z_{\rm phot} < 1.9$. The large squares show the median values for the MUSYC ECDFS points, binned by color; all other symbols are as in the left panel. Here, the right envelope of the colored points defines our mass completeness limit at $z_{\rm phot} \approx 1.8$ as a function of color. For comparison with Figures 4 and 5, the hatched region shows estimated completeness limits based on synthetic SSP spectra. While we may well miss galaxies considerably redder than the predicted red sequence (see Section 7), this empirical argument suggests that at $z_{\rm phot} \approx 1.8$, we are approximately complete ($\gtrsim 85$ %) for galaxies with $M_* > 10^{11}$ $M_\odot$ and $(u - r) < 2$.

one, they do suggest slightly higher incompleteness. Of the $21.5 < K < 22.0$ sources with $1.6 < z_{\rm phot} < 1.8$, 23 % (20/87) would have $M_* > 10^{11}$ $M_\odot$ when scaled down to $K = 22$, indicating that our completeness for $K = 22$, $M_* = 10^{11}$ $M_\odot$ galaxies is $\sim 75$ % for $1.6 < z_{\rm phot} < 1.8$. However, the $21.5 < K < 22.0$ subsample shown here represents only 30 % of our full $K < 22$ sample in this mass and redshift range, suggesting that the overall completeness is more like $> 90$ %.

As a second and complementary check on this conclusion, the right panel of Figure 3 shows the color–stellar mass diagram for a narrow redshift slice at $1.7 < z_{\rm phot} < 1.9$. Here, the large squares show the median values from the MUSYC ECDFS points, binned by color; the other symbols are the same as in the other panel of this Figure. As before, the points in this panel represent objects at our detection limit; the right envelope of these points thus describes our mass completeness limit at $z_{\rm phot} \approx 1.8$, this time as a function of restframe color.

Again, the 'internal' and 'external' analyses broadly agree. For blue galaxies,

both tests suggest that MUSYC should be very nearly complete for $M_* > 10^{11}$ $M_\odot$ and $z_{\rm phot} < 1.8$. For $(u-r) > 1.5$ galaxies, however, the down-scaled MUSYC points again suggest slightly lower completeness than those scaled up from deeper catalogs: 45 % (13/29) of these galaxies would fall foul of one of our selection criteria if their masses were scaled down to $M_* = 10^{11}$ $M_\odot$. Using an argument analogous to that above, this suggests that our completeness fraction for galaxies with $M_* > 10^{11}$ $M_\odot$, $z_{\rm phot} = 1.8$, and $(u-r) > 1.5$ is at least $\sim 85$ %.

We therefore adopt $M_* > 10^{11}$ $M_\odot$ and $z_{\rm phot} < 1.8$ as our approximate completeness limits, corresponding to our $K < 22$ detection/selection limit. In Section 9.3.2, we apply simple completeness corrections to determine the extent to which our results may be affected by incompleteness. As a final caveat, however, there remains the concern of additional incompleteness due to our $K$ S:N $> 5$ criterion, which we will address in Section 9.3.1.

## 5 The Color–Magnitude and Color–Stellar Mass Diagrams for $z_{\rm phot} \lesssim 2$

In this section, we present our basic observational results: the color–magnitude and color–stellar mass diagrams for $z_{\rm phot} \lesssim 2$.

### 5.1 The Color–Magnitude Diagram for $z_{\rm phot} \lesssim 2$

**Figure 4** shows the color–magnitude diagram (CMD), plotted in terms of $(u-r)$ color and absolute $r$ magnitude, $M_r$, for $z_{\rm phot} \lesssim 2$.

The first panel of Figure 4 is for $z \approx 0$ galaxies from the 'low-$z$' comparison sample; these data and our analysis of them are described in Appendix A. The basic features of the CMD — the red sequence, blue cloud, and green valley — are all easily discernible. The dotted line shows our characterization of the $z \approx 0$ CMR for red galaxies (Equation A.1), also discussed in Appendix A.

The other eight panels show the $0.2 < z_{\rm phot} < 1.8$ MUSYC ECDFS data. Except for the two highest redshift bins, which are twice as large, the $z_{\rm phot} \gg 0$ bins have been chosen to have equal comoving volume.[5] For the $z \approx 0$ bin, we plot only a random subsample of the low-z catalog, chosen to effectively match the volume of the higher redshift bins. Thus, the density of points in the color–magnitude plane is directly related to changes in the bivariate comoving number density.

In the bottom-right corner of each panel, we show representative error bars for a $M_* \approx 10^{11}$ $M_\odot$ galaxy with $(u-r) \approx 2.0$, near the mean redshift of each bin. In order to derive these errors, we have created 100 Monte Carlo realizations of our catalog, in which we have perturbed the catalog photometry according to the photometric errors, and repeated our analysis for each: the error bars show the scatter in the values so derived. The shaded grey regions show approximately how our $K < 22$ completeness limit projects onto color–magnitude space through each redshift bin, derived using synthetic single stellar population (SSP) spectra.

[5]The exact redshift limits we use are $z_{\rm phot} = 0.200$, 0.609, 0.825, 0.987, 1.127, 1.254, 1.373, 1.486, 1.595, 1.700, 1.804, 1.906, and 2.000; elsewhere we will round these values as convenient.

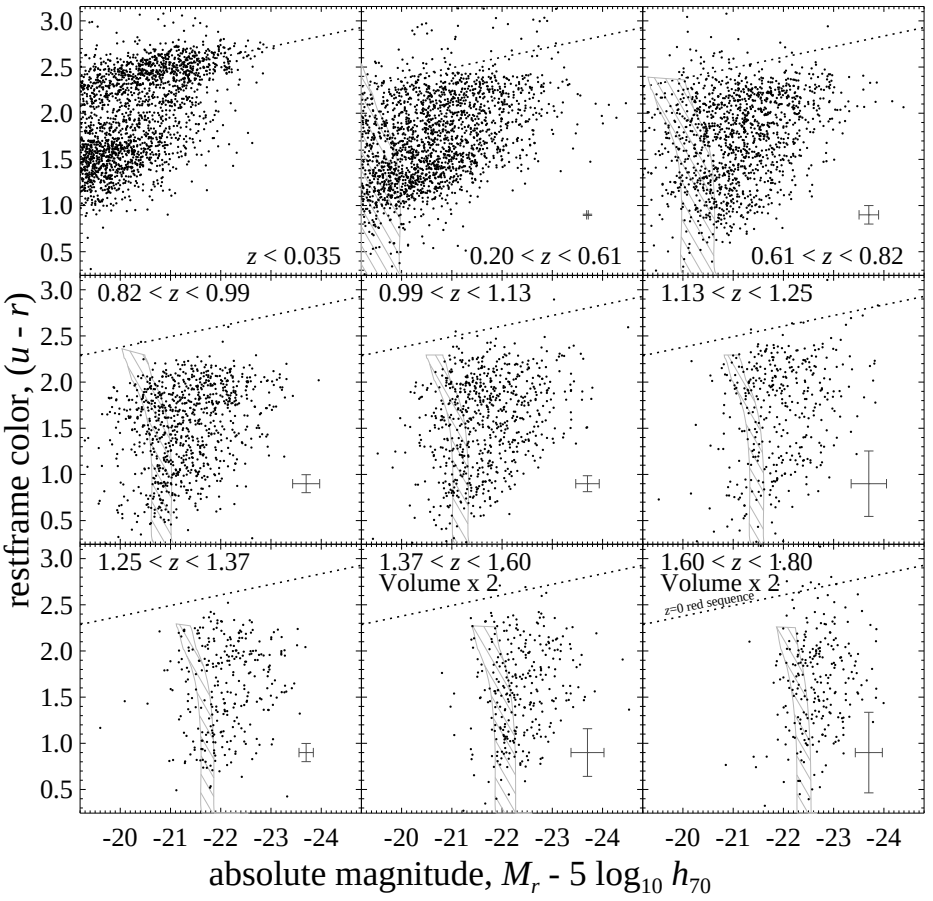


**Figure 4.** — The color–magnitude diagram (CMD) for galaxies with $z_{\rm phot} \lesssim 2$. — The first panel shows a random selection from the NYU VAGC's 'low-z' sample, based on DR4 of the SDSS (Blanton et al., 2005b), discussed in Appendix A; the other panels show the MUSYC ECDFS data, discussed in the main text. Except where marked, bins are of equal comoving volume; for the low-z sample, we have plotted a random subsample to yield the same effective volume: the density of points is thus directly related to bivariate comoving number density. The shaded area shows the approximate $K = 22$ detection/selection limits, based on synthetic spectra for an SSP; the error bars show representative errors for a $M_* \approx 10^{11}$ $M_\odot$, $(u-r) \approx 2.0$ galaxy at the mean redshift of each bin. The dotted line in each panel shows our fit to the CMR for bright, red sequence galaxies at $z \approx 0$, derived in Appendix A. In this work, we prefer to use stellar mass, rather than absolute magnitude, as a basic parameter — accordingly, we focus our attention on the CM$_*$D, presented in Figure 5.

Examining this diagram it is clear that, in the most general terms possible, bright/massive galaxies were considerably bluer in the past. At a fixed magnitude, the entire $z \sim 1$ galaxy population is a few tenths of a magnitude bluer than at $z \approx 0$. At the same time, particularly for $z \gtrsim 1$, there is a growing population of galaxies with $M_r < -22$ and $(u - r) < 2$ that has no local analogue. While there are some indications of a red sequence within the $z \gg 0$ data, particularly for $z_{\rm phot} \lesssim 1$, it is certainly not so easily distinguishable as locally.

### 5.2 The Color–Stellar Mass Diagram for $z_{\rm phot} \lesssim 2$

There are a number of advantages to using stellar mass as a basic parameter, rather than absolute magnitude. Principal among these is the fact that stellar mass is more directly linked to a galaxy's growth and/or assembly: while a galaxy's brightness will wax and wane with successive star formation episodes, a galaxy's evolution in stellar mass is more nearly monotonic. On the other hand, it must be remembered that the necessary assumptions in the derivation of stellar mass estimates produce greater systematic uncertainties than for absolute luminosities. With this caveat, we will focus on the color–stellar mass diagram in this and following sections.

**Figure 5** shows the color–stellar mass diagram ($\rm CM_*D$) for $z \lesssim 2$. As in Figure 4, the first panel shows a random subsample of the low-z sample; the other panels show the MUSYC ECDFS data. The dotted line in each panel shows the $z \approx 0$ color–stellar mass relation ($\rm CM_*R$), which we have derived in Appendix A, given in Equation A.1.

Each of the basic features of the CMD are also seen in the $\rm CM_*D$. We see an increasing number of galaxies with $z_{\rm phot} \gtrsim 1$ and with $M_* > 10^{11}$ $\rm M_\odot$ and $(u - r) \lesssim 2$ which have no analogues in the local universe. For a SSP, the colors of these galaxies would suggest ages of $\lesssim 1$ Gyr: these massive galaxies appear to be in the throes of their final star formation episodes. For $z_{\rm phot} \gtrsim 1.2$, these galaxies may even dominate the massive galaxy population. Some evidence for a distinct red sequence is visible in the $\rm CM_*D$ for $z_{\rm phot} \lesssim 1$, but not much beyond.

The next three sections are devoted to more quantitative discussion of each of the following three specific observations:

1. We see evidence for a red galaxy sequence for $z_{\rm phot} \lesssim 1.2$; beyond this redshift, whether due to physical evolution or to observational errors, it becomes impossible to unambiguously identify a distinct red sequence on the basis of the present data (Section 6).

2. At a fixed mass, a red sequence galaxy at $z \sim 1$ is a few tenths of a magnitude bluer than its $z \sim 0$ counterpart (Section 7).

3. At higher redshifts, there appear to be fewer massive galaxies on the red sequence. Further, it appears that the proportion of blue cloud galaxies among the most massive galaxies increases; conversely, the red fraction is lower at higher redshifts. (Section 8).

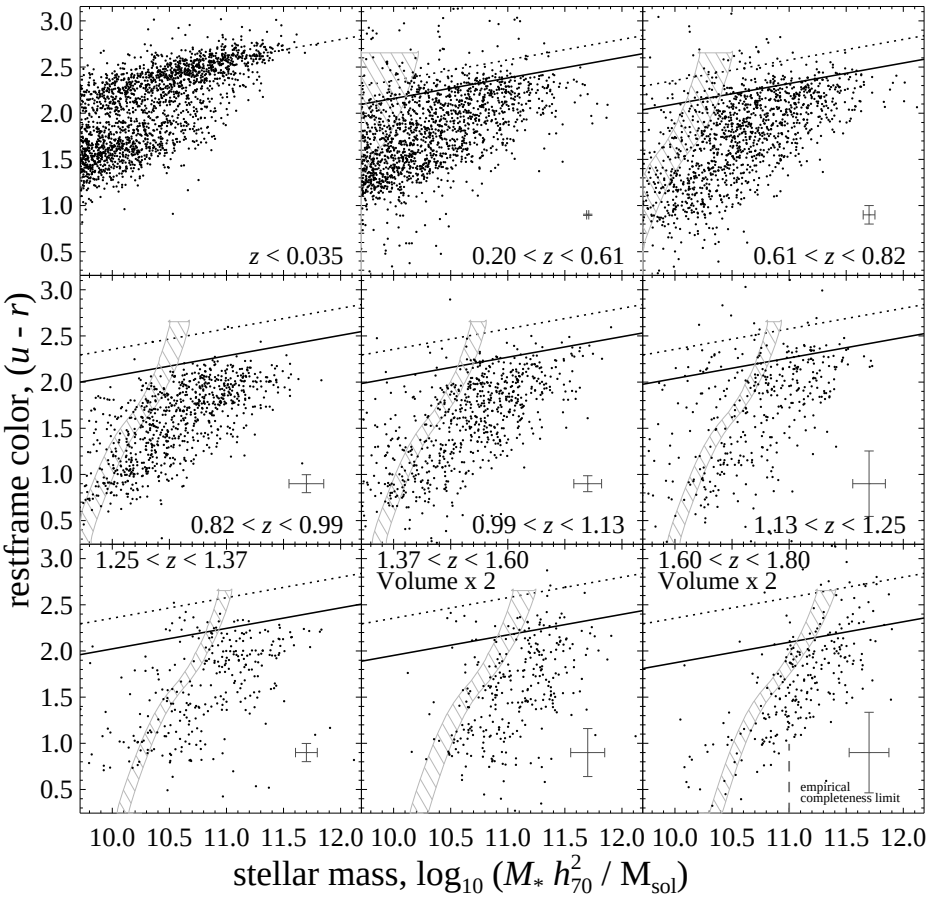


**Figure 5.** — The color–stellar mass diagram ($CM_*D$) for galaxies with $z_{\rm phot} \lesssim 2$. — As in Figure 4, the $z \approx 0$ bin is based on the low-z sample of SDSS galaxies, discussed in Appendix A; the $z_{\rm phot} \gg 0$ points based on the MUSYC ECDFS data, described in the main text. The hatched area shows approximate selection limits, based on synthetic spectra for an SSP; our empirical completeness limit is marked in the last bin. The error bars show representative errors for a $M_* \approx 10^{11}$ $M_\odot$, $(u-r) \approx 2.0$ galaxy near the bin's mean redshift, based on 100 Monte Carlo realizations of the catalog data, including photometric redshift errors. Within each panel, the dotted line shows our fit to the $CM_*R$ for bright red sequence galaxies at $z \approx 0$, derived as per Appendix A; for the $z_{\rm phot} \gg 0$ bins, the solid lines show our fit to the color evolution of the massive red galaxy population, derived in Section 7; the dashed line shows our red galaxy selection criterion, introduced in Section 8. We analyze the key features of this diagram further in Figures 6, 8, 9, and 10.

## 6 The Color Distribution of Massive Galaxies for $z_{\rm phot} < 2$

In an attempt to quantitatively separate the massive galaxy population into distinct red and blue subpopulations, **Figure 6** plots the color distribution of the $M_* > 10^{11}$ $M_\odot$ galaxy population, after subtracting out the slope of the $z \approx 0$ $CM_*R$, and using the same redshift bins as in Figures 4 and 5. The grey histograms in each panel show the data themselves. Note that for the $z \approx 0$ panel of this plot, we have used the full low-z sample. Also, recall that for this mass regime, we are approximately complete (volume limited) to $z_{\rm phot} = 1.8$.

### 6.1 The Massive Galaxy Red Sequence at $z_{\rm phot} \lesssim 1.2$

Locally, red sequence galaxies totally dominate the massive galaxy population: what bimodality exists between the red and blue populations is weak. (As the name suggests, 'bimodality' implies two distinct local maxima in the distribution.) This is a reflection of the apparent 'transition mass' between red and blue galaxies observed by Kauffmann et al. (2003a); the bimodality is stronger in a luminosity limited sample, including a greater proportion of bright but less massive blue cloud galaxies (see also Figure A.1; Appendix A). At slightly higher redshifts, where some progenitors of $z \approx 0$ red sequence galaxies are still forming stars in the blue cloud, we may then expect the bimodality to actually become stronger, before weakening again as the fraction of those galaxies already on the red sequence becomes small at moderate–to–high redshifts. In general, however, the color distributions shown in Figure 5 are not clearly bimodal.

With this in mind, as a simple means of separating red from blue galaxies, we have fit the observed distributions in each redshift bin with double Gaussian functions. These fits are shown by the smooth curves in each panel of Figure 6. For the most part, these fits provide a reasonable description of the $0.2 < z_{\rm phot} \lesssim 1.2$ data (see also Borch et al., 2006).

### 6.2 A Massive Galaxy Red Sequence at $z_{\rm phot} \gtrsim 1.3$?

In contrast to lower redshifts, for $z_{\rm phot} \gtrsim 1.3$, we are no longer able to reliably fit the color distributions in this manner: on the basis of the sensitivity tests presented in Section 9, neither the red/blue separation nor the fits to the distributions of these populations is robust. Looking at the typical errors shown in each panel of Figure 5, the measurement errors on the $(u-r)$ colors of red galaxies rises sharply from $\Delta(u-r) \lesssim 0.1$ mag for $z_{\rm phot} \lesssim 1.2$ to 0.2 mag for $z_{\rm phot} \sim 1.3$, 0.3 mag for $z_{\rm phot} \sim 1.5$, and 0.4 mag for $z_{\rm phot} \sim 1.7$. This would suggest that our inability to reliably distinguish both a red and a blue population at these higher redshifts may very well be due to the large errors in restframe colors for higher redshift galaxies.

In order to help interpret our $z_{\rm phot} \gtrsim 1$ results, we have tested our ability to recover a single-color galaxy population, given our observational and analytical errors. To this end, we have generated a mock photometric catalog containing only red sequence galaxies: beginning with our main galaxy sample, we have replaced each galaxy's photometry using synthetic spectra for a SSP formed over a short

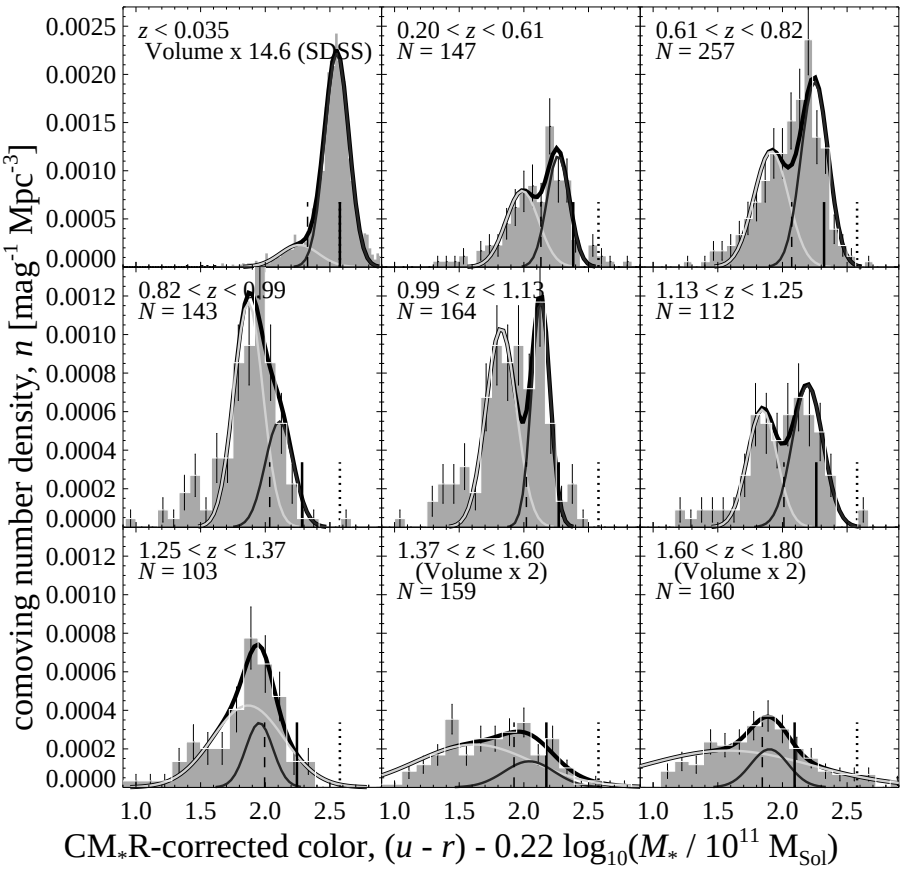


**Figure 6.** — Color distributions for $M_* > 10^{11}$ $M_\odot$ galaxies at $z \lesssim 2$. — In each panel, the histograms show the color distribution for $M_* > 10^{11}$ $M_\odot$ galaxies, after subtracting out the slope of the local $CM_*R$, normalized at $M_* = 10^{11}$ $M_\odot$. The shaded distributions show the data themselves. The smooth curves show double Gaussian fits to the observed distributions. For $z_{\rm phot} \gtrsim 1.2$, the Gaussian fits to the observed color distributions are not robust; our inability to reliably distinguish separate red and blue populations for $z_{\rm phot} \gtrsim 1.2$ is likely due to insufficient S:N in our NIR data (see Figure 7). In each panel, the vertical dotted line shows the location of the $z \approx 0$ red sequence; the vertical solid lines show our fit to the observed color evolution of the red sequence, derived in Section 7; the dashed lines show our red galaxy selection criterion, introduced in Section 8.

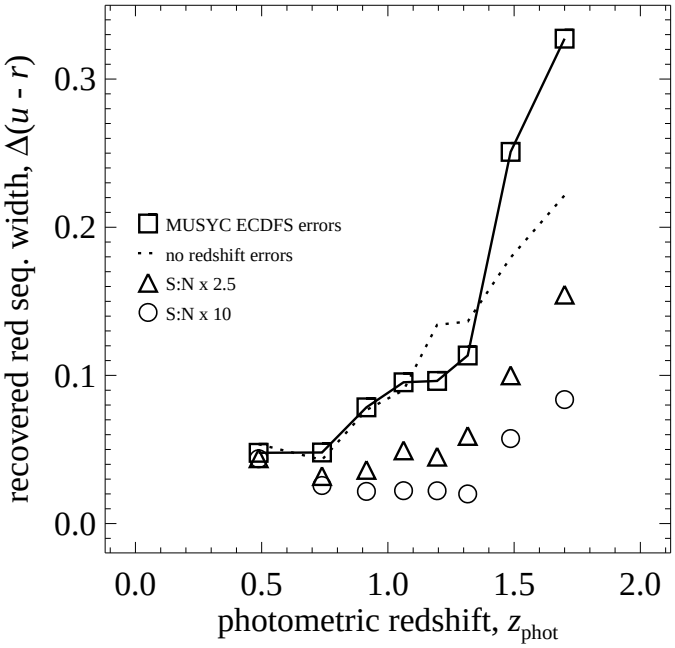


**Figure 7.** – The effects of photometric errors on our ability to recover a single-color (red) galaxy population. — These results are based on mock galaxy catalogs containing only passively evolving galaxies, taking redshifts and stellar masses of galaxies from the main sample, which we have analyzed in the same manner as the actual data: we are thus testing our ability to recover a single-color galaxy population as a function of redshift. The squares, triangles, and circles show the width of the recovered color distributions assuming typical errors for the MUSYC ECDFS data divided by 1, 2.5, and 10, respectively; the dotted line shows the results assuming MUSYC ECDFS photometric errors, but no redshift errors. Even with spectroscopic redshifts, the depth of the MUSYC ECDFS NIR data precludes the detection of a $z_{\rm phot} \gtrsim 1.4$ red sequence.

period beginning at $z_{\rm form} = 5$, assuming the catalog values of $z_{\rm phot}$ and $M_*$, and then adding typical MUSYC ECDFS photometric errors. We have then reanalyzed this catalog using the same methods and procedures as for the main analysis, including recomputing photometric redshifts, restframe colors, and stellar masses.

The results of these tests are shown in **Figure 7**, which plots the observed width of the color distribution for this intrinsically single color population, as a function of photometric redshifts, assuming photometric errors typical for the MUSYC ECDFS (squares). The measurement errors on the $(u-r)$ colors of red galaxies rise sharply from 0.05–0.07 mag for $z_{\rm phot} \lesssim 1$ to 0.10 mag for $z_{\rm phot} \approx 1$, and then continue to increase for higher redshifts. In order to demonstrate that this is not a product of redshift errors, but of photometric errors *per se*, we have also repeated this analysis holding the redshifts of each object fixed; the results of this test are shown as the dotted line. Even with spectroscopic redshifts, the depth of our NIR data would seem to preclude the detection of a distinct red sequence at $z \gtrsim 1.3$.

What then would be required in order to confirm the non/existence of a red sequence at $z_{\rm phot} \gtrsim 1.5$? We have also constructed mock galaxy catalogs with S:N that is 2.5 and 10 times greater than typical values for the MUSYC ECDFS catalog; the results of these tests are shown as the triangles and circles, respectively. Even pushing a full magnitude deeper, it would be difficult to identify a red sequence at $z \sim 1.5$ (assuming that observational errors of $\Delta(u-r) \lesssim 0.1$ mag would be required to robustly identify a red sequence). In order to probe $z \gtrsim 1.5$, an order of magnitude improvement is required. This would suggest that the detection of a red sequence at $z \gtrsim 1.5$ would require a $J$ band ($5\sigma$ point source) limit of $\sim 25.8$, roughly the final target depth for the Ultra Deep component of the UKIRT Infrared Deep Sky Survey (Lawrence et al., 2007).

It is clear from this analysis that we cannot confirm or exclude the existence of a red galaxy sequence at $z \gtrsim 1.3$ on the basis on the present data. This is in good accord with the recent detections of a $z \lesssim 1.5$ red galaxy sequence by Cirasuolo et al. (2007) and Franzetti et al. (2007; see also Kriek et al. 2008). Moreover, we note in passing that if we were to subtract away the broadening effect of photometric errors, as derived from the test described above, then the implied intrinsic width of the red sequence is $\approx 0.1$ mag for all $z_{\rm phot} \lesssim 1.2$, consistent with the $z_{\rm spec} < 1.0$ findings of Ruhland et al. (2009).

## 7 The Color Evolution of Massive Red Galaxies

Our next task is to quantify the color evolution of the massive red galaxy population. We have addressed this question based on the double Gaussian fits to the ($\rm CM_*R$–corrected) color distributions of $M_* > 10^{11}$ $\rm M_\odot$ galaxies shown in Figure 6. In **Figure 8**, we plot the fit centers of the blue (*lower points*) and red (*upper points*) galaxy color distributions as a function of redshift; the errors on the locus of the massive red galaxy population shown in this Figure have been obtained by bootstrap resampling.

From this plot, it is clear that the red galaxy population as a whole has become progressively redder by $\sim 0.4$ mag in $(u-r)$ over the past 9 Gyr; the evolution in the blue cloud is similar. Making a linear fit to the (robust) $z_{\rm phot} < 1.1$ measurements, we find $\Delta(u-r)_{\rm corr} = 2.57 - (0.44 \pm 0.02)\ z_{\rm phot}$. (Note that this fit is constrained to match the $z \approx 0$ point.) These results do not change significantly if we also include the point $z_{\rm phot} \sim 1.2$, but it is clear that if we were to fit to the $z_{\rm phot} \gtrsim 1.3$ points, we would find slightly less strong evolution.

Kriek et al. (2008) report a $3.3\sigma$ detection of a red sequence in the spectrally derived $(U-B)$ color distribution of a mass-selected sample of 36 $z_{\rm phot} \gtrsim 2$ galaxies, 12 of which lie in the ECDFS. The square at $z = 2$ in Figure 8 shows the approximate $(u-r)$ color equivalent of their red sequence detection. While the Kriek et al. (2008) point is slightly redder than an extrapolation of our linear fit, the two results agree rather well.

In **Figure 9**, we compare the observed color evolution of the red galaxy population with naïve expectations from passive evolution of synthetic spectra. For this purpose, we have used Pégase V2.0 (Le Borgne & Rocca-Volmerange, 2002)

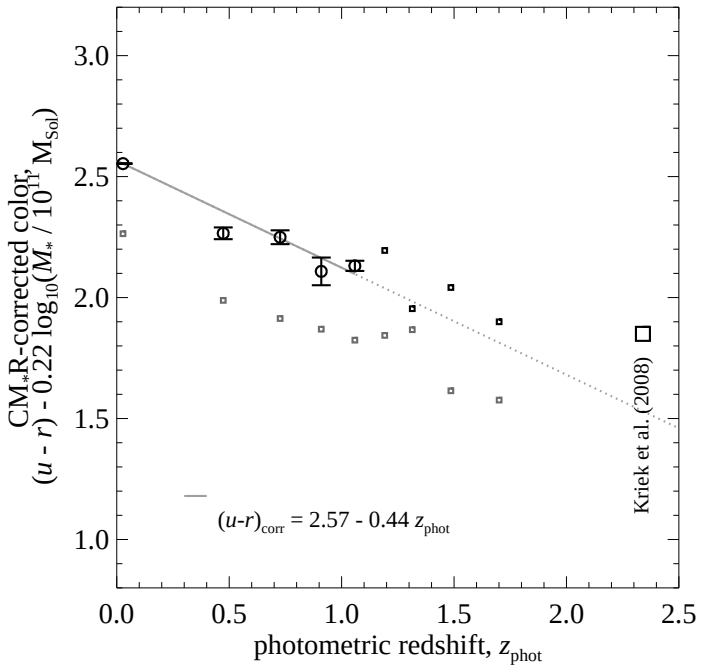


**Figure 8.** — The color evolution of massive galaxies for $z \lesssim 2$. — Points show the fit centers of the color distributions for the red and blue galaxy subpopulations (see Figure 6). Only the $z_{\rm phot} \lesssim 1.2$ points (*circles with error bars*) were used when fitting for the color evolution of the red galaxy sequence (*solid line*); the errors on each point have been derived from bootstrap resampling. The large square at $z = 2$ shows the approximate equivalent $(u - r)$ color of the $3.3\sigma$ detection of a $(U - B)$ red sequence among $z \sim 2$ galaxies from Kriek et al. (2008), based on NIR spectra. We see rather a rather smooth reddening of the red sequence from $z \sim 1.2$ to the present day, which is well described by the linear fit given.

models with an initial metallicity of $Z = 0.004$, and assuming a short burst of star formation ($e$–folding time of 100 Myr), beginning at (*from bottom to top*) $z_{\rm form} = 2$, 3, 5, or 10.

The $z_{\rm form} \gtrsim 3$ tracks do an adequate job of describing the amount of observed evolution for $z_{\rm phot} < 1.3$. However, to get this level of agreement, the model (luminosity weighted) metallicities at $z = 0$ must be roughly solar ($Z \approx 0.2$), whereas the local mass–metallicity relation would suggest that these abundances should be super-solar by a factor of 3 or more (Tremonti et al., 2004). If we were to adopt $Z = 0.2$ initially, leading to final metallicities that are super-solar by approximately 50 %, the model colors would become nearly 0.5 mag too red. The problem is even worse for SSP models: while solar metallicities lead to colors which are too red by 0.25 mag, an acceptable description of the data is only possible assuming roughly half solar abundances (see also Bell et al. 2004.)

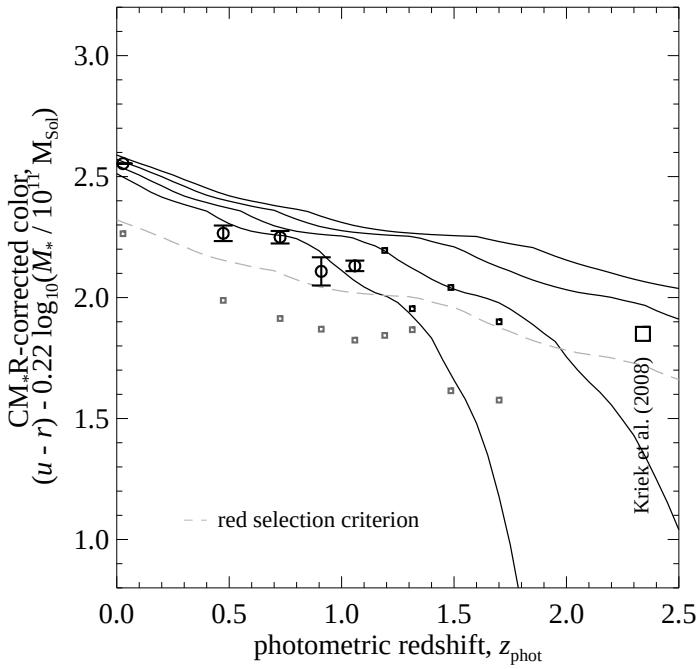


**Figure 9.** — Comparing the color evolution of massive galaxies for $z \lesssim 2$ to passively evolving stellar population models. — The data in this panel are identical to Figure 8. The overlaid curves show simple expectations for the passive evolution of a SSP, formed in a short burst (*e*-folding time of 100 Myr) beginning at $z_{\rm form} = 2$, 3, 5, and 10 (*bottom to top*). While the $z_{\rm form} \gtrsim 3$ curves provide a good qualitative description of the observed evolution, they have substantially lower metallicity than would be expected from the local mass–metallicity relation. Moreover, these very simple models make no attempt to account for progenitor bias, and other important effects.

Nevertheless, these simple models do provide an acceptable description of the rate or amount of color evolution of the red galaxy population, $\Delta(u-r)$, if not the $(u-r)$ colors *per se*. Using these extremely simple models to interpret the color evolution of the red galaxy population, the observations suggest that the bulk of the stars in massive red galaxies were formed at $z_{\rm form} \gtrsim 3$; including the point from Kriek et al. (2008) would suggest $z_{\rm form} \gtrsim 5$.

This is not to say, however, that the observations are consistent with all massive red galaxies being formed at $z \gtrsim 5$, or even 3: even among $M_* > 10^{11}$ $M_\odot$ galaxies, there are simply not enough stars at $z \gtrsim 2$ to build the $z \approx 0$ red sequence population (Fontana et al., 2006; Conselice et al., 2007, see also Figure 10). Instead, what we see is that the colors of massive red galaxies are consistent with being dominated by ancient stars at all redshifts. This implies both extended star formation histories among red galaxies (and/or their progenitors), as well as

a large spread in the times at which galaxies of a given (stellar) mass join the red population — a kind of long migration of galaxies, occurring over many Gyr (*cf.* Brown et al., 2008). A proper description of red sequence evolution would therefore have to account for, among other things, progenitor bias (van Dokkum & Franx, 2001): the continual skewing of the population by new additions (see also Faber et al., 2007; Ruhland et al., 2009). This is beyond the scope of this work.

## 8 The Rise of Red Galaxies Over $z_{\mathbf{phot}} \lesssim 2$

In this section, we turn our attention to the third of our basic results. Whereas our focus until now has been on the properties of the red galaxy population as a whole, we now look at how the number of red galaxies within the total massive galaxy population — *i.e.* the red galaxy fraction — evolves with time.

### 8.1 Defining a Red Galaxy Selection Criterion

Our inability to robustly distinguish separate red and blue galaxy populations on the basis of the observed color distributions for $z_{\rm phot} \gtrsim 1.2$ forces us to devise some alternate means of separating 'red' galaxies from the general field population.

We have already seen that the color evolution of the red galaxy population is roughly consistent with ancient stars at all redshifts (Figure 9). Our simple solution is therefore to use the predicted color evolution for a passively evolving stellar population formed at high redshift to define a redshift-dependent 'red' selection criterion, *viz.*:

$$\begin{aligned}(u-r) > 2.57 \quad &+ \quad 0.24 \times \log_{10}(M_*/10^{11}{\rm M}_\odot) \\ &+ \quad \delta(z_{\rm phot}) - 0.25 \ , \end{aligned} \tag{2}$$

where $\delta(z_{\rm phot})$ is the $(u-r)$ color evolution of a SSP with $z_{\rm form} = 5$, as shown in Figure 9. This selection limit is shown as the dashed lines in each of Figures 5, 6, and 8.

How does this definition of 'red' relate to things like membership of the red sequence, star formation rate and/or history, etc.? As we remarked in the first paragraph, a galaxy's optical color is a reflection of its mean stellar age, modulo the complicating factors of metallicity and dust extinction. In addition to 'red and dead' galaxies, therefore, simply selecting 'red' galaxies can potentially catch a significant number of star forming galaxies with high dust obscuration. In other words, while all passive galaxies are red, not all red galaxies are passive.

In this sense, it is not unreasonable to interpret the redshift evolution of the number and fraction of red, massive galaxies as placing an upper limit on the numbers of 'fully formed' (in the sense that they have essentially completed their star formation and/or assembly) massive galaxies. These results can thus be used to constrain the epoch at which the star formation quenching mechanism operates.

### 8.2 The Number Density Evolution of Massive Galaxies for $z_{\mathbf{phot}} < 1.8$

**Figure 10** shows the evolving number density of $M_* > 10^{11}$ M$_\odot$ galaxies for $z_{\rm phot} < 1.8$. As before, the $z \approx 0$ point comes from our analysis of the low-

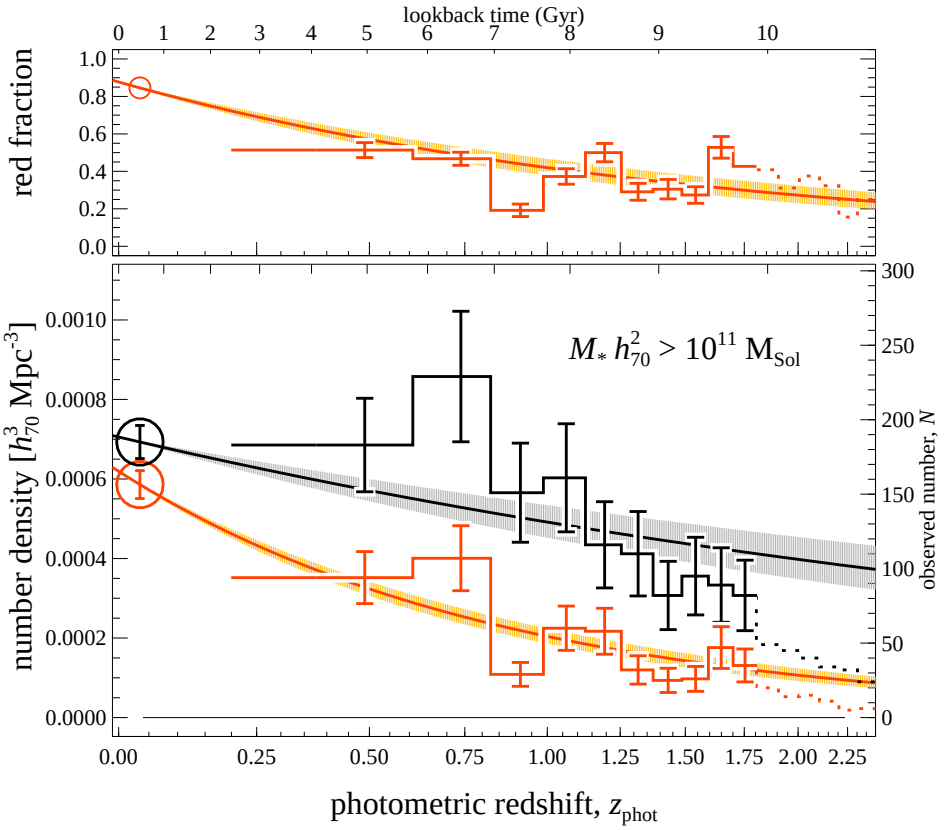


**Figure 10.** — The rise of massive, red galaxies over cosmic time. — *Lower panel*: evolution in the number density of all galaxies (*black histograms*) and of red galaxies (*red histograms*) with $M_* > 10^{11}$ $M_\odot$ in the MUSYC ECDFS catalog; The error bars shown reflect the expected field–to–field variation, calculated as per Somerville et al. (2004). *Upper panel*: evolution in the red galaxy fraction among $M_* > 10^{11}$ $M_\odot$ galaxies, as a function of photometric redshift. In this panel, the error bars have been derived by bootstrap resampling. In both panels, the dotted histograms show where we are significantly affected by incompleteness. The $z \approx 0$ points (*circles*) are derived from the low-z sample, discussed in Appendix A. The smooth curves show our fits to the observations; the shaded regions show the statistical $1\sigma$ uncertainties in the fits: we find $\gamma_{\rm tot} = -0.52 \pm 0.12$, $\gamma_{\rm red} = -1.70 \pm 0.14$, and $\gamma_{\rm frac} = -1.17 \pm 0.18$. Note that, in Section 9, we find that the accuracy of these fits is limited by systematic, rather than statistical, uncertainties.

z sample discussed in Appendix A; the histograms are for the main MUSYC ECDFS sample. For the $z \gg 0$ galaxies, since we have used bins of equal comoving volume, the observed numbers (*right axes*) can be directly related to a comoving number density (*left axes*), modulo uncertainties in the cosmological model. The black point/histograms refer to the total $M_* > 10^{11}$ $M_\odot$ population; the red point/histograms refer to red galaxies only; the dotted lines show where our results are significantly affected by incompleteness.

The error bars on the $z \gg 0$ histograms include the estimated measurement uncertainty due to field–to–field variation, derived as in Somerville et al. (2004), but modified for cuboid rather than spherical volumes (R Somerville, 2006; private communication). For any single measurement, this is the dominant source of uncertainty: typically $\sim 30$ %, as compared to random photometric errors, which are at the $\sim 10$ % level. Note, however, that each of the $z_{\rm phot} \gtrsim 1$ bins contains its own 'spike' in the $z_{\rm spec}$ distribution (Vanzella et al., 2008); the $0.6 < z_{\rm phot} < 0.8$ bin contains two, and the $0.8 < z_{\rm phot} < 1.0$ bin none. Further, note that for $z_{\rm phot} \gtrsim 1$, our redshift errors are comparable to the size of the bins themselves; in this sense, the densities in neighboring bins are correlated, and the variations due to large scale structure are to a certain extent masked.

The evolution in the number density of massive galaxies is only moderate: the number of $M_* > 10^{11}$ $M_\odot$ galaxies at $1.5 < z_{\rm phot} < 1.8$ is $52 \pm 8$ % of the $z \approx 0$ value. While we see a very similar trend in mass density, a handful of galaxies with inferred $M_* \gg 10^{12}$ $M_\odot$ make this measurement considerably noisier.

In order to quantify the observed evolution, we have made a parametric fit to our measured number densities of the form:

$$n_{\rm tot}(z) = n_0 \ (1+z)^{-\gamma_{\rm tot}} \ . \tag{3}$$

In practice, since the dominant source of statistical uncertainty in any single point is from field–to–field variance, we perform a linear fit in $\log n$–$\log(1+z)$ space, weighting each point according to its uncertainty as shown in Figure 10. Further, we constrain the fits to pass through the $z \approx 0$ point, effectively eliminating $n_0$ as a free parameter. In this way, we find $\gamma_{\rm tot} = -0.52 \pm 0.12$. The fit itself is shown in Figure 10 as the smooth black curve; the shaded region shows the $(1\sigma)$ statistical uncertainty in the fit. Note, however, that we show in Section 9 that the accuracy of these fits is limited by systematic, rather than statistical, uncertainties.

### 8.3 Evolution in the Red Galaxy Fraction from $z_{\rm phot} = 1.8$ to the Present Day

We now turn our attention to the red galaxy population. From Figure 10, it is clear that the observed evolution is much stronger for red galaxies than it is for the total population: the number density of red galaxies at $1.5 < z_{\rm phot} < 1.8$ is $18 \pm 3$ % of the $z \approx 0$ value. Making a fit to the red galaxy number densities to quantify this evolution, in analogy to the previous section, we find $\gamma_{\rm red} = -1.60 \pm 0.14$ (*smooth red curve*).

A complementary way of characterizing the rise of massive red galaxies is to look at the evolution of the red galaxy fraction. There are several advantages to

focusing on the red galaxy fraction, rather than the number density of red galaxies *per se*. First and foremost, the uncertainty in the red fraction due to large scale structure and field–to–field variation should be considerably smaller than those in the number density, especially at $z \gtrsim 1$ (Cooper et al., 2007).

We show the red fraction as a function of photometric redshift in the upper panel of Figure 10. Fitting these results (*smooth red curve*) with the same functional form as in Equation 3, we find $\gamma_{\rm frac} = -1.06 \pm 0.16$. Taken together, the results encapsulated in Figure 10 suggest that $\lesssim 20$ % of $M_* > 10^{11}$ $M_\odot$ galaxies in the local universe were already on the red sequence by $z_{\rm phot} \approx 1.6$ (9.5 Gyr ago). By the same token, approximately 50 % of these galaxies only (re-)joined the red sequence after $z = 0.5$ (5.0 Gyr ago).

### 8.4 The Importance of the $z \approx 0$ Comparison Point

It is clear from Figure 10 that almost all the signal for $z \gg 0$ evolution in the red fraction comes from the $z \approx 0$ comparison point (see also Borch et al., 2006). Fitting to the $z \gg 0$ data alone, we find $f_{\rm red} = (0.53 \pm 0.05)\,(1 + z_{\rm phot})^{-0.29 \pm 0.17}$; that is, less than a $2\sigma$ signal of evolution. The same is also true for the number density measurements (see also Borch et al., 2006); the reasons for this are not just the relatively mild evolution, but also to the relative size of the error bars on the low– and high–redshift measurements. Fitting only to the $z \gg 0$ points, we find $\gamma_{\rm tot} = -1.55 \pm 0.84$ and $\gamma_{\rm red} = -1.77 \pm 1.84$; these fits 'overpredict' the $z \approx 0$ number densities by factors of 2.2 and 1.3, respectively.

In this sense, then, rather than quantifying the absolute evolution in the $z \gg 0$ population, we are performing a difference measurement between the situations at $z \approx 0$ and $z \gg 0$. For this reason, it is imperative that care is taken when deriving the $z \approx 0$ comparison values to ensure that the low– and high–redshift samples have been analyzed in a uniform way (Appendix A).

In comparison to more sophisticated analyses by Bell et al. (2003) and Cole et al. (2001), both of which combine the 2MASS and SDSS datasets, our $z \approx 0$ number densities are approximately 10 % higher, and 15 % lower, respectively. Adopting these values in place of our own determinations, we find $\gamma_{\rm tot} = -0.72 \pm 0.12$ using the Bell et al. (2003) mass function, and $\gamma_{\rm tot} = -0.42 \pm 0.12$ using the Cole et al. (2001) mass function, changes of $-0.2$ and $+0.1$ with respect to our default analysis. These rather large discrepancies underline the importance of uniformity in the analysis of high– and low–redshift galaxies.

### 8.5 Comparison with Other Works

In **Figure 11**, we compare our results to a selection of the steadily growing number of similar measurements. We show results from: the COMBO-17 survey (Borch et al., 2006), the GOODS-MUSIC catalog (Fontana et al., 2006), the K20 survey (Fontana et al., 2004), the VVDS (Pozzetti et al., 2007), the DEEP2 survey (Bundy et al., 2006), and MUNICS (Drory, 2004). (Note that all of these results have come within the last five years.) In all cases, the data in Figure 11 have been derived by integrating up a fit mass function, taking into account different choices

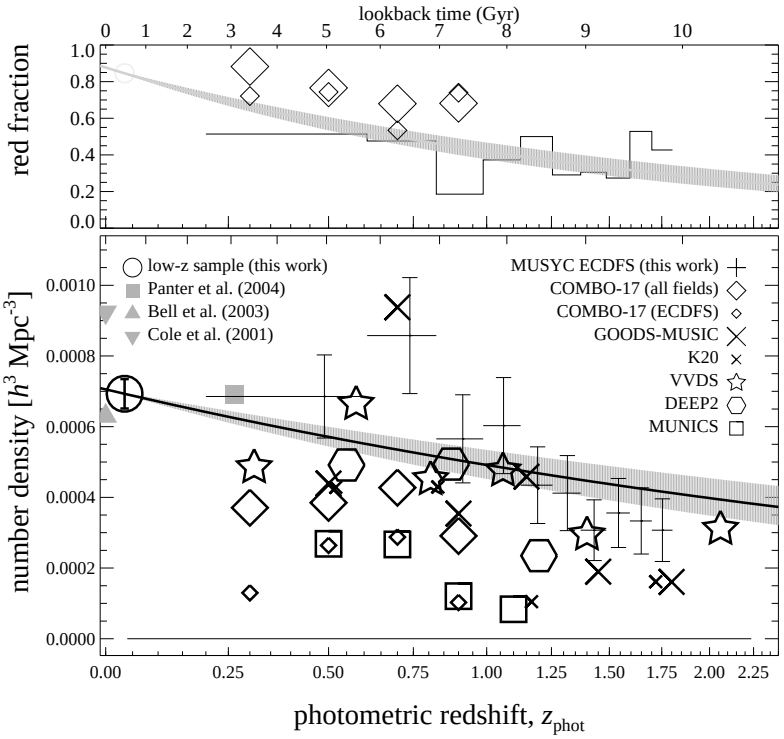


**Figure 11.** — Comparison with other works. — The different symbols correspond to different authors and surveys as marked: the filled symbols refer to other authors' analyses of SDSS data; references for other $z \gg 0$ surveys are given in the main text. As in Figure 10, the smooth curves show our fits to the MUSYC ECDFS data; the shaded regions show the statistical ($1\sigma$) uncertainties on the fits. Excepting the MUSYC ECDFS data points, all points have been derived by integrating up Schechter function fits to the observed mass functions, in redshift bins. Apart from the MUNICS and the COMBO-17 ECDFS data points, there is good agreement between the many different groups' results; the cause for the discrepancy between the COMBO-17 and MUSYC results in the ECDFS is the subject of Appendix B.

of cosmology and IMF. We note that the strong redshift spike at $z_{\rm phot} \sim 0.7$ is also present in the GOODS-MUSIC results, which are based on the CDFS.

The comparison with the COMBO-17 results deserves some further comment. While our results agree reasonably well with the combined results from three of the COMBO-17 fields (*large diamonds*), there is significant disagreement between our results and the COMBO-17 results in the ECDFS (*small diamonds*). In Appendix B, we present a detailed comparison between the MUSYC and COMBO-17 datasets, and demonstrate that the discrepancy between the two results is due to systematic calibration errors in the COMBO-17 photometry, rather than differences in our analyses. We note in passing that these calibration errors have only a small effect on the COMBO-17 photometric redshifts, and affect the ECDFS only; not the other COMBO-17 fields (Wolf et al., 2008).

Particularly given the significant uncertainties in these measurements, and with the possible exception of the MUNICS results, the degree of agreement between these surveys is impressive. However, in all of these surveys, field–to–field variance is a — if not the — major source of (random) uncertainty. More and larger fields are necessary to lock down the growth rate of massive galaxies.

# 9 Quantifying Potential Systematic Errors

Having now described our experiment in full, in this section we describe a wide array of sensitivity analyses, which are designed to determine how sensitively our results depend on specific choices in our experimental design. The basic idea is to vary individual aspects of our analysis, and see what effects these changes have on the results presented in Section 8; in particular, we will focus on how the best-fit values of $\gamma_{\rm tot}$ and $\gamma_{\rm frac}$ depend on our experimental design. The results of many of these tests are summarized in Tables 2 and 3. Note that we have used the same red galaxy selection criterion (Equation 2) throughout all of these tests.

We will consider, in turn, how sensitive or robust our results are to basic photometric calibration errors (Section 9.1), the methods used for the photometry itself (Section 9.2), incompleteness (Section 9.3), and variations on our photometric redshift computation (Section 9.4). We then discuss possible systematic effects arising from our method for estimating stellar masses in Section 9.5, and other concerns in Section 9.6. Our findings in the section are summarized in Section 9.7.

## 9.1 Photometric Calibration

How sensitive or robust are our results to errors in the basic photometric calibration? And how accurate are the calibrations of each of the different bands, relative to one another, and in an absolute sense? We address these two questions in this section.

### *9.1.1 The Effect of Perturbing Individual Bands*

In order to gauge the effect of photometric calibration errors on our final results, we have perturbed the photometry in each band in turn by $\pm 0.05$ mag, and repeated our full analysis, from the derivation of photometric redshifts and stellar masses to fitting the $\gamma$s. Roughly speaking, these perturbations can affect our results in two separate ways: either through direct changes in the SEDs themselves, or indirectly, through changes in the derived photometric redshifts, and so the transformation from observed to restframe quantities. In order to disentangle these two effects, we have also repeated our analysis with these $\pm 0.05$ mag shifts, but while holding the photometric redshifts fixed to their default values. The results of these tests are summarized in **Table 2**.

For the $UBV$ bands, the effect of changing the photometric calibration on our main results is driven almost exclusively by changes in the photometric redshifts. This is simply because our stellar mass estimates typically do not depend directly on the observed $UBV$ photometry: the restframe $B$ has already shifted past the observed $V$ by $z \approx 0.25$. (Recall that we use the $(B-V)$ color to infer $M_*/L_V$,

and thus $M_*$.) Similarly, the effect that changing the reddest $H$ and $K$ bands has on $\gamma_{\rm tot}$ is small: $\pm < 0.02$ for both. The situation changes for the $RIz'J$ bands, where the direct effect comes to dominate over the indirect effect from changes in the redshifts.

In terms of the SED photometry, our results are most sensitive to errors in the $R$ and $J$ band photometric calibrations: a $\pm 0.05$ mag shift in the $R$ or $J$ zeropoint changes the value of $\gamma_{\rm tot}$ by $\mp 0.08$ or $\pm 0.14$, respectively. For the $R$ band, the effect on restframe properties is focused on the $z_{\rm phot} \approx 0.7$ bin, where we happen to measure the highest density; this is also the $z \gg 0$ bin that has the single greatest effect on our measurement of $\gamma_{\rm tot}$. The critical importance of the $J$ band stems from the fact that it plays a role in the derivation of restframe $V$ photometry, and thus $M_*$, for all $z_{\rm phot} \gtrsim 1$; *i.e.*, roughly two thirds of our surveyed volume. It seems plausible that our sensitivity in this regard might be substantially reduced if we were to use a different method for estimating stellar masses (see also Section 9.5).

While the SED modeling results are not particularly sensitive to the $K$ band calibration, it still plays a critical role in normalizing each SED through the total $K$ flux measurement. Although changing the $K$ calibration by $\pm 0.05$ mag effectively rescales the stellar masses by just $\mp 0.02$ dex, this can change the inferred number densities by as much as $\mp 10$ %. In terms of $\gamma_{\rm tot}$, the effect is $\mp 0.07$.

### *9.1.2 Testing the MUSYC ECDFS Photometric Calibration*

We now turn our attention to identifying and quantifying potential calibration errors in the MUSYC ECDFS dataset. We will then be able to use this information to estimate the extent to which our results may actually be affected by such errors.

To address this question, we have used EAZY to fit main sequence stellar spectra from the BPGS stellar atlas (Gunn & Stryker, 1983), fixing $z_{\rm phot} = 0$, to the observed photometry for stars (see Section 5.2.3 of Chapter II). Assuming that whatever calibration errors do exist do not affect the choice for the best fit template spectrum, we can then interpret the mean difference between the best fit and the observed photometry as being the product of calibration errors. We note that since this test considers only SED shapes, it cannot comment on the absolute calibration of any given band; instead it assesses the relative– or cross–calibration of the ten band photometry that makes up the MUSYC ECDFS dataset.

The offsets we derive in this way are given in Col. (10) of Table 2. The most notable offsets derived in this way are $-0.048$ in $(U - U_{38})$, 0.033 in $(I - z')$, 0.064 in $(J - H)$, and 0.032 in $(J - K)$. These offsets give an indication of the plausible size of any calibration errors or inconsistencies in the MUSYC SED photometry. If we recalibrate our photometry to eliminate these apparent offsets (holding the $K$ band fixed), and repeat our full analysis, we find that the best fit values of $\gamma_{\rm tot}$, $\gamma_{\rm red}$, and $\gamma_{\rm frac}$ change by $+0.00$, $+0.02$, and $+0.04$, respectively.

We have also checked the absolute calibration of each band in comparison to the FIREWORKS catalog (Wuyts et al., 2008) in the region of overlap (Section 5.2.1 of Chapter II). Using $(B - z')$–$(z' - K)$ selected stars, we have constructed

| Band (1) | Holding $z_{\rm phot}$ fixed: change in $\gamma_{\rm tot}$ +0.05 mag (2) | change in $\gamma_{\rm tot}$ −0.05 mag (3) | change in $\gamma_{\rm tot}$ +0.05 mag (4) | change in $\gamma_{\rm tot}$ −0.05 mag (5) | Recalculating $z_{\rm phot}$: change in $\gamma_{\rm red}$ +0.05 mag (6) | change in $\gamma_{\rm red}$ −0.05 mag (7) | change in $\gamma_{\rm frac}$ +0.05 mag (8) | change in $\gamma_{\rm frac}$ −0.05 mag (9) | Photometric offset *cf.*: stellar SEDs (10) | FIREWORKS (11) |
|---|---|---|---|---|---|---|---|---|---|---|
| $U$ | +0.00 | +0.00 | +0.01 | −0.01 | +0.02 | +0.02 | +0.001 | +0.041 | −0.004 | +0.013 |
| $U_{38}$ | +0.00 | +0.00 | +0.00 | +0.01 | +0.01 | +0.02 | −0.002 | +0.009 | −0.051 | −0.020 |
| $B$ | +0.00 | +0.00 | +0.00 | +0.00 | +0.03 | +0.01 | +0.061 | −0.006 | −0.017 | +0.050 |
| $V$ | +0.01 | +0.00 | −0.02 | +0.00 | +0.04 | −0.10 | +0.037 | −0.068 | −0.006 | +0.038 |
| $R$ | +0.05 | −0.03 | +0.08 | −0.08 | +0.06 | −0.05 | +0.071 | −0.053 | +0.017 | +0.016 |
| $I$ | +0.04 | −0.03 | +0.01 | −0.02 | +0.03 | −0.03 | −0.051 | +0.066 | +0.023 | +0.055 |
| $z'$ | −0.02 | +0.01 | −0.03 | +0.02 | −0.01 | −0.01 | −0.011 | −0.024 | −0.011 | −0.004 |
| $J$ | −0.10 | +0.09 | −0.13 | +0.10 | −0.14 | +0.11 | +0.022 | −0.033 | +0.032 | +0.015 |
| $H$ | −0.03 | +0.03 | −0.02 | +0.01 | −0.02 | +0.05 | −0.009 | +0.039 | −0.032 | −0.012 |
| $K_{\rm SED}$ | +0.00 | +0.02 | −0.02 | +0.01 | −0.07 | +0.08 | −0.078 | +0.096 | — | — |
| $K_{\rm tot}$ | −0.07 | +0.07 | −0.07 | +0.07 | −0.03 | +0.03 | +0.057 | −0.025 | — | −0.017 |

**Table 2.** — Photometric calibration sensitivity tests. — For each band (Col. 1), Col.s (2)—(9) give how our parametric fits to the $z_{\rm phot} \lesssim 2$ evolution of the $M_* > 10^{11}$ M$_\odot$ popuation changes with $\pm$0.05 mag perturbations to individual bands' zeropoints: Col.s (2) and (3) show the effect on the total number density evolution of massive galaxies due to the change in the photometry only. Col.s (4) and (5) show the same information, but including the effect of changes in the photometric redshifts that come with changing the calibration; similarly, Col.s (6) and (7) show the effect on the number density measurements for red, massive galaxies, and Col.s (8) and (9) show the effect n the red galaxy fraction. Col.s (10) and (11) give zeropoint offsets for each band based on stellar colors and comparison with the FIREWORKS catalogue (Wuyts et al., 2008), respectively.

empirical 'color transforms' between FIREWORKS and MUSYC filters, as a function of HST color. Comparing these diagrams to predictions derived from the BPGS stellar spectral atlas (Gunn & Stryker, 1983), we treat any offset between the predicted and observed stellar sequences as a calibration error in the MUSYC photometry. (This is the same method we used to identify the disagreement between the COMBO-17 and MUSYC calibrations discussed in Appendix B; see Section 5.2.2 of Chapter II.)

The offsets we have derived in this way are given in Col. (11) of Table 2. The biggest offsets derived in this way are $\approx 0.05$ mag in ($B$-F435W) and ($I - F850LP$). The offset in ($K$-$K_{\rm ISAAC}$) is $-0.017$ mag; for the crucial $J$ band, the offset is just $+0.015$ mag. In agreement with the stellar SED fitting test, this analysis also finds an inconsistency between the $I$ and $z$ bands at the level of 0.05 mag. If we recalibrate our photometry to match the FIREWORKS catalog, we find that our values of $\gamma_{\rm tot}$, $\gamma_{\rm red}$, and $\gamma_{\rm frac}$ change by $-0.03$, $-0.13$, and $-0.08$ with respect to our default results. The sizes of these changes are in excellent agreement with the results of the previous section; more than half of these changes can be ascribed just to the 0.02 mag rescaling of total $K$ magnitudes.

We estimate that the systematic uncertainty in our main results due to photometric calibration errors is at the level of $\Delta\gamma_{\rm tot} \lesssim 0.05$ and $\Delta\gamma_{\rm frac} \sim 0.10$.

## 9.2 Photometric Methods

While we rely on SExtractor for our basic photometry, we have introduced three sophistications in our analysis. In this section, we investigate the effects that these three changes have on our results.

### *9.2.1 Background Subtraction*

We have applied a correction to account for the tendency of SExtractor to overestimate the background level; for individual objects, this correction is typically on the order of $-0.03$ mag. To test the sensitivity of our results to background subtraction errors, we have repeated our analysis without applying this correction. Repeating our analysis relying on our own background estimation we find changes in $\gamma_{\rm tot}$ and $\gamma_{\rm red}$ of $-0.06$ and $-0.05$, respectively; the change in $\gamma_{\rm frac}$ is just $-0.02$.

### *9.2.2 Total Magnitudes*

Our total flux measurements are based on SExtractor's `AUTO` flux measurement. It is well known that the `AUTO` aperture misses a significant amount of flux, especially for the faintest sources (Bertin & Arnouts, 1996; Labbé et al., 2003; Brown et al., 2007). Following Labbé et al. (2003), we partially redress this by applying a minimal correction to correct for light laying beyond the `AUTO` aperture, treating each object as if it were a point source. For individual sources, this correction amounts to as much as 20 %. Experiments with synthetic $r^{1/4}$ sources placed in our own image suggest that for the specific example of a $K = 22$ elliptical galaxies

with $R_e = 0\farcs4$ ($\approx 3$ kpc at $z = 1$), this correction reduces the missed flux from 0.33 mag to 0.15 mag.

In terms of the measured number densities, the effect of adopting this correction is 1–5 % for $z_{\rm phot} \lesssim 1.5$, rising to 5–10 % for $1.5 < z_{\rm phot} < 1.8$. The use of uncorrected `AUTO` fluxes as total flux measurements thus produces a rather mild spurious evolutionary signal. Repeating our analysis without this correction, we find slightly steeper evolution in the number densities: the values of $\gamma_{\rm tot}$, $\gamma_{\rm red}$ change by $-0.07$ and $-0.06$, respectively. By comparison, the red fraction measurements remain almost unchanged: $\gamma_{\rm frac}$ changes by $-0.01$.

### *9.2.3 SED Apertures and Color Gradients*

When constructing multicolor SEDs, we have used the larger of SExtractor's `ISO` aperture (based on the $1\farcs0$ FWHM $K$ mosaic) and a fixed $2\farcs5$ diameter aperture. This aperture flexibility is intended to guard against potential biases due to color gradients in individual galaxies. This problem is presumably the most severe for the largest, relatively low redshift galaxies with significant bulge components, leading to overestimates in both $(B - V)$ and $M_*$.

Consistent with this expectation, when repeating our analysis relying exclusively on the fixed $2\farcs5$ aperture fluxes to construct SEDs, we find the measured number densities of massive galaxies increases by $\sim 5$ % for $z_{\rm phot} < 1.1$. The increases in these low redshift bins brings them closer to the $z \sim 0$ point, leading to a slight decrease in the measured evolution: $\gamma_{\rm tot}$ and $\gamma_{\rm red}$ change by $+0.05$ and $+0.13$, respectively; $\gamma_{\rm frac}$ changes by just $+0.02$.

We estimate that the systematic uncertainty in our results associated with our photometric methods are on the order of $\gamma_{\rm tot} \lesssim 0.07$ and $\gamma_{\rm frac} \approx 0.00$.

## 9.3 Correcting for Incompleteness

In Section 4, we have argued that we are approximately complete for $M_* > 10^{11}$ $M_\odot$ galaxies for $z_{\rm phot} < 1.8$. In this section, we examine the effects of incompleteness due to both the $K < 22$ detection threshold and the $K$ S:N $> 5$ 'analysis' selection, by deriving simple completeness corrections.

### *9.3.1 Incompleteness Due to Low S:N*

In addition to incompleteness due to our $K < 22$ detection limit, discussed below, there is the concern of incompleteness due to the $K$ S:N $> 5$ selection limit, which we have imposed to ensure against extremely poorly constrained photometric redshifts. This cut affects 4.5 % of all $K \leq 22$ sources in the ECDFS catalog, with 43 % of those galaxies laying in the slightly shallower Eastern pointing. At a fixed magnitude, there is not a strong dependence of the fraction of S:N $< 5$ detections as a function of $(J - K)$ color or — with the caveat that this cut is designed to remove poorly constrained redshifts — as a function of $z_{\rm phot}$.

In order to assess the impact of this cut on our results, in this section we attempt to correct for this additional source of incompleteness. Our procedure is as follows: we determine the completeness fraction in the face of this selection,

as function of total $K$ magnitude, $f(K)$; this can then be used to weight each retained galaxy according to its $K$ magnitude, $w(K) = 1/f(K)$. Of galaxies in our main sample, 8 % of all galaxies, and 11 % of $M_* > 10^{11}$ $M_\odot$ galaxies at $1 \leq z_{\rm phot} \leq 2$ are given $w(K) > 1.25$.

Recalculating our number density measurements using these weights, we find that the number density of massive galaxies at $1.2 < z_{\rm phot} < 1.8$ changes by less than 5 %; the same is true for red galaxies alone. Repeating our analysis with these corrections, we find $\gamma_{\rm tot} = -0.49$, $\gamma_{\rm red} = -1.56$, and $\gamma_{\rm frac} = -1.05$, amounting to differences of $+0.03$, $+0.04$, and $+0.01$ with respect to our fiducial analysis. We get similar results adopting a more stringent S:N > 10 criterion. To put this change into perspective, it is comparable to that due to uncertainty in the background subtraction for our basic photometry.

### *9.3.2 Correcting for Undetected Sources*

We have trialed using a correction for incompleteness due to our $K < 22$ detection limit, which we have derived as follows: taking the $M_* > 10^{11}$ $M_\odot$ galaxy population in a given redshift bin, we then predict what the observed $K$ magnitude for each galaxy would be if the galaxy were shifted through the next most distant redshift bin. In other words, we effectively 'K correct' each galaxy's observed $K$ flux from its fiducial $z_{\rm phot}$ to higher redshifts, using the same machinery as for the interpolation of restframe fluxes. This makes it possible to determine the fraction (in terms of volume) of the next redshift bin over which a given galaxy would remain detectable; the overall completeness is then simply the average of this quantity for all galaxies in the bin. We have performed this correction for red and blue galaxies separately.

Based on this analysis, we are indeed 100 % complete for $z_{\rm phot} < 1.25$; for $1.7 < z_{\rm phot} < 1.8$, we are more than 80 % complete overall, and at least 75 % complete for red galaxies. This agrees reasonably well with our completeness estimates in Section 4. For higher redshifts, our estimated completeness drops to 75 % for $1.9 < z_{\rm phot} < 2.0$, and to 65 % for $2.1 < z_{\rm phot} < 2.2$. The estimated incompleteness correction factor to the measured number densities is thus less than 1.5 for all $z_{\rm phot} < 2.2$. Using these completeness corrections to extend our analysis to $z_{\rm phot} < 2.2$, we find $\gamma_{\rm tot} = -0.47$ and $\gamma_{\rm frac} = -0.90$; changes of $-0.05$ and $-0.16$ with respect to our default values for $z_{\rm phot} < 1.8$.

Fitting to these incompleteness-corrected measurements for the number density of massive galaxies with $z_{\rm phot} < 1.8$ (*i.e.*, repeating our default analysis, but with a completeness correction), we find $\gamma_{\rm tot} = -0.37$, $-1.42$, and $\gamma_{\rm frac} = -1.06$. In comparison to our uncorrected results for $z_{\rm phot} < 1.8$, these represent changes of $-0.15$, $-0.18$, and $-0.00(3)$, respectively. However, we also note that if we were to assume that we are 100 % complete for $z_{\rm phot} < 1.4$, these changes become $-0.10$, $-0.12$, and $+0.01$; that is, the uncertainties on these incompleteness corrections are comparable to in size to the corrections themselves.

## 9.4 Photometric Redshifts

The next major aspect of our analysis that we will explore in detail is systematic effects associated with the derivation of photometric redshifts; we split this discussion into two parts. In the first part (Section 9.4.1) we explore how our results depend on the choice of templates used in the $z_{\rm phot}$ calculation. Then, in the second (Section 9.4.2), we investigate how our results depend on the exact method used for deriving photometric redshifts by varying individual aspects of the EAZY algorithm. Some illustrative results from a selection of these sensitivity tests are given in **Figure 12**.

### *9.4.1 Trialing Different Template Sets*

*Using the Fontana et al. (2006) template set* — Using a library of $\sim 3000$ synthetic Pégase V2.0 (Le Borgne & Rocca-Volmerange, 2002) spectra described by Fontana et al. (2006), Grazian et al. (2006) obtain $\sigma_z \approx 0.045$ for $K$ selected galaxies from their catalog of the GOODS-CDFS data. In fact, this library provides the parent catalog for EAZY's default template set (see Brammer et al., 2008). If we use this template library in place of the EAZY default, and do not allow template combination, we find $\sigma_z = 0.039$; the fraction of objects having $|\Delta z|/(1+z) > 0.1$, $f_{0.1}$, is 16 % (*cf.* 12.4 % for our default redshifts). Adopting these redshifts in place of our default determinations, we find $\gamma_{\rm tot} = -0.55$ and $\gamma_{\rm frac} = -1.00$; differences of $-0.03$ and $-0.06$, respectively (see Figure 12).

*Using the Rudnick et al. (2003) template set* — In the past, our group has tended to determine photometric redshifts as per Rudnick et al. (2003), which considers linear combinations among: four empirical template spectra from Coleman, Wu & Weedman (1980); two starburst spectra from Kinney et al. (1996); and two Bruzual & Charlot (2003) synthetic spectra for SSPs with ages of 10 Myr and 1 Gyr. If we use these templates with EAZY, we find $\sigma_z = 0.055$, with $f_{0.1} = 27$ %. Looking at Figure 12, these photometric redshift determinations suffer from a number of systematics whereby there are clear preferred redshift solutions at, e.g., $z_{\rm phot} \sim 0.4$, as well as a larger systematic underestimate of the true redshift: in addition to the larger random error, we also find a systematic offset in $\Delta z/(1+z)$ of $-0.054$. The effect this has on the $z_{\rm phot} \lesssim 0.8$ number densities is strong, and is dominated by the shifting of many of the galaxies from the $0.6 \lesssim z_{\rm phot} \lesssim 0.8$ bin down into the $0.2 \lesssim z_{\rm phot} \lesssim 0.6$ bin. The net effect of this change on our final results is considerably less, however: using these redshifts, we find $\gamma_{\rm tot} = -0.58$ and $\gamma_{\rm frac} = -1.11$; changes of $-0.06$ and $-0.05$ with respect to our fiducial results.

*Emulating hyperz with empirical templates* — As a final test for this section, we have also repeated our photometric redshift analysis emulating the popular hyperz (Bolzonella, Miralles & Pello, 2000) code in its default configuration; that is, $\chi^2$ minimization between the observed photometry and synthetic photometry using four Coleman, Wu & Weedman (1980) empirical template spectra, and making no allowance for dust extinction beyond what is included in the empirical spectra. Using the spectroscopic redshift sample shown in Figure 1, we find $\sigma_z = 0.060$, and again a substantial systematic offset: $\Delta z/(1+z) = -0.051$.

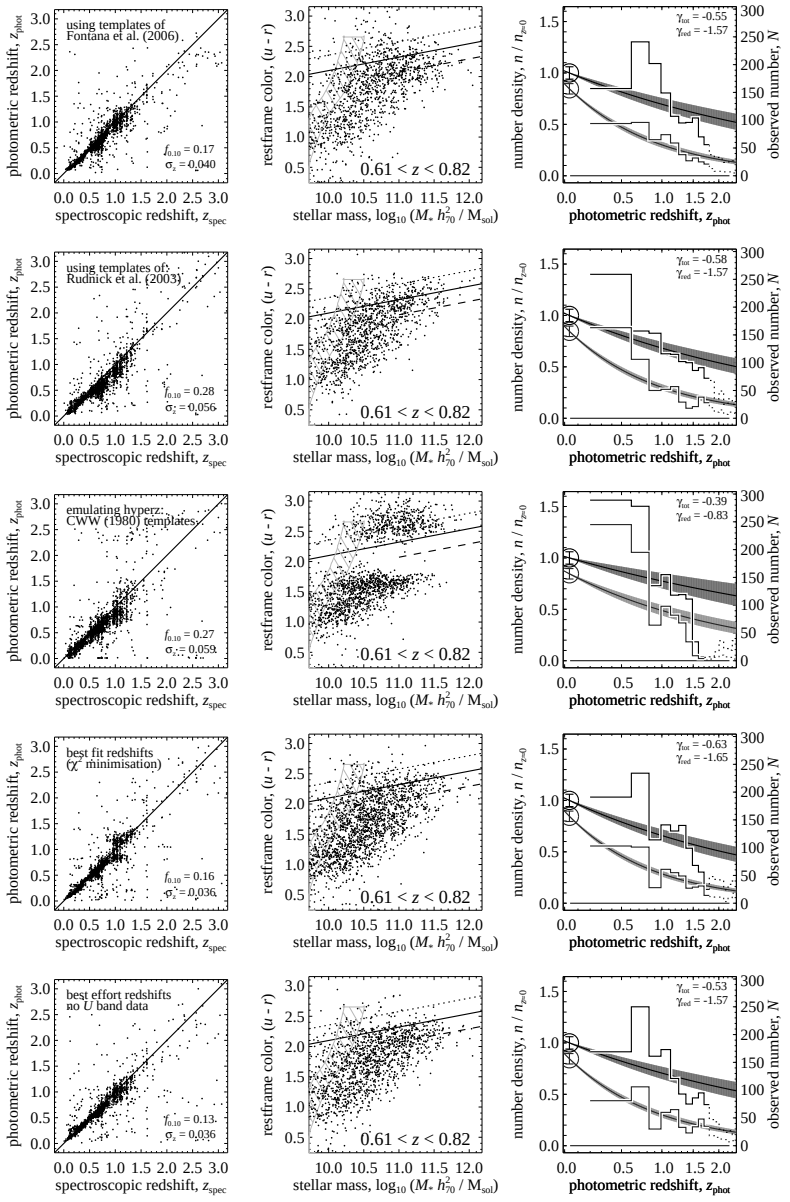


**Figure 12.** — The effect of different photometric redshift analyses on our results. — In rows, we show how our results would change (*top to bottom*): adopting the template set used by Rudnick et al. (2003); adopting the template set described by Fontana et al. (2006); emulating hyperz in its default configuration; adopting the maximum likelihood (*i.e.*, $\chi^2$ minimization) photometric redshift value; neglecting the $U$ band photometry. In each row, we show (*left to right*): the analogs of Figure 1; the third panel of Figure 5; and Figure 10.

One very striking consequence of using so few template spectra is a very pronounced biasing of galaxy restframe colors, corresponding to the colors of the template spectra themselves (see Figure 12). This of course has a very strong effect on our stellar mass estimates. For this reason, it is critical when relying on photometric redshifts to derive restframe properties to use a template basis set that spans the full range of galaxy colors.

### *9.4.2 Variations in our Default Photometric Redshift Calculation*

*Adopting the best fit redshift* — By default, EAZY assigns each object a redshift by taking a probability weighted integral over the full redshift grid. However, it also outputs the single most likely redshift, determined by $\chi^2$ minimization. For these redshifts, we find $\sigma_z = 0.035$, and $f_{0.1} = 14.7$ %. For $1 < z_{\rm spec} < 2$, however, we find $\sigma_z = 0.086$. We also see evidence for rather strong systematic effects in the photometric redshifts, such that there are preferred photometric redshift solutions, for example at $z_{\rm phot} \sim 0.8$ and 1.2 (see Figure 12), corresponding to the points where the optical breaks fall between the observed filters. With these redshifts, the values of $\gamma_{\rm tot}$, $\gamma_{\rm red}$, and $\gamma_{\rm frac}$ change by $-0.11$, $-0.05$, and $-0.14$, respectively.

*No Luminosity Prior* — Our default analysis makes use of a Bayesian $K$ band luminosity prior. If we do not include this feature, we find $\sigma_z = 0.038$ and $f_{0.1} = 13.5$ %. The best fit values of $\gamma_{\rm tot}$, $\gamma_{\rm red}$, and $\gamma_{\rm frac}$ change by $-0.11$, $-0.38$, and $-0.54$, respectively. These changes are large; however, we note that these results are not internally consistent, in that it is not true that $\gamma_{\rm frac} \approx \gamma_{\rm red}/\gamma_{\rm tot}$. Further, these results are not consistent with the test we describe in Section 10.

*No Template Error Function* — A novel feature of the EAZY code is the inclusion of a template error function; *i.e.*, a systematic error, as a function of restframe wavelength, designed to down-weight those parts of the spectra like the restframe near-UV where galaxies show an greater intrinsic variation. If we do not include this feature, we find $\sigma_z = 0.034$, and $f_{0.1} = 15.3$ %; the best fit values of $\gamma_{\rm tot}$, $\gamma_{\rm frac}$, and $\gamma_{\rm frac}$ change by $-0.04$, $-0.14$, and $-0.13$, respectively.

*Two Template Combination* — By default, we allow non-negative combinations of all six EAZY templates when fitting to the observed SEDs to determine the redshift. If we instead allow combinations of only two (but any two) of the template spectra, we find $\sigma_z = 0.038$ and $f_{0.1} = 14.4$ %. The fit values of $\gamma_{\rm tot}$ and $\gamma_{\rm frac}$ change by $+0.02$ and $-0.02$, respectively.

*'Best Effort' Photometric Redshifts* — It is well known that the WFI $U$ filter suffers from a red leak beyond 8000 Å. We find that we get our best $z_{\rm phot}$–$z_{\rm spec}$ agreement if we ignore the $U$ band photometry in the computation of photometric redshifts. For the spectroscopic sample shown in Figures 1 and 2, we find $\sigma_z = 0.035$ and $f_{0.1} = 12.0$ %; this translates into an uncertainty in the stellar mass estimates due to redshift uncertainties of just 0.1 dex. This modest improvement in our photometric redshift estimates changes the fit values of $\gamma_{\rm tot}$ and $\gamma_{\rm frac}$ change by just $-0.01$ and $+0.05$, respectively.

These tests show that our results do depend rather sensitively on the details of the photometric redshift calculation, typically at the level of $\Delta\gamma \lesssim 0.15$. In

particular, the inclusion of a luminosity prior is crucial in shaping our results. However, we note that in most cases there are objective reasons to favor our default redshifts and/or results. Similarly, the use of different template sets changes our results at the level of $\Delta\gamma \approx 0.06$. The effect of minor refinements to our photometric redshift calculation (two template combination versus unrestricted; the inclusion/omission of the $U$ band data) is small: $\Delta\gamma \lesssim 0.05$. Note that, in any case, changing the photometric redshift calculation cannot change our qualitative results: a gradual increase in the red galaxy fraction with time.

### 9.5 Stellar Mass Estimates

We have relied on a rather simple method for estimating stellar mass–to–light ratios, based on restframe $(B-V)$ colors. Certainly this method does not make use of the full amount of information available in the full, ten-band SED; on the other hand, it does ensure that the same information is used for all galaxies, irrespective of redshift.

Nevertheless, we have trialed reanalyzing our data using standard population synthesis SED fitting techniques to estimate stellar masses for our main sample, in order to assess the potential scale of biases arising from this aspect of our experimental design. Specifically, we have fit Bruzual & Charlot (2003) synthetic spectra to the observed photometry, with redshifts fixed to the fiducial $z_{\rm phot}$, using the hyperzspec utility, which is a part of the hyperz v1.2 release package (M Bolzonella 2007; private communication).

While the random scatter in comparison to the color-derived stellar mass estimates is 0.18 dex, after correcting for IMF, cosmology, etc., the SED-fit masses are systematically lower by 0.3 dex. This is also true for both the FIREWORKS (Wuyts et al., 2008) or GOODS-MUSIC (Grazian et al., 2006) catalogs of the CDFS-GOODS data. Besides this offset, we do not see evidence for strong evolution in the normalization of the color relation in comparison to the SED fit masses (but see Lin et al., 2007). Since we have not been able to identify the source of this offset, we have simply corrected for it. Note, however, that we have not refit the $z \approx 0$ SEDs for this test.

Using these SED-fit stellar masses, we find $\gamma_{\rm tot} = -0.49$ $\gamma_{\rm red} = -1.62$, and $\gamma_{\rm frac} = -1.11$. These represent differences of just $+0.03$, $-0.02$, and $-0.05$ with respect to our fiducial values. Similarly, using the prescription for $M/L$ as a function of $(B-V)$ from Bell et al. (2003) rather than that from Bell & de Jong (2001), we find $\gamma_{\rm tot} = -0.41$, $\gamma_{\rm red} = -1.55$, and $\gamma_{\rm frac} = -1.12$; changes of $+0.11$, $+0.05$, and $-0.06$, respectively.

This suggests that our results are not grossly effected by our choice of method for estimating stellar masses, and especially not by the use of color– as opposed to SED–derived stellar masses.

### 9.6 Other Potential Systematic Effects

*Large Scale Structure and Field–to–Field Variance* — One interesting test of the effect of field–to–field variance is to restrict our attention only to the MUSYC

ECDFS coverage of the GOODS-CDFS region. For this 143 □' sub-field, we find $\gamma_{\rm tot} = -0.62$, $\gamma_{\rm red} = -1.37$, and $\gamma_{\rm frac} = -1.00$, differences of –0.10, +0.23, and +0.06 with respect to the full 818 □' ECDFS. In comparison to the full ECDFS, the dearth of galaxies at $z_{\rm phot} \sim 0.8$ is even more pronounced, and seems to extend over the range $0.8 < z_{\rm phot} < 1.2$; this appears to be the main driver of the strong change in $\gamma_{\rm red}$. Conversely, if we exclude the significantly underdense GOODS area, we find $\gamma_{\rm tot} = -0.49$, $\gamma_{\rm red} = -1.60$, and $\gamma_{\rm frac} = -1.10$; differences of +0.03, −0.00, and −0.04, respectively. Again, this analysis underscores the fact that measurements of the red fraction are more robust against the effects of large scale structure than those of the number densities themselves.

*The $z \approx 0$ value* — In Section 8.4, we have shown that the $z \approx 0$ comparison point is critical in providing most of the signal for $z \gg 0$ evolution in the massive galaxy population. We have also considered our results vary if we change the way we treat the $z \approx 0$ point. In our default analysis, we constrain the fits so that they pass through the $z \approx 0$ points, effectively using the $z \approx 0$ point to normalize the high-redshift measurements. Since the (Poisson statistical) errors on the $z \approx 0$ points are just a few percent, our results do not change significantly if we include the $z \approx 0$ points in the fits.

For our default analysis, we have approximately accounted for the effect of photometric redshift errors, as they apply to the $z \gg 0$ sample, on the $z \approx 0$ measurements; we have done this by randomly perturbing the masses and colors of $z \approx 0$ galaxies using the typical uncertainties for $z \gg 0$ galaxies, due to the use of photometric redshifts, and given in Figure 2. Systematic and random errors in the number densities both of all and of red galaxies are on the order of 5 %. If we do not account for the Eddington bias in this way, we find that $\gamma_{\rm tot} = -0.46$ and $\gamma_{\rm frac} = -1.16$, changes of 0.06 and -0.10, respectively.

*Correcting for Galactic Dust Extinction* — Note that we have not included specific corrections for Galactic foreground dust extinction. The CDFS was specifically chosen for its very low Galactic gas and dust column density; the suggested corrections for the optical bands are $\lesssim 0.05$ mag. These corrections are typically as large or larger than the uncertainties on the photometric zeropoints themselves; it is for this reason that we have chosen not to apply these corrections. If we were to include these corrections, however, our derived values of $\gamma_{\rm tot}$ and $\gamma_{\rm frac}$ change by −0.03 and +0.04, respectively.

*Including Spectroscopic Redshift Determinations* — From among the thousands of spectroscopic redshift determinations that are publicly available in the ECDFS, we have robust $z_{\rm spec}$s for just over 20 % (1669/7840) of galaxies in our main sample, and for 20 % (269/1297) of those with $0.2 < z_{\rm phot} < 2.0$ and $M_* > 10^{11}$ M$_\odot$. Repeating our analysis with this additional information included, our results do not change significantly: we find $\gamma_{\rm tot} = -0.50$ and $\gamma_{\rm red} = -1.70$, differences of just +0.02 and +0.00, respectively. Similarly, we show in Appendix B that our $z \lesssim 0.8$ results do not change by more than a few percent if we adopt photometric redshifts from COMBO-17 (Wolf et al. 2004; $\sigma_z = 0.035$).

*Excluding X-ray-Selected Galaxies* — Note that we have made no attempt to exclude Type I AGN or QSOs from our analysis. In order to discover how

significant an omission this is, we have repeated our analysis excluding all those objects appearing in the X-ray selected catalogs of Szokoly et al. (2004) and Treister et al. (2008). This excludes 3.2 % (250/7840) of galaxies from our main sample, and 6.5 % (96/1482) from our $M_* > 10^{11}$ $M_\odot$ sample. Given these numbers, it is perhaps unsurprising that the exclusion of these objects does not greatly affect our results: $\gamma_{\rm tot}$ and $\gamma_{\rm frac}$ change by $-0.07$ and $+0.03$, respectively.

### 9.7 Quantifying Potential Systematic Errors — Summary

How, and how much, do our results depend on our analytical methods and experimental design? We have now described a rather large number of sensitivity tests, designed both to identify which aspects of our experimental design are crucial in determining our results, as well as to estimate the size of lingering systematic errors in our results. The results of many of these tests are summarized in Table 3. Clearly, these sorts of tests can provide a staggering array of 'metadata', offering insights to guide not only the interpretation of the present data and results, but also the design of future experiments.

As we have shown in Section 8.4, the systematic uncertainties in the $\gamma$s due to discrepancies between the treatment of $z \approx 0$ and $z \gg 0$ galaxies is on the order of $\pm 0.2$; this is the single greatest potential source of systematic uncertainty. Since we have treated these two samples in a uniform manner, we do not consider this as a lingering source of systematic uncertainty.

The dominant sources of systematic uncertainty are, in order of importance: the method of deriving photometric redshifts ($\Delta\gamma_{\rm tot} \approx 0.15$, $\Delta\gamma_{\rm frac} \approx 0.15$); photometric calibration errors ($\Delta\gamma_{\rm tot} \approx 0.05$; $\Delta\gamma_{\rm frac} \approx 0.10$); incompleteness ($\Delta\gamma_{\rm tot} \approx 0.1$, $\Delta\gamma_{\rm frac} \approx 0$); the choice of photometric redshift template set ($\Delta\gamma_{\rm tot} \approx 0.06$; $\Delta\gamma_{\rm frac} \approx 0.06$); photometric methods ($\Delta\gamma_{\rm tot} \approx 0.07$; $\Delta\gamma_{\rm frac} \approx 0$); and stellar mass estimates ($\Delta\gamma_{\rm tot} \approx 0.05$; $\Delta\gamma_{\rm frac} \approx 0.05$). We can also, for example, dismiss incompleteness due to our $K$ S:N criterion, contamination of the sample by QSOs, and minor details of our photometric redshift calculation as significant sources of systematic uncertainty.

In summary, we estimate the systematic errors on the measured values to be $\Delta\gamma_{\rm tot} = 0.21$ and $\Delta\gamma_{\rm frac} = 0.20$; these values have been obtained by adding the dominant sources of systematic error in quadrature. We note that previous studies have not generally taken (all) sources of systematic error into account in their analysis.

## 10 A Final Independent Consistency Check: Quantifying the Evolution of Bright, Red Galaxies Without Redshifts

Our goal in this Chapter has been to quantify the growth of massive galaxies in general, and of massive red galaxies in particular, over the past 10 Gyr, or since $z \sim 2$. We have now seen — not completely unexpectedly — that the use of photometric redshifts is potentially a major source of systematic error in such measure-

| Test description | $1.5 < z_{\rm phot} < 1.8$ | | $z_{\rm phot} < 1.8$ | | | $z \approx 2$ | |
|---|---|---|---|---|---|---|---|
| | $n_{\rm tot}/n_0$ | $n_{\rm red}/n_0$ | $\gamma_{\rm tot}$ | $\gamma_{\rm red}$ | $\gamma_{\rm frac}$ | $n_{\rm tot}/n_0$ | $n_{\rm red}/n_0$ |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Fiducial Results* | | | | | | | |
| Default analysis ($z_{\rm phot} < 1.8$) | $0.52 \pm 0.08$ | $0.18 \pm 0.03$ | $-0.52 \pm 0.12$ | $-1.60 \pm 0.14$ | $-1.06 \pm 0.16$ | 0.57 | 0.15 |
| Correcting for incompleteness ($z_{\rm phot} < 2.2$) | 0.57 | 0.20 | $-0.47 \pm 0.09$ | $-1.39 \pm 0.10$ | $-0.90 \pm 0.11$ | 0.61 | 0.19 |
| *Photometric Calibration* (§9.1.2) | | | | | | | |
| Recalibration to match stellar SEDs | 0.54 | 0.18 | $-0.52 \pm 0.12$ | $-1.58 \pm 0.14$ | $-1.02 \pm 0.17$ | 0.58 | 0.16 |
| Recalibration to match FIREWORKS$^{(*)}$ | 0.49 | 0.15 | $-0.55 \pm 0.12$ | $-1.73 \pm 0.14$ | $-1.14 \pm 0.15$ | 0.56 | 0.13 |
| *Photometric Methods* (§9.2) | | | | | | | |
| No correction for background oversubtraction | 0.49 | 0.16 | $-0.58 \pm 0.12$ | $-1.65 \pm 0.14$ | $-1.08 \pm 0.17$ | 0.54 | 0.15 |
| No correction for missed flux | 0.49 | 0.16 | $-0.59 \pm 0.12$ | $-1.66 \pm 0.14$ | $-1.07 \pm 0.17$ | 0.53 | 0.14 |
| Fixed $2\farcs5$ apertures for SEDs | 0.52 | 0.17 | $-0.47 \pm 0.12$ | $-1.47 \pm 0.14$ | $-1.04 \pm 0.14$ | 0.61 | 0.18 |
| *Incompleteness Corrections* (§9.3) | | | | | | | |
| Correcting for incompleteness ($z_{\rm phot} < 1.8$) | 0.57 | 0.20 | $-0.37 \pm 0.12$ | $-1.42 \pm 0.13$ | $-1.06 \pm 0.15$ | 0.67 | 0.19 |
| Correcting for $K$ S:N $< 5$ incompleteness | 0.54 | 0.18 | $-0.49 \pm 0.12$ | $-1.56 \pm 0.14$ | $-1.05 \pm 0.16$ | 0.59 | 0.16 |
| *Using Other $z_{phot}$ Template Sets* (§9.4.1) | | | | | | | |
| Using Rudnick et al. (2003) templates | 0.52 | 0.14 | $-0.58 \pm 0.12$ | $-1.57 \pm 0.14$ | $-1.11 \pm 0.15$ | 0.54 | 0.16 |
| Using Fontana et al. (2006) templates$^{(*)}$ | 0.53 | 0.21 | $-0.55 \pm 0.12$ | $-1.57 \pm 0.14$ | $-1.00 \pm 0.09$ | 0.56 | 0.16 |
| Emulating hyperz with CWW (1980) templates | 0.48 | 0.24 | $-0.39 \pm 0.14$ | $-0.83 \pm 0.15$ | $-0.69 \pm 0.15$ | 0.66 | 0.35 |
| *Variations on our Default $z_{phot}$ Analysis* (§9.4.2) | | | | | | | |
| Adopting the best fit redshift | 0.55 | 0.22 | $-0.63 \pm 0.13$ | $-1.65 \pm 0.14$ | $-0.92 \pm 0.14$ | 0.51 | 0.15 |
| No $K$ luminosity prior | 0.41 | 0.09 | $-0.63 \pm 0.12$ | $-1.98 \pm 0.15$ | $-1.59 \pm 0.22$ | 0.51 | 0.10 |
| No template error function | 0.53 | 0.15 | $-0.56 \pm 0.12$ | $-1.74 \pm 0.14$ | $-1.19 \pm 0.20$ | 0.55 | 0.13 |
| Two template combinations | 0.54 | 0.16 | $-0.50 \pm 0.12$ | $-1.56 \pm 0.14$ | $-1.08 \pm 0.15$ | 0.59 | 0.16 |
| Neglecting $U$ data ('best effort' $z_{\rm phot}$s) | 0.46 | 0.19 | $-0.53 \pm 0.12$ | $-1.59 \pm 0.14$ | $-1.01 \pm 0.16$ | 0.57 | 0.16 |
| *Other Variations* (§9.5 and §9.6) | | | | | | | |
| SED–fit stellar masses | 0.48 | 0.19 | $-0.49 \pm 0.12$ | $-1.62 \pm 0.14$ | $-1.11 \pm 0.14$ | 0.59 | 0.15 |
| Including spectroscopic redshifts | 0.51 | 0.18 | $-0.54 \pm 0.12$ | $-1.60 \pm 0.14$ | $-1.06 \pm 0.14$ | 0.56 | 0.15 |
| Excluding X-ray detections | 0.50 | 0.18 | $-0.59 \pm 0.12$ | $-1.63 \pm 0.14$ | $-1.03 \pm 0.15$ | 0.53 | 0.15 |
| GOODS area only | 0.47 | 0.22 | $-0.63 \pm 0.17$ | $-1.37 \pm 0.24$ | $-1.00 \pm 0.25$ | 0.51 | 0.20 |

**Table 3.** Summary of sensitivity test results. — Each row describes a different sensitivity test, designed to determine how and how much our main results depend on our choice of experimental design. These tests are described in §9. Each row gives an identifier (Col. 1) for a different test; section references are given for a description of each. For the *Dissecting our Default $z_{phot}$ Analysis* tests, additional $z_{\rm phot}$ information is given in §9.4.2. For each test, we give the measured number density of $M_* > 10^{11}$ $M_\odot$ galaxies with $1.5 < z_{\rm phot} < 1.8$: these values are given for the total massive galaxy samples (Col. 2), and the red subpopulation (Col. 3), and are normalized relative to the $z \approx 0$ value for all galaxies. We also give the best fit values for the total $0.2 < z_{\rm phot} < 1.8$ evolution in the number density of all massive galaxies (Col. 4), in the number density of red, massive galaxies (Col. 5), and in the red fraction among massive galaxies (Col. 6). Finally, we give values for the number density of massive galaxies (Col. 7) and of red, massive galaxies (Col. 8), obtained by extrapolating the fits to $z = 2$. Dominant effects are marked with an asterisk; a summary of our interpretations of these results can be found in §9.7.

ments. For this reason, in this section we present a complementary and completely independent measurement which does not rely on redshift information *at all*.

### 10.1 Selecting Red Galaxies in Redshift Intervals

**Figure 13** shows evolutionary tracks for a passively evolving stellar population, in terms of observed colors in the MUSYC ECDFS bands. Specifically, we have used Pégase V2.0 (Le Borgne & Rocca-Volmerange, 2002) models with a short star formation episode ($e$-folding timescale of 100 Myr), beginning with $Z = 0.004$ at $z_{\rm form} = 5$, and ending with $Z \approx {\rm Z}_{\odot}$. This (approximately) maximally old model describes a 'red envelope' for the color–redshift relation for observed galaxies: at a given redshift only extreme dust extinction can produce observed colors redder than these tracks.

As is evident in this figure, as the Balmer and/or 4000 Å breaks are first redshifted between the observed bands, there is a sharp rise in the observed color — by selecting objects that are very red in a certain color, it is thus possible to select galaxies that lie beyond a certain redshift. This is completely analogous to the ERO (McCarthy, 2004a, and references therein) or DRG (Franx et al., 2003) selection criteria for red galaxies at moderate–to–high redshifts. The particular selections we have adopted in Figure 13 (and given in Col. 1 of Table 4) translate to minimum redshifts of approximately 0.16, 0.40, 0.66, 1.12, and 2.05.

By applying several of these color criteria in concert, with the reddest criteria given primacy, it is then possible to select red galaxies in rough redshift intervals: the highest redshift galaxies are selected by the criterion $(J - K) > 1.4$; then, the next highest redshift galaxies are then selected from the remaining set (*i.e.* $(J - K) < 1.4$) by the criterion $(I - J) > 1.4$, and so on. In the current context, this can be thought of as a two-color, binned photometric redshift.

The prevalence of objects selected in this way are given in Col. 4 of **Table 4**. To our limit of $K < 22$, we find 386, 655, 690, 304, and 286 objects selected by this tiered set of color criteria.

### 10.2 Interpreting the Numbers of Color Selected Galaxies

It is clear that the exact redshift range over which a galaxy might satisfy a given color criterion depends on that galaxy's SED: as well as passively evolving galaxies, these selections will also include galaxies whose red observed colors are due to, e.g., dust obscuration or considerably higher redshifts. Thus, the details of the evolving, bivariate color–magnitude distribution is folded into the numbers of color-selected galaxies. For this reason, we are forced to interpret the numbers of color-selected galaxies with reference to a particular model for the evolution of red galaxies.

This is done as follows: we construct a mock catalog in which galaxies' luminosities are distributed according to the $z \approx 0$ luminosity function for red galaxies, which we have determined from the low-z sample, as analyzed in Appendix A. This luminosity function is nearly identical to that of Blanton et al. (2005c); our results do not change significantly adopting the red galaxy luminosity functions

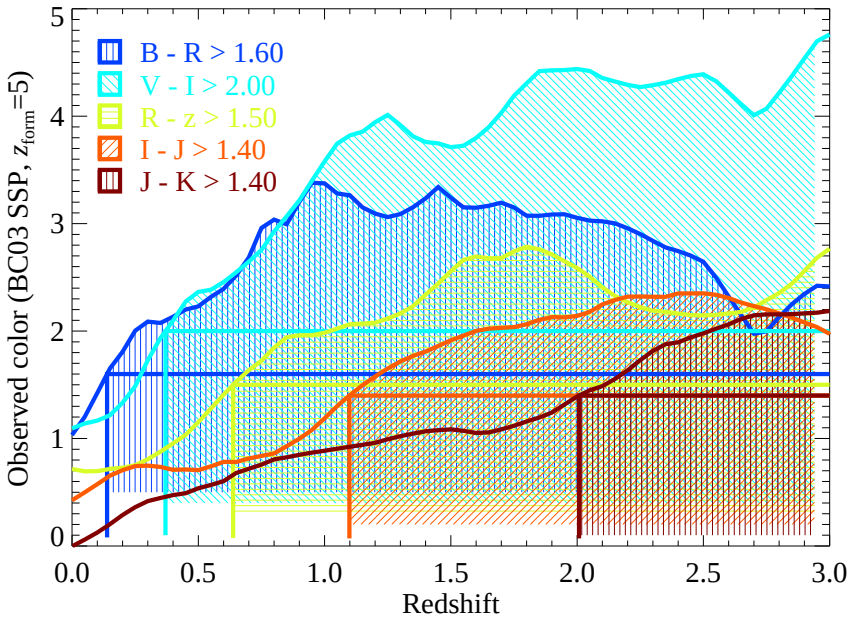


**Figure 13.** — Evolutionary tracks, in terms of observed colors, for a passively evolving stellar population. — Each track is for a synthetic Pégase V2.0 (Le Borgne & Rocca-Volmerange, 2002) stellar population that evolves passively from $z_{\rm form} = 5$; each of these tracks describe an approximate 'red envelope' for the true $z \lesssim 2$ color–redshift distribution. The sharp rise in each track is due to the Balmer and/or 4000 Å breaks becoming redshifted between the two observed filters. The shaded regions show where each of the four observed color criteria given will select red galaxies; taken together, they thus provide a means for selecting passive galaxies in rough redshift bins (see also Section 10.2).

| Color Criterion | Redshift | | Prevalence | | |
|---|---|---|---|---|---|
| | $z_{\rm low}$ | $\langle z \rangle$ | Observed | Expected | Relative |
| (1) | (2) | (3) | (4) | (5) | (6) |
| $B-R>1.6$ | 0.1 | 0.3 | 0.47 / □” | 0.76 / □” | 62 % |
| $V-I>2.0$ | 0.4 | 0.6 | 0.80 / □” | 1.80 / □” | 45 % |
| $R-z>1.5$ | 0.6 | 0.9 | 0.84 / □” | 1.94 / □” | 43 % |
| $I-J>1.4$ | 1.1 | 1.4 | 0.37 / □” | 1.29 / □” | 29 % |
| $J-K>1.4$ | 2.0 | 2.6 | 0.35 / □” | 1.55 / □” | 23 % |

**Table 4.** — Quantifying the evolution of bright, red galaxies without redshifts. — We have used a tiered set of color selection criteria (Col. 1) to select red galaxies in approximate redshift intervals. Here, it is understood that redder criteria are given primacy; for example, objects counted in the second row may or may not satisfy the first criterion, but do not satisfy any of the three subsequent criteria. We give the minimum and mean redshifts of color-selected galaxies from the synthetic, 'passive evolution' catalog in Col.s (2) and (3), respectively. The observed prevalence of color-selected objects in the MUSYC ECDFS catalog are given in Col. (4); Col. (5) gives the same quantity for the synthetic catalog. The ratio of the two, given in Col. (6), which gives an estimate of the relative number of (potentially) passively evolving galaxies at different redshifts.

from either Bell et al. (2003) or Cole et al. (2001). Galaxies are placed randomly (*i.e.* uniform comoving density) in the volume $z < 5$. Next, we generate synthetic photometry for each object in the catalog using the set of Pégase V2.0 (Le Borgne & Rocca-Volmerange, 2002) models shown in Figure 13, which are essentially passively evolving from $z_{\rm form} = 5$. We also include typical errors for the MUSYC ECDFS catalog, to approximately account for photometric scatter. Finally, we apply our color selection criteria to this catalog, exactly as for the observed ECDFS catalog. The predicted prevalence of color selected galaxies, to be compared to those observed, are given in the Col. 5 of Table 4.

Whereas all red galaxies are assumed to evolve passively in the synthetic reference catalog, the real color selected samples will include dusty or high-redshift star-forming galaxies: while all passive galaxies are red, not all red galaxies are passive. As we argue in Section 8.1, the number of color selected galaxies can therefore be used to place an upper limit on the number of passively evolving galaxies, since these color criteria should select a complete but contaminated sample of genuinely passive galaxies.

The results presented in Table 4 thus suggest that the number density of bright, passively evolving galaxies has increased by a factor of at least $\sim 2$ since $z \sim 1$. Moreover, we find that the number density of passive galaxies in the range $1 \lesssim z \lesssim 2$ is at most $\sim 25$ % of the present day value. Taken at face value, this would suggest that at most 1/4 present day red sequence galaxies can have evolved passively from $z \sim 2$; the remainder were still forming, whether through star formation or by the hierarchical assembly of undetected, and thus fainter, subunits, or some combination of the two.

Finally, we emphasize that, while this analysis is model dependent, it is either insensitive or totally immune to the three systematic uncertainties that dominate for our more sophisticated analysis; *viz.*, systematic differences between the analysis of $z \gg 0$ and $z \approx 0$ galaxies; details of the photometric redshift calculation; and photometric calibration errors (Section 9). We also emphasize the essential simplicity of this analysis, and thus its suitability for comparisons between different datasets, which may use different strategies for the computation of photometric redshifts, restframe colors, stellar masses, and so on.

### 10.3 Tying it all Together

Our final task is to compare the results of this section with those in Section 8; we do this in **Figure 14**. In this plot, the ordinate is the effective wavelength of the bluer of the two filters used in each color criterion; the abscissa is the number of color selected galaxies observed in the MUSYC ECDFS catalog, relative to the expectation for passive evolution. The yellow squares refer to the $z_{\rm form} = 5$ model shown in Figure 13 and presented in Table 4. The blue triangles and red pentagons show how these results would vary had we assumed $z_{\rm form} = 4$ and $z_{\rm form} = 6$, respectively.

The specific redshift range that is selected by each color criterion depends on the particular tracks used. For a given color criterion, the mean redshift of

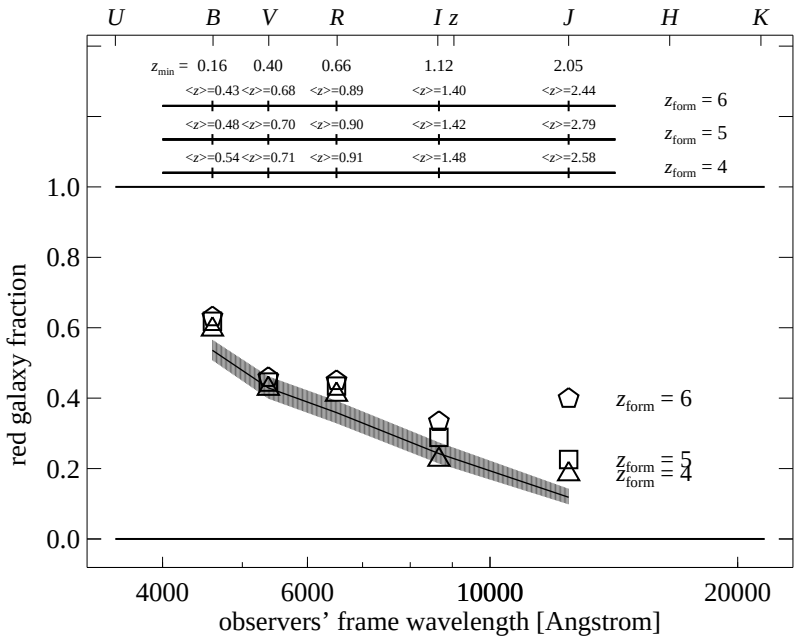


**Figure 14.** — Quantifying the number evolution of bright, red galaxies without redshifts. — The observed numbers of color-selected galaxies, relative to the number expected for passive evolution, plotted as a function of the effective wavelength of the bluer filter used in the selection criterion; the criteria themselves are given in Table 4. The ordinate in this plot is essentially the number of galaxies with colors that are consistent with passive evolution from $z_{\rm form}$, relative to the number of $z \approx 0$ red sequence galaxies; we have plotted results assuming $z_{\rm form} = 4$ (*blue triangles*), 5 (*yellow squares*), and 6 (*red pentagons*). The mean redshift of color selected galaxies, as derived from the synthetic 'passive evolution' catalog, does depend mildly on the choice of evolutionary scenario. The mean redshift of galaxies selected by each criterion, $\langle z \rangle$, is given separately for each choice of $z_{\rm form}$ at the top of the Figure. The minimum redshift is more robust; this is given as $z_{\rm min}$. Lastly, the solid line and shaded region show the fit to the $z_{\rm phot} \lesssim 2$ evolution in the number density of massive red galaxies, derived in Section 8 (see Figure 10); the shaded region shows the ($1\sigma$) statistical uncertainty in the fit. (To put this relation on this plot, we have used the $z_{\rm form} = 5$ catalog to map each color criterion to $\langle z \rangle$.) The agreement between these two very different calculations is impressive.

selected sources (derived from the synthetic 'passive evolution' catalog described above) thus varies for different choices of $z_{\rm form}$: approximate transformations are given for each scenario at the top of the Figure.

Finally, the solid line shows our fit results for the $z_{\rm phot} \lesssim 2$ number density evolution of red galaxies, shown in Figure 10 . In order to put this curve on this plot, we have used the mean redshift of (synthetic, passive) color-selected sources assuming passive evolution from $z_{\rm form} = 5$. Given the very different assumptions and methods lying behind these two results, the agreement is very impressive.

Both analyses clearly indicate that the number of red galaxies at $z \approx 0.7$ is approximately 50 % of the present day value; at $z \approx 1.5$, it is approximately

25 %. In other words, at most, approximately 1/2 local red sequence galaxies were already 'in place' by $z \approx 0.7$, and have evolved passively since that time; at $z \approx 1.5$, at most 1/4 were in place.

## 11 Summary and Conclusions

We have presented the color–magnitude (Figure 4) and color–stellar mass diagrams (Figure 5) for $z \lesssim 2$, based on a sample of $K < 22$ galaxies drawn from the MUSYC catalog of the ECDFS. On the basis of the ten band SEDs, we have achieved a photometric redshift accuracy of $\sigma_z = 0.036$ (Figure 1); this figure represents the current state of the art for broadband photometric lookback surveys. We have used an empirical argument to demonstrate that our main galaxy sample is approximately complete (volume limited) for $M_* > 10^{11}$ and $z_{\rm phot} < 1.8$ (Figure 3). Based on the joint color–stellar mass–redshift distribution of this mass-selected sample, we make the following conclusions:

1. The color distribution of the massive galaxy population is well described by the sum of two separate Gaussian distributions for $z_{\rm phot} \lesssim 1.2$ (Figure 6). Beyond this redshift, the depth of our NIR data makes it impossible to identify distinct red and blue subgroups from within the general massive galaxy population on the basis of the observed color distribution. The question as to the existence or otherwise of a $z_{\rm phot} \gtrsim 1.5$ red sequence will require data approximately an order of magnitude deeper than our own (Figure 7).

2. The colors of red sequence galaxies have become progressively redder by $\sim 0.5$ mag since $z_{\rm phot} \approx 1.1$ (Figure 8). Making a linear fit to the observed evolution, we find $\Delta(u-r) \propto (-0.44 \pm 0.02)\ z_{\rm phot}$. Simple models can only describe the observations assuming sub-solar metallicities (Figure 9): it remains a challenge to consistently describe the observed colors and the mass–metallicity relation.

3. While the number density of massive galaxies evolves mildly for $z_{\rm phot} \lesssim 2$, we see strong evolution in the red galaxy fraction. That is, the massive galaxy population appears to be *changing* more than it is *growing*. We have quantified this evolution using parametric fits of the form $(1+z)^{\gamma}$. For $z_{\rm phot} < 1.8$, we find $\gamma_{\rm tot} = -0.52 \pm 0.12 (\pm 0.20)$, $\gamma_{\rm red} = -1.60 \pm 0.14 (\pm 0.21)$, and $\gamma_{\rm frac} = -1.06 \pm 0.16 (\pm 0.21)$.

The systematic errors (given in brackets above) have been derived on the basis of a whole raft of sensitivity analyses, and are due primarily to photometric calibration errors and systematic effects arising from the photometric redshift calculation (Section 9).

Finally, in Section 10, we showed that these results are completely consistent with an independent analysis based only on directly observed quantities; that is, without deriving redshifts, etc., for individual galaxies. Note that this complementary analysis is almost completely insensitive to the systematic uncertainties that affect our more sophisticated analysis.

The two major advances in this work are the quantification of the $z_{\rm phot} \lesssim 1.2$ color evolution of the red galaxy sequence as a whole, and the quantification of the $z_{\rm phot} \lesssim 2$ evolution of the red fraction among $M_* > 10^{11}$ $M_\odot$ galaxies.

Knowing that the vast majority of $10^{11}$ $M_\odot$ galaxies in the local universe fall on the red sequence (Figure 6), we can identify our massive galaxy sample as the immediate progenitors of local red sequence galaxies. Further, by extrapolating the observed $z_{\rm phot} < 1.4$ color evolution of the red sequence to higher redshifts, we can identify our 'red' galaxies as those which *potentially* have already found their place on the red sequence: the high redshift analogs of local red sequence galaxies. However, simply selecting 'red' galaxies may include a significant number of galaxies whose red colors are due to heavy dust obscuration, we argue that our red galaxy measurements provide an approximate upper limit on the number of passive or quiescent galaxies.

With this assumption, our results suggest that at least 1/2 of all $M_* \gtrsim 10^{11}$ $M_\odot$ galaxies joined (or rejoined) the red sequence only after $z \sim 1$, and that at most 1/5 massive galaxies have resided on the red sequence since $z \sim 2$. These results provide new constraints for quenching models, such as quasar and "radio mode" AGN feedback (see, e.g., Croton et al., 2006; Cattaneo et al, 2006). Establishing which of our 'red' galaxies are genuinely 'red and dead' offers the opportunity to considerably tighten these constraints.

**Acknowledgments.** This work was supported through grants by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), the Leids Kerkhoven-Bosscha Fonds (LKBF), and National Science Foundation (NSF) CAREER grant AST 04-49678. We also wish to thank the organizers and participants of the several workshops hosted by the Lorentz Center, where many aspects of this work were developed and refined. SW gratefully acknowledges support from the W M Keck Foundation.

# Appendices

## A The $z = 0$ Comparison Point

In this Appendix, we describe the process by which we have derived the $z \approx 0$ number density of massive galaxies in general, and of red galaxies in particular, using the low-z subsample of the New York University (NYU) Value Added Galaxy Catalog (VAGC) from the Sloan Digital Sky Survey (SDSS), Data Release 4 (DR4), which has been compiled and described by Blanton et al. (2005b). We will refer to this sample simply as the 'low-z' sample, as opposed to the 'high-z', $K$-selected sample in the main text. The overarching concern is uniformity in the analysis of the low-z and high-z samples, where possible and appropriate, in order to make as fair as possible a comparison between the high-z and low-z data. Our discussion proceeds in three parts: first, we describe our analysis of low-z galaxies; we then describe our characterization of the galaxy red sequence; finally, we give our derived number density of massive galaxies at $z \approx 0$, including mimicking the effects of the random photometric redshift errors typical for the $z \gg 0$ sample.

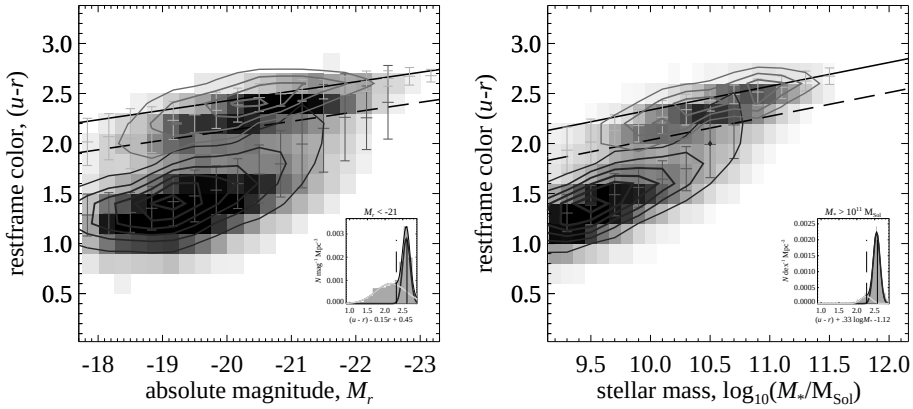


**Figure A.1.** — The 'low-z' sample from the NYU VAGC of the SDSS (DR4), used to derive the $z \approx 0$ comparison point for the main, 'high-z' sample discussed in the text. — Color–magnitude (*left*) and color–stellar mass (*right*) diagrams — The greyscale shows the density of points. The errors bars show the center and width of the color distributions of the red and blue populations, obtained from double Gaussian fits in narrow magnitude/mass slices. These fits to the populations are also represented by the logarithmic (0.5 dex) contours, to give some idea of the data density of red/blue galaxies.

## A.1 Low-z Galaxy Analysis

For each galaxy in the low-z sample, we have constructed *ugriz* SEDs using `model` magnitudes (measured by fitting either an exponential or $r^{1/4}$ profile in each band, with structural parameters derived from the $i$ band image) as given in the low-z catalog. In analogy to high-z galaxies, these SEDs are then scaled up to the 'Petrosian' or `petro` $r$ magnitude. The `petro` aperture is designed to measure a fixed fraction of a galaxy's light, irrespective of brightness or distance, but with some dependence on the shape of the surface brightness profile. Neglecting the effects of seeing, the `petro` aperture captures 99 % of the total light for an exponential profile, compared to $\gtrsim 80$ % for a de Vaucouleurs profile (Blanton et al., 2005b; Strauss et al., 2002). We make no attempt to correct for missed flux, but note that similar effects are present at a similar level in our own data (see Section 9.2.2). Following (Blanton et al., 2005b), we use the following factors to correct the basic SDSS calibration to AB magnitudes: (-0.042, +0.036, +0.015, +0.013, -0.002) for ($u$, $g$, $r$, $i$, $z$).

We have derived restframe photometry and stellar masses for low-z galaxies using exactly the same machinery as for the high-z sample, adopting the heliocentric redshifts given in the low-z catalog. Our interpolated restframe *ugriz* photometry agrees quite well with that given by Blanton et al. (2005b) in the low-z catalog. (Whereas we use galaxy color–color relations, based on six empirical galaxy template spectra, to interpolate restframe photometry, Blanton et al. (2005b) determine restframe fluxes by fitting a non-negative linear combination of five carefully chosen synthetic spectra, and then integrating the best fit spec-

trum.) On average, in comparison to Blanton et al. (2005b), our $(u-r)$ colors are $\sim 0.02$ mag bluer for blue galaxies and $\sim 0.03$ mag redder for red galaxies. We calculate the distance modulus using the proper motion corrected redshifts given in the catalog. Finally, we have derived stellar masses based on the interpolated restframe $B$ and $V$ fluxes, using Equation 1.

### A.2 The Red Galaxy Sequence at $z = 0$

**Figure A.1** shows our CMD (left panel) and CM$_*$D (right panel) diagrams for the general $z \approx 0$ field galaxy population; in each panel, the logarithmic greyscales show the data density. We note that these two plots are very nearly equivalent: the CM$_*$D can be seen as simply a 'sheared' version of the CMD, with the bluer galaxies dragged further towards the left.

A separate red galaxy sequence is easily discernible in each panel. We have characterized this red galaxy sequence, as a function of magnitude or of mass, by taking a narrow slices in either absolute magnitude or stellar mass, and fitting double Gaussians to the color distribution in each bin. The results of these fits are shown in each panel of Figure A.1 as the red and blue contours. This gives some indication of the relative numbers of red and blue galaxies as a function of magnitude/mass. In order to characterize the separate red and blue populations, we also show the center and width of each Gaussian fit as the yellow and cyan points and error bars.

In order to determine the slope of the high luminosity/mass end of the red sequence CMR and CM$_*$R, we have simply made linear fits to the centers of the $M_r < -21$ and $M_* > 10^{10.5}$ M$_\odot$ 'red' distributions. The results of these fits are:

$$(u-r) = 2.612 - 0.090 \times (M_r + 22) = 2.573 + 0.237 \times \log_{10}(M_*/10^{11}\ \mathrm{M}_\odot)\ . \quad \text{(A.1)}$$

These relations are shown as the solid black lines in each panel. Finally, in order to make some distinction between 'red' and 'blue' galaxies which could be easily applied to the high-$z$ sample, we simply use a cut 0.25 mag bluer than this relation. In each panel, this cut can be seen to approximately coincide with the so–called 'green valley' between the red and blue galaxy populations.

Finally, in the inset of each diagram, we show the color distribution for the brightest ($M_r < -21$) or most massive ($M_* > 10^{11}$) galaxies, after subtracting away the color–magnitude or color–stellar mass relations given by Equation A.1. (See also Figure 6.) We also show double Gaussian fits to these distributions, which are the basis for the $z \approx 0$ points shown in Figures 8 and 9. We note in passing that, at least for $-19 \gtrsim M_r \gtrsim -21.5$, the width of the color distributions of red and blue galaxies is not a strong function of magnitude: $\Delta(u-r) \sim 0.12$, and $\Delta(u-r) \approx 0.20$ for red and blue galaxies, respectively.

### A.3 The Number Density of Massive Galaxies at $z \approx 0$

For our purposes (see Section 8), the crucial final quantity is the number density of massive galaxies at $z \approx 0$. We find that the number density of $M_* > 10^{11}$ M$_\odot$ galaxies is $(6.64 \pm 0.38) \times 10^{-4}\ h^3$ Mpc$^{-3}$; for $(u,\ r)$ red galaxies, the number is

$(6.07 \pm 0.36) \times 10^{-4}$ $h^3$ Mpc$^{-3}$. This measurement is not significantly affected by either volume or surface brightness incompleteness. Compared to the more sophisticated analyses of Bell et al. (2003) and Cole et al. (2001), our value for the number density of massive galaxies is 10 % higher and 25 % lower, respectively.

As we have repeatedly stressed, our prime concern is uniformity in the analysis of the low- and high-z samples. Photometric redshift errors are a major source of uncertainty for the high-z sample — how would comparable errors affect the low-z measurements? Given the redshift range of our low-z galaxies ($0.003 < z < 0.05$) and the estimated photometric redshift errors among high-z galaxies ($\Delta z/(1+z) \sim 0.035$), it would be inappropriate to simply apply typical redshift errors and repeat our calculations: this would effectively throw away all distance information. Instead, what we have done is to apply the typical uncertainties on $M_r$, $(u-r)$ and $M_*$ due to photometric redshift errors, as shown in Figure 2. (For simplicity, we have ignored correlations between these errors.)

In line with our earlier findings, including the effects of photometric redshift errors makes less of a difference to measurements based on stellar masses than to those based on absolute magnitudes. Our fits to the red sequence become:

$$(u-r) = 2.608 - 0.113 \times (M_r + 22) = 2.577 + 0.223 \times \log_{10}(M_*/10^{11}\ \mathrm{M}_\odot)\ . \quad \text{(A.2)}$$

Our measurements of the $z \approx 0$ number density of galaxies more massive than $10^{11}$ M$_\odot$ becomes $(6.93 \pm 0.38) \times 10^{-4}$ $h^3$ Mpc$^{-3}$ in total, and $(5.86 \pm 0.35) \times 10^{-4}$ $h^3$ Mpc$^{-3}$ for red sequence galaxies. That is, the effects of the redshift errors expected among $z \gg 0$ galaxies affect the measurement of the $z \approx 0$ number density by less than 5 %. These numbers, approximately accounting for the Eddington bias due to photometric redshift errors (as they apply to the $z \gg 0$ sample), are then what have used as a local reference value to compare to the $z \gg 0$ data.

## B A Detailed Comparison with COMBO-17

In the ECDFS, the COMBO-17 and MUSYC datasets represent parallel analytical paths describing the same physical reality — indeed the COMBO-17 broadband observations form the foundation of the MUSYC raw optical data. Ideally, then, the two surveys' results should agree identically; instead, in comparison to COMBO-17, the MUSYC results suggest 40 % more massive ($M_* > 10^{11}$ M$_\odot$) galaxies at $0.2 < z_{\mathrm{phot}} < 0.8$ — a degree of difference more on a par with that expected from field–to–field variation. In this Appendix, we identify the main cause of this discrepancy.

### B.1 Photometric Redshift Accuracy: MUSYC photometry with COMBO-17 redshifts

The essential differences between the MUSYC and COMBO-17 datasets are: 1.) COMBO-17's twelve medium bands, which allow much more precise photometric redshift determinations for $z \lesssim 1$, and 2.) the MUSYC $z'JHK$ data, which open the door to the $z \gtrsim 1$ universe. The first and most obvious concern is thus that the difference between the two surveys' $z_{\mathrm{phot}} \lesssim 1$ results are a product of

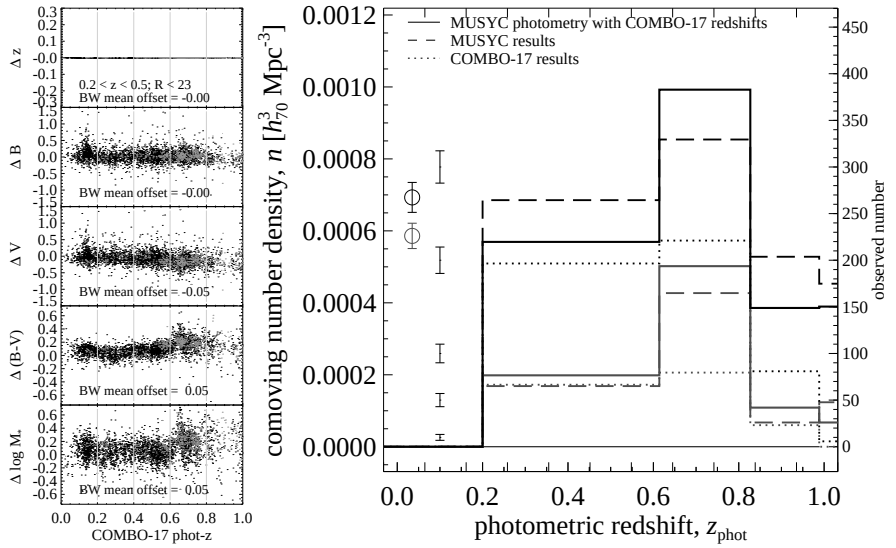


**Figure B.1.** — Reanalysis of the MUSYC data adopting redshifts from COMBO-17. — *Main panel*: The measured number density of $M_* > 10^{11}$ $M_\odot$ galaxies for $z \lesssim 1$ based on the MUSYC photometry, but adopting redshifts from the COMBO-17 catalog in place of our own (*solid histograms*). These results should be compared to the fiducial results from MUSYC (*dashed histograms*) and from COMBO-17 (*dotted histograms*). The red and black histograms are for red and all galaxies, respectively. Where the COMBO-17 and MUSYC results differ by 40 %, the effect of using of COMBO-17 redshifts is less than 10 %; the difference between the two surveys' results cannot be explained by differences in the photometric redshifts. *Left panels*: the difference between the (*from top to bottom*) photometric redshifts, absolute $B$ and $V$ magnitudes, restframe $(B-V)$ colors, and stellar masses used to produce the results shown in the main panel, in comparison to those found in the COMBO-17 catalog, each plotted as a function of the COMBO-17 redshift. In comparison to COMBO-17, even using the same redshifts, the MUSYC data and/or data analysis make galaxies appear brighter and redder, and so more massive.

their different photometric redshift accuracies. To test this, we have tried simply adopting the COMBO-17 photometric redshifts in place of our own determinations, and repeated our analysis.

The results of this test are shown in **Figure B.1**: the main panel shows our primary result (*viz.*, the number density of $M_* > 10^{11}$ $M_\odot$ galaxies as a function of redshift; see Figure 10). The solid histograms show this trial re-analysis of the MUSYC data adopting COMBO-17 redshifts; this should be compared to the fiducial results from MUSYC (*dashed histogram*) and COMBO-17 (*dotted histograms*). We have derived these 'COMBO-17' results from the public catalog presented by Wolf et al. (2004), supplemented with the stellar mass determinations used by Borch et al. (2006).

At least for $z_{\rm phot} < 0.8$, the use of COMBO-17 redshifts in place of our own does not have a great effect on our results: the solid histograms lie quite close to the dashed ones. Quantitatively, adopting COMBO-17 redshifts in place of

our own leaves the number of $M_* > 10^{11}$ $M_\odot$ galaxies over $0.2 < z_{\rm phot} < 0.8$ unchanged to within 1 %; the effect is slightly greater for the red population, leading to a red fraction that is 17 % higher over the same interval. This implies that, at least for $z_{\rm phot} \lesssim 1$, our results are not seriously affected by our lower photometric redshift accuracy ($\sigma_z \approx 0.03$ for MUSYC, versus 0.02 for COMBO-17, for $R < 24$ and $z_{\rm spec} < 1$).

To the left of the main panel in Figure B.1, we show the difference in (*from top to bottom*) the photometric redshifts, absolute B, and V magnitudes, restframe $(B-V)$ colors, and stellar masses used to produce the solid histograms, and those from the COMBO-17 catalog. In all cases, the '$\Delta$' is in the sense of 'MUSYC reanalysis' minus 'COMBO-17 catalog', and is plotted as a function of the COMBO-17 catalog redshift. Within each panel, we also give the biweight mean offset for each quantity, evaluated for sources with $R < 23$ (to reduce random scatter due to photometric uncertainties) and $0.2 < z_{\rm phot} < 0.5$ (where the COMBO-17 photometry still samples restframe $V$) in the COMBO-17 catalog.

Looking now briefly at these panels, we see that for $z_{\rm phot} \gtrsim 0.5$, where the COMBO-17 value for the restframe $V$ magnitude comes from an extrapolation of the best-fit template spectrum, there is a progressive offset between the restframe $V$ luminosities inferred by the MUSYC and COMBO-17 data and analyses — even while using the same redshifts. Even for $z_{\rm phot} < 0.5$, however, we see a systematic offset of 0.05 mag in $(B-V)$ color. Finally, we note that the greater number of $0.6 < z_{\rm phot} < 0.8$ galaxies in comparison to $0.2 < z_{\rm phot} < 0.6$ that we see in the MUSYC data is also present in the COMBO-17 photometric redshift distribution; the difference is that these galaxies do not have $M_* > 10^{11}$ $M_\odot$ in the COMBO-17 catalog.

We therefore conclude that the difference between the COMBO-17 and MUSYC results is not a product of the two surveys' different photometric redshift accuracies: the use of COMBO-17 rather than MUSYC redshifts affects our results only by a few percent. This leaves the possibilities that the difference between the two results is due to the different methods used to infer galaxies' restframe properties, or to differences in the data themselves.

## B.2 Derivation of Restframe Properties: MUSYC analysis of the COMBO-17 photometry

Are the discrepancies between the COMBO-17 and MUSYC results due to different systematic effects inherent in our different methods for deriving restframe photometry and stellar masses? In order to address this concern, we have reanalyzed the COMBO-17 data using the procedures described in the main text. For this test, in order to isolate the effect of the different analytical methods, we also adopt the COMBO-17 redshift determinations; the difference between this test and the previous one is thus the use of the COMBO-17, rather than the MUSYC, photometry.

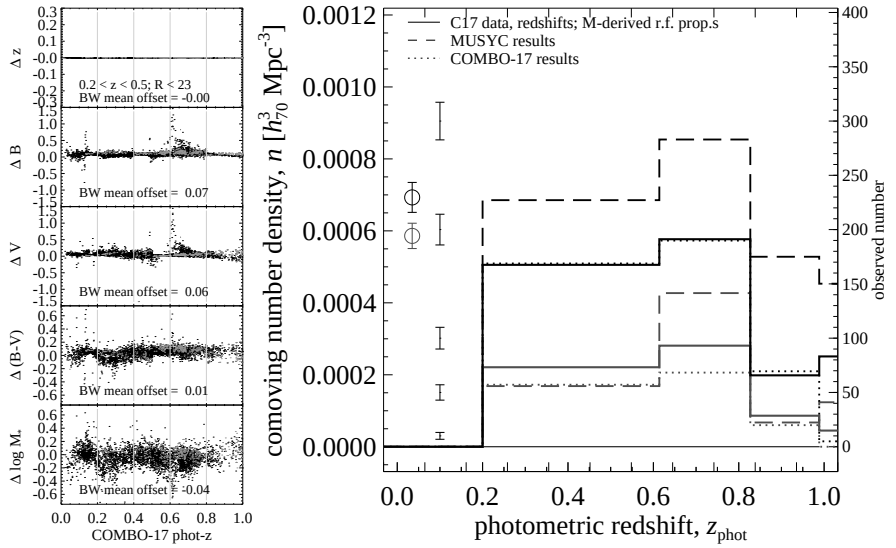


**Figure B.2.** — MUSYC analysis of the COMBO-17 photometry. — The solid histograms in the main panel show the measured number density of $M_* > 10^{11}$ galaxies, applying the MUSYC analysis to the COMBO-17 photometry. In order to isolate the effect on our measurements due to the different methods for deriving restframe properties, we continue to adopt the COMBO-17 redshifts for this test. All other symbols and their meanings are as in Figure B.1. Applying the MUSYC analysis to the COMBO-17 photometry, we reproduce the COMBO-17 result to better than 1 %; the difference between the COMBO-17 and MUSYC results cannot be explained solely by our different analytical methods.

The results are shown as the solid histograms in **Figure B.2**; all other symbols and their meanings are as in Figure B.1. The results of the MUSYC re-analysis of the COMBO-17 data agree very well with the COMBO-17 analysis proper: the solid black histogram lies very near the dotted black histogram. Quantitatively, the MUSYC analysis of the COMBO-17 data leads to a red fraction which is 33 % higher than for COMBO-17's own analysis; the total number density of $M_* > 10^{11}$ $M_\odot$ galaxies agree to better than 1 %.

Comparing the MUSYC– and COMBO-17–derived restframe properties — again, based on the same photometry and redshifts — we do see systematic differences. We make three specific observations: 1.) there appear to be discrete redshift regimes where the MUSYC– and COMBO-17–derived restframe fluxes compare differently; 2.) for $0.2 \lesssim z_{\rm phot} \lesssim 0.35$, we see a 'bimodal' offset in the MUSYC– and COMBO-17–derived $(B-V)$ colors, corresponding to red and blue galaxies; 3.) the MUSYC–derived restframe fluxes appear to be systematically brighter than those derived by COMBO-17, even for the same photometry. Remarkably, even despite these differences, our stellar mass estimates agree to within 0.04 dex, with a scatter of just 0.11 dex. Finally, we note that the progressive offset in extrapolated $V$ luminosities for $z_{\rm phot} \gtrsim 0.5$ seen in the previous test is not seen here; that is, applied to the same data, the two techniques for extrapolating

restframe photometry yield similar results.

We therefore conclude that the difference between the COMBO-17 and MUSYC results in the ECDFS cannot be explained by differences in the analytical methods employed by each team: the MUSYC reanalysis of the COMBO-17 data agrees with the COMBO-17 fiducial results. Instead, it seems that the different results are really in the data themselves.

### B.3 Photometric Calibration: MUSYC analysis of the recalibrated COMBO-17 photometry

A direct, object–by–object comparison of COMBO-17 and MUSYC photometry reveals significant differences in the photometric calibration of the two surveys (see Section 5.2.2 of Chapter II). Specifically, we see differences of 0.00, 0.06, 0.08, 0.08, and 0.14 mag between the MUSYC and COMBO-17 $UBVRI$ band photometry for galaxies, such that the MUSYC photometry is systematically brighter and redder. This has subsequently been confirmed to be due to an error in the photometric calibration of the COMBO-17 data (Wolf et al., 2008). Can this difference in photometric calibration explain the different results found by COMBO-17 and MUSYC?

To address this issue, we have simply scaled the COMBO-17 photometry to match the MUSYC photometry, and repeated our analysis. The COMBO-17 team has calibrated each of the medium bands relative to the nearest broad band; accordingly we have scaled each of the medium bands by the MUSYC–minus–COMBO-17 offset for the nearest broad band. The results of this test are shown in **Figure B.3**. The MUSYC re-analysis of the recalibrated COMBO-17 photometry agrees extremely well with the fiducial MUSYC results: the measured number density of $M_* > 10^{11}$ $M_\odot$ galaxies at $0.2 < z_{\rm phot} < 0.8$ agree to better than 1 %.

There are two separate aspects to this recalibration: galaxies are both brighter and redder in the MUSYC catalog than in COMBO-17. Since COMBO-17 measures total fluxes from their $R$ band image, the 0.08 mag offset between the COMBO-17 and MUSYC zeropoints makes galaxies appear 7.6 % brighter (and thus more massive) in the MUSYC catalog; this effect is responsible for approximately 70 % of the change in the measured number density. At the same time, the reddening of the SED shape implies a higher mass–to–light ratio after recalibration; this effect is responsible for the other 30 % of the change in the measurements.

We therefore conclude that the primary cause for the disagreement between the results of the MUSYC and COMBO-17 surveys in the ECDFS is the different photometric calibrations of the two surveys: reanalyzing the COMBO-17 data set, recalibrated to match the MUSYC photometry, the results agree with the MUSYC fiducial analysis.

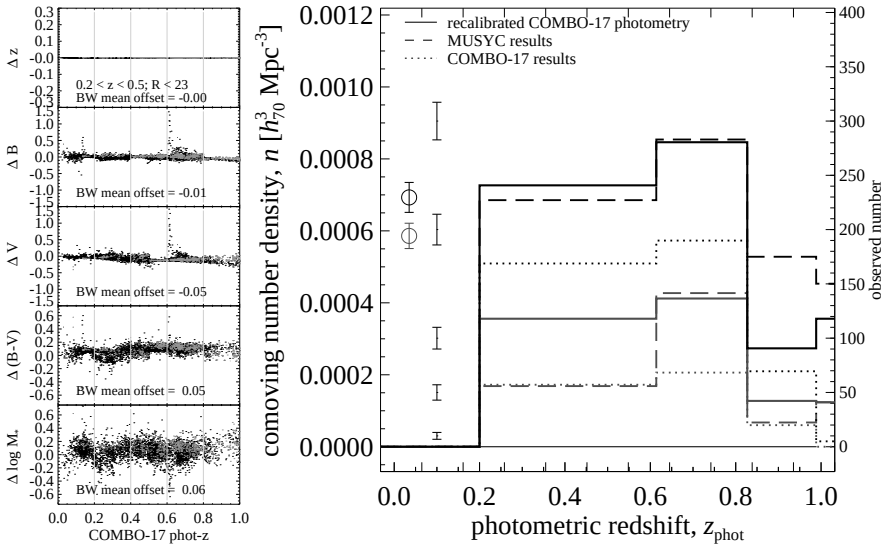


**Figure B.3.** — MUSYC analysis of the COMBO-17 photometry, recalibrated to match MUSYC. — the solid histograms in the main panel show the measured number density of $M_* > 10^{11}$ galaxies, applying the MUSYC analysis to the COMBO-17 photometry, recalibrated to match the MUSYC photometry. Specifically, the $UBVRI$ bands have been scaled up by 0.00, 0.06, 0.08, 0.08, and 0.14 mag; each of the medium bands has been scaled to match the nearest broad band. We continue to adopt the COMBO-17 redshifts for this test. All other symbols and their meanings are as in Figure B.1. The effect of this rescaling is to increase the measured values by a further 30 %, in comparison to the previous test. Together, the combined effect of this recalibration is nearly 50 %, almost completely explaining the difference between the COMBO-17 and MUSYC results.

# References


Arnouts S et al., 2001, A&A 379, 740

Arnouts S et al., 2007, A&A 476, 137

Baldry I K, Glazebrook K, Brinkmann J, Ivezić Ž, Lupton R H, Nicol R C, Szalay A S, 2004, ApJ 600, 694

Baldry I K, Balough M L, Bower R G, Glazebrook K, Nichol R C, Bamford S P, Budavari T, 2006, MNRAS 373, 469

Balough M L, Baldry I K, Nichol R, Miller C, Bower R, Glazebrook K, 2003, ApJ 615, L101

Bell E F, de Jong R S, 2001, ApJ 550, 212

Bell E F, McIntosh D H, Katz N, Weinberg M D, 2003, ApJS 149, 289

Bell E F et al., 2004, ApJ 600, L11

Bell E F et al., 2004, ApJ 608, 752

Bell E F, Zheng X Z, Papovich C, Borch A, Wolf C, Meisenheimer K, 2007, ApJ 663, 834

Benìtez N, 2000, ApJ 536, 571

Bertin E & Arnouts S, A&A 117, 393

Blanc G A et al., 2008, ApJ, submitted

Blanton M R et al., 2003, ApJ 594, 186

Blanton M R, Eisenstein D, Hogg D W, Schlegel B K, Brinchmann J, 2005, ApJ 629, 143

Blanton M R, Schlegel D J, Strauss M A, Brinkmann J, Finkbeiner D, Fukugita M, Gunn J E, Hogg D W, Ivezić Ž, Knapp G R, Lupton R H, Munn J A, Schneider D P, Tegmark M, Zehavi I, 2005, AJ 129, 2562

Blanton M R, Lupton R H, Schlegel D J, Strauss M A, Brinkmann, Fukugita M, Loveday J, 2005, ApJ 631, 208

Blanton M R & Roweis S, 2007, ApJ 133, 734

Bolzonella M, Miralles J-M & Pelló R, 2000, A&A 363, 476

Borch A, Meisenheimer K, Bell E F, Rix H-W, Wolf C, Dye S, Kleinheinrich M, Kovacs Z, Wisotzki L, 2006, A&A 453, 869

Brammer G B, van Dokkum P G, Coppi P, 2008, submitted

Brinchmann J, Charlot S, White S D M, Tremonti C, Kauffmann G, Heckman T, 2004, MNRAS 351, 1151

Brown M J I, Dey A, Jannuzi B T, Brand K, Benson A J, Brodwin M, Croton D J, Eisenhardt P R, 2007, ApJ 164, 858

Brown M J I, Zheng Z, White M, Dey A,

Jannuzi B T, Benson A J, Brand K, Brodwin M, Croton D J

Bruzual G, Charlot S, 2003, MNRAS 344, 1000

Bruzual G A, 2008, astro-ph/0703052 (v2)

Bundy K et al., 2006, ApJ 651, 120

Cattaneo A, Dekel A, Devriendt J, Guiderdoni B, Blaizot J, 2006, 370, 1651

Cattaneo A, Dekel A, Faber S M, Guiderdoni B, 2008, MNRAS 389, 567

Cimatti A, Mignoli M, Daddi E, Pozzetti L, Fontana A, Saracco P, Poli F, Renzini A, Zamorani G, Broadhurst T, Cristiani S, D'Odorico S, Giallongo E, Gilmozzi R, Menci N, 2002, A& A 392, 395

Cirasuolo M, McLure R J, Dunlop J S, Almaini O, Foucaud S, Smail I, Sekiguchi K, Simpson C, Eales S, Dye S, Watson M G, Page M J, Hist P, 2007, MNRAS 380, 585

Cole S et al., 2001, MNRAS 326, 255

Coleman G D, Wu C C, Weedman D W, 1980, ApJS 43, 393

Connolly A J, Szalay A S, Dickinson M,Subbarao M U, Brunner R J, 1997, ApJ 486, L11

Conselice C J, Bundy K, Trujillo I, Coil A, Eisenhardy P, Ellis R S, Georgakakis A, Huang J, Lotz J, Nandra K, Newman J, Papovich C, Weiner B, Willmer C, 2007, MNRAS 381, 962

Cooper M J et al., 2007, MNRAS 376, 1445

Croton D J, Springel V, White S D M, De Lucia G, Frenk C S, Gao L, Jenkins A, Kauffmann G, Navarro J F, Yoshida N, 2006, MNRAS 365, 11

Daddi E et al., 2005, ApJ 626, 680

Damen M et al., 2008, ApJ (submitted)

Dekel A, Birnboim Y, 2006, MNRAS 368, 2

De Lucia G et al., 2007, MNRAS 374, 809

Dickinson M E, Giavalisco M, et al., 2002, astro-ph/0204213 (v1)

Driver S P, Allen P D, Graham A W, Cameron E, Liske J, Ellis S C, Cross N J G, De Propris R, Phillipps S, Couch W J, 2006, MNRAS 368, 414

Drory N et al., 2004, ApJ 608, 742

Ellis S C, Driver S P, Allen S D, Liske J, Bland-Hawthorn J, De Propris R, 2005, MNRAS 363, 1257

Erben T et al., 2005, AN 326, 432

Faber S M et al., 2007, ApJ 665, 265

Fontana A et al., 2004, A&A 424, 23

Fontana A et al., 2006, A&A 459, 745

Förster-Schreiber N M, Franx M, Labbé I, Rudnick G, van Dokkum P G, Illingworth G D, Kuijken K, Moorwood A F M, Rix, HW., Röttgering H, van der Werf P, 2006, AJ 131, 1891

Franx M et al., 2003, ApJ 587, L79

Franzetti P et al., 2007, A&A, 165, 711

Gawiser E et al., 2006, ApJS 162, 1

Giacconi R, Rosati P, Tozzi P, Nonino M, Hasinger G, Norman C, Bergeron J, Borgani S, Gilli R, Gilmozzi R, Zheng W, 2001, ApJ 551, 624

Grazian A, Fontana A, De Santis C, Nonino M, Salimbeni S, Giallongo E, Cristiani S, Gallozzi S, Vanzella E, 2006, A& A 449, 951

Gronwall C et al., 2007, ApJ, 667. 79

Gunn J E & Stryker L L, 1983, ApJSS 52, 121

Hildebrandt H et al., 2006, A&A 452, 1121

Hogg D W, Blanton M R, Eisenstein D J, Gunn J E, Schlegel D J, Zehavi I, Bahcall N A, Brinkmann J, Csabai I, Schneider D P, Weinberg D H, York D G, 2003, ApJ 585, L5

Hopkins A M, 2004, ApJ 615, 209

Im M et al., 2002, ApJ 571, 136

Juneau S et al., 2005, ApJ 619, L135

Kauffman G, Heckman T M, White S D M, Charlot S, Tremonti C, Peng E W, Seibert M, Brinkmann J, Nichol R C, SubbaRao M, York D, 2003, MNRAS 341, 54

Kinney A L, Calzetti D, Bohlin R C, McQuade K, Storchi-Bergmann T, Schmitt H R, 1996, ApJ 467, 38

Kriek M, et al., 2006, ApJ, 649, L71

Kriek M, van der Wel A, van Dokkum P G, Franx M, Illingworth G D, 2008, ApJ 682, 896

Kroupa P, 2001, MNRAS 322, 231

Labbé I et al., 2003, AJ 125, 1107

Labbé I et al., 2005, ApJ 624, L81

Labbé I, Franx M, Rudnick G, Forster Schrieber N M, van Dokkum P G, Moorwood A, Rix H-W, R ottgering H, Trujillo I, van der Werf P, 2007, ApJ 665, 964

Lawrence A et al., 2007, MNRAS 379, 1599

Le Borgne D & Rocca-Volmerange B, 2002, A&A 386, 466

Le Fèvre O et al., 2004, A& A 428, 1043

Lilly S, Le Fèvre O, Hammer F, Crampton D, 1996, ApJ 460, L1

Lin L et al., 2007, ApJ 660, L51

Madau P, Ferguson H C, Dickinson M E, Giavalisco M, Steidel C C, Fruchter A, 1996, MNRAS 283, 1388

Maraston C, 2005, MNRAS 362, 799

McCarthy P J, 2004, ARAA, 42, 477

McGrath E J, Stockton A, Canalizo G, 2007, ApJ 669, 241

Menci N, Fontana A, Giallongo E, Salimbeni S, 2005, ApJ 632, 49

Moy E et al., 2003, A&A 403, 493

Nagamine K, Ostriker J P, Fukugita M, Cen R, 2006, ApJ 653, 881

Panter B, Jimenez R, Heavens A F, Charlot S, 2007, MNRAS 378, 1550

Pérez-Gonzaléz P G et al., 2008, ApJ 675, 234

Popesso P et al., 2009, A&A 494, 442

Pozzetti L et al., 2007, A&A, 474, 447

Quadri R et al., 2007, AJ 134, 3

Ravikumar C D et al., 2006, A& A 465, 1099

Reddy N A, Steidel C C, Fadda D, Yan L, Pettini M, Shapley A E, Erb D K, Adelberger K L, 2006, ApJ 644, 792

Rix H-W et al., 2004, ApJSS 152, 163

Rudnick G, Rix H-W, Franx M, Labbé I et al., 2003, ApJ 599, 847

Ruhland C, Bell E F, Häußler B, Taylor E N, Barden M, McIntosh D, 2009, ApJ 1058, 1069

Salpeter E E, 1955, ApJ 121, 161

Scarlata C et al., 2008, ApJSS, 172, 510

Somerville R S, Lee K, Ferguson H, Gardner J P, Moustakas L A, Giavalisco M, 2004, ApJ 600, L171

Strateva I et al., 2001, AJ, 122, 1861

Strauss M A, Weinberg S H, Lupton R H, Narayanan V K et al., 2002, AJ, 124, 1810

Szokoly G P et al., 2004, ApJSS 155, 271

Taylor E N et al., 2008 (Chapter II), ApJSS (submitted)

Treister E, Virani S, Gawiser E, Urry C M, Lira P, Francke H, Blanc G A, Cardamone C N, Damen M, Taylor E N, Schawinski K

Tremonti C A et al., 2004, ApJ 613, 898

Tresse L et al., A&A 472, 403

van der Wel A, Franx M, van Dokkum P G, Rix H-W, Illingworth G D, Rosatti P, 2005, ApJ 631, 145

van der Wel A, 2008, ApJ 675, L13

van Dokkum P G & Franx M, 2001, ApJ 553, 90

van Dokkum P G et al., 2006, ApJ 638, 59

Vanzella et al., 2008, A& A 471, 83

Weiner B J et al., 2005, ApJ 620, 595

Willmer C N A et al., 2006, ApJ, 647, 853

Wolf C et al., 2004, A&A 421, 913

Wolf C, Hildebrandt H, Taylor E N, Meisenheimer K, 2008, A&A 492, 933

Wuyts S et al, 2008, ApJ (in press), arXiv:0804.0615v2

Wyder T K et al., 2007, ApJSS 173, 293

York D G et al., 2000, ApJ 120, 1579

# Chapter IV

## On the Dearth of Compact, Massive, Red Sequence Galaxies in the Local Universe

In this Chapter, we test the claim that the recently identified population of compact, massive, and quiescent galaxies at $z \sim 2.3$ must undergo significant size evolution to match the properties of galaxies found in the local universe. Using data from the Sloan Digital Sky Survey (SDSS; Data Release 7), we have conducted a search for local red sequence galaxies with sizes and masses comparable to those found at $z \sim 2.3$. The SDSS spectroscopic target selection algorithm excludes high surface brightness objects; we show that this makes incompleteness a concern for such massive, compact galaxies, particularly for *low* redshifts ($z \lesssim 0.05$). We have identified 63 $M_* > 10^{10.7}$ $M_\odot$ ($\approx 5 \times 10^{10}$ $M_\odot$) red sequence galaxies at $0.066 < z_{\rm spec} < 0.12$ which are smaller than the median size–mass relation by a factor of 2 or more. Consistent with expectations from the virial theorem, the median offset from the mass–velocity dispersion relation for these galaxies is 0.12 dex. We do not, however, find any galaxies with sizes and masses comparable to those observed at $z \sim 2.3$, implying a decrease in the comoving number density of these galaxies, at fixed size and mass, by a factor of $\gtrsim 5000$. This result cannot be explained by incompleteness: in the $0.066 < z < 0.12$ interval, we estimate that the SDSS spectroscopic sample should typically be $\gtrsim 75\%$ complete for galaxies with the sizes and masses seen at high redshift, although for the very smallest galaxies it may be as low as $\sim 20\%$. In order to confirm that the absence of such compact massive galaxies in SDSS is not produced by spectroscopic selection effects, we have also looked for such galaxies in the basic SDSS photometric catalog, using photometric redshifts. While we do find signs of a slight bias against massive, compact galaxies, this analysis suggests that the SDSS spectroscopic sample is missing at most a few objects in the regime we consider. Accepting the high redshift results, it is clear that massive galaxies must undergo significant structural evolution over $z \lesssim 2$ in order to match the population seen in the local universe. Our results suggest that a highly stochastic mechanism (e.g., major mergers) cannot be the primary driver of this strong size evolution.

Taylor E N, Franx M, Glazebrook K,
Brinchmann J, van der Wel A, van Dokkum P G
for publication in *the Astrophysical Journal*
(submitted July 2009; arXiv:0907.4766)

## 1 Introduction

In the simplest possible terms, the naïve expectation from hierarchical structure formation scenarios is that the most massive galaxies form relatively late. This is in contrast to the observation that the bulk of cosmic star formation occurs in galaxies with progressively lower stellar masses at later times (e.g. Juneau et al., 2005; Zheng et al., 2007; Damen et al., 2008); the so–called downsizing of galaxy growth. These observations have been accommodated within the ΛCDM framework with the introduction of a quenching mechanism (e.g. Menci et al., 2005; Croton et al., 2006; Cattaneo et al, 2008), which operates to shut down star formation in the most massive galaxies; this mechanism is also required to correctly predict the absolute and relative numbers of red galaxies at $z \lesssim 1$ (Dekel & Birnboim, 2006; Bell et al., 2007; Faber et al., 2007). With this inclusion, models thus predict that a significant fraction of massive galaxies finish their star formation relatively early in the history of the universe, with later mergers working to build up the most massive galaxies.

There is thus a crucial distinction to be made between a galaxy's mean stellar age, and the time since that galaxy has assumed its present form (see, e.g., De Lucia et al., 2006): the most massive galaxies are expected to be both the oldest *and* the youngest galaxies. They are the oldest in the sense that their progenitors are expected to form first in the highest cosmic overdensities. However, these stars are only assembled into their $z = 0$ configuration relatively recently; in this sense, massive galaxies are expected to be rather younger than their constituent stellar populations.

This leaves (at least) two open questions relating to the quenching of star formation and the formation and evolution of massive galaxies: 1.) when does star formation stop in massive galaxies, and then 2.) What happens to galaxies after they have stopped forming stars?

In connection with the first of these questions, deep spectroscopic surveys have identified massive galaxies with little or no ongoing star formation at $1 \lesssim z \lesssim 2$ (e.g. Cimatti, 2004; Glazebrook et al., 2004; McCarthy, 2004a; Daddi et al., 2004). At the same time, color selection techniques like the ERO (McCarthy, 2004b, and references therein), DRG (Franx et al., 2003), or BzK (Daddi et al., 2005) criteria have been used to identify massive, passive galaxies at moderate– to high–redshifts. While these techniques are deliberately biased towards certain kinds of galaxies and certain redshift intervals, advances in techniques for photometric redshift estimation and stellar population modeling have allowed the selection of mass-limited samples, and so the construction of representative samples of the high redshift massive galaxy population (e.g. van Dokkum et al., 2006).

By obtaining very deep restframe optical spectra of a photometric redshift selected sample of massive galaxies at $z \gtrsim 2$, Kriek et al. (2008a) made a significant advance on previous spectroscopic and photometric studies. Of the 36 $z_{\rm spec} > 2$, $M_* > 10^{11}$ $M_\odot$ galaxies in the Kriek et al. (2008a) sample, 16 were shown unambiguously to have evolved stellar populations and little or no ongoing star formation. These galaxies also seem to form a red sequence in $(B - V)$ color,

although at low significance ($3.3\sigma$; Kriek et al., 2008b). In other words, these massive galaxies appear both to have assembled stellar populations similar to galaxies of comparable mass in the local universe, and to have had their star formation effectively quenched.

Using Keck laser guide-star assisted adaptive optics and *Hubble Space Telescope* imaging, van Dokkum et al. (2008, hereafter vD09) measured sizes for 9 of the 16 strongly quenched galaxies from the Kriek et al. (2008a) sample. They found (restframe optical) effective radii in the range 0.5–2.4 kpc; that is, smaller than typical galaxies of the same mass in the local universe by factors of 3–10. These galaxies have stellar mass densities, measured within the central 1 kpc, which are 2–3 times higher than typical local galaxies of the same mass (Bezanson et al., 2009). Similar sizes and densities have been found for a larger sample of 82 massive galaxies at $1.7 < z < 3.0$ from the GOODS survey by Buitrago et al. (2008). Cimatti et al. (2008) and Damjanov et al. (2009, hereafter D09) have found similarly compact sizes for massive galaxy samples drawn from $1 < z < 2$ spectroscopic surveys. Further, van Dokkum, Kriek & Franx (2009) have recently measured a velocity dispersion of $510^{+165}_{-95}$ km/s for one of the galaxies in the vD08 sample, based on a 29 hr NIR spectrum; this extremely high value is consistent with the galaxy's measured mass and size. (See also Cappellari et al. 2009, who have measured velocity dispersions for two $z \sim 1.4$ galaxies and a stacked spectrum of 7 massive galaxies at $1.6 < z < 2.0$, and Cennaro & Trujillo 2009, who measured a velocity dispersion for a stacked spectrum of 13 massive galaxies at $1.4 < z < 2.0$.) These results confirm and consolidate the work of Daddi et al. (2005), Trujillo et al. (2006), Trujillo et al. (2007), Zirm et al. (2007), and Toft et al. (2007), as well as $1 < z < 2$ results from, e.g., Longhetti et al. (2007) and Saracco et al. (2009), and $z \lesssim 1$ results from van der Wel et al. (2008).

The significance of these results is that, in terms of their stellar populations, these $z \gtrsim 2$ galaxies appear to be more or less 'fully formed'. Not only have they already assembled stellar populations comparable to local early type galaxies, but they have also already had their star formation strongly quenched, to the extent that they may even form a red sequence (see also Williams et al., 2009; Brammer et al., 2009). On the other hand, in order to develop into galaxies like the ones seen in the local universe, it would seem that they must *each* undergo significant structural evolution. Taken together, these results thus paint a consistent picture of strong size evolution among massive, early type and/or red sequence galaxies[1] — both both individually and as a population — even after their star formation has been quenched (see also Franx et al., 2008). Whatever the mechanism for this growth in size (see, e.g., Fan et al., 2008; Hopkins et al., 2009; Naab et al., 2009; Khochfar & Silk, 2009), it would seem that the formation of massive, passive galaxies is not monolithic.

[1]There is considerable, but not total, overlap between color–selected samples of red sequence galaxies, and morphology–selected samples of early type galaxies. While it is common to use these terms as if they were more or less interchangeable, it should be remembered that they are not.

Our aim in this Chapter is to test the claim that there are no galaxies in the local universe with properties consistent with their being the passively evolved counterparts to the massive, passive, compact galaxies seen at $z \gtrsim 2$. In doing so, we aim to confirm (or otherwise) the idea that each of the galaxies seen at $z \gtrsim 2$ must undergo significant structural evolution between then and now — this is the crux of the argument against the 'monolithic' formation of massive galaxies. Our search is based on many of the latest data products from the Sloan Digital Sky Survey (SDSS; York et al., 2000; Strauss et al., 2002). In particular, we will focus on the possibility that such galaxies have been overlooked in SDSS due to selection effects associated with the construction of the spectroscopic target sample.

The structure of this Chapter is as follows: We describe the basic SDSS data that we have used in Section 2. In Section 3, we define our sample of compact galaxy candidates, and present several checks to confirm that these galaxies are indeed unusually small for their stellar masses. Then, in Section 4, we consider the importance of the SDSS spectroscopic selection for massive, compact galaxies. In this Section, we also compare our $z \sim 0.1$ compact galaxy candidates with the vD08 and D09 samples. Finally, in Section 5, we compare our results to two similar, recent studies, and briefly examine the properties of our compact galaxies' stellar populations in comparison to the general $z \sim 0.1$ red sequence galaxy population.

We also provide a complementary analysis in Appendix A, in which we search for massive, compact, red sequence galaxies the full SDSS photometric sample, using photometric redshifts. In this way, we test our conclusion that the apparent differences between the high- and low-redshift samples cannot be explained by selection effects, and derive an estimate for the number of compact galaxies that may be missing from the SDSS spectroscopic sample.

A summary of our main results is given in Section 6. Throughout this work, we assume the concordance cosmology (*viz.*: $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70$ km/s/Mpc) and a Chabrier (2003) stellar initial mass function (IMF).

## 2 Basic Data and Analysis

The present work is based on Data Release 7 (DR7; Abazajian et al., 2009) of the SDSS, accessed via the Catalog Archive Server[2] (CAS; Thakar et al., 2008). In this section, we describe the different SDSS datasets that we have used, and our analysis of them. We will search for compact galaxy candidates in the SDSS spectroscopic catalog; to this end, we will only consider `sciencePrimary` objects (a flag indicating a "science-grade" spectrum, and weeding out multiple observations of individual objects) with either a `star` or `galaxy` photometric `type` (*i.e.*, a genuine astronomical source). The details of the SDSS spectroscopic sample selection are given in Strauss et al. (2002); we will summarize the most relevant aspects of this process in Section 4.1.

[2] http://casjobs.sdss.org/CasJobs/

### 2.1 The Basic SDSS Catalog

For the basic SDSS catalog, there are two different methods for performing photometry. The first, the Petrosian or `petro` magnitude, is derived from the observed, azimuthally averaged (1D) light profile. The Petrosian radius is defined as the point where the mean surface brightness in an annulus drops to a set fraction (*viz.* 0.2) of the mean surface brightness within a circular aperture of the same radius. Within SDSS, the `petro` aperture is defined to be twice the Petrosian radius; this aperture will contain 99 % of the total light for a well resolved exponential disk, but may miss as much as 18 % of the light for a de Vaucouleurs $R^{1/4}$ profile (Strauss et al., 2002; Blanton et al., 2005).

The second photometric measure is derived from fits to the observed (2D) distribution of light in each band, using a sector-fitting technique, in which concentric annuli are divided into 12 30° sectors (see Appendix A.1 of Strauss et al., 2002). These fits are done assuming either an exponential or a de Vaucouleurs profile, convolved with a fit to the appropriate PSF. For each profile, the structural parameters (*viz.* axis ratio, position angle, and scalelength) are determined from the $r$ band image. The more likely (in a $\chi^2$ sense) of the two profile fits is used to define `model` magnitudes for each galaxy. For the *ugiz* bands, these parameters are then held fixed, and only the overall normalization (*i.e.* total flux) is fit for.

The basic catalog also provides two different measures of size, associated with these two magnitude measurements. The Petrosian halflight radius, $R_{50}$, is defined as the radius enclosing half the 'total' `petro` flux. The catalog also contains best fit structural parameters, including the effective radius, from a separate set of fits to each band independently, again for both an exponential and a de Vaucouleurs profile. Note that whereas the `petro` magnitude and size are derived from the observed, PSF-convolved radial profile, the `model` values provide a PSF-corrected measure of the intrinsic size.

We use `model` magnitudes to construct *ugriz* SEDs for each object, since these measurements are seeing–corrected. From DR7, the basic SDSS photometric calibration has been refined so that the photometry is given in the AB magnitude system without the need for any further corrections (Padmanabhan et al., 2008). For measuring sizes, we will rely on the best-fit `model` effective radius, $R_{\rm e}$, as determined from the $z$ band. We also adopt a minimum measured size of $0\farcs75$, corresponding to half the median PSF FWHM for the SDSS imaging; we will plot all galaxies with observed sizes smaller than $0\farcs75$ as upper limits. (None of our conclusions depend on the choice of this limit, which ultimately affects only 5 of our lower-mass compact galaxy candidates.)

### 2.2 Derived Quantities

We have derived restframe photometry for each object, based on its observed *ugriz* SED and redshift, using the IDL utility InterRest (Chapter II), using a redshift grid of $\Delta z = 0.001$. In order to minimize the k-corrections and their associated errors, we determine restframe photometry through the *ugriz* filters redshifted to $z = 0.1$, which we denote with a superscript 0.1 (see, e.g. Blanton

& Roweis, 2007). We estimate that the systematic uncertainties are at the level of $\lesssim 0.02$ mag. The agreement between our interpolated restframe photometry and that derived using the SDSS `kcorrect` algorithm (Blanton & Roweis, 2007) is very good: our derived $(u-r)$ colors are $\sim 0.02$ mag bluer for blue galaxies, and $\sim 0.03$ mag redder for red galaxies.

We use the stellar mass estimates derived by the MPIA Garching group.[3] JB has fit the *ugriz* `model` photometry of all galaxies using the synthetic stellar population library described by Gallazzi et al. (2005), based on Bruzual & Charlot (2003) models and assuming a Chabrier (2003) IMF in the range 0.1–100 $M_\odot$. The Gallazzi et al. (2005) library contains a large number of Monte Carlo realizations of star formation histories, parameterized by a formation time ($1.5 < t_{\rm form}/[{\rm Gyr}] < 13.5$), an exponential decay rate ($0 < \gamma/[{\rm Gyr}^{-1}] < 1$), and including a number of random star formation bursts (randomly distributed between $t_{\rm form}$ and 0, normalized such that 10 % of galaxies experience a burst in the last 2 Gyr). In the fitting, the photometry has been corrected for emission lines under the assumption that the global emission line contribution is the same as in the spectroscopic fiber aperture.

The agreement between these SED-fit mass estimates and those of Kauffmann et al. (2003a), which were derived from spectral line indices, are excellent: the median offset is -0.01 dex, with a scatter on the order of 0.1 dex. For the highest masses, however, the SED-fit results are slightly less robust: for $M_* > 10^{11}$ $M_\odot$, the median formal error is $\lesssim 0.10$ dex, compared to $\lesssim 0.06$ dex for the Kauffmann et al. (2003a) estimates. (Note that these uncertainties do not include, for example, uncertainties intrinsic to the stellar population models, and are thus underestimated; see Conroy, Gunn & White, 2009)

In the upper panel of **Figure 1**, we show the stellar mass to light ratios, $M_*/L_i$, for $0.066 < z < 0.12$ galaxies as a function of their ${}^{0.1}(g-i)$ color (again, $L_i$ should be understood as referring to the $i$-band filter redshifted to $z = 0.1$, or ${}^{0.1}i$). Notice that, at least for these mass estimates, $M_*/L$ is very tightly correlated with color. In the main panel of this Figure, the red line shows the median $M_*/L_i$ in narrow color bins. Making a simple linear fit to these points, we find:

$$\log(M_*/L_i) = -0.82 + 0.83 \times {}^{0.1}(g-i) \ , \qquad (1)$$

where both $M_*$ and $L_i$ are in solar units. (The absolute magnitude of the sun in the ${}^{0.1}i$ band is 4.58.) This relation is shown in Figure 1 as the solid blue line. We present this relation as an alternative to the popular Bell & de Jong (2001) or Bell et al. (2003) relations.

In the lower panel of Figure 1, we show the dispersion around the median relation; in this Figure, the error bars show the 16/84 percentiles in the $M*/L$ distribution in narrow bins of color. Overall, the dispersion around this relation is just 0.032 dex. Note that while the simple linear relation given above provides an acceptable description of the 'true' relation, systematic offsets exist at the 0.02–0.04 dex level. The global mean and random offset from this linear relation are 0.002 dex and 0.040 dex, respectively.

[3]Available via http://www.mpa-garching.mpg.de/SDSS/DR7/Data/stellarmass.html .

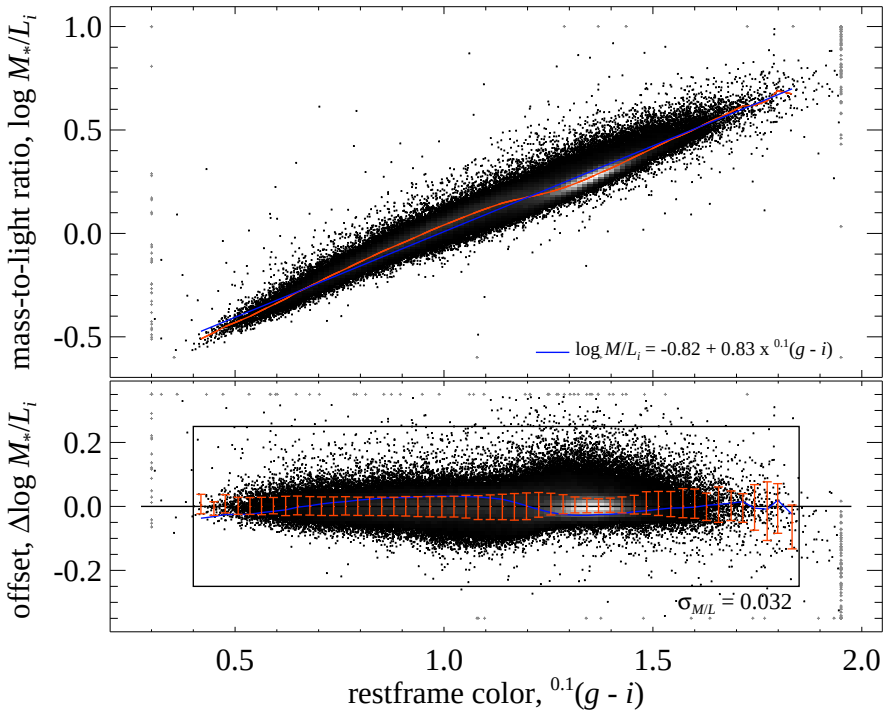


**Figure 1.** — The mass–to–light ratios, $M_*/L_i$, of $0.066 < z < 0.12$ galaxies as a function of their $^{0.1}(g-i)$ color. — The greyscale shows the (linear) data density in cells, where the data density is high. In the main panel, the red line shows the median $M_*/L_i$ in narrow bins of $^{0.1}(g-i)$ color; the blue line is a linear fit to these points. (Here, $M_*/L_i$ should be understood to relate to the $i$ band redshifted to $z = 0.1$.) In the lower panel, we have simply subtracted away the median relation; in this panel, the error bars show the 16/84 percentiles in color bins. The simple linear relation shown provides an acceptable description of the observed relation, to within 0.02–0.04 dex; the global RMS offset from this relation is 0.032 dex. In order to avoid selecting 'catastrophic failures' in terms of stellar mass estimates, we will consider only those galaxies that have $0.4 < {}^{0.1}(g-i) < 1.8$, and that fall within 0.25 dex of the median $M_*/L_i$–$^{0.1}(g-i)$ relation; this selection is shown by the box in the lower panel.

In both panels, the small grey pluses show points that fall outside the plotted range. Notice that a small fraction of galaxies lie well off the main $M_*/L$–color relation, some by an order of magnitude or more. These galaxies also lie significantly off the main stellar mass–dynamical mass relation and are very likely to represent catastrophic failures of the stellar mass SED-fitting algorithm (see Section 3.1 below). This presents a problem when it comes to looking for outliers in the mass–size plot: selecting the most extreme objects may well include those objects with the largest errors. For this reason, we will restrict our attention to those objects that fall within 0.25 dex ($\approx 7.8\sigma$) of the main $M_*/L$–color relation, and with $0.2 < {}^{0.1}(g-i) < 1.8$, as shown by the box in the lower panel of Figure 1. This selection excludes just under 600 of the 223292 galaxies shown in Figure 1.

## 3 Searching for Massive, Compact, Early-Type Galaxies in the Local Universe

### 3.1 Identifying Massive, Compact Galaxy Candidates

**Figure 2** shows the size–mass plot for a sample of massive, red-sequence galaxies drawn from the SDSS DR7 spectroscopic catalog; this sample has been selected to have ${}^{0.1}(u-r) > 2.5$ in the range $0.066 < z < 0.12$. Since we are interested primarily in potential passively-evolved analogues to the $z \gtrsim 1.5$ galaxies seen by vD08 and D09, most of our analysis will focus on this red sequence sample; we will briefly consider massive, compact, blue galaxies in Section 5.2.

We have chosen our redshift limits to minimize the importance of selection effects and measurement biases, which we will discuss in Section 4.1. For now, we note that, mapping the $m_r < 17.77$ spectroscopic limit onto $M_*(z)$, we should be highly complete (volume limited) for $M_* > 10^{10.7}$ $\mathrm{M}_\odot$ and $z < 0.12$. As a very simple check on this, we note that for this sample, the median redshift in narrow bins of stellar mass is within the range $z = 0.098$–$0.102$ for all $M_* > 10^{10.7}$ $\mathrm{M}_\odot$; the volumetric center of the $0.066 < z < 0.12$ bin is $z = 0.10$.

The yellow points in this Figure show the median size in narrow bins of stellar mass; the error bars show the 14/86 percentiles. For comparison, the long-dashed line shows the local size–mass relation for early-type galaxies from Shen et al. (2003), corrected for differences in assumed IMF and cosmology. Contrary to the findings of Valentinuzzi et al. (2009), a simple fit to the size–mass relation for *red sequence* galaxies (${}^{0.1}(u-r) > 2.5$) shown in Figure 2 is consistent with the Shen et al. (2003) relation for *early type* ($n > 2.5$) galaxies, albeit offset by $-0.05$ dex in size, or, equivalently, by $-0.09$ dex in mass. At fixed mass, the mode of the distribution is similarly offset (see Figure 7); this does not appear to be due to large numbers of late type galaxies in our sample of red sequence galaxies.

We next select and study very compact galaxies from within the red sequence sample shown in Figure 2. At first glance, it appears that there may be a few galaxies that lie well below the main size–mass relation. However, it must be remembered that by selecting the most extreme outliers, we will also be selecting those objects with most egregious measurement errors.

For this reason, we have individually visually inspected all $M_* > 10^{10.7}$ $\mathrm{M}_\odot$ galaxies with inferred sizes that are less than half the size predicted from the Shen et al. (2003) relation; *i.e.* $\Delta R_\mathrm{e} < -0.3$ dex. For sizes smaller than the median relation, the distribution of sizes around the Shen et al. (2003) relation is very well described by a log-normal with $\sigma = 0.11$ dex; this $\Delta R_\mathrm{e}$ cut thus corresponds to selecting those galaxies whose sizes are smaller than the mean size (at fixed mass) at the $\gtrsim 2.7\sigma$ level. (Adopting our own fit to the size–mass relation, this selection translates to $\Delta R_\mathrm{e} < -0.35$ dex; our results are otherwise unchanged.)

We have inspected 280 such objects, and discarded those where there are obvious reasons to distrust the size measurements. The most common reasons for discarding galaxies are confusion with other galaxies (99 galaxies, including 19 good merger candidates, and two possible lenses), or with the extended halos, diffraction spikes, and/or reflections of bright stars (62 galaxies). Another 19

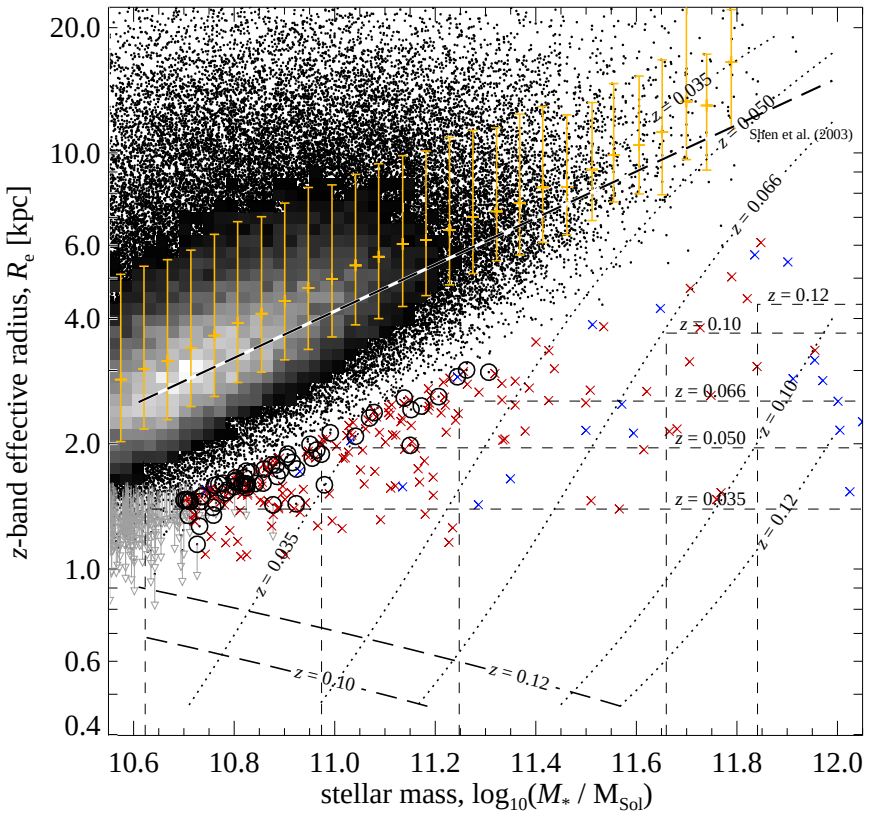


**Figure 2.** — The size–mass relation for massive, red sequence galaxies, illustrating the importance of the SDSS spectroscopic selection criteria. — The points show SDSS galaxies ($0.066 < z < 0.12$) selected to have ${}^{0.1}(u-r) > 2.5$. The yellow points show the median size in narrow bins of stellar mass; the error bars show the 16/84 percentiles in each bin. A fit to this median size–mass relation for red sequence galaxies is consistent with the Shen et al. (2003) relation for early type galaxies (dashed line), albeit offset by 0.05 dex in $R_e$. Individual galaxies that we have visually inspected ($M_* > 10^{10.7}$ $M_\odot$; $\Delta \log R_e < -0.3$ dex) are marked with large symbols. Galaxies with $M/L$s that differ significantly from the main color–$M/L$ relation shown in Figure 1 are marked with small blue crosses. Galaxies with obvious problems in their photometry (especially those affected by the presence of a bright nearby star or blended with other galaxies) are marked with a small red cross; those that look okay are plotted as circles. Galaxies with observed sizes smaller than $0\farcs75$ are plotted as upper limits, assuming a size of $0\farcs75$. The different lines show how the principal selection limits for spectroscopic followup translate onto the $(M_*, R_e)$ plane for $z = 0.12$, 0.10, 0.066, 0.050, and 0.35 (top to bottom): the diagonal, long-dashed lines show the star/galaxy discriminator; the short-dashed boxes show the 'saturation' selection limit, and the diagonal dotted lines show the 'cross-talk' selection limit (see Section 4.1 for a detailed discussion). Galaxies lying below these lines will not be targeted spectroscopically. Note that, at least for massive galaxies, the star/galaxy separation criterion is not a major source of incompleteness for $z \lesssim 0.10$. Because both the 'cross talk' and 'saturation' criteria exclude high surface brightness objects, the SDSS spectroscopic sample is potentially highly incomplete for bright, compact galaxies for *low* redshifts. For instance, the 'cross-talk' selection criterion cuts out a large fraction of the massive galaxy population, including the majority of $M_* \gtrsim 10^{11.5}$ $M_\odot$ red sequence galaxies, at $z \lesssim 0.35$. In the $0.066 < z < 0.12$ that we consider, the 'saturation' criterion is the biggest potential cause of incompleteness.

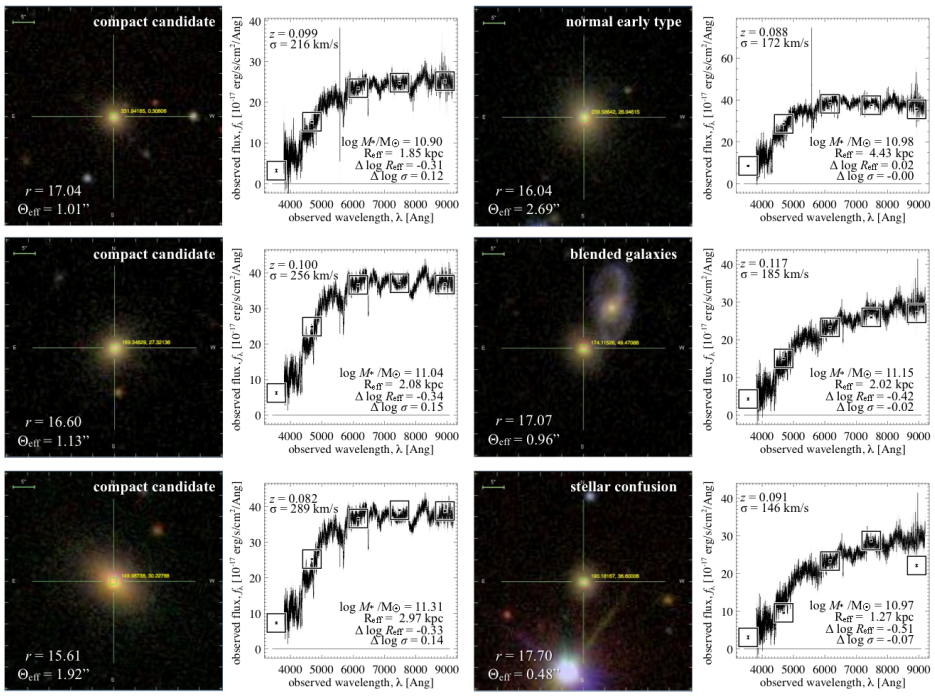


**Figure 3.** — Illustrative examples of the galaxies we consider. — For each galaxy, we show the SDSS SkyServer thumbnail image used for visual inspection, as well as the galaxies' observed spectra; the boxes show the SEDs from the photometry, scaled to match the spectroscopic flux in the $r$ band. Clockwise from the top-right, we show a 'normal', massive early type galaxy that lies very close to the median size–mass and velocity dispersion–mass relations, two compact galaxy candidates where visual inspection suggested problematic size measurements, and three of our compact galaxy candidates. Each of the three compact galaxy candidates shown in this Figure have observed velocity dispersions that are approximately consistent with their small measured sizes (see Section 3.3).

galaxies were clearly disk-like, 5 showed marked asymmetries, and 1 had a very strong AGN spectrum; these candidates were also discarded. We discarded a further 3 objects with bad or missing data.

In **Figure 3**, we show several illustrative examples of the galaxies we are considering. On the right-hand side of this Figure, we show a 'normal' early type galaxy, with $M_* \approx 10^{11}$ M$_\odot$, which falls very close to the Shen et al. (2003) relation. Below this, we show two of the compact galaxy candidates that we have rejected on the basis of visual inspection. On the left-hand side of this Figure, we show three of the compact galaxy candidates of different stellar masses that we have retained after visual inspection. For each galaxy, we show the thumbnail image from the SDSS SkyServer[4], used for visual inspection. We also show each galaxy's observed spectrum and photometry; here, we have scaled the photometry

[4]Also accessible via CAS at http://cas.sdss.org.

| RA (1)[a] | dec (2)[a] | $z$ (3)[a] | $(u-r)_{\rm obs}$ (4)[a] | $^{0.1}(u-r)$ (5) | $\log M_*$ (6)[b] | $\Theta_{50,z}$ (7)[a] | $n$ (8)[c] | $\Theta_{50,z}$ (9)[c] | $R_{\rm e}$ (10) | $\sigma$ (11)[b] | $T$ (12)[d] | $Z$ (13)[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 190.16652 | 13.81563 | 0.0855 | 2.577 | 2.642 | 10.700 | 0.912 | 4.31 | 0.917 | 1.462 | 160 | — | — |
| 127.02722 | 55.37988 | 0.0665 | 2.828 | 2.999 | 10.701 | 1.152 | 3.26 | 1.211 | 1.468 | 191 | — | — |
| 225.31708 | 30.58266 | 0.0980 | 2.858 | 2.877 | 10.705 | 0.808 | 3.87 | 0.773 | 1.464 | 195 | — | — |
| 227.08531 | 7.25325 | 0.0764 | 2.963 | 3.113 | 10.709 | 0.929 | 5.13 | 1.188 | 1.345 | 199 | — | — |
| 215.41043 | 40.03233 | 0.0998 | 2.803 | 2.813 | 10.709 | 0.795 | 4.11 | 0.670 | 1.464 | 176 | 9.775 | 0.035 |
| 222.12988 | 26.48791 | 0.1058 | 2.549 | 2.540 | 10.712 | 0.750 | 4.65 | 0.722 | 1.453 | 155 | — | — |
| 118.81702 | 33.22864 | 0.0980 | 2.680 | 2.697 | 10.713 | 0.803 | 2.67 | 0.739 | 1.454 | 154 | -99 | -99 |
| 143.05707 | 11.70454 | 0.0811 | 2.690 | 2.776 | 10.726 | 0.750 | 5.52 | 0.830 | 1.146 | 166 | 9.255 | 0.132 |
| 204.66577 | 59.81854 | 0.0704 | 2.857 | 3.012 | 10.731 | 0.943 | 3.92 | 0.974 | 1.267 | 235 | 9.845 | 0.229 |
| 230.28553 | 24.21978 | 0.0809 | 2.922 | 3.028 | 10.733 | 0.952 | 5.90 | 1.298 | 1.453 | 153 | — | — |

**Table 1.** — Properties of our compact galaxy candidates. — Col.s (1) and (2): position in decimal degrees (J2000); Col. (3) spectroscopic redshift; Col.s (4) and (5): observed and restframe colors; Col. (6): stellar mass in units of solar masses; Col. (7): apparent De Vaucouleurs `model` effective radius in arcsec; Col.s (8) and (9): Sérsic index and apparent Sérsic effective radius in arcsec; Col. (10): physical De Vaucouleurs effective radius in kpc; Col. (11): velocity dispersion in km/s; Col. (12): luminosity-weighted age in Gyr; Col. (13): mean metallicity. In Col.s (12) and (13), data are only given for those objects that appear in DR4; values of -99 indicate unsuccessful fits to the spectra. Note that we give only the first ten candidates here; the properties of the full sample of 63 galaxies is available as an electronic table via http://www.strw.leidenuniv.nl/∼ent/.Sources. — [a] the default SDSS (York et al., 2000; Strauss et al., 2002) catalog for DR7 (Abazajian et al., 2009), accessed via CAS (Thakar et al., 2008); [b] the MPA-JHU catalog for DR7 (accessible via http://www.mpa-garching.mpg.de/SDSS/DR7/); [c] the NYU VAGC (Blanton et al., 2005) for DR7; [d] the stellar age and metallicity catalog of Gallazzi et al. (2005) for DR4.

to match to the integrated $r$ band flux from the observed spectrum.

In addition to these galaxies with suspect size measurements, we have excluded 27 galaxies whose SED-fit $M_*/L$s are offset from the main color–$M_*/L$ relation shown in Figure 1 by more than 0.25 dex. If we use Equation 1 to derive new stellar mass estimates for these galaxies, all of these galaxies move back into the main cloud in both Figure 2 and a stellar mass-dynamical mass plot, with mean/median offsets of $\lesssim 0.02$ dex in both cases. This strongly suggests the SED-fit $M_*/L$s for these 27 galaxies are simply wrong.

The 190 galaxies discarded on the basis of inspection are shown in Figure 2 as small red crosses; the small blue crosses show the 27 galaxies with discrepant $M_*/L$s. As a function of $\Delta R_{\rm e}$, the fraction of inspected sources that have been discarded goes fairly smoothly from 60 % for $\Delta R_{\rm e} \sim -0.3$ dex to $\sim 100$ % for $\Delta R_{\rm e} < -0.5$ dex. The discarded fraction has a similar dependence on mass: it is $\sim 60$ % for $M_* \sim 10^{10.7}$ M$_\odot$, rising to $\sim 85$ % for $M_* \sim 10^{11}$ M$_\odot$, and 100 % for $M_* > 10^{11.4}$ M$_\odot$.

This leaves us with a sample of 63 massive, compact, early-type and red sequence galaxy candidates; these are are marked in Figure 2 with heavy black circles. Of those galaxies that we have retained, 8 % (5/63) have observed sizes smaller than $0\farcs75$; all of these have $M_* < 10^{11}$ M$_\odot$. We have provided the properties of our compact galaxy candidates in **Table 1**.

### 3.2 Are the Size Measurements Wrong?

We have performed a number of checks to validate the small measured sizes of our compact galaxy candidates. The compact galaxy candidates do not have significantly larger size measurement errors in comparison to the full sample shown in Figure 2. For both the $r$- and $z$-bands, our candidates are not anomalous in a plot of `petro` halflight radius versus `model` effective radius, nor are they anomalous in a plot of $r$ band size versus $z$ band size. For all but two of the candidates, the `petro` and `model` magnitudes agree to within 0.15 mag. The mean offset between `model` and `petro` magnitudes is -0.06 mag for the compact galaxies, compared to $-0.08$ mag for the full sample shown in Figure 2. That is, the compact candidates appear to be well described by the de Vaucouleurs `model` fits.

For the New York University (NYU) Value Added Galaxy Catalog (VAGC), Blanton et al. (2005) have mode Sérsic-profile fits to the radially-averaged light profiles of each object, allowing the Sérsic index $n$ to vary over the range $0 \leq n < 6$. In order to explore further the issue of the quality of the de Vaucouleurs `model` fits, we have compared the `model` effective radii to those given in the VAGC.

In **Figure 4**, we show the distribution of Sérsic parameters for our candidates, as well as a comparison between the Sérsic and de Vaucouleurs sizes. First, we note that nearly all (59/63) of our compact galaxy candidates have $n > 3$; these are not late-type (exponential) galaxies. It is therefore unsurprising — but nonetheless reassuring — that the two size measures agree quite well: for the median galaxy among our candidates, the de Vaucouleurs size is $\sim 10\%$ smaller than the Sérsic size; the RMS dispersion is 0.10 dex. For comparison, the median quoted error

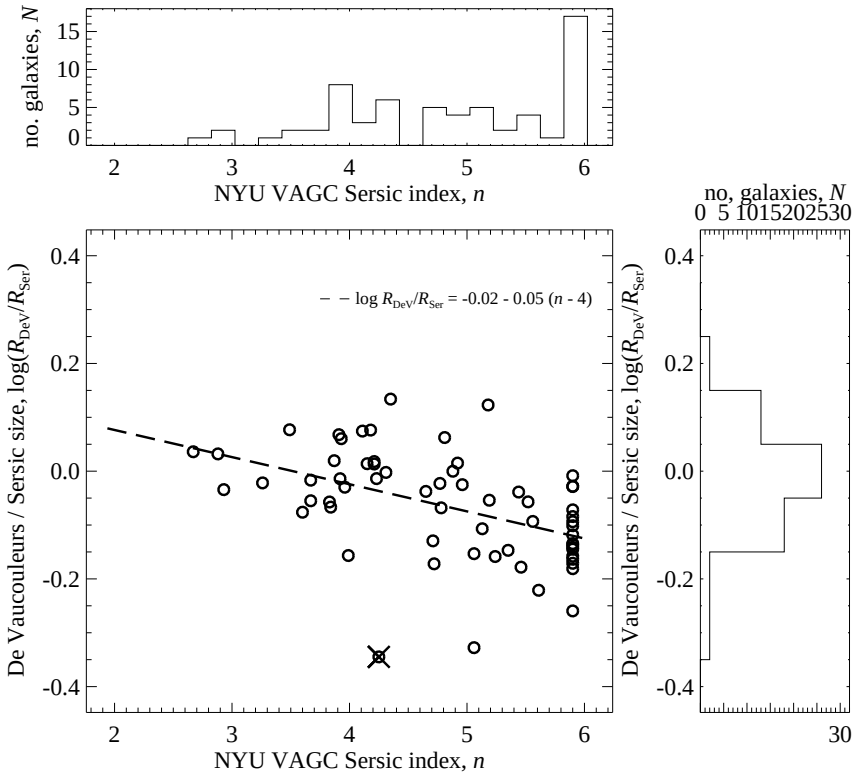


**Figure 4.** — Comparison between effective radii derived assuming a de Vaucouleurs ($n = 4$) profile and assuming a Sérsic profile ($0 < n < 6$). — Whereas the basic SDSS catalog uses a sector-fitting technique to fit either an exponential ($n = 1$) or a de Vaucouleurs ($n = 4$) profile, for the NYU VAGC, Blanton et al. (2005) have fit the radial profiles of each object assuming a general Sérsic model ($0 < n < 6$). This Figure shows the ratio of these two sizes for our compact galaxy candidates, based on the $z$ band data, as a function of Sérsic index $n$. Almost all candidates have $n > 3$ — these galaxies are not exponentials. However, approximately 25 % have $n = 5.9$; the maximum value allowed in the Blanton et al. (2005) fits. For these galaxies, the median ratio between the two size measurements is 0.88, with an RMS scatter of 0.1 dex. In the main panel, we show a least-squares fit to the data, the dispersion around this relation is $\lesssim 0.1$ dex.

for the de Vaucouleurs size measurements is 4.6%.

Notice that $\approx 25$ % (17/63) of our candidates have $n = 5.9$ in the NYU VAGC (the maximum value allowed in the fits). These galaxies are considerably more centrally-concentrated than the canonical de Vaucouleurs $R^{1/4}$-law profile. However, the trend with increasing Sérsic index is for the de Vaucouleurs size, $R_{\rm DeV}$, to be systematically lower than the Sérsic size, $R_{\rm Ser}$: making a least-squares fit to the data shown in Figure 4, we find $\log R_{\rm DeV}/R_{\rm Ser} = -0.02 - 0.05\,(n-4)$. If these galaxies do have $n > 6$, then we may well be underestimating their sizes by $\gtrsim 25$ %.

Guo et al. (2009) have recently demonstrated that as a result of biases in the way the background sky level is estimated for the Sérsic fits, the NYU-VAGC sizes are systematically underestimated at the $\gtrsim 15$ % level for $n \gtrsim 5$. This problem becomes progressively worse for large sizes ($\Theta_{\rm e} \gtrsim 1''$) and bright magnitudes ($m_r \lesssim 16$); for our compact galaxy candidates, the effect is likely to be at the $\sim 20$ % level. But note this implies that the difference between the de Vaucoleurs and Sérsic sizes is even greater than Figure 4 might suggest: the sizes of the $n \gtrsim 5$ compact galaxies may be underestimated by as much as $\gtrsim 30$ %.

As a final check, therefore, we have also rederived Sérsic effective radii for our compact galaxy candidates using GALFIT (Peng et al., 2002) and done a similar comparison as for the NYU VAGC sizes. The agreement between the GALFIT and VAGC Sérsic indices is quite good, with an rms difference in $n$ of 1.1. Again the vast majority of objects have $n > 3$. There are 19 objects that are assigned the maximum allowed value of $n = 8$, but only 9 of these have $n = 5.9$ in the VAGC. Making a similar fit to the difference between the `model` De Vaucouleurs and the GALFIT Sérsic effective radii, we find $\log R_{\rm DeV}/R_{\rm Ser} = 0.08 - 0.08\ (n - 4)$. As before, we may be underestimating the sizes of high $n$ galaxies by 10–35 %; at the same time this comparison does also suggest that we may well be overestimating the sizes of the few candidates with $n < 4$. The median galaxy has a GALFIT Sérsic effective radius 15 % smaller than the default De Vaucouleurs value. Lastly, we note that there is a definite mass-dependence to the agreement between the GALFIT Sérsic and default De Vaucouleurs effective radii, such that all but one of the galaxies for which the sizes agree to within 20 % have $M_* > 10^{11}$ M$_\odot$.

To summarize the results of this section: comparisons to 1D and 2D Sérsic fits do not suggest that the De Vaucouleurs `model` effective radii are catastrophically wrong for our compact galaxy candidates. If anything, we may in fact be *underestimating* the sizes of these galaxies by 10–30 %. Note that using underestimates of the sizes of local galaxies will effectively *reduce* the size of any apparent discrepancy in comparison to the compact, high-redshift galaxies.

### 3.3 A Consistency Check Based on Velocity Dispersions

Assuming that elliptical galaxies are structurally self-similar, the virial theorem implies that $M_* \propto R_{\rm e}\sigma^2$. At fixed mass, galaxies with small sizes should therefore have higher velocity dispersions, with $\Delta\sigma \propto (\Delta R_{\rm e})^{-1/2}$, where $\Delta \log \sigma$ and $\Delta \log R_{\rm e}$ are the offsets from the $M_*$–$\sigma$ and size–mass relations for local early type galaxies, respectively.

To determine whether the observed velocity dispersions of our compact galaxy candidates are consistent with their inferred sizes and masses, in the righthand panel of **Figure 5** we plot $\Delta \log R_{\rm e}$ against $\Delta \log \sigma$. The $M_*$–$\sigma$ relation itself is shown in the lefthand panel of the Figure. (Note that for this plot, we have shifted the Shen et al. (2003) relation upwards in size by 0.05 dex to be consistent with the present data set; our conclusions do not depend on this decision.) The greyscale and points show those $0.066 < z < 0.12$ galaxies with ${}^{0.1}(u - r) > 2.5$ and $M_* > 10^{10.7}$ M$_\odot$; the red circles indicate our compact galaxy candidates.

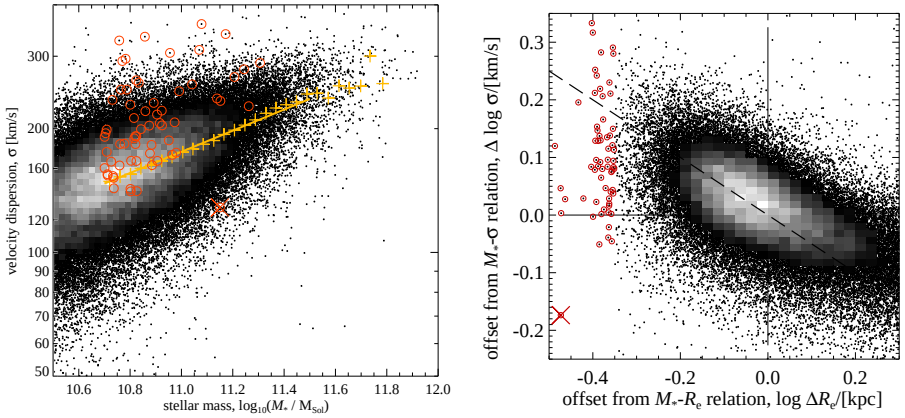


**Figure 5.** — Using observed velocity dispersions to validate the measured sizes of our compact galaxy candidates. — *Left panel*: the mass–velocity dispersion relation for red sequence galaxies at $0.066 < z < 0.12$. The points and greyscale show the SDSS data. The yellow plusses show the median velocity dispersion in narrow bins of stellar mass; the solid yellow line shows a simple fit to these points for $10.7 < \log M_* < 11.5$. The red circles highlight our compact galaxy candidates. *Right panel*: the offset from the $M_*$–$R_\mathrm{e}$ relation, plotted against the offset from the $M_*$–$\sigma$ relation for $M_* > 10^{10.7}$ $\mathrm{M}_\odot$ galaxies with $^{0.1}(u-r) > 2.5$. If the offsets from these two relations is a function of galaxy size, then we expect $\Delta \log \sigma = -0.5 \times \Delta \log R_\mathrm{e}$ (long dashed line). Our compact galaxy candidates are shown as the red circles. In general, the observed velocity dispersions support the idea that our compact galaxy candidates are indeed relatively small; this is particularly true for those with $\sigma > 200$ km/s. There is one clear exception, marked with a cross in both panels; this galaxy is also the most extreme outlier in Figure 4.

For the galaxies that we have identified as 'compact', the median offset from the size–mass relation is $\Delta \log R_\mathrm{e} = -0.38$ dex. We would therefore expect a median offset from the $M_*$–$\sigma$ relation relation of $\Delta \log \sigma = -0.5 \times -0.38 = 0.19$ dex. The median value for $\Delta \log \sigma$ is 0.12 dex–roughly 85 % of the expected value, and $\sim 1.5$ times greater than the intrinsic scatter in the relation. Overall, these results are fairly consistent, although they do indicate that the sizes may be underestimated and/or the masses may be overestimated at the level of 10–20 %. We note that the difference between the default SDSS and the NYU VAGC size measurements can account for at least half of this effect (see Section 3.2).

There is one of our compact galaxy candidates, marked in Figure 5 with a cross, whose velocity dispersion is clearly inconsistent with its being massive and compact; indeed, it has the lowest observed velocity dispersions of all of our compact galaxy candidates. This galaxy is also the biggest outlier in Figure 4. We will discuss this object in more detail in Section 4.2.

We also note that the observed velocity dispersions of the most extreme outliers from the size–mass relation ($\Delta \log R_\mathrm{e} \lesssim -0.4$) are only marginally higher than for galaxies with 'normal' sizes. Only one of these candidates ($\log M_* = 10.73$) has $\Delta \log \sigma > 0.18$ dex; the median value of $\Delta \log \sigma$ is 0.03 dex. It would seem that the effects of 'outlier noise' (*i.e.* objects being pushed to the edge of the observed

distribution by measurement errors, rather than by virtue of their true, intrinsic properties) become dominant at these very extreme values of $\Delta \log R_{\rm e}$.

With these caveats, the observed velocity dispersions generally support the idea that the offsets from both the $M_*$–$R_{\rm e}$ and $M_*$–$\sigma$ relations for our compact galaxy candidates can be explained by their having small sizes for their masses/velocity dispersions.

## 4 The Importance of Selection Effects for Massive, Compact Galaxies

### 4.1 SDSS Spectroscopic Sample Selection

In order to be targeted for SDSS spectroscopic follow-up (and thus to appear in Figure 2), galaxies have to satisfy a complicated set of selection criteria (Strauss et al., 2002). In brief, there is a magnitude cut: objects must be detected at $> 5\sigma$ significance in the $r$ band, and have $m_{{\rm Pet},r} < 17.77$. Any objects that have been marked as blended and then segmented into smaller objects are rejected, as are any objects that include saturated pixels, or have been deblended from objects with saturated pixels. There are also a series of (low) surface-brightness-dependent criteria that are not relevant here. There are three further selection criteria that are particularly important for the relatively bright, compact galaxies we are interested in here.

The first of these is the star/galaxy separation criterion. Star/galaxy separation is done on the basis of the difference between the `PSF` and `model` magnitudes in the $r$ band. Here, the `PSF` magnitude $m_{\rm PSF}$, is derived by fitting the surface brightness profile of each object with the appropriate PSF model. In analogy to the exponential/de Vaucouleurs `model` fits described in Section 2.1, the profile shape is kept fixed in the fitting, so that only the overall normalization is allowed to vary. The value of the $m_{\rm PSF}$ is then defined as the flux implied by this fit within a $7\farcs4$ aperture. Specifically, objects are only selected for spectroscopic follow-up where:

$$\Delta_{\rm SG} \equiv m_{{\rm PSF},r} - m_{{\rm mod},r} \geq 0.3 \text{ mag}. \qquad (2)$$

Here, $m_{{\rm mod},r}$ is the `model` $r$ band magnitude, as described in Section 2.1. The star/galaxy discriminator thus selects against unresolved objects.

Further to this selection, there are two separate selections that operate against high surface brightness objects. This first is designed to avoid cross-talk between spectroscopic fibers: any objects with `fiber` magnitudes $m_{{\rm fib},g} < 15$, $m_{{\rm fib},r} < 15$, and $m_{{\rm fib},i} < 14.5$ are rejected. Finally, all objects with $m_{{\rm Pet},r} < 15$ and a Petrosian radius $\Theta_{\rm Pet} < 2''$ are rejected. This last criterion was introduced to eliminate "a small number of bright stars that that managed to satisfy equation [2] during the commissioning phase of the survey, when the star/galaxy separation threshold was $\Delta_{\rm SG} = 0.15$ mag, and was retained for later runs to avoid saturating the spectroscopic CCDs" (Strauss et al., 2002). Strauss et al. (2002) also note that of the approximately 240000 $m_{{\rm Pet},r} < 17.77$ objects in runs 752 and 756, none were rejected by the $m_{{\rm Pet},r} < 15$, $\Theta_{\rm Pet} < 2''$ criterion alone.

In order to model these selections, we need to relate the relevant observed quantities (*viz.*, the apparent `petro` magnitude, $m_{\mathrm{Pet},r}$, *gri* `fiber` magnitudes, the apparent Petrosian size, $\Theta_{Pet}$, and the star/galaxy separation parameter, $\Delta_{\mathrm{SG}}$) to intrinsic size and stellar mass.

For a given redshift/distance, the intrinsic size can be trivially related to the observed effective radius, $\Theta_{\mathrm{e}}$. In order to relate $m_{\mathrm{Pet},r}$ to $M_*$, we have made a simple fit to the relation between stellar mass and absolute magnitude in the observers frame $r$ band (*i.e.* with no K-correction) for red sequence galaxies at $0.066 < z < 0.12$ with $M_r > -21$. Note that this method naturally accounts for mass-dependent trends in, e.g., metallicity along the red sequence. The scatter around this relation is $\sim 0.06$ dex, with no obvious magnitude dependence. We have derived similar relations for both $m_{\mathrm{Pet},g}$ and $m_{\mathrm{Pet},i}$. We have derived similar empirical relations for $\Theta_{\mathrm{Pet}}$, $\Delta_{\mathrm{SG}}$, and the difference between the `petro` and `fiber` magnitudes, $\Delta_{\mathrm{fib}} = m_{\mathrm{Pet}} - m_{\mathrm{fib}}$, as functions of $\Theta_{\mathrm{e}}$ and $m_{\mathrm{Pet},r}$, using the sample of massive, red sequence galaxies shown in Figure 2. The scatter around these relations is 0.059 dex (15 %) , 0.18 mag (18 %), and 0.11 mag (9 %) respectively, with no obvious systematic residuals.

There is a danger of circularity in this argument: any objects that do not satisfy the selection criteria will not be present in the sample that we use to model the selection criteria. The crucial assumption here, then, is that we can extrapolate the functions for $\Theta_{\mathrm{Pet}}(\Theta_{\mathrm{e}},\ m_{\mathrm{Pet},r})$, $\Delta_{\mathrm{SG}}(\Theta_{\mathrm{e}},\ m_{\mathrm{Pet},r})$, and $\Delta_{\mathrm{fib}}(\Theta_{\mathrm{e}},\ m_{\mathrm{Pet},r})$ down past the limits of the spectroscopic sample. In this regard, it is significant both that the derived functions are smooth all the way down to the selection limits, and that we do not see obvious cut-offs in the data associated with these limits.

In Figure 2, we show how these selection criteria translate onto the $(M_*, R_{\mathrm{e}})$ plane for several example redshifts between 0.035 and 0.12. The thicker, roughly diagonal, long-dashed lines represent the star/galaxy separation criterion; the dotted lines represent the 'cross-talk' `fiber` magnitude selection; the thinner, short-dashed boxes represent the effect of the 'saturation' selection against bright, compact objects.

With reference to this Figure, let us now consider each of these three selection criteria in turn. First, it turns out that the star/galaxy separation criterion does not play an important role in terms of incompleteness. This is simply due to the fact that the massive galaxies we are interested are bright enough that they have enough flux in their wings to make them clearly distinguishable from stars, even despite their small intrinsic sizes.

In fact, they are so bright that for $z \lesssim 0.05$, they would induce cross-talk in the spectrograph. For $z \lesssim 0.035$, many — perhaps even most — of the most massive ($M_* \gtrsim 10^{11.3}$ $\mathrm{M}_\odot$) red sequence galaxies would not be considered as spectroscopic targets, because of their high surface brightness.

This is why we have chosen to focus on the $0.066 < z < 0.12$ interval; at lower redshifts, this cross-talk selection means that galaxies like those of vD08 and D09 would not be targeted for spectroscopic followup. In this regime, it turns out that the most important selection effect is due to the $m_{\mathrm{Pet},r} < 15$, $\Theta_{\mathrm{Pet}} < 2''$ saturation criterion.

We stress the fact that incompleteness becomes *less* of an issue at *higher* redshifts. For example, any galaxies with $M_* \gtrsim 10^{11.3}$ $M_\odot$ and $R_e \lesssim 2$ kpc would not be selected as spectroscopic targets if placed at $z < 0.066$.

### 4.2 Compact Galaxies at High and Low Redshifts

In **Figure 6** we again show the size–mass relation for our sample of massive, red sequence galaxies at $0.066 < z < 0.12$, with the exception that we have not plotted those galaxies rejected as described in Section 3.1. Furthermore, in contrast to Figure 2, we have used the selection limits derived in Section 4.1 to estimate the relative completeness of the SDSS spectroscopic sample across the $0.066 < z < 0.12$ volume; these are shown by the contours. These completeness estimates also include the $m_{\mathrm{Pet},r} < 17.77$ selection limit, which can be seen to affect galaxies with $M_* \lesssim 10^{10.6}$ $M_\odot$ at the distant end of our redshift window.

For comparison, we have also overplotted the high-redshift samples of D09 and vD08, blue points (blue points). Where we have used size measurements from the $z$ band for the SDSS galaxies, these high-redshift studies use the NICMOS F160W filter, which corresponds to restframe $r$ at $z = 1.6$, moving close to $g$ by $z = 2.3$. Locally, the difference between $z$- and $r$-band measured sizes leads to a slightly different slope to the size–mass relation for red sequence galaxies (a slope of 0.65, rather than 0.56). The $r$- and $z$-band size–mass relations intersect at around $M_* \sim 10^{10}$ $M_\odot$; the mean $r$ band size at $M_* \sim 10^{11}$ $M_\odot$ is 15 % larger than in the $z$ band. That is, by using $z$ band derived effective radii, we are, if anything, *underestimating* the sizes of the local galaxies in comparison to those at high redshift. Similarly, our decision to use the De Vaucouleurs effective radii given in the basic SDSS catalog, rather than more general Sérsic ones appears to lead to an underestimate of galaxy sizes. In other words, adopting $r$- or $g$-band derived sizes, or using Sérsic instead of De Vaucouleurs effective radii, would *increase* the discrepancy between the high– and low–redshift samples.

There is one of our candidates (marked with a cross) that appears to have similar properties to one of the larger of the vD08 galaxies. This turns out to be the galaxy whose observed velocity dispersion is inconsistent with its being genuinely compact (Section 3.3). Where we would predict $\Delta \log \sigma = 0.24$ dex, or $\sigma = 310 \pm 70$ km/s, what we observe is $\Delta \log \sigma = -0.17$ dex and $\sigma = 129 \pm 14$ km/s. This is also the galaxy with the largest difference between the Sérsic– and De Vaucouleurs–sizes ($\log R_{\mathrm{Dev}}/R_{\mathrm{Ser}} = -0.34$; see Section 3.2). Adopting the NYU VAGC Sérsic size measurement is not sufficient to reconcile the observed size and mass with the velocity dispersion: the observed velocity dispersion would still be too small by $\sim 0.2$ dex. This galaxy also sits nearly 0.25 dex above the median color–mass-to-light relation shown in Figure 1; using the Bell & de Jong (2001) prescription for $M_*/L$ as a function of $(B - V)$ leads to a stellar mass estimate that is 0.17 dex lower. Adopting both this mass estimate and the NYU VAGC size estimate, we do find consistency between $\Delta \log R_e$ and $\Delta \log \sigma$. In this sense, this galaxy is by far the weakest of our compact galaxy candidates — it seems to have had its size underestimated and/or its mass overestimated.

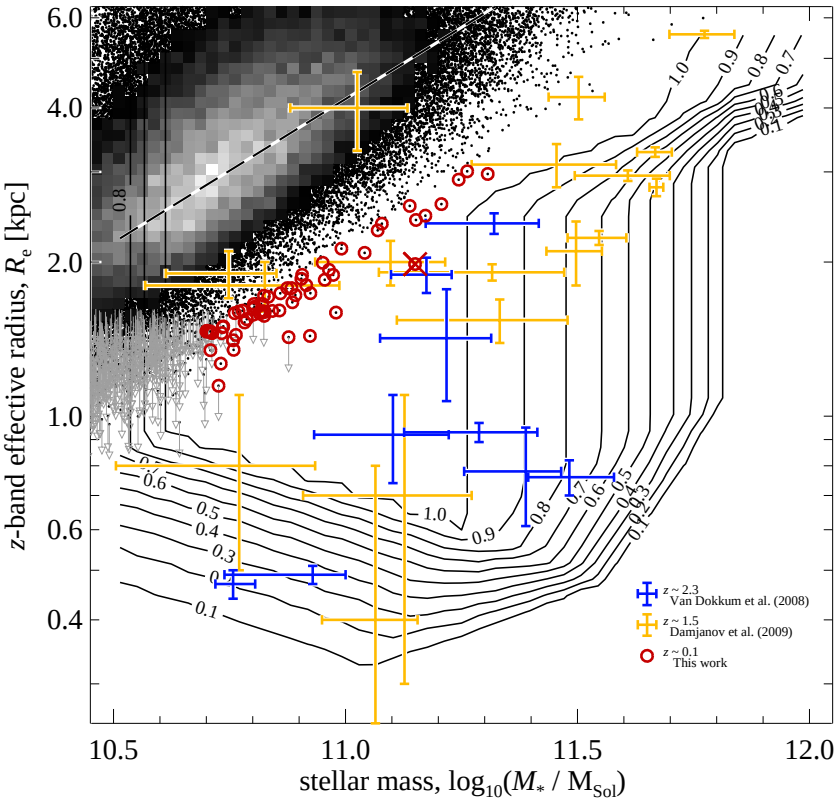


**Figure 6.** — The size–mass relation for massive, red sequence galaxies at low and high redshifts. — As in Figure 2, the points and circles are for SDSS galaxies; those galaxies that we have rejected as described in Section 3 are not shown. The contours show the relative volume completeness of the SDSS spectroscopic sample for $0.066 < z < 0.12$, as marked. The orange points with error bars are the D09 sample of $1.2 < z < 2.0$ galaxies from the GDDS and MUNICS. The blue points with error bars are the vD08 sample of $z \sim 2.3$ galaxies from MUSYC. While 3/10 of the $z \sim 1.5$ galaxies are comparable to local galaxies, there are no red sequence galaxies in the local universe with sizes and masses comparable to the compact galaxies found at $z \sim 2.3$. This lack cannot be explained by selection effects: the minimum SDSS completeness for the vD08 and D09 galaxies placed at $0.066 < z0.12$ is 20–40 %; the average completeness is 80 %. If the vD08 galaxies were to be passively evolved into the local universe, we would expect to detect on the order of $\sim 6500$ such galaxies.

We also stress that the observed velocity dispersions of the candidates that lie furthest from the main size–mass relation suggest that these galaxies have had their sizes significantly underestimated (see Section 3.3).

If the vD08 galaxies were placed at $0.066 < z < 0.12$, the SDSS spectroscopic completeness would typically be $\gtrsim$ 75 %. Note, however, that there are two $R_{\rm e} < 0.5$ kpc galaxies from the vD08 sample and one from the D09 sample for which the SDSS completeness is just 20–40 %. If the Kriek et al. (2008a)/vD08 galaxies were not to evolve in either size or number density from $z \sim 2$ to the present day, we would expect there to be $\sim 6500$ $M_* > 10^{11}$ $\mathrm{M}_\odot$ galaxies with $\Delta \log R_{\rm e} < -0.4$ dex at $0.066 < z < 0.12$, of which $\sim 5250$ should appear in the SDSS spectroscopic sample. Instead, we have only one (very weak) candidate: this galaxy appears to have large errors in its size and mass measurements.

Moreover, we stress that those galaxies which we have identified as 'compact' are not qualitatively similar to the compact galaxies found at higher redshifts, which are offset from the local size–mass relation by at least twice as much again as our local compact galaxy candidates.

As an interim conclusion, then, we have shown that there are no galaxies in the local universe (at least as probed by the SDSS spectroscopic sample) that are directly analogous to the compact galaxies found at high redshift. This dearth of compact galaxies cannot be explained by selection effects. In Appendix A, we confirm and extend this conclusion by searching for compact galaxy candidates from within the SDSS photometric sample, on the basis of photometric redshifts.

### 4.3 The Number Density of Massive, Compact Galaxies

In **Figure 7**, we provide a more quantitative statement of our conclusion with respect to the size evolution of massive galaxies from $z \sim 2$ to $z \sim 0.1$ by plotting the size distribution for massive, red galaxies in different mass bins. In this figure, the filled histograms represent the main SDSS spectroscopic sample described above. For comparison, the horizontal-hatched histograms show the situation at $z \sim 1.6$, based on the ten D09 galaxies drawn from the GDDS; similarly, the diagonal-hatched histograms show the nine $z \sim 2.3$ Kriek et al. (2008a) galaxies with sizes from vD08.

The Kriek et al. (2008a)/vD08 sample is representative, but not complete. In order to derive the densities shown in Figure 7, we have scaled each of the vD08 galaxies as follows. First, we have normalized the distribution to have a density of $1.5 \times 10^{-4}$ $\mathrm{Mpc}^{-3}$; this corresponds to the total number density of all $2 < z < 3$ galaxies to the mass limit of Kriek et al. (2008a), derived from the mass function fit given by Marchesini et al. (2008). We have then scaled this distribution by a factor of 16/36 to count only those galaxies with little or no ongoing star formation from Kriek et al. (2008a) that seem to form a red sequence (Kriek et al., 2008b). For the D09 sample, we are able to use $1/V_{\rm max}$ scalings from Glazebrook et al. (2004).

The location of each individual high-redshift galaxy is marked in Figure 7 with an arrow: the slightly lower blue arrows show the vD08 galaxies; the slightly higher yellow arrows are for the D09 galaxies. Clearly, given the small numbers,

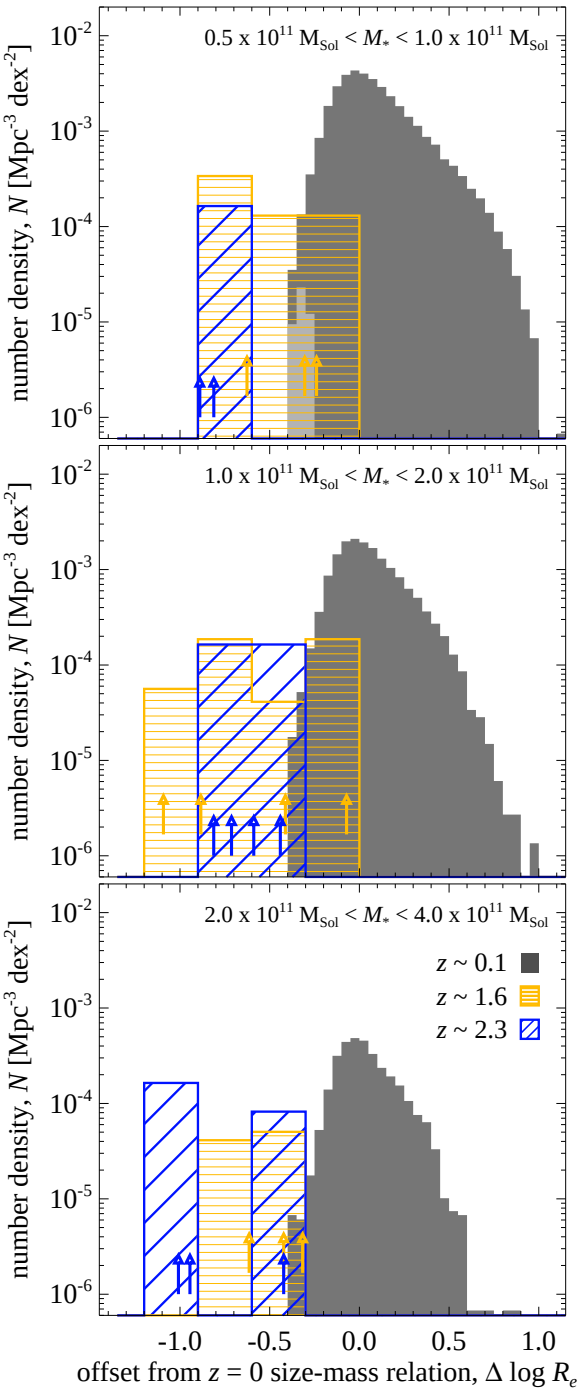


**Figure 7.** — The observed size distribution of massive, red galaxies at $z \sim 0.1$, $z \sim 1.6$, and $z \sim 2.3$. — Each panel of this Figure is for a different mass range, as marked. Within each panel, the solid histogram represents the SDSS spectroscopic sample. The blue, diagonally-hatched histogram represents the vD08 sample of nine massive, passive galaxies at $z \sim 2.3$; the yellow, horizontally-hatched histogram represents the 10 $z \sim 1.6$ GDDS galaxies from D09. The arrows at the bottom of each panel indicate the positions of the individual high redshift galaxies. The $z \sim 2.3$ galaxies have to undergo significant structural evolution over $z \lesssim 2.3$ to match the properties of local universe galaxies; at least part of this evolution has already occurred by $z \sim 1.6$.

the uncertainties on these high redshift values are quite large; even so, they do provide a useful order of magnitude estimate for comparison to the local values.

The clear implication from the comparison between the $z \sim 0.1$ and $z \sim 2.3$ data in Figure 7 is that — in line with the conclusions of vD08 — not one of the vD08 galaxies is consistent with the properties of the $z \sim 0$ galaxy population. With the results we have now presented, we can extend this conclusion by confirming that this discrepancy cannot be explained by selection effects in the low redshift sample. There are local analogs for less than half of the $z \sim 1.6$ galaxies, albeit with considerably higher number densities. This would imply that at least some ($\lesssim 50$ %) of the $z \lesssim 2.3$ evolution has already occurred by $z \sim 1.6$.

## 5 Discussion

### 5.1 Compact Galaxy Properties

In **Figure 8**, we plot the properties of our compact galaxies in comparison to the general massive, red galaxy population. In each panel, the circles highlight our compact galaxies, while the points and greyscale show all galaxies with $M_* > 10^{10.7}$ M$_\odot$, $^{0.1}(u-r) > 2.5$, and $0.066 < z < 0.12$. The large grey boxes with error bars show the mean and standard deviation of each plotted property in quintiles of the velocity dispersion distribution. Similarly, the red boxes with error bars show the mean and standard deviations for our compact galaxy candidates in two bins, separated at $\sigma = 200$ km/s; the median for this sample.

In each of the panels of Figure 8 (from left to right), we show the equivalent width of the H$\delta$ line (where negative values imply emission), the luminosity-weighted mean stellar metallicity, and the luminosity weighted mean stellar age, as derived by Gallazzi et al. (2005). Because these estimates are available only for DR4, only around half of our compact candidates, and only 3/10 of those with $M_* > 10^{11}$ M$_\odot$, can be plotted in this Figure.

We have also matched our compact galaxy sample to the AGN sample described by Kauffmann et al. (2003b), for SDSS DR4. These AGN hosts have been selected by their [OIII]/H$\beta$ and [NII]/H$\alpha$ emission line ratios; *i.e.* the Baldwin, Phillips & Terlevich (1981, BPT) diagram. 34 of our 63 galaxies appear in the DR4 catalog; of these, 11 are classified as AGN on the basis of their emission line ratios. This is slightly higher than the AGN fraction of the parent sample, which is in the range 20–26 % for the mass range we are considering. Of the 11 galaxies identified as AGN hosts, five sit on or slightly above the main $M_*$–$L_{[\rm OIII]}$ relation, with $L_{[\rm OIII]} \approx 10^6$ L$_\odot$, four have $L_{[\rm OIII]} \sim 10^{7-8}$ L$_\odot$, and one is quite high luminosity, with $L_{[\rm OIII]} = 10^{8.7}$ L$_\odot$. These 11 galaxies are marked in each panel of Figure 8 with a small blue cross.

Kauffmann et al. (2003b) also provide revised stellar mass and velocity dispersion measurements for these AGN-host galaxies. Accounting for the presence of an AGN does not have a major impact on these measurements: the masses and velocity dispersions change at the level of 0.05 dex and 16 km/s, respectively. That is, while it is possible that an optically-bright point source may bias the

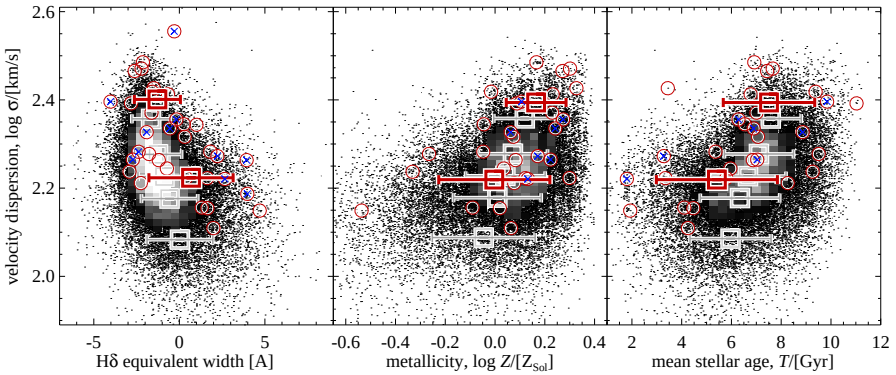


**Figure 8.** — The properties of our compact galaxy candidates in comparison to the general population of massive, red sequence galaxies. — As in other Figures, the black points and greyscale show the data density of all galaxies with $M_* > 10^{10.7}$ $M_\odot$, $^{0.1}(u-r) > 2.5$, and $0.066 < z < 0.12$; the red circles show our compact galaxy candidates. The grey boxes with error bars show the mean and rms scatter in each quantity for quintiles of the velocity dispersion distribution; the red boxes show the same for those of our compact galaxy candidates with $\sigma < 200$ km/s and $\sigma > 200$ km/s separately. At fixed velocity dispersion, our compact galaxies have slightly lower than average mean ages and slightly higher metallicities — however, this result is only significant at the $2\sigma$ level.

measured sizes of these galaxies downwards, within the stated errors, the AGN does not significantly affect the derived values of $M_*$ or $\sigma$. (It is relevant here that only one of our compact galaxy candidates shows a significant residual point-source after subtracting off the best-fit Sérsic profile, as produced by GALFIT; see Section 3.2.)

Looking now at Figure 8, it is clear that the majority of our compact galaxy candidates have quite old stellar populations. For the $\sigma > 200$ km/s bin, all but one of our candidates have $T > 6$ Gyr. For the $\sigma < 200$ km/s galaxies, the median age is 6 Gyr (although the ages do range from 2 to 10 Gyr). Among the lower velocity dispersion candidates, there is a clear tendency towards relatively high equivalent widths for H$\delta$ absorption, suggestive of a relatively recent ($\lesssim 2$ Gyr) star formation event. That is, at fixed velocity dispersion, our compact galaxy candidates appear to have slightly higher metallicities, and be slightly younger than average.

That said, using bootstrap-resampling on similar sized samples drawn randomly from the mass-limited sample, and controlling for velocity dispersion, these results are weakly significant at best: $1.9\sigma$ for the age, and $< 1\sigma$ for metallicity. Considering only the higher velocity dispersion candidates ($\sigma > 200$ km/s), the significance of these differences become $1.9\sigma$ and $2.2\sigma$ for the age and metallicity offsets, respectively. This weakly significant result should be contrasted with the results of Shankar & Bernardi (2009) and van der Wel et al. (2009), who find that, on average and at fixed dynamical mass, early type galaxies with higher velocity dispersions (or, equivalently, smaller sizes) have older mean stellar ages.

While the younger mean stellar ages and lower metallicities of our compact galaxy sample are only weakly significant, both would imply a relatively late start to star formation for these galaxies and/or their progenitors. But if these galaxies grow in size through mergers (for example) then it is possible that these galaxies are small not because their formation is delayed relative to other galaxies of the same mass or velocity dispersion, but rather because they have had fewer mergers overall, or perhaps just fewer *recent* mergers. That is, it may be that, at fixed mass, these compact galaxies are in fact *older*, in the sense that they have been *assembled earlier*, and existed in (more or less) their present form for longer than other galaxies of the same mass or velocity dispersion.

### 5.2 Comparison to Other Recent Works

In a similar study to this, using sizes and photometry from the NYU VAGC for SDSS DR6, Trujillo et al. (2009, hereafter T09) have recently reported the detection of 28 $z < 0.2$ compact galaxies. These galaxies were selected to have $M_* > 8 \times 10^{10}$ $M_\odot$ and $R_e < 1.5$ kpc. To the same size and mass limits, we find just one galaxy in our $0.066 < z < 0.12$ red sequence sample. T his implies a difference in volume density of a factor of 5.5. Whereas we are primarily interested in passive (red sequence) galaxies, T09 did not preselect their compact galaxy candidates by color. If we expand our sample to include blue galaxies, after visual inspection of an additional 34 objects, we are left with only 7 plausible blue compact galaxy candidates. None of these galaxies have masses greater than $8 \times 10^{10}$ $M_\odot$. On the other hand, applying our ${}^{0.1}(u-r) > 2.5$ criterion to the T09 sample, only 30 % (9/28) of these galaxies are red. In other words, to the same limits, our results are in good agreement with those of T09 for red galaxies: based on the single galaxy in our sample that satisfies the T09 size–mass selection, our inferred number densities agree to within 65 %. At the same time, however, T09 find a number of blue compact galaxies which we do not.

In an attempt to better understand the cause of this discrepancy, we have compared the $r$ band `model` effective radii from the basic SDSS catalog to the $r$ band Sérsic-fit sizes from the NYU VAGC that are used by T09. For all four of the galaxies in the T09 sample that lie at $z < 0.14$, the agreement is better than 25 %. For the galaxies at $z > 0.14$, however, the agreement becomes progressively worse; the median NYU VAGC size is approximately 40 % larger than that given in the basic SDSS catalog. The random scatter between the two size measurements does not obviously depend on the observed sizes, but does appear to be significantly higher for $m_r \gtrsim 17$; all of the $z < 0.14$ T09 galaxies have $m_r < 17$. This suggests that the T09 size measurements may be affected by larger random or systematic size measurement errors at higher redshifts, due to lower signal-to-noise.

We have also compared our mass estimates for the T09 sample to those given in the NYU VAGC and used by T09. We find systematically lower masses for the galaxies in the T09 galaxies, at the level of 0.05–0.10 dex. Interestingly, all but one of the T09 galaxies lie below the median color–$M/L$ relation shown in Figure 1. (The exception lies 0.16 dex above it.)

There are thus significant systematic differences between both the size and mass estimates used here, and those used by T09. Without followup observations, however, it is not possible to make any further conclusions — we are, after all, analyzing the same data. That said, we stress the fact that the observed velocity dispersions for our candidates with $\Delta \log R_{\rm e} < -0.4$ dex imply that their very small inferred sizes are produced by large errors in the measured sizes (see Section 3.3). In contrast to this, T09 find a median velocity dispersion which is only 0.04 dex higher than their control sample, even though the mean size and mass are offset from the Shen et al. (2003) relation by $-0.5$ dex. This discrepancy can only be explained by either very large structural differences, or if the T09 sample is disproportionally affected by large measurement errors in size and/or mass. Spatially resolved dynamical studies of the two samples may offer a practical means of resolving this issue.

While this discrepancy in the numbers of blue compact galaxies remains a concern, it has no impact on our conclusions. As well as being blue, the T09 galaxies have mean stellar ages of $\lesssim 2$ Gyr; that is, their *current* ages are comparable to the ages of the vD08 galaxies at $z \gtrsim 2$. In this sense, the T09 sample may represent close analogues to the vD08 and D09 galaxies, but they cannot be the passively evolved counterparts to the galaxies seen at $z \gtrsim 2$ that we are interested here.

Even more recently, Valentinuzzi et al. (2009, hereafter V09) have described a sample of 122 compact galaxies selected from the WIde-field Nearby Galaxy-cluster Survey (WINGS) of X-ray selected clusters at $0.04 < z < 0.07$. Unlike in this work, V09 do find galaxies with properties comparable to the 3 largest vD08 galaxies; similarly, there are local WINGS analogs for 8 of the 10 GDDS galaxies from D09. However, this result relies on their scaling the high redshift galaxies' masses down by 0.15 dex to account for the poor treatment of NIR-luminous thermally pulsating asymptotic giant branch (TPAGB) stars in the Bruzual & Charlot (2003) stellar population models. While both Kriek et al. (2009) and Muzzin et al. (2009) show that the stellar masses for the $z \sim 2.3$ implied by different models vary by $\sim 0.1$ dex, we have not applied such a correction here.

That said, however, we stress that the high- and low-redshift samples have been treated consistently here, including the fact that all masses were derived from the restframe optical. Further, we note that van der Wel et al. (2006) have shown that stellar masses derived from the restframe optical and using Bruzual & Charlot (2003) models are consistent with the dynamical masses of $z < 1$ galaxies, and are unaffected by the TPAGB uncertainties. Further, we have verified that the vD08 galaxies follow the same $M_*/L$–color relation as our SDSS sample.[5]

The V09 compact galaxy sample is selected by effective surface mass density, $\Sigma_{\rm e} = M_*/2\pi R_{\rm e}^2 > 3 \times 10^9$ M$_\odot$ kpc$^{-2}$, in the mass range 3–50 $\times 10^{10}$ M$_\odot$. (Our compact galaxy selection is roughly equivalent to $\Sigma_{\rm e} \gtrsim 3.6 \times 10^9$ M$_\odot$ kpc$^{-2}$.) For their sample, V09 derive a number density of $1.31 \times 10^{-5}$ Mpc$^{-3}$; to our mass limit of $M_* > 10^{10.7}$ M$_\odot$, this value becomes $0.94 \times 10^{-5}$ Mpc$^{-3}$. No volume

[5]Of course, this is no guarantee that the stellar masses for either sample are accurate; all this says is that they have been derived self-consistently.

corrections have been applied to these values — in other words, they assume that no such galaxies exist outside of the clusters observed by WINGS. In this sense, they are concrete lower limits.

For our sample, however, the number density of $M_* > 10^{10.7}$ $M_\odot$ galaxies with $\Sigma_e > 3\times10^9$ $M_\odot$ kpc$^{-3}$ is just $0.23\times10^{-5}$ Mpc$^{-3}$. That is, after correcting as best we can for the different stellar mass limits, our number densities are inconsistent by a factor of more than 4 with the hard lower limit given by V09. Using a more extreme density criterion of $\Sigma_e > 4 \times 10^9$ $M_\odot$ kpc$^{-2}$, this discrepancy grows to a factor of 16. Again, our use of $z$ band effective radii leads to smaller measured sizes than for bluer bands; this discrepancy would only increase using $r$- or $g$-band measured sizes. There is an offset of $\sim 0.1$ dex between the V09 stellar mass estimates and the ones that we use here, but this can explain at most 33 % of the discrepancy. This leaves two possibilities: either our results are badly affected by unexplained selection effects, or there are large discrepancies between our size estimates and those of V09.

We have considered possible spectroscopic selection effects that could produce a bias against bright, compact objects in Section 4.1, and shown these to be relevant for $z \lesssim 0.05$. These effects may well explain why V09 were able to match only a small fraction of their compact galaxies (which have $0.04 < z < 0.07$) to objects in the (DR4) SDSS spectroscopic catalog. We have shown, however, that our $0.065 < z < 0.12$ results are not strongly affected by these kinds of selection effects (Figure 6, see also Appendix A). Our estimated completeness is more than 60 % for all galaxies in the V09 sample and greater than 90 % for 90 % of the sample. The selection effects considered in Section 4.1 thus cannot explain the difference in our inferred number densities.

An alternative explanation is that the V09 galaxies only exist in rich clusters, and that SDSS suffers much higher spectroscopic incompleteness in such dense fields. The most obvious selection effect that might become important in dense environments is due to fiber collisions, but it is also possible that the selection against galaxies that the selection against blended/segmented objects (see Section 4) may play a role. However, for this explanation to work (assuming that this selection did not act preferentially against compact galaxies) SDSS would have to be less than 25 % complete for the kinds of clusters targeted by WINGS. Visual inspection of the SDSS spectroscopic coverage of several WINGS clusters (using the SkyServer web application) suggests that this level of incompleteness is implausible, but not necessarily impossible. Again, if galaxies like those selected by V09 exist outside of X-ray bright clusters, this would imply even greater SDSS incompleteness.

We have also looked at the SDSS data for the Coma cluster, to explore the idea that compact galaxies live in rich clusters. Starting from the photometric catalog, we have selected photometrically-classified galaxies with $m_{\mathrm{Pet},r} < 17.88$ within 1.5 degrees of NGC 4899 (RA = $3^{\mathrm{h}}02^{\mathrm{m}}0^{\mathrm{s}}.6$, Dec = $27°58'32''.3$). Using the SkyServer web application, we have confirmed that the SDSS spectroscopic completeness is high, even near the cluster center. More quantitatively, selecting the cluster red sequence out of the $u$–$r$ observed color–magnitude diagram, 94.5 % (636/673) of galaxies are spectroscopic targets, and 89 % (599/673) have

`sciencePrimary` spectra. After visual inspection to exclude stars, if we then place those objects without spectra at the cluster redshift and use Equation 1 to estimate stellar masses, we have 6 $M_* > 10^{10.5}$ $M_\odot$ candidate cluster red sequence galaxies without spectroscopic redshifts, compared to 72 spectroscopically confirmed ones. None of these galaxies, with or without spectroscopic redshifts, have $\Sigma_e > 3 \times 10^9$ $M_\odot$. In other words, we see no evidence of galaxies like those found by V09 in the Coma cluster. Again, this lack cannot be explained by selection effects: the SDSS spectroscopic completeness for massive red sequence galaxies in Coma is at least 92.5 %.[6] We also note that several studies (e.g. Kauffmann et al., 2004; Park et al., 2007; Weinmann et al., 2009) have found little or no evidence for an environmental dependence of the size–mass relation within SDSS.

A completely independent estimate can be obtained using the Faber et al. (1989) sample of early type galaxies: we find that 5/319 of these galaxies have sizes that are smaller than the mass-size relation by a factor of 2 or more. This fraction is approximately 5 times higher than what we find for the SDSS. While not conclusive, this does suggest that SDSS may suffer from additional incompleteness or biases beyond what we have considered here.

Thus we can find no easy explanation for the difference between the V09 results and our own. Despite these differences, we note that V09 conclude that — barring large systematic errors in the high-redshift measurements — at least 65 % (*cf.* our value of 100%) of the $z \sim 2.3$ galaxies from vD08 and at least 20 % (*cf.* our value of 60 %) of the $z \sim 1.6$ galaxies from D09 have disappeared from the local universe. Accepting the high-redshift results, these galaxies simply cannot evolve passively and statically into the red sequence and/or early type galaxies found in the local universe: a significant number must still undergo substantial size evolution. Early type galaxies do not (all) form monolithically.

## 6 Summary and Conclusions

The central question of this work has been the existence or otherwise of massive, compact, quiescent and/or early type galaxies in the local universe, and particularly the importance of selection effects in the SDSS spectroscopic sample for such galaxies. We have shown that, especially for lower redshifts ($z \lesssim 0.05$), galaxies with the masses and sizes of those found at $z \gtrsim 2$ would not be targeted for spectroscopic followup (Figure 2). The main reason for this is not the star/galaxy separation criterion, but rather the exclusion of bright and compact targets in order to avoid saturation and cross-talk in the spectrograph (see Section 4.1).

Bearing this in mind, we have conducted a search for massive, compact galaxies at $0.066 < z < 0.12$; we estimate that, in this redshift interval, the average completeness for galaxies like those from vD08 and D09 would be $\gtrsim 20$ % in the worst case, and $\sim 80$ % on average (Figure 6).

[6]Note that, being at low redshift, Coma galaxies are well separated on the sky, and so fiber collisions are less of an issue than for higher redshifts. Further, while the high-surface brightness selection effects discussed in Section 4 are, in principle, a concern, we have verified that all six of these candidate cluster members would satisfy these high surface brightness criteria.

Starting from a sample of massive ($M* > 10^{10.7}$ $M_\odot$) red sequence ($^{0.1}(u-r) > 2.5$) galaxies, we have selected the 280 galaxies with inferred sizes that are a factor of 2 or more smaller than would be expected from the Shen et al. (2003) $M_*$–$R_e$ relation for early type galaxies. In order to confirm their photometry and size measurements, we have visually inspected all of these objects. Unsurprisingly, by selecting the most extreme outliers, a large fraction of these objects ($\sim 70\%$) appear to be instances where the size and/or stellar mass estimates are unreliable (Section 3.1).

For the 63 galaxies with no obvious reason to suspect their size or stellar mass estimates, there is good agreement between the default SDSS size measurement (based on the 2D light distribution, using a sector-fitting algorithm, and assuming a de Vaucouleurs profile), and those given in the NYU VAGC (based on the azimuthally average growth curve, assuming a more general Sérsic profile). However, particularly for galaxies with high $n$, the de Vaucouleurs size measurement is systematically *smaller* than the Sérsic one, at the level of $\lesssim 25$ % (Section 3.2).

In general, and as expected, our 63 compact galaxy candidates have significantly higher than average velocity dispersions (Figure 5). While it remains possible that the sizes of at least some of our compact galaxy candidates may have had their sizes underestimated by $\sim 30$ %, in general, the relatively high observed velocity dispersions support the notion that they are indeed unusually compact given their stellar masses (Section 3.3).

Among our compact galaxy candidates, there are no galaxies with sizes comparable to those found $z \sim 2.3$ by vD08; we find analogs for $\lesssim 50$ % of the D09 galaxies at $z \sim 1.6$ (Figures 6 and 7). This lack cannot be explained by selection effects: if the vD08 galaxies were passively evolved into the local universe, we would expect to detect on the order of $\gtrsim 5000$ galaxies in this region of the size–mass diagram.

To test this conclusion, in Appendix A, we have also used photometric redshifts to construct the size–mass diagram for the full photometric sample, and compared this to the size–mass diagram for the spectroscopic subsample, also analyzed using photometric redshifts. In this way, we can also make an estimate for the number of massive, compact galaxies that are missing from the spectroscopic sample. While it is conceivable that SDSS is missing (at most) a handful of massive, compact galaxies, there are again no signs of galaxies comparable to those of vD08 or D09.

It is not impossible that some systematic errors in the estimation of $M_*/Ls$ for the high redshift galaxies (e.g., an evolving stellar IMF) mean that their stellar masses are vastly overestimated, however it would require an overestimate of $\gtrsim 0.7$ dex to reconcile the vD08 galaxies with the sizes of the smallest galaxies we have identified in the SDSS catalog. Further, we stress that we have verified that the masses of the $z \sim 0$ and $z \sim 2.3$ galaxies have been derived consistently. By the same token, there may be systematic errors in the measured sizes of the high- and low-z galaxies. vD08 point out that complex morphologies, and in particular the possibility of low surface brightness wings, could mean that the sizes of the $z \sim 2.3$ galaxies are underestimated. We point out, however, that all indications

are that the sizes that we have used here are *underestimates*; using different size measures for the low-z galaxies would only *increase* the discrepancy between the high- and low-redshift samples, so strengthening our conclusions.

Accepting the high redshift observations at face value, then, our results confirm that massive galaxies, both individually and as a population, must undergo considerable structural evolution over the interval $z \lesssim 2.3$ in order to develop into the kinds of galaxies seen locally — even after star formation in these galaxies has effectively ended. We see some hints that a significant amount of this evolution ($\lesssim 50$ %) may have already occurred by $z \sim 1.6$.

The fact that each and every one of the vD08 galaxies must undergo significant structural evolution to match the properties of present-day galaxies implies that the mechanism that drives this growth must apply more or less evenly to all galaxies. To see this, let us assume that some external process drives the size evolution of these galaxies, and that even a single event is sufficient to move an individual galaxy onto the main size–mass relation. Then, we can assume some simple probability distribution for the number of events, $N$, among individual galaxies. (For example, we could assume that events occur randomly across the time interval $z < 2.3$, or that each galaxy experiences $N \pm \sqrt{N}$ events.) Now, our results suggest that the number density of vD08–like galaxies drops by at least a factor of 5000 since $z \sim 2.3$. In order to ensure that at most 1/5000 galaxies have $N = 0$ after $z \sim 2.3$, simple probabilistic arguments imply that the average galaxy must undergo $\gtrsim 20$ events. This would imply that a strongly stochastic process like major mergers cannot be the primary mechanism for the strong size evolution of massive galaxies.

Apart from their small sizes and high velocity dispersions, our compact galaxy candidates are not obviously distinct from the general population (Figure 8). If anything, at fixed velocity dispersion, our compact galaxies have stellar populations that are slightly younger than average (at $\sim 1.9\sigma$ significance). Even so, the majority of these galaxies' stellar populations are definitely 'old', with luminosity-weighted mean stellar ages typically in the range 6–10 Gyr. But if some external mechanism drives the size evolution of these galaxies, we speculate that their small sizes may indicate that they have assumed their present form comparatively *early*, and in this sense they may actually be relatively *old* (see also, e.g., van der Wel et al., 2009). If so, with better understanding of the processes that determine the sizes of early type galaxies, and in particular the role of merging, the properties of these galaxies could provide a means of constraining the evolution of massive galaxies after they have completed their star formation, including their late-time merger histories.

**Acknowledgments.** We wish to thank both Tiziano Valentinuzzi and Ignacio Trujillo for helpful discussions in trying to reconcile our results. This work was supported through grants by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), the Leids Kerkhoven-Bosscha Fonds (LKBF).

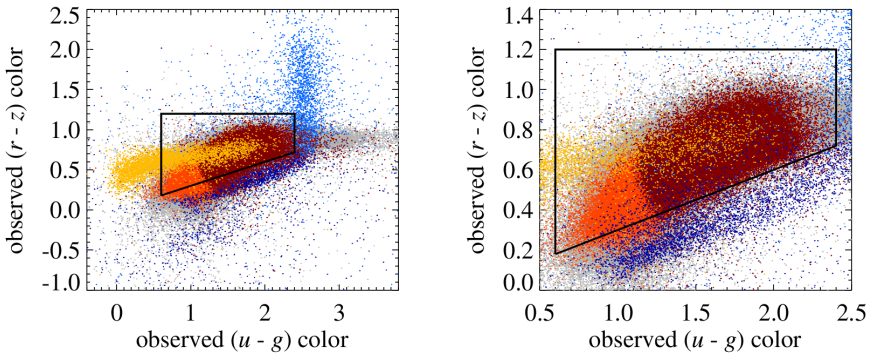


**Figure A.1.** — Selecting $z \lesssim 0.1$ galaxies based on color alone. — Each panel shows the observed $(u-g)$–$(r-z)$ color–color plot for objects in the SDSS spectroscopic sample; the right panel simply shows the central region in greater detail. Points are color–coded according to their spectral classification, *viz.*: galaxies (*grey*), galaxies with $0.066 < z < 0.10$ (*light red*), galaxies with $0.066 < z < 0.10$ and ${}^{0.1}(u-r) > 2.5$ (*dark red*), quasars (*yellow*), late-type stars (*light blue*), and ordinary stars (*dark blue*), . The box shows the color selection that we use to select $z \lesssim 0.1$ galaxies. This selection should produce a reasonably complete sample of $z \lesssim 0.1$ galaxies, with some contamination from both stars and quasars. In particular, later-type stars and some quasars have observed SEDs that are very similar to red sequence galaxies at $z \sim 0.1$.

# Appendix

## A Looking for Massive Compact Galaxies in the SDSS Photometric Sample

In this Appendix, we present a complementary analysis in which we directly compare the spectroscopic and photometric SDSS samples, in order to test our conclusion that the apparent lack of massive, compact galaxies in the local universe cannot be explained by incompleteness in the spectroscopic sample.

### A.1 Selecting Galaxies by Color Alone

Before we can address the question of massive compact galaxies in the SDSS photometric sample, we must first devise a means of separating stars and galaxies without selecting on the basis of observed size or light profile. Our method for doing so is shown in **Figure A.1**, which plots the observed (extinction-corrected) $ugrz$ colors of different classes of objects from the spectroscopic sample; we show: all galaxies (grey), $0.066 < z < 0.12$ galaxies (bright red), and those with ${}^{0.1}(u-r) > 2.5$ (dark red), O–K stars (dark blue), M-type or later stars (light blue), and quasars (yellow).

The black box shown in Figure A.1 shows our criteria for selecting $0.066 < z < 0.12$ galaxies based on their $ugrz$ colors:

$$0.6 < (u-g) < 2.4 \quad \text{and} \quad 0.3 \times (u-g) < (r-z) < 1.2 \; . \tag{A.1}$$

Again, we apply this selection in terms of `model` colors. Note how, whereas the stellar sequence is reasonably well separated from the region of color space occupied by galaxies for $(u-g) \lesssim 2.5$, beyond this point, the late-type stellar sequence turns up, such that late-type stars and galaxies are blended. In the most general terms possible, the mean galaxy redshift increases towards redder $(u-g)$ colors. This means that our ability to distinguish red galaxies from late-type stars on the basis of their optical SEDs is limited to $z \lesssim 0.12$.

In the right-hand panel of Figure A.1, we zoom in on this selection region. From this panel, it is clear that a large proportion of quasars will also be included in our color-selected 'galaxy' sample. Similarly, it is clear that this color selection is not 100 % efficient in excluding stars from our sample: more quantitatively, with this selection we are able to exclude more than 80 % of spectroscopically identified stars that are given $0.066 < z_{\rm phot} < 0.12$, while retaining more than 97 % of all spectroscopically targeted and confirm galaxies that are given $0.066 < z_{\rm phot} < 0.12$. Furthermore, it should be remembered that stars are already heavily selected against for the spectroscopic sample plotted in Figure A.1; the relative number of stellar 'contaminants' may well be considerably higher for the photometric sample.

### A.2 Photometric Redshifts and Stellar Mass Estimates

A major improvement in DR7 is a complete revision in how the basic (`photoz`) photometric redshifts are derived (Abazajian et al., 2009). Rather than using some combination of synthetic template spectra to fit galaxies' SEDs, the new `photoz` algorithm directly compares the observed colors of individual galaxies to those of a reference sample of galaxies with spectroscopic redshifts. Specifically, for each individual object, the algorithm finds the 100 closest neighbors in *ugriz* color space, and fits a hyper-plane to these points, rejecting outliers; the `photoz` is then determined by interpolating along this 4D surface. In comparison to the DR6 algorithm, this reduces the RMS redshift error by more than 75 % ($\langle \Delta z \rangle = 0.025$), and significantly reduces systematic errors (Abazajian et al., 2009).

For this exercise, rather than derive SED-fit stellar mass estimates assuming the photometric redshifts, we will simply use the empirical relation between ${}^{0.1}(g-i)$ color and $M_*/L$ (Equation 1). In this way, we are able to recover the $z_{\rm spec}$-derived, SED-fit $M_*/L$s of the sample of galaxies shown in Figure 1 to 0.045 dex ($1\sigma$); including the effects of photometric redshift errors, k-corrections, and $M_*/L$ errors, the total ($1\sigma$) error in $M_*$ is 0.13 dex. This should be compared to the median formal error on the original SED-fit stellar mass estimates (0.10 dex). That is, the errors on $M_*$ based on photometric masses (adding these two errors in quadrature) are only about 60 % greater than those based on spectroscopic redshifts.

### A.3 The Size Distribution of Massive, Red Sequence Galaxies

In **Figure A.2**, we show three size–mass diagrams corresponding to, from top to bottom: (*a.*) the spectroscopic sample, analyzed using spectroscopic redshifts; (*b.*) the spectroscopic sample, analyzed using photometric redshifts; and (*c.*) the photometric sample, analyzed using photometric redshifts. In all three cases, the only

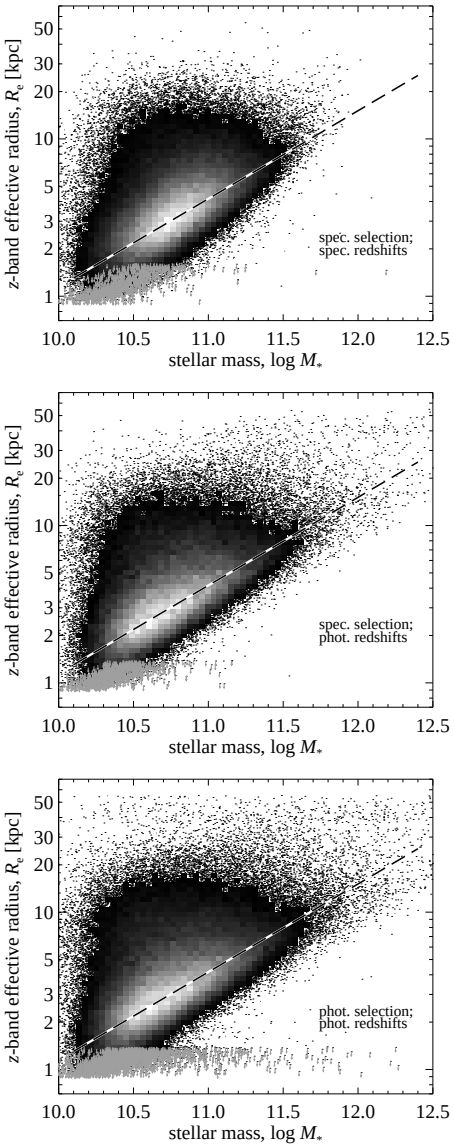


**Figure A.2.** — The size–mass plot for massive, red sequence galaxies at $0.066 < z < 0.12$, based on the spectroscopic and the full photometric SDSS catalogs. — Each panel shows the sizes and masses of galaxies based on, from top to bottom, the spectroscopic sample using spectroscopic redshifts, the spectroscopic sample using photometric redshifts, and the photometric sample using photometric redshifts; in each case, only those objects inferred to have ${}^{0.1}(u-r) > 2.5$, and $0.066 < z < 0.10$ are shown. In panel 3, many more objects with inferred sizes $\lesssim 0.3$ kpc can be seen; these are largely stars misclassified (in terms of their photometric redshifts) as galaxies. For $M_* \lesssim 10^{10.8}$ $M_\odot$ and $R_e \lesssim 10^{-0.2}$ kpc, comparison between panels 2 and 3 suggest that there may be a few additional galaxies in the photometric sample that do not appear in the spectroscopic sample.

selections applied to each sample are on photometric `type` (to exclude optical artifacts, etc., we require either a `star` or `galaxy` `type` classification) and $ugrz$ color (to exclude stars); then, as in Figures 2 and 6, we are only showing those galaxies inferred to have $0.066 < z < 0.12$ and ${}^{0.1}(u-r) > 2.5$. Again, objects with measured sizes smaller than $0\farcs75$ are shown as upper limits, assuming a size of $0\farcs75$. When comparing these three different analyses, the difference between (*a.*) and (*b.*) shows the effect of using spectroscopic versus photometric redshifts, and the difference between (*b.*) and (*c.*) shows the difference between the SDSS spectroscopic and photometric selection. That is, the comparison between (*b.*) and (*c.*) gives a direct indication of the level of incompleteness in the spectroscopic sample.

Looking at panels (*a.*) and (*b.*), it is clear that the use of photometric redshifts produces a considerably greater scatter in the size–mass diagram, including a rather large number of galaxies with inferred stellar masses of $10^{12}$ $M_\odot$ or greater. There is a clear excess of unresolved objects with inferred stellar masses greater than $\sim 10^{11}$ $M_\odot$ in panel (*b.*) in comparison to panel (*a.*) However, we already know from section 3 that there are no objects in the spectroscopic sample with these sizes and masses — these objects cannot be genuine compact galaxies. Of the 34 with inferred $M_* > 10^{11}$ $M_\odot$, 16 of these objects are spectrally identified as being stars, and one as a quasar at $z = 0.102$. Of the 17 spectrally confirmed galaxies, all have $|z_{\rm phot} - z_{\rm spec}| \gtrsim 0.02$. Of these, 15 have had their redshifts, and thus stellar masses, seriously overestimated; the other two are at $z > 0.12$, and so have had their intrinsic sizes underestimated.

Turning now to the comparison between panels (*b.*) and (*c.*), the first point to make is that the excess of unresolved sources is even more pronounced. We have matched all of these objects to the 2MASS point source catalog in order to investigate their NIR colors. 90 % of these objects fall in the stellar region of the $(J-K)$–$K$ color–magnitude plot; similarly, 80 % fall in the stellar region of a $(g-z)$–$(J-K)$ color–color plot.

Further, we have visually inspected the 434 objects with inferred $M_* > 10^{11}$ $M_\odot$ and with sizes smaller than the main $M_*$–$R_{\rm e}$ relation by 0.4 dex or more. Roughly 70 % of these objects are obviously stars: 133 come from crowded Galactic fields covered as part of SEGUE; 126 are double stars; 49 have clear diffraction spikes and/or are clearly saturated. Another 12 objects have been cross-matched with the USNO-B star catalog (within $1''$), and have measured proper motions of 1–$4''$/yr. 19 objects are the central point sources of very large spiral galaxies; most of these are also found in the ROSAT and/or FIRST catalogs. We also note that there are 17 very small disk or irregular galaxies with red point sources at or very near their centers. Most of these also have proper motion measurements from the USNO-B catalog, and several are spectrally identified as late type stars; it seems plausible that these galaxies simply have foreground stars coincidentally superposed very near their centers.

In short, of the 434 objects from the full photometric sample that, on the basis of photometric redshifts, are inferred to have $M_* > 10^{11}$ $M_\odot$ and $\Delta R_{\rm e} < -0.4$ dex, not one remains as a viable compact galaxy candidate.

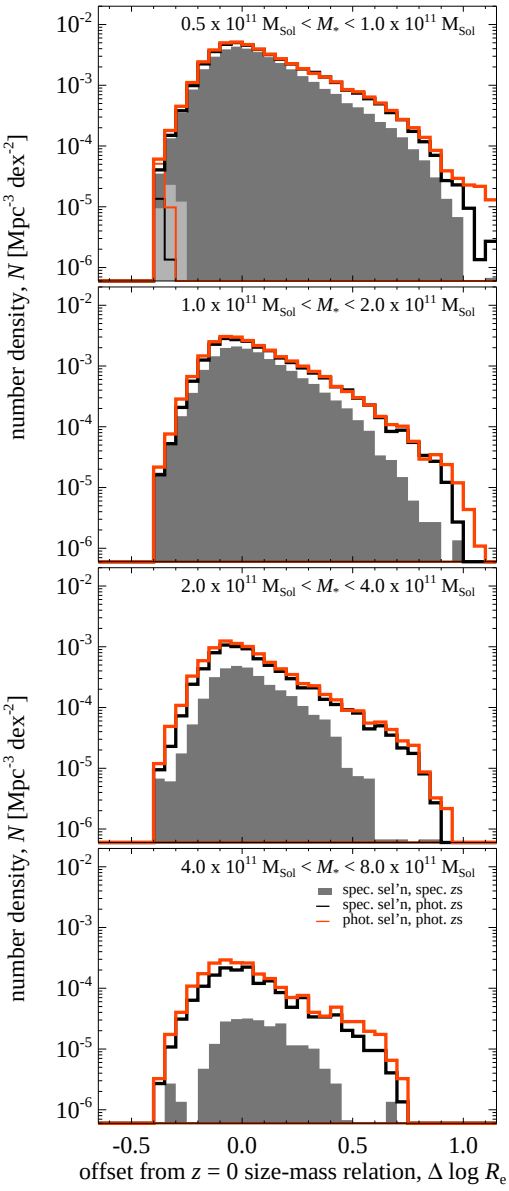


**Figure A.3.** — The observed size distribution of massive, red galaxies at $0.066 < z < 0.12$. — Each panel is for a different mass range as marked. In each panel, the solid histogram represents the SDSS spectroscopic sample, analyzed using spectroscopic redshifts. The black and red histograms are the SDSS spectroscopic and photometric samples, respectively, analyzed using photometric redshifts. We have visually inspected all objects with inferred $M_* > 10^{11}$ $M_\odot$ and $\Delta R_e < -0.5$ dex; not one of these objects is a plausible massive, compact galaxy candidate. The fact that the shape of the red histogram does not differ significantly from that of the black histogram for $\Delta R_e > -0.5$ dex indicates that the spectroscopic sample is not significantly biased against compact galaxies.

### A.4 Estimating the Importance of Spectroscopic Selection Effects

The conclusion from both the analyses that we have now presented is that there are no galaxies in the local universe with sizes and masses comparable to the compact galaxies found at higher redshifts. In **Figure A.3**, we provide a more quantitative statement of this conclusion, by plotting the size distribution for massive, red galaxies in different mass bins.

In this figure, the filled histograms represent the main SDSS spectroscopic sample, analyzed using spectroscopic redshifts, as in Section 3. The heavy black and red histograms represent the spectroscopic and photometric samples, respectively, analyzed using photometric redshifts, as in Section A.3. In all cases, objects excluded on the basis of visual inspection are not plotted; this accounts for the sharp cutoffs at $\Delta \log R_{\rm e} = -0.3$ and at $\Delta \log R_{\rm e} = -0.4$ for the filled and open histograms, respectively. Immediately above these cutoffs, where we have done no visual inspection but where there is likely to still be significant contamination, these distributions should be regarded as upper limits on the true distribution. In the upper panel, we plot those objects with observed sizes smaller than $0\farcs75$ separately as the light grey filled histogram, and the thin black and red histograms.

As in Section A.3, the difference between the filled and solid black histogram, both of which are derived from the spectroscopic sample, shows the increased scatter due to the use of photometric redshifts.

Similarly, the difference between the black and red histograms show the difference between the spectroscopic and photometric samples, and so allow a quantification of the bias in the spectroscopic sample. By simply tallying the numbers of galaxies with $-0.4 < \Delta \log R_{\rm e} < -0.3$, we find that the 'completeness' (the ratio between the number of galaxies in the spectroscopic sample compared to the full photometric sample) is 75 %, 68 %, 67 %, and 43 % for each of these mass bins, from lowest to highest.

In order to improve on these estimates, we have done the following. Using the approach described above, we have assigned each object a weight according to its $z_{\rm phot}$–derived mass and size. Then, going back to the spectroscopic sample, we use these to compute the mean weight in cells of $z_{\rm spec}$–derived mass and size. The completeness contours we derive in this way are in good qualitative agreement with those shown in Figure 2, although they suggest incompleteness at the 2–5 % level for mean–sized galaxies with $M_* \gtrsim 10^{11}$ M$_\odot$. Using these values to estimate the number of $M_* > 10^{11}$ M$_\odot$ galaxies with $\Delta R_{\rm e} < -0.3$ dex, this suggests that the spectroscopic sample is missing on the order of 4 such galaxies.

## References


Abazajian K N, et al., 2009, ApJS 182, 543

Baldwin, J.A., Phillips, M.M., Terlevich, R., 1981, PASP, 93, 5

Bell E F, de Jong R S, 2001, ApJ 550, 212

Bell E F, Zheng X Z, Papovich C, Borch A, Wolf C, Meisenheimer K, 2007, ApJ 663, 834

Bell E F, McIntosh D H, Katz N, Weiberg M D, 2003, ApJSS 149, 289

Bezanson R, van Dokkum P G, Tal T, Marchesini D, Kriek M, Franx M, Coppi P, 2009, ApJ 697, 1290

Blanton M R, Schlegel D J, Strauss M A et al., 2005, AJ 129, 2562

Blanton M R & Roweis S, 2007, ApJ 133, 734

Brammer G, Whitaker K E, van Dokkum P G et al., 2009, ApJ (accepted; arXiv:0910.2227)

Bruzual G & Charlot S, 2003, MNRAS 344, 1000

Buitrago F, Trujillo I, Barro G, Gallego J, Zamorano J, Conselice C J, 2008, ApJL 687, 61

Cappellari M, di Serego Alighieri S, Cimatti A, Daddi E, Renzini A, Kurk J et al., 2009, ApJ 704, L34

Cattaneo A, Dekel A, Faber S M, Guiderdoni B, 2008, MNRAS 389, 567

Cennaro A J, Trujillo I, 2009, ApJ 696, L43

Chabrier G, 2003, ApJ 586, L133

Cimatti A, Daddi E, Renzini A, Cassata P, Vanzella E, Pozzetti L, Cristiani S, Fontana A, Rodighiero G, Mingoli M, Zamorani G, 2004, Nature 430, 184

Cimatti A, Cassata P, Pozzetti L, Kurk J, Mignoli M, Renzini A, Daddi E et al., 2008, A&A 482, 21

Croton D J, Springel V, White S D M, De Lucia G, Frenk C S, Gao L, Jenkins A, Kauffmann G, Navarro J F, Yoshida N, 2006, MNRAS 365, 11

Daddi E, Renzini A, Pirzkal N, Cimatti A, Malhotra S, Stiavelli M, Xu C, Pasquali A, Rhoads J E et al., 2004, ApJ 617, 746

Daddi E, Renzini A, Cimati A, Malhotra S, Stiavelli M, Xu C et al., 2005, ApJ 626 680

Damen M et al., 2009, ApJ 690, 937

Damjanov I, McCarthy P J, Abraham R G, Glazebrook K, Yan H et al., 2009, ApJ 695, 101 (D09)

Dekel A, Birnboim Y, 2006, MNRAS 368, 2

De Lucia G, Springel V, White S D M, Croton D, Kauffmann G, 2006, MNRAS 366, 2, 499

Faber S M Wegner G, Burnstein D, Davies R L, Dressler A, Lynden-Bell D, Terlevich R J, 1989, ApJS 69, 763

Faber S M et al., 2007, ApJ 665, 265

Fan L, Lapi A, De Zotti G, Danese L, 2009, ApJ 689, L101

Franx M et al., 2003, ApJ 587, L79

Franx M, van Dokkum P G, Fórster-Schreiber N M, Wuyts S, Labbé I, Toft S, 2008, ApJ 688, 770

Gallazzi A, Charlot S, Brinchmann J, White S D M, Tremonti C A, 2005, MNRAS 362, 41

Glazebrook K, Abraham R G, McCarthy P J, Savaglio S, Chen H-W, Crampton D, Murowinski R, Jørgensen I, Roth K, Hook I, Marzke R O, Carlberg R G, 2004, Nature 430, 181

Guo Y, McIntosh D H, Mo H J, Katz N, van den Bosch F C, Weinberg M, Weinmann S M, Pasquali A, Yang X, 2009, MNRAS 398 1129

Hopkins P F, Bundy K, Murray N, Quataert E, Lauer T, Ma C-P, 2009, MNRAS 398, 898

Juneau S et al., 2005, ApJ 619, L135

Khochfar S & Silk J, 2009, MNRAS 397, 506

Kriek M, van Dokkum P G, Franx M, Quadri R, Gawiser E, Herrera D Illingworth G D, Labbé I, Lira P, Marchesini D, Rix H-W, Rudnick G, Taylor E N, Urry M C, Wuyts S, 2008, ApJL 649, 71

Kriek M, van Dokkum P G, Franx M, Illingworth G D, Marchesini D, Quadri R, Rudnick G, Taylor E N, Förster-Schreiber N M, Gawiser E, Labbé I, Lira P, Wuyts S, 2008, ApJ 677, 219

Kriek M, van der Wel A, van Dokkum P G, Franx M, Illingworth G D, 2008, ApJ 682, 896

Kriek M, van Dokkum P G, Labbé I, Franx M, Illingworth G D, Marchesini D, Quadri R, 2009, ApJ 700, 221

Kauffmann G, Heckman T M, White S D M et al., 2003, MNRAS 341, 33

Kauffmann G, Heckman T M, Tremonti C et al., 2003b, MNRAS 346, 1055

Kauffmann G, White S D M, Heckman T, Mé nard B, Brinchmann J, Charlot S, Tremonti C, Brinkmann J, 2004, MNRAS 353, 713

Longhetti M, Saracco P, Severgnini P, Della Caca R, Mannucci F, Bender R, Drory N, Feulner G, Hopp U, 2007, MNRAS 374, 614

Marchesini D, van Dokkum P G, Förseter-Schreiber N M, Franx M, Labbé I, Wuyts S, 2008, ApJ 701, 1765

McCarthy P J, Le Borgne D, Crampton D, Chen H-W, Abraham R G, Glazebrook K, Savaglio S et al., 2004, ApJ 614, L9

McCarthy P J, 2004b, ARAA, 42, 477

Menci N, Fontana A, Giallongo E, Salimbeni S, 2005, ApJ 632, 49

Muzzin A, Marchesini D, van Dokkum P G, Labbé I, Kriek M, Franx M, 2009, ApJ (submitted; arXiv 0906.2012)

Naab T, Johansson P H, Ostriker J P, 2009, ApJ 699, L178

Padmanabhan N, Schlegel D J, Finkbeiner D P et al. 2008, ApJ 674, 1217

Park C, Choi Y-Y, Vogeley M S, Gott J R, Blanton M R, 2007, ApJ 658, 898

Peng C Y, Ho L C, Impey C D, Rix H-W, 2002, AJ 124, 266

Saracco P, Longhetti M, Andreon S, 2009, MNRAS 392, 718

Shankar S & Bernardi M, 2009, MNRAS 396, L76

Shen S, Mo H J, White S D M et al., 2003, NMRAS 343, 978

Stoughton C et al., 2002, AJ 123, 485

Strauss M A, Weinberg S H, Lupton R H, Narayanan V K et al., 2002, AJ, 124, 1810

Taylor E N, Franx M, van Dokkum P G et al., 2009, ApJS 183, 295

Thakar A R, Szalay A, Fekete G, Gray J, 2008, CSE 10, 30

Toft S, van Dokkum P G, Franx M et al., 2007, ApJ 671, 285

Trujillo I, Fórster-Schreiber N M, Rudnick G et al., 2006, ApJ 650, 18

Trujillo I, Conselice C, Bundy K, Cooper M C, Eisenhardt P, Ellis R C, 2007, MNRAS 382, 109

Trujillo I, Cenarro A J, de Lorenzo-Cáceres A,

Vazdekis A, de la Rosa I G, Cava A, ApJL 692, L118

Valentinuzzi T, Fritz J, Poggianti B M, Bettoni D, Cava A, Fasano G, Onofrio M, Couch W J, Dressler A, Moles M, Moretti A, Omizzolo A, Kjaergaard P, Vanzella E, Verela J, 2009, ApJ (submitted; arXiv:0907.2392)

van der Wel A, Franx M, Wuyts S, van Dokkum P G, Huang J, Rix H-W, Illingworth G D, 2006, ApJ 652, 97

van der Wel A, Holden B, Zirm A W, Franx M, Rettura A, Illingworth G D, Ford H C, 2008, ApJ 688, 48

van der Wel A, Bell E F, van den Bosch F C, Gallazzi A, Rix H-W, 2009, ApJ 698, 1232

van Dokkum P G et al., 2006, ApJ 638, 59

van Dokkum P G, Franx M, Kriek M et al., 2008, ApJL 677, L5 (vD08)

van Dokkum P G, Kriek M, Franx M, 2009, *Nature* 460, 717

York D G et al., 2000, AJ 120, 2131

Weinmann S M, Kauffmann G, van den Bosch F C, Pasquali A, McIntosh D H, Mo H, Yang X, Guo Y, 2009, MNRAS 394, 1213

Williams R J, Quadri R F, Franx M, van Dokkum P G; Labbé I, 2009, ApJ 691, 1879

Wuyts S, Labbé I, Franx M, Rudnick G, van Dokkum P G, Fazio G G, Förster-Schreiber N M F, Huang J, Moorwood A F M, Rix H-W, Rö ttgering H, van der Werf P, 2007, ApJ 655, 51

Zheng X Z, Bell E F, Papovich C, Wolf C, Meisenhemier K, Rix H-W, Rieke G H, Sommerville R, 2007, ApJL 661, 41

Zirm A W, van der Wel A, Franx M et al.,2007, ApJ 656, 66

# Chapter V

## On the Masses of Galaxies in the Local Universe

We compare estimates of stellar mass, $M_*$, and dynamical mass, $M_{\rm d}$, for a sample of galaxies from the Sloan Digital Sky Survey (SDSS). Under the assumption of dynamical homology (*i.e.*, $\tilde{M}_{\rm d} \sim \sigma_0^2 R_{\rm e}$, where $\sigma_0$ is the central velocity dispersion and $R_{\rm e}$ is the effective radius), we find a tight but strongly non-linear relation between the two mass estimates: the best fit relation is $M_* \propto \tilde{M}_{\rm d}^{0.73}$, with an observed scatter of 0.15 dex. We also find that, at fixed $M_*$, the ratio $M_*/\tilde{M}_{\rm d}$ depends strongly on galaxy structure, as parameterized by Sérsic index, $n$. The size of the differential effect is on the order of 0.6 dex across $2 < n < 10$. Further, both the size and shape of the dependence is very close to the expectations from simple, spherical and isotropic dynamical models. This indicates that assuming homology gives the wrong dynamical mass. To explore this possibility, we have also derived dynamical mass estimates that explicitly account for differences in galaxies' structures. Using this 'structure-corrected' dynamical mass estimator, $M_{{\rm d},n}$, the best fit relation is $M_* \propto M_{{\rm d},n}^{0.92\pm0.08}$ with an observed scatter of 0.13 dex. While the data are thus consistent with a linear relation, they do prefer a slightly shallower slope. Further, we see only a small residual trend in $M_*/M_{{\rm d},n}$ with $n$. We find no statistically significant systematic trends in $M_*/M_{{\rm d},n}$ as a function of observed quantities (e.g., apparent magnitude, redshift), or as a function of tracers of stellar populations (e.g., H$\alpha$ equivalent width, mean stellar age), nor do we find significantly different behavior for different kinds of galaxies (*i.e.*, central versus satellite galaxies, emission versus non-emission galaxies). At 99 % confidence, the net differential bias in $M_*/M_{{\rm d},n}$ across a wide range of stellar populations and star formation activities is $\lesssim 0.12$ dex ($\approx 40$ %). The very good agreement between stellar mass and structure-corrected dynamical mass strongly suggests, but does not unambiguously prove, that: 1.) galaxy non-homology has a major impact on dynamical mass estimates, 2.) the stellar-to-dynamical mass ratio $M_*/M_{\rm d}$ has a relatively weak mass-dependence, and 3.) there are not strong systematic biases in the stellar mass-to-light ratios derived from broadband optical SEDs.

Taylor E N, Franx M, Brinchmann J,
van der Wel A, van Dokkum P G
for publication in *the Astrophysical Journal*
(to be submitted November 2009)

## 1 Introduction

Techniques for estimating galaxies' stellar masses are a crucial tool for understanding galaxies and their evolution. There are tight and well-defined correlations between stellar mass and many other important global properties like color, size, structure, metallicity, star formation activity, and environment (see, e.g., Kauffmann et al., 2003b; Shen et al., 2003; Blanton et al., 2005b; Gallazzi et al., 2006). Given a galaxy's stellar mass, $M_*$, it is thus possible to predict a wide variety of global properties with considerable accuracy. In this sense, stellar mass appears to be a key parameter in determining (or at least describing) a galaxy's current state of evolution. Moreover, since the growth of stellar mass (*cf.* absolute luminosity, color, etc.) is relatively slow and approximately monotonic, stellar mass is a particularly useful parameter for quantifying galaxy evolution.

Stellar mass estimates, whether derived from spectroscopic or photometric spectral energy distributions (SEDs), are plagued by a variety of random and systematic errors. These include a generic degeneracy between mean stellar age, metallicity, and dust obscuration. It is typical to make the simplifying assumptions that galaxies' stellar populations can be described *en masse* (*i.e.*, neglecting age/metallicity gradients and complex dust geometries), and that galaxies' complex star formation histories can be described parametrically. It is rare to attempt to account for active galactic nucleus (AGN) emission. The stellar initial mass function (IMF), including its universality or otherwise, remains a major 'known unknown'. Then there is the complication that different wavelengths probe different aspects of the stellar population; the inclusion of restframe UV or NIR data can thus, in principle and in practice, have a large impact on the estimated stellar mass. These effects are compounded by uncertainties in the stellar evolution models themselves. A topical example is the importance of NIR-luminous thermally pulsating asymptotic giant branch (TP-AGB) stars: for the same data and stellar population parameters, the use of Bruzual & Charlot (2003) or Maraston (2005) models can change the derived value of $M_*$ by a factor of 3 for galaxies that host young ($\lesssim 1$ Gyr) stars, but only if restframe NIR data is included (van der Wel et al., 2006; Kannappan & Gawiser, 2007). Conroy, Gunn & White (2009) have argued that the total random uncertainties in $M_*$ are on the order of $\sim 0.3$ dex for galaxies at $z \sim 0$.

For these reasons, it is essential to devise some way of assessing the quality of stellar mass estimates through comparison to some other fiducial mass estimate — this is the primary motivation for the present Chapter. Specifically, using a number of the publicly available 'value added' catalogs of the Sloan Digital Sky Survey (SDSS; York et al., 2000; Strauss et al., 2002), we will compare stellar mass estimates to total mass estimates derived from galaxy dynamics.

From the outset, we note that a difference between two quantities shows only that: a difference. With no definitive standard to use as a basis for comparison, the best that we can hope for is consistency between the two mass estimates. Further, if and when there are differences, it is impossible to unambiguously identify where

the 'fault' lies — or even if there is indeed a fault. For example, it is likely that the ratio between stellar and total mass varies as a function of mass, and/or some other global property/ies. We will also explore this issue in some detail.

This kind of comparison has been done for SDSS galaxies by Drory, Bender & Hopp (2004), who considered both stellar mass estimates derived from the SDSS spectra, as described by Kauffmann et al. (2003a), and those derived using SED-fitting techniques that are commonly used at for higher-redshift studies. These authors find a relatively tight correlation between the two stellar mass estimates, with a mild systematic bias depending on H$\alpha$ equivalent width. This bias suggests a potential problem with the stellar mass estimates as a function of specific star formation rate. Further, both stellar mass estimates correlated well with the simple dynamical mass estimate, $\tilde{M}_{\rm d}$ (defined below), but showed a clear trend in $M_*/\tilde{M}_{\rm d}$ with mass, such that less massive galaxies had higher values of $M_*/\tilde{M}_{\rm d}$. Padmanabhan et al. (2004) used the mass dependence of $M_*/\tilde{M}_{\rm d}$ to argue for an increasing stellar-to-dark mass ratio for elliptical galaxies with higher masses, as did Gallazzi et al. (2005). Both Rettura et al. (2006) and van der Wel et al. (2006) have performed similar comparisons for $z \lesssim 1$ galaxies, with similar conclusions.

It is common practice to derive a simple dynamical mass estimate based on the velocity dispersion, $\sigma_0$, and the effective radius, $R_{\rm e}$, via the scalar virial theorem:

$$G\tilde{M}_{\rm d} \approx k\sigma_0^2 R_{\rm e} \ . \tag{1}$$

(This is the dynamical mass estimator used for each of the studies cited in the previous paragraph.) The constant $k$ is usually assumed to be in the range 3–5, and is intended to account for the 'degree of virialization', including the effects of dark matter and the intrinsic shape of the velocity dispersion profile (see, e.g., Cappellari et al., 2006; Gallazzi et al., 2006; van der Wel et al., 2006). By assuming a constant $k$ for all galaxies, this expression implicitly assumes that all galaxies are dynamically homologous, or self-similar.[1]

But it is important to remember that the observed velocity dispersion is actually the luminosity-weighted mean of the true, radially-dependent velocity dispersion, projected onto the line of sight, and within the spectroscopic aperture. The shape of mass profile has a strong influence on the spatial and dynamical distribution of stellar orbits: in general, the relation between the observed velocity dispersion and the underlying mass profile thus depends on structure as well as size. As a dynamical mass estimator, $\tilde{M}_{\rm d}$ can therefore only be considered as approximate. (The tilde in $\tilde{M}_{\rm d}$ is intended to remind the reader of this fact.)

Bertin, Ciotti & Del Principe (2002) provide an analytic expression that makes it possible to approximately account for this effect. Using their formulation of the problem, the dynamical mass can be expressed as:

$$GM_{{\rm d},n} = K_V(n)\sigma_0^2 R_{\rm e} \ , \tag{2}$$

[1]But see also Wolf et al. (2009), who derive a mathematically identical relation from the spherical Jeans equation for a system in dynamical equilibrium.

Here, the term $K_V(n)$ encapsulates the effects of structure on stellar dynamics. (The subscript $n$ in $M_{\mathrm{d},n}$ is intended to make it clear that non-homology has been accounted for as a function of Sérsic index, $n$.) For convenience, we will refer to $M_{\mathrm{d},n}$ as a 'structure corrected' dynamical mass estimator, but we note that the inclusion of a structure-dependent term is not strictly a correction. Bertin, Ciotti & Del Principe (2002) also provide an analytical approximation for $K_V(n)$:

$$K_V(n) \cong \frac{73.32}{10.465 + (n - 0.95)^2} + 0.954 \ . \tag{3}$$

This expression for $K_V(n)$ has been derived assuming a spherical mass distribution that is dynamically isotropic and non-rotating, and which, in projection, follows a Sérsic (1963, 1968) surface density profile. For this (admittedly simple) scenario, this approximate expression for $K_V(n)$ is accurate at the percent level for $1 \leq n \leq 10$. Substituting trial values of $n = 2$ and $n = 8$ into Equation 3 suggests that the differential effect of non-homology on the inferred value of the dynamical mass is as much as a factor of 3, or 0.5 dex. Our first task in this paper, then, will be to explore the importance of structural differences between galaxies, using this prescription.

Before we begin, note that there are alternative approaches to exploring the consistency between stellar and dynamical mass estimates. In particular, a number of authors have considered the relation between galaxies' stellar and dynamical masses in the context of well known scaling relations between luminosity/mass and dynamics. For example, Bell & de Jong (2001) considered the relation between baryonic (*cf.* stellar) mass, $M_{\mathrm{bar}}$, and circular rotation velocity, $V_C$, for disk galaxies — the baryonic Tully-Fisher relation. These authors showed that stellar mass estimates based on different passbands (*i.e.* $M_*/L_V$ versus $M_*/L_K$) produced consistent $M_{\mathrm{bar}}$–$V_c$ relations. Furthermore, for a fixed IMF, they argued that it was possible use a single color to estimate stellar mass-to-light ratios with an accuracy of 0.1–0.2 dex.

There have also been a number of analogous studies for elliptical galaxies, based on the fundamental plane (Djorgovsky & Davis, 1987; Dressler et al., 1987), which can be understood as a correlation between the dynamical mass-to-light ratio, $M_{\mathrm{d}}/L$, and surface brightness. These studies (see, e.g., Cappellari et al. 2006; and references therein) have tended to focus on the 'tilt' of the fundamental plane — that is, the deviation of the observed relation from the expectation assuming both a constant $M_*/L$ and structural and dynamical homology for all early type galaxies. The tilt of the fundamental plane thus offers a means of probing variations in $M_{\mathrm{d}}/L$ (including both variations in $M_*/L$ due to different stellar populations, and variations in $M_*/M_{\mathrm{d}}$ due to, e.g., different dark-to-luminous mass ratios) and/or the degree of non/homology. While the relative importance of these different effects remains an open question, it seems clear that, at least for early type galaxies, both $M_*/L$ and $M_*/M_{\mathrm{d}}$ vary systematically with mass (see, e.g., Prugniel & Simien, 1996; Bertin, Ciotti & Del Principe, 2002; Trujillo et al., 2004; Cappellari et al., 2006; La Barbera, 2008; Allanson et al., 2009).

This Chapter is structured as follows: in Section 2, we describe the various SDSS-derived catalogs that we will use, including the definition of our galaxy sample. We validate the velocity dispersion measurements used to derive dynamical masses in Appendix A. In Sections 3 and 4, we present two parallel comparisons between stellar and dynamical mass estimates for galaxies in our sample. First, in Section 3, we use the simple dynamical mass estimate $\tilde{M}_{\rm d}$; then, in Section 4 we show how the comparison changes using the structure-corrected dynamical mass estimate, $M_{{\rm d},n}$. In Section 5, we explore the consistency between $M_{{\rm d},n}$ and $M_*$. In particular, we will show that there are no statistically significant trends in the ratio $M_*/M_{{\rm d},n}$ that would indicate measurement biases in $M_*$ and/or $M_{{\rm d},n}$; this is not the case for the simple dynamical mass estimate $\tilde{M}_{\rm d}$. We show in Appendix B that these results are not unique to the sample we consider in the main text. We discuss the interpretation and implications of this result in Section 6, before providing a summary of our main results and conclusions in Section 7.

For this work, we will assume the concordance cosmology; *viz.*., ($\Omega_m$, $\Omega_\Lambda$, $\Omega_0$) = (0.3, 0.7, 1.0), and $H_0 = 70$ km/s/Mpc, and adopt a Chabrier (2003) IMF.

## 2 Data

The work presented in this Chapter is based on data drawn from several publicly available catalogs based on the SDSS dataset. Our analysis is based on redshifts and velocity dispersions from the basic SDSS catalog for DR7 (Abazajian et al., 2009)[2]. We use Sérsic-fit structural parameters from Gou et al. (2009, hereafter G09) and SED-fit stellar mass-to-light ratio measurements from the DR7 Max-Planck-Institute for Astrophysics (MPA)/Johns Hopkins University (JHU) value added catalog[3]. In Appendix B, we repeat our analysis using the Sérsic -fit structural parameters given in the New York University (NYU) Value Added Galaxy Catalog (VAGC; Blanton et al., 2005a) for DR7. Each of these catalogs have been well described and documented in the references given; in this Section, we only briefly summarize the most relevant aspects of each catalog for the present work.

### 2.1 Redshifts and Velocity Dispersions

There are two sets of redshift and velocity dispersion measurements given in the basic SDSS catalog for DR6 and DR7: the 'spectro1D' values produced by the Chicago group, and the 'specBS' values produced by the Princeton group. In terms of redshifts, the two algorithms produce virtually identical results. The major difference between the two algorithms is that, whereas the Chicago pipeline only gives velocity dispersion measurements to those galaxies that are spectrally classified as being 'early type', all galaxies are given a velocity dispersion measurement by the Princeton pipeline. From DR6, both the Princeton and Chicago velocity dispersion measurement algorithms have been updated, so as to eliminate the systematic bias at low dispersions identified by Bernardi (2007) for the DR5

[2]Accessed via the Catalog Archive Server (CAS; Thakar et al., 2008): http://casjobs.sdss.org/CasJobs/

[3]Available via http://www.mpa-garching.mpg.de/SDSS/

values.[4] In Appendix A, we compare both sets of velocity dispersions to those given by Faber et al. (1989) for bright, early type galaxies: in both cases, the values agree with the Faber et al. (1989) catalog values with an rms difference of $\sim 18$ km/s and no discernible systematic bias.

The default redshifts and velocity dispersions for in the SDSS catalog (specifically, using SDSS parlance, the parameters `z` and `veldisp` given in the table `specObjAll`) are the Chicago values. For the sole reason that Princeton velocity dispersions are given for all galaxies (rather than only the spectroscopically early-types) we have elected to use the Princeton values instead; these are also the measurements adopted for both the NYU and the MPA/JHU value added catalogs.

As we have mentioned in the Introduction, the observed velocity dispersion is the luminosity weighted average within the (projected) spectroscopic aperture. In order to account for aperture effects, we have scaled the observed value, $\sigma_{\rm ob}$, for each galaxy to a central velocity dispersion, $\sigma_0$, which is defined to be that that would be observed within a circular aperture with a radius equal to 1/8 times the apparent effective radius, $\Theta_{\rm e}$ Jørgensen, Franx & Kjaergaard (see, e.g. 1995). This correction has been made assuming $\sigma(R) \propto R^{-0.066}$; *i.e.*, $\sigma_0/\sigma_{\rm ob} = (8\Theta_{\rm ap}/\Theta_{\rm e})^{-0.066}$, where $\Theta_{\rm ap} = 1\farcs5$ is the radius of the SDSS spectroscopic aperture. The scaling of $\sigma(R)$ has been derived by taking a luminosity-weighted integral of the spatially resolved velocity dispersions of galaxies from the SAURON survey (Cappellari et al., 2006). The corrections themselves are small — the median correction is 0.02 dex, with an rms scatter of 0.02 dex — and does not have a major impact on our results. Our qualitative conclusions do not change if we assume the slightly weaker radial dependence $\sigma(R) \propto R^{-0.04}$ as found by Jørgensen, Franx & Kjaergaard (1995), or if we neglect this correction altogether.

### 2.2 Sérsic Parameters: Size, Flux, and Structure

G09 have derived $r$-band structural parameters including total magnitude, $m_{\rm tot}$, effective radius, $\Theta_{\rm e}$, and Sérsic index, $n$, for a modest sized sample of SDSS galaxies. (We discuss the specific sample selection in Section 2.4 below.) These values have been derived via parametric fits to the (2D) $r$-band surface brightness distribution of each galaxy, assuming a Sérsic (1963, 1968) profile, and convolved with the appropriate PSF, using the publicly available code `galfit` (Peng et al., 2002). In order to account for blending, where two galaxies are very close, both the target and companion(s) are fit simultaneously. Through analysis of simulated data, G09 show that the median error in each of $m_{\rm tot}$, $\Theta_{\rm e}$, and $n$ to be less than 10 %.

In Appendix B, we will also make use of Sérsic-fit structural parameters from the NYU VAGC (Blanton et al., 2005a). Whereas `galfit` considers the 2D surface brightness distribution, the VAGC algorithm makes fits to the 1D azimuthally averaged curve of growth. The analysis of simulated data presented by Blanton et al. (2005a) shows the VAGC Sérsic parameters to be systematically biased

[4] See http://www.sdss.org/dr7/algorithms/veldisp.html for a discussion of the spectro1D and specBS algorithms, as well as a comparison between these values and those from Bernardi et al. (2003a,b) and SDSS DR5.

towards low fluxes, sizes, and Sérsic indices. This problem becomes progressively worse for larger $n$, such that sizes are underestimated by $\gtrsim 20$ % and fluxes by $\gtrsim 10$ % for $n \gtrsim 5$. G09 have shown that this bias is produced by background over-estimation and over-subtraction in the VAGC Sérsic fits, owing to the use of a 'local', rather than a 'global' background estimator.

### 2.3 Stellar Masses

We note that there are rather large differences between the the Sérsic magnitudes given by G09 and the default `model` magnitudes given in the SDSS catalog. The `model` photometry is derived by making parametric fits to the 2D surface brightness distribution in each band, using the sector fitting technique described by Strauss et al. (2002). These fits assume either an exponential or a De Vaucouleurs profile; the profile shape is chosen based on the best-fit $\chi^2$ in the $r$-band. For galaxies that are best fit by a De Vaucouleurs `model` we find $(m_{\mathrm{G09},r} - m_{\mathrm{DeV},r}) \approx -0.26 + 0.11(n-4)$, where $n$ is the Sérsic index reported by G09; the scatter around this relation is at the level of 0.15 mag ($1\sigma$). That is, even where G09 find $n = 4$, their flux is approximately 0.26 mag brighter than the SDSS (De Vaucouleurs) `model` flux; this discrepancy is larger for larger values of $n$. For this reason, we take the $r$-band Sérsic magnitude from G09 as a measure of total flux.

To derive a stellar mass, we then use $M_*/L$s taken from the MPA-JHU catalog (DR7), which is maintained by the Garching group.[5] Note that, unlike previous MPA-JHU catalogs (e.g. Kauffmann et al., 2003a; Brinchmann et al., 2004; Gallazzi et al., 2005), which were based on the SDSS spectroscopy, these masses are derived from fits to the *ugriz* `model` SEDs.[6] Note, however, that the SED photometry has been corrected for emission lines, according to the line–to–continuum flux ratio in the spectroscopic fiber aperture (we discuss the importance of this correction in Section 5.3 below). The SED fits are based on the synthetic stellar population library described by Gallazzi et al. (2005), which have been constructed using the Bruzual & Charlot (2003) stellar population models, assuming a Chabrier (2003) stellar IMF. These $M_*/L$s have been shown to be in good agreement (rms in $\Delta \log M_*$ of $\lesssim 0.1$ dex) with the spectrally-derived values described by Kauffmann et al. (2003a) for DR4.[7]

### 2.4 Sample Definition

Our decision to rely on the Sérsic structural parameters derived by G09 restricts us to working with their sample. Our rationale for this decision is that, in our estimation, the G09 Sérsic fits are the most robust that are presently available.

The G09 sample was constructed with the specific goal of exploring differences in the sizes and structures of 'central' and 'satellite' galaxies in groups

[5] Available via http://www.mpa-garching.mpg.de/SDSS/DR7/

[6] Note that in the SDSS algorithm, when deriving the *ugiz* `model` photometry, the structural parameters in the fits are held fixed to the $r$-band values; only the overall normalization (*i.e.*, total flux) is allowed to vary. The fits in each band are also convolved with the appropriate PSF. In this sense, the `model` SEDs are both aperture matched, and PSF-corrected.

[7] See http://www.mpa-garching.mpg.de/SDSS/DR7/mass_comp.html

and clusters. To this end, they selected 911 $z < 0.08$ 'centrals' as the first-ranked (in terms of $M_*$) group/cluster members from the Yang et al. (2007) group catalog, which was in turn constructed from the DR4 NYU VAGC. These galaxies were selected to have a flat logarithmic distribution in halo mass in the range $11.85 < \log M_{\rm halo}/{\rm M}_\odot < 13.85$ (800 galaxies), plus 100 galaxies in the range $13.85 < \log M_{\rm halo}/\,{\rm M}_\odot < 14.35$, and all 11 central galaxies in clusters with $\log M_{\rm halo}/{\rm M}_\odot > 14.35$. In this way, the central galaxy sample was constructed to span a representative range of (large) halo masses.

G09 also construct two $z < 0.08$ 'satellite' control samples, in which the satellite galaxies are selected to match the central galaxies. For the first of these, satellites are chosen to match centrals in $M_*$ to within 0.08 dex; in the second, satellites are also required to match centrals to within 0.03 mag in ${}^{0.1}(g-i)$ color. Because more massive galaxies are more likely to be (counted as) centrals, not every central has a satellite counterpart within these limits: the matching is more than 90 % successful for $M_* < 10^{10.85}$ ${\rm M}_\odot$, and less than 10 % successful for $M_* > 10^{11.15}$ ${\rm M}_\odot$. The two satellite samples, so constructed, consist of 769 and 746 galaxies, respectively.

G09 exclude a number of these galaxies from their analysis because of confusion, leaving a sample of 879 central galaxies, and two samples of 704 and 696 satellites each. While duplicates are not allowed within the individual satellite samples, some galaxies do appear in both samples; combining the two satellite samples we have 1167 unique galaxies. We exclude a further 71 galaxies whose spectra are not deemed 'science worthy' by the SDSS team (*i.e.* the flag `sciencePrimary` is set to zero). In order to avoid very large errors in the dynamical mass estimates, we also exclude 160 galaxies that have relative errors in their velocity dispersion measurements that are greater than 10 %. This requirement excludes mostly low-$n$ and low-$M_*$ galaxies: the vast majority of these 160 galaxies have $M_* < 10^{10.8}$ ${\rm M}_\odot$ and $n < 1.5$. Our results do not depend on these selections. We are thus left with a sample of 1816 galaxies, of which 784 have been selected as central galaxies, and 1032 have been selected as being satellites of comparable mass.

The major disadvantage to using the G09 sample is that the relative number of central/satellite galaxies is not at all representative of the general galaxy population. However it is worth noting that G09 have shown that, at least for structurally early type galaxies, and after matching both color and mass, there are no differences in the sizes and structures of central and satellite galaxies. They thus conclude that the distinction between central and satellite has no impact on galaxy structure. This already suggests that the G09 sample may be adequately statistically representative of the massive galaxy population.

Even so, we will explicitly examine the possible role of sample selection effects in shaping our results in Section 5.1 by comparing different subsamples from within the combined G09 sample. Further, in Appendix B, we analyze a more general galaxy sample, using structural parameters from either the NYU VAGC Sérsic fits or the SDSS De Vaucouleurs/exponential `model` fits.

## 3 Results I — Comparing Stellar and Dynamical Mass Estimates Assuming Dynamical Homology

In this Section and the next, we present parallel comparisons between stellar mass and two different estimates of dynamical mass. As we have said in the Introduction, it is common practice to obtain a simple dynamical mass estimate based on $\sigma_0$ and $R_{\rm e}$ alone, using the scalar virial theorem; *viz.* $\tilde{M}_{\rm d} \approx k\sigma_0^2 R_{\rm e}$. In Section 3.1, we directly compare the values of $M_*$ and $\tilde{M}_{\rm d}$ for the G09 galaxies; we will assume $k = 4$. We will then argue in Section 3.2 that the agreement between stellar and dynamical mass estimates may be significantly improved if we allow for non-homology. To test this idea, in Section 4, we will perform the same comparisons using the structure corrected dynamical mass estimator, $M_{{\rm d},n}$.

### 3.1 The Relation Between Stellar and Dynamical Mass

In **Figure 1**, we compare the values of the simple dynamical mass estimator, $\tilde{M}_{\rm d}$, to the values of $M_*$ for galaxies in the G09 sample. The first thing to notice is that there is a relatively tight but clearly nonlinear relation between $\tilde{M}_{\rm d}$ and $M_*$, such that $M_* \propto \tilde{M}_{\rm d}^a$ with $a < 1$. Moreover, this simple analysis suggests that for many galaxies, including the majority of galaxies with $\tilde{M}_{\rm d} \lesssim 10^{10.5}$ M$_\odot$, $M_*$ actually *exceeds* $\tilde{M}_{\rm d}$. This is logically inconsistent, and necessarily implies a problem in the calculation of $M_*$ and/or $\tilde{M}_{\rm d}$.

Before discussing this result further, it is appropriate to make a few comments about the random errors in our estimates of $M_*$ and $\tilde{M}_{\rm d}$. In particular, it is important to realize that the errors in $\Theta_{\rm e}$, $m_{\rm tot}$, and $n$ are strongly covariant: for example, an error in the structural index will affect the values of both $\Theta_{\rm e}$ and $m_{\rm tot}$. Because $M_*$ depends on the measured value of $m_{\rm tot}$, and $\tilde{M}_{\rm d}$ on the measured value of $\Theta_{\rm e}$, $M_*$ and $\tilde{M}_{\rm d}$ are thus also covariant. This makes the seemingly trivial task of fitting a line to the observed $M_*$–$\tilde{M}_{\rm d}$ relation rather problematic. To do this properly would require full and consistent treatment of the covariant uncertainties in the Sérsic-fit parameters, but this information is not given by G09.

Our solution to this problem is simply to minimize the mean absolute perpendicular distance between the data and the fit. When doing so, we also use a 'sigma-clipping' algorithm to avoid the influence of the most egregious outliers; specifically, we iteratively exclude points that lie off the best-fit line by more than 5 times the rms offset. While the gradient of the best-fit line does depend on the fitting scheme used (we will explore this in more detail in Section 4.1), the best fit parameters are not strongly dependent on how aggressively we sigma-clip. In order to avoid strong covariances between the slope and intercept of the best-fit line, we actually compute the fit in terms of $\log(\tilde{M}_{\rm d}/10^{11}{\rm M}_\odot)$; that is, we fit a relation of the form $y = a(x-11)+b_{11}$. Statistical uncertainties on the fit parameters have been derived from bootstrap resampling. The best fit to the $M_*$–$\tilde{M}_{\rm d}$ relation, so derived, is shown as line heavy dashed line in Figure 1. The best fit parameters are $a = 0.73 \pm 0.007$ and $b_{11} = -0.14 \pm 0.003$.

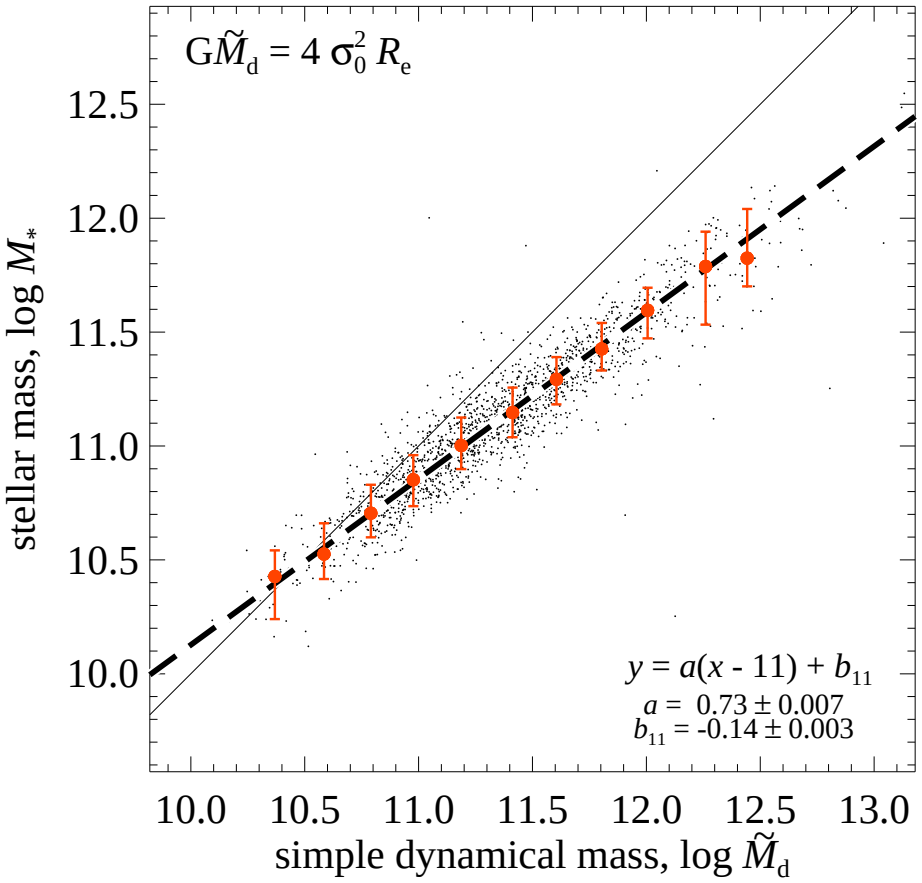


**Figure 1.** — Comparing stellar and dynamical mass estimates under the assumption of dynamical homology. — This Figure shows the relation between stellar mass and a simple estimate of dynamical mass, $G\tilde{M}_d = 4\sigma_0^2 R_e$, for galaxies in the G09 sample. The black points show the data themselves; the red points with error bars show the median and 16/84 percentile values of $M_*$ in narrow bins of $\tilde{M}_d$; the heavy dashed line shows a linear fit to the data, with the form and parameters as given. *While there is a relatively tight correspondence between $M_*$ and $\tilde{M}_d$, the relation is clearly non-linear.* Further, for $\tilde{M}_d \lesssim 10^{10.5}$ $M_\odot$, $M_*$ appears to exceed $\tilde{M}_d$ for the majority of galaxies, which is logically inconsistent. We explore these results in greater detail in Figures 2 and 3. In Figure 4, we show how these results change if we account for structural and dynamical non-homology in our estimates of dynamical mass.

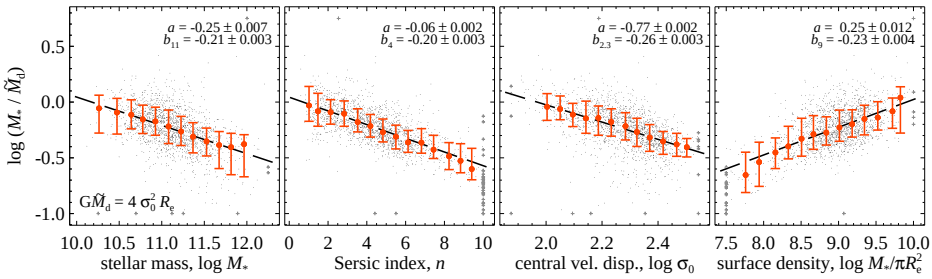


**Figure 2.** — Comparing stellar and dynamical mass estimates under the assumption of dynamical homology. — Each panel of this Figure plots the stellar-to-dynamical mass ratio, $M_*/\tilde{M}_{\rm d}$, as a function of (left to right) mass, structure, velocity dispersion, and effective surface density. Within each panel, the black points show the data themselves; points that fall outside the plotted range are shown with a small grey plus; the large points with error bars show the median and 16/84 percentiles of $\log(M_*/\tilde{M}_{\rm d})$ in narrow bins of the quantity shown on the $x$-axis. The dashed lines show fits to the data of the form $y = a(x - X) + b_X$, in analogy to Figure 1. *At least when using this (overly) simple way of estimating galaxies' dynamical masses, there are strong trends in $M_*/\tilde{M}_{\rm d}$ with both mass and structure.* We see similarly tight and strong trends with velocity dispersion and effective surface density, as well as with other parameters like size (not shown).

In **Figure 2**, we explore the relation between $M_*$ and $\tilde{M}_{\rm d}$ in greater detail. The different panels of this Figure show the difference between $M_*$ and $\tilde{M}_{\rm d}$ as a function of several interesting global properties: (from left to right) galaxy mass, structure, dynamics, and surface density. It clear that that $M_*/\tilde{M}_{\rm d}$ is strongly correlated with all four of these parameters. For each of the parameters shown, the size of the median trend in $M_*/\tilde{M}_{\rm d}$ across the sample is on the order of 0.5 dex, although it is slightly lower for $M_*$ and slightly higher for effective surface density.

To quantify this statement, we have again made fits to the data, assuming the form $y = a(x - X) + b_X$, where $X$ is an arbitrary value chosen to be close to the median value of the quantity $x$ for our sample. For these fits, in contrast to the previous Section, we have minimized the mean absolute *vertical* offset between the data and the fit. Again, we use a non-aggressive sigma-clipping scheme to exclude extreme outliers. (In all that follows, when considering the stellar-to-dynamical mass ratio, we will always fit in this way; we will only use the minimum perpendicular distance algorithm described above when fitting the relation between stellar and dynamical masses.) The best fit lines to the data, so derived, are shown as the heavy dashed lines in each panel; the best-fit parameters are given in each panel. In the case of Sérsic index, the scatter around the best fit relation is $\approx 0.12$ dex; for the other three parameters it is slightly higher: $\approx 0.15$ dex.

### 3.2 The Importance of Galaxy Structure in Dynamical Mass Estimates

There are of course strong correlations between mass, velocity dispersion, surface density, and structure. It is thus possible that the apparent trend with any given parameter in Figure 2 is 'spurious', in the sense that it is driven by a trend

in another more 'fundamental' parameter. We note that galaxies' star formation activity and histories have been shown to correlate closely with all of mass, velocity dispersion, and surface density (see, e.g., Kauffmann et al., 2003a,b, 2006; Franx et al., 2008; Graves, Faber & Schiavon, 2009). Indeed, with the assumption that $M_* \sim M_{\mathrm{d},n}$, these three quantities are all related by factors of $R_\mathrm{e}$, which is also closely correlated with $M_*$ (Shen et al., 2003; Franx et al., 2008). But the fact that $M_*/\tilde{M}_\mathrm{d}$ depends on galaxy *structure* — and particularly the agreement between the observed trend and expectations derived from a simple dynamical model — immediately suggests that structure-dependent differences in galaxy dynamics may play a role in the results shown in Figure 1. With this in mind, in **Figure 3** we attempt to separate out the $M_*$– and $n$–dependences of $M_*/\tilde{M}_\mathrm{d}$. Specifically, we want to test the hypothesis that departure from linearity in the $M_*$–$\tilde{M}_\mathrm{d}$ relation seen in Figure 1 is at least in part a function of structure, and not mass.

Figure 3a shows $M_*/\tilde{M}_\mathrm{d}$ as a function of $M_*$; the colored lines show the median relation in bins of Sérsic index. The median relation between $M_*/\tilde{M}_\mathrm{d}$ and $M_*$ has a rather similar slope for each different $n$ bin: $M_*/\tilde{M}_\mathrm{d}$ does depend on mass. If the dynamical mass-to-light ratio were to depend on mass only, however, we would expect the relations for different Sérsic indices to overlap. Instead, the relations for each bin are clearly offset from one another. That is, at fixed mass, the scatter in $M_*/\tilde{M}_\mathrm{d}$ is closely correlated with galaxy structure.

In Figure 3b, we do the opposite: in this panel, we plot $M_*/\tilde{M}_\mathrm{d}$ as a function of Sérsic index, and the different lines show the median relation in bins of stellar mass. Again, it is clear that $M_*/\tilde{M}_\mathrm{d}$ depends on both $M_*$ and $n$: the median relations for each different mass bin are roughly parallel, but offset from one another.

Further, the different mass bins in Figure 3b cover different ranges in Sérsic index: where the lowest mass bin is dominated by galaxies with $1 < n < 4$, the highest mass bin is dominated by $n > 4$ galaxies. Similar behavior can be seen in Figure 3a: the lowest $n$ bin contains very few galaxies with $\log M_* > 11.3$ $\mathrm{M}_\odot$, and virtually all $\log M_* > 11.5$ $\mathrm{M}_\odot$ galaxies are in the $n > 7$ bin. That is, there is a correlation between $M_*$ and $n$ within our sample. Because the trend is towards lower values of $M_*/\tilde{M}_\mathrm{d}$ with increasing $n$, this correlation contributes to the apparent mass dependence of the dynamical-to-stellar mass ratio.

Again, the dynamical mass estimates used thus far have been derived under the assumption of dynamical homology (*i.e.* $K_V$ is equal to 4 for all galaxies). The solid line in Figure 3b shows the expected shape of the relation between $M_*/\tilde{M}_\mathrm{d}$ and $n$, derived from Equation 3. The agreement between the observed relation between $M_*/\tilde{M}_\mathrm{d}$ and $n$ and the expectations from this (admittedly simple) dynamical model immediately suggests that non-homology has an important effect on dynamical mass estimates. We note that both Prugniel & Simien (1996) and Trujillo et al. (2004) have made a similar argument for the importance of non-homology based on the fundamental plane of elliptical galaxies (see also Section 6 below).

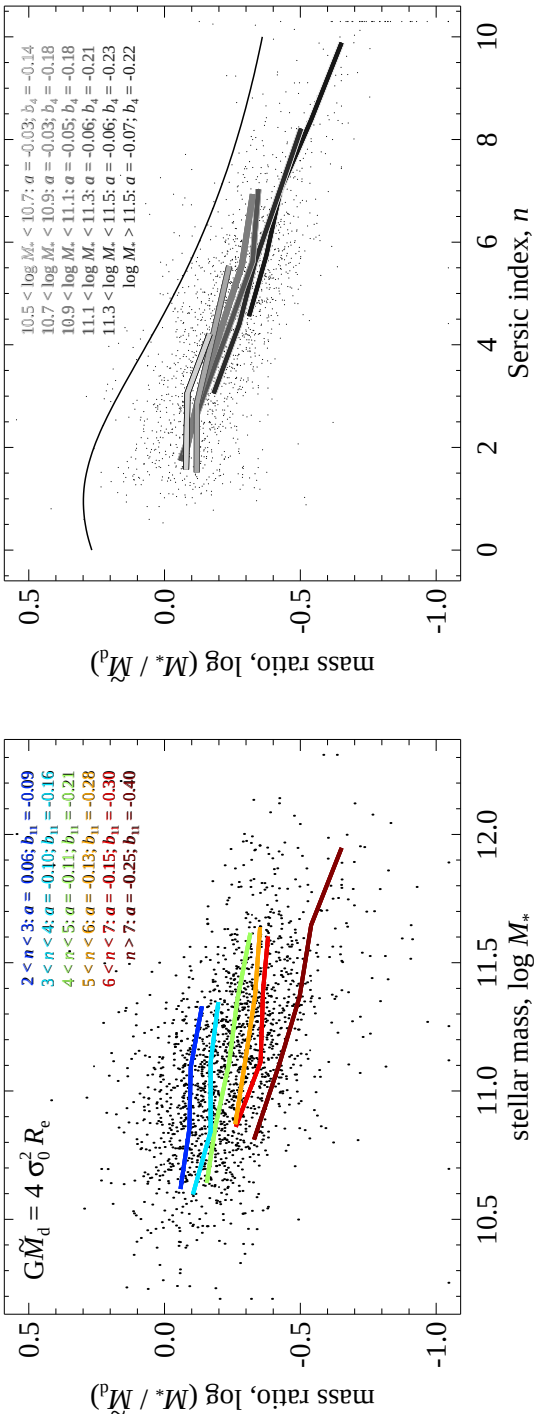


**Figure 3.** — Separating out the mass- and structure-dependences of the stellar-to-dynamical mass ratio, $M_*/\tilde{M}_\mathrm{d}$. — In the left panel, we plot $M_*/\tilde{M}_\mathrm{d}$ as a function of $M_*$; the thick lines in this panel show the median relation in bins of $n$. In the right panel, we do the reverse: $M_*/\tilde{M}_\mathrm{d}$ is plotted as a function of $n$, with the solid lines showing the median relation in bins of $M_*$. In both panels, we give the parameters of the best fit relation for galaxies in each of the bins shown. It is clear from both panels that *at fixed* $M_*$, *the scatter in* $M_*/\tilde{M}_\mathrm{d}$ *is directly related to* $n$. It is also true that, even at fixed $n$, $M_*/\tilde{M}_\mathrm{d}$ appears to vary with $M_*$; we explore this issue further in Figure 6. For this Figure, we have assumed $G\tilde{M}_\mathrm{d} = 4\sigma_0^2 R_\mathrm{e}$; in the right-hand panel, the solid curve shows the expected shape of the $M_*/\tilde{M}_\mathrm{d}$ relation for a spherical and dynamically isotropic system that follows a Sérsic profile, derived using Equation 3. The general agreement between the shape of this curve and that of the $M_*/\tilde{M}_\mathrm{d}$–$n$ relation suggests that including this term may significantly improve the correspondence between stellar and dynamical mass estimates. We explore this issue further in Figures 4 and 5.

# 4 Results II.— Comparing Stellar and Dynamical Mass Estimates Accounting for Dynamical Non-Homology

In this Section, we investigate the potential impact of non-homology on the agreement between stellar and dynamical mass estimates. To this end, we repeat the comparisons between stellar and dynamical mass estimates presented above, using the structure corrected dynamical mass estimator, $M_{\rm d,}{}_n$, in place of the simple estimate $\tilde{M}_{\rm d}$. We quantify the relation between $M_*$ and $M_{\rm d,}{}_n$ for our sample in Section 4.1, and show in Section 4.2 that allowing for non-homology considerably improves the correspondence between stellar and dynamical mass estimates.

## 4.1 The Relation Between Stellar and Dynamical Mass

In **Figure 4**, we show the relation between stellar and dynamical mass for the G09 sample, using the structure corrected dynamical mass estimator, $M_{\rm d,}{}_n$; this Figure should be compared to Figure 1. It is immediately obvious that the correlation between $M_*$ and $M_{\rm d,}{}_n$ is much closer to linear than that between $M_*$ and the simple dynamical mass, $\tilde{M}_{\rm d}$. Further, we note that the results are now logically consistent, in that $M_* < M_{\rm d,}{}_n$ for almost all galaxies. This is our most basic result: *structure-dependent differences in galaxy dynamics can have a big impact on the inferred dynamical mass*, and so the stellar-to-dynamical mass ratio.

The best-fit parameters for the $M_*$–$M_{\rm d,}{}_n$ relation are $a = 0.92 \pm 0.007$ and $b_{11} = -0.23 \pm 0.004$, where we have used bootstrap resampling to estimate the statistical uncertainty. While the statistical errors in the fit parameters are impressively small, systematic errors are sure to dominate. To see this, consider what would happen if we were to impose a stellar mass limit $M_* > M_{\rm lim}$ in Figure 4: we would only include those galaxies with $M_{\rm d,}{}_n < M_{\rm lim}$ that have high values of $M_*/M_{\rm d,}{}_n$; similarly, we would exclude those galaxies with $M_{\rm d,}{}_n < M_{\rm lim}$ that have low values of $M_*/M_{\rm d,}{}_n$. This would lead to a significantly shallower best-fit slope to the $M_*/M_{\rm d,}{}_n$ relation. As a specific example, if we were only to consider galaxies from the G09 sample with $M_* > 10^{10.8}$ M$_\odot$, we would find $M_* \propto M_{\rm d,}{}_n^{0.86}$. Note that, while we have phrased this in terms of a mass selection effect, at least part of this effect is related to how we have fit for the slope of the $M_*$–$M_{\rm d,}{}_n$ relation; *viz.*, by minimizing the mean perpendicular distance between the data and the best-fit line. If we were instead to fit by minimizing the mean offset in $M_*/M_{\rm d,}{}_n$ as a function of $M_*$, we would reduce our vulnerability to this effect.

There are thus two effects that have the potential to systematically bias the measured slope of the $M_*$–$M_{\rm d,}{}_n$ relation. We can obtain a simple estimate for the systematic error on the parameters $a$ and $b_{11}$ by re-fitting the $M_*$–$M_{\rm d,}{}_n$ relations in different ways. If we fit by minimizing the mean vertical offset, $\Delta M_*$, from the best-fit $M_*/M_{\rm d,}{}_n$ relation, we find $a = 0.85$ and $b_{11} = -0.20$. If instead we fit by minimizing the mean horizontal offset, $\Delta M_{\rm d,}{}_n$, we find $a = 1.00$ and $b_{11} = -0.26$. That is, the systematic uncertainties related to the technique used to fit the $M_*$–$M_{\rm d,}{}_n$ relation (for this sample) are on the order $\Delta a = 0.08$ and $\Delta b_{11} = 0.03$.

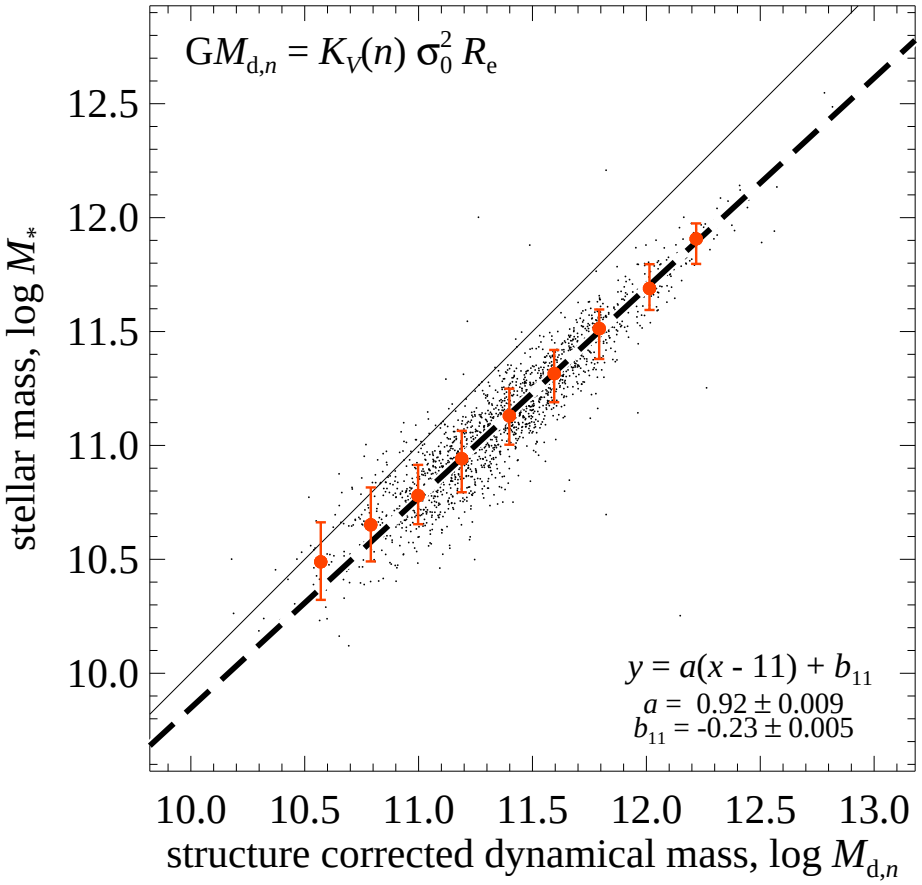


**Figure 4.** — Comparing stellar and dynamical mass estimates, accounting for both structural and dynamical non-homology. — The difference between this Figure and Figure 1 is that we have used $GM_{\rm d,n} = K_V(n)\sigma_0^2 R_{\rm e}^2$, with $K_V(n)$ defined as in Equation 3; otherwise all symbols and their meanings are as in Figure 1. The key point to be made from this Figure, in comparison to Figure 1, is that *allowing for non-homology makes a big difference to the inferred dynamical masses*, and so to the correspondence between stellar and dynamical masses. Further, we note that the apparent inconsistency whereby $M_* > \tilde{M}_{\rm d}$ for $M_* \lesssim 10^{10.5}$ $M_\odot$ galaxies seen in Figure 1 is no longer apparent. After accounting for structure-dependent differences in galaxies' dynamics, the relation between $M_*$ and $M_{\rm d,n}$ is much more nearly linear. However it remains true that the difference between stellar and dynamical mass appears to grow with increasing mass. -1

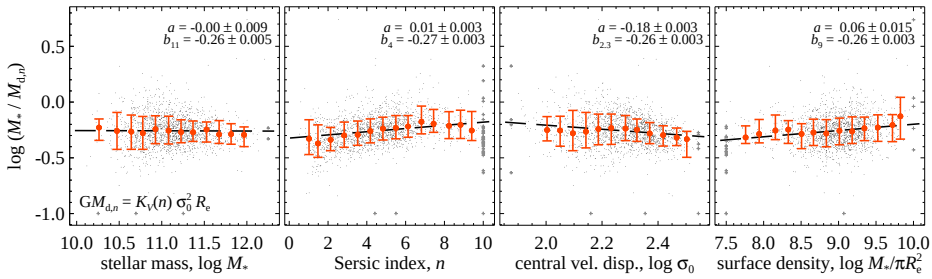


**Figure 5.** — Comparing stellar and dynamical mass estimates accounting for structure-dependent differences in galaxy dynamics. — The difference between this Figure and Figure 2 is that we have used $GM_{\mathrm{d},n} = K_V(n)\sigma_0^2 R_{\mathrm{e}}^2$, with $K_V(n)$ defined as in Equation 3; otherwise all symbols and their meanings as is in Figure 1. *After accounting for structure dependent differences in galaxy dynamics, the apparent trends in $M_*/M_{\mathrm{d},n}$ with stellar mass and Sérsic index are substantially reduced.* The apparent trends with other properties, including velocity dispersion, surface density, size, and color, are also substantially reduced, or effectively disappear (see also Figures 7 and 9).

What about the systematic biases due to the particular mass distribution of galaxies in the G09 sample? To explore the importance of these effects in our measurement of the slope of the $M_*$–$M_{\mathrm{d},n}$ relation, we have tried re-fitting the $M_*$–$M_{\mathrm{d},n}$ relation, weighting each point according to its stellar mass. The specific weights have been derived through a comparison between the mass distribution of galaxies within the G09 sample, in bins of $\Delta M_* = 0.1$ dex, and the $z \sim 0$ mass function of Bell et al. (2003). We have chosen the weight for each galaxies so that that the weighted stellar mass distribution of the G09 sample matches the 'real' stellar mass function. This weighting scheme is akin to $1/V_{\mathrm{max}}$ weighting, inasmuch as if one were able to derive $V_{\mathrm{max}}$ values for the sample, one would hope to obtain similar values.

Re-fitting the G09 sample using these weights, we find $a = 1.00 \pm 0.05$. The larger random error on this value in comparison to our fiducial values stems from the fact that the lower mass galaxies are given much greater weights (by several orders of magnitude); the inclusion/exclusion of these points in the bootstrap resampling thus has a major impact on the best-fit slope. The fact that using these weightings gives an almost perfectly linear relation between $M_*$ and $M_{\mathrm{d},n}$ is striking, but it is important to remember that this fit is based almost entirely on the relatively small number of $M_* \lesssim 10^{11}$ $\mathrm{M}_\odot$ galaxies in the sample. In Appendix B, we perform the same analysis for a more general galaxy sample, drawn from the NYU VAGC, which provides a much better sampling of the true galaxy mass function. Using the weighting scheme described above, the relative weights of galaxies with $10.2 < \log M_*/\mathrm{M}_\odot < 11.7$ differ by a factor of only 10. For this galaxy sample, we find $a = 0.930 \pm 0.004$ $(^{+0.03}_{-0.07})$ without weighting, compared to $a = 0.933 \pm 0.007$ with weighting.

While the data are consistent with a linear relation between $M_*$ and $M_{\mathrm{d},n}$, they thus prefer a slightly shallower relation. For the G09 sample, we find $a =$

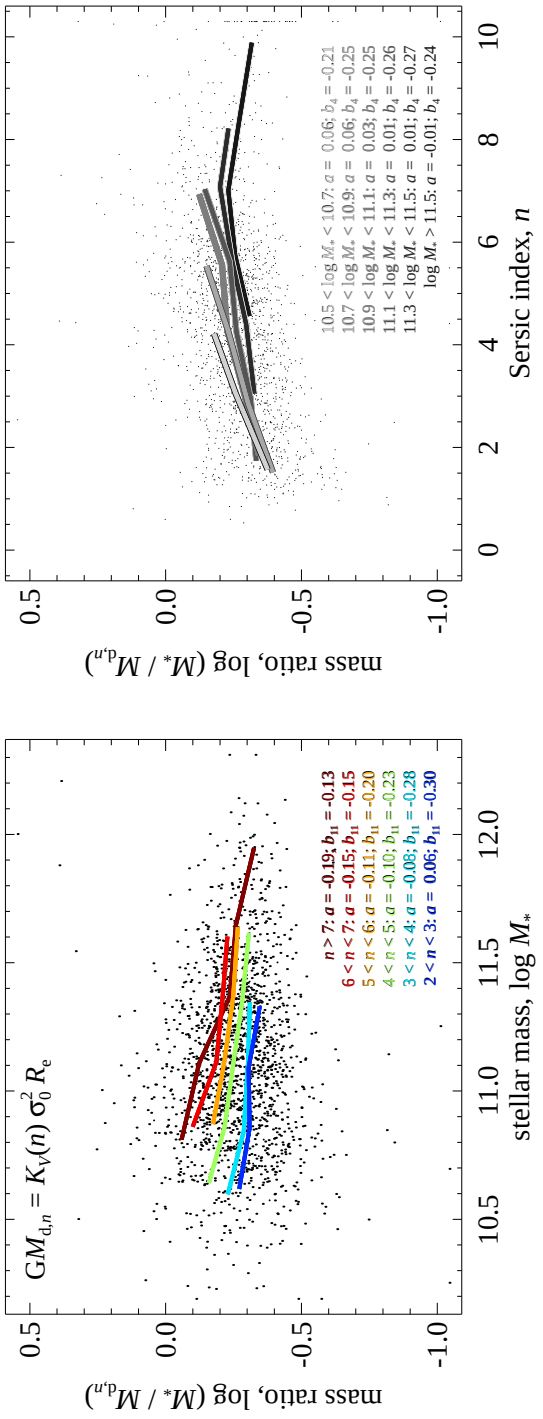


**Figure 6.** — Does $M_*/M_{\mathrm{d},n}$ vary with $M_*$, $n$, or both? — The difference between this Figure and Figure 3 is that we have accounted for structure-dependent differences in galaxy dynamics through the $K_V(n)$ term defined in Equation 3; otherwise all symbols and their meanings as is in Figure 3. As in Figure 3, it is clear that the scatter in $M_*/M_{\mathrm{d},n}$ at fixed $M_*$ depends strongly on $n$; conversely, at fixed $n$, there is also a strong correlation between $M_*$ and $M_*/M_{\mathrm{d},n}$. This effect appears to be stronger for higher values of $n$ and $M_*$. In comparison to Figure 3, the inclusion of an $n$-dependent dynamical term obviously changes the slope of the $M_*/M_{\mathrm{d},n}$–$n$ relation, but cannot affect the slope of the $M_*/M_{\mathrm{d},n}$–$M_*$ relation at fixed $n$. In other words, the apparent $n$-dependence of $M_*/M_{\mathrm{d},n}$ at fixed $M_*$ is sensitive to the specific model used to derive $K_V(n)$; on the other hand, the result that, at fixed $n$, $M_*/M_{\mathrm{d},n}$ varies with $M_*$ is insensitive assumed form of $K_V(n)$.

$0.93 \pm 0.007$ ($\pm 0.07$), and $b_{11} = -0.23 \pm 0.004$ ($\pm 0.03$). This should be compared to the values of $a = 0.73 \pm 0.006$ ($^{+0.07}_{-0.03}$) and $b_{11} = -0.14 \pm 0.003$ ($^{+0.01}_{-0.03}$) for the simple dynamical mass estimate, $\tilde{M}_{\rm d}$. While the mass distribution of galaxies within the G09 sample can in principle induce a large bias in the measured slope of the $M_*$–$M_{{\rm d},n}$ relation, our best fit value is in fact consistent with that derived from a more general galaxy sample, in which these effects play a far smaller role. We will explore the potential role of other sample selection effects in Section 5.2.

In **Figure 5**, we show the trends in $M_*/M_{{\rm d},n}$ with mass, structure, dynamics, and density; this Figure should be compared to Figure 2. For each of these four parameters, the trends in $M_*/M_{{\rm d},n}$ are significantly weaker than what we have seen for $M_*/\tilde{M}_{\rm d}$. The net differential trend across the sample is now on the order of 0.2 dex or so, as compared to 0.5 dex for $M_*/\tilde{M}_{\rm d}$. While there is still a strongly statistically significant trend in $M_*/M_{{\rm d},n}$ with $\sigma$, the trend with surface density is now only significant at the $4\sigma$ level. While we do still see signs of a trend in $M_*/M_{{\rm d},n}$ with $n$, this trend is not statistically significant, at least for the sample as a whole. We discuss this point further in the next Section.

### 4.2 Does $M_*/M_{{\rm d},n}$ Depend on Mass, or Structure, or Both?

In **Figure 6**, we return to the issue of the $n$- and $M_*$-dependence of $M_*/M_{{\rm d},n}$; this Figure should be compared to Figure 3. In Figure 6a, we show the median relation between $M_*/M_{{\rm d},n}$ and $M_*$ in bins of $n$. Again, the trends in $M_*/M_{{\rm d},n}$ with $M_*$ for the different $n$-bins are parallel, but offset from one another. In each of the $3 \lesssim n \lesssim 7$ bins, we find that $M_*/M_{{\rm d},n}$ scales approximately as $M_*^{0.1}$; that is, roughly consistent with the scaling that we see for the sample as a whole. Figure 6b shows the median relation between $M_*/M_{{\rm d},n}$ as a function of $n$ for different bins in $M_*$. While the trend in $M_*/M_{{\rm d},n}$ with $n$ is substantially weaker than we saw using the simple dynamical mass, $\tilde{M}_{\rm d}$, we still see that $M_*/M_{{\rm d},n}$ varies with $n$; if anything, it would appear that by using the prescription for $K_V(n)$ given in Equation 3, we have overcorrected for the effects of non-homology. Without detailed dynamical modeling, however, we have no means of refining the model used to derive Equation 3. (We will discuss this point further in Section 6.)

In other words, *we have shown that accounting for structural and dynamical homology significantly improves the agreement between stellar and dynamical mass estimates as a function of Sérsic index, n,* but we have not unambiguously shown whether or not $M_*/M_{{\rm d},n}$ depends on galaxy structure — nor can we.

## 5 Results III — Exploring Potential Biases in $M_*/M_{{\rm d},n}$

In this Section we discuss three general classes of biases that may affect the results we have presented in Section 4: first, systematic biases in the Sérsic fits that we use to derive $M_*$ and $M_{{\rm d},n}$ (Section 5.1; see also Appendix B); then, the possibility of severe selection effects for the G09 sample (Section 5.2; see also Appendix B); and finally, systematic effects associated with the estimation of stellar mass-to-light ratios (Section 5.3). We will show very good consistency between

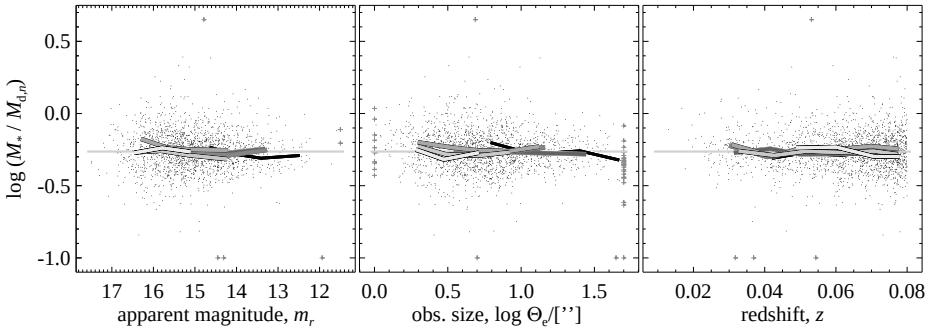


**Figure 7.** — Signs of observational biases?. — Each panel shows $M_*/M_{\mathrm{d},n}$ a function of a direct observable. Within each panel, the solid lines show the median relation in bins of stellar mass; these bins are the same as those shown in the right-hand panels of Figures 3 and 6. The points show the data themselves; points that fall outside the range of each panel are shown as a small grey plus. Within the G09 sample, there are correlations between Sérsic index and each of the observed quantities shown in this Figure; however, there are only very weak trends in $M_*/M_{\mathrm{d},n}$ with any of these observables. Moreover, each of the stellar mass bins follows essentially the same median relation. This suggests that *neither the stellar nor dynamical mass estimates are obviously seriously biased by systematic errors in the Sérsic fits.*

the values of $M_*$ and $M_{\mathrm{d},n}$; this is not the case for the simple estimate $\tilde{M}_{\mathrm{d}}$. As in the previous Section, this consistency provides strong circumstantial evidence — but not proof beyond a reasonable doubt — that there are no significant biases in either measurement.

### 5.1 Looking for Possible Observational Biases

We explore the possibility of serious observational biases in **Figure 7**. In each panel of this Figure, we plot $M_*/M_{\mathrm{d},n}$ as a function of a basic observable: namely (left to right), apparent magnitude, apparent size, and redshift. The lines in each panel show the median relation for the same bins of stellar mass as are shown in Figures 3b and 6b.

By a similar argument to the one given in Section 4.2, if the apparent mass dependence of the ratio $M_*/M_{\mathrm{d},n}$ were driven by observational effects that are strong functions of apparent brightness or size, then we would expect there to be clear trends in $M_*/M_{\mathrm{d},n}$ for each of the stellar mass bins in Figure 7. This is not obviously the case. Fitting to the data in each stellar mass bin, the gradients of the best-fit relation between $M_*/M_{\mathrm{d},n}$ and all of apparent magnitude, apparent size, and redshift are consistent with zero; this is true for each stellar mass bin individually (typically within $1\sigma$), as well as for the sample as a whole (within 1.2–$1.4\sigma$).

We note that the same is not true using, for example, the Sérsic structural parameters given in the NYU VAGC to derive $M_*$ and $M_{\mathrm{d},n}$ (as we do in Appendix B). In this case, we do see a weakly statistically significant gradient (at the level of $4\sigma$) with observed size. That is, this kind of test is indeed able to (weakly) detect mild systematic errors in the Sérsic-fit parameters on the order of 10–20 %. The results in Figure 7 thus argue against the idea that there are any serious biases

affecting the measurement of $M_*$ or $M_{\mathrm{d},n}$ (or, more accurately, the ratio $M_*/M_{\mathrm{d},n}$) associated with the Sérsic-fit structural parameters used to derive these values.

### 5.2 Looking for Potential Sample Selection Effects

As we have described in Section 2.4, the G09 sample has been constructed in such a way that massive galaxies in general, and in particular central galaxies in very massive halos, are drastically overrepresented in comparison to the general field population. If there are systematic differences in $M_*/M_{\mathrm{d},n}$ as a function of, for example, environment or star formation activity, there is thus a very real danger that sample selection effects may play an important role in shaping our results.

We explore this issue in **Figure 8**, in which we have divided the G09 sample in central/satellite and non-/emission subsamples, in order to look for differences between these populations. Here, we have made the non-/emission distinction using the spectral classification scheme of Brinchmann et al. (2004), which is based on the BPT diagram (that is, on a comparison between the [NII]6584/H$\alpha$ and [OII]5007/H$\beta$ emission line ratios; Baldwin, Phillips & Terlevich 1981). What we call 'non-emission' galaxies in Figure 8 are actually those galaxies that that are 'unclassifiable' on the basis of the BPT diagram, because they have little or no emission in one or more of these lines; the 'emission' subsample comprises both star forming galaxies and AGN.

In the upper panels of Figure 8, it is clear that we find essentially identical relations between $M_*$ and $M_{\mathrm{d},n}$ for the central and satellite galaxy subsamples. Given that, as we have shown in Figure 6, $M_*/M_{\mathrm{d},n}$ varies with both $M_*$ and $n$, it is not all that surprising that central and satellite galaxies show the same $M_*$–$M_{\mathrm{d},n}$ relation: not only have the two subsamples have been constructed to be matched in $M_*$, G09 have shown that there are no structural differences between satellites and centrals at fixed mass.

We do find very slightly different $M_*$–$M_{\mathrm{d},n}$ relations for the emission and non-emission subsamples, however. Formally, the two relations are slightly offset from one another, at the level of 0.05 dex; the gradients of the two relations differ at the $2\sigma$ level. That said, for the $10.5 < \log M_* < 12$ range spanned by the sample, the difference between the best-fit relations for each subsample and that for the sample as a whole are $\lesssim 0.03$ dex. We also note that the apparent offset disappears (at least within $2\sigma$) if we consider only the $n > 2$ galaxies in both subsamples. Moreover, we do not find any differences in $M_*/M_{\mathrm{d},n}$ at fixed $M_*$ and $n$ between these different subsamples — within statistical uncertainties, each subsample shows the same behavior as is seen in Figure 6. This suggests that the apparent offset between for the emission and non-emission galaxies is driven by the different distributions of Sérsic indices within the emission and non-emission subsamples, rather than any difference in $M_*/M_{\mathrm{d},n}$ intrinsically related to galaxy activity. (See also Section 5.3).

That is, our results do not appear to be strongly influenced by the relative numbers of central/satellite galaxies or of non-/emission galaxies in our sample. The above analysis does suggest, however, that the measured relation between

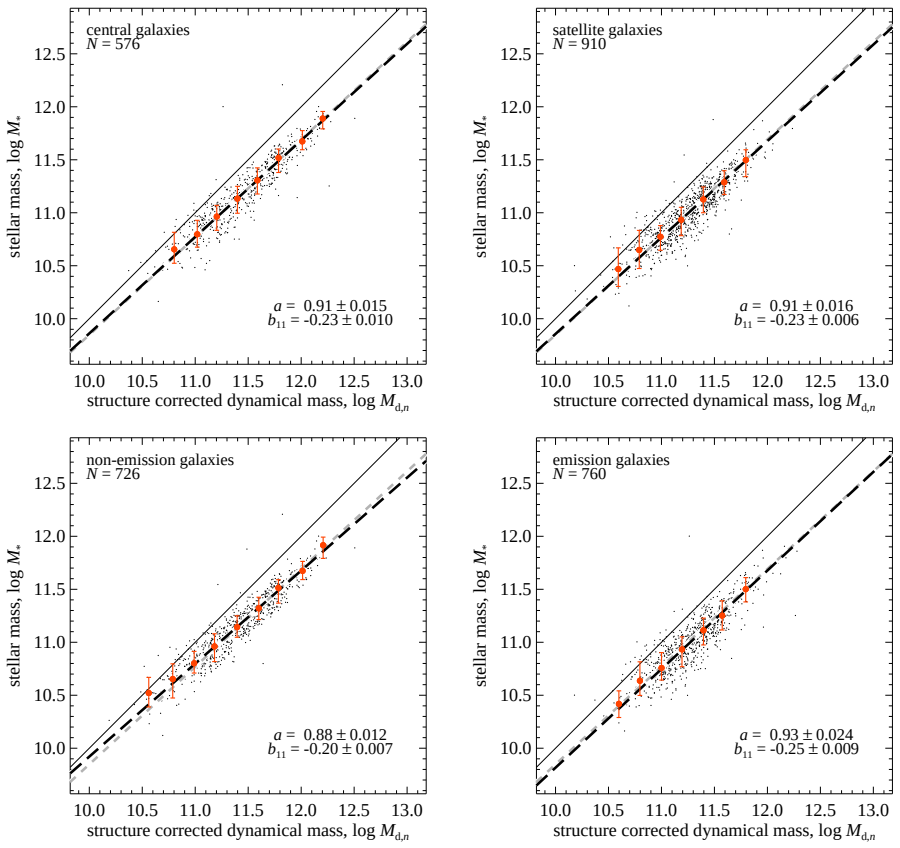


**Figure 8.** — Possible sample selection effects? — The G09 sample is not necessarily representative of the general galaxy population, in terms of the distribution of masses, environments, or star formation activity. To explore the potential role of these effects, each panel of this Figure shows the $M_*$–$M_{\rm d,}n$ comparison for different subsets of the G09 sample, distinguishing between central and satellite galaxies (upper panels) and between spectroscopically emission and non-emission galaxies (lower panels). For the lower panels, the spectroscopic classification is based on the BPT diagram, following the scheme of Brinchmann et al. (2004); the 'emission' sample includes both AGN-dominated and composite spectra galaxies. In the upper panels, we do not show those satellite/central that do not have Brinchmann et al. (2004) spectral classifications; that is, the same samples are plotted in both the upper and the lower panels. The dashed grey line in each panel shows the best fit relation for the entire G09. *We find a consistent $M_*$–$M_{\rm d,}n$ relation for all four of these subsamples, as well as the G09 as a whole.* While there is possibly a slight offset in $M_*/M_{\rm d,}n$ between star-forming and passive galaxies, we note that this disappears if we only consider $n > 2$ galaxies. That is, this offset appears to be due to the different $n$-distributions of the star-forming and passive samples, rather than an intrinsic difference between the values of $M_*/M_{\rm d,}n$ for emission and non-emission galaxies (see also Figure 9). We also note that for each of the subsamples shown, within statistical errors, we find consistent behavior in $M_*/M_{\rm d,}n$ at fixed $M_*$ and $n$ as is shown in Figure 6. We therefore conclude that selection effects do not play a major role in shaping our results (see also Appendix B).

$M_*$ and $M_{\mathrm{d},n}$ is sensitive to the joint $M_*$–$n$ distribution within the sample. This is a direct consequence of the fact that $M_*/M_{\mathrm{d},n}$ depends on both mass and Sérsic index (Figure 6). We have considered biases associated with the mass distribution in Section 4.1. But note that if, at fixed mass, the distribution of $n$ within the G09 sample differs significantly from the 'true' distribution for a general field population, then we may therefore find a very different slope for the $M_*/M_{\mathrm{d},n}$ relation.

For this reason, in Appendix B, we repeat our $M_*$–$M_{\mathrm{d},n}$ comparisons for a more general sample of $0.035 < z < 0.08$ field galaxies, using the structural parameters given in the NYU VAGC.[8] The best fit logarithmic slope of the $M_*$–$M_{\mathrm{d},n}$ relation for $0.035 < z < 0.08$ field galaxies is $a = 0.91 \pm 0.003$, compared to $a = 0.92 \pm 0.009$ for the G09 sample. This suggests that the G09 sample is not grossly biased in terms of the distribution of $n$ at fixed mass. (Here again, it is significant that G09 have found that, at fixed mass, there are no structural differences between satellite and central galaxies.)

To summarize the results of this section, then, separate analysis of central/satellite and non-/emission galaxies suggest that our results are not strongly affected by selection effects associated with these properties. Furthermore, although the observed slope of the $M_*/M_{\mathrm{d},n}$ relation is in principle sensitive to the joint $M_*$–$n$ distribution within the sample, we find very little difference between the G09 sample and a more general field galaxy sample. Finally, we stress that we do not find any evidence that selection effects have an important impact on the results shown in Figure 6; *i.e.*, the observation that the ratio $M_*/M_{\mathrm{d},n}$ depends on both $n$ (at fixed $M_*$) and on $M_*$ (at fixed $n$).

### 5.3 Looking for Biases in the Stellar Mass-to-Light Ratio Estimates

As we have stated in the Introduction, our primary motivation for comparing stellar and dynamical mass estimates is to validate the stellar mass estimates. We explore this issue in **Figure 9**. In each panel of this Figure, we plot the ratio $M_*/M_{\mathrm{d},n}$ as a function of a different property of the stellar population. The solid lines in each panel show the median relation for the same stellar mass bins shown in Figures 3 and 6. The upper panels plot $M_*/M_{\mathrm{d},n}$ as a function of a direct observable; the lower panels plot $M_*/M_{\mathrm{d},n}$ as a function of a derived property. Note that the stellar mass estimates we have used were derived from the *ugriz* photometry, rather than spectra. The measured values of $M_*/M_{\mathrm{d},n}$ are thus formally independent of the three spectral measurements shown in the top panels. Further, note that the age estimate that we show is taken from Kauffmann et al. (2003a); these values are also derived from the spectra. The dust obscuration

[8] As we have already remarked, the VAGC Sérsic fits have been shown to suffer from systematic errors arising from background oversubtraction. But, as we also show in Appendix B, it turns out that the ratio $M_*/M_{\mathrm{d},n}$ is extremely robust to random or systematic errors in the fitting of structural parameters, so long as the apparent magnitude, effective radius, and Sérsic index are derived consistently, and the term $K_V(n)$ is included in the definition of the dynamical mass (See also Section 6). Further, we find very similar $M_*$–$M_{\mathrm{d},n}$ relations for the G09 sample using either VAGC or the G09 values for the structural parameters. Any large differences in the measured $M_*$-$M_{\mathrm{d},n}$ relation between the G09 sample and the general field sample would therefore necessarily be a product of selection effects.

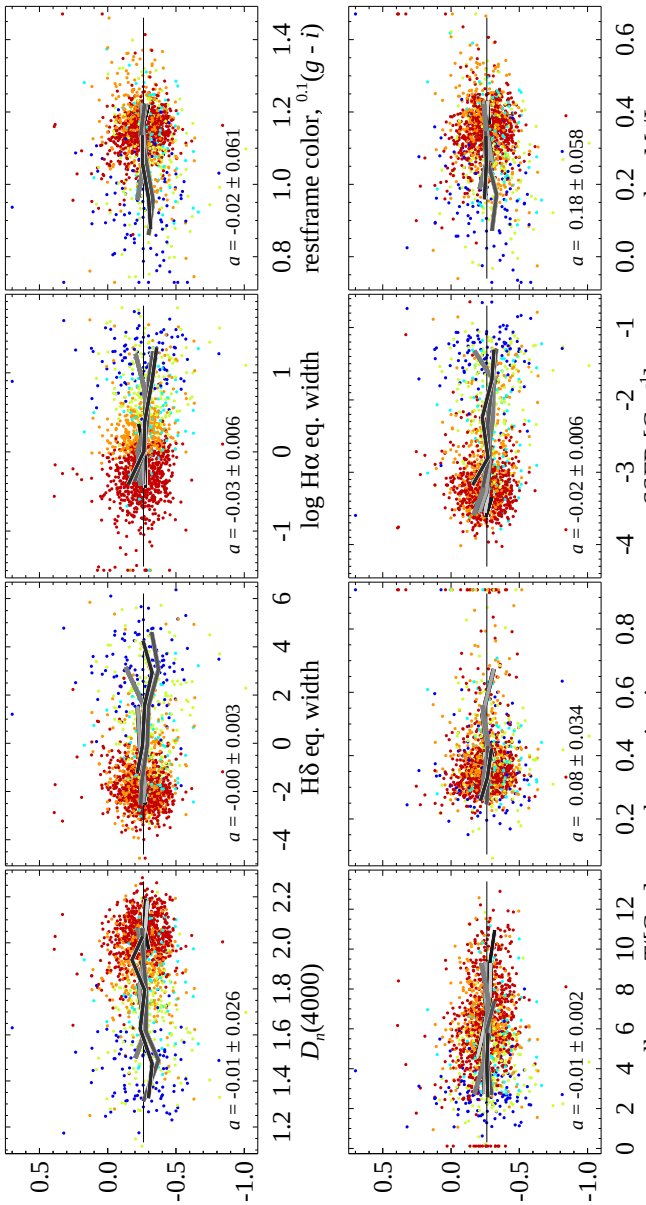


**Figure 9.** — Stellar population-dependent effects? — Each panel plots $M_*/M_{\mathrm{d},n}$ as a function of a stellar population diagnostic. The upper panels show directly observed quantities; the lower panels show stellar population parameters derived from synthetic stellar population modeling. In these panels, the mean stellar age shown is that given by Kauffmann et al. (2003a), which are based on spectra, and are thus formally independent of $M_*/L$. The dust extinction, specific star formation rates, and $M_*/L$s are all self-consistently derived from the same SED fits, as described in Section 2.3. Within each panel, individual points are color-coded according to spectral classifications; specifically: star forming (blue), low S/N star forming (cyan), composite (yellow), AGN-dominated (orange), and 'unclassifiable' (*i.e.* non-emission; red) galaxies. Within each panel, the black points with error bars show the median and 16/84 percentiles of $M_*/M_{\mathrm{d},n}$ in bins. We see no statistically significant systematic differences in $M_*/M_{\mathrm{d},n}$ for galaxies with different stellar populations or star formation histories. This argues against there being major problems with the stellar population models used to estimate $M_*/L$. At 99 % confidence, these results suggest that any differential biases as a function of the parameters shown is $\lesssim 0.12$ dex ($\approx 40$ %).

and specific star formation rates (SSFRs) shown come from the SED fits used to derive $M_*/L$; these values are thus self-consistently derived.

We have color-coded the data in Figure 9 according to their spectral classification as given by Brinchmann et al. (2004); *viz.*: star forming (blue); low S:N star forming (cyan); composite (yellow), AGN (orange), and non-emission (red). In general, we see little if any differences in the values of $M_*/M_{\mathrm{d},n}$ for different stellar populations. There are not obviously large differences between the mean values of $M_*/M_{\mathrm{d},n}$ for AGN hosts, star forming galaxies, or non-emission galaxies.

Within each panel we give the gradient of the best-fit line for the whole sample. These values are all statistically consistent with zero. The only possible exceptions to this rule are the gradients in $M_*/M_{\mathrm{d},n}$ as a function of H$\alpha$ equivalent width (EW) and as a function of $M_*/L_r$, both of which are non-zero at the $\sim 3\sigma$ level. Just as there is little if any trend for the sample as a whole, there are no statistically significant trends for any of the individual mass bins.

We can quantify the degree of correspondence between stellar and dynamical mass estimates by considering the differential bias between galaxies over the range of each of the properties shown in Figure 9. Looking at the median relations shown for each mass bin suggests that the magnitude of such differential biases are at most 0.2 dex. We can obtain similar estimates for the sample as a whole using the fit parameters given in each panel of Figure 9. Taking the $3\sigma$ statistical limits on the slopes of these relations, we find that the differential effects across the full range of the sample are $\lesssim 0.1$ dex for $D_n(4000)$, H$\delta$ EW, and restframe color; and $\lesssim 0.15$ dex for H$\alpha$ EW, age, dust extinction, SSFR, and $M_*/L$.

The same is not true using the simple estimate of dynamical mass, $\tilde{M}_{\mathrm{d}}$, in place of $M_{\mathrm{d},n}$. As might be expected from comparing Figures 1 and 4, we find that the $M_*/\tilde{M}_{\mathrm{d}}$ relations for different mass bins are largely parallel, but significantly offset from one another. Then, because of correlations between mass and activity, we also find significant gradients in the $M_*/\tilde{M}_{\mathrm{d}}$ relations for the sample as a whole; typically at the 10–20$\sigma$ level. The size of differential biases as a function of all of $D_n(4000)$, H$\delta$ EW, H$\alpha$ EW, age, and SSFR are on the order 0.2–0.4 dex. We note in particular that the relatively strong gradient in $\tilde{M}_{\mathrm{d}}/M_*$ with H$\alpha$ EW noticed by Drory, Bender & Hopp (2004) disappears when we use the structure corrected dynamical mass estimator, $M_{\mathrm{d},n}$, in place of the simple estimate $\tilde{M}_{\mathrm{d}}$; this apparent bias seems to be more closely linked to structure than to H$\alpha$ emission *per se*. *We thus find a very good correspondence between $M_*$ and $M_{\mathrm{d},n}$ for galaxies in the G09 sample, but only provided we account for structural and dynamical non-homology.*

As a final point, Bell & de Jong (2001) have shown that $M_*/L$ and color are strongly correlated in both the optical and the NIR. This implies that a single color is enough to make a reasonable estimate of $M_*/L$ (see also, e.g., Longhetti & Saracco, 2009; Gallazzi & Bell, 2009). These kinds of color relations have since been widely used for high redshift studies. In Chapter IV, we have shown that the values of $M_*/L$ that we use here correlate very strongly with ${}^{0.1}(g-i)$ color (the scatter around this relation is just 0.10 dex). If we use this relation to predict $M_*/L$ for galaxies in the G09 sample, we again find very good correspondence between $M_*$ and $M_{\mathrm{d},n}$ — in fact, the rms scatter in $M_*/M_{\mathrm{d},n}$ is unchanged.

That is, at least from the point of view of consistency between $M_*$ and $M_{\mathrm{d},n}$, it would seem that $M_*/L$s estimated on the basis of a single color are not significantly worse than estimates based on full SED fits. This is significant because the SEDs that were used to derive the $M_*/L$s were corrected for emission lines using the SDSS spectra. This kind of correction is not practicable for, for example, high redshift studies. Further, we have repeated our analysis using the $M_*/L$s derived from the SDSS spectra by Kauffmann et al. (2003a), and find similarly good agreement between $M_*$ and $M_{\mathrm{d},n}$: at least on average, it would appear that color-derived $M_*/L$s are just as good as those derived from optical spectroscopy (see also Gallazzi & Bell, 2009).

# 6 Discussion

## 6.1 Comparison to Previous Studies

In comparison to other studies of the relation between stellar and dynamical mass estimates based on SDSS data, we find considerably less variation in $M_*/M_{\mathrm{d},n}$ with mass. In the case of Drory, Bender & Hopp (2004), this difference is simply due to the fact that we account for non-homology in the derivation of $M_{\mathrm{d},n}$; using the simple mass estimate $\tilde{M}_{\mathrm{d}}$, we have verified that we are able to reproduce their results. In the case of Gallazzi et al. (2006), there is the additional complication that they use qualitatively different measures of total flux and size. Specifically, the use the Petrosian magnitudes and half-light radii given in the basic SDSS catalog, which are derived directly from the observed curves of growth. Again, we have verified that we can reproduce their results using the same measurements.

Gallazzi et al. (2006) also split their galaxy sample into bins of Sérsic index, and find similar slopes to the $M_*$–$\tilde{M}_{\mathrm{d}}$ relation for each subsample; they find that the logarithmic slope of the relation varies from 0.847 to 0.801 between $n = 3$ and $n = 5.5$. The size of this variation is entirely consistent with the results we have shown in Figure 3. They use this fact to argue that non-homology does not have a significant impact on the slope of the global $M_*$–$M_{\mathrm{d},n}$ relation. However, as we have also shown in Figure 3, while the $M_*$–$\tilde{M}_{\mathrm{d}}$ relations for each bin in $n$ are parallel, they are significantly offset from one another. It is this through this offset, combined with a correlation between $M_*$ and $n$, that non-homology affects the slope of the $M_*$–$M_{\mathrm{d},n}$ relation; Gallazzi et al. (2006) make no mention of such an offset. While our conclusions differ with those of Gallazzi et al. (2006), our results are thus not obviously inconsistent.

Cappellari et al. (2006) have also argued against the idea that non-homology has an important impact on dynamical mass estimates. This argument was based on dynamical mass-to-light ratios derived from detailed 2D and 3D modeling 25 structurally early type galaxies from the SAURON sample. Cappellari et al. (2006) compared the dynamical mass-to-light ratios, so derived, to the simple virial mass estimator $(\tilde{M}_{\mathrm{d}}/L) \propto \sigma^2 R_{\mathrm{DeV}}/L_{\mathrm{DeV}}$, where $R_{\mathrm{DeV}}$ and $L_{\mathrm{DeV}}$ were derived from De Vaucouleurs profile fits, and found no evidence for an $n$-dependent offset between these two quantities. The fact that Cappellari et al. (2006) use De Vaucouleurs-fit sizes and magnitudes is significant. As we have shown in Ap-

pendix B, the covariances between $m_{\rm tot}$, $\Theta_{\rm e}$, and $n$ mean that the combination $M_{{\rm d},n}/L \propto K_V(n)\sigma^2 R_{\rm e}/L$ is remarkably robust to errors in the Sérsic- fit parameters, provided that 1.) $m_{\rm tot}$, $\Theta_{\rm e}$, and $n$ are self-consistently derived, and 2.) non-homology is taken into account when deriving $M_{{\rm d},n}$ via the term $K_V(n)$. Assuming a De Vaucouleurs profile (*i.e.*, $n = 4$) to derive $m_r$ and $\Theta_{\rm e}$ for all galaxies therefore guarantees that $\tilde{M}_{\rm d} \approx M_{{\rm d},n}$. (This point is amply demonstrated in Figure B.1.) The fact that Cappellari et al. (2006) did not find evidence for an $n$-dependent offset between these two dynamical mass-to-light estimates is thus directly linked to their use of De Vaucouleurs-fit sizes and masses.

Note that both Prugniel & Simien (1996) and Trujillo et al. (2004) have made a very similar argument for the importance of non-homology in estimating dynamical masses as we have made in Sections 3 and 4, based on dynamical mass-to-light ratios derived from the fundamental plane. Our analysis based on the correspondence between $M_*$ and $M_{{\rm d},n}$ is complementary to theirs in two ways. First, their analyses were specific to early type galaxies; we have thus extended their result to the general galaxy population. Secondly, both authorss focussed on $M_{\rm d}/L$, rather than $M_*/M_{{\rm d},n}$; that is, neither of these authors considered the relation between galaxies' stellar and dynamical masses.

## 6.2 Interpretation

Turning now to the interpretation of our results, the remarkable consistency between stellar and dynamical mass estimates shows two things. First, it strongly suggests that the measurements of $M_*$ and $M_{{\rm d},n}$ are both meaningful and relatively robust. In particular, our results indicate that it is possible to derive stellar mass estimates without strong differential biases as a function of age, dust, SSFR, or $M_*/L$, based only on broadband optical photometry (or indeed on a single optical color); *i.e.*, without fitting to spectra or including restframe NIR data.

Secondly, it implies that intrinsic variations in the stellar-to-dynamical mass ratio (due to, e.g., variations in the dark–to–stellar mass ratio, variations in the IMF, or dynamical differences beyond the simple non-homology considered here) as a function of stellar mass, galaxy structure, and star formation rate/history are either small, or conspire to leave the inferred values of $M_*/M_{{\rm d},n}$ relatively unchanged. We discuss three separate aspects of this result below.

### *6.2.1 Inferences From the $M_*$–$M_{{\rm d},n}$ Relation*

Since $M_{{\rm d},n}$ is an estimate of total mass, it can only be interpreted as placing a solid upper limit on the true stellar mass. Because we have no *a priori* means of separating out the relative contributions of luminous and non-luminous mass to $M_{{\rm d},n}$, the interpretation of the $M_*$–$M_{{\rm d},n}$ relation is complicated by degeneracies between the relative contributions of gas and dark matter, as well as uncertainties in the low mass shape of the IMF. We address each of these points in turn below.

The simplest way to interpret the non-linearity of the $M_*$–$M_{{\rm d},n}$ relation is as indicating a greater central dark matter fraction for higher mass galaxies, in qualitative agreement with theoretical expectations. Using simple arguments based on

the observed dynamics of elliptical galaxies, Franx (1993) and Kochanek (1994) have argued that accounting for a dark matter halo implies $|\log M_*/M_{\mathrm{d},n}| \sim 0.14$–$0.18$ dex. This would go a long way towards explaining the $\sim 0.23 \pm 0.03$ offset that we have observed.

We can estimate gas masses using the prescription given by Zhang et al. (2009). These authors have used a sample of relatively low mass SDSS galaxies with literature HI masses to derive a prescription for $M_{\mathrm{HI}}/M_*$ as a function of $(g-r)$ color and stellar surface density. Using this prescription to derive baryonic mass estimates, $M_{\mathrm{bar}} = M_* + M_{\mathrm{HI}}$, reduces the size of the offset between $M_{\mathrm{bar}}$ and $M_{\mathrm{d},n}$ by 0.05 dex to $-0.18$ dex, and brings the logarithmic slope of the $M_{\mathrm{bar}}/M_{\mathrm{d},n}$ relation to 0.95. The fact that the Zhang et al. (2009) relation has been derived for very different galaxies to the ones we consider here means that this result should be interpreted with caution. Even so, it is striking that, taken together, the estimated contributions of HI and dark matter almost perfectly explain the observed offset between $M_*$ and $M_{\mathrm{d},n}$, and imply only a mild trend in $M_{\mathrm{bar}}/M_{\mathrm{d},n}$ with mass: $M_{\mathrm{bar}}/M_{\mathrm{d},n} \propto M_{\mathrm{bar}}^{-0.05}$.

Then there is the matter of the IMF. The effect of adopting a Salpeter (1955) IMF rather than that of Chabrier (2003) would be approximately to scale all our values of $M_*$ up by 0.22 dex. For a linear $M_*$–$M_{\mathrm{d},n}$ relation (which our data are only marginally consistent with), this would leave virtually no room for dark matter or gas in the centers of galaxies in our sample. For the slightly less-than-linear relation preferred by our data, this would imply that $M_* > M_{\mathrm{d},n}$ for galaxies with $M_* \lesssim 10^{11}$ $\mathrm{M}_\odot$, which is logically inconsistent. Thus we can say that, at best, our results are only marginally consistent with a Salpeter (1955) IMF. Accounting for dark matter, our results are also weakly inconsistent with a 'diet Salpeter' IMF, and completely consistent with a Kroupa (2001) or Chabrier (2003) IMF. (Parenthetically, we also note that the results shown in Figure 9 can also provide a weak constraint on variations in the IMF as a function of star formation rate/history.)

Finally, we note that the observed scatter around the $M_*$–$M_{\mathrm{d},n}$ relation is rather small: just 0.13 dex. We argue in Appendix B that the ratio $M_*/M_{\mathrm{d},n} \propto \sigma^2 R_{\mathrm{e}}/L$ is remarkably insensitive to errors in the Sérsic-fit parameters, provided that they are consistently derived, and that dynamical non-homology is taken into account. This implies that the uncertainties in $M_*/M_{\mathrm{d},n}$ are dominated by errors in the measurement of $\sigma_0$ and $M_*/L$. The mean formal uncertainty in $\sigma$ for our galaxy sample is 0.034 dex. We estimate the mean random error in $M_*/L$ to be on the order of 0.1 dex; this is the random scatter between the SED-fit $M_*/L$s used here and the spectrally derived $M_*/L$s given by Kauffmann et al. (2003a).[9] Adding these errors in quadrature (*i.e.* neglecting correlations between $\sigma$ and $M_*/L$ at fixed $M_*$) produces an uncertainty in $M_*/M_{\mathrm{d},n}$ of 0.12 dex. This would imply that, at fixed $M_*$, the intrinsic scatter in $M_*/M_{\mathrm{d},n}$ is potentially very small indeed: $\lesssim 0.04$ dex.

[9]Note that this is almost certainly an underestimate of the 'true' random uncertainty in $M_*/L$. Including NIR data (where the stellar population models are the most uncertain), and properly accounting for propagation of uncertainties in stellar population models and the IMF, Conroy, Gunn & White (2009) argue that the uncertainties in $M_*/L$ are on the order of 0.3 dex.

### *6.2.2 Comparison with Fundamental Plane Studies*

As we have mentioned in the Introduction, the fundamental plane can be thought of as measuring the variation in the dynamical mass-to-light ratios of early type galaxies as a function of velocity dispersion, luminosity, or mass (see, e.g., Dressler et al., 1987; Jørgensen, Franx & Kjaergaard, 1996). How do our derived values of $M_{\mathrm{d},n}$ compare to those derived from the fundamental plane? To address this question, we selected the non-emission galaxies from within the G09 sample with $n > 2.5$. For these galaxies, we find $M_{\mathrm{d},n}/L \propto \sigma^{0.88\pm0.06}$.

For comparison, Jørgensen, Franx & Kjaergaard (1996) find $\sigma^2 R_{\mathrm{DeV}}/L_{\mathrm{DeV}} \propto \sigma^{0.86}$, where again, $R_{\mathrm{DeV}}$ and $L_{\mathrm{DeV}}$ have been derived via De Vaucouleurs fits. As we argued in Section 6.1 above, the covariance between the fit values of $m_r$, $\Theta_{\mathrm{e}}$, and $n$ mean that $\sigma^2 R_{\mathrm{DeV}}/L_{\mathrm{DeV}} \approx M_{\mathrm{d},n}/L$. Our structure-corrected dynamical mass estimates are thus in good agreement with those derived from the fundamental plane. (Parenthetically, we also note that Cappellari et al. (2006) found that their dynamical mass-to-light ratios derived from detailed dynamical fits scaled as $\sigma^{0.82}$, which is also consistent with our results.)

We have also considered how the dynamical-to-stellar mass ratio, $M_{\mathrm{d},n}/M_*$ varies with $\sigma$ for this same sample of early type galaxies: we find $M_{\mathrm{d},n}/M_* \propto \sigma^{0.50\pm0.06}$. This would suggest that less than half of the tilt of the fundamental plane is due to variations in the mass-to-light ratios of early type galaxies as a function of $\sigma$ (*cf.*, e.g., Prugniel & Simien, 1996; Trujillo et al., 2004; Allanson et al., 2009). We present this result only for completeness; proper interpretation of this result requires much more detailed analysis, and is beyond the scope of this work.

### *6.2.3 $K_V(n)$ and Dynamical Mass Estimates*

The inclusion of the term $K_V(n)$ makes the structure-corrected dynamical mass estimator $M_{\mathrm{d},n}$ explicitly model dependent.[10] Further, this prescription for $K_V(n)$ has been derived under very simple and idealized assumptions (*viz.* a single component, spherical, and dynamically isotropic distribution), and so can only be regarded as approximate.

That said, more sophisticated dynamical models can give an indication as to how large these effects might be. For the case of anisotropy, the effects on the value of $K_V(n)$ are on the order of $\lesssim 0.1$ dex, and become less important for larger $n$ (see, e.g., Ciotti & Lanzoni, 2001). Further, Bertin, Ciotti & Del Principe (2002) argue that the galaxy dynamics (or, more accurately, the value of $K_V$) close to the galaxy center are in principle rather sensitive to the precise shape of the total mass distribution. Their results suggest that this effect is on the order of $\lesssim 0.1$ dex (see their Figure D.1).

In order to probe the dependence of our conclusions on the assumed form of $K_V(n)$, we have also trialed using an alternate prescription for $K_V(n)$, given by Cappellari et al. (2006, see their Equation 20). The main difference between this prescription and the one given in Equation 3 is that it is phrased in terms of

[10] By the same token, the simple dynamical mass estimator $\tilde{M}_{\mathrm{d}}$ is also model dependent, inasmuch as it assumes homology, which is patently wrong.

the observed velocity dispersion within the effective radius, $\sigma_{\rm e}$, rather than the central velocity dispersion $\sigma_0$. This prescription thus has a different dependence on dynamical isotropy and the dark matter profile, and so provides an indirect means of probing the importance of these effects. Using the Cappellari et al. (2006) prescription, we find qualitatively and quantitatively similar results: we find that $M_* \propto M_{{\rm d},n}^{0.85}$; we still see that $M_*/M_{{\rm d},n}$ depends on $M_*$ at fixed $n$, and on $n$ at fixed $M_*$ (although this dependence is somewhat shallower); we see no trends in $M_*/M_{{\rm d},n}$ with apparent magnitude or redshift, but a weak trend with observed size; and, at fixed mass, we see no statistically significant trends in $M_*/L$ with stellar population parameters.

That is, while it is virtually certain that the model used to derive the prescription for $K_V(n)$ given in Equation 3 is wrong in several important respects, it seems unlikely that accounting for those effects that are ignored in the model would have a drastic effect on our results and conclusions. As we have repeatedly stressed, detailed dynamical modeling is necessary better constrain the 'true' values of $K_V$ for individual galaxies.

## 7 Summary

The central focus of this work has been the degree of consistency between stellar and dynamical mass estimates, based on the latest generation of data products from the SDSS. We have shown that structural differences in galaxy dynamics can have a large impact on the estimated values of dynamical mass, and so on the degree of correspondence between stellar and dynamical mass (Section 3; Figures 1 and 4). Provided we account for structure-dependent differences in galaxy dynamics (using the term $K_V(n)$, as defined in Equations 2 and 3), we find very good agreement between the inferred stellar and dynamical masses of galaxies within the G09 sample.

Our analysis is based on the carefully-constructed satellite/central galaxy sample of G09, making selection effects a major potential concern. However, we find no signs of major differences in the relation between $M_*$ and $M_{{\rm d},n}$ for central/satellite or non-/emission galaxies within the sample, suggesting that our results are not seriously affected by selection effects (Section 5.2; Figure 8). Moreover, we find qualitatively and quantitatively similar results analyzing a more general sample of $0.035 < z < 0.08$ galaxies, using the best-fit Sérsic parameters given in the NYU VAGC, or using the De Vaucouleurs/exponential `model` fit parameters given in the basic SDSS catalog (Appendix B).

We find that the ratio $M_*/M_{{\rm d},n}$ varies with both $M_*$ and $n$ (Sections 3.2 and 4.2; Figures 3 and 6). While the apparent $n$-dependence of $M_*/M_{{\rm d},n}$ is sensitive to the assumed form of $K_V(n)$, changing $K_V(n)$ cannot affect the result that $M_*/M_{{\rm d},n}$ varies with $M_*$ at fixed $n$. Without spatially resolved dynamical information for individual galaxies, however, we cannot determine whether the apparent mass-dependence of $M_*/M_{{\rm d},n}$ is caused by some mass-dependent difference in galaxy dynamics, rather than a genuine physical difference in the stellar-to-dynamical mass ratios of galaxies with different masses.

Similarly, while we have shown very good agreement between stellar and dynamical masses for SDSS galaxies, we cannot unambiguously prove that neither of these quantities suffers from systematic biases. On the other hand, using the Sérsic-fit parameters given by G09, we do not see any systematic variation in $M_*/M_{\mathrm{d},n}$ with observed properties like apparent magnitude, apparent size, or redshift (Section 5.1; Figure 7). This is not true if we use the Sérsic-fit parameters given in the NYU VAGC, which has been shown to suffer from systematic errors arising due to background over-subtraction. That is, we have the ability to detect these sorts of errors, and do not see evidence for such errors for our sample. Further, we do not see any signs of variation in $M_*/M_{\mathrm{d},n}$ for galaxies with different stellar populations, or for galaxies in different states of activity (*i.e.*, AGN hosts, star forming galaxies, or non-emission galaxies).

These results, together with the good general agreement between $M_*$ and $M_{\mathrm{d},n}$ provide strong circumstantial evidence (but not proof beyond a reasonable doubt) that there are no serious systematic biases in the values of $M_*$ and $M_{\mathrm{d},n}$ that we use here. This implies that the assumption of non-homology gives the wrong dynamical mass. Further, this suggests that there are not strong biases in the $M_*/L$s we have used here: at 99 % confidence, the consistency between $M_*$ and $M_{\mathrm{d},n}$ implies any differential biases in the estimate of $M_*/L$ across a wide range of stellar populations are at the level of $\lesssim 0.12$ dex ($\approx 40$ %).

**Acknowledgments.** This work was supported through grants by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), the Leids Kerkhoven-Bosscha Fonds (LKBF).

# Appendices

## A Validating the SDSS DR7 Velocity Dispersion Measurements

Bernardi (2007) showed that there was an inconsistency between the $\sigma$–$L$ relations for early type galaxies derived using the early data release (EDR) and DR5 SDSS catalogs. Further, she was able to show that the cause for this discrepancy was systematic biases in the DR5 velocity dispersions: in comparison to literature values from HyperLeda, the DR5 measurements slightly but systematically over-estimated the velocity dispersions of intrinsically low-$\sigma$ galaxies. For DR6+, partially in response to the findings of Bernardi (2007), the SDSS velocity dispersion pipelines were substantially revised. The new dispersions have been shown to agree well with the EDR velocity dispersions used by Bernardi et al. (2003a,b), and thus, by implication, with the improved estimates for DR5 derived by Bernardi (2007).[11]

In this Appendix, in order to validate the DR7 velocity dispersions, we present a comparison between the velocity dispersions given in the basic SDSS DR7 catalog to those given by Faber et al. (1989) for elliptical galaxies in their sample.

[11] See http://www.sdss.org/dr7/algorithms/veldisp.html

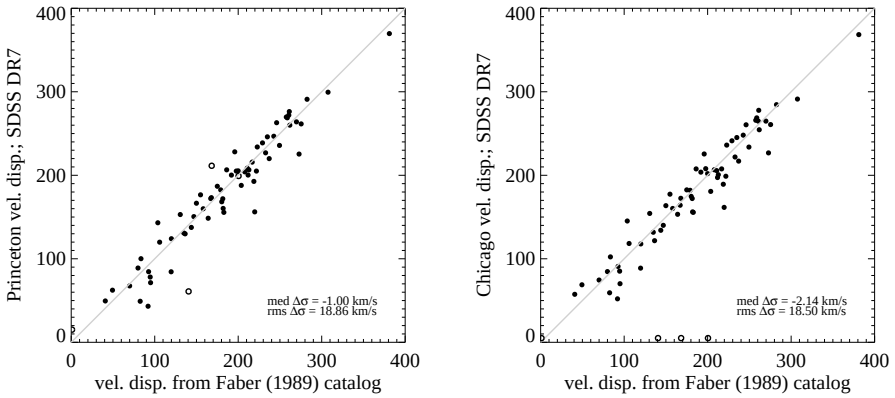


**Figure A.1.** — Comparison between the two different velocity dispersions measurements given in the SDSS DR7 catalog and those given by Faber et al. (1989). — Bernardi (2007) have shown that the DR5 SDSS velocity dispersions suffered from systematic biases in comparison to those from the ENEAR sample as well as earlier SDSS releases. For this reason, the algorithms for estimating velocity dispersions from SDSS spectra were substantially revised for DR6 and later. In each panel of this Figure, we compare one of the two SDSS velocity dispersion measurements to those in the 'seven samurai' catalog (Faber et al., 1989). There are no signs of any systematic problems with the DR7 SDSS velocity dispersions. (Note however that the scatter in these comparisons is significantly higher than would be expected from the formal measurement uncertainties, which are on the order of 3.5 km/s.)

The results of this comparison are shown in **Figure A.1**. The left panel of this Figure shows the comparison for the Princeton or SpecBS values of $\sigma$; the right panel shows that for the Chicago or spectro1d values of $\sigma$. Note that the Chicago algorithm only outputs values of $\sigma$ for those galaxies that are spectroscopically classified as being early type; the three Faber et al. (1989) galaxies at the bottom of the right-hand panel are not classified as being early type, and so are not given Chicago velocity dispersions.

Within both panels, we give the median and rms difference between the SDSS and Faber et al. (1989) velocity dispersion measurements. It is clear from this Figure that neither of the DR7 velocity dispersions suffers from serious systematic biases in comparison to the Faber et al. (1989) measurements. We note, however, that the rms scatter, which is on the order of 19 km/s, is considerably higher than the median formal measurement uncertainty given in the SDSS catalog, which is on the order of 3.5 km/s for the galaxies shown in Figure A.1. That is, it seems that the formal uncertainties on the SDSS velocity dispersions significantly underestimates the true error, at least for the relatively bright galaxies shown here.

## B Selection Effects and Systematic Biases: Repeating our Analysis for a General Galaxy Sample

As we have repeatedly stressed in the main text, the G09 sample that we analyze in the main text is heavily selected. In order to make sure that our conclusions are not unique to the G09 sample, in this Appendix we repeat our analysis for a more general galaxy sample. For this exercise, we have selected $m_{\mathrm{Pet},r} < 17.5$ galaxies with `sciencePrimary` spectra in the range $0.035 < z < 0.08$. As in our main analysis, we also require that the relative error on the velocity dispersion is less than 10 %, and that $\sigma_{\mathrm{ob}} > 75$ km/s; these selections effectively limit the sample to $M_* \gtrsim 10^{10}$ $\mathrm{M}_\odot$. The additional incompleteness due to our velocity dispersion criteria is less than 10 % for all $M_* > 10^{10.3}$ $\mathrm{M}_\odot$ and $n \gtrsim 3$, but is significant for $n \lesssim 1$ at all masses.

There is one complicating factor in the comparison between this general field sample and the G09 sample that we discuss in the main text. For the field sample, we are forced to rely on either the De Vaucouleurs/exponential `model` fits provided in the basic SDSS catalog, or the Sérsic fits given in the NYU VAGC. (Recall that we need a measure of total magnitude to derive $M_*$, and both an effective radius and a Sérsic index measurement to derive $M_{\mathrm{d},n}$.) Both of these sets of measurements have their faults. The SDSS `model` fits are overly simplistic in that they assume that $n$ is equal to either 1 or 4; this will clearly introduce systematic errors in the fit quantities as a function of (intrinsic) profile shape. The VAGC Sérsic fits are also known to suffer systematic errors (Blanton et al., 2005a), due to background over-subtraction (G09).

Bearing both these issues in mind, in **Figure B.1** we show the relation between $M_*$ and $M_{\mathrm{d},n}$ for our general, field galaxy sample, using either the Sérsic fit parameters from the NYU VAGC (left panel) or the `model` fit parameters from the basic SDSS catalog (right panel). Using either set of parameters, the results for this general sample agree really very well with what we have found for the G09 sample in Figure 4. Moreover, the two panels in Figure B.1 agree remarkably well with one another, even despite the significant and very different systematic errors that each set of measurements suffers from.

How can this be? It turns out that the covariance between Sérsic parameters leaves ratio $M_*/M_{\mathrm{d},n}$ is remarkably robust to both random and systematic errors in the Sérsic fits, provided $M_{\mathrm{d},n}$ is calculated as per Equation 2. To illustrate this, let us compare the G09 and VAGC measurements. Although there are large differences in all three parameters individually, there are tight correlations between $\Delta n$, $\Delta m_{\mathrm{tot}}$, and $\Delta R_{\mathrm{e}}$. (Here and in what follows the '$\Delta$' implies the difference between the VAGC and Gou et al. (2009)-derived value, in the sense of VAGC-minus-G09.) Now, $M_*$ scales directly with total flux; fitting to $\Delta M_*$ as a function of $\Delta n$, we find that $\Delta \log M_* \propto 0.04 \Delta n$, with an rms scatter in $\Delta \log M_*$ of 0.07 dex. $M_{\mathrm{d},n}$, at least as defined in Equation 2, depends on both the effective radius and Sérsic index. Considering the change in $M_{\mathrm{d},n}$ due to changes in size alone, we find $\Delta \log M_{\mathrm{d},n} \propto 0.11 \Delta n$, with a scatter of 0.07 dex; for the effect due

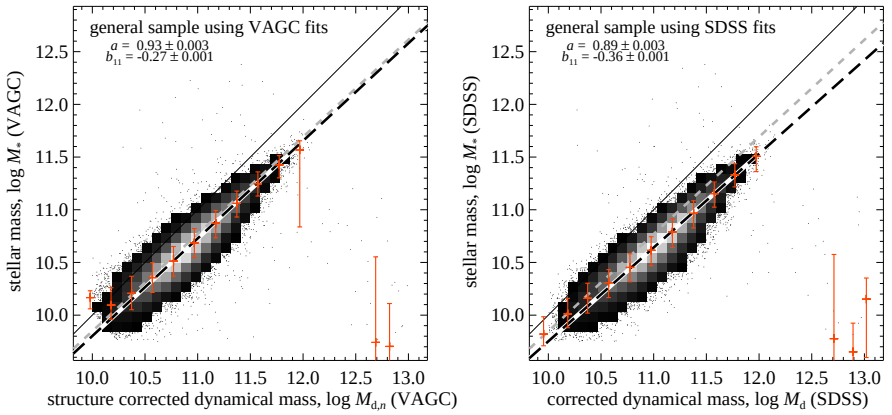


**Figure B.1.** — Comparing dynamical and stellar mass for a general galaxy sample. — For this Figure, we have selected $0.035 < z < 0.08$ galaxies with `sciencePrimary` spectra, $\sigma_{\rm ob} > 75$ km/s, and $\Delta\sigma_{\rm ob}/\sigma_{\rm ob} < 0.1$. For each panel, we have derived $M_{{\rm d},n}$ using either Sérsic structural parameters from the NYU VAGC (Blanton et al., 2005a, left panel), or using De Vaucouleurs/exponential `model` structural parameters from the basic SDSS catalog (right panel). As in other Figures, the solid lines show fits to the data; the points with error bars show the median relations in bins. The grey dashed line shows the $M_*$–$M_{{\rm d},n}$ relation we derive for the G09 catalog, using their Sérsic structural parameters. The general $M_*$–$M_{{\rm d},n}$ relation for field galaxies is very similar to the one we find for the heavily-selected G09 sample.

to changes in the Sérsic index alone, we find $\Delta \log M_{{\rm d},n} \propto -0.06\Delta n$, with an rms scatter of 0.03 dex. Taken together, the overall change in $M_{{\rm d},n}$ scales with $\Delta n$ as $\Delta \log M_{{\rm d},n} \propto 0.04\Delta n$.

Thus we see that the changes in $M_*$ and $M_{{\rm d},n}$ thus have virtually the same dependence on $\Delta n$, leaving the ratio $M_*/M_{{\rm d},n}$ virtually unchanged. Further, the scatter in $\Delta(M_*/M_{{\rm d},n})$ is just 0.04 dex. Using the basic SDSS `model` fits, while we find slightly stronger dependences with $\Delta n$, we still find that the ratio $M_*/M_{{\rm d},n}$ remains very robust. We stress that the above argument only holds if we account for the dynamical effects of structure in the calculation of $M_{{\rm d},n}$: if we removed the $n$-dependence of $M_{{\rm d},n}$ that enters via $K_V(n)$, then we would find that the ratio $\Delta \log(M_*/\tilde{M}_{\rm d}) \propto -0.06\Delta n$, in agreement with the expectation from the analysis immediately above.

To explicitly demonstrate that the observed relation between $M_*$ and $M_{{\rm d},n}$ is not particularly sensitive to the measurements used to derive the values of $M_*$ and $M_{{\rm d},n}$, in **Figure B.2** we show the $M_*$–$M_{{\rm d},n}$ relation for the G09 sample analyzed using structural parameters from the NYU VAGC (left panel) or from the SDSS catalog (right panel). In comparison to Figure 4, the slope of the $M_*$–$M_{{\rm d},n}$ relation for the G09 sample is very similar using any of the three sets of structural parameters: 0.92 for the G09 fits, compared to 0.89 for the NYU VAGC fits, and 0.88 for the SDSS `model` fits. The normalization of the $M_*$–$M_{{\rm d},n}$ relation is slightly more sensitive: at $M_{{\rm d},n} = 10^{11}$ M$_\odot$, we find that $\Delta \log(M_*/M_{{\rm d},n}) =$

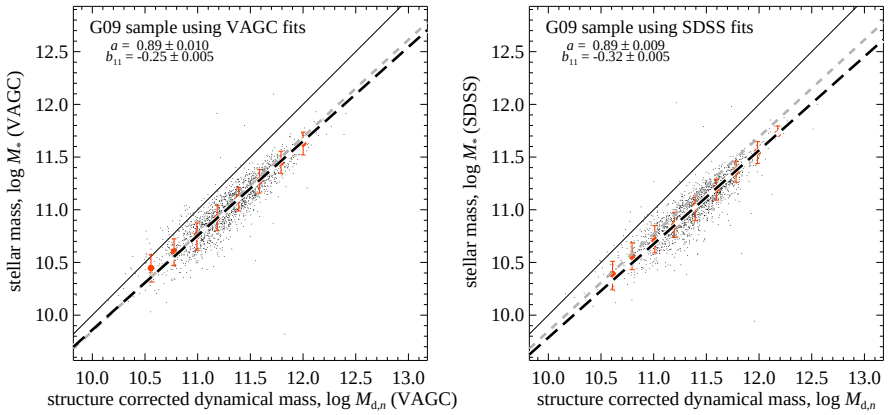


**Figure B.2.** — Comparing dynamical and stellar mass for the G09 sample, using structural fit parameters from the NYU VAGC or the basic SDSS catalog. — All symbols and their meanings are directly analogous to Figures 4 and B.1. In comparison to Figure 4, this Figure differs only in that we have used the Sérsic-fit parameters from either the NYU VAGC (Blanton et al., 2005a) or the De Vaucouleurs/exponential `model` fit parameters from the basic SDSS catalog. The fact that these results agree very well with those shown in Figure 4 shows that the ratio $M_*/M_{\mathrm{d},n}$ is not extremely sensitive to errors in the structural fit parameters. In comparison to Figure B.1, the difference is that we only show galaxies in the G09 sample. The fact that the results in this Figure agree very well with those shown in Figure B.1 show that sample selection effects do not play an important role in our results.

$-0.23$, $-0.24$, and $-0.32$ dex using the G09, VAGC, and SDSS fits, respectively.

In comparison to Figure B.1, the results in Figure B.2 also demonstrate that the $M_*$–$M_{\mathrm{d},n}$ relation for the G09 is very similar to that for a more general field galaxy sample. For example, using structural parameters from the VAGC, the logarithmic slope and intercept of the $M_*$–$M_{\mathrm{d},n}$ relation are $a = 0.89$ and $b_{11} = -0.24$ for the G09 sample, compared to $a = 0.91$ and $b_{11} = -0.27$ for the general galaxy sample.

In **Figure B.3**, we show that galaxies in different states of activity follow very similar $M_*$–$M_{\mathrm{d},n}$ relations. In this Figure, we have split the general galaxy sample into non-emission, star forming, and AGN and composite spectra subsamples using the Brinchmann et al. (2004) spectral classification scheme described in Section 5.2). The logarithmic slopes of the $M_*$–$M_{\mathrm{d},n}$ relation for each subsample agree with one another, as well as with that for the sample as a whole, to within a few percent. We do find that the $M_*$–$M_{\mathrm{d},n}$ relation for the star forming subsample is offset from that for the non-emission and AGN/composite subsamples, at the level of 0.07 dex. However, as for the G09 sample (see Section 5.2), these small differences disappear if we consider only $n > 2.5$ galaxies. We thus conclude that these differences are principally driven by the different distribution of $n$ values within the star forming sample, rather than intrinsic differences in the stellar-to-dynamical mass ratios of star forming galaxies.

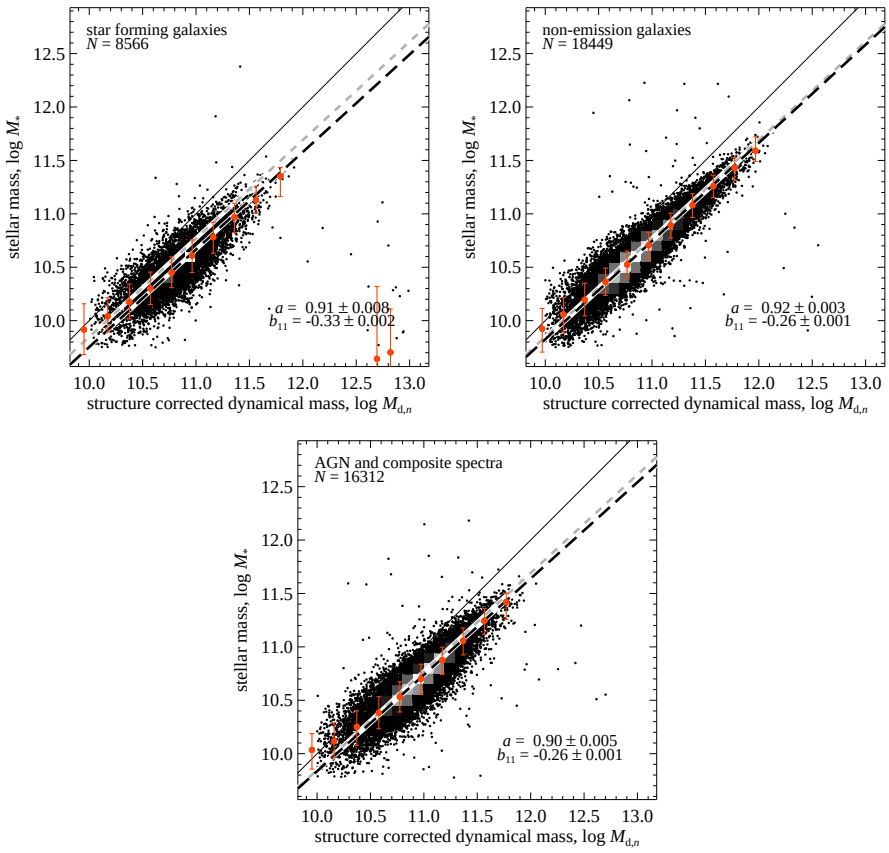


**Figure B.3.** — The $M_*$–$M_{\mathrm{d},n}$ relation for galaxies in different states of activity. — Each panel of this Figure shows the relation between $M_*$ and $M_{\mathrm{d},n}$ for different subsamples of the general $0.035 < z < 0.08$ galaxy population, split according to their spectral classification, and analyzed using the Sérsic fits given in the NYU VAGC. From left to right, we show non-emission galaxies, star forming galaxies, and AGN/composite spectra galaxies; the spectral classifications are those of Brinchmann et al. (2004), which are based on the BPT diagram. In each panel, the heavy dashed line shows the best fit $M_*$–$M_{\mathrm{d},n}$ relation; for comparison, the grey short-dashed line shows the best fit relation for the sample as a whole. While each subsample follows a similar $M_*$–$M_{\mathrm{d},n}$ relation, there is an offset between the different relations, on the order of 0.07 dex. As for the G09 sample, the differences between the different subsamples disappears if we consider only $n > 3$ galaxies. That is, these offsets appear to be due to the different distributions of $n$ within each subsample, rather than intrinsic differences in the values of the stellar-to-dynamical mass ratio for galaxies in different states of activity. This argues the idea that selection effects play a major role in shaping our results.

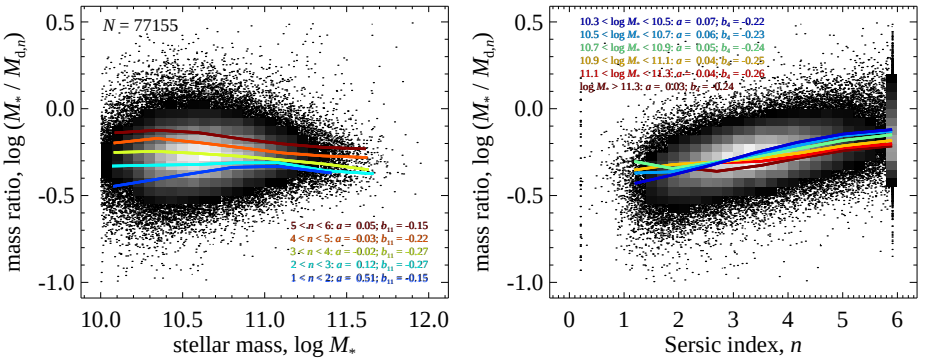


**Figure B.4.** — Separating out the mass- and structure-dependence of the mass ratio $M_*/M_{\mathrm{d},n}$ of a field sample of $0.035 < z < 0.08$ galaxies, showing active and passive galaxies separately. — In analogy to Figure 6, the colored lines in the left panels show the mass-dependence of the mass ratio $M_*/M_{\mathrm{d},n}$ in bins of Sérsic index; those in the right panel shows how $M_*/M_{\mathrm{d},n}$ varies with $n$ in bins of $M_*$. The precise bins are given within each panel, along with the best fit parameters for the relation for each bin. The results in this panel have been derived using the NYU VAGC Sérsic fit parameters. In comparison to Figure 6, we see qualitatively similar behavior for $M_*/M_{\mathrm{d},n}$ as a function of both $M_*$ and $n$ for the general galaxy sample as we do for the heavily-selected G09 sample. Moreover, we point out that this is true for non-emission, star forming, AGN and composite spectra galaxies separately, as well as for the general sample as a whole. Because the NYU fits suffer systematic biases, there are quantitative differences in the results shown in this Figure and those in Figure 6. Despite these quantitative differences, the relatively weak dependence of $M_*/M_{\mathrm{d},n}$ on $n$ for each bin in $M_*$ supports our main result; *viz.*, that accounting for non-homology leads to reasonably good consistency between stellar and dynamical mass estimates.

Finally, in **Figure B.4**, we separate out the $M_*$- and $n$-dependences of $M_*/M_{\mathrm{d},n}$ for the general galaxy sample, analyzed using the NYU VAGC Sérsic-fit parameters. Again, we find that the ratio $M_*/M_{\mathrm{d},n}$ depends on both $M_*$ (at fixed $n$) and on $n$ (at fixed $M_*$). The results in this Figure suggest that the mass-dependence of $M_*/M_{\mathrm{d},n}$ may flatten considerably for $n \lesssim 2$ and $10 \lesssim \log M_*/\mathrm{M}_\odot \lesssim 10.5$ (*i.e.* below the mass limit of the G09 sample).

As we have noted above, there are significant differences in the values of $M_*$ and $n$ given by G09 and in the VAGC. These differences in $M_*$ and $n$ mean that the results in this Figure are not in quantitative agreement with those shown in Figure B.4, even though both datasets show good agreement in the global $M_*$–$M_{\mathrm{d},n}$ relation. In particular, the slope of the $M_*/M_{\mathrm{d},n}$–$n$ relation at fixed $M_*$ is significantly steeper than we find for the G09 sample. This is at least partially due to the bias in the NYU values of $n$; the NYU values are systematically lower than the G09 values, which has the effect of steepening the $M_*/M_{\mathrm{d},n}$–$n$ relation.

With these caveats, the main conclusion to be drawn from Figure B.4 is that accounting for non-homology in the derivation of dynamical masses leads to considerably better consistency between $M_*$ and $M_{\mathrm{d},n}$ (as a function of $n$, and at fixed $M_*$), in agreement with our findings in Section 4. Further, we note that we find similar and consistent behavior in $M_*/M_{\mathrm{d},n}$ at fixed $M_*$ and $n$ for each of the three subsamples shown in Figure B.3, in agreement with our conclusions above.

In summary, then, in this Appendix we have demonstrated two things. First, we have shown that we find very similar results for the G09 sample, analyzed using the results of the Sérsic fits given by G09, and for a more general galaxy sample, analyzed using either the Sérsic fits given in the NYU VAGC or the De Vaucouleurs/exponential `model` fits given in the basic SDSS catalog. Secondly, we have shown that we find very similar results for the G09 sample analyzed using any of these three sets of structural parameters. The most important conclusion to be drawn from these results is that the results we have presented in the main text are not driven, nor particularly sensitive to, selection effects.

## References


Abazajian K N, et al., 2009, ApJS 182, 543

Allanson S P, Hudson M J, Smith R J, Lucey J R, 2009, ApJ 702, 1275

Baldwin, J.A., Phillips, M.M., Terlevich, R., 1981, PASP, 93, 5

Bell E F, de Jong R S, 2001, ApJ 550, 212

Bell E F, McIntosh D H, Katz N, Weinberg M D, 2003, ApJS 149, 289

Bernardi M, 2007, AJ 133, 1954

Bernardi M, Sheth R K, et al., 2003a, AJ 125, 1817

Bernardi M, Sheth R K, et al., 2003b, AJ 125, 1849

Bertin G, Ciotti L, Del Principe M, 2002, A&A 386, 149

Blanton M R, Schlegel D J, Strauss M A, et al., 2005, AJ 129, 2578

Blanton M R, Eisenstein D, Hogg D W, Schlegel D J, Brinkmann J, 2005, ApJ 629, 143

Brinchmann J, Charlot S, White SDM, Tremonti C, Kauffmann G, Heckman T, Brinkmann J, 2004, MNRAS 351, 1151

Bruzual G & Charlot S, 2003, MNRAS 344, 1000

Cappellari M et al., 2006, MNRAS 366, 1126

Chabrier G, 2003, ApJ 586, L133

Ciotti L & Lanzoni B, 2001, A&A 321, 724

Connolly A J, Szalay A S, 1999, AJ 117, 2052

Conroy C, Gunn J E, White M, 2009, ApJ 699, 486

Djorgovsky S & Davis M, 1987, ApJ 313, 59

Dressler A, Lynden-Bell D, Burstein D, Davies R L, Faber S M, Terlevich R, Wegner G, 1987, ApJ 313, 42

Drory N, Bender R, Hopp U, 2004, ApJ 616, L106

Faber S M Wegner G, Burnstein D, Davies R L, Dressler A, Lynden-Bell D, Terlevich R J, 1989, ApJS 69, 763

Franx M, 1993, in *Galactic Bulges* (Dordrecht: Kluwer), Dejonghe H & Habing H J, ed.s, 243

Franx M, van Dokkum P G, Fórster-Schreiber N M, Wuyts S, Labbé I, Toft S, 2008, ApJ 688, 770

Gallazzi A, Charlot S, Brinchmann J, White S D M, Tremonti C A, 2005, MNRAS 362, 41

Gallazzi A, Charlot S, Brinchmann J, White S D M, 2006, MNRAS 370, 1106

Gallazzi A & Bell E F, 2009, ApJS (accepted; arXiv:0910.1591)

Graves G J, Faber S M, Schiavon R P, 2009, ApJ 698, 1950

Gott J R, 1977, ARA&A 15, 235

Graham A & Colless M, 1997, MNRAS 287, 221

Guo Y, McIntosh D H, Mo H J, Katz N, van den Bosch F C, Weinberg M, Weinmann S M, Pasquali A, Yang Xiaohu, 2009, MNRAS 398, 1129

Jørgensen I, Franx M, Kjaergaard P, 1995, MNRAS 276, 1341

Jørgensen I, Franx M & Kjaergaard P, 1996, MNRAS 280, 167

Kannappan S J & Gawiser E, 2007, ApJ 657, L5

Kauffmann G, Heckman T M, White S D M, Charlot S, Tremonti C, Brinchmann J, Bruzual G, Peng E W, Seibert M et al., 2003, MNRAS 341, 33

Kauffmann G, Heckman T M, White S D M, Charlot S, Tremonti C, Peng E W, Seibert M, Brinkmann J, Nichol R C, SubbaRao M, York D, 2003, MNRAS 341, 54

Kauffmann G, Heckman T M, De Lucia G, Brinchmann J, Charlot S, Tremonti C, White S D M, Brinkmann J, 2006, MNRAS 367, 1394

Kochanek C S, 1994, ApJ 436, 56

Kroupa P, 2001, MNRAS 322, 231

La Barbera F, Busarello G, Merluzzi P, de la Rosa I G, Coppola G, Haynes C P, 2008, ApJ 689, 913

Longhetti M & Sarraco P, 2009, MNRAS 394, 774

Maraston C, 2005, MNRAS 362, 799

Padmanabhan N, Seljak U, Strauss M A, Blanton M R, Kauffmann G, Schlegel D J, Tremonti C, Bahcall N A, Bernardi M, Brinkmann J, Fukugita M, Ivezić, Ž, 2004, New Astronomy 9, 329

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Prugniel Ph & Simien F, 1996, A&A 321, 111

Rettura A, Rosati P, Strazzullo V, Dickinson M, Fosbury R A E, Rocca-Volmerange B et al., 2006, A&A 658, 717

Salpeter E E, 1955, ApJ 121, 161

Shen S, Mo H J, White S D M, Blanton M R, Kauffmann G, Voges W, Brinkmann J, Csabai J, 2003, MNRAS 343, 978

Sérsic, J-L, 1963, Boletin de la Asociacion Ar gentina de Astronomia, vol.6, p.41

Sérsic, J-L, 1968, 418, 617 Atlas de Galaxias Australes (Cordoba: Observatorio Astronomico)

Strauss M A, Weinberg S H, Lupton R H, Narayanan V K et al., 2002, AJ, 124, 1810

Taylor E N, Franx M, van Dokkum P G et al., 2009, ApJ 694, 1171 (Chapter II)

Taylor E N, Franx M, van Dokkum P G et al., 2009, ApJS 183, 295 (Chapter III)

Taylor E N, Franx M, Glazebrook K, Brinchmann J, van der Wel A, van Dokkum, 2009, ApJ (submitted)

Thakar A R, Szalay A, Fekete G, Gray J, 2008, CSE 10, 30

Trujillo I, Burkert A, Bell E F, 2004, ApJ 600, L39

van der Wel A, Franx M, Wuyts S, van Dokkum P G, Huang J, Rix H-W, Illingworth I, 2006, ApJ 652, 97

Wolf J, Martinex G D, Bullock J S, Kaplinghat M, Geha M, Muñoz R R, Simon J D, Avedo F F, 2009, MNRAS (submitted; arXiv:0902.2995v3)

Yang X, Mo H J, van den Bosch F C, Pasquali A, Li C, Barden M, 2007, ApJ 671, 153

Yip C W, Connolly A J, Szalay A S et al., 2004, AJ 128, 585

York D G et al., 2000, AJ 120, 2131

Zhang W E, Chang L, Kauffmann G, Hu Z, Catinella B, Shen S, Guo Q, Chang R, 2009, MNRAS 397, 1243

# Hoofdstuk VI

## Nederlandse Samenvatting

### 1 Achtergrond

De vorming en de evolutie van sterrenstelsels is de moeder van alle sterrenkundige problemen. In haar bereik overbrugt ze de kloof tussen kosmologie en de vorming van sterren en planeten binnen sterrenstelsels. De studie naar de evolutie van sterrenstelsels verbindt op die manier de twee grote vraagstukken in de astronomie: hoe is het heelal begonnen, en hoe is voor het eerst leven ontstaan? Elk van de processen die naar vermoed belangrijk zijn in het vormen en ontwikkelen van sterrenstelsels — bijvoorbeeld de vorming van kosmische structuur, het afkoelen en aantrekken van gas; stervorming; sterevolutie; de vorming en groei van zwarte gaten, zogenaamde 'feedback' door zwarte gaten en supernova's; de invloed van de omgeving; interactie tussen sterrenstelsels — vormt een actief veld van theoretische of observationele studie. Op deze manier maakt het onderzoek naar de evolutie van sterrenstelsels niet alleen gebruik van veel astronomische disciplines, maar weet deze ze ook in zich te verenigen.

Sterrenstelsels zijn ingewikkelde monsters. Wat hun voornaamste eigenschappen betreft, zoals massa, lichtkracht, afmeting en stervormingssnelheid, beslaat de verscheidenheid van sterrenstelsels verschillende ordes van grootte. Maar ondanks de brede variatie in ontstaansgeschiedenissen, gedragen sterrenstelsels zich tegelijkertijd zeer fatsoenlijk. Er zijn veel nauwe en goed gedefinieerde 'schalingsrelaties' die deze eigenschappen terugbrengen tot slechts een of twee blijkbaar 'fundamentele' eigenschappen. De stellaire massa (of wellicht dichtheid) lijkt in dit opzicht bijzonder belangrijk: wanneer de massa van een sterrenstelsel bekend is, is het mogelijk om vrijwel alle andere eigenschappen van het stelsel met een opmerkelijke precisie te voorspellen. Het doorgronden van de fysische processen die aan deze schalingsrelaties ten grondslag liggen, is al lange tijd een van de belangrijkste doelen voor theoretici die de evolutie van sterrenstelsels modelleren.

De huidige modellen zijn nog niet in staat om de vorming en evolutie van sterrenstelsels op basis van grondprincipes te beschrijven. De fysische schalen van de verschillende processen reiken van megaparsecs tot luttele kilometers. Rekenkundig gezien kunnen deze processen alleen bij benadering beschreven worden. Erger nog, van nagenoeg geen van de belangrijke processen is duidelijk hoe het zich in isolatie manifesteert, laat staan in relatie tot andere processen; de onzekerheden zijn enorm. In plaats daarvan proberen de huidige modellen een statistische beschrijving van de eigenschappen van sterrenstelsels af te leiden (zoals de massa-

verdeling en stervormingssnelheden). Hiertoe gebruiken ze empirische resultaten (zoals bijvoorbeeld de stervormings'wet') en verstelbare parameters (bijvoorbeeld de hoeveelheid energie die vrijkomt bij een supernova explosie). Wanneer de resultaten van het model verschillen van wat wordt waargenomen, wordt het model bijgesteld om de discrepantie terug te brengen. Met de grote onzekerheden en vele benaderingen die er zijn, hebben theoretici geen tekort aan knoppen en hendels om hun modellen bij te stellen. Op dit moment is het veld van de vorming en evolutie van sterrenstelsels wetenschappelijk gezien voornamelijk gebaseerd op waarnemingen.

In deze context is het doel van dit proefschrift om te voorzien in nieuwe observationele limieten voor de evolutie van sterrenstelsels, waarmee de zojuist beschreven modellen getest kunnen worden. In het bijzonder *is het doel geweest om de groei en evolutie van sterrenstelsels gedurende 80 % van de geschiedenis van het heelal in kaart te brengen en te quantificeren, in termen van aantallen, stervormingsactiviteit en structuur.*

## 2 Terugkijken in de Tijd

Omdat de snelheid van het licht eindig is, kost het tijd voor licht om van de ene plek naar de andere te reizen. Dit betekent dat als je verder weg kijkt, je delen van het heelal ziet, die jonger zijn dan het hier en nu. *Door de eigenschappen van sterrenstelsels over een groot bereik afstanden te vergelijken, is het dus mogelijk om direct te zien hoe sterrenstelsels zich ontwikkelen.*

De technische moeilijkheid van dit soort dataverzameling, ofwel 'terugblik onderzoek' genaamd, zit hem in het bepalen van de afstand van een object. Op hun tocht door de kosmos, worden fotonen beinvloed door de uitdijing van het heelal; hun golflengte wordt hierdoor uitgerekt, ofwel roodverschoven. Het is mogelijk om uit het waargenomen lichtspectrum van een sterrenstelsel de roodverschuiving '$z$' te bepalen, en daarmee dus ook de afstand. Maar het verzamelen van grote hoeveelheden van dit soort spectra is observationeel gezien duur. Het alternatief is om waar te nemen op zoveel mogelijk golflengten door brede filters te gebruiken. Hoe meer (of hoe nauwere) filters gebruikt worden, des te beter kan het spectrum nagebootst worden. Deze methode wordt 'fotometrie' genoemd. Roodverschuivingen kunnen dan worden afgeleid aan de hand van een wat grovere weergave van het spectrum. Vergeleken met spectroscopie maakt fotometrie het mogelijk om veel grotere verzamelingen van zwakke, afgelegen sterrenstelsels te analyseren, zij het met aanzienlijk grotere onzekerheden.

De technieken die gebruikt worden om fotometrische roodverschuivingen te bepalen, zijn voornamelijk gebaseerd op de optische kenmerken in het spectrum. Van bronnen die zich op roodverschuiving bevinden groter dan $z \sim 1$ (wat overeenkomt met hoe het heelal er 7 Gyr geleden uitzag, op 50 % van zijn huidige leeftijd), zijn deze kenmerken roodverschoven naar het nabij-infrarode (NIR) gebied. *Waarnemingen in het nabij infrarood zijn dus de sleutel om toegang te krijgen tot het heelal op* $z > 1$.

Met dit in het achterhoofd, beschrijft **Hoofdstuk II** een nieuwe, publiekelijk toegankelijke catalogus van bronnen die geselecteerd zijn in het NIR in het Extended Chandra Deep Field South (ECDFS), samengesteld als deel van het MUltiwavelength Survey by Yale–Chile (MUSYC). Het ECDFS is een van de meest vooraanstaande velden voor het terugblik onderzoek. Het is het doelwit geweest van tal van grote projecten, waardoor er nu waarnemingen beschikbaar zijn in het ultraviolet en het optische regime (breed-, medium-, en nauwe-filters), afbeeldingen verkregen met behulp van de *Hubble Space Telescope* en de *Spitzer Space telescope* en ook radio interferometrie. In een poging om de waarde van de MUSYC ECDFS waarnemingen te maximaliseren, is er een aanzienlijke hoeveelheid tijd en moeite in gestoken om de kwaliteit van de data te optimaliseren en testen. Door NIR waarnemingen te verzamelen van hoge kwaliteit, heeft MUSYC een cruciaal gat gevuld in de tot nu toe aanwezige data van dit belangrijke veld. De MUSYC NIR data vormen het uitgangspunt van drie andere, nog lopende, projecten die aan de huidige hoeveelheid waarnemingen diepe Spitzer data zullen toevoegen, alsook medium-filter optische en NIR data. *Voornamelijk in combinatie met de rijkdom aan aanvullende data, bieden de MUSYC ECDFS data een uitgelezen mogelijkheid om het heelal op $z \lesssim 2$ te verkennen.*

De catalogus bevat meer dan 10.000 bronnen, geselecteerd in het NIR, inclusief een compilatie van spectroscopische roodverschuivingen voor bijna 2.000 sterrenstelsels. Naast de fotometrie, zijn ook catalogi van fotometrische roodverschuivingen en van intrinsieke, ofwel ware kleuren uitgebracht. De fotometrische roodverschuivingen zijn afgeleid met behulp van een nieuw algoritme dat door MUSYC teamleden is ontwikkeld. Deze waarden zijn tot op heden de meest nauwkeurig bepaalde roodverschuivingen gebaseerd op breed-band fotometrie en zijn op dit gebied momenteel het neusje van de zalm. De restframe fotometrie is bepaald met behulp van een nieuwe voorziening genaamd InterRest, die in Hoofdstuk II beschreven en getest is. Ook InterRest is voor het publiek toegankelijk gemaakt.

## 3 De Opkomst van de Rode Sterrenstelsels

In het lokale heelal, zijn de meeste sterrenstelsels 'rood en dood'. Er worden in deze stelsels geen nieuwe sterren meer gevormd en hun licht wordt daarom gedomineerd door oude, rode sterren. Deze passieve sterrenstelsels vormen een nauwe 'rode reeks' in een kleur–magnitude diagram, die goed te onderscheiden is van de 'blauwe wolk' van minder massieve sterrenstelsels die nog steeds sterren vormen. Optische terugblik projecten hebben aangetoond dat deze gescheiden rode en blauwe populaties al bestaan sinds $z \sim 1$. Bovendien hebben ze laten zien dat het aantal rode stelsels met een factor $\sim 2$ is toegenomen tussen $z \sim 1$ en het heden. De totale massa van de blauwe stelsels is daarentegen redelijk constant gebleven. Deze waarneming — een groeiend aantal passieve stelsels en een constant aantal stervormende stelsels — kan enkel begrepen worden als een mechanisme wordt ingeroepen dat de stervorming in massieve stelsels ontwricht en verder voorkomt, zodat een overgang van actief naar passief, ofwel van blauw naar rood mogelijk wordt.

Met het doel om beperkingen op te leggen aan het tijdstip en ook het precieze proces van dit smoren van de stervorming is in Hoofdstuk III de MUSYC ECDFS catalogus gebruikt om de evolutie te quantificeren sinds $z \lesssim 2$, van massieve stelsels in het algemeen en massieve rode stelsels in het bijzonder. De totale hoeveelheid sterrenstelsels met een massa groter dan $10^{11}$ zonsmassa's neemt tussen $z \sim 2$ en $z \sim 1$ met een factor $\sim 2$ toe en is daarna ongeveer constant. De fractie rode stelsels daarentegen groeit van $\sim 33\%$ op $z \sim 2$ naar $\sim 50\%$ op $z \sim 1$ en $\sim 95$ % op $z \sim 0$.

Alle passieve sterrenstelsels zijn rood, maar niet alle rode sterrenstelsels zijn passief. Het aantal massieve, rode stelsels kan daarom gezien worden als een bovenlimiet voor het aantal sterrenstelsels waarvan de vorming van sterren werkelijk gestopt is. De resultaten van Hoofdstuk III laten dus zien *dat maximaal 1/6 van de massieve sterrenstelsels in het lokale heelal, reeds op $z \sim 2$ passief was; tenminste de helft van die stelsels vormden op $z \sim 1$ nog actief nieuwe sterren.* Wat het precieze mechanisme dat verantwoordelijk is voor het stopzetten van de stervorming ook mag zijn, dit is het moment waarop het werkzaam is.

De studie die in Hoofdstuk III gepresenteerd wordt, was de eerste in zijn soort die een gedetailleerde, systematische studie bood van de onzekerheden die gepaard gaan met de experimentele werkwijze. Deze analyse is niet alleen nuttig voor het interpreteren van de specifieke resultaten die in Hoofdstuk III gepresenteerd worden, maar ook als leidraad voor het ontwerp van toekomstige projecten en experimenten. Zo wordt er bijvoorbeeld aangetoond dat de nauwkeurigheid van de resultaten gelimiteerd wordt door systematische fouten die samenhangen met de analyse van de data en door onzekerheden in de fundamentele fotometrische calibratie van de waarnemingen — en niet door statistische fouten die gerelateerd zijn aan de hoeveelheid of kwaliteit van de data. *Dit betekent dat, in plaats van betere waarnemingen, toekomstige projecten en experimenten allereerst betere methoden en technieken vereisen om de resultaten die in Hoofdstuk III gepresenteerd worden, te verbeteren.*

## 4 De Groei van Passieve Sterrenstelsels

De zojuist beschreven resultaten laten zien dat op een roodverschuiving $z \sim 2.3$, sommige sterrenstelsels het overgrote deel van hun massa al verzameld hadden, *en* ook gestopt waren met het vormen van nieuwe sterren. In termen van stellaire populaties, lijken deze passieve stelsels volledig ontwikkeld, en dat op een tijdstip dat het heelal slechts op 20 % van zijn huidige leeftijd was. Er is echter een groot verschil tussen de massieve sterrenstelsels op $z \sim 2.3$ en die van $z \sim 0$: de stelsels op hoge roodverschuiving zijn veel kleiner. Ze hebben afmetingen die 3–10 keer kleiner zijn dan lokale stelsels met dezelfde massa. Dit betekent dat *elk* van die stelsels behoorlijk moet groeien om zich te kunnen ontwikkelen tot het soort stelsels dat we in het lokale heelal zien — en dat nadat hun stervorming is gestopt.

Deze bewering is getest in **Hoofdstuk IV**, dat een speurtocht beschrijft naar stelsels in het lokale heelal met afmetingen en massa's die te vergelijken zijn met die van de compacte stelsels op $z \sim 2.3$. Deze speurtocht is uitgevoerd aan de

hand van data van de Sloan Digital Sky Survey (SDSS); een uitgebreid spectroscopisch overzicht van de lokale populatie sterrenstelsels. Het SDSS algoritme dat ervoor zorgt dat bronnen met een hoge oppervlaktehelderheid worden uitgesloten van de spectroscopische selectie gaat uit van twee selectiecriteria. Deze criteria zorgen ervoor dat, zelfs als er lokale stelsels zouden zijn met afmetingen en massa's die vergelijkbaar zijn met de compacte stelsels op $z \sim 2.3$, ze niet in de SDSS catalogus van bronnen op $z < 0.05$ zouden verschijnen. Om deze reden beschrijft hoofdstuk IV een zoektocht naar massieve, compacte sterrenstelsels in het roodverschuivingsgebied tussen $0.066 < z < 0.12$, waar de SDSS catalogus vrijwel compleet zou moeten zijn.

Nadat sterrenstelsels die duidelijk verdacht waren wat betreft hun afmetingen/of massabepaling, uit de selectie verwijderd waren, bleef een verzameling van 63 compacte kandidaten over. De snelheidsdispersies van deze stelsels zijn grofweg consistent me hun afgeleide kleine afmetingen en hoge massa's. Echter, geen van deze stelsels heeft een afmeting of massa die overeenkomt met de stelsels die op $z \sim 2.3$ gevonden zijn. De resultaten in Hoofdstuk IV bevestigen dus dat *de massieve, compacte sterrenstelsels die op* $z \sim 2.3$ *worden waargenomen, niet zomaar ontbreken in de SDSS catalogus, maar dat ze simpelweg niet voorkomen in het lokale heelal.*

De grote systematische onzekerheden in de $z \sim 2.3$ metingen negerend, bieden deze resultaten overtuigend bewijs dat massieve sterrenstelsels niet 'monolithisch', dat wil zeggen niet in een klap gevormd worden: de compacte stelsels op $z \sim 2.3$ zijn nog niet volledig gevormd, maar moeten tussen toen en nu nog behoorlijk in grootte groeien. Het blijkt dat een aanzienlijke fractie ($\lesssim 50$ %) van deze groei al rond $z \sim 1.5$ heeft plaatsgevonden. Het mechanisme verantwoordelijk voor deze sterke, structurele evolutie is niet bekend. Wel kan aan de hand van een simpel statistisch argument, namelijk het feit dat *elk* van de $z \sim 2.3$ sterrenstelsels een aanzienlijke groei moet doormaken, een stochastisch proces zoals grote samensmeltingen van sterrenstelsels uitgesloten worden als oorzaak.

## 5 Stellaire Massa's van Sterrenstelsels

Al het werk dat hierboven is beschreven, is afhankelijk van onze bekwaamheid om stellaire massa's af te leiden uit de breedband fotometrie van sterrenstelsels. Het bepalen van de massa van een sterrenstelsel is een lastige aangelegenheid. De gangbare methode is om modellen te gebruiken die de stellaire evolutie beschrijven, gecombineerd met een geparametriseerde beschrijving van een reeks plausibele stervormingsgeschiedenissen. Hiermee kan een acceptabele beschrijving van de waargenomen vorm van het spectrum worden verkregen. De massa en lichtkracht van het model kan dan worden gebruikt om de stellaire massa van het waargenomen stelsel af te leiden. De nauwkeurigheid van dit soort technieken wordt gelimiteerd door bijvoorbeeld dupliciteit van verschillende parameters in de modellen en onzekerheden inherent aan de modellen zelf.

Helaas is het zo goed als onmogelijk om de stellaire massa op een onafhankelijke manier te bepalen, om de fotometrisch afgeleide stellaire massa te testen. In

plaats daarvan, wordt in **Hoofdstuk IV** een vergelijking gepresenteerd tussen de stellaire massa en de *totale* massa, die afgeleid kan worden van de dynamica van het stelsel en ook dynamische massa wordt genoemd. Bij een dergelijke vergelijking is het niet mogelijk om aan te wijzen dat een van de bepalingen goed of fout is, het is alleen te hopen dat de twee methoden consistent met elkaar zijn.

De overeenkomst tussen stellaire en dynamische massa's van een verzameling lokale stelsels van het SDSS blijkt zeer goed te zijn, maar alleen wanneer tijdens het afleiden van de dynamische massa's rekening gehouden wordt met verschillen in de structuur van de sterrenstelsels. Dit heeft het ongelukkige gevolg dat er een expliciete modelafhankelijkheid wordt geintroduceerd tijdens de afleiding van dynamische massa's. Onder dit voorbehoud werden geen significante verschillen gevonden tussen de stellaire en dynamische massa's ten opzichte van een breed scala aan waarneembare grootheden (zoals schijnbare magnitude of lichtkracht), of modelparameters (zoals leeftijd, metalliciteit, of stofgehalte). De afwijking is hoogstens $< 0.12$ dex of 40 % (99 % betrouwbaar). Dit resultaat levert een sterke aanwijzing dat de twee methoden betrouwbaar zijn, inclusief de modellen en metingen waar deze op gebaseerd zijn.

## 6 Conclusies

De belangrijkste resultaten van dit proefschrift zijn:

1. Het samenstellen van een publiekelijk toegankelijke NIR-geselecteerde catalogus van een belangrijk veld voor de terugblik wetenschap, inclusief het opstellen van een lijst afgeleide grootheden zoals fotometrische roodverschuivingen, ware kleuren, en stellaire massa's. Het is te hopen dat deze verzameling data een nuttige hulpbron voor de sterrenkundige gemeenschap zal zijn. Derhalve is de kwaliteit van de data door middel van uitgebreide interne en externe tests zorgvuldig gewaarborgd.

2. Het quantificeren van de evolutie in de hoeveelheid massieve sterrenstelsels in het algemeen en van rode (passieve) sterrenstelsels in het bijzonder, over de laatste 10 Gyr. Deze resultaten bieden belangrijke limieten voor het tijdstip waarop massieve stelsels hun stellaire massa verzamelen en ook wanneer hun stervorming gesmoord wordt. Op deze manier bieden deze resultaten de mogelijkheid om beter in kaart te brengen wat de rol is van samensmeltingen van sterrenstelsels, de feedback van AGN, en het aantrekken van gas tijdens de evolutie van massieve sterrenstelsels.

3. De bevestiging dat massieve sterrenstelsels een sterke groei in afmeting doormaken, zelfs nadat ze gestopt zijn met het vormen van sterren. Het mechanisme voor deze groei is nog niet bekend. Verdere studie naar de evolutie vanaf $z \lesssim 2$ van de afmeting–massa–relatie voor massieve sterrenstelsels zal de gelegenheid bieden om niet alleen af te tasten wat het relatieve belang is van kleine en grote samensmeltingen voor de evolutie van sterrenstelsels, maar ook een beter beeld te krijgen van de recente samensmeltingen van massieve sterrenstelsels.

4. Een demonstratie van het feit dat stellaire massa's die bepaald zijn met behulp van optische fotometrie wederzijds consistent zijn. Omdat de technieken die gebruikt worden om de massa's van sterrenstelsels te bepalen van cruciaal belang zijn in het onderzoek naar de evolutie van sterrenstelsels, is het al even belangrijk om deze technieken te testen en te valideren.

Een terugkerend thema in dit werk was het belang van systematische fouten en onzekerheden. In de afgelopen decennia heeft een reeks technologische en technische doorbraken onze kennis en ons begrip van de evolutie van sterrenstelsels over een tijdspanne die 80 % van de leeftijd van heelal beslaat, simpelweg gerevolutioneerd. Het lijkt er echter op dat deze verkennende fase ten einde loopt; de volgende 10 jaar zullen waarschijnlijk voornamelijk bestaan uit consolidatie en verfijning. In die context reflecteert het groeiende bewustzijn van het belang van systematische fouten de wasdom waartoe het veld van vorming en evolutie van sterrenstelsels gekomen is.

*Translatie:* Maaike Damen

# Curriculum vitæ

I was born on April 2, 1980 in Baltimore, MD (USA) to Australian parents. In 1990, my family moved (back) to Melbourne (AUS). I graduated from Scotch College in Hawthorn in 1997. In 1998, I began a Bachelor of Arts degree at the University of Melbourne, with a major in History. I was recognized on the Dean's Honours List in 1999. The following year, I changed my degree to a combined Bachelor of Arts/Bachelor of Science, with the intention of becoming an astrophysicist. In 2003, I completed an Honours year in the School of Physics; my Honours research project was supervised by Prof. Rachel Webster.

In the course of my undergraduate degree, I spent two (Australian) summers as a vacation student. In 2001/02, I worked at the Australia Telescope National Facility (ATNF; Epping, AUS.) with Bärbel Koribalski and Erwin de Blok. Then, in 2002/03 I went to the Space Telescope Science Institute (STScI; Baltimore, USA), where I worked with Mark Dickinson and Harry Ferguson.

I began my doctoral research at the Leiden Observatory (Sterrewacht Leiden) in September 2004, with Marijn Franx as my supervisor. Parts of this work were completed during visits to Yale University (USA), the Harvard-Smithsonian Center for Astrophysics (Cambridge; USA), Universidad de Chile (Santiago, CHI), and the Max-Planck Institut für Astronomie (Heidelberg, DEU).

I have attended graduate student schools in Brisbane (AUS), Dwingeloo, and Obergurgl (AUT). I have presented posters at international conferences in Marseille (FRA), Santa Cruz (USA), Durham (GBR), Heidelberg (DEU), and Kuala Lumpur (MYS). I have also given talks at meetings in Prague (CZE), Cardiff (GBR), Hayama (JAP), and Melbourne (AUS), as well as colloquia at the University of Melbourne, Swinburne University, and the University of Sydney.

I am deeply committed to teaching. As an Honours student, I was a lab demonstrator in the School of Physics at the University of Melbourne in 2003 and 2004. At the Observatory, I was a teaching assistant for the bachelor subject 'Introduction to Modern Research' in 2005 and 2006. In addition, I have tutored at several of the Colleges connected to the University of Melbourne, and was a tutor for the undergraduate 'Introduction to Climate Change' subject at the University of Melbourne in 2008.

I will soon begin a postdoctoral fellowship with Andrew Hopkins (Head of Science; Anglo-Australian Observatory, AUS) and Prof. Joss Bland-Hawthorn (University of Sydney, AUS), within the GAMA collaboration. This position will be funded by the University of Sydney; I will be based at the University of Melbourne.

Outside of astronomy, I have wide and deep interests across music, books, food, fashion, design, street art, 'art' art, theatre, and film. At different times, I have been a photographer, a political advocate, a tutor, a research scientist, a climbing instructor, an actor and director, and a radio comic.

# Acknowledgments

My thanks go to Eric Gawiser and the entire MUSYC collaboration for giving me the opportunity to work on and with such a fantastic dataset. Ryan, thankyou for showing me the joys of NIR data reduction; Guillermo, thankyou for giving me a chance to pass that on. Thankyou also to Gabriel Brammer and Greg Rudnick for teaching me more than I ever wanted to know about phot-$z$s. In particular, the work presented in Chapter III would never have been possible without the EAZY photometric redshift code developed primarily by Gabriel Brammer; Gabe, I cannot thank you enough.

I also want to express my gratitude to and admiration for the SDSS team; your ongoing contribution to the field is extraordinary. I am particularly grateful to Jarle Brinchman for deriving the stellar mass-to-light ratios used in Chapters IV and V, and to Yicheng Guo for the structural parameters used in Chapter V.

I thank both Eric Bell and Karl Glazebrook for their close collaboration on Chapters II and III and on Chapter IV, respectively.

A good part of this thesis was completed more than halfway around the world. I am deeply grateful to both the Sterrewacht and to the University of Melbourne for making this not only possible, but easy. In particular, I thank Jan Lub for his gracious and generous support; I also thank Prof. Rachel Webster for her ongoing support and mentorship.

I thank the Nederlandse Organisatie voor Wetenschappelijk Onderzoek for funding my graduate research position. I am very grateful to the Leidsche Kerkoven-Bosscha Fonds for providing generous travel support, which allowed me to develop and present my thesis material on five continents. I also thank the Lorentz Center for hosting our regular FIRES 'family meetings', which were always extremely enjoyable and productive events.

My thanks go to my scientific siblings — Ivo, Arjen, Mariska, Stijn, and Maaike — and to Andrew, Greg, Sune, Natascha, Danilo, Tracy, Vy, Gabe, Adam, and Rik, for being both scientific partners and excellent drinking companions.

Olja, Nina, Niruj, Sarah, Simon, Ryan, Rik, Maaike, Isa, Remco, Andrew, Kirsten, Yuri: thank you for your friendship, and your patience, compassion and support in (extended) times of need. I don't know what I would've done without you.

To my R&R buddies Dylan, Katie, and Issy: I so value the friendships we have forged, and look forward to sipping iced tea in rocking chairs some day.

There is one person above all others who I am greatly indebted to; I am profoundly grateful for and humbled by your insight, intellect, wisdom and humanity.

Last by but no means least, I thank my family — Hugh, Liz, Katie, Bart, Claire, Phoebe, Isabel, Jimmy, Emeline, and the Bump — for their boundless and unconditional love, trust, support, and laughter.

To those whom I haven't space to mention, you're not forgotten. Thankyou.