

Spatially unresolved SED fitting can underestimate galaxy masses: a solution to the missing mass problem

Robert Sorba^{1,2★} and Marcin Sawicki^{1★†}

¹*Department of Astronomy and Physics, Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia B3H 3C3, Canada*

²*Physics Department, Mount Allison University, 62 York Street, Sackville, New Brunswick E4L 1E2, Canada*

Accepted 2018 January 17. Received 2018 January 13; in original form 2017 May 16

ABSTRACT

We perform spatially resolved, pixel-by-pixel Spectral Energy Distribution (SED) fitting on galaxies up to $z \sim 2.5$ in the Hubble eXtreme Deep Field (XDF). Comparing stellar mass estimates from spatially resolved and spatially unresolved photometry we find that unresolved masses can be systematically underestimated by factors of up to 5. The ratio of the unresolved to resolved mass measurement depends on the galaxy's specific star formation rate (sSFR): at low sSFRs the bias is small, but above $\text{sSFR} \sim 10^{-9.5} \text{ yr}^{-1}$ the discrepancy increases rapidly such that galaxies with $\text{sSFRs} \sim 10^{-8} \text{ yr}^{-1}$ have unresolved mass estimates of only one-half to one-fifth of the resolved value. This result indicates that stellar masses estimated from spatially unresolved data sets need to be systematically corrected, in some cases by large amounts, and we provide an analytic prescription for applying this correction. We show that correcting stellar mass measurements for this bias changes the normalization and slope of the star-forming main sequence and reduces its intrinsic width; most dramatically, correcting for the mass bias increases the stellar mass density of the Universe at high redshift and can resolve the long-standing discrepancy between the directly measured cosmic SFR density at $z \gtrsim 1$ and that inferred from stellar mass densities ('the missing mass problem').

Key words: galaxies: high-redshift – galaxies: luminosity function, mass function – galaxies: statistics – galaxies: stellar content.

1 INTRODUCTION

Spatially resolved (i.e. pixel-by-pixel) broad-band Spectral Energy Distribution (SED) fitting provides an exciting tool to study the spatial distribution of stellar matter in a galaxy, particularly for high-redshift galaxies where obtaining 2D spectra is cost prohibitive. Resolving features in a galaxy's spatial distribution allows us to uncover details that are obscured and blotted out by integrated photometry. The technique of pixel-by-pixel SED fitting was first developed by Abraham et al. (1999), who matched spectral synthesis models to resolved multicolour data of $z \sim 1$ galaxies in the Hubble Deep Field in order to study ages and evolutionary histories of stellar populations in those galaxies. An analysis of how pixel-by-pixel fitting affected mass estimates was performed by Zibetti, Charlot & Rix (2009) for nine nearby ($D < 30$ Mpc) SINGS galaxies. They used median stellar mass-to-light (M_*/L) ratios (derived from a large SPS library) to create stellar mass maps of the galaxies and found that the unresolved mass estimate (i.e. from integrating all

the light of the galaxy) could be underestimated compared to the resolved mass by up to 40 per cent (0.22 dex). They postulated that the presence of large dust lanes contributed to the mass discrepancy. Sorba & Sawicki (2015) corroborated this by using pixel-by-pixel analysis on a larger set of 67 nearby galaxies in the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2011). They found that the presence of strong dust lanes could cause unresolved mass estimates to be underestimated by 45 per cent (0.25 dex) for an individual galaxy, but only in extreme cases, and that dust did not affect mass estimates on average. However, by being able to disentangle older stellar populations from younger ones, Sorba & Sawicki (2015) were able to find an increasing bias in a galaxy's mass estimation with specific star formation rate (sSFR). They found the effect was small (unresolved masses were underestimated by 13 per cent or 0.06 dex on average), but strongly correlated with sSFR leading to a maximum bias of 25 per cent (0.12 dex) at sSFRs of around 10^{-8} yr^{-1} , and they concluded this bias was due to outshining. Essentially, young massive stars have a much smaller M_*/L than older stars. They are orders of magnitude brighter than solar-type stars. But they are also short-lived, meaning they are only present in currently star-forming regions. When all the light from a galaxy is integrated together, however, the light from young stars dominates

* E-mail: rsorba@mta.ca (RS); marcin.sawicki@smu.ca (MS)

† Canada Research Chair.

a galaxy’s SED at optical wavelengths. Thus, model SED fits to broad-band photometry preferentially fit the M_*/L of the younger stellar population, often missing mass from older components (for further discussion on outshining and its effects, see Sawicki & Yee 1998; Papovich, Dickinson & Ferguson 2001; Maraston et al. 2010; Pforr, Maraston & Tonini 2012).

Spatially resolved stellar property maps have also been created for high-redshift ($0.5 < z < 2.5$) star-forming galaxies in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (Grogin et al. 2011; Koekemoer et al. 2011) by Wuyts et al. (2012). They used the maps to examine star-forming clumps and variations in rest-frame colour, stellar surface mass density, age, and extinction as a function of radius, finding results consistent with inside-out disc-growth. They also compared unresolved to pixel-by-pixel mass estimates, but found no discrepancy. However, their binning of low signal-to-noise (S/N) pixels may have impacted their effective spatial resolution, leaving an open question as to how outshining may change at higher redshifts. In this work we set out to examine the influence of outshining as a function of cosmic time. We do so following the pixel-by-pixel method laid out in Sorba & Sawicki (2015). It is important to note that this work is a comparative study between the results given by pixel-by-pixel SED fitting versus the standard integrated-light SED fitting approach. When we discuss any bias in mass estimation, this is only in relation to how the two methods compare. We do not presume to know, and for our purposes do not need to know, the true stellar mass of the galaxies studied.

Throughout this work we assume a *WMAP7* (Komatsu et al. 2011) flat Λ cold dark matter cosmology ($H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.272$) and use the AB magnitude system.

2 METHOD

2.1 Data

To examine pixel-by-pixel SED fitting at higher redshifts we used the Hubble eXtreme Deep Field (XDF; Illingworth et al. 2013) publicly available data set, which combines data from a decade of observations made with the *Hubble Space Telescope* (*HST*) creating optical/NIR images unsurpassed in depth. The extreme depth of this data set is particularly important for a pixel-by-pixel analysis, increasing the number of pixels with acceptable S/N and enabling us to probe fainter regions of galaxies. The XDF data set contains mosaic images covering approximately 4 arcmin^2 in nine passbands (ACS/WFC *F435W*, *F606W*, *F775W*, *F814W*, and *F850LP*; WFC3/IR *F105W*, *F125W*, *F140W*, and *F160W*) that have all had their background flux removed, been astrometrically aligned, and drizzled on to the same 60 mas pixel scale.

It is important when doing pixel-by-pixel photometry that all images have the same point spread function (PSF). To measure the PSF of each bandpass, we started by visually inspecting stars from the catalogue of stars in the *Hubble Ultra Deep Field* (Pirzkal et al. 2005) and selecting those that were not contaminated by noise or nearby galaxies. These stars were re-centred to their intensity-weighted sub-pixel centres, normalized, and median-stacked to form the PSF. The WFC3 PSFs were augmented at larger radii using model PSFs created using the *TINYTIM* software package (Krist 1995; Krist, Hook & Stoehr 2011) to account for the relatively large amount of flux contained in the broad wings of the NIR PSFs (van der Wel et al. 2012; Rafelski et al. 2015). The broadest PSF belonged to the *F160W* image, which had a PSF full width at half-maximum of approximately 0.20 arcsec . For each bandpass other than *F160W*, we convolved the images to match the broadest PSF by creating an

11×11 convolution kernel using the method described in Alard & Lupton (1998), employing a delta function at each pixel in the kernel as the set of basis functions (Becker et al. 2012). The size of the kernel provided a good compromise, limiting the tendency of our choice of basis function to overfit the noise in the wings of the PSF, but was still large enough to capture all the features of the transformation kernel.

After convolving the images to the same PSF, we created a multi-band detection image by summing each image weighted by the inverse variance on a pixel-by-pixel basis. This process created an image where galaxies that are bright at the bluest wavelengths but faint in the NIR (such as low-redshift starbursts) are present, but so are galaxies with very faint observer-frame optical fluxes but stronger observer-frame NIR flux (such as faint quiescent galaxies or galaxies at higher redshifts). The multiwavelength detection image maximized the number of galaxies present in the image, and limited any bias in choosing which wavelength to select pixels from. We included in our pixel-by-pixel catalogue any pixels that had a S/N ratio greater than five in the detection image.

We labelled galaxies and assigned redshifts using the spectroscopic and photometric redshift catalogues provided by Rafelski et al. (2015). The photometric redshifts were derived using 11 bandpasses covering the near-ultraviolet to NIR from WFC3 observations in the Ultra-Deep Field. Rafelski et al. (2015) provide photometric redshifts found using two different codes – *BPZ* (Benitez 2000; Benitez et al. 2004; Coe et al. 2006) and *EAZY* (Brammer, van Dokkum & Coppi 2008). We chose to use the *BPZ* redshifts, which had a slightly smaller scatter when compared to spectroscopic redshifts as measured by the normalized median absolute deviation ($\sigma_{\text{NMAD}} = 0.028$ for *BPZ* versus 0.030 for *EAZY*), and a slightly lower outlier fraction (2.4 per cent versus 5.9 per cent for *EAZY*). Although averaging the two redshift measurements is possible, it runs the risk of increasing the catastrophic outlier fraction, and thus we chose to only use a single set of measurements for the photometric redshifts. The spectroscopic redshifts were compiled from nine different sources and contain robust measurements for 169 galaxies.

We assigned pixels to separate galaxies by converting the RA and Dec. given in the Rafelski et al. (2015) catalogue to pixel coordinates, and used these as initial seeds for basin markers for a watershed segmentation (Soille & Ansoult 1990). Whereas traditional segmentation through deblending with, e.g. *SExtractor* (Bertin & Arnouts 1996), requires multiple run-throughs with different parameters in order to accurately segment both galaxies that are blended together and large galaxies that should not be split into separate objects (see Barden et al. 2012; Rafelski et al. 2015), the watershed algorithm for segmentation can be done straight-forwardly when one already knows roughly where objects are located. Briefly, one can imagine the light profile from our detection image flipped upside down to create ‘basins’ which are labelled by the initial markers and bounded by our S/N acceptance criteria. The basins are flooded with water starting from the initial seeds, and where one basin overflows into another marks the delineation between the two objects. The use of initial markers is important for the success of this algorithm; simply using local minima leads to oversegmentation due to noise.

From this segmentation map we created a pixel-by-pixel flux catalogue and an unresolved total flux catalogue. The fluxes were corrected for extinction due to foreground dust according to Schlafly & Finkbeiner (2011). Similarly to Sorba & Sawicki (2015), in order to compare the same light with the same light, the unresolved flux catalogue was made by summing all of the flux from each pixel

for each galaxy. We removed any objects flagged as stars in Pirzkal et al. (2005) and removed by eye any objects too close to the borders of the image, or with obvious artefacts. In total, the analysis in the following sections included 162 galaxies with spectroscopic redshifts and 1060 with photometric redshifts. The minimum, median, and maximum number of pixels in the spectroscopic sample were 72, 921, and 17 499 pixels, respectively. For the photometric sample, we restricted the minimum number of pixels required in order to be included in the analysis to 64. Below this threshold, we found that uncertainties in mass estimates (and specifically the ratio of the unresolved to resolved masses) became inordinately large. The median and maximum number of pixels for the photometric redshift galaxies were 182 and 12 228 pixels, respectively.

2.2 Models

We created model SED templates using the Flexible Stellar Population Synthesis (FSPS; Conroy, Gunn & White 2009; Conroy & Gunn 2010) code. The model set assumed solar metallicity, a Chabrier (2003) IMF, and a Calzetti et al. (2000) dust law, and were corrected for extinction due to intergalactic hydrogen following the Madau (1995) prescription. Nebular line emission was also included in these model templates (Byler et al. 2017), which is a relatively new feature available to FSPS.

While the assumption of solar metallicity becomes less accurate at higher redshifts, both Bolzonella et al. (2010) and Pforr et al. (2012) found that the differences in mass estimation between different metallicity setups is small. Since in this work we are most interested in recovering stellar masses, fitting with solar metallicity should be sufficient, and helps to reduce the parameter space.

Departing from Sorba & Sawicki (2015), we chose to parametrize the SFH of our model SEDs as an exponential decay (i.e. τ -models), in part for easy comparison with other pixel-by-pixel works that used an exponentially declining SFH (Wuyts et al. 2012; Welikala et al. 2008). Whereas our previous work found that photometry of nearby galaxies in the Sloan Digital Sky Survey was well matched by two-component burst SFH models, we found the burst SFH models were poor representations for higher redshift galaxies (further details of how pixel-by-pixel studies are affected by assumed SFH and model grid spacing are located in Appendix A).

Our model grid consisted of $\log(\text{ages/yr})$ ranging from 7.7 to 10.1 in steps of 0.05, a dust parameter ranging from 0 to 0.7 in steps of 0.1, and 20 different τ values spaced roughly evenly logarithmically between 0.3 and 10. Our minimum τ was chosen based on the recommendation from Wuyts et al. (2011), who found that forcing the e -folding time-scale to be greater than several 100 Myr was most effective at reproducing UV+IR measured SFRs.

For each galaxy's pixel, we shifted the model spectrum to the redshift given by Rafelski et al. (2015), defaulting to the spectroscopic redshift when available. Model broad-band fluxes were created from each spectrum by convolving with broad-band filter transmission curves, and the best-fitting model was found using the SEDFIT software package (Sawicki 2012a), which finds the minimum χ^2 in our model parameter space. Fits were restricted to agree with the age of the Universe at the given redshift. Example mass and SFR maps are shown in Fig. 1 for three different galaxies. The 2D maps are smooth and lack artificial discontinuities, providing confidence that our model grid is finely sampled enough to account for the range of colours observed. We reiterate that whether or not the models necessarily provide the most accurate mass or SFR measurements is irrelevant for this work due to its comparative nature. However, we have made every effort to follow current best practices, and our

stellar property estimates should be within systematic uncertainty of other works (see Conroy 2013 for a review of parameter estimations from SEDs).

We use SEDFIT to determine uncertainties by constructing probability density functions (PDFs) via several Monte Carlo (MC) realizations. For each MC realization, the observed flux in each bandpass is changed by adding a normally distributed perturbation (with standard deviation equal to the uncertainty in the observed bandpass and centred on the original measurement). We performed 300 of these realizations for each pixel and used the median mass estimate to avoid quirks arising from minimal χ^2 best fits (Kauffmann et al. 2003; Taylor et al. 2011). Using the PDF to study the mass (or other fitted parameters) allows for a fuller understanding of the interplay between observations and model space. While the mass of the minimum χ^2 model may lie anywhere within the PDF, the median value is typically very close to the most-probable-value, defined as the peak of the PDF (Davidzon et al. 2017).

3 RESULTS

3.1 Outshining at high redshift

To see if the results of Sorba & Sawicki (2015) extend to galaxies at higher redshift, we plot in the left-hand panel of Fig. 2 the ratio of stellar mass estimates derived from unresolved photometry to those determined from pixel-by-pixel SED fitting as a function of sSFR for galaxies with spectroscopic redshifts. The colour of each data point is related to the galaxy's redshift (bluer colours for lower redshifts, redder for higher redshifts). If the pixel-by-pixel and unresolved stellar mass estimates were the same, all points would lie along the one-to-one correspondence horizontal dashed black line. For the large majority of galaxies, the mass ratio is below one, meaning that unresolved photometry SED fitting tends to underestimate the true mass, if we assume the pixel-by-pixel mass is closer to the true mass. Looking at the colours of the points, there is no discernible difference in where a galaxy's mass ratio falls as a function of redshift, other than the fact that higher redshift galaxies tend to have higher sSFRs (Feulner et al. 2005). Note that this does not necessarily exclude differing behaviour as a function of redshift, only that the large uncertainty in stellar masses obscures any relation if it were present. The typical 1σ and 2σ confidence regions are shown by the dark and light green ellipses and were measured by centring all the MC estimates for each galaxy on the median values and finding the ellipses based on the eigenvectors of the covariance of the point cloud. There is a slight correlation introduced by having the unresolved mass present in both axes, but the skew is small compared to the range of data covered, and this correlation could not explain the observable trends in the figure.

Significantly, there appear to be two regimes in Fig. 2: a roughly linear decrease in the mass ratio at sSFRs less than -8.75 , followed by a much sharper decrease beyond. A broken power-law fit is shown by the magenta dashed line to demonstrate this dual nature. Sorba & Sawicki (2015) and Zibetti et al. (2009) found that the presence of dust lanes can drastically affect the mass estimates of galaxies compared to their pixel-by-pixel mass estimates, but a visual inspection revealed no obvious dust lanes or other peculiarities present that could explain the mass discrepancy for the most egregious outliers. False-colour images, as well as mass and SFR maps are provided as an online supplement to this work. For demonstration purposes, the left-hand panel of Fig. 2 is also reproduced using false-colour images in rest-frame ugr and shown in Fig. 3.

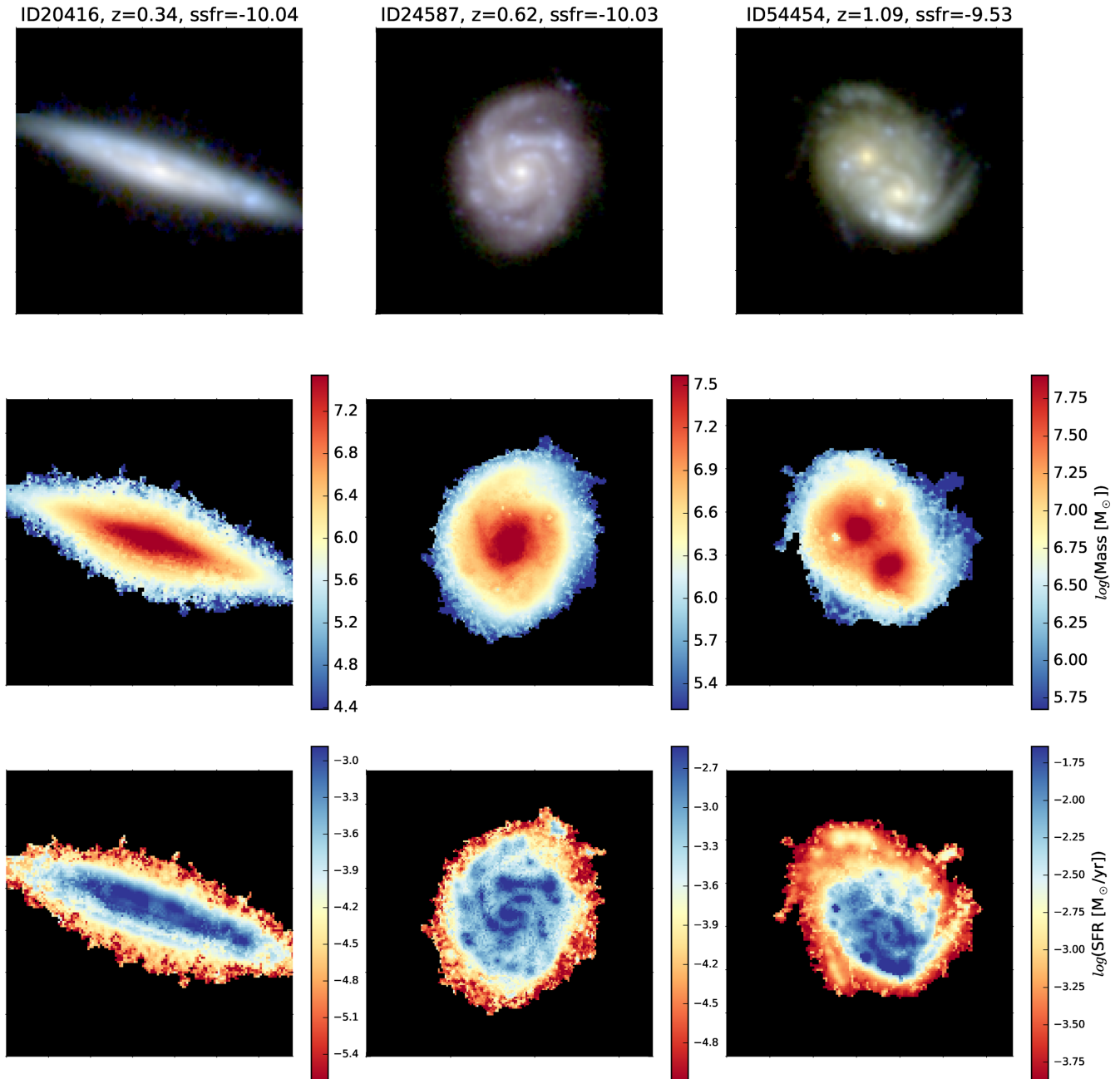


Figure 1. Two-dimensional maps of three representative XDF galaxies created using our model SED template set. From top to bottom we show false-colour images in approximately rest-frame *ugr*, logarithmic stellar mass maps, and logarithmic SFR maps. The galaxies have, from left to right, RA = {53.171 2061°, 53.169 9419°, 53.147 858 33°}, Dec. = {−27.814 710 20°, −27.771 0194°, −27.774 036 11°}, $z_{\text{spec}} = \{0.337, 0.622, 1.088\}$, and $\log(M_*/M_\odot) = \{10.3, 10.7, 10.9\}$ where the mass measurement is the median pixel-by-pixel mass from our final catalogue.

The right-hand panel of Fig. 2 a similar effect, but this time using the photometric redshift for galaxies with $z_{\text{phot}} < 2.5$ (a limit chosen for ease of comparison with other works; see Section 4.2). The same broken power-law trend is visible with these galaxies as with the spectroscopic redshift galaxies. Again, no discernible difference in behaviour with redshift is distinguishable, but this does not necessarily preclude an offset that steepens with redshift. We chose to focus our analysis on a redshift invariant offset that changes behaviour at a certain sSFR, but it is possible that future surveys with greater galaxy numbers could extract a redshift dependence.

Although the shallow linear decrease at low sSFRs is consistent with Sorba & Sawicki (2015), the sharper exponential decay at higher sSFRs is surprising, and perhaps alarming. Taken at face

value, this would mean that a large number of stellar mass measurements would be underestimated by a factor of 2–3 or some as high as 5 (0.3–0.5 or 0.7 dex, respectively). The abrupt change in the nature of the mass ratio/sSFR relation hints that there may be other factors involved.

3.2 Is the effect due to poor signal-to-noise ratio?

It is possible that the poorer S/N ratios for these galaxies is biasing the SED fits, an effect described by Gallazzi & Bell (2009). Indeed if we restrict our examination to only galaxies that have an unresolved median stellar mass estimate greater than $10^{10} M_\odot$ then the second

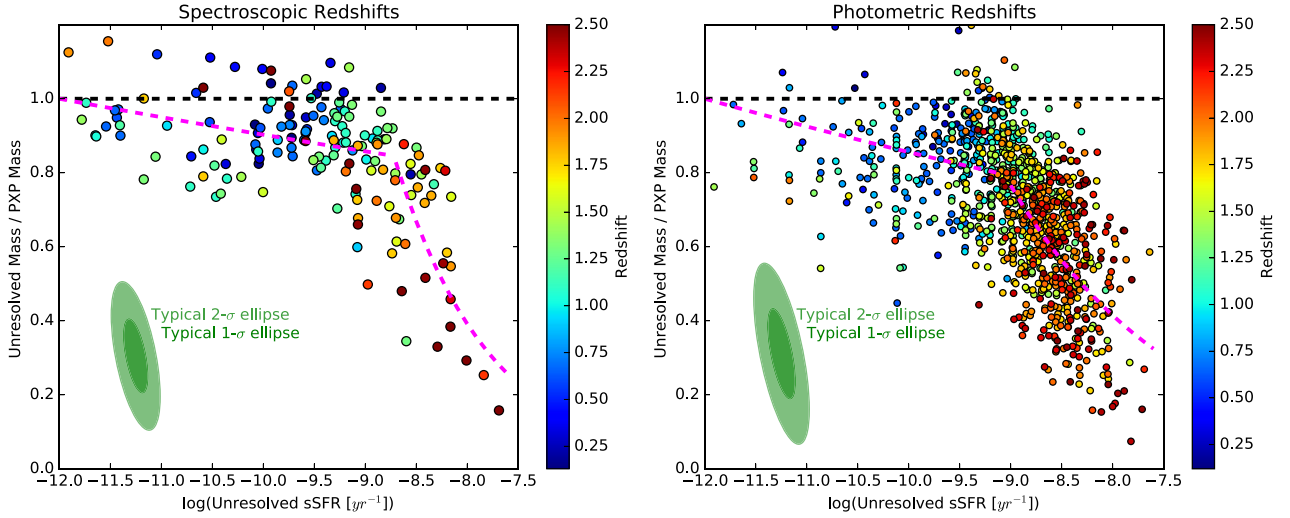


Figure 2. Ratio of the median unresolved mass estimate and the median pixel-by-pixel mass estimate as a function of specific SFR for galaxies in the XDF with spectroscopic redshifts (left) and photometric redshifts (right). Points are coloured based on their redshift. The horizontal black dashed line shows where the two mass estimates are equal. The magenta dashed line shows a broken power-law fit to the data to demonstrate the two regime behaviour.

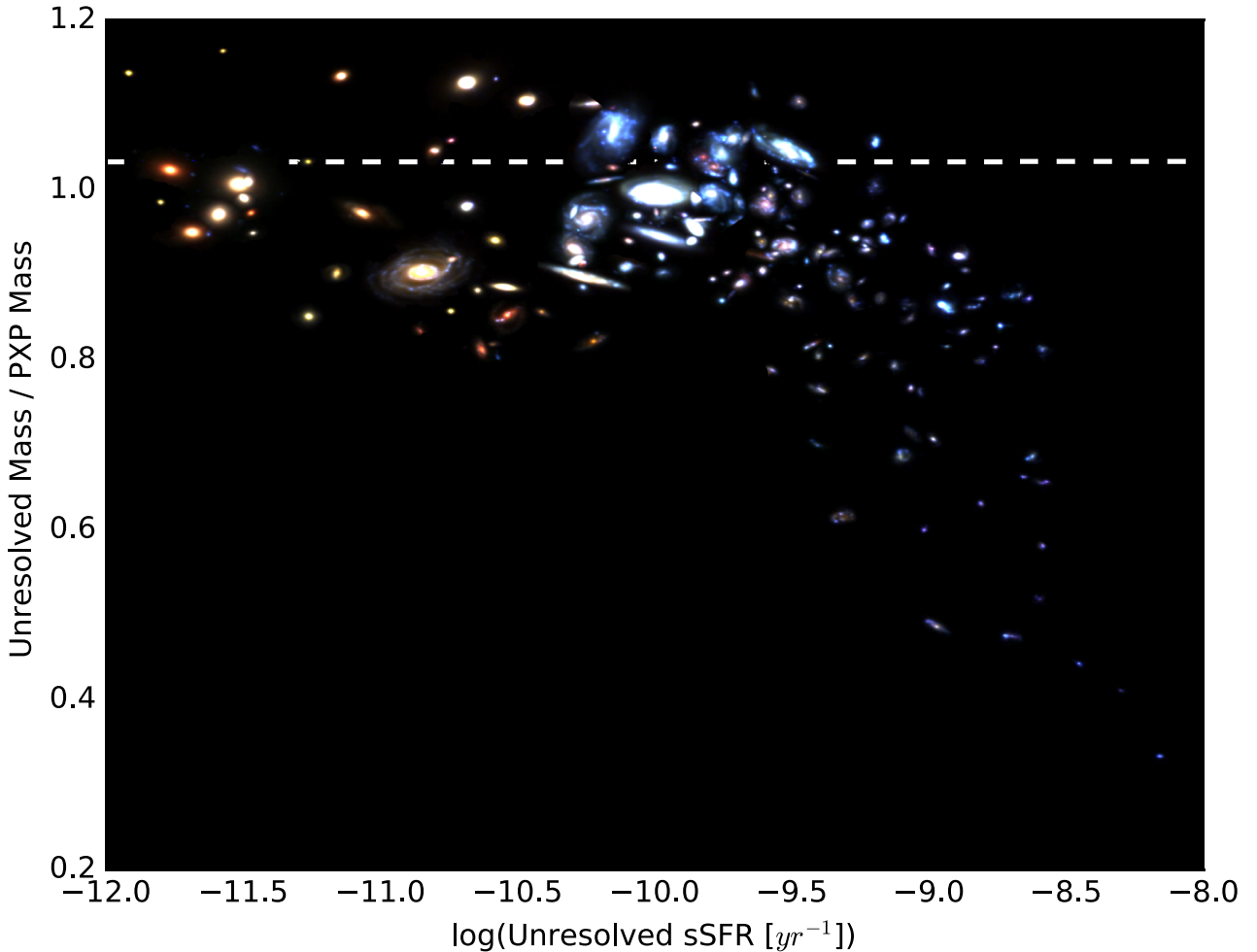


Figure 3. Ratio of the median unresolved mass estimate and the median pixel-by-pixel mass estimate as a function of specific SFR for galaxies in the XDF with spectroscopic redshifts. The false-colour cutouts are made such that blue, green, and red correspond to roughly rest-frame u , g , and r , respectively. No dust lanes or other artefacts seem to be responsible for the large discrepancy at high sSFRs.

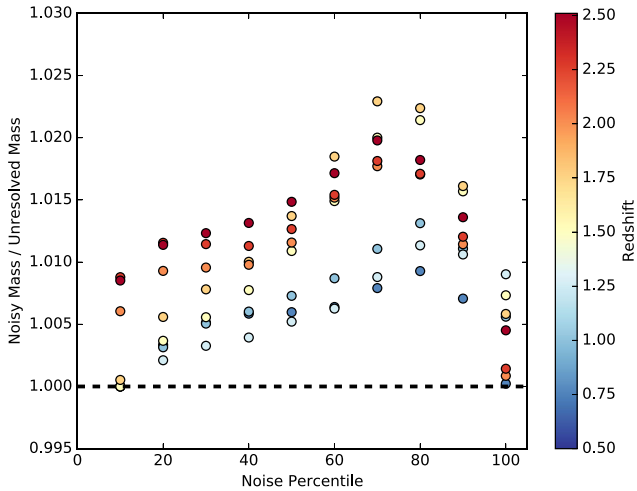


Figure 4. Ratio of mass estimates from unresolved photometry with artificially increased photometric uncertainties to those with the lowest possible uncertainties. The noise level is set by the i th percentile of the photometric uncertainties of every pixel in every galaxy separated in eight redshift bins from 0.5 to 2.5.

steeper power law is not visible, but only because these galaxies all have sSFRs lower than the best-fitting break sSFR of approximately -9 dex. To see how our mass estimates could be biased by the relatively low S/N ratios of pixels, we first found how the noise in each bandpass grows, finding the i th percentiles of the photometric magnitude error of every pixel in galaxies split into eight redshift bins. This yielded the typical magnitude uncertainty (denoted as $\Delta m(B, i, z)$), which is inversely proportional to the S/N and is a function of bandpass (B), redshift, and ‘noisiness’ (i), where i ranges from 10 to 99 in steps of 10 with the 99th percentile being the noisiest pixels. We next took the unresolved photometric catalogues and refit them to our model SED grid 10 different times, each time replacing the photometric uncertainties with successively noisier and noisier Δm . Fig. 4 compares the masses of these noisy fits compared to the mass of the most accurate unresolved mass.

The increase in mass estimate with increasing noise shows that our model SED set does introduce a prior bias, increasing mass by a maximum of approximately 2.5 per cent for the 70th noisiest percentile pixels at the higher redshifts. Interestingly, the mass bias begins to decrease as the pixels become even noisier than that. There is a dichotomy between pixels with redshifts above approximately 1.5; although the curves all have a similar qualitative rise and fall, the higher redshift bias increases more steeply. Now, this is a fractional bias, and the noisiest pixels will also in general be the least massive, and thus contribute less to the total mass bias of the galaxy. For each galaxy, we calculate a predicted fractional mass bias (F) by

$$F = \frac{\sum_x M_x^* B_{p(x),z} - \sum_x M_x^*}{\sum_x M_x^*} \quad (1)$$

where M_x^* is the median stellar mass in the x th pixel and $B_{p(x),z}$ is the bias interpolated from Fig. 4 where $p(x)$ is the average photometric error percentile in that pixel and z is the galaxy’s redshift. The maximum mass bias from low S/N pixels was 2.2 per cent of the mass of the galaxy, which is not significant compared to the factor of 2 (or more) discrepancy. For demonstration purposes, we removed from our sample any galaxies with predicted fractional mass biases greater than 1 per cent. Doing so left us with 684 galaxies shown in Fig. 5, and shifted the median number of pixels in these galaxies

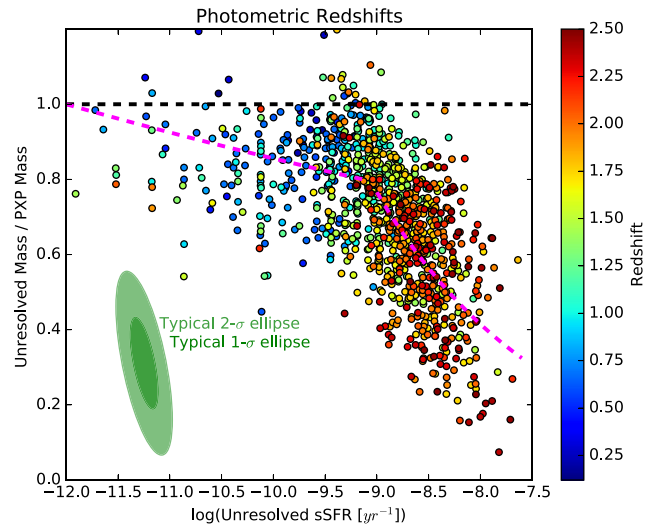


Figure 5. Same as the right-hand panel of Fig. 2, except removing galaxies whose masses could be unduly biased by noisy pixels. The overall kink at high sSFRs is unchanged, which is unsurprising since our maximum estimated bias was only 2.2 per cent, much lower than could account for large relative mass differences at high sSFRs.

to 242 pixels. The double power law is still present, and the feature remains even with more stringent cuts of $F < 0.005$. We conclude that this effect is not due to poor S/N.

3.3 Is the effect due to poor lack of redder bandpasses?

The sharp downturn happens almost exclusively for high-redshift ($z > 1.5$) galaxies, where our reddest bandpass ($F160W$) is only probing rest-frame visual regions of the galaxies’ spectra. A large number of works (Maraston et al. 2006; Kannappan & Gawiser 2007; Bolzonella et al. 2010; Lee et al. 2009; Ilbert et al. 2010; Pforr et al. 2012) found that coverage in the rest-frame NIR was crucial to stellar mass estimates from SED fitting, with Pforr et al. (2012) stating that excluding NIR bandpasses can lead to masses that are underestimated by up to 2 dex. However, Shapley et al. (2005), van der Wel et al. (2006), and Muzzin et al. (2009) did not find as strong a need for NIR coverage. Moreover, our pixel-by-pixel versus unresolved comparison should have both the pixel-by-pixel and unresolved masses affected in a similar manner by the lack of NIR data. It should be emphasized that our method examines relative mass differences, and is unconcerned with evaluating the true total stellar mass in each galaxy.

To see if our lack of redder bandpasses for high-redshift galaxies could be affecting our mass estimates, we make use of the photometric catalogues provided by Lundgren et al. (2014), who included *Spitzer*/IRAC photometry. We cross-correlated that catalogue with that of Rafelski et al. (2015) based on a search in RA and Dec. utilizing a maximum matching distance of 3 arcsec (approximately five pixels in the XDF images), matching only galaxies that had IRAC 3.6 and 4.5 measurements and with a photometric redshift between 1.5 and 2.5. To ensure that the objects in each catalogue were in fact representing the same galaxy, we compared the $F606W$ flux in each catalogue and removed any objects where this did not match to within the 1σ uncertainties, leaving 71 galaxies. We performed (unresolved) SED fitting on these galaxies once using only photometry from the Rafelski et al. (2015) catalogue in the nine XDF bandpasses, and a second time also including the two IRAC

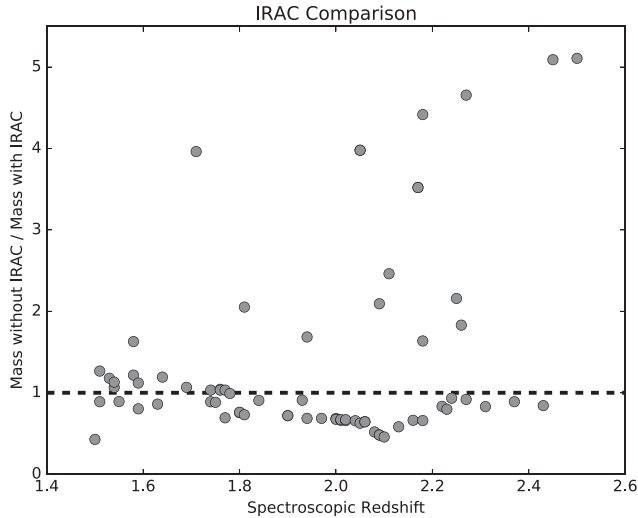


Figure 6. Comparison of mass estimates using photometry from Rafelski et al. (2015) and either including or excluding IRAC photometry from Lundgren et al. (2014).

bandpasses from Lundgren et al. (2014). As shown in Fig. 6, we found no consistent bias in the median MC mass estimates of these $z \sim 2$ galaxies when either including or excluding the two IRAC bandpasses. The median mass ratio ($M_{\text{no IRAC}}^*/M_{\text{with IRAC}}^*$) of the 71 galaxies was 0.89 with upper and lower inner quartile bounds at 1.24 and 0.68, respectively. Examining solely galaxies with $1.9 < z < 2.2$, the median mass ratio does drop to 0.68, leading one to think the effect may become more dramatic as redshift increases, but then the median mass ratio rises up to 0.93 again for galaxies with $z > 2.2$, suggesting that the previous dip was likely noise. Although the masses were, in general, slightly smaller without IRAC photometry, the discrepancy was nowhere near as drastic as that found in the pixel-by-pixel analysis. There were also no discernible trends with either sSFR or redshift. It is further questionable whether any differences in the mass estimates are due to the presence of IRAC photometry, or due simply to systematic differences between the two catalogues used here.

To more deeply probe the effects of including rest-frame NIR photometry on our mass estimates, we refit galaxies with spectroscopic redshifts less than 1.2, only this time excluding the four WFC3 bandpasses. This removes their rest-frame NIR bands, allowing us to test whether NIR data is important. The results are shown in Fig. 7. Similarly to the mass estimates with and without IRAC photometry, the mass estimates of these lower redshift galaxies show no obvious systematic bias when the WFC3 photometry is included/excluded. Again no obvious trend with redshift is present. The median mass ratio when comparing the mass estimates without the four WFC3 bandpasses versus those with is 1.11, with a lower quartile value of 0.89 and an upper quartile value of 1.35. Interestingly, this slight offset is in the opposite direction of that found with the IRAC bandpasses. Here, mass estimates including rest-frame NIR photometry are typically lighter than when NIR photometry is excluded.

In either case, even if rest-frame NIR photometry does have an effect on the absolute mass estimate, the relative mass ratio from comparing pixel-by-pixel to integrated mass estimates for any one galaxy should not be altered, since both mass estimates are made using the same set of bandpasses. Despite the lack of influence from including rest-frame NIR observations, the large body of work stat-

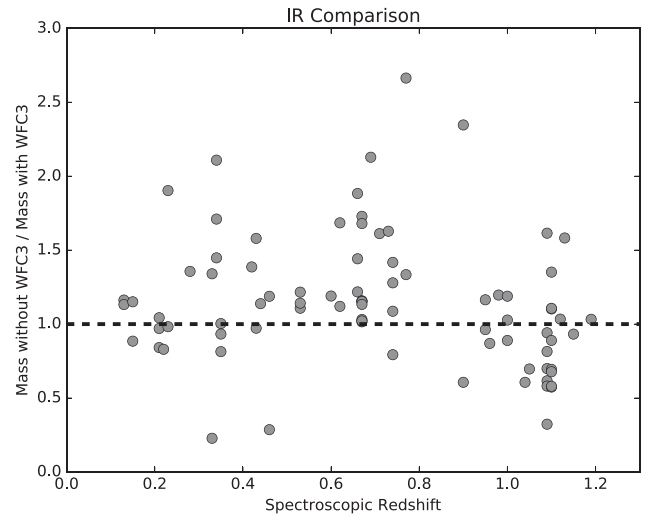


Figure 7. Comparison of mass estimates for $z_{\text{spec}} < 1.2$ galaxies made both including WFC3 photometry and excluding it.

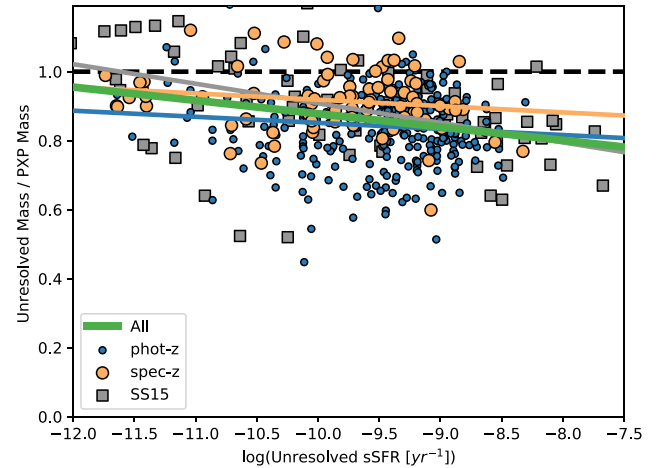


Figure 8. Linear fits to the resolved pixel-by-pixel mass ratio as a function of sSFR. Yellow (blue) points represent XDF galaxies with spectroscopic (photometric) redshifts less than 1.2. The grey squares are $z \sim 0$ SDSS galaxies from Sorba & Sawicki (2015). The yellow, blue, and grey coloured lines show linear fits to their corresponding data set. The thick green line shows a linear fit to the combined data sets. It represents our new best relation for correcting outshining for galaxies in the range $0 < z < 1.2$.

ing their significance has led us to divide our analysis into two parts. In Section 3.4, we restrict our analysis to galaxies with redshifts less than 1.2, meaning all galaxies have photometric observations up to at least rest-frame I band, and compare to previous results from Sorba & Sawicki (2015). In Section 3.5, we examine the increasing discrepancy for galaxies up to $z = 2.5$, and discuss the implications if the effect is real.

3.4 Outshining up to redshift 1.2

Fig. 8 shows the XDF $z_{\text{spec}} \leq 1.2$ and $z_{\text{phot}} \leq 1.2$ samples as yellow and blue circles, respectively, and the SDSS galaxies from Sorba & Sawicki (2015) as grey squares. The best-fitting parameters can be found in Table 1, and uncertainties were found by refitting lines for each of our MC instances and taking the 16th and 84th percentile values. The slopes and intercepts of all three data sets agree to

Table 1. Best-fitting parameters of linear fits $y = mx + b$. All galaxies have $0 < z \leq 1.2$. Here, z_s and z_p , respectively, refer to spectroscopic and photometric redshift samples.

Data set	m	b
SS15	$-0.057^{+0.022}_{-0.019}$	$0.34^{+0.20}_{-0.18}$
XDF z_s	$-0.017^{+0.008}_{-0.021}$	$0.75^{+0.07}_{-0.22}$
XDF z_p	$-0.019^{+0.005}_{-0.012}$	$0.66^{+0.05}_{-0.11}$
XDF ($z_s + z_p$)	$-0.027^{+0.007}_{-0.007}$	$0.60^{+0.07}_{-0.06}$
XDF ($z_s + z_p$) + SS15	$-0.039^{+0.010}_{-0.008}$	$0.48^{+0.10}_{-0.07}$

within uncertainty. The fact that the $z \sim 0$ sample falls so well along the same relation as the XDF samples despite differences in photometry and modelling strongly implies that any effects of outshining do not change significantly up to $z \sim 1$. However, there may be a slight steepening of the slope as redshift decreases from approximately one to the present day, accompanied by a subsequent rise in the intercept. The best-fitting slope of the XDF spec- z and photo- z samples together gives a slope of -0.027 which is shallower than the best-fitting SDSS slope of -0.057 , and the slope fit to just the spectroscopic XDF sample is shallower still at -0.017 . There is a great deal of scatter in the data about the linear trend. Partly, this is due to the effect of dust lanes on pixel-by-pixel fitting (Zibetti et al. 2009; Sorba & Sawicki 2015), and partly due to the inherent noise in stellar mass estimates. Additionally, the linear fits to the XDF data may be affected by the artificial cut-off at $z = 1.2$, as well as the relative lack of low sSFR galaxies at higher redshifts. Combining the XDF and Sorba & Sawicki (2015) galaxies, we fit an updated linear trend shown as the solid red line in Fig. 8. This represents our new best estimate at a correction factor for outshining:

$$M_{\text{resolved}}^* = \frac{M_{\text{unresolved}}^*}{-0.039 \log(\text{sSFR}) + 0.50}, \quad (2)$$

where the M^* are given in M_{\odot} and sSFR is in units of yr^{-1} .

3.5 Outshining up to redshift 2.5

For galaxies at higher redshifts (and thus generally higher sSFRs), the unresolved versus pixel-by-pixel mass ratio decreases sharply at sSFRs greater than roughly 10^{-9}yr^{-1} . In Fig. 9, we show a piecewise fit to this behaviour for galaxies at both low and high redshifts in our sample. The piecewise function is described in log-log space as

$$f(x) = \begin{cases} \mu(x + 12) + \beta & : x \leq p \\ k_2 x^2 + k_1 x + k_0 & : x > p, \end{cases} \quad (3)$$

which is a linear relation to the left of the break point p , and a parabola to the right. The free parameters are p , μ , β , and k_2 . To ensure continuity and differentiability at p , k_0 , and k_1 are constrained to be

$$k_1 = -2k_2 p + \mu, \quad (4)$$

$$k_0 = \beta + 12\mu + (\mu - k_1)p - k_2 p^2. \quad (5)$$

Because of the nature of a piecewise function that can have a shallow slope abruptly changing to a extremely steep slope, we use orthogonal distance regression (ODR) to find the best-fitting parameters, listed in Table 2. ODR minimizes the orthogonal distance from the fitted line to each point (i.e. the line going through the point that intersects the fitted line at a right angle), rather than minimizing the

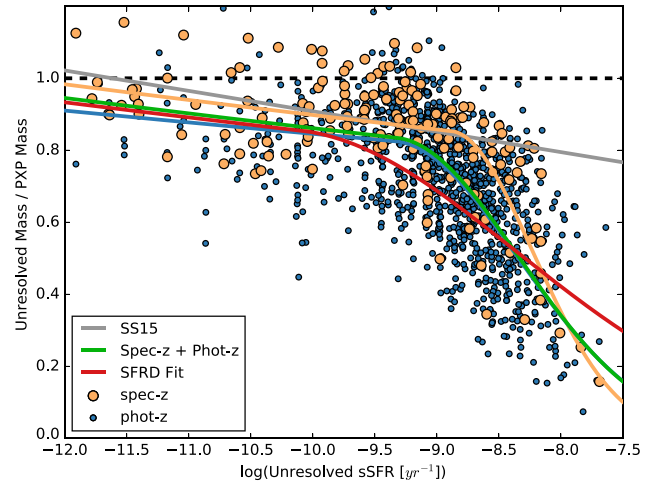


Figure 9. Same as Fig. 8, but including galaxies up to redshift 2.5. Piecewise linear-parabolic fits in log-log space are shown in yellow (fit to the spectroscopic redshift sample) and green (fit to both the spectroscopic and photometric redshift samples). The linear relation from Sorba & Sawicki (2015) is shown for comparison in grey, but is not included in any of the piecewise fits. The red line shows the correction that best matches the observed SFRD history from Madau & Dickinson (2014) (see Section 4.1.4).

Table 2. Best-fitting parameters of piecewise fits. Here, z_s and z_p , respectively, represent the spectroscopic and photometric XDF samples.

Data set	p	μ	β	k_2
XDF z_s	$-8.88^{+0.04}_{-0.18}$	$-0.020^{+0.010}_{-0.004}$	$-0.007^{+0.010}_{-0.018}$	$-0.47^{+0.14}_{-0.05}$
XDF z_p	$-9.34^{+0.49}_{-0.28}$	$-0.016^{+0.006}_{-0.008}$	$-0.041^{+0.043}_{-0.016}$	$-0.20^{+0.12}_{-0.31}$
XDF ($z_s + z_p$)	$-9.32^{+0.47}_{-0.26}$	$-0.020^{+0.011}_{-0.004}$	$-0.024^{+0.027}_{-0.001}$	$-0.21^{+0.12}_{-0.31}$

vertical distance from each point to the fitted line as is done in ordinary least-squares minimization. In this manner, galaxies which are close to the break point p are matched to the part of the piecewise function they are nearest to, rather than whichever part they happen to be above or below. Uncertainties were again found by performing the same fitting procedure on each of the Monte Carlo instances and taking the 16th and 84th percentiles for each parameter.

4 DISCUSSION

4.1 Implications

4.1.1 Corrections to stellar mass measurements

In Section 3, we have shown that spatially resolved SED fitting can give significantly higher stellar mass estimates than unresolved photometry. Taking the spatially resolved fits as more accurate than the cruder, spatially unresolved ones, this suggests that stellar mass estimates done using spatially unresolved photometry need to be corrected upwards. We recommend that those working with stellar masses derived from integrated photometry apply the empirical correction based on equation (3) with constants derived for the spectroscopic+photometric case (third row in Table 2). The correction then is

$$M_{\text{corrected}}^* = M_{\text{unresolved}}^* \times \begin{cases} 10^{0.02s+0.264} & : s \leq -9.23 \\ 10^{0.21s^2+3.9344s+18.505} & : s > -9.23, \end{cases} \quad (6)$$

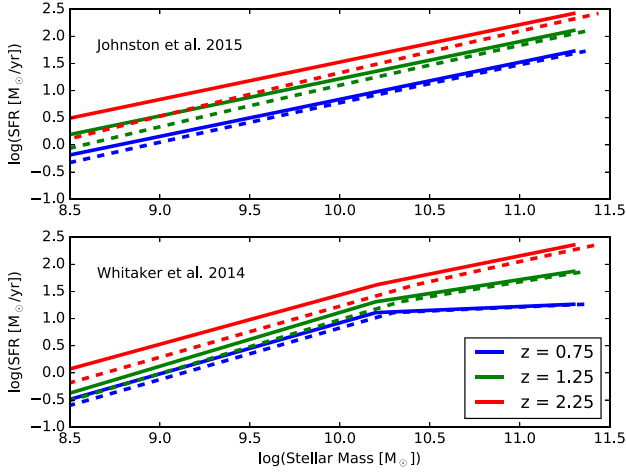


Figure 10. Corrections to previously published star-forming main-sequence relations at various redshifts made using our piecewise fit to the XDF ($z_s + z_p$) galaxies. The top panel shows relations from Johnston et al. (2015) and the bottom panel those of Whitaker et al. (2014). The solid lines show the original relationships, and the dashed lines the corrections based on this work. The blue, green, and red colours show redshift 0.75, 1.25, and 2.25, respectively.

where s is the logarithm of the sSFR in units of yr^{-1} and the M^* quantities are in solar masses.

In Sections 4.1.2–4.1.4 we apply this correction to three concrete situations.

4.1.2 Slope of the main sequence

The increasing disparity between unresolved and pixel-by-pixel mass estimates, if real, has implications for several previously published results derived using masses from SED template fitting. Perhaps most obviously, the relationship between a galaxy’s mass and SFR (when it is actively star forming), known as the star-forming main sequence, is directly affected. In Fig. 10 we show how correcting the galaxies’ masses for outshining would change two recent parametrizations of the star-forming main sequence from Whitaker et al. (2014) and Johnston et al. (2015) at three different redshifts. The sharp decline at higher sSFRs leads to certain areas of the mass–SFR plane changing more drastically than others. For the Johnston et al. (2015) parametrization, the lowest redshift main sequence is barely affected at all, but there is a much larger correction for the higher redshift galaxies because of their increased SFR. The larger correction at higher sSFRs leads to a main sequence which is no longer linear. Non-linear main sequences were also proposed by Whitaker et al. (2014), who fit broken power laws to star-forming galaxies in different redshift ranges. The corrections due to outshining are less dramatic for this parametrization because the kink in the main sequence brings it down out of the highest sSFR region of the mass–SFR plane. Once again, the lowest redshift relation is barely affected, but the $z = 2.25$ has a higher offset.

4.1.3 Intrinsic scatter in the main sequence

Additionally, the intrinsic scatter of the star-forming main sequence is linked to smoothness (or, conversely, burstiness) of the SFH of the Universe (Abramson et al. 2014; Kurczynski et al. 2016). The small observed scatter around the SFR– M_* correlation implies a

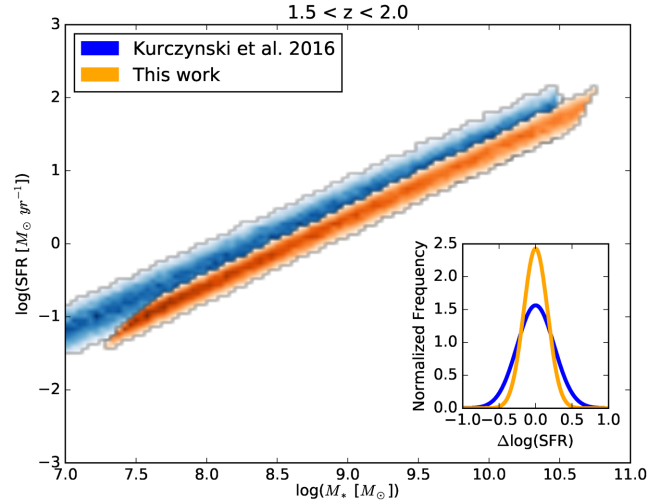


Figure 11. Effects of our mass correction on the intrinsic scatter about the star-forming main sequence for a redshift range $1.5 < z < 2.0$. The blue density map is generated from Kurczynski et al. (2016) relations and converted to Chabrier IMF. The orange density map shows the correction to stellar mass based on this work. The shearing caused by the sSFR-dependent correction leads to a lower intrinsic scatter in $\log(\text{SFR})$ (0.17 versus 0.26), as shown by the distributions about the central relation shown in the inset.

Table 3. Corrections to intrinsic scatter in $\log(\text{SFR})$ around SFMS for various redshift ranges. Original standard deviations are from Kurczynski et al. (2016).

Redshift	σ_{int}	$\sigma_{\text{int_corrected}}$
$0.5 < z < 1.0$	0.43	0.34
$1.0 < z < 1.5$	0.27	0.21
$1.5 < z < 2.0$	0.26	0.17
$2.0 < z < 2.5$	0.28	0.21

steady, gradual assembly of stellar material rather than a history dominated by bursts. Because the bias in mass estimation found here is linked to a galaxy’s sSFR, it causes a shear in the intrinsic scatter. Galaxies located in higher sSFR regions of the SFR– M_* plane are affected more so than others. In Fig. 11, we show that the shearing leads to a smaller intrinsic scatter around the SFR– M_* relation for the redshift range $1.5 < z < 2$. The blue density map shows 500 000 galaxies randomly produced using the intrinsic scatter and star-forming main-sequence parameters of Kurczynski et al. (2016) adapted to a Chabrier IMF rather than a Salpeter. These galaxies were corrected for the effects of outshining based on this work, and the result shown as the orange density map. The shearing and subsequent tightening of the correlation can be seen by the standard deviations about the central relation shown in the inset. Although only one redshift range is shown here, a similar effect happens at all redshifts. The results are summarized in Table 3. The tighter scatter around the star-forming main sequence implies that the formation of stars in galaxies is even smoother than previously thought.

4.1.4 The missing mass problem

Finally, the effects of outshining may shed some light on a long-standing tension between observations of the stellar mass density (SMD) of the Universe and the star formation rate density (SFRD).

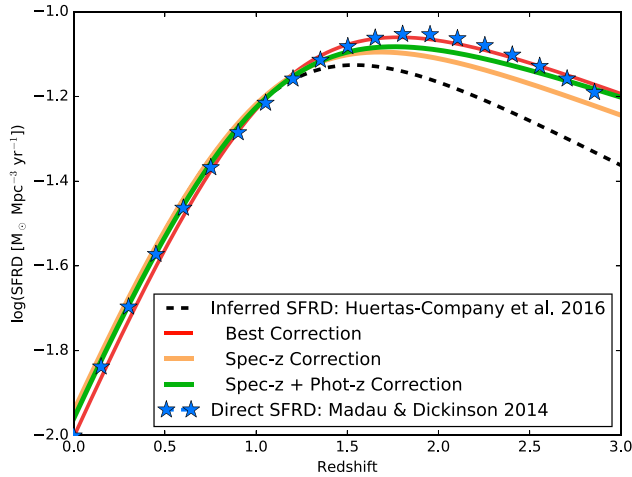


Figure 12. Corrections to the observed cosmic star formation history. The dashed black line shows the SFRD curve inferred from the observed SMD (we show the results of Huertas-Company et al. 2016), and the solid coloured lines the corrected values based on our work. The corrections bring the inferred SFRD into much better agreement with the direct observations of the SFRD compiled by Madau & Dickinson (2014), shown as the blue stars

The SMD at any time should be the integral of the SFRD corrected for mass lost and returned to the interstellar medium. A large body of work has reported a discrepancy between the inferred SFRD and the directly measured one, particularly at higher redshifts. Wilkins, Trentham & Hopkins (2008) found instantaneous indicators of the SFRD were 0.6 dex higher than those fit to the stellar mass history at $z > 2$. Ilbert et al. (2013) compared their inferred star formation history to data compiled by Behroozi, Wechsler & Conroy (2013) and found a difference of 0.2 dex at $z > 1.5$ (although the discrepancy was still within the expected uncertainties). Huertas-Company et al. (2016) recently inferred a star formation history extremely similar to Ilbert et al. (2013) despite a completely different data set, but when they compared their results to a recent compilation of direct SFRD measurements by Madau & Dickinson (2014), they found the direct measurements were ~ 1.25 times larger than their inferred SFRD at $z > 2$ (blue stars versus broken black line in Fig. 12). Reddy & Steidel (2009) and Sawicki (2012b) attributed this difference to poor constraints on the low-mass tail of the stellar mass functions, but the issue has persisted even with more recent and deeper surveys, and an alternate explanation may be required.

To test how outshining would affect SMD measurements, we first took the best-fitting Schechter function parameters for all galaxy morphologies listed in table 2 of Huertas-Company et al. (2016), which were divided at different redshifts into star-forming and passive mass functions. We estimated the SFR of the star-forming galaxies of a given mass and redshift using the main-sequence relation of Johnston et al. (2015). Then, with estimates of the sSFRs of the star-forming population of galaxies, we used equation (6) to correct the star-forming mass function at each redshift for the mass bias found in this work. Following Huertas-Company et al. (2016), we integrated below the corrected curve at each redshift between 10^8 and $10^{12} M_{\odot}$, and added the integral of the passive mass function in that same mass range to find the SMD as a function of redshift. Following the procedure laid out by Wilkins et al. (2008), we related the corrected SMD to the SFRD using

$$\text{SMD} = \int_0^t \text{SFRD}(t') \left(1 - 0.05 \ln \left(1 + \frac{t-t'}{0.3 \text{ Myr}} \right) \right) dt', \quad (7)$$

where the parametrization for the return fraction is assumed to be that given by Conroy & Wechsler (2008) for a Chabrier IMF.

The functional form of the SFRD is taken to be that given by Behroozi et al. (2013), namely

$$\text{SFRD}(z) = \frac{C}{10^{A(z-z_0)} + 10^{B(z-z_0)}}. \quad (8)$$

The SFRD was fit holding A fixed at -1 as done in the previous works. The parameters B , C , and z_0 were determined through χ^2 minimization of equation (7), yielding a redshift-dependent SFRD inferred from observations of stellar mass functions. Our best-fitting parameters were $B = 0.15$, $C = 0.12$, and $z_0 = 0.99$ using the $z_s + z_p$ correction. Because it is not our intent to actually measure the star formation history of the Universe, only show how a pixel-by-pixel mass correction would affect previous results, we have refrained from performing any uncertainty analysis on these parameters.

The results of inferring the SFRD of the Universe are shown in Fig. 12. The dashed black curve shows the original best-fitting star formation history of Huertas-Company et al. (2016), and the most recent compilation of direct SFRD measurements from Madau & Dickinson (2014) is shown as blue stars. The discrepancy between these two different measurements is evident, particularly above $z \sim 1.5$. Our correction to the inferred SFR history is shown as the yellow and green curves for fits to the z_s and $(z_s + z_p)$ XDF samples, respectively. They follow the original curve of Huertas-Company et al. (2016) closely at low redshift, but begin to diverge significantly around $z > 1.5$, and become more inline with the direct SFRD measurements. There is still a maximum offset of 0.03 dex at $z \sim 2$ between our $z_s + z_p$ correction and the direct measurements, and the slope at the highest redshifts may be too shallow compared to that found by Madau & Dickinson (2014). These differences may arise due to the different functional parametrization used by Madau & Dickinson (2014), or due to our assumed conversion between a galaxy's mass and SFR given by Johnston et al. (2015), which had a higher sSFR for lower mass galaxies than the relation from Whitaker et al. (2014). In general, it is clear that the mass correction presented in this work acts to increase the SMD at higher redshifts, and thus greatly reduce the tension between inferred (from stellar mass functions) star formation histories and directly measured ones.

We perform a complementary test to see how close a mass correction of the same piecewise form given by equation (3) could come to the direct observations. For this, we converted the Madau & Dickinson (2014) SFRD into a SMD as a function of redshift. We then found the best piecewise parameters that would make the Huertas-Company et al. (2016) mass functions match this SMD. Note that this is in a sense the reverse process of that given earlier in this section: instead of having corrected stellar mass functions and fitting a functional form of the SFRD, we now are given a functional form of the SFRD from Madau & Dickinson (2014) and are going to determine what correction needs to be applied to the stellar mass functions. The resulting best correction is shown in Figs 9 and 12 as the solid red curve, and it is generated using the parameters p , m , b , $k_2 = -10.1, -0.02, -0.03, -0.06$. The best-fitting correction to match the directly measured SFRD has a linear component that is within uncertainty of the XDF fits, but a break-point (p) that lies just to the left the 1σ confidence region, and a parabolic component that is shallower than would be expected from the XDF data. The general similarity between the form of the best-fitting correction and the XDF corrections (if not the exact details) lends further credence to the observed kink in the mass discrepancy at high

sSFRs. Overall, we find that correcting the observed stellar mass functions for mass missed by unresolved SED fitting eliminates the ‘missing mass’ problem that is otherwise seen when comparing direct measurements of the SFRD with those inferred from the SMD.

A potential caveat regarding these results is that the Hubble XDF is a small slice of the sky compared to the scale that most stellar mass functions are measured from [4 arcmin² compared to 880 arcmin² for Huertas-Company et al. (2016), or 1.62 deg² for Muzzin et al. (2013)]. The (relatively) small number of galaxies limit how sweeping our conclusions can be. For example, there are only a small number of galaxies (55) with masses between 10^{10.4} and 10^{11.4} M_⊙, which corresponds roughly to plus or minus three times M^* . While the M^* galaxies in our sample all follow the trend shown in Fig. 9, they all have sSFRs that lie to the left of the break point, meaning it is unclear if the population of M^* galaxies would also turn sharply downward. Fortunately, the XDF does include a large number of sub- M^* galaxies, which is where the bulk of stellar mass lies [approximately 80 percent as shown in fig. 1 of Sawicki & Thompson (2006) for a faint end slope α of -1.4 , consistent with deep SMF estimates from Tomczak et al. (2014)]. In order to study how pixel-by-pixel mass estimates perform with varying galaxy populations, greater numbers are needed, and must wait for the onset of next generation high-resolution wide-field astronomical imagers, such as *WFIRST*.

4.2 Why was this bias not seen before?

As discussed above, Sorba & Sawicki (2015) observed a bias in stellar mass measurement due to outshining for nearby galaxies at low redshift. This work shows a similar trend for galaxies up to $z = 1.2$, and a more egregious offset in mass for galaxies from $1.2 < z < 2.5$. However, previous pixel-by-pixel analysis at high redshift by Wuyts et al. (2012) found zero mass discrepancy when comparing the pixel-by-pixel and unresolved mass estimates for star-forming galaxies at $0.5 < z < 1.5$ and $1.5 < z < 2.5$. Although their selection criteria differed from ours (they only performed pixel-by-pixel SED fitting on galaxies with masses greater than 10¹⁰ M_⊙ and log sSFRs greater than -9.76 and -9.51 yr⁻¹ for their low- z and high- z samples, respectively), our results indicate that they should have seen an unresolved versus resolved mass bias regardless. It is necessary to understand what led to our disparate results. Perhaps the most obvious factor that could have contributed to the outshining bias not being measured is that it is a small effect, at least for galaxies with lower sSFRs. When averaged over all sSFRs, our XDF low-redshift ($z \leq 1.2$) spectroscopic sample displays a mean bias of only 0.04 dex. It is only the increasing difference with sSFR that makes the bias particularly notable. Nevertheless, here we examine the differences in our methodologies to see what other factors would influence the unresolved to pixel-by-pixel mass ratios.

The first difference is Wuyts et al. (2012) constructed their model SED templates with Bruzual & Charlot (2003) models, whereas we used FSPS models. Although the manner in which various codes treat certain phases of stellar evolution such as the Thermally Pulsating Asymptotic Giant Branch (TP-AGB) can have strong effects on M_*/L ratios, this should not affect the mass ratios in each work, as both the pixels and integrated light are fit consistently within that work.

Second, Wuyts et al. (2012) used a modified χ^2 statistic which incorporated additional integrated fluxes (U , K_s , and the four IRAC

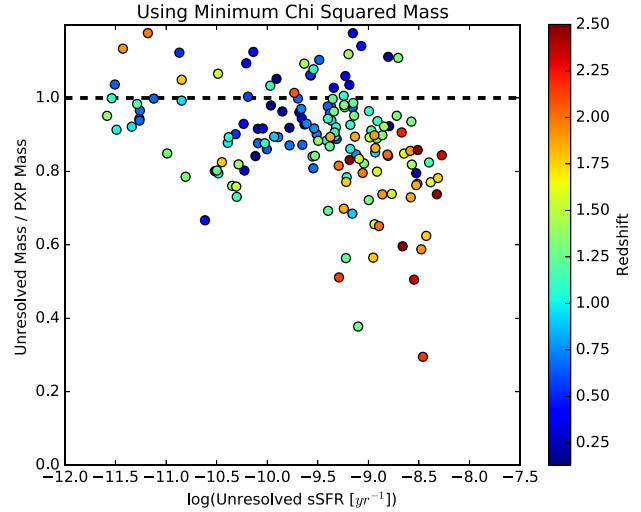


Figure 13. Same as the left-hand panel of Fig. 2, but showing results of using the minimum χ^2 parameter estimates. The linear trend at lower sSFRs is de-emphasized and the mass discrepancy plummets more abruptly with several galaxies still remaining near the one-to-one line at higher sSFRs.

bandpasses) in order to find the best-fitting model. The statistic they minimized was

$$\chi_{\text{res+int}}^2 = \chi_{\text{res}}^2 + \sum_{j=1}^{N_{\text{int}}} \frac{\left(F_j - \sum_{i=1}^{N_{\text{bin}}} M_{i,j}\right)^2}{E_j^2}, \quad (9)$$

where χ_{res}^2 is the sum of all the χ^2 values as normally found in SED fitting for each spatial bin (essentially each pixel) using only bandpasses that are fully spatially resolved, N_{int} is the number of spatially integrated-only bandpasses, N_{bin} is the number of spatial bins, F_j and E_j are the observed flux and error in the j th integrated-only bandpass, and $M_{i,j}$ the model flux in the j th integrated only bandpass and i th spatial bin. Because the parameter space of this statistic is very large, they started with an initial guess using the best-fitting model of each spatial bin found using only the resolved bandpasses and then iteratively adjusted one spatial bin’s model at a time to improve the $\chi_{\text{res+int}}^2$. This iteration was continued until no improvement was found or a maximum of 500 iterations. In contrast, our pixels were treated essentially as their own, independent, objects. This would seem to be a major difference in our approaches, but Wuyts et al. (2012) clearly state that the good correspondence between pixel-by-pixel and unresolved mass estimates remains even if constraints from integrated fluxes were ignored.

Perhaps their choice to use the model with the minimum χ^2 as their best-fitting mass estimate versus our use of the median mass from several hundred Monte Carlo iterations makes a difference. Taylor et al. (2011) have shown how using the best-fitting model to define the stellar mass can induce strong systematic effects. A simple check of finding the mass ratios of our best-fitting model masses (rather than median masses) for the spectroscopic sample gives an average mass ratio of 0.94 ± 0.10 which is higher than the average of the median mass ratios of 0.91 ± 0.06 , but not significantly. More telling, as shown in Fig. 13 the minimum χ^2 mass ratios show a lack of a linear trend with sSFR, with a slope of only -0.014 ± 0.11 , until sharply dropping at practically a right angle. The downward plunge in mass ratio is much more sudden, and driven by the high sSFR galaxies, while the rest remain close to the one-to-one line. This means that for a random selection of

galaxies, the increasing bias with sSFR would not be as obvious when using mass derived from minimized χ^2 fits, and emphasizes the value of using the median mass of many perturbed SED fits. However, by selecting galaxies with only high sSFRs, the overall bias should still be present in the sample used in Wuyts et al. (2012), even if the minimum χ^2 masses would not show the bias as strongly as median mass estimates.

Finally, in order to increase S/N in their spatial bins, Wuyts et al. (2012) combine pixels together using the Voronoi two-dimensional binning technique of Cappellari & Copin (2003). The pixel binning degrades their spatial resolution, which is a key factor in allowing pixel-by-pixel analysis to uncover any bias due to outshining. The pixel-by-pixel SED fitting allows one to separately fit spatially segregated stellar populations, thus better constraining the M_*/L ratio of the galaxy built up over its entire star-forming history. As the spatial resolution gets poorer and poorer, the ability to see fainter stars behind the bright stars necessarily goes away. Sorba & Sawicki (2015) studied the effects of resolution on pixel-by-pixel mass estimates and found that a minimum spatial resolution of 2–3 kpc was required to reveal a mass discrepancy. At the pixel scale of the XDF (60 mas) and redshift two, this would mean binning together 16–25 pixels would be enough to obscure the effects of outshining, an amount commonly needed to meet the S/N criteria of 10 required by Wuyts et al. (2012), particularly in the intermediate and outer regions of galaxies. It is likely then, that the spatial binning performed by Wuyts et al. (2012) pushed the pixel-by-pixel mass estimates into a regime where outshining could not be measured. In fact, when they compared resolved to unresolved mass estimates without binning, they found the pixel-by-pixel masses were systematically heavier by 0.2 dex (about 27 per cent). They state that the binning to a minimum S/N is necessary to avoid biases in SED fits that lead to erroneously large M_*/L ratios. This is a true concern as discussed above, exacerbated by their use of minimum χ^2 best fits. Their use of binning was entirely justified, especially considering the focus of their analysis on star-forming clumps and overall surface profiles as a function of half-mass/half-light radii. The binning did, however, along with the use of minimum χ^2 masses, act to obscure evidence of the outshining bias found in our work.

5 CONCLUSIONS

In this work we fit spatially resolved galaxies in the Hubble XDF with spectral synthesis models. We focused on comparing galaxy stellar mass estimates from spatially resolved model fits to those from spatially unresolved, integrated light. Our key result is that spatially unresolved fitting underestimates galaxy masses compared to the more realistic, spatially resolved fits. For high sSFR galaxies the discrepancy can be large, as large as a factor of ~ 5 in some cases.

In detail, our findings are as follows:

(i) Following Sorba & Sawicki (2015), we examined the ratio of the unresolved mass estimate to the pixel-by-pixel mass estimate as a function of sSFR. We found that a bias in the mass estimate of galaxies due to outshining still persists up to $z = 2.5$. Moreover, the bias appears to have two distinct components:

(a) At low sSFRs (less than approximately $10^{-9.2} \text{ yr}^{-1}$), the ratio of unresolved to pixel-by-pixel mass changes slowly and roughly linearly with $\log(\text{sSFR})$. The slope and intercept of a linear fit is commensurate with the results found for nearby galaxies in Sorba & Sawicki (2015).

(b) Above $\text{sSFR} = 10^{-9.2} \text{ yr}^{-1}$, the mass ratio turns sharply downward, and can no longer be represented by the shallow linear trend found at lower redshifts. Because sSFRs tend to be larger at higher redshift, the sharply increasing bias implies that masses for some galaxies between redshifts 1.5 and 2.5 could systematically be underestimated by factors of 2–5.

(ii) The piecewise form of the mass correction (shallow at low sSFRs and becoming much steeper as sSFR increases) naturally resolves a long-standing tension between the directly observed SFRD and that derived from the observed SMD. Correcting stellar mass functions from Huertas-Company et al. (2016) brings the derived SFRD much more in line with the directly observed SFRD compilation of Madau & Dickinson (2014).

(iii) Correcting mass estimates for the bias found above slightly alters the slope of the star-forming main sequence. It also reduces the intrinsic scatter about the star-forming main sequence, suggesting that star formation in galaxies proceeds in a smooth fashion.

Spatially resolved SED fitting is a powerful tool, and differences between resolved and unresolved fits highlight the importance of treating galaxies as composite structures, rather than one cohesive whole.

In practical terms, the more realistic, spatially resolved fits can give significantly higher masses than the unresolved fits that are necessarily used with unresolved, ground-based data. We recommend that studies using unresolved fits correct their stellar mass estimates using a formula such as the one we provided in equation (6).

ACKNOWLEDGEMENTS

We thank the anonymous referee for many insightful comments that greatly improved the quality of this paper. We are indebted to the high-quality data products made publicly available by the Hubble XDF Team. We are very grateful to Anneya Golob, Liz Arcila-Osejo, and Adam Muzzin for useful discussions related to this work. RS wishes to thank his colleagues at both institutions for their support.

This work was supported financially by the Natural Sciences and Engineering Research Council (NSERC) of Canada, including an NSERC Graduate Scholarship to RS and an NSERC Discovery Grant to MS.

Computational facilities are provided by ACEnet, the regional high performance computing consortium for universities in Atlantic Canada. ACEnet is funded by the Canada Foundation for Innovation (CFI), the Atlantic Canada Opportunities Agency (ACOA), and the provinces of Newfoundland and Labrador, Nova Scotia, and New Brunswick.

This research made use of *ASTROPY*, a community-developed core PYTHON package for Astronomy (Robitaille et al. 2013), as well as many of the routines available in the *SCIPY* and *NUMPY* software packages.

REFERENCES

- Abraham R. G., Ellis R. S., Fabian A. C., Tanvir N. R., Glazebrook K., 1999, *MNRAS*, 303, 641
 Abramson L. E., Kelson D. D., Dressler A., Poggianti B. M., Gladders M. D., Oemler A., Vulcani B., 2014, *ApJ*, 785, L36
 Alard C., Lupton R. H., 1998, *ApJ*, 503, 325
 Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *MNRAS*, 422, 449

- Becker A. C., Homrighausen D., Connolly A. J., Genovese C. R., Owen R., Bickerton S. J., Lupton R. H., 2012, *MNRAS*, 425, 1341
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Benitez N., 2000, *ApJ*, 536, 571
- Benitez N. et al., 2004, *ApJS*, 150, 1
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bolzonella M. et al., 2010, *A&A*, 524, A76
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Byler N., Dalcanton J. J., Conroy C., Johnson B. D., 2017, *ApJ*, 840, 44
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Cappellari M., Copin Y., 2003, *MNRAS*, 342, 345
- Chabrier G., 2003, *PASP*, 115, 763
- Coe D., Benitez N., Sanchez S. F., Jee M., Bouwens R., Ford H., 2006, *AJ*, 132, 926
- Conroy C., 2013, *ARA&A*, 51, 393
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Wechsler R. H., 2008, *ApJ*, 696, 620
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Davidzon I. et al., 2017, *A&A*, 605, A70
- Eisenstein D. J. et al., 2011, *AJ*, 142, 72
- Feulner G., Gabasch A., Salvato M., Drory N., Hopp U., Bender R., 2005, *ApJ*, 633, L9
- Gallazzi A., Bell E. F., 2009, *ApJS*, 185, 253
- Grogin N. A. et al., 2011, *ApJS*, 197, 35
- Huertas-Company M. et al., 2016, *MNRAS*, 462, 4495
- Ilbert O. et al., 2010, *ApJ*, 709, 644
- Ilbert O. et al., 2013, *A&A*, 556, A55
- Illingworth G. D. et al., 2013, *ApJS*, 209, 6
- Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, *MNRAS*, 453, 2540
- Kannappan S. J., Gawiser E., 2007, *ApJ*, 657, L5
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Koekemoer A. M. et al., 2011, *ApJS*, 197, 36
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Krist J., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E., eds, *ASP Conf. Ser. Vol. 77, Astronomical Data Analysis Software and Systems IV*. Astron. Soc. Pac., San Francisco, p. 349
- Krist J. E., Hook R. N., Stoehr F., 2011, in Kahan M. A., ed., *Proc. SPIE, Vol. 8127, Optical Modeling and Performance Predictions V*. SPIE, Bellingham, p. 81270J
- Kurczynski P. et al., 2016, *ApJ*, 820, L1
- Lee S.-K., Idzi R., Ferguson H. C., Somerville R. S., Wiklund T., Giavalisco M., 2009, *ApJS*, 184, 100
- Lundgren B. F. et al., 2014, *ApJ*, 780, 34
- Madau P., 1995, *ApJ*, 441, 18
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Maraston C., Daddi E., Renzini A., Cimatti A., Dickinson M., Papovich C., Pasquali A., Pirzkal N., 2006, *ApJ*, 652, 85
- Maraston C., Pforr J., Renzini A., Daddi E., Dickinson M., Cimatti A., Tonini C., 2010, *MNRAS*, 407, 830
- Muzzin A., Marchesini D., van Dokkum P. G., Labbé I., Kriek M., Franx M., 2009, *ApJ*, 701, 1839
- Muzzin A. et al., 2013, *ApJ*, 777, 18
- Papovich C., Dickinson M., Ferguson H. C., 2001, *ApJ*, 559, 620
- Pforr J., Maraston C., Tonini C., 2012, *MNRAS*, 422, 3285
- Pirzkal N. et al., 2005, *ApJ*, 622, 319
- Rafelski M. et al., 2015, *AJ*, 150, 31
- Reddy N. A., Steidel C. C., 2009, *ApJ*, 692, 778
- Robitaille T. P. et al., 2013, *A&A*, 558, A33
- Sawicki M., 2012a, *PASP*, 124, 1208
- Sawicki M., 2012b, *MNRAS*, 421, 2187
- Sawicki M., Thompson D., 2006, *ApJ*, 648, 299
- Sawicki M., Yee H. K. C., 1998, *AJ*, 115, 1329
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, *ApJ*, 626, 698
- Soille P. J., Ansoult M. M., 1990, *Signal Process.*, 20, 171
- Sorba R., Sawicki M., 2015, *MNRAS*, 452, 235
- Taylor E. N. et al., 2011, *MNRAS*, 418, 1587
- Tomczak A. R. et al., 2014, *ApJ*, 783, 85
- 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. et al., 2012, *ApJS*, 203, 24
- Welikala N., Connolly A. J., Hopkins A. M., Scranton R., Conti A., 2008, *ApJ*, 677, 970
- Whitaker K. E. et al., 2014, *ApJ*, 795, 104
- Wilkins S. M., Trentham N., Hopkins A. M., 2008, *MNRAS*, 385, 687
- Wuyts S. et al., 2011, *ApJ*, 738, 106
- Wuyts S. et al., 2012, *ApJ*, 753, 114
- Zibetti S., Charlot S., Rix H.-W., 2009, *MNRAS*, 400, 1181

SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://academic.oup.com/mnras/article/476/2/1532/4828361) online.

[supplementalFigs.tar.gz](#)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: MODEL PARAMETER SPACE STUDY

To state the obvious, the parameter estimates derived from SED fitting depend on the model template grid used; a poorly suited model grid will give poor results. This is true with pixel-by-pixel SED fitting as well, but because the pixel-by-pixel allows for the creation of 2D property maps, we have an extra check on the validity of our models that is not possible when fitting the entire galaxy's light as one. This is shown in Fig. A1 where we examine property maps created when using two-component burst SFHs as in Sorba & Sawicki (2015). We can see the two-dimensional stellar mass and SFR distributions are not at all smooth as would be expected from the false-colour images. The stellar mass maps contain wispy filaments of pixels much less massive than the surrounding stellar material, and the SFR maps show no discernible structure when we would expect them to roughly trace the blue stars in the spiral arms. Why two-component burst models would work well for SDSS galaxies, but fail for XDF galaxies is an open question. It could be that the SDSS observations probe smaller physical scales, but it may also be related to the results of Gallazzi & Bell (2009), who found that including a large fraction of bursty models could underestimate the mass-to-light ratio of galaxies with smoother SFHs. As we observe galaxies at greater redshifts, it becomes more likely that they will have had smoother SFHs, as any recent bursts will have stellar populations not too dissimilar from the underlying older stellar populations.

It could also be that the two-component burst models are too degenerate at higher redshifts. Specifically, the mass fraction of the young stellar component cannot be constrained by the observations, and not only could the models fail to capture a true underlying older population of stars, they could erroneously add a large fraction of mass to a strictly younger population of stars by overestimating the older stellar component. This would not be the case for low-redshift galaxies in the SDSS, since it is very unlikely to observe a pixel actively forming stars that did not have an underlying older stellar component. The models may overfit the older component somewhat at low redshift, but not nearly as

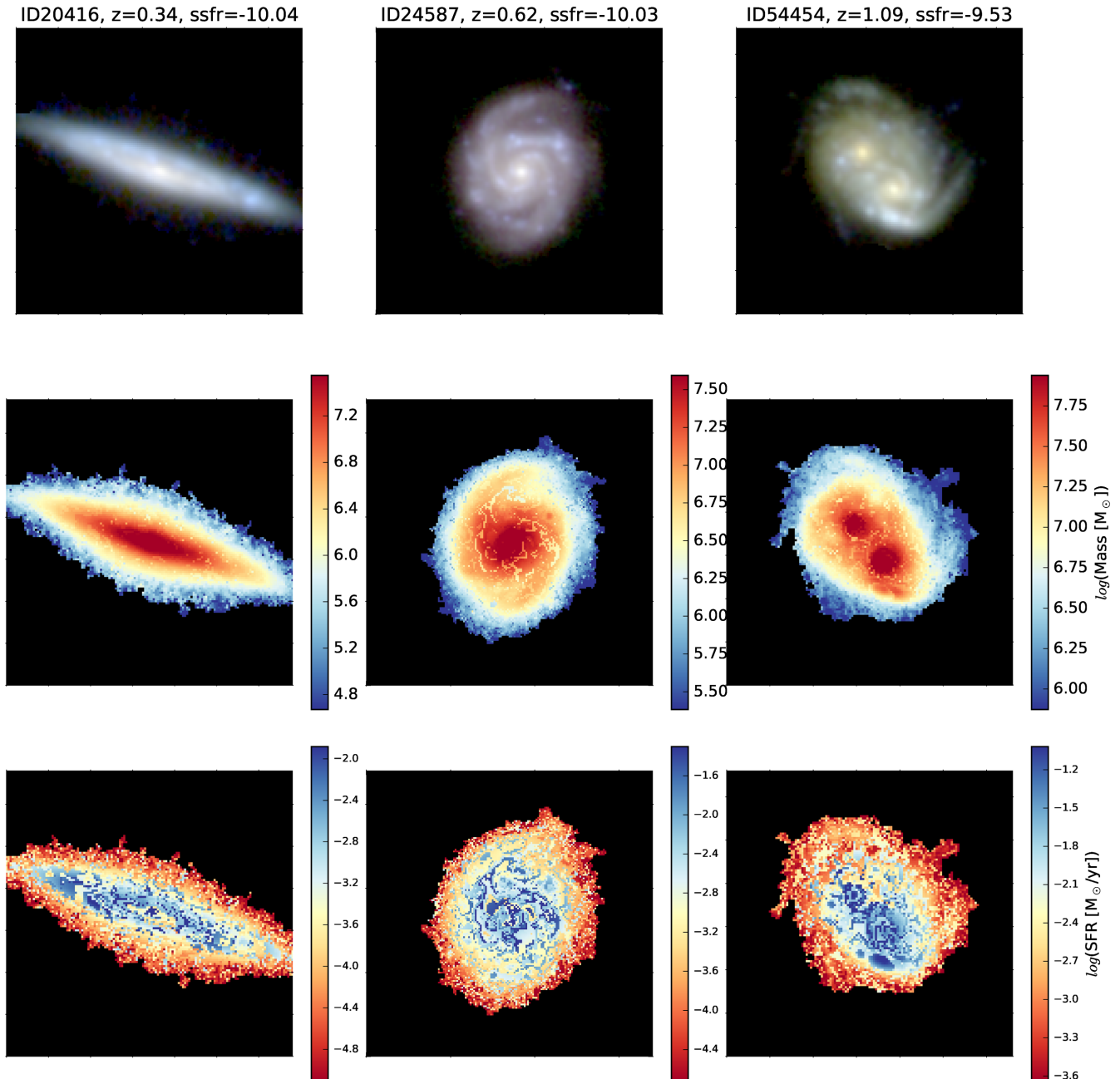


Figure A1. Two-dimensional maps of three representative XDF galaxies created using the same model SED set as Sorba & Sawicki (2015). From top to bottom we show false-colour images in approximately rest-frame *ugr*, stellar mass, and SFR. This model set did not provide adequate pixel-by-pixel SED fitting. The galaxies have, from left to right, RA = {53.171 2061°, 53.169 9419°, 53.147 858 33°}, Dec. = {−27.814 710 20°, −27.771 0194°, −27.774 036 11°}, $z_{\text{spec}} = \{0.337, 0.622, 1.088\}$, and $\log(M_*/M_\odot) = \{10.3, 10.7, 10.9\}$ where the mass measurement is the median pixel-by-pixel mass from our final catalogue.

drastically as is possible at higher redshifts, where there has not been as much time to assemble mass in the older population. The overestimation of the older stellar component would also affect SFR measurements since each component’s SFH is modelled as an exponential decay. This would explain the poor reproduction of the star-forming regions seen in the bottom row of Fig. A1. To ensure realistic spatially resolved stellar property maps, we use one-component τ -models, which were shown to give qualitatively similar mass estimates as two-component burst models for galaxies in the SDSS (Sorba & Sawicki 2015). Note that with one-component models, outshining will always act to hide the mass of older stars.

We found that it is imperative to have a parameter space that is well sampled with regards to SFH when doing pixel-by-pixel SED fitting. Figs A2 and 1 show stellar mass and SFR maps when using coarse and fine model grids, respectively. The coarse grid contains six different values for τ , ranging from 0.3 to 10 Gyr and the spacing between each step roughly doubling each time, and ten different ages between $10^{7.75}$ and 10^{10} yr, with the exponent increasing in steps of 0.25. As can be seen in Fig. A2, although the stellar population property maps are smoother than those of the burst models, the SFR map in particular still shows sharp discontinuities that seem unrelated to any features visible in the false-colour images. These discontinuities disappear when using finer spacings for the SFH

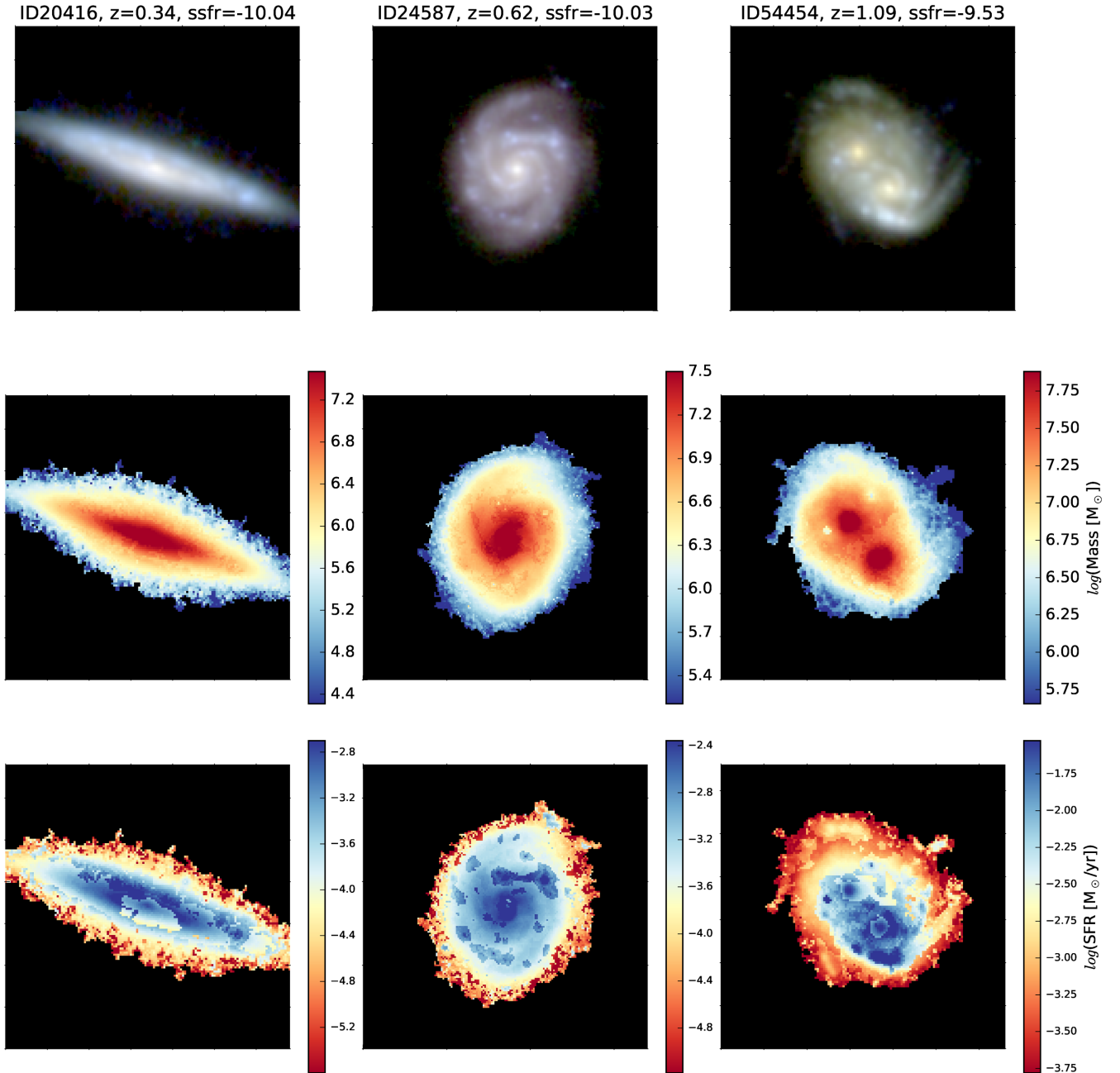


Figure A2. Same as Fig. A1, except created using a coarse grid of τ -models. These show a slight improvement over the burst models.

parameters, as shown in Fig. 1 of the main text. Here the model grid consists of 20 different τ values spaced roughly evenly logarithmically between 0.3 and 10, and $\log(\text{ages}/\text{yr})$ ranging from 7.7 to 10.1 in steps of 0.05. The finer model grid displays much smoother property maps with no unphysical discontinuities. It is this set of models that we chose to use as our ultimate template grid. Note that the need for a finely spaced model grid may also have played a role in the poor performance of the two-component models above. However, because two-component models have many more free parameters, the total number of models grows quickly as the resolution of each parameter is increased, and can become cumbersome to work with.

Studying the right-most mass map of Fig. 1 we can see an example of outshining happening on a small scale as evidenced by the two ‘dimples’ of less massive pixels to the left and right of the upper bulge. It is unlikely that there are massive holes in the

galaxy when the rest of the mass map shows a very smooth gradient. These two areas correspond to the bright blue star-forming regions visible in the false-colour image, indicating that the youngest stars are obscuring an older stellar population, leading to an erroneous M_*/L ratio. Similar dimples can also be seen for the brightest blue spots in the spiral arms of the left-most galaxy. This effect seems unavoidable with τ -models, and even in SED fitting general since the light from very different SFHs can be dominated by the younger stellar populations. With no easy way to distinguish between SFHs with different M_*/L ratios, noisy observations can easily be fit to models that underestimate the stellar mass, even if the correct SFH was present in the model grid. We do not correct our pixel-by-pixel masses for this dimpling effect, but estimating the missing mass by interpolating the surrounding pixels shows that any correction would be of order less than 1 per cent of the total mass of the galaxy.

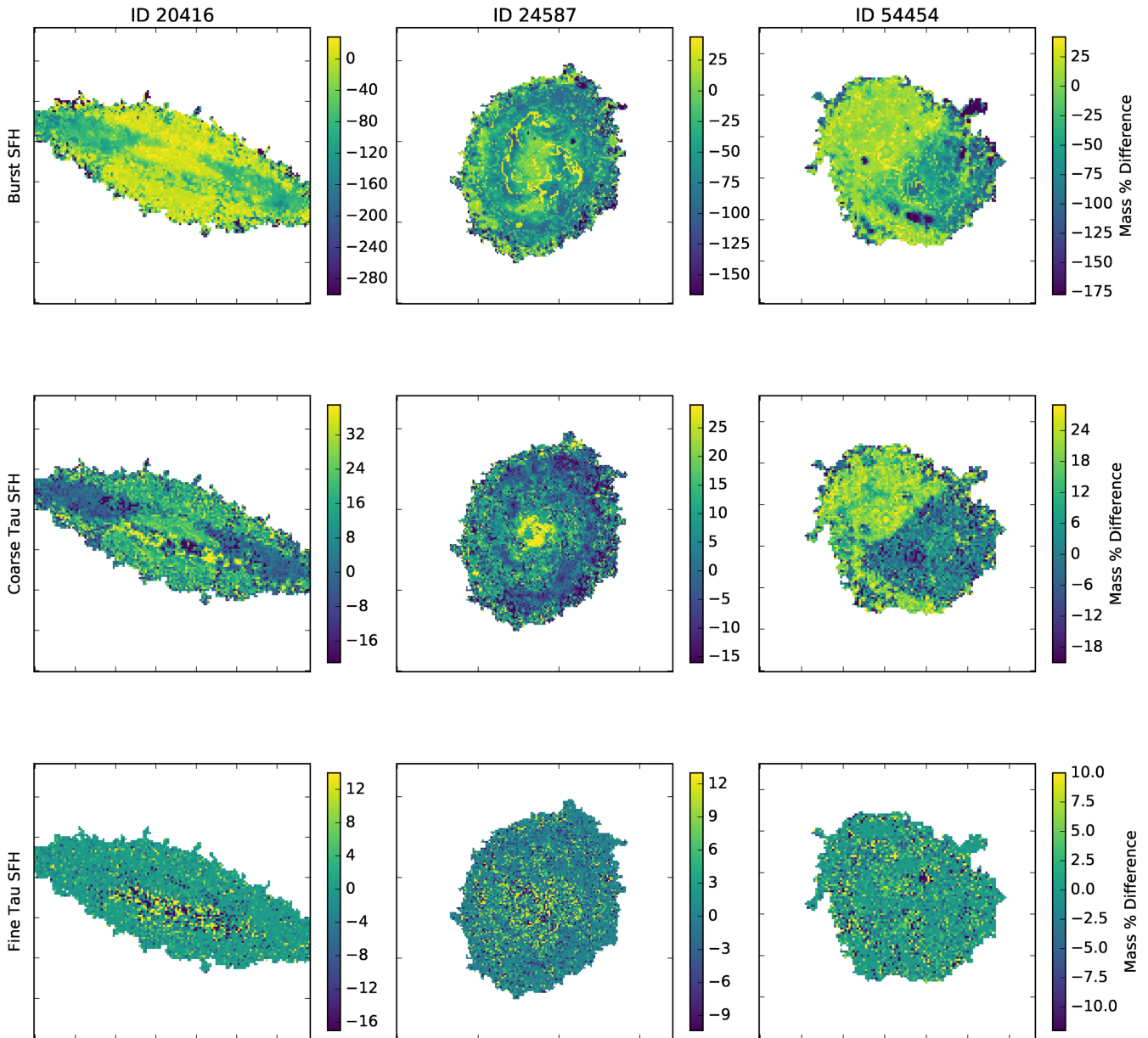


Figure A3. Relative residual mass maps created by dividing the mass map of the specified model grid (burst SFH in the top row, coarse tau SFH in the middle, and fine resolution tau in the bottom) by the ‘best’ mass map created with the doubly-fine tau SFH grid. The results for three galaxies are shown as columns, demonstrating that the fine grid has roughly the same results as the doubly-fine.

To ensure that our model grid had adequate resolution, we created an even finer set of model templates by doubling the number of values used for each parameter (tau, age, and dust) while keeping the upper and lower limits the same. We compared the mass maps made with this doubly-fine template grid to those of the other template grids by taking the ratio. These residual maps are shown in Fig. A3 for our three test galaxies. As expected, the burst (top row) and coarse model (middle row) grids show strong spatial correlations in the residuals as well as large relative differences. The fine model grid (bottom row), however, compares well to the doubly-fine grid, with the residuals looking more like random noise and relative differences typically less than approximately 10 per cent. As a further check that the quality of fits had converged, we investigated how the average χ^2 value of the best-fitting model changed with our model grid. These values are shown in Table A1. Of particular note is that

Table A1. Average χ^2 values of various model grid templates fit to pixels of three different galaxies.

Galaxy ID	χ^2			
	Burst	Coarse τ	Fine τ	Doubly-fine τ
20416	17.59	13.80	13.70	13.60
24587	23.99	9.24	8.07	8.06
54454	16.37	13.27	11.36	11.21

the fine and doubly-fine values are virtually equivalent, demonstrating that our chosen resolution is enough to represent the galaxy pixels.

This paper has been typeset from a \LaTeX file prepared by the author.