Star formation trends in high-redshift galaxy surveys: the elephant or the tail?

Martin Stringer,1* Shaun Cole,1 Carlos S. Frenk1 and Daniel P. Stark2

1Institute of Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE
2Kavli Institute of Cosmology & Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA

Accepted 2011 February 8. Received 2011 January 31; in original form 2010 November 12

ABSTRACT
Star formation rate and accumulated stellar mass are two fundamental physical quantities that describe the evolutionary state of a forming galaxy. Two recent attempts to determine the relationship between these quantities, by interpreting a sample of star-forming galaxies at redshift of \( z \sim 4 \), have led to opposite conclusions. Using a model galaxy population, we investigate possible causes for this discrepancy and conclude that minor errors in the conversion from observables to physical quantities can lead to a major misrepresentation when applied without awareness of sample selection. We also investigate, in a general way, the physical origin of the correlation between star formation rate and stellar mass within the hierarchical galaxy formation theory.

Key words: galaxies: evolution – galaxies: formation – galaxies: high redshift.

1 INTRODUCTION
As more distant galaxy populations become accessible to modern surveys, astronomers are striving to estimate their physical properties, despite the challenges inherent in such pioneering tasks. Even light that barely registers on our instruments is analysed to infer the stellar mass and star formation activity of its source, providing valuable stepping stones on which our physical picture of structure formation can progress.

For example, Stark et al. (2009) produced estimates of stellar mass for 1038 galaxies from the GOODS survey, grouped into three populations by redshift: \( z \approx 4 \), 5 and 6. These stellar masses were estimated using a population synthesis model (Bruzual & Charlot 2003; Bruzual 2007) which searches for the stellar population which best fits the observed spectral energy distribution of each galaxy (see Section 3.4).

Star formation rates were specifically not derived for this sample, because of uncertainties in the extinction correction. In lieu of this, the galaxies’ ‘emerging’ UV luminosities were computed (the luminosity at 1550 Å without any dust correction). However, fig. 9 of Stark et al. (2009) does include the star formation rates that would be inferred if a standard proportionality between UV luminosity and star formation rate is assumed (Madau, Pozzetti & Dickinson 1998):

\[
\log \left( \frac{M_*}{M_\odot \text{yr}^{-1}} \right) = -M_{50} + 18.45 \frac{2.5}{M_*^\odot].
\]

The resulting figure, for the galaxies in the nearest of those three samples, is reproduced for reference in the upper panel of our Fig. 1.

Despite only a fleeting appearance in the observational paper, these star formation rate estimates have since been the subject of a quite detailed theoretical analysis. Dutton et al. (2010) summarize the trend given by the sample in Fig. 1 as

\[
\frac{M_*}{M_\odot} \approx \frac{1}{0.62 \text{Gyr} \left( \frac{M_*}{10^{10} M_\odot} \right)^{-0.2}},
\]

which implies that the specific star formation rate \( \dot{M}_*/M_* \) is only weakly dependent on the stellar mass.

Meanwhile, the same sample of observational estimates has been subject to analysis by Khochfar & Silk (2011). Having chosen to plot the information on different axes, with a derived quantity, the specific star formation rate (SSFR), on the y-axis (as in the lower panel of Fig. 1) these authors perceive there to be a ‘strong observed mass-dependence’ with stellar mass.

So the same sample has been interpreted, on the one hand, as having a strong correlation with stellar mass and, on the other hand, a weak correlation.\(^1\) What is the reader to conclude from this literature?

The confusion can be appreciated by comparing the two panels in Fig. 1. The trend (2) does not seem unreasonable when looking at the top panel, but the problem is that the observational limit in

\(^1\) Both groups of authors agree on the relative evolution in specific star formation rate implied by the data when compared with equivalent relationships at low redshifts, and that this evolution seems to cease (appear constant) for \( z \gtrsim 4 \). Dutton et al. (2010) explain this in terms of high gas densities, and thus higher star formation rates, for a galaxy of a given mass at higher redshift. Khochfar & Silk (2011) look for modulated models of accretion-driven star formation. In this paper, we focus on the extent to which the data may or may not reveal the true underlying evolution (Section 4).
In summary, many independent studies find similar trends, with SSFR increasing with redshift and decreasing with stellar mass. However, the strength of these relationships varies considerably and is likely to be due to a complicated combination of sample-selection criteria, particularly wavelength and depth. Many authors make strong statements about down-sizing, based on the evolution of SSFR with redshift in different stellar mass bins, and on the oft-strong correlation of SSFR with stellar mass. We would urge caution before over-interpreting this type of plot as it is influenced strongly by selection biases.

On this note, it is clear that further theoretical interpretation of patterns in the $M_\star-M_{\star}$ plane requires an understanding of how the points arrive on these axes.

In Section 3, we confront the pitfalls in the process of translation between observable and physical properties by looking at a plausible model galaxy sample, both in its entirety and through the restrictions of an observational survey. The model we use to generate this mock population is a version of GALFORM (see Appendix A) which is a development of the model applied in Bower et al. (2006). No claim is made, in this context, that this is the correct physical picture. All that matters for this exercise is that we are using a realistic model, based on the current understanding of the physical processes involved, and that this particular observational sample could have been drawn from the model population. This caveat allows us to focus on the investigation in hand: the difference between what would be inferred from a sample and the true characteristics of the population from which the sample was drawn. In Section 4, this investigation is extended to look at the appearance of populations at a range of redshifts.

Before carrying out these simple but instructive exercises, we momentarily set the issue aside to consider what general hierarchical galaxy formation theory brings to bear on the $M_\star-M_{\star}$ relationship, in the absence of any particular model or agenda (Section 2). Clearly, this is required for a proper interpretation of the observational data:

Constraining the nature of the physical processes by which specific star formation rates are kept approximately constant in star-forming galaxies of wildly different mass [presents] substantial challenges for theoretical models to reproduce. (Pannella et al. 2009)

### 2 Specific Star Formation Rates

The two quantities that are presented in Fig. 1 are stellar mass, $M_\star$, and star formation rate, $\dot{M}_\star$. These are related by the definition

$$M_\star(z) = \int_0^z f(t_z - t) \dot{M}_\star(t) \, dt,$$

where $f(\Delta t)$ is the fraction of the initial stellar mass that has been retained by a stellar population after time, $\Delta t$, since formation. This is related to the recycled fraction, $R$, that usually appears in galaxy formation models. In the instantaneous recycling approximation, $R$ is a constant: $R \equiv 1 - \int_0^\infty f(t) \, dt$.

Of course, when referring to (3), one must not forget the hierarchical assembly of the final system. Some stars that are present in a galaxy at time $t_z$ would have formed in a separate, smaller system at earlier times. So a more explicit version of (3) would be

$$M_\star(z) = \int_0^z \sum_i f(t_z - t) \dot{M}_\star_i(t) \, dt,$$

where the index, $i$, runs over all the progenitor galaxies which merge into the final host. From here on, this bulky notation is assumed rather than repeatedly stated.
To provide some preliminary insight into the likely final relationship that emerges between $M_\star$ and $\dot{M}_\star$, equation (3) can be rewritten:

$$M_\star(z) = M_\star(t_0) \int_0^{t_0} f(t_0 - t) \frac{M_\star(t)}{M_\star(t_0)} \, dt.$$  \hfill (4)

The integral which remains in (4) is just the integral of the star formation history in units of its current value. This is a quantity which could potentially be largely system independent, if the hierarchical assembly of a halo can be approximated as self-similar. This would be expected over a certain mass range in a cosmology with conditional mass function index close to unity (Parkinson, Cole & Helly 2008).

For the galaxies within the haloes, such underlying similarity can be broken by various factors. At low mass, it is broken when a significant fraction of the final baryonic mass has never been in a halo massive enough to be heated above the temperature threshold for atomic cooling (White & Rees 1978; Stringer, Cole & Frenk 2010). At high mass, it is broken by long cooling times and processes such as active galactic nuclei. However, at intermediate masses, the underlying self-similarity in the halo assembly should carry through quite well to the galaxy population. In other words, hierarchical formation theory leads us to suspect that

$$\int_0^{t_0} f(t_0 - t) \frac{M_\star(t)}{M_\star(t_0)} \, dt \approx (1 - R) \tau_c,$$  \hfill (5)

where $\tau_c$ is some time-scale that is common to all galaxies at that redshift. This would lead to a direct proportionality between stellar mass and star formation rate at a given redshift:

$$\dot{M}_\star = \frac{M_\star}{(1 - R) \tau_c}.$$  \hfill (6)

To explore this supposition, we look at a model galaxy population generated using the GALFORM model. The underlying population of haloes (with total host masses above $10^6 h^{-1} M_\odot$ at $z = 0$) is generated in a comoving volume of $10^9 h^{-1}$ Mpc, assuming standard cosmological parameters (Komatsu et al. 2010). These haloes are then populated with galaxies using parameters and assumptions detailed in Appendix A. The evolution of subhaloes is followed by the model, right down to a subhalo mass of $10^5 h^{-1} M_\odot$, but we do not include satellite galaxies in this part of the discussion as they represent a distinct population governed by their own particular evolutionary characteristics. The main panel of Fig. 2 shows all the central galaxies in this population in the $\dot{M}_\star - M_\star / M_\star$ plane, as they are at redshift, $z = 4$.

Now, equation (6) says that the ratio of the current star formation rate to the mean star formation rate in a galaxy rarely differs significantly from one system to another. It does not say that the systems have a constant star formation rate. Far from it. This important point is illustrated in the peripheral panels of Fig. 2 which show the star formation rate history of five galaxies in the model population, plotted as a fraction of their final value. What brings three of them on to the main trend is not that they have constant or similar star formation histories, but that their varied and sporadic histories are subject to hierarchical assembly in the same cosmology. Put another way, if the star formation rate had persistently differed from its current value, the system would have just ended up with a different stellar mass (and be in a different position but on the same trend).

The lines of relative star formation rate must all converge on unity at $t = t_0$. Though all deviate significantly from this over the course of their history, as mergers and instabilities create bursts and lulls of star formation, the integral under this line invariably ends up being about the same value (unless you happen to catch a system at the height or tail of one of these episodes, as can be seen in the left corner panels of Fig. 2). This integral is just our time-scale from (5), $\tau_c$.

2 For example, if all galaxies formed at $t = 0$ and each had (its own particular) constant star formation rate, this common time-scale would just be the age of the universe: $\tau_c = t$.

© 2011 The Authors, MNRAS 414, 1927–1936

Monthly Notices of the Royal Astronomical Society. © 2011 RAS

2 Star formation trends in Galaxy Surveys

The lines of relative star formation rate must all converge on unity at $t = t_0$. Though all deviate significantly from this over the course of their history, as mergers and instabilities create bursts and lulls of star formation, the integral under this line invariably ends up being about the same value (unless you happen to catch a system at the height or tail of one of these episodes, as can be seen in the left corner panels of Fig. 2). This integral is just our time-scale from (5), $\tau_c$.

2.1 Insensitivity to star formation processes

Equation (6) describes the trend which connects the star formation rate of a galaxy at redshift, $z$, to the host stellar mass. It is motivated solely by the argument that the star formation histories of all central galaxies are subject to the same principal constraints: the age of the universe and natural statistical fluctuations due to mergers and instabilities.

This argument is not an attempt to evade the important complexities of star formation. Rather, we wish to be realistic about what the observed $M_\star - \dot{M}_\star$ trend can teach us. The fact that details were not needed to support the argument suggests that documenting the main trend at any particular redshift, may not help us distinguish between different proposed theories of star formation and feedback.

Over a range of redshifts, the evolving position of the trend (i.e. the value of $\tau_c$) may provide more clues (see Section 4), but only given an assumed halo merger history. The information that is really being provided by such surveys concerns the typical mass assembly history. Part of this story is indeed the star formation process, but it is very difficult to separate this from the dominant influence of the structure formation process which is, as we have argued, ultimately responsible for (6).

The characteristic which is sensitive to star formation and feedback physics, even over a small redshift range, is the position of the population along the trend. If star formation and/or feedback had been different, each halo would have ended up hosting a different stellar mass. The galaxies would still have every reason to appear on our trend, but to the left or right of their original position, not perpendicular to it. This has been previously noted (Stringer 2008; Dutton et al. 2010). So, given a known or assumed halo population, it is the luminosity function that can be used to constrain theories of star formation and feedback, whilst direct measurements of star formation rates, oddly, may not.

Having made these few simple points about theoretical expectations, we now move to the more pertinent matter of how the reality of these physical properties might be revealed. The model from Fig. 2 can be approached as if it were real data and the process of observation followed to find out if and where pitfalls in the interpretation of the data may appear.

3 True and inferred populations

To illustrate the importance of sampling effects in high-redshift surveys, the generated population of galaxies from Section 2 can be analysed under the same observational constraints, and using the same techniques as were applied by Stark et al. (2009) to the real data.

3 The positions of satellites in the $\dot{M}_\star - M_\star$ plane are more sensitive to the details of star formation physics, and are mostly found off the main trend (Lagos et al., in preparation), hence our restriction of the above argument to central galaxies.
The model follows the formation of each galaxy through hierarchical assembly, computing the build-up of the stellar population, interstellar medium, and their metallicities. The intrinsic luminosities of the galaxies are then obtained using the stellar population synthesis method of Bruzual & Charlot (2003).

Fig. 3 shows the real and model samples on a plot of observed quantities: rest-frame visible magnitude versus rest-frame UV magnitude. The band filters used to calculate the magnitudes in the figures were matched precisely to those in the observational survey.

The two sets of points in Fig. 3 are somewhat offset from each other, but the statistical significance of this difference is low: the majority of the population (i.e. the fainter galaxies) are overlapping. So, for the purposes of this purely illustrative exercise, we consider this model to be an acceptable match to the data. (For a discussion of discrepancies that exist between current semi-analytic models and recent observations, the reader is referred to Lacey et al. 2010.)

With this caveat, we proceed to follow our model sample all the way through from the ‘real’ physical parameters to the magnitudes that would be observed, and then back again to the inferred physical parameters. This process from physical quantities to observables, and back, is shown as a sequence of panels in Fig. 4. Each transition (clockwise) from one panel to the next introduces one part of this chain, as follows.

---

4 Throughout this paper, we use the simple magnitude notation $M_\lambda$ (Absolute Magnitude, rest-frame wavelength) and $m_\lambda$ (apparent magnitude, observed wavelength).

---

© 2011 The Authors, MNRAS 414, 1927–1936
Monthly Notices of the Royal Astronomical Society © 2011 RAS

Downloaded from https://academic.oup.com/mnras/article-abstract/414/3/1927/1035015 by guest on 30 July 2018
3. Dust

The top-left panel of Fig. 4 shows the star formation rates and stellar masses of the model galaxy sample. Immediately to the right of this is shown the mapping to absolute rest-frame UV magnitude. Whilst the scatter in the relation from star formation rate (SFR) to initial UV emission is worth understanding (panel B), it is very minimal. The real problem in any efforts to derive the SFR is the effect of intervening dust on the UV emission.

In the model, the mass of dust is assumed to be proportional to the mass of metals in the interstellar medium. The obscuration is computed adopting the scheme of Ferrara et al. (1999), by following the radiative transfer of light (at all wavelengths) through dust assumed to be distributed smoothly in the galactic disc. Inclination angles are assigned to each galaxy at random. For full details, the reader is directed to Cole et al. (2000) and Lacey et al. (2010). This model predicts changes in UV magnitude between 0 and 2, with an average of about 0.5 (depending on intrinsic luminosity).

The top-right panel shows the correlation between the absolute UV magnitude and star formation rate. For comparison, a dashed line (Section 3.3) shows the relationship (1) that will be assumed when mapping back from the UV to the SFR. As this relationship is based on the same stellar population synthesis model assumed in GALFORM, the simple estimate ‘backwards’ from the UV to the star formation rate (dashed line) compares well with the ‘forward’ calculation from the full hierarchical model (points). However, as would be expected, the effect of dust has been both to introduce scatter and to reduce the UV luminosities with respect to this estimate (see Appendix B). The systematic effect of applying the estimate, regardless of this, is described in Section 3.3.

3.2 Observational limits

The middle-right panel of Fig. 4 shows our model galaxy population in terms of two estimated observables, the UV and visible absolute magnitudes. This is the point at which we can turn the process around and analyse the sample to see how well we can recover the physical properties of the population.

The observational limits used for this illustrative exercise are chosen from Stark et al. (2009), namely that galaxies are included in the survey if their apparent magnitudes in the rest-frame UV satisfies $m_{UV} < 27$. Those sources that are fainter than $m_{3600} \approx 27$ are not detected in that filter, but will still be included in the sample and their stellar masses computed using the measured $2\sigma$ upper limits.

3.3 Inferred star formation rate

The bottom-right panel of Fig. 4 begins the mapping back to physical parameters. We apply the direct proportionality (1) used to produce fig. 9 in Stark et al. (2009), between absolute uncorrected UV magnitude, $M_\ast$, and inferred star formation rate, $\dot{M}_\ast$. This simple conversion is well founded by our knowledge of stellar evolution, as explained by its proponents:

The UV continuum emission from a galaxy with significant ongoing star formation is entirely dominated by late-O/early-B stars on the main sequence (Madau et al. 1998).

The need to critically examine the criterion of ‘significant ongoing star formation’ is addressed in Appendix B. The principal conclusion of this section is that this concern is very minor, particularly at the visible end of the luminosity range.

Much more of a problem is the scatter due to dust, which was discussed in Section 3.1, particularly when (1) is used to recover the star formation rates in spite of this scatter. Because of the increased number of galaxies towards the limits (fainter galaxies are more abundant), this working assumption has quite drastic effects on the appearance of the population when the final step is made back to inferred physical properties.

Because of the relative abundance of fainter systems, there is an accumulation of points at the intersection between the UV limit and the assumed SFR–UV relation (the two lines in the bottom-right panel of Fig. 4). When the final step is taken to map back to the original axes (the lower-left panel), this concentration of data points spreads out across a range of values for stellar mass and creates the impression of constant star formation rate that we saw in Fig. 1. Crucially, the proportionality shown in the top-left panel (reproduced again as a line in the bottom-left corner) has been completely lost.

Additionally, because the inferred SFRs are lower than the original values, not only has the $M_\ast-\dot{M}_\ast$ correlation been lost, but the overall position of points in the plane has shifted. This systematic underestimation of star formation rates will be particularly important when mapping the evolution of the relationship with redshift, as we show in Section 4.

3.4 Inferred stellar mass

To find the stellar mass that would be inferred for the model galaxies, the estimated apparent magnitudes from the mock sample were processed using the same model (Bruzual & Charlot 2003) that was applied by Stark et al. (2009). The ‘observed’ magnitudes in

\[ \text{Since both the galaxy formation model and the post-observational analysis assume the same cosmology and model for IGM absorption, we ignore these parts of the process here.} \]
Figure 4. An illustration of how a high-redshift galaxy population could be misrepresented by a magnitude-limited sample. The top-left panel shows the physical properties (total stellar mass, $M_\star$, and current star formation rate, $\dot{M}_\star$) of a model galaxy population within a volume of $10^5 \text{ Mpc}^3 h^{-3}$ at redshift $z = 4$. The panel immediately to the right shows the translation to absolute magnitude, $M_\lambda$. The central panel shows the same rest-frame UV magnitudes together with their optical counterparts. Highlighted in red are the systems that would be seen in a survey such as that by Stark et al. (2009), with approximate magnitude limits as indicated by the dotted lines. The bottom panels show the assumed mapping from rest-frame UV magnitude back to star formation rate and then to stellar mass. Diagonal dashed lines indicate the position of equation (1), the $M_{\text{UV}}-\dot{M}_\star$ relationship of Madau et al. (1998). Solid diagonal lines highlight the approximate trend in the top-left panel (6).

With a complete loop from physical properties to observables, and back, we can estimate the correlation of the ‘observational’ estimates (bottom-left panel of Fig. 4) with the original population (top-left panel). The trend in the population (solid line) has been lost in the process, and overwhelmed by patterns due to the limits on the sample (both observational and natural).

In particular, the loss of correlation in the $M_\star-\dot{M}_\star$ plane will appear as a ‘strong mass dependence’ in $\dot{M}_\star/M_\star$, as was apparent to Khochfar & Silk (2011), and can also be seen in the lower panel.

four optical HST/ACS filters (F435, F606, F775, F850), three near-infrared filters ($J, H, K_s$) and the four Spitzer/IRAC filters (3.6, 4.5, 5.8 and 8.0 $\mu$m) are used to estimate the stellar masses of the real galaxies.

This valuable exercise produced stellar mass estimates which were in excellent agreement (The fractional difference, $\epsilon \equiv$ Inferred mass / ‘True’ mass, was found to have a standard deviation of just 0.3). These inferred masses are plotted in the final, bottom-right panel of Fig. 4.
Star formation trends in Galaxy Surveys

4 EVOLUTION OF TRENDS WITH REDSHIFT

Section 2 established the argument in support of a common specific star formation rate for all central galaxies. Of course, this only applies at any one particular time; the constant of proportionality in equation (6) will be different for samples at different redshifts, hence the notation $\tau_z$.

To understand this evolution, we return to the picture used to support (6), namely that $\tau_z$ is just the integral under the star formation rate.

of Fig. 1. Existing concerns about this effect can be found in Dunne et al. (2009).

Of course, we do not know how closely our model galaxy population matches reality. However, with rigorous application of what we do know, particularly about the process of observational analysis, this exercise still provides a clear demonstration of the potential pitfalls. It is clearly incorrect to presume that the correlation apparent in the sample will be that with the underlying population.

Figure 5. An illustration of how specific star formation rate could be independent of stellar mass but dependent on redshift, and how such evolution can be misinterpreted without a proper understanding of selection effects. Both rows of panels shows contour lines which represent densities in this plane of 2, 10 and 50 galaxies/dex$^2$(10 Mpc)$^3$. The dots in the upper row pick out those galaxies that would be visible in a current survey (see Section 3.2), plotting their ‘true’ star formation rates. The lower row highlights these same galaxies, but now shows the inferred specific star formation rate that would have been derived from their uncorrected UV luminosity (see Section 3.3). The characteristic rates for both the inferred sample and the true population are shown as points in Fig. 6, along with equivalent points for intermediate redshifts.

Figure 6. Characteristic growth rates, $1/(\tau_z)$, for the true galaxy population and ‘observational’ samples shown in Fig. 5. Open squares show the mean value for all the galaxies in the population. Filled squares show the value for just the galaxies that would be visible to observers under an example UV limit of $m_{775} > 27$ (derived from the dots in the upper row of Fig. 5). Dots show the mean that would be inferred using the UV magnitude, without correction, to derive the star formation rate (derived from the dots in the lower row of Fig. 5).
rate history, in units of its current value. Two limiting cases are easy to identify immediately: at the highest redshifts, galaxies will be seen near birth and $\tau_\star \rightarrow \tau_\star$, the characteristic time-scale for star formation itself. At the other extreme, as gas is exhausted and the growth of $\Omega_\star$ suppresses further accretion, $\tau_\star \rightarrow \infty$.

In between times, while gas is plentiful but the age of the universe, $t(z) \gg \tau_\star$, the fluctuations in star formation are ironed out, leading to much less scatter about the trend, which becomes comparable to $\tau_\star \sim t(z)$. This is the era we are in at present.

Fig. 5 illustrates this argument using the same model population considered in Figs 1 and 4, but now seen at a range of redshifts. As time progresses (from right to left), we see the trend narrow and drop in accordance with the argument above. This evolution is also seen directly in Fig. 6, which shows $\tau_\star$ against redshift.

Highlighted in the upper row of Fig. 5 are the galaxies in the model population which might be visible in an observational survey. At higher redshifts, this sample consists of brighter and brighter subsets of the population, which leads to a very strong bias towards selecting galaxies with high specific star formation rates. The mean rate, $1/(\tau_\star)$, is plotted in Fig. 6 as a function of redshift, and it can be seen that the value based on this visible sample diverges wildly from the true mean after about $z \sim 2$.

A very separate issue, which is apparent in Fig. 6, is the evolution of the characteristic rates that would be inferred from this reduced sample. Using the process outlined in Section 3, the emission that is predicted by the model for each magnitude-selected sample is converted into stellar masses and star formation rates that would be inferred using, for example, the correlation (1).

Also worth noting from this figure is the fact that none of the three lines follows the pattern that actually has been put together from collected observations (Dutton et al. 2010, and references therein), where the inferred characteristic rate is seen to flatten out at high redshifts. For the model to be consistent with these observations, it is of course the inferred line in Fig. 6 that should broadly agree with the published results. Clearly this is not the case, and this could be due to shortcomings of the model, or of the analysis applied to the observations.

As far as the exercise in this paper is concerned, it is more important to emphasize, again, the discrepancy between all three lines in Fig. 6, which calls into question how well the evolution in specific star formation rates has really been captured by surveys thus far.

5 SUMMARY

Very different trends of stellar mass and star formation rate have been attributed to the same observational sample (Dutton et al. 2010; Khochfar & Silk 2011). In this paper, we have carried out a more rigorous investigation into the origin of these two key estimated physical quantities.

A simple argument, based on hierarchical galaxy formation theory, was presented to understand why a strong trend might exist between these two quantities, and that such a relationship can result from the self-similar nature of galaxy assembly, independently of star formation or feedback processes.

Using a model galaxy population as a guide, we have shown that, due to the combined effects of selection bias and physical scatter in the relations between observable and physical properties, this underlying trend can be easily misrepresented. There can also be a large discrepancy between inferred and true parameter values.

These results highlight the importance of using realistic, physical galaxy formation models to guide the interpretation of high-redshift surveys. By subjecting model-generated galaxy populations to the same analysis as the real data, observation can be compared with competing theories on an even footing. In this way, new surveys can lead to more incisive quantitative conclusions about the true underlying galaxy population.

ACKNOWLEDGMENTS

The authors would like to thank Andrew Benson, Cedric Lacey, Claudia Lagos and the referee for all their helpful comments. CSF acknowledges a Royal Society Wolfson Research Merit Award and SC acknowledges the support of the Leverhulme Trust Research Fellowship. DPS acknowledges support from an STFC postdoctoral research fellowship. This work was supported by an STFC rolling grant to the Institute for Computational Cosmology.

REFERENCES

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Benson A. J., 2005, MNRAS, 358, 551
Benson A. J., Bower R., 2010, MNRAS, 405, 1573
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Bruzual G., 2007, in Vazdekis A., Peletier R. F., eds, Proc. IAU Symp. 241, Stellar Populations as Building Blocks of Galaxies. Cambridge, Cambridge Univ. Press, p. 125
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Cole S., Lacey C., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Dunne L. et al., 2009, MNRAS, 394, 3
Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690
Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 437
Khochfar S., Silk J., 2011, MNRAS, 410, L42
Komatsu E. et al., 2011, ApJS, 192, 18
Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2011, MNRAS, doi:10.1111/j.1365-2966.2010.18021.x
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
Pannella M. et al., 2009, ApJ, 698, L116
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Rodighiero G. et al., 2010, A&A, 518, L25
Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, ApJ, 697, 1493
Stringer M. J., 2008, DPhil thesis, Univ. Oxford
Stringer M. J., Cole S., Frenk C. S., 2010, MNRAS, 404, 1129
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

APPENDIX A: THE GALAXY FORMATION MODEL

The mock galaxy population in the figures in this article was generated using a version of the GALFORM semi-analytic model which is currently under development. The goal of this new version is to combine the most realistic aspects of the two previously published versions (Baugh et al. 2005; Bower et al. 2006), whilst achieving a better match, than either of these, to current observational constraints.

© 2011 The Authors, MNRAS 414, 1927–1936

Monthly Notices of the Royal Astronomical Society © 2011 RAS
Table A1. Changes to the GALFORM model.

| Parameter | Bower et al. (2006) | This version |
|-----------|---------------------|--------------|
| $\alpha_{\text{hot}}$ | 3.2 | 2.5 |
| $v_{\text{hot}}$ | 485 km s$^{-1}$ | 300 km s$^{-1}$ |
| $\alpha_{\text{reheat}}$ | 1.26 | 0.3 |
| $\alpha_{\text{reheat}}$ | 1.26 | 0.3 |
| $\tau_{\star}$ | 350 Gyr | 4 Gyr |
| $\alpha_{\text{cool}}$ | 0.58 | 0.78 |
| $t_{\text{mrg}}$ | 1.5 | 1.0 |
| $v_{\text{cut}}$ | 50 km s$^{-1}$ | 30 km s$^{-1}$ |

The development model is most closely related to the model published by Bower et al. (2006), but uses parameters for star formation and feedback that are more realistic; closer to those favoured by Baugh et al. (2005). A list of parameter changes appears in Table A1. The parameters $\tau_{\star}$ and $\alpha_{\star}$ apply to star formation rate $\psi$, as follows:

$$\psi = \frac{M_{\text{gas}}}{\tau_{\star}} \left(\frac{v_c}{200 \text{ km s}^{-1}}\right)^{\alpha_{\star}}.$$  \hspace{1cm} (A1)

These, and all other parameters, are as defined in Bower et al. (2006) and references therein.

Other changes are the distribution of orbital parameters, which has been updated to follow Benson (2005), and the treatment of the cooling of hot halo gas, which now follows Benson & Bower (2010). The full details of this version will appear in Lacey et al. (in preparation).

**APPENDIX B: UV LUMINOSITY AS A STAR FORMATION TRACER**

To explore this relationship, we return to the model galaxy population that was shown in Fig. 2. Consider, first, their star formation rate versus rest-frame UV magnitude, which is shown in the left main panel.

![Figure B1](https://example.com/fig_b1.png)

**Figure B1.** The scatter in the relationship between star formation rate and UV magnitude by incongruous star formation histories. The dots in the left-central panel show the galaxies from the same model population as in Figs 3 and 4, and the dashed line is the correlation of Madau et al. (1998) (equation 1) which can be used to derive star formation rates from observed magnitudes. Four particular galaxies are circled and their star formation histories shown in smaller panels. For reference to physical properties, these same galaxies are also highlighted on a plot of star formation rate versus stellar mass (right-central panel).
To clarify our understanding of this, the same star formation rates are shown in the right-hand panel as a function of stellar mass. This shows the very strong correlation between current star formation rate, and mean star formation rate, which was discussed in Section 2. Now, for galaxies that lie on or above this trend, the massive star population is large enough that they will indeed be the main contributors to the total UV luminosity, and the strong correlation (1) holds.

For galaxies below the main trend, this approximation breaks down; less massive stars are comparatively so abundant that they are responsible for most of the total UV output, despite their poor individual contribution to this part of the spectrum.

This is further illustrated by the small peripheral panels which show the star formation histories of four particular galaxies. In the lower two panels, past star formation episodes were so productive that the stars produced then are outshining the recently formed stars, even at this high-energy end of the spectrum.

The main conclusion of this exercise is positive; hierarchical formation theory predicts that only a small fraction of galaxies would differ from the assumed correlation. Furthermore, such scatter occurs mostly on the lower side; unusually high star formation rates may still be estimated correctly as it only serves to accentuate the underlying assumption (1).

This one-sided nature of this error means that the characteristic star formation rates would be systematically overestimated. However, the practical consequences of this are negligible, particularly in the context of the expected obscuration due to dust, discussed in Section 3.1 and shown in Fig. B2.