San Jose State University

SJSU ScholarWorks

Faculty Research, Scholarly, and Creative Activity

11-1-2022

Still at odds with conventional galaxy evolution: the star formation history of ultradiffuse galaxy Dragonfly 44

Kristi A. Webb
University of Waterloo

Alexa Villaume
University of Waterloo

Seppo Laine
Infrared Processing & Analysis Center

Aaron J. Romanowsky
San Jose State University, aaron.romanowsky@sjsu.edu

Michael Balogh
University of Waterloo

See next page for additional authors

Follow this and additional works at: https://scholarworks.sjsu.edu/faculty_rsca

Recommended Citation

Authors

Kristi A. Webb, Alexa Villaume, Seppo Laine, Aaron J. Romanowsky, Michael Balogh, Pieter Van Dokkum, Duncan A. Forbes, Jean Brodie, Christopher Martin, and Matt Matuszewski

Still at odds with conventional galaxy evolution: the star formation history of ultradiffuse galaxy Dragonfly 44

Kristi A. Webb [ID],[1,2]⋆ Alexa Villaume,[1,2] Seppo Laine,[3] Aaron J. Romanowsky,[4,5] Michael Balogh,[1,2] Pieter van Dokkum,[6] Duncan A. Forbes,[7] Jean Brodie,[7] Christopher Martin[8] and Matt Matuszewski[8]

[1]*Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L3G1, Canada*
[2]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*
[3]*IPAC, Mail Code 314-6, Caltech, 1200 E. California Blvd., Pasadena, CA 91125, USA*
[4]*Department of Physics and Astronomy, San José State University, One Washington Square, San Jose, CA 95192, USA*
[5]*Department of Astronomy and Astrophysics, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA*
[6]*Astronomy Department, Yale University, 52 Hillhouse Avenue, New Haven, CT 06511, USA*
[7]*Centre for Astrophysics and Supercomputing, Swinburne University, Hawthorn, VIC 3112, Australia*
[8]*Cahill Center for Astrophysics, California Institute of Technology, 1216 East California Boulevard, Mail Code 278-17, Pasadena, CA 91125, USA*

ABSTRACT

We study the star formation history (SFH) of the ultradiffuse galaxy (UDG) Dragonfly 44 (DF44) based on the simultaneous fit to near-ultraviolet to near-infrared photometry and high signal-to-noise optical spectroscopy. In fitting the observations, we adopt an advanced physical model with a flexible SFH and discuss the results in the context of the degeneracies between stellar population parameters. Through reconstructing the mass-assembly history with a prior for extended star formation (SF) (akin to methods in the literature), we find that DF44 formed 90 per cent of its stellar mass by $z \sim 0.9$ ($\sim$7.2 Gyr ago). In comparison, using a prior that prefers concentrated SF (as informed by previous studies of DF44's stellar populations) suggests that DF44 formed as early as $z \sim 8$ ($\sim$12.9 Gyr ago). Regardless of whether DF44 is old or very old, the SFHs imply early SF and rapid quenching. This result, together with DF44's large size and evidence that it is on its first infall into the Coma cluster, challenges UDG formation scenarios from simulations that treat all UDGs as contiguous with the canonical dwarf population. While our results cannot confirm any particular formation scenario, we can conclude from this that DF44 experienced a rare quenching event.

Key words: galaxies: evolution.

1 INTRODUCTION

Matching predictions to observations of how, and when, galaxies assemble serves as an important test for our greater understanding of cosmology and baryonic physics. Modern theories that suggest galaxy evolution is determined by the growth of their dark matter haloes, as well as the regulation of their gas processes (i.e. infall and star formation histories, SFHs; e.g. White & Frenk 1991; Schaye et al. 2010; Davé, Finlator & Oppenheimer 2012; Wechsler & Tinker 2018), have successfully replicated some observed relations between galaxy properties – for example, the tight connection between stellar mass and halo mass (i.e. the SMHM relation; Moster et al. 2010). A number of outstanding issues remain, however. A particularly challenging problem is explaining the increasing number of galaxies that cease forming stars (i.e. 'quench') over time (Renzini 2006; Faber et al. 2007). While simulations correctly predict scaling relations for massive galaxies [e.g. the mass–metallicity relation; MZR, and star formation (SF) main sequence], there are still fundamental discrepancies at lower stellar masses.

In the low-mass regime, observations have shown that quenched galaxies associated with massive host haloes are rare (Geha et al. 2012), such that quenching at $z < 1$ is thought to predominantly be a result of environmental effects (e.g. Boselli & Gavazzi 2006; Fillingham et al. 2018; Mao et al. 2021). Rather than remain quenched, recent studies instead suggest that isolated quiescent dwarfs may in fact oscillate between 'star forming' and 'quenched' states (e.g. Polzin et al. 2021). Yet, cosmological simulations typically overpredict the abundance of quiescent field dwarfs (e.g. Dickey et al. 2021).

The recently discovered ultradiffuse galaxies (UDGs) potentially exemplify our limited understanding of the true diversity of galaxy evolution and quenching. UDGs were initially noted for their surprisingly large sizes, given their low surface brightnesses ($R_{\rm eff} \geq 1.5$ kpc and $\mu_0(g) \geq 24$ mag arcsec^{-2}; van Dokkum et al. 2015) that, along with their red colours, distinguished them from classical Low Surface Brightness (LSB) galaxies (e.g. Dalcanton, Spergel & Summers 1997).

Several current cosmological model predictions suggest that conventional processes can explain the UDG population, thus maintaining standard dark matter halo occupancy relations (e.g. Tremmel et al. 2020). Such models typically focus on the mechanisms that increase the size of otherwise canonical dwarf galaxies to make

⋆ E-mail: kristi.webb@uwaterloo.ca

them 'ultra-diffuse' (for a summary of UDG origins, see Jiang et al. 2019a). Simulations have shown that unusual SF or galaxy evolution processes can 'puff up' canonical dwarfs (e.g. high-spin scenarios, Amorisco & Loeb 2016; Rong et al. 2017; energetic SF feedback, Di Cintio et al. 2017; Chan et al. 2018; Jackson et al. 2021) or dynamically redistribute their stellar populations (e.g. tidal heating and/or stripping; Jiang et al. 2019a; Liao et al. 2019; Carleton et al. 2019; Sales et al. 2020). Alternatively, UDGs may represent the tail of galaxy evolution processes, such that only minor differences in their evolution (e.g. when they infall or have major mergers) distinguish their final properties from normal dwarfs (e.g. Tremmel et al. 2020; Wright et al. 2021).

Despite these differences, nearly all models rely on environmental processes to explain the lack of SF in the subset of UDGs that are quiescent (e.g. via ram pressure stripping; Yozin & Bekki 2015; Rong et al. 2017; Chan et al. 2018; Tremmel et al. 2020). Accordingly, all of the scenarios follow a dichotomy related to when UDGs infall into a cluster environment: whether the proto-UDGs surpassed the size-threshold prior-to or post infall, is tied to whether they infall 'late' or 'early'. While UDGs are found both in the field and clusters, those that are quiescent are usually located in clusters (the few exceptions may be on backsplash orbits; e.g. Papastergis, Adams & Romanowsky 2017; Benavides et al. 2021). Explaining the origin of UDGs and the diversity of their properties in the context of their environments remains a key question in understanding galaxy formation and evolution.

Testing the predicted UDG properties (e.g. kinematics, Amorisco & Loeb 2016; stellar populations, Rong et al. 2017; Ferré-Mateu et al. 2018; globular cluster (GC) properties, Carleton et al. 2021; infall versus quenching times, Gannon et al. 2022) from these scenarios against the observed properties, however, has revealed a number of discrepancies. And while some UDGs are found with very large sizes ($R_{eff} > 4.5$ kpc), these exotic objects are beyond the predictions of most models (Di Cintio et al. 2017; Carleton et al. 2019). Along the same lines, models that accurately predict the distribution of UDG sizes fail to reproduce the distribution of sizes among normal dwarfs (e.g. Rong et al. 2017; Jiang et al. 2019a; Tremmel et al. 2020).

On the other hand, van Dokkum et al. (2015) proposed that some UDGs originate similar to today's massive galaxies (and have sizes reflecting their massive haloes), but lost their gas early in their histories. As a result of their early quenching, these 'failed' galaxies did not build up the stellar mass expected for their haloes. This scenario deviates from the expected galaxy–halo connection, in that either these failed galaxies do not follow the SMHM relation or at least have a larger scatter than the standard relation.

A particularly interesting UDG is Dragonfly 44 (DF44), which is the largest galaxy in the original van Dokkum et al. (2015) sample, with $R_{eff} = 4.7 \pm 0.2$ kpc (van Dokkum et al. 2017). High signal-to-noise ratio (S/N) spectroscopy has revealed an extremely old and metal-poor stellar population ($\sim$2.3σ below the canonical dwarf MZR; Villaume et al. 2022), implying that DF44 quenched very early and over a short time-scale. Moreover, while DF44 appears to have very low rotation (van Dokkum et al. 2019) characteristic of dwarf spheroidal galaxies, the stellar population gradients are 'inverted' compared to the gradients typical of dwarf spheroidals (Villaume et al. 2022). Regardless of whether DF44 has an overmassive halo or not (van Dokkum et al. 2017; Wasserman et al. 2019; Bogdán 2020; Lee, Hodges-Kluck & Gallo 2020; Saifollahi et al. 2021), this UDG is inconsistent with the majority of UDG formation models.

Late-quenching (after infall into a dense environment) scenarios can be ruled out for DF44 given its old age (e.g. Rong et al. 2017;

Chan et al. 2018; Liao et al. 2019; Jiang et al. 2019a; Jackson et al. 2021). Moreover, DF44's low rotation conflicts with high-spin scenarios (e.g. Rong et al. 2017; although the rotation could increase at larger radii, Grishin et al. 2021). Yet, given the uncertainty in establishing the cluster infall time for an individual galaxy, we cannot preclude early-infall scenarios (e.g. Yozin & Bekki 2015; Carleton et al. 2019, 2021; Liao et al. 2019; Tremmel et al. 2020). While some evidence (e.g. Alabi et al. 2018; van Dokkum et al. 2019) suggests that DF44 is on its first infall into Coma, this is difficult to prove.

There is more to be learned, however, as UDG formation scenarios can be tested via their inferred SFHs. The time-scales of SF reveal important epochs (e.g. mergers, infall, and/or quenching), which can be compared against observations. A number of studies have investigated the ages and mass assembly histories of UDGs, relying either only on broad-band colours, or low to moderate S/N spectroscopy (e.g. Kadowaki, Zaritsky & Donnerstein 2017; Ferré-Mateu et al. 2018; Gu et al. 2018; Pandya et al. 2018; Ruiz-Lara et al. 2018; Martín-Navarro et al. 2019; Buzzo et al. 2022). While these studies provide important first steps, comparisons with predictions are not necessarily straightforward. This is primarily because constraining the detailed shape of a galaxy's SFH is a complex problem.

Several galaxy properties can conspire to alter the spectral energy distribution (SED) in similar ways (e.g. stellar age, metallicity, and dust), which are particularly difficult to disentangle with low spectral resolution data (e.g. with photometry alone; Bell & de Jong 2001). Recovering the SFHs for old stellar populations is particularly difficult – the integrated spectrum evolves non-linearly with age (Serra & Trager 2007) such that old populations appear relatively similar (for a complete discussion, see the review by Conroy 2013). Moreover, a late burst of SF can 'outshine' a (dominant) older population (e.g. Papovich, Dickinson & Ferguson 2001; Allanson et al. 2009). While broad wavelength coverage is needed to precisely determine the dust absorption (and emission, with mid-infrared coverage), high-resolution data of select spectral features are needed to precisely constrain the stellar metallicity and age. Both observations are necessary to break the degeneracy between these parameters (e.g. Vazdekis 1999; Trager et al. 2000). Using spectra that span a relatively wide wavelength range, full-spectrum fitting has proven to be effective in this respect (e.g. MacArthur, González & Courteau 2009; Sánchez-Blázquez et al. 2011). However, this technique requires a well-calibrated spectral continuum. Simultaneously fitting photometry and spectra can bypass this issue, as the photometry provides a means to fit the continuum[1] and increases the wavelength coverage.

In fitting the data, it is necessary to impose 'prior knowledge',[2] such as the flexibility of the SFH. The choice of a prior for the shape of the SFH can significantly impact age estimates, particularly for older stellar populations, and for low-resolution and/or low-S/N data (as shown in, e.g. Maraston 2005; Leja et al. 2017, 2019; Han & Han 2018; Carnall et al. 2019). In order to draw connections between the predicted and observed properties of UDGs it is necessary to give due attention to the choice of a prior. While it is advantageous to use flexible models together with physically motivated priors, a 'good prior' is not necessarily known a priori. Therefore, results should

---

[1] In practice it is generally easier to calibrate photometry to standard filters than to calibrate a spectrum.

[2] 'Prior' here is used in the Bayesian sense, where the probability of a model given the data (i.e. the 'posterior') is proportional to both the likelihood of the data (given the model) and the prior knowledge about the model.
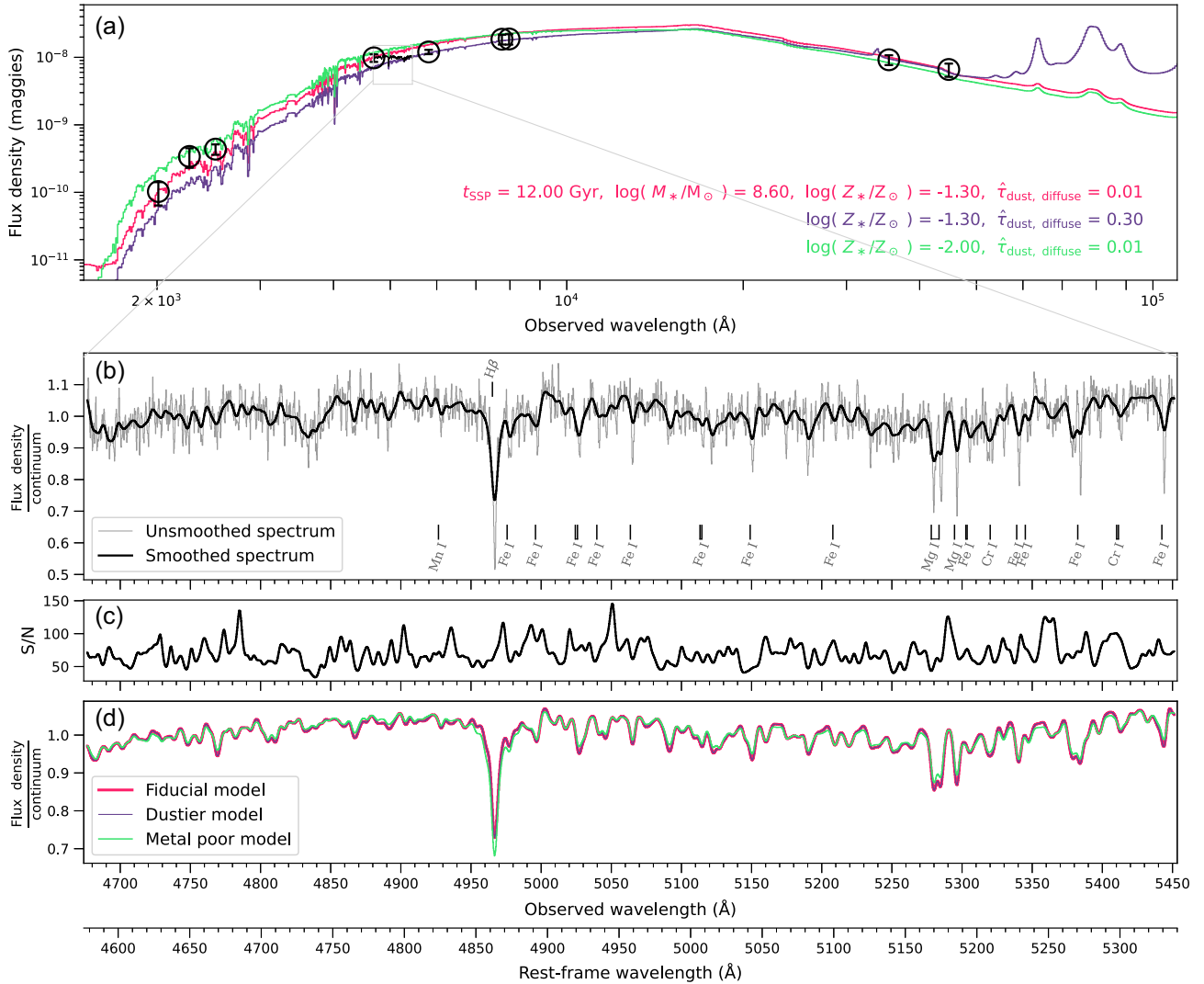
Figure 1. Observations of DF44 and three simple stellar populations (SSPs) shown to briefly illustrate the effect of the dust attenuation and metallicity on the shape of the SED. Panel (a): the photometry with error bars (black markers) and model SEDs (coloured lines) with listed parameter values. Panel (b): the original (continuum removed) KCWI spectrum (thin grey line) and the smoothed spectrum (thick line) for fitting purposes. Significant spectral features are labelled for reference. Panel (c): the S/N of the smoothed spectrum. Panel (d): the (continuum removed) model spectra, shown with the same resolution as the smoothed KCWI spectrum. While changing the dust attenuation affects the shape of the overall continuum, changing the metallicity affects both the continuum shape and the absorption features.

be discussed in the context of the prior used (which may not be as 'uninformative' as intended; e.g. Leja et al. 2019).

In this work, we simultaneously fit near-ultraviolet (NUV) to near-infrared (NIR) photometry (nine bands) with high S/N ($\sim$96 Å^{-1}) rest-frame optical spectroscopy (from KCWI, the Keck Cosmic Web Imager). The same data set was used in van Dokkum et al. (2019) and Villaume et al. (2022) to study the stellar kinematics and populations of DF44. We adopt flexible SFHs in our fiducial model that do not assume a certain shape with time. Moreover, we compare the results between SFH priors of different degrees of 'smoothness' in order to identify which results are fully constrained by the observations. We address the unique stellar population properties of this UDG, and its epoch of formation and quenching, in order to test models of UDG formation.

The data are described in Section 2, and Section 3 details how we fit the data with an advanced physical model. In Section 4, we discuss the results, and put the results in the context of the literature. What our

results imply about the origins of DF44 in the context of theoretical models is discussed in Section 5. A summary of the key results is provided in Section 6. The SFHs of DF44 determined by this work are listed in full in Appendix A. We provide additional details on the above discussion in the Appendix, touching on systematic biases in measuring SFHs in Appendix B, and degeneracies between dust extinction and the flux from old stellar populations in Appendix C.

The magnitudes reported follow the AB magnitude system. We use a Chabrier (2003) initial mass function (IMF), and adopt a flat Λ cold dark matter (ΛCDM) cosmology with $\Omega_m = 0.3$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

2 DATA

Our data for DF44 include both rest-frame optical spectroscopy and NUV to NIR photometry, shown in Fig. 1, and described in more

detail below. We assume the spectroscopic redshift measured by van Dokkum et al. (2017): $z = 0.021\,32 \pm 0.000\,02$.

2.1 Spectroscopy

The spectroscopy is described in detail in van Dokkum et al. (2019) and analysed further in Villaume et al. (2022); we summarize the relevant details here.

Of particular note is the sky subtraction, as the sky is much brighter than the UDG. Sky exposures were obtained 1.5 arcmin away from DF44 intermittently between DF44 observations. The wavelength-dependent time variation in the sky spectrum was obtained from the spatially collapsed individual sky spectra, as parametrized by principal component analysis (PCA). The sky in each science cube was determined from a linear combination of templates, where the best-fitting sky spectrum for the given exposure was subtracted from each spatial pixel. Additional details are provided in van Dokkum et al. (2019).

KCWI integral field spectroscopy was obtained for DF44 and spectra were extracted in nine elliptical apertures after masking the ten brightest point sources. The apertures were sized $9 \times 6\,\mathrm{arcsec}^2$, to match the UV photometry; see the following section. The integrated spectrum was determined through bootstrapping the individual spectra, where we used the 50th percentile of the bootstrapped flux distribution and the average of the 16th and 84th percentile as the uncertainty. With 17 h of exposure on-target, the integrated spectrum reaches an S/N $\sim 96\,\text{Å}^{-1}$ (see the third panel in Fig. 1).

The KCWI Medium slicer with BM grating was used, yielding a spectral resolution of $R \sim 4000$. After masking and interpolating over regions badly affected by sky transmission, the spectrum was smoothed to a resolution of 110 km s^{-1}, for the purpose of later comparing with templates at this resolution. The final spectrum is shown in Fig. 1 (the unsmoothed spectrum shown with grey lines), covering 4578–5337 Å rest-frame, with notable absorption features labelled. Also shown is the S/N of the spectrum as a function of wavelength.

Given the challenge of precisely flux-calibrating the spectrum (e.g. due to residuals from the spectral extraction), we instead rely on the calibration of the photometry to provide constraints on the SED continuum when fitting the galaxy properties and SFH (see Section 3). For this reason, we do not flux calibrate the spectrum, and the continuum shape therefore reflects primarily the instrument response function and not the galaxy SED. We then effectively flatten the continuum by dividing through by a polynomial fit. In the fitting routine, we therefore need to marginalize over the shape of the spectral continuum in comparing the models to the observations (see Section 3.3).

Lastly, we chose to mask the spectrum between 4700 and 4750 Å rest-frame where there is a broad dip in the spectrum that does not appear in the models. We note that the blue end of the spectrum ($\lesssim 4800$ Å) was not fitted by either van Dokkum et al. (2019) or Villaume et al. (2022). Our results are not impacted by masking this region of the spectrum, although the χ^2 values are slightly higher without masking.

2.2 Photometry

Photometry in all the broad-band images was performed by measuring fluxes within a $9 \times 6\,\mathrm{arcsec}^2$ elliptical aperture, with a position angle of 65°, to be consistent with the UV photometry reported by Lee et al. (2020). As this is significantly larger than the image

resolution in all filters, no point spread function homogenization was applied, though appropriate aperture corrections are made to the *Spitzer* and *GALEX* images to account for light lost outside the aperture due to the point spread function. Details on the reduction and analysis of each image is described in more detail, below. The photometric measurements in each broad-band filter were corrected for foreground extinction in the Milky Way in the direction of the Coma Cluster using the website http://argonaut.skymaps.info/usage and table 6 of Schlafly & Finkbeiner (2011) with $R_V = 3.1$.

2.2.1 Spitzer-IRAC NIR imaging

Spitzer-IRAC (Fazio et al. 2004; Werner et al. 2004) observations of DF44 were taken on 2017 May 12 starting at 07:19 (UT). Both 3.6 and 4.5 μm (channels 1 and 2, respectively) observations were taken. 50 medium-scale (median dither separation 53 pixels) cycling dither pattern 100-s frames were taken in each channel. The total exposure time was $93.6 \times 50 = 4680$ s in channel 1 and $96.8 \times 50 = 4840$ s in channel 2.

We removed the 'first-frame correction' (to address imperfect bias subtraction; see Section 5.1.10 of the IRAC Instrument Handbook). The rectification of each individual data frame for history effects in the IRAC arrays was performed in two steps that are explained in detail in Pandya et al. (2018). In short, we first performed a per pixel correction that was based on IRAC idling time characteristics in the IRAC skydarks, matched to those that took place before our observations. The typical magnitude of the per pixel correction was about 4 kJy sr^{-1} in channel 1 and 1 kJy sr^{-1} in channel 2. The typical corrections are much smaller than the read noise error and we do not add any systematic magnitude uncertainties due to these first-frame corrections.

In the second step, a mean background is calculated for each frame, and a function fitted to these means is subtracted. The typical function consisted of a constant term plus terms that are declining exponentially with time. The uncertainties in these first-frame effect corrections are negligible compared to other sources of systematic error. We also formed a median image after doing a 3σ clipping from all the frames on the source in each channel and subtracted that median image separately in each frame. Such a median image will subtract the residual images that have been formed on the detector from previous observations. We determined that the uncertainty in the final magnitudes added by this step is less than 0.01 mag.

The DF44 frames include a point source on top of the faint galaxy. We used *Spitzer* Science Center provided software MOPEX, specifically the APEX and APEX-QA modules, to subtract this point source using point response function (PRF) fitting. The estimated uncertainty due to this step is about 0.5 mJy in both channels.

We used the contributed *Spitzer*/IRAC software IMCLEAN (Hora 2021) to remove leftover column pulldown artefacts from the CBCD frames. We then used the *Spitzer* custom software package MOPEX to create mosaics of the 50 frames in each channel, using the default parameters and the north up, east left orientation. Before mosaicking, we ran the overlap correction module to adjust for background offsets among the CBCD frames (one number per frame). We used only the multiframe outlier rejection scheme in MOPEX to reject outlier pixels in the input frames.

Next we manually created masks of other sources (including point-like sources on the galaxy) in both channels with the custom software GIPSY (van der Hulst et al. 1992). We then measured the 'sky background' in five empty areas of sky close to DF44 in channels 1 and 2, and from the results we estimated an average sky background

Table 1. DF44 photometry.

Filter	m_0 (AB)	λ_{eff} (Å)
UVOT UV1	23.40 ± 0.19	2516.7
UVOT UV2	24.97 ± 0.41	2010.4
GALEX NUV	23.67 ± 0.35	2271.1
GMOS g_G0301	20.02 ± 0.14	4687.6
GMOS i_G0302	19.33 ± 0.18	7751.6
WFC3 *F606W*	19.80 ± 0.08	5813.0
WFC3 *F814W*	19.32 ± 0.19	7972.9
IRAC1	20.09 ± 0.18	35 439.4
IRAC2	20.45 ± 0.24	44 840.9

(0.004 08 and 0.004 15 MJy sr^{-1} in channels 1 and 2, respectively) to be subtracted at the position of DF44, applying the mask and using ASTROPY PYTHON library commands in a $9 \times 6\,\mathrm{arcsec}^2$ (PA $+\,65°$) aperture centred on the coordinates given by van Dokkum et al. (2015): RA $= 13^\mathrm{h}00^\mathrm{m}58\overset{\mathrm{s}}{.}0$, Dec. $= 26°58'35''$. We corrected the results with the appropriate aperture corrections from the IRAC Instrument Handbook.

The uncertainty in aperture photometry was estimated by performing aperture photometry on several positions in empty sky and taking the rms scatter in these measurements. This gave 0.05 and 0.10 mag in IRAC channels 1 and 2, respectively. We estimated the uncertainty due to masking by replacing the pixel values under the masks by the average pixel values within the unmasked aperture, and performed the photometry again, and took the difference between this measurement and the measurement using the masks as the uncertainty. The channel 1 masking uncertainty is thus 0.14 mag, and 0.18 mag for channel 2.

The sky background subtraction uncertainty is estimated by taking the maximum difference in the sky background measurements in three areas of empty sky around DF44 in the images and adding this difference to all the pixels within the photometry aperture and summing them up. This method gives 0.01 mag and 0.11 mag as the sky uncertainty in channels 1 and 2, respectively.

The calibration uncertainty was estimated to be 2 per cent in IRAC channels 1 and 2, amounting to 0.02 mag in systematic uncertainty. There is an additional uncertainty of 9 per cent in channel 1 and 2 per cent fractional flux in channel 2 due to the uncertainty in integrated aperture flux correction factor (limiting case is infinite aperture). These convert into 0.09 and 0.02 mag in channels 1 and 2. In addition, there is the point-source subtracting uncertainty of 0.01 mag.

We list the final AB magnitudes for channels 1 and 2 and their respective uncertainties from the quadrature sum of the magnitude uncertainties in Table 1.

2.2.2 Gemini GMOS g- and i-band imaging

DF44 was observed on 2017 May 12 with the Gemini Multi-Object Spectrometer (GMOS) for a total of 3000 s in both the *g* and *i* bands. The observations have been described by van Dokkum et al. (2016). We flux-calibrated the images with SDSS, accounting for a $g - i$ colour term and using four SDSS catalogued stars in our images. The data were obtained in photometric conditions, and we adopt an absolute calibration magnitude uncertainty to be 3 per cent, amounting to 0.03 mag in the *g* and *i* bands, based on https://www.gemini.edu/instrumentation/gmos/calibrations. The sky background uncertainty was calculated as above for the IRAC channels, and amounted to 0.03 mag in the *g* band and 0.09 mag in the *i* band.

Aperture photometry was performed using the coordinates from van Dokkum et al. (2015) and the ASTROPY PYTHON library commands.

We list the final AB magnitudes for the *g* and *i* bands and the respective uncertainties in Table 1.

2.2.3 HST/WFC3/UVIS F606W *and* F814W *imaging*

Additional visual images of DF44 were taken on 2017 April 23 with the *Hubble Space Telescope* using the WFC3 camera and its UVIS detector and broad-band filters *F606W* and *F814W*. van Dokkum et al. (2017) reported 5σ AB depths of *F606W* = 28.4 and *F814W* = 26.8 for DF44. A total of 2430 and 2420 s were spent on the source in *F606W* and *F814W* filters. In both filters, we calculated the sky mode in five different 'empty' regions of the sky and took an average and subtracted those values from the images. We also manually masked out point sources in the images. We used the image headers to calculate the conversion from electrons/s to AB magnitudes and performed elliptical aperture photometry within the same apertures as mentioned above for IRAC.

The uncertainties were estimated in the following way: We estimate a photometric calibration offset uncertainty of 0.03 mag, and the uncertainty due to background subtraction (estimated as above) is 0.05 mag in *F606W* and 0.13 mag in *F814W*. The uncertainty due to masked point sources within the aperture is estimated to be 0.03 and 0.01 mag in *F606W* and *F814W*. The uncertainty in performing aperture photometry was estimated as above and results in an additional 0.05 and 0.14 mag in *F606W* and *F814W*.

We list the final AB magnitudes for *F606W* and *F814W* and the respective uncertainties in Table 1.

2.2.4 Ultraviolet

The UV data reduction and analysis was presented in Lee et al. (2020). This consists of two filters observed with *Swift* UVOT (UV1 at 2600 Å and UV2 at 1928 Å), and *GALEX* NUV images. The UVOT data include a correction for red leakage and scattered light, where the correction (14 per cent) was comparable to the flux uncertainty. Again, we list the final results in Table 1.

3 STELLAR POPULATION MODELLING AND FITTING

In this section we describe how we fit the DF44 observations using the fully Bayesian inference code PROSPECTOR[3] (v1.0; Leja et al. 2017; Johnson et al. 2019, 2021b). The photometry and spectroscopy are fitted simultaneously, incorporating the information on the stellar properties and SFH from both data sets. In Section 3.1, we describe the advanced physical model, which includes a non-parametric SFH and a flexible dust attenuation law. We additionally include a white noise and spectral outlier model described in Section 3.2, and a spectrophotometric calibration model that marginalizes out the shape of the spectral continuum, in Section 3.3. A summary of the parameters and priors of our physical model is shown in Table 2. Section 3.4 briefly describes the sampling method.

3.1 The physical model

The physical models are based on the stellar population synthesis (SPS) models from the Flexible Stellar Population Synthesis library

---

[3] https://github.com/bd-j/prospector.

Table 2. SFH parameters and priors.

Note	Parameter	Description	Prior
	SFH		
1	f_n	sSFR fraction	Dirichlet (α_D)
2	z_{obs}	Redshift	Uniform (min $= z_{spec} - 0.01$, max $= z_{spec} + 0.01$)
3	$\log(M_*/M_\odot)$	Total stellar mass formed	Uniform (min $= 8$, max $= 12$)
4	$\log(Z/Z_\odot)$	Stellar metallicity	Uniform (min $= -2$, max $= 0.19$)
	Dust attenuation		
5	$\hat{\tau}_{dust,\,diffuse}$	Diffuse dust optical depth (equation 2)	Uniform (min $= 0$, max $= 1.5$)
	$\hat{\tau}_{young}/\hat{\tau}_{dust,\,diffuse}$	Ratio of diffuse to birth-cloud dust optical depth (equation 1)	Clipped Normal ($\mu = 1$, $\sigma = 0.3$, min $= 0$, max $= 1.5$)
	n_{dust}	Diffuse dust attenuation index	Uniform (min $= -2$, max $= 0.5$)
	Dust emission		
6	Q_{PAH}	Per cent mass fraction of PAHs in dust	Uniform (min $= 0.5$, max $= 7$)
	$U_{min,dust}$	Minimum starlight intensity to which the dust mass is exposed	Uniform (min $= 0.1$, max $= 25$)
	γ_{dust}	Mass fraction of dust in high radiation intensity	LogUniform (min $= 0.001$, max $= 0.15$)
	Noise model		
7	spec_jitter	Multiplicative spectral noise inflation term	Uniform (min $= 1$, max $= 3$)
8	$f_{outlier,\,spec}$	Fraction of spectral pixels considered outliers	Uniform (min $= 10^{-5}$, max $= 0.5$)
	Spectrophotometric calibration		
9	c_n	Chebyshev polynomial coefficients, $n = 4$	

Notes. (1) Fraction of SFR in a given time bin, where the SFH is a piece-wise constant function with N parameters ($N - 1$ free parameters). The prior is a Dirichlet function, controlled by the parameter α_D; see Section 3.1.1. (2) Redshift, with a tight prior about the measured spectroscopic redshift, z_{spec}. (3) The total stellar mass is the integral of the SFH, which includes the mass lost to outflows. To convert into stellar mass *remaining* at the time of observation, we regenerate the spectral templates and subtract the mass lost as calculated by FSPS. (4) The total stellar metallicity where scaled-Solar α-abundance is assumed. (5) Parameters for the two-component Charlot & Fall (2000) dust absorption model, with an adjustable attenuation curve slope from Noll et al. (2009) with a UV bump based on Kriek & Conroy (2013). (6) Parameters for the Draine et al. (2007) dust emission model. (7) The uncertainty on the spectra can be increased by a given factor, with a likelihood penalty for factors giving reduced $\chi^2 < 1$. (8) An outlier pixel model can increase the errors for individual pixels by a factor of 50, to accommodate for poor matches between the data and spectral templates. (9) A fourth-degree Chebyshev polynomial is fit (via optimization) to the residual of the normalized ratio between the observed spectrum and the proposed model spectrum and multiplied out prior to each likelihood calculation. This effectively accounts for the lack of flux-calibration in the spectrum.

(FSPS; Conroy, Gunn & White 2009; Conroy & Gunn 2010) with MESA Isochrones and Stellar Tracks (MIST; Choi et al. 2016; Dotter 2016, based on the MESA stellar evolution code; Paxton et al. 2011, 2013, 2015, 2018), and MILES[4] spectral templates (Sánchez-Blázquez et al. 2006).

The dust is modelled with the two-component dust attenuation model from Charlot & Fall (2000), which separates the dust components between those associated with the birth-cloud, and a uniform dust screen. While we expect DF44 to have an old stellar population with very little dust content, we prefer to include a flexible dust model and marginalize over the parameters rather than assume a simplistic model. This avoids the assumption that dust attenuation in DF44 is the same as dust attenuation in the local Universe. The birth-cloud dust acts to only attenuate stellar emission for stars younger than 10 Myr,

$$\tau_{dust,\,birth}(\lambda) = \hat{\tau}_{dust,\,birth}\left(\frac{\lambda}{5500\,\text{Å}}\right)^{-1}, \quad (1)$$

while the diffuse-dust acts as a uniform screen with a variable attenuation curve (Noll et al. 2009),

$$\tau_{dust,\,diffuse}(\lambda) = \frac{\hat{\tau}_{dust,\,diffuse}}{4.05}\left(k'(\lambda) + D(\lambda)\right)\left(\frac{\lambda}{5500\,\text{Å}}\right)^n, \quad (2)$$

where n is the diffuse dust attenuation index, $k'(\lambda)$ is the attenuation curve from Calzetti et al. (2000), and $D(\lambda)$ describes the UV bump based on Kriek & Conroy (2013). The diffuse dust is given a uniform prior (min $= 0$, max $= 1.5$). We note that the diffuse dust optical

depth is related to the dust extinction via $A_\lambda = 2.5 \log_{10}(e)\,\tau_\lambda$, where τ_λ is the sum of the diffuse and birth dust components.

We use a joint prior for the ratio of diffuse to birth-cloud dust, rather than a direct prior on birth-cloud dust, to avoid degeneracies between the two parameters. The prior for $\hat{\tau}_{young}/\hat{\tau}_{dust,\,diffuse}$ is a clipped normal with $\mu = 1$, $\sigma = 0.3$, min $= 0$, and max $= 1.5$, which broadly follows results from the literature for massive galaxies while allowing some variation. Lastly, the prior on the attenuation index is uniform (min $= -2$, max $= 0.5$).

Dust emission is calculated assuming energy conservation, i.e. all the energy attenuated by dust is re-emitted at infrared wavelengths (da Cunha, Charlot & Elbaz 2008). As our photometry is limited to $<4.4\,\mu$m (rest-frame) there is no significant information in the SED constraining dust emission. We chose to include the full dust model and marginalize over the unconstrained parameters, rather than a more simplistic model, in order to avoid biasing the result.

The stellar metallicity is a free parameter; however we assume a constant metallicity for all the stars and for the entire history of the galaxy. This single metallicity has a uniform prior in $\log(Z_*/Z_\odot)$ (min $= -2$, max $= 0.19$), where $Z_\odot = 0.0142$ (Asplund et al. 2009). In addition, we assume scaled-Solar abundances, which is a current limitation of the FSPS models. Lastly, a Chabrier (2003) IMF is used.

3.1.1 Non-parametric SFH

To characterize the SFH, we use a non-parametric[5] model of the form of a piece-wise constant function with $N = 12$ time bins. The

---

[4]http://miles.iac.es/.

[5]*Non-parametric* here means that the SFH has no specified functional form.

benefits of such a flexible SFH (relative to parametric functions, e.g. declining exponential or lognormal) have been well characterized by Leja et al. (2019) and Lower et al. (2020), among others. The time bins are defined in lookback time, spaced so that the first seven bins correspond to 0–30 Myr, 30–100 Myr, 100–500 Myr, 500 Myr–1 Gyr, 1.0–2.0 Gyr, 2.0–3.0 Gyr, and 3.0–4.0 Gyr. There are four bins spaced logarithmically between 4 and $0.95 \times t_{\rm univ}$ (4.0–5.4, 5.4–7.2, 7.2–9.6, and 9.6–12.6 Gyr), and the last bin covers $0.95 \times t_{\rm univ}$–$t_{\rm univ}$, where $t_{\rm univ}$ is the age of the Universe at the time of observation. Defining the time bins this way reflects the non-linear evolution in the SEDs: The narrower time bins at recent lookback times allow a sufficient precision in capturing recent SF, while the wider bins at later lookback times reflect the modest evolution of older stellar populations. The last time bin is included to permit a maximally old population.

Fitting SEDs to recover SFHs is an ill-defined problem, and prone to overfitting (e.g. Moultaka & Pelat 2000; Moultaka et al. 2004; Ocvirk et al. 2006a). In order to recover a physically plausible SFH, it is common to invoke 'regularisation'. There a number of ways that this can be done, which differ in technical detail. One approach is to impose Gaussian-like priors on the SFH and/or the age–metallicity relation (e.g. as in the commonly used code STECKMAP; Ocvirk et al. 2006a,b), and another is to penalize sharp transitions in the SFH (e.g. the continuity prior; Leja et al. 2019). In this work, we use a third method, which is to control the degree of concentration of fractional specific SFR (sSFR) between the time bins of the non-parametric function. While these approaches differ in detail, they all attempt to avoid non-physical solutions by imposing constraints on the variability of the SFH over time.

We adopt a Dirichlet prior that includes a concentration parameter, $\alpha_{\rm D}$, that controls the preference to distribute the fractional sSFR in one bin ($\alpha_{\rm D} < 1$) or evenly between all bins ($\alpha_{\rm D} \geq 1$), respectively. A detailed description of this prior is provided in Leja et al. (2017). Without direct physical motivation to inform a choice of $\alpha_{\rm D}$, we consider both $\alpha_{\rm D} = 1$ and 0.2 as valid options, labelling them as 'extended' and 'concentrated' versions of the SFH prior. In comparing the results produced from these two choices of SFH prior, we explore the dependence of the results on the degree of regularization.

Fig. 2 shows random draws (thin lines) for priors with $\alpha_{\rm D} = 1$ (extended) and 0.2 (concentrated), with the time bins as defined above. The median and 68 per cent credible regions (CRs) of the priors are shown with a thick line and shaded regions, respectively. The corresponding implicit prior on the mass-weighted age is shown in the bottom panel for reference. The mass-weighted stellar age ($t_{\rm age}$, sometimes referred to as the mean stellar age, broadly describes the average formation time of stars in a given galaxy in units of the lookback time) is calculated from the SFH:

$$t_{\rm age} = \frac{\int_{t_{\rm obs}}^{0} t \; {\rm SFR}(t) \, {\rm d}t}{\int_{t_{\rm obs}}^{0} {\rm SFR}(t) \, {\rm d}t}, \qquad (3)$$

where $t_{\rm obs}$ is the age of the Universe at the time of observation. The implicit age prior for an extended SFH is centred at half the age of the Universe with a 99.9 per cent CR between 3.08 and 9.98 Gyr, and thus is a strong prior against both very old and very young ages. In comparison, the concentrated SFH also peaks around half the age of the Universe (although offset given the varying widths of the time bins), but the prior is not as tight (99.9 per cent CR between 0.83 and 12.17 Gyr) such that old ages are less strongly disfavoured.
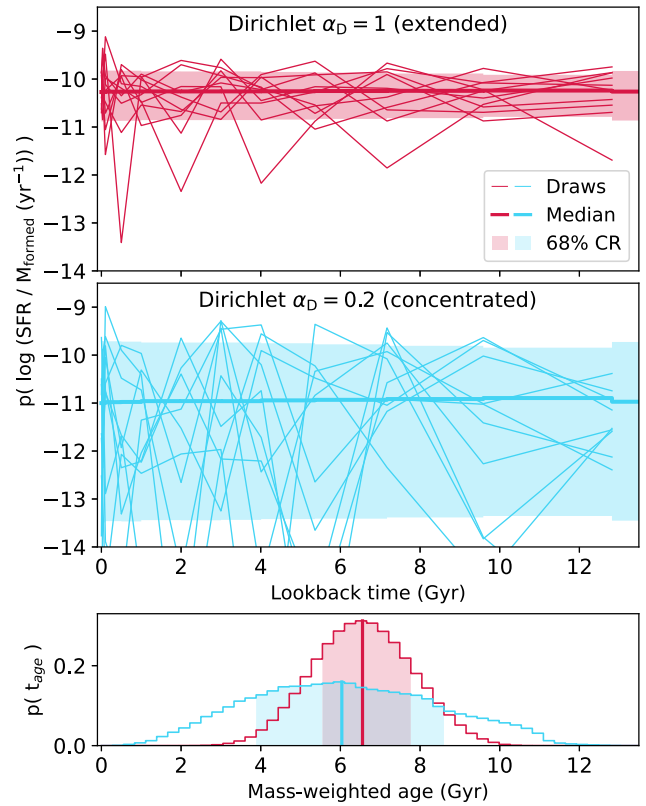

Figure 2. Comparison of SFH priors between sSFR (SFR per unit mass; first and second panels) and corresponding mass-weighted age (third panel). The $\alpha_{\rm D} = 1$ SFH prior prefers solutions where the SF is equally weighted between the time bins, hence we label it 'extended'. In comparison, the $\alpha_{\rm D} = 0.2$ SFH prior prefers solutions where the SF is unequally distributed between time bins, which we label as 'concentrated'. The sSFR is shown as a function of lookback time for ten random draws (thin lines) from the Dirichlet SFH priors with $\alpha_{\rm D} = 1$ (extended; red) and 0.2 (concentrated; blue). The medians and 68 per cent CRs of the prior are indicated with thick lines and shaded regions, respectively. The implicit mass-weighted age priors are shown in the lower panel, with vertical lines indicating the median, and shaded regions indicating the 68 per cent CRs.

3.2 Noise and outlier models

A noise model is used to account for possible under- or overestimates of the spectral uncertainties, where the noise is uniformly inflated (or deflated). This effectively modifies the spectral uncertainty by a multiplicative factor, but is counterbalanced by a penalty in the likelihood calculation for larger uncertainties. This down-weights spectra where the uncertainties are otherwise low, but there is a mismatch between the spectrum and the models. The minimum uncertainty in the photometry is 7 per cent, and as we expect this to be large enough to account for deviations with the template SEDs, we do not include a noise model for the photometry in our model.

A mixture model is used to identify and mask pixels in the spectra that have large deviations from the model. The purpose here is to avoid being overly sensitive to outlier pixels in the spectrum. This is again relevant where the S/N is large and significant residuals can result from poor matches to the models where the model itself is inaccurate (e.g. due to differences in α-enhancement). PROSPECTOR uses the mixture model approach described in Hogg, Bovy & Lang (2010).

The spectral outlier model finds that less than 1 per cent of the pixels are inconsistent with the model templates beyond the specified

uncertainty. Note that the spectral white noise model prefers to inflate the uncertainties by ∼1–3 per cent, which is not unexpected, given that the S/N of the spectrum is high, 40–140 (median 96) and that the models are not flexible enough to precisely match the metallicity- and α-abundance sensitive spectral features (i.e. the Mg triplet).

3.3 Spectrophotometric calibration

We rely on the calibration of the photometry to constrain the shape of the SED continuum. The DF44 spectrum is not flux-calibrated such that neither the normalization nor the shape of the spectral continuum provides information about the stellar properties. In fact, the spectrum was flattened prior to fitting (see Section 2.1). For this reason, we ignore the shape of the spectrum when computing the likelihood of the SED model (relative to the spectrum). We do this by following the routine provided through PROSPECTOR that fits (via optimization) a polynomial to the residual between the spectrum and the model, which is then multiplied to the model. We use an $n = (\lambda_{max} - \lambda_{min})/100$ Å ∼eighth-order Chebyshev polynomial, which is flexible enough to remove the broad continuum shape without overfitting absorption features (e.g. Conroy et al. 2018). We test our results using several different orders of the polynomial, and find that we are generally insensitive to the choice of n as long as $n > 4$ (otherwise the dust attenuation pdf is skewed).

3.4 Sampling

The complete model includes 19 free parameters (11 of which describe the shape of the SFH), which are summarized in Table 2. We follow the sampling procedure outlined in Johnson et al. (2021b) (see also Tacchella et al. 2021), using the dynamic nested sampling (Skilling 2004; Higson et al. 2019) algorithm DYNESTY[6] (Speagle 2020) to efficiently sample the high-dimensional parameter space of the model and build posterior pdfs. This approach provides full posterior distributions of the model parameters together with their degeneracies. A useful primer on Bayesian methods can be found in van de Schoot et al. (2021).

Throughout this work, we report the uncertainties as 68 per cent CRs (which corresponds to the 16th to 84th per cent range) of the posterior pdfs as the majority of the distributions are non-symmetric.

3.5 Simultaneously fitting the photometry and spectroscopy

In fitting both the photometry and spectroscopy we consider the log-likelihood of the model, conditioned on the observation, to be the sum of the two individual likelihood functions:

$$\ln \mathcal{L}(d_s, d_p | \theta, \phi, \alpha) = \ln \mathcal{L}(d_s | \theta, \phi, \alpha) + \ln \mathcal{L}(d_p | \theta), \quad (4)$$

where d_s is the spectroscopic data, d_p is the photometric data, the parameters θ describe the physical model used in PROSPECTOR, the parameters α describe the spectroscopic noise model (Section 3.2), and the parameters ϕ include the spectro-photometric calibration (Section 3.3). The parameters of the physical model are summarized in Table 2. We apply no relative weighting between fitting the spectroscopy and photometry in assessing the match between the observations and SEDs.

The basic likelihood calculation is effectively a χ^2 calculation for both the spectral and the photometric data. We alter the likelihood calculation for the spectroscopy to include the noise model and outlier

[6] https://dynesty.readthedocs.io/en/latest/.

model described in Section 3.2, following the procedure outlined in appendix D of Johnson et al. (2021b).

4 RESULTS

Given the sensitivity of modelling ages of old stellar populations, and their dependence on both the flexibility of the assumed SFH and the choice of SFH prior (e.g. Leja et al. 2017, 2019), we present the results for two 'extremes' of the SFH prior: (i) an 'extended' SFH, preferring equal distribution of fractional sSFR between the time bins ($\alpha_D = 1$), and (ii) a so-called 'concentrated' SFH, preferring an unequal distribution of fractional sSFR between time bins ($\alpha_D = 0.2$). The difference between these priors is discussed in Section 3.1.1.

In assuming the SFH is extended, there is a preference for ages of half the age of the Universe and against old ages ($\gtrsim 10$ Gyr; see Fig. 2). However, the results of Villaume et al. (2022) suggest that DF44 formed its stellar population early, and shortly thereafter rapidly quenched, as determined from its inverted stellar population gradients and low iron metallicity for its mass. The 'concentrated' SFH prior has a higher likelihood for such an SFH. Moreover, the concentrated SFH prior has an overall broader implicit prior on mass-weighted age as there is no preference for where the fraction sSFR is concentrated between the time bins.

The results from the full-spectral modelling of DF44 are shown in Figs 3–5, where the fits for the extended SFH prior are shown in red, and for the concentrated SFH prior in blue. In Fig. 3, the observations are shown with the 'bestfit' models (the maximum a-posteriori model; i.e. with the highest probability of the set of samplings) and the 68 per cent CR of 500 random draws from the posteriors.

Overall, the fits to the photometry are similar between the two SFH priors; the extended SFH model has marginally smaller residuals at NUV wavelengths. Similarly, the best-fitting model spectra (multiplied by the spectrophotometric calibration polynomial) compared to the spectroscopy are nearly identical, with differences only at the <1 per cent level. Given the degeneracy between the age, dust, and metallicity, the subtle differences in these features lead to the differences in the predicted stellar population parameters.

4.1 SFH and stellar population parameters at $z = 0$

Fig. 4 shows the median (solid line) and 68 per cent CR (shaded) of the posterior pdfs for the sSFR, and corresponding SFR and mass-assembly history. Similarly, the median (dashed line) and 68 per cent CRs (hatched) for the explicit and implicit priors are shown (see also Fig. 2). The SFHs of the best-fitting models (shown in Fig. 3) are indicated with open crosses. Dotted lines are drawn at the 50th, 70th, and 90th per cent levels of the cumulative stellar mass for reference. In Table 3, we summarize the two SFH results by listing the times at which different percentiles of the final stellar mass were in place. The SFH within the time bins of the non-parametric model are provided in Appendix A.

The SFHs determined from both priors suggest that DF44 formed early, having 90 per cent of its stellar mass in place at least ∼7.2 Gyr ago ($z \sim 0.9$). Using the extended SFH prior, we find that it took ∼3.5 Gyr for DF44 to assemble between 50 and 90 per cent of its mass, suggesting a relatively fast transition between star-forming and quiescent states. The SFH determined with the concentrated prior is extreme in that more than 90 per cent of the mass formed within the first time bin, i.e. ∼12.8 Gyr ago ($z \sim 8$). During the last 5 Gyr, the
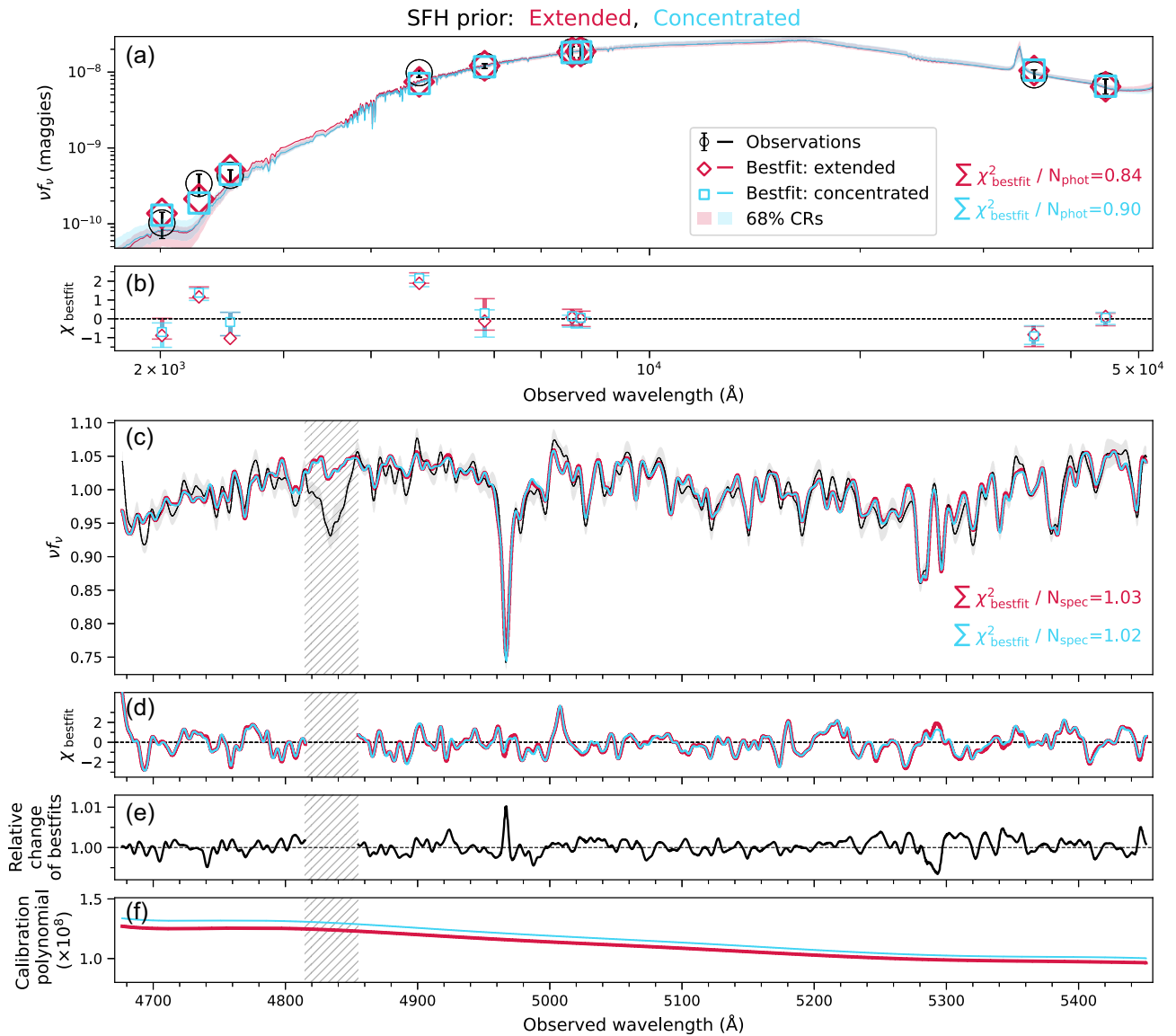
Figure 3. Summary of the fitting results for DF44. The observed data (black) is compared to the best-fitting models (red, extended SFH prior; blue, concentrated SFH prior) and the 68 per cent CR of 500 randomly drawn models from the posteriors (red/blue shaded). The corresponding posteriors are shown in Figs 4 and 5. Panel (a) shows the observed (circles) and best-fitting (diamonds and squares) photometric points, where the χ^2/N_{data} of the best-fitting SED is listed and panel (b) shows χ ([data $-$ model]/σ) of the best-fitting points. Panel (c): the observed (uncertainties shown in grey) spectrum and best-fitting spectra (multiplied by the spectrophotometric calibration polynomial). The light grey region indicates the spectral region masked throughout the fitting process. Panel (d): the χ of the best-fitting spectra as a function of wavelength. Panel (e): the relative change of the best-fitting models, i.e. the ratio of the two best-fitting spectra. Panel (f): the spectrophotometric calibration polynomials.

two results are otherwise similar, with low levels of SF until the last 100 Myr.[7]

A curious feature of both SFHs is the rise in SFR within the last 100 Myr (corresponding to the first two time bins and shaded grey in Fig. 4; by 1.8–2.4 dex). Although residual SF appears to be common for massive early-type galaxies, where ∼0.5 per cent of their mass formed within the last 2 Gyr, the fraction decreases at lower stellar

masses, consistent with galaxy 'downsizing' (e.g. Salvador-Rusiñol et al. 2020). The recent rise in DF44's SFH accounts for ≲ 1 per cent of the total stellar mass, assuming either SFH prior. While DF44 shows no indication of recent SF from the photometry, and similarly lacks emission lines in the spectrum, it is possible that H α emission (perhaps related to SF ignited by a late infall in to the Coma cluster) recently stopped. This is perhaps unlikely, however, given the lack of blue regions within the galaxy. Lee et al. (2020) concluded, based on the difference in NUV and UVW2 bands, that the light traces older stars (on ∼ Gyr time-scales, as opposed to young stars that evolve of the order of ∼ Myr time-scales). The 'recent burst' is not a consequence of an artefact in the KCWI spectrum; the same feature is apparent when fitting the MaNGA data from Gu et al. (2018). Rather, we expect this recent SF to be an artefact of the stellar models not

---

[7]Additional testing of the prior-sensitivity of the SFH showed that using $\alpha_D = 0.5$ (mildly concentrated) produced parameter values between the results from $\alpha_D = 0.2$ and 1, as expected. The mass-weighted age was found to be 11.9 Gyr, which indicates that the very old age is not overly sensitive to the choice of the α_D value.
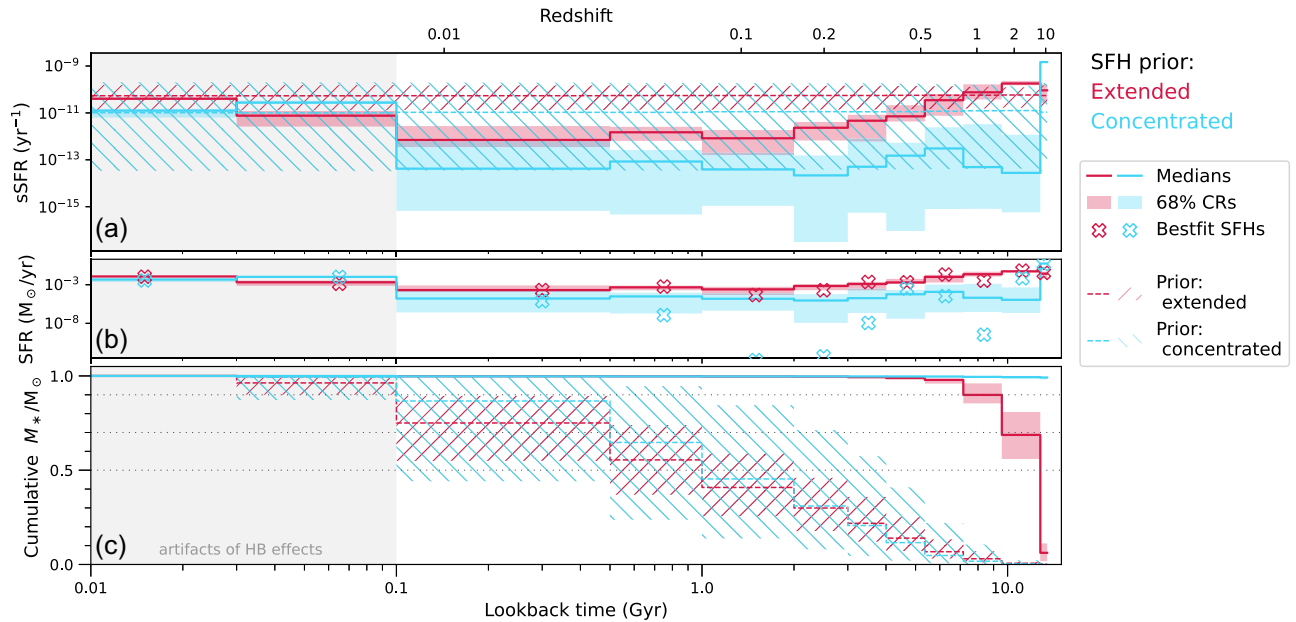
Figure 4. Posteriors for the (panel a) sSFR, (panel b) derived SFR, and (panel c) derived cumulative stellar mass for the fits shown in Fig. 3. The results and priors shown in red (blue) colours correspond to the fit with an extended (concentrated) SFH prior. Solid lines indicate the median, and the shaded regions the 68 per cent CR. These values are listed in Table A1. The SFHs of the best-fitting models are indicated with open crosses. The median (dashed line) and 68 per cent CRs (hatched) of the priors are shown for reference. Note that the cumulative mass and mass-weighted age priors are implicit, as they are derived from the sSFR prior. Dotted lines are drawn at 50, 70, and 90 per cent of the cumulative mass for reference. The last 100 Myr are shaded grey to indicate that the SFH is affected by artefacts such as HB stars (see the text).

being flexible to the contribution of blue horizontal branch (HB) stars (discussed in Appendix B1) or non-solar Mg-abundances.

We none the less test the sensitivity of the models to the presence of a very young stellar population by re-defining the time bins of our SFH, only allowing for SF older than 1 Gyr. This places a strong prior against recent SF to counteract the inability of the SPS models to correctly model the influence of the blue HB stars. In excluding SF younger than 1 Gyr, the models are better able to recover the shape of the SED, particularly in the NUV, but are marginally worse in matching the spectrum. With this revised model, we recover SFHs equivalent to that of our primary results (at times >1 Gyr), with statistically consistent but less dust and metallicity, and slightly higher stellar mass. Interestingly, with the extended SFH prior, the revised age estimate is ~2.4 Gyr older. With the concentrated SFH prior, the revised age estimate is unchanged from that of our main result. As such, we conclude that the presence of the 'recent burst' of SF does not affect our conclusion that DF44 formed and quenched very early in the history of the Universe.

Fig. 5 shows the posteriors for the normalization of the diffuse dust attenuation curve, stellar metallicity, stellar mass, and mass-weighted age. The parameters marked with an asterisk are not directly fit in our physical model, but derived from the posterior distributions. We calculate the dust extinction following equations (1) and (2) in the V band, where we use $\lambda = 5500$ Å. We note that 'total stellar mass formed' is a free parameter in our model, which we convert into 'stellar mass' by subtracting the mass lost throughout the SFH, as calculated by FSPS. The median and uncertainties of the marginalized posteriors for extended (concentrated) SFH priors are as follows:

(i) $\hat{\tau}_{\rm dust,\,diffuse} = 0.24^{+0.03}_{-0.05}$ $\left(0.20^{+0.04}_{-0.03}\right)$,
(ii) $^{*}A_{\rm V} = 0.51^{+0.11}_{-0.13}$ $\left(0.45^{+0.10}_{-0.08}\right)$,
(iii) $\log(Z_{*}/Z_{\odot}) = -1.18^{+0.01}_{-0.01}$ $\left(-1.27^{0.03}_{-0.02}\right)$,

(iv) $^{*}\log(M_{*}/M_{\odot}) = 8.23^{+0.02}_{-0.06}$ $\left(8.33^{+0.03}_{-0.03}\right)$,
(v) $^{*}t_{\rm age}\,/\,{\rm Gyr} = 10.20^{+0.34}_{-0.48}$ $\left(13.06^{+0.02}_{-0.04}\right)$,

as labelled above the one-dimensional histograms. In both cases, DF44 has a very old, modestly dusty, and metal-poor stellar population.

Contrary to our expectation that old (e.g. Péroux & Howk 2020) and metal-poor (e.g. Galliano, Galametz & Jones 2018) populations are devoid of dust (see also Barbosa et al. 2020), DF44 appears to have a non-negligible amount: The normalization of the diffuse-dust attenuation curve is $\hat{\tau}_{\rm dust,\,diffuse} \gtrsim 0.2$ and $A_{\rm V} \gtrsim 0.5$. The origin of such dust is not clear; however, Buzzo et al. 2022 recently measured similar extinction values from optical to mid-infrared photometry for a sample of quiescent UDGs. The overall shape of the SED constrains the dust content, however there are degeneracies with both metallicity and age. If we instead fix $\hat{\tau}_{\rm dust,\,diffuse} = 0$ and refit DF44 (with an extended SFH prior), the posterior pdfs are statistically consistent with that of our main result, although we note that the age increases (as expected) by ~0.23 Gyr. In Appendix B2, we discuss the fit to just the photometry, which prefers an even dustier solution ($\hat{\tau}_{\rm dust,\,diffuse} \sim 0.36$ and $A_{\rm V} \sim 0.8$, although the photometry provides no direct constraint for the metallicity, and little constraint for the age). While the spectroscopy breaks the degeneracy between dust and metallicity, the degeneracy with age remains; adding either more dust or a stellar population older than ~3 Gyr lowers the flux at wavelengths < 5000 Å (see Appendix C). Additional observations in the mid-infrared would provide better constraints on the dust content, as age and dust affect the flux in opposite directions at this wavelength range.

Other than $\hat{\tau}_{\rm dust,\,diffuse}$, the posteriors of the dust model parameters largely reflect their priors – which is to be expected, given the lack of constraining data. None the less, to check that our results do not depend on the particular dust model, we also fit the data with the dust
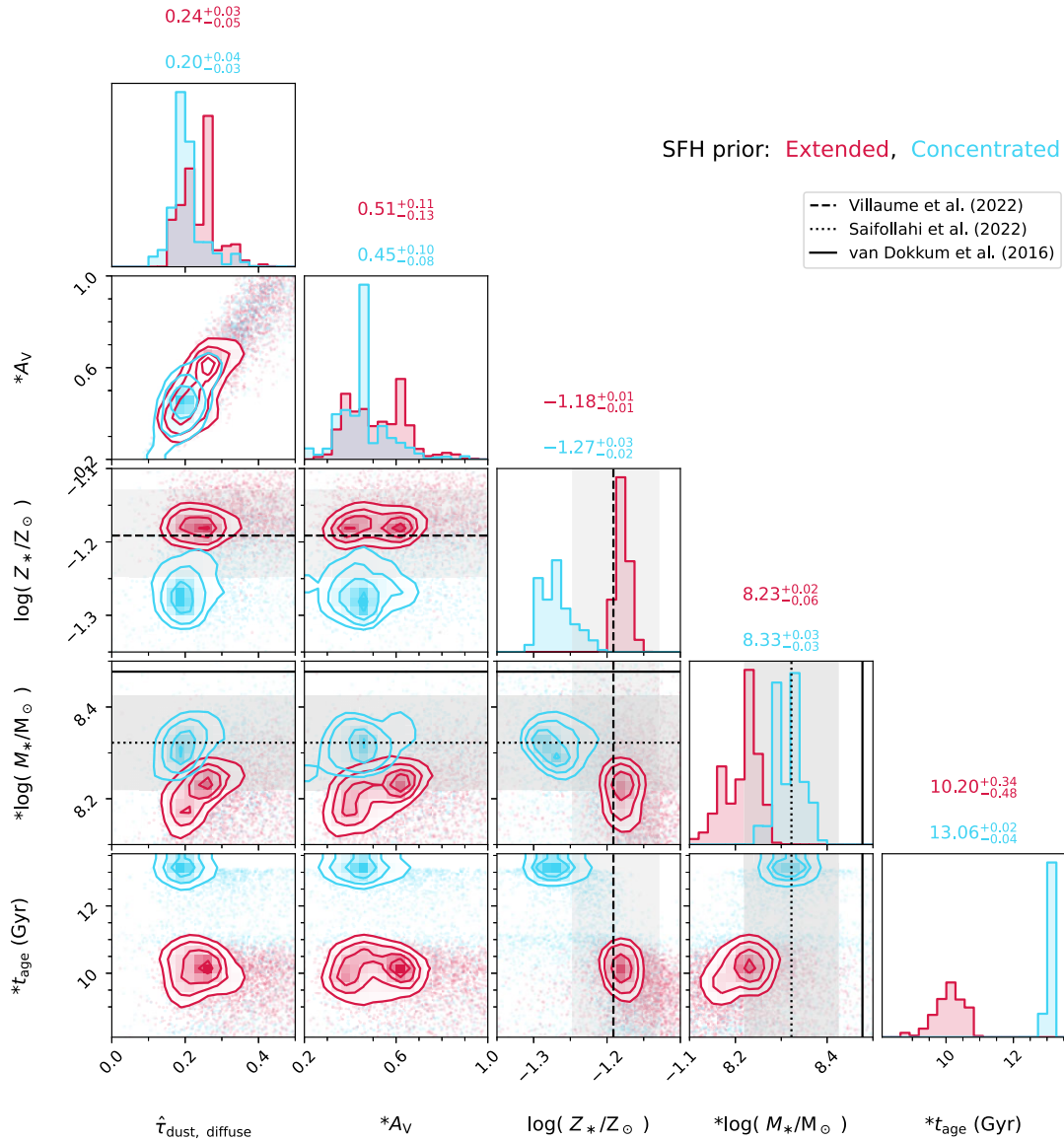
Figure 5. Posteriors of selected fitted and derived parameters (indicated with an asterisk) for the fits shown in Fig. 3. Contours are shown smoothed with a $n = 1$ Gaussian kernel, where red (blue) contours show the fits with an extended (concentrated) SFH prior. Black lines denote the expected results given the analysis by Villaume et al. (2022) for $\log(Z_*/Z_\odot)$, and stellar mass from van Dokkum et al. (2016) and Saifollahi et al. (2022). Grey shaded regions indicate the uncertainties on these values. The median and uncertainties from the 68 per cent CR for our results are listed along the top of the one-dimensional histograms.

model of Gordon et al. (2003) based on the SMC Bar (thought to have similar dust properties to dwarf elliptical galaxies, i.e. without a UV bump in the extinction curve), and find no change to our result. A degeneracy between the dust normalization and stellar mass can be seen in the joint posterior in Fig. 5, where an increase in dust suggests a higher stellar mass. As a point of comparison, a solid black line indicates the estimated stellar mass from van Dokkum et al. (2016), and a dotted black line indicates that measured by Saifollahi et al. (2022) with uncertainties reflecting the systematics of the model fitting. Both of our fits produce stellar masses lower than (and statistically inconsistent within their 68 per cent CRs) with the van Dokkum et al. (2016) value, but consistent with Saifollahi et al. (2022). Given that the photometry included in our fits is measured within an aperture, and thus does not include all of the light of the galaxy, it is not unexpected that the stellar mass we recover underestimates that from the literature.

There is an $\sim$0.1-dex difference in $\log(Z_*/Z_\odot)$ between the fits with an extended or concentrated SFH prior, where the sense of the metallicity difference is consistent with that of the age difference ($\sim$2.9 Gyr) with respect to the age–metallicity degeneracy. This indicates that we are not able to fully break the age–metallicity degeneracy with the data at hand. While in Fig. 5 we show the stellar 'isochrone' metallicity measured by Villaume et al. (2022) as a black dashed line for comparison, there are several caveats to their comparison that are discussed in the following section.

At this point the dichotomy of DF44 being 'old' or 'very-old' is subject to the choice of SFH prior. We remind the reader that the extended SFH prior behaves analogously to regularization methods used throughout the literature. While the concentrated SFH prior provides more flexibility to better recover the short and early SF expected for DF44, it is not necessarily a 'good' prior; we provide no physical information for the shape of the SFH. We simply tune

Table 3. Summary of results.

Time-scale (Gyr)	Extended SFH prior			Concentrated SFH prior		
	16	50	84	16	50	84
t_{10}	0.60	0.87	1.12	0.068	0.068	0.069
t_{20}	1.18	1.39	1.64	0.135	0.136	0.137
t_{30}	1.63	1.79	2.33	0.203	0.204	0.206
t_{40}	2.07	2.33	2.92	0.271	0.272	0.274
t_{50}	2.55	2.85	3.53	0.339	0.340	0.343
t_{60}	2.93	3.41	4.17	0.406	0.408	0.411
t_{70}	3.37	4.07	4.92	0.474	0.475	0.480
t_{80}	3.86	5.39	5.70	0.542	0.543	0.549
t_{90}	5.73	6.32	7.01	0.610	0.611	0.617
t_{95}	6.25	7.57	7.93	0.643	0.645	0.652

Notes. Time-scales t_x correspond to the time at which x per cent of stellar mass had formed, in units of Gyr since the big bang. Given that our SFH is a step function, we interpolate to estimate t_x. We provide the 16th, 50th, and 84th percentiles of the posterior (i.e. the 68 per cent CR) as an estimate of the uncertainty, although this is likely an underestimate, given the width of the time bins in our SFH. Given the observed redshift of DF44 and the adopted cosmology, the age of the universe is 13.47 Gyr. Time-scales in units of lookback time are therefore $t_{Lookback} = 13.47$ Gyr $- t_x$. The fractional SFH within the time bins of the non-parametric model (i.e. not interpolated) are listed in Table A1 in Appendix A.
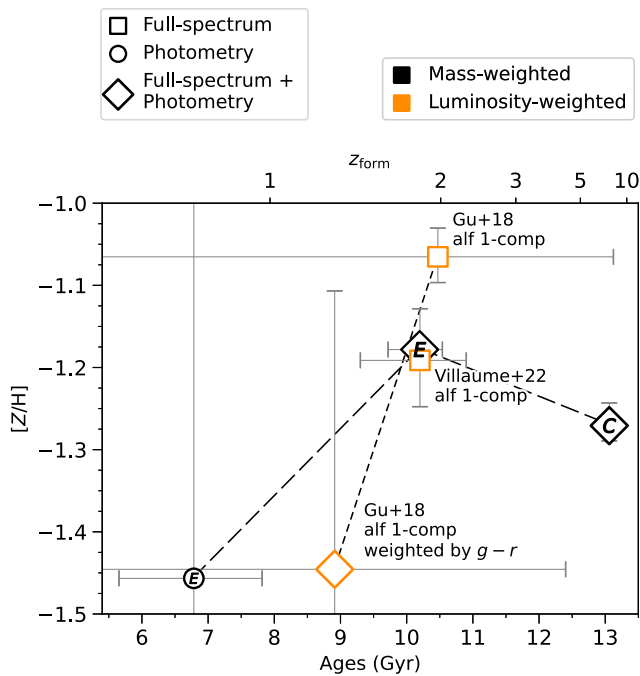

Figure 6. Comparison of stellar metallicity and age for DF44 from this work, Gu et al. (2018), and Villaume et al. (2022) (the latter using the same spectroscopic data set as ours). Both results from the literature derived values using ALF, and thus are not directly comparable to our results using PROSPECTOR (see the text). Black coloured points show mass-weighted ages, while orange points show luminosity-weighted ages. Marker shapes indicate the data used in fitting the stellar properties. Dashed lines connect results obtained from the same study. We mark the results from this work derived with an extended SFH with an 'E', or with the concentrated SFH with a 'C'.

the prior such that it prefers to distribute the SF within fewer time bins (see Section 3.1.1).

This prior-dependence problem is exacerbated with less complete or lower S/N data sets. As a brief example, in Fig. 6, we compare

the stellar metallicities and ages determined through fitting both the spectrum and photometry (diamond), with that fitted to only the photometry (circle) for the extended SFH prior (points marked with an 'E'). While the NUV–NIR photometry provides information on the dust in DF44 (see Appendix B), the age estimate is more heavily weighted by the SFH prior than are the full spectrum fitting results. Accordingly, the photometry-only fit gives a median age ∼3.4 Gyr younger than the fit to the spectrum and photometry together.[8]

4.2 Which SFH prior is preferred?

There is little statistical evidence to decide whether the results from either SFH prior better reflects the 'true' properties (or SFH) of DF44.[9] The distributions of SED models shown in Fig. 3 are similar between the fits with each prior, and the models have similar residuals.

There are subtle differences, however, particularly around the H β and Mg II features where the concentrated SFH gives a (statistically) lower χ^2. The H β line is sensitive to recent SF (and to HB stars, as discussed in Section 4.1), while Mg II is to sensitive the α-abundance of the stellar population. The FSPS models that we use are currently limited to fixed solar α-abundance. However, Villaume et al. (2022) found that DF44 has [Mg/Fe]$= 0.11^{+0.06}_{-0.04}$ through fitting the same spectrum of DF44 as this work with the full-spectrum fitting code ALF (Conroy et al. 2018), which includes response functions to measure the non-solar chemical abundance variations. Given the relationship between both features and the age of the stellar population, this points to the need to include more complex stellar populations variables, e.g. α-abundance, in models in order to break this degeneracy.[10]

Fig. 6 compares the stellar metallicities and ages measured for DF44 by this work, Villaume et al. (2022), and Gu et al. (2018).[11] Both previous studies fitted rest-frame optical spectra of DF44 with the full-spectrum fitting code ALF. We caution that there are fundamental differences between ALF and PROSPECTOR that make their results only broadly comparable, e.g. the inclusion of non-solar

---

[8]If instead of the non-parametric model, we assume the SFH follows a delayed exponential form (a common parametric model adopted within the literature) we find similar results. With a logarithmically uniform prior on the e-folding time, τ, and linearly uniform prior for the delay time, t_{age}, the implicit age prior has a complex form with 16th, 50th, and 84th percentiles of 1, 3.8, and 8.4 Gyr, respectively – preferring younger ages than the extended SFH prior results. The implicit age skews even younger if instead τ is linearly sampled. Fitting the photometry of DF44 suggests the age is ∼8.2 Gyr, and slightly less dusty than using the extended SFH model. Fitting both the photometry and spectroscopy suggests the age is ∼13.6 Gyr, and slightly less dusty and more metal-poor than our main result. We note that the photometry-only results with the delayed parametric model appear particularly sensitive to the S/N – if we inflate the photometric uncertainties by a factor of 2, the age posterior decreases by ∼2 Gyr. The same is not true when using the non-parametric models.

[9]Comparing the Bayesian evidence of the two fits (as derived from the nested sampling described in Section 3.4), we find a strong preference (according to the Jefferys scale, see e.g. Kass & Raftery 1995) for the concentrated SFH prior (ln $Z_{concentrated} = 62590$ is much larger than ln $Z_{extended} = 62542$), where Z is the Bayesian evidence. However, this likely reflects the fact that the old age of DF44 is more disfavoured by the extended SFH prior (see Section 3.1.1) more than a preference of the data itself.

[10]While FSPS does include an option to set the fraction of blue HB stars, for technical reasons, we cannot include it as a free parameter in our models.

[11]The stellar 'isochrone' metallicity (distinct from that which includes the response function for individual elements) [Z/H] values from Villaume et al. (2022) and Gu et al. (2018) were provided via private communication.

abundance patterns (as mentioned above), and ALF fits a single-age stellar component (with a uniform prior with minimum age of 1 Gyr) rather than an SFH. That said, the luminosity- and mass-weighted ages should be comparable, given that DF44 is old.

Villaume et al. (2022) fitted the same KCWI spectrum as this work, while Gu et al. (2018) fitted a MaNGA spectrum that covers a broader wavelength range (including several additional age diagnostics: Hδ, Hγ, Ca II H&K, and G band). The MaNGA spectrum has an S/N $\sim$ 8 Å^{-1}, however, which is only $\sim$12 per cent the S/N of the KCWI spectrum. Despite differences in data, the two studies both found the age of DF44 to be $\sim$10.5 Gyr, although the stellar metallicities are formally discrepant.[12] Notably, Gu et al. (2018) also considered the $g - r$ colour of DF44 from Dragonfly imaging, and re-weighted their posteriors, which significantly lowers their metallicity value (and is then consistent with Villaume et al. 2022 owing to its large uncertainty).

Considering that we fit DF44 in a completely independent way compared to these studies, it is at least encouraging that the results are fairly similar. Significant variations among age and metallicity measurements for the same object, measured between different studies, is not unique to DF44. In Appendix B4, we outline two additional examples and discuss the reasons behind their differences.

The comparison shown in Fig. 6 demonstrates the difficulty in measuring the stellar properties of old stellar populations, related both to limitations of data and modelling. As discussed in the previous section, a solution is within reach as the inclusion of a variable α-abundance or the addition of mid-IR photometry would help to break degeneracies between the stellar population properties.

We conclude that DF44 has an age of $\sim$10–13 Gyr. Without clear statistical evidence to favour one SFH model over the other, throughout the remainder of this work, we present both sets of results. In the next section, we discuss the implications of such a large sized galaxy having formed the bulk of its stellar mass very early.

5 DISCUSSION

In this work, we sought to measure the detailed SFH of DF44 as a means to distinguish between UDG formation scenarios, which predict a variety of quenching times (i.e. SFHs). The consistent narrative among theoretical simulations is that UDGs are contiguous with the canonical dwarf population. However, Villaume et al. (2022) established that DF44 is dissimilar to canonical dwarf galaxies with respect to both the stellar population gradients, stellar metallicity, and kinematics. In measuring the SFH of DF44, we can further test this scenario.

Previous analyses of DF44 found that its stellar population is old, having an age of $\sim$10 Gyr (see Fig. 6; Gu et al. 2018; Villaume et al. 2022). In this work, we have shown that DF44 formed the majority of its mass early, where we consider the galaxy 'quenched' after it forms $\sim$90 per cent of its mass. In using an extended SFH prior we obtain a lower limit of the quenching epoch of $z \sim 0.9$ ($\sim$6.3 Gyr after the big bang). Alternatively, in using a concentrated SFH prior (motivated by the results of Villaume et al. 2022), we recover an extremely early quenching epoch of $z \sim 8$ ($\sim$0.6 Gyr after the big bang). In either case we find that DF44 is old, the distinction being that a concentrated SFH prior suggests that it is *very* old. Without clear statistical evidence to favour one prior over
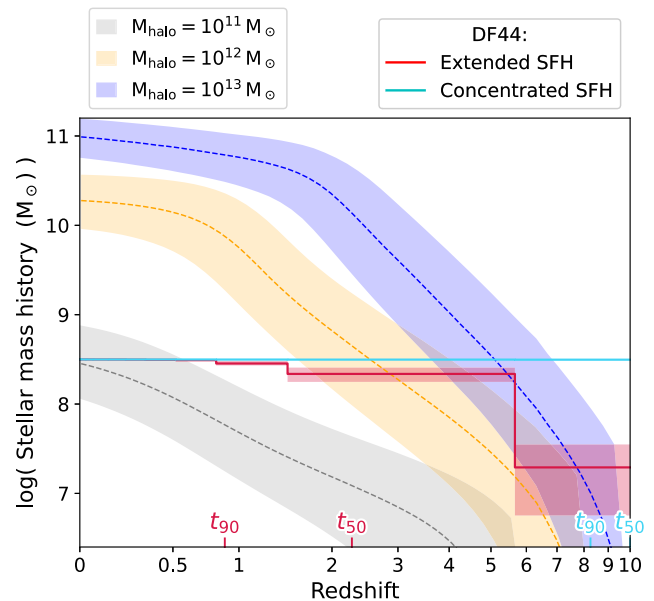
Figure 7. DF44's mass assembly history compared to total stellar mass histories of the main progenitor for haloes of several masses from the empirical model of Behroozi et al. (2019). Lines indicate the median values, and shaded regions correspond to the 68 per cent CRs. The halo mass bins have width of ±0.25 dex. Time-scales of DF44's SFH are labelled on the plot: when 50 per cent and 90 per cent of the final mass had been formed. At the time that DF44 quenched ($z > 0.9$), it was already more massive than expected for most normal dwarf galaxies.

the other (see Section 4.2) we instead focus on providing a qualitative comparison of the implications of the two results.

For either of our two results, the bulk formation of DF44 occurs during an epoch where the evolution of galaxies in dwarf-scale dark matter haloes ($\lesssim 10^{11}$ M$_\odot$) significantly differs from that of galaxies in more massive haloes. The mass assembly histories expected for average galaxies with dark matter halo masses between 10^{11} and 10^{13} M$_\odot$ are shown in Fig. 7, from the empirical model of Behroozi et al. (2019).[13] The mass assembly history of DF44 (as shown in Fig. 4) is shown for comparison.

While the current stellar mass of DF44 falls within the range expected for the $z = 0$ canonical central dwarf population, and its halo mass is in the neighbourhood of $\sim 10^{11}$ M$_\odot$ (e.g. van Dokkum et al. 2016, 2019; Wasserman et al. 2019; see also Bogdán 2020), its mass assembly history is not necessarily compatible with this population. In fact, the mass of DF44 at $z \sim 8$ was typical for galaxies destined to become brightest cluster galaxies (BCGs) – however, the mass growth was halted. This provides our first evidence that DF44 may not originate among the canonical field dwarf population.

We now further investigate our results in the context of the predictions of UDG formation scenarios from theoretical work.

---

[12]Villaume et al. (2022) considered the presence of a second young population (aged 1–3 Gyr), which lowers their age estimate by 0.6 Gyr but is consistent with their fiducial fit.

[13]The population averages likely overestimate the SF time-scales of galaxies in dense environments (e.g. Thomas et al. 2010). Indeed, dEs and UDGs in clusters are found to be old (e.g. Weisz et al. 2011; Ferré-Mateu et al. 2018; Ruiz-Lara et al. 2018), and satellite dwarfs in simulations appear to form much earlier than central dwarfs (Digby et al. 2019; Garrison-Kimmel et al. 2019; Joshi et al. 2021). However, DF44 is likely only on its first infall into the Coma cluster (discussed further below).
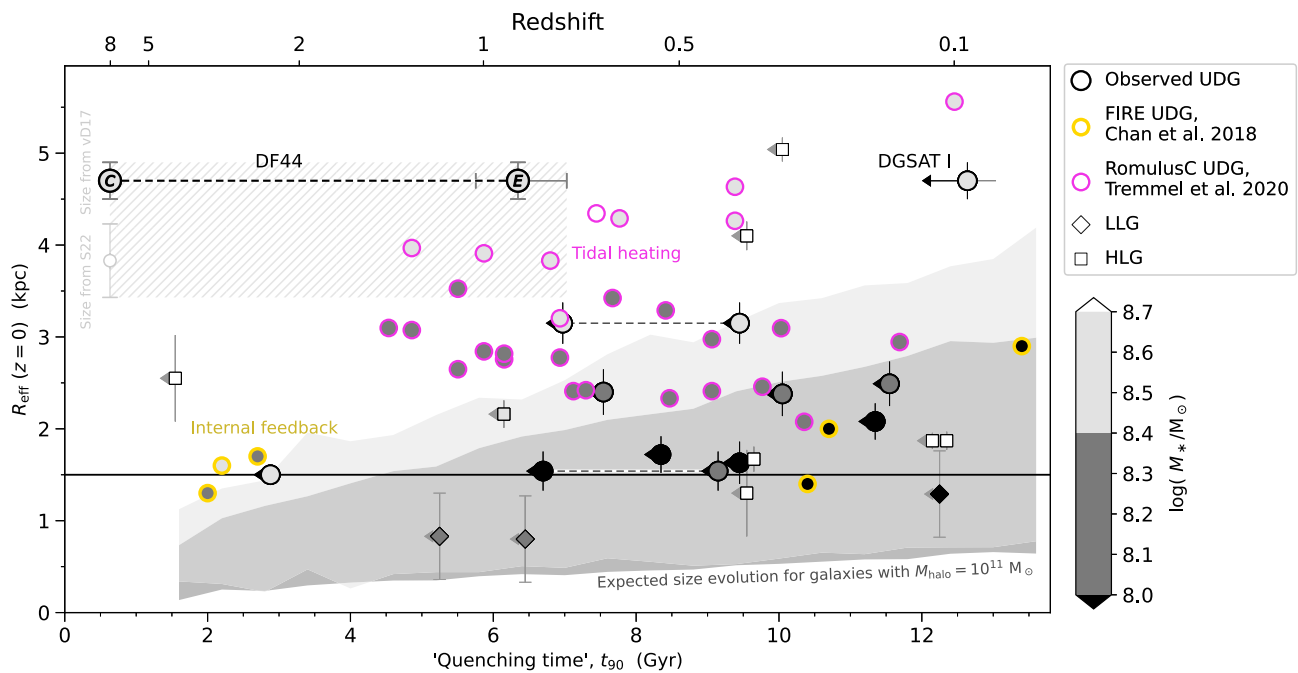
Figure 8. Comparison of effective radius and quenching time of UDGs (circles) between observations (black outlines) and predictions from two UDG formation scenarios (coloured outlines). Quenching times are shown in units of age of the Universe, such that $t_{90} = 0$ is the big bang. We show the quenching time results for DF44 as points with a 'C' (concentrated SFH prior) and with an 'E' (extended SFH prior). A hatched region fills the parameter space covered by the range in DF44's size and quenching time measurements. The distribution of dwarf galaxies in two mass ranges is shown with grey shaded regions. A horizontal line shows the size threshold for UDGs, $R_{eff} > 1.5$ kpc. We show UDGs from two simulations: FIRE (yellow outlines; Chan et al. 2018) and RomulusC (magenta outlines; Tremmel et al. 2020), which are described in the main text of Section 5. *The large size and early quenching of DF44 are inconsistent with the internal feedback model of FIRE, but is not inconsistent with the tidal-heating model of* RomulusC (although see the text). Observations from the literature are shown with arrows indicating that they may in fact be upper limits, given the potential bias in their measured SFH time-scales, see Section 5.2. Most of these UDGs fall within the expected size range of normal or tidally enlarged dwarfs. We also include low- (LLGs; diamonds) and high-luminosity galaxies (HLGs; squares) from the literature. Dashed lines connect values obtained for the same object. Points are coloured according to their stellar mass.

5.1 DF44 in tension with UDG formation scenarios

There have been many scenarios that have come out of cosmological simulations and SAMs that satisfy the size and surface brightness constraints of the observed UDG population. As these scenarios all work under the same constraints of conventional galaxy evolution physics, they follow the same dichotomy: whether the significant size growth necessary to transform a canonical dwarf galaxy into a UDG occurs pre- or post-infall (which is to say, whether the cluster environment is necessary to the process) is related to whether they are late or early infallers into that environment. Nearly[14] all of these models require the infall into the dense, hot environment of a cluster to quench the UDGs such that the quenching time is directly linked to the infall time.

This dichotomy is demonstrated in Fig. 8 where we show the relation between the effective radius and the quenching time for three different scenarios: i) what is typically expected under 'normal' conditions (grey shaded regions),[15] (ii) the predictions for the 'internal feedback' scenario from the FIRE simulations (gold outlined circles; Chan et al. 2018), and (iii) the predictions from the RomulusC

simulations (magenta outlined circles; Tremmel et al. 2020).[16] The symbols are additionally colour-coded by stellar mass.

We show our quenching time results for DF44 in this figure with the size measured by van Dokkum et al. (2017) based on the deepest imaging available ($R_{eff} = 4.7 \pm 0.2$ kpc; with black outlines). However, Saifollahi et al. (2022) obtained a smaller size ($R_{eff} = 3.83 \pm 0.4$ kpc; grey point) based on the same data. A hatched region covers the parameter space of the various quenching times from our results, and the two size measurements.

The FIRE UDGs largely follow the expected size growth trend, where the distinction is that their internal feedback causes bursts of SF (which puff up the sizes of the galaxies, prior to quenching) place them on the top end of the size distribution. While Chan et al. (2018) predicted that there are objects with quenching times as early as measured for DF44, these objects barely reach the nominal size of UDGs ($R_{eff} < 2$ kpc). Indeed, they could not reach the size of DF44 without significantly more time to form stars, which would then violate the stellar mass/surface brightness constraint.

A formation scenario that can explain both large and early-quenched UDGs is tidal heating, where the expected size–quenching time trend is the exact opposite of the FIRE scenario, i.e. the earliest infallers/quenchers will be the largest because they have spent the longest time expanding due to the cluster environment. We see this

---

[14]The exception is Wright et al. (2021) in which a small fraction of isolated UDGs that experience major mergers are quenched at $z = 0$. This scenario is discussed further below.

[15]The expected size growth was determined from the stellar mass assembly histories of Behroozi et al. (2019) and the size–mass relation of Sales et al. (2020).

[16]These values were taken from figs 2 and 11 of Tremmel et al. (2020).

effect demonstrated in the RomulusC points.[17] While Tremmel et al. (2020) show that some objects reach the nominal sizes of UDGs prior to infall into a cluster, in order to reach the large end of UDG sizes requires the additional effect of tidal heating from the cluster environment. With the extended SFH prior, the quenching time of DF44 is reasonably consistent with the tidally heated RomulusC UDGs.

Certainly tidal heating is happening on some level to some galaxies in clusters. Evidence of such has been observed among proto-UDGs in clusters (e.g. Grishin et al. 2021). Carleton et al. (2019) interpreted the radial alignment of UDGs in Coma (Yagi et al. 2016, which includes DF44) as evidence that these galaxies have been tidally influenced. While we cannot discount the tidal heating scenario in explaining the size and quenching times of DF44, this scenario conflicts with other properties of DF44.

Measurements of the kinematics and dynamics of DF44 indicate that it has not been in the cluster environment long enough to be impacted by tidal effects. Its position in phase-space points to a late infall into the Coma cluster (<2 Gyr ago; Alabi et al. 2018). Moreover, DF44 appears to be part of a dynamically cold group that would have surely been disrupted if tidal heating had taken place (van Dokkum et al. 2019), and there is no distortion in its ellipticity that would be a marker of tidal heating (Mowla et al. 2017).

Together with the above points, the SFH provides evidence that DF44 certainly quenched prior to cluster infall. This would suggest that its progenitor was larger than a dwarf galaxy or that a process unrelated to environment caused an expansion. This interpretation is consistent with the conclusion of Saifollahi et al. (2022), who find that the elevated GC populations at a given stellar mass (N_{GC}/M_*) of large UDGs (including DF44) are inconsistent with scenarios that explain the sizes of UDGs via redistributing the stars to larger radii (i.e. tidal interactions, stellar feedback, or high-spin). Villaume et al. (2022) similarly ruled out such scenarios given DF44's 'inside-out' stellar population gradients. Therefore, *how* DF44 quenched is the crucial question to answer to understand its origins.

From simulations, only Wright et al. (2021, based on ROMU-LUS25; Tremmel et al. 2020) have proposed a scenario, 'early major mergers',[18] in which UDGs can form and quench[19] without relying on environmental quenching mechanisms. The UDGs in ROMULUS25 had their star-forming gas and SF moved outwards from the central cores of the galaxies to larger radii by major mergers $\sim$8–11 Gyr ago. For most of the simulated UDGs, SF continued in the galaxy outskirts, while the central core passively dimmed, leading to negative radial age gradients.

Considering that DF44 quenched $\gtrsim 7$ Gyr ago, this may suggest that a major merger is responsible for (or at least concurrent with) its quenching – and that there would be a flat age gradient. The central (<0.5 kpc) SFH predicted for ROMULUS25 UDGs is broadly

consistent with DF44's SFH when assuming an extended SFH prior, although not when assuming a concentrated prior (which quenches much earlier). Villaume et al. (2022) measured a flat-to-negative [Mg/Fe] gradient out to $\sim$2.5 kpc, which taken as a proxy for an age gradient is not strictly inconsistent with this scenario.[20]

Further work is needed in order to establish whether DF44 is the product of an early major merger. For instance, the mechanism that quenches $\lesssim 5$ per cent of the ROMULUS25 UDGs is not fully described, providing no point of comparison with DF44's SFH or stellar population gradients. Moreover, when this quenching occurs, or whether the galaxies remain quenched, is unclear. While Wright et al. (2021) and Van Nest et al. (2022) explored the predictions of 'early major mergers' in differentiating average UDGs and non-UDGs, the fact that DF44 is a rare case warrants more detailed comparisons.

The results of this work show that DF44 has been shaped by some rare galaxy evolution process, no matter whether the 'true' SFH resembles our result with an extended or concentrated SFH prior, or falls somewhere in between. As was shown in Fig. 7, the early SFR of DF44 is more typical of normal (MW-like) star-forming galaxies at $z > 3$ (Rinaldi et al. 2021). The implication is that it is not the early, extreme SFH that makes DF44 unusual among $z = 0$ galaxies, but rather its sudden quenching. Given the lack of galaxies like DF44 in cosmological simulations, this would imply that galaxy evolution models are not capturing the true diversity of quenching mechanisms.

In fact, cosmological simulations already struggle to reconcile the opposing stellar mass–effective radius constraints for objects like DF44 in the context of the broader galaxy population. A common problem among cosmological simulations is that they do not accurately reproduce the population of normal sized dwarfs (e.g. Chan et al. 2018; El-Badry et al. 2016; Lupi, Volonteri & Silk 2017; Tremmel et al. 2020; Benavides et al. 2021; see also Jiang et al. 2019a). Since this points to issues in the implementation of SF and related feedback, the evidence from this work and Villaume et al. (2022) that there are objects like DF44 that require even more intense SF feedback exacerbates this problem.

Analytic and semi-analytic models can avoid such issues to some degree. With respect to size, several UDG formation scenarios apply empirical distributions (e.g. Carleton et al. 2019; Sales et al. 2020) but they are then subject to the likely bias of 'getting out what they put in' (see Jiang et al. 2019b). With respect to SF and feedback, Danieli et al. (2022) analysed the large number of GC candidates hosted by NGC 5846_UDG1 (Forbes et al. 2021) with a model that connects the evolution of a galaxy with its dark matter halo and GC populations (Trujillo-Gomez, Reina-Campos & Kruijssen 2019) to show that it is plausible that clustered supernova feedback could significantly increase the mass loading factor of gas outflows. However, these models miss an important component of galaxy evolution – the impact of the different environments a galaxy moves through over its lifetime. DF44's very early quenching and relatively late infall into the Coma cluster invokes the question of what has it been doing for the last $\sim$10 billion years? Given the potential 'pre-processing' by group environments or filaments that can affect everything from the size of a galaxy's dark matter halo, to its SFH and present-day GC population, makes it vital to understand this aspect of galaxy evolution in general.

---

[17]We note that other simulations and SAMs that invoke tidal heating (e.g. Carleton et al. 2019; Liao et al. 2019; Jiang et al. 2019a; Sales et al. 2020) with slightly different prescriptions (i.e. cuspy versus cored dark matter haloes, how early satellite infall begins) could change the exact predictions of the sizes of UDGs. We note that RomulusC appears to overpredict the sizes of cluster dwarfs by a factor of $\sim$2 (compared to observations from Eigenthaler et al. 2018).

[18]We note that Saifollahi et al. (2022) refer to this scenario as a 'lack of late mergers'.

[19]Less than 5 per cent of the simulated UDGs with masses $M_* > 10^8$ M$_\odot$ are quenched, in the sense that they are gas-poor. This population is dominated by galaxies that have had an interaction with a more massive halo and/or AGN activity.

[20]While Villaume et al. (2022) measured a flat age gradient, they note that given the limitations of modelling granular differences in old stellar populations, the [Mg/Fe] gradient is more sensitive to age variations.

5.2 DF44 in context

The prior-dependence of the SFH for old stellar populations, even with high-S/N data, means that further work is needed to understand what 'good' SFH priors are for these systems. The problem is amplified at lower S/N, where the prior will have a stronger influence on the posterior pdfs (see Appendix B3 for an example). Consequently, it is not straightforward to compare results between studies in the literature. With this caveat in mind, we also show in Fig. 8 the quenching times and sizes of UDGs from three studies (Ferré-Mateu et al. 2018; Ruiz-Lara et al. 2018; Martín-Navarro et al. 2019), and for comparison high- and low-luminosity dwarfs in Coma (squares and diamonds, respectively; Ferré-Mateu et al. 2018). Arrows attached to these points indicate that they are perhaps upper limits, given potential biases from the use of regularized SFHs (akin to the extended SFH prior used in this work; see the discussion in Appendix B4). We note that the UDGs from the literature are shown with effective radii from the catalogue of Alabi et al. (2020) when possible, where DF44 was found to have a size of 3.74 ± 0.23 kpc in the *Subaru*/Suprime-Cam *R* band.

Regardless of potential biases in the SFHs, there are still interesting conclusions to draw from this data set. DF44 stands out as an outlier among the largest observed UDGs with an early quenching time, for any of the discussed quenching times or sizes. On the other hand, the UDG DGSAT I stands out with both the largest size and latest quenching time among the literature values shown in Fig. 8, and it is also the only non-cluster member. Unlike the rest of the UDGs, DGSAT I is similar to a subset of the RomulusC UDGs that follow a trend in size-quenching time in distinct disagreement with the standard expectations of tidal heating. Its size is also well outside of what is plausible for the concentrated SFH scenario, or normal expectations of size growth given its late quenching time.

While it is outside the scope of this work to examine DGSAT I in detail, it is relevant to this discussion in that it further provides evidence that multiple observed objects, all of which are 'UDGs', in fact, have distinct formation pathways.

That DF44 attained a similar stellar mass and size as the other large galaxies, but much earlier, supports the idea that it is either the product of unconventional galaxy evolution processes, or it was interrupted from becoming a much more massive galaxy by some catastrophic quenching event. Speculation of the latter has also been drawn on the basis of the wide range of GC counts among UDGs, and the range of implied dark matter halo masses (with some having little to no dark matter). This is the first time this diversity has been shown in the SFHs of the galaxies' field star populations.

6 SUMMARY

In this work, we simultaneously fit NUV to NIR photometry and high S/N rest-frame optical spectroscopy of the UDG DF44 with an advanced physical model. Our model includes non-parametric SFHs, a flexible dust attenuation law, a white noise model, and an outlier model, which we fit to the observations in a fully Bayesian framework with PROSPECTOR.

We find that DF44 formed the majority of its stellar mass (>90 per cent) early, although how early is sensitive to the choice of the SFH prior and degeneracies between stellar population parameters. Using an extended SFH prior akin to similar studies in the literature (which strongly favours ages of half the age of the Universe, and therefore disfavours very old ages) we find that DF44 formed by $z \gtrsim 0.9$. If we instead adopt prior knowledge from DF44's stellar population gradients that the DF44 formed early and rapidly

quenched (Villaume et al. 2022), such that its SFH is concentrated within a short time-scale, we find that DF44 assembled as early as $z \sim 8$. Neither of these priors encode physical information of the shape of the SFH based on a priori knowledge, and thus neither are necessarily 'good' priors. Further work is needed to understand what 'good' SFH priors are for such old galaxies from a theoretical standpoint. Even with the high-S/N spectral data used in this work (~ 96 Å^{-1}), the data showed no statistical preference for either result. Improved age constraints are possible with the inclusion of observations in the mid-infrared in that this would pin down the dust attenuation, which, in the NUV, is degenerate with the contribution of old stellar populations. Improvements in the models (e.g. including variable α-abundance) to replicate old and complex stellar populations are also needed.

DF44's early and short SFH determined from this work, together with previous results that DF44 is very metal poor for its mass, and that stellar population gradients indicate 'inside-out' formation (unlike kinematically- and morphologically-similar dwarfs; Villaume et al. 2022), points towards an unusual origin, likely distinct from the canonical dwarf population. UDG formation scenarios outlined in simulations only predict the SFH and size of DF44 through invoking prolonged environmental effects, yet we conclude that DF44 quenched prior to accretion into the Coma cluster. While analysis of the ROMULUS25 simulation by Wright et al. (2021) proposes early major mergers as a means to produce UDGs in the field, it is not yet clear if the properties of DF44 are fully consistent with this scenario. Instead, DF44 may be a 'failed galaxy' with its initial size, or whatever processes that expanded it, being unrelated to its environment. In summary, early quenching an late infall taken together rules out most UDG formation scenarios except for the failed-galaxy and early-major-mergers (with the caveats above). Additional work is needed to explain the old quiescent UDGs from a theoretical standpoint, while reproducing the observed stellar properties beyond general size–mass trends.

This work made use of the following softwares: ASTROPY (The Astropy Collaboration et al. 2018, 2013), DYNESTY (Speagle 2020), FSPS (Conroy et al. 2009; Conroy & Gunn 2010), IPYTHON (Pérez & Granger 2007), MATPLOTLIB (Hunter 2007), NUMPY (van der Walt, Colbert & Varoquaux 2011), PYTHON-FSPS (Johnson et al. 2021a), PROSPECTOR (Leja et al. 2017; Johnson et al. 2019, 2021b), and SCIPY (Virtanen et al. 2020).

DATA AVAILABILITY

The data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Alabi A. B., Romanowsky A. J., Forbes D. A., Brodie J. P., Okabe N., 2020, MNRAS, 496, 3182
Alabi A. et al., 2018, MNRAS, 479, 3308
Allanson S. P., Hudson M. J., Smith R. J., Lucey J. R., 2009, ApJ, 702, 1275
Amorisco N. C., Loeb A., 2016, MNRAS, 459, L51
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Barbosa C. E. et al., 2020, ApJS, 247, 46
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212
Benavides J. A., et al., 2021, Nat. Astron., 1, 5
Bogdán Á., 2020, ApJ, 901, L30
Boselli A., Gavazzi G., 2006, PASP, 118, 517
Buzzo M. L. et al., 2022, MNRAS, preprint (arXiv:2208.11819)
Cabrera-Ziri I., Conroy C., 2022, MNRAS, 511, 341
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Carleton T., Errani R., Cooper M., Kaplinghat M., Peñarrubia J., Guo Y., 2019, MNRAS, 485, 382
Carleton T., Guo Y., Munshi F., Tremmel M., Wright A., 2021, MNRAS, 502, 398
Carnall A. C., Leja J., Johnson B. D., McLure R. J., Dunlop J. S., Conroy C., 2019, ApJ, 873, 44
Chabrier G., 2003, PASP, 115, 763
Chan T. K., Kereš D., Wetzel A., Hopkins P. F., Faucher-Giguère C. A., El-Badry K., Garrison-Kimmel S., Boylan-Kolchin M., 2018, MNRAS, 478, 906
Charlot S., Fall S. M., 2000, ApJ, 539, 718
Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Conroy C., 2013, ARA&A, 51, 393
Conroy C., Gunn J. E., 2010, Astrophysics Source Code Library, record ascl: 1010.043
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Conroy C., Villaume A., van Dokkum P. G., Lind K., 2018, ApJ, 854, 139
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659
Danieli S. et al., 2022, ApJ, 927, L28
Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98
Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, MNRAS, 466, L1
Dickey C. M. et al., 2021, ApJ, 915, 53
Digby R. et al., 2019, MNRAS, 485, 5423
Dotter A., 2016, ApJS, 222, 8
Draine B. T. et al., 2007, ApJ, 663, 866
Eigenthaler P. et al., 2018, ApJ, 855, 142
El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131
Faber S. M. et al., 2007, ApJ, 665, 265
Fazio G. G. et al., 2004, ApJS, 154, 10
Ferré-Mateu A. et al., 2018, MNRAS, 479, 4891

Fillingham S. P., Cooper M. C., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., Wheeler C., 2018, MNRAS, 477, 4491
Forbes D. A., Gannon J. S., Romanowsky A. J., Alabi A., Brodie J. P., Couch W. J., Ferré-Mateu A., 2021, MNRAS, 500, 1279
Galliano F., Galametz M., Jones A. P., 2018, ARA&A, 56, 673
Gannon J. S. et al., 2022, MNRAS, 510, 946
Garrison-Kimmel S. et al., 2019, MNRAS, 487, 1380
Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, ApJ, 757, 85
Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, ApJ, 594, 279
Greco J. P., Goulding A. D., Greene J. E., Strauss M. A., Huang S., Kim J. H., Komiyama Y., 2018, ApJ, 866, 112
Grishin K., Chilingarian I., Afanasiev A., Fabricant D., Katkov I., Moran S., Yagi M., 2021, Nat. Astron., 5, 1308
Gu M. et al., 2018, ApJ, 859, 37
Han Y., Han Z., 2018, ApJS, 240, 3
Higson E., Handley W., Hobson M., Lasenby A., 2019, Stat. Comput., 29, 891
Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)
Hora J. L., 2021, jhora99/imclean: imclean 20210323, Zenodo
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jackson R. A. et al., 2021, MNRAS, 502, 4262
Jiang F. et al., 2019b, MNRAS, 488, 4801
Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A. A., Macciò A. V., Di Cintio A., 2019a, MNRAS, 487, 5272
Johnson B. D., Leja J. L., Conroy C., Speagle J. S., 2019, Astrophysics Source Code Library, record ascl:1905.025
Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021b, ApJS, 254, 22
Johnson B. et al., 2021a, Dfm/Python-Fsps: Python-Fsps v0.4.1rc1, Zenodo
Joshi G. D., Pillepich A., Nelson D., Zinger E., Marinacci F., Springel V., Vogelsberger M., Hernquist L., 2021, MNRAS, 508, 1652
Kadowaki J., Zaritsky D., Donnerstein R. L., 2017, ApJ, 838, L21
Kass R. E., Raftery A. E., 1995, J. Am. Stat. Assoc., 90, 773
Kriek M., Conroy C., 2013, ApJ, 775, L16
Lee C. H., Hodges-Kluck E., Gallo E., 2020, MNRAS, 497, 2759
Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019, ApJ, 876, 3
Leja J., Johnson B. D., Conroy C., van Dokkum P. G., Byler N., 2017, ApJ, 837, 170
Liao S. et al., 2019, MNRAS, 490, 5182
Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, ApJ, 904, 33
Lupi A., Volonteri M., Silk J., 2017, MNRAS, 470, 1673
MacArthur L. A., González J. J., Courteau S., 2009, MNRAS, 395, 28
Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85
Maraston C., 2005, MNRAS, 362, 799
Maraston C., Thomas D., 2000, ApJ, 541, 126
Martín-Navarro I. et al., 2019, MNRAS, 484, 3425
Monaco L., Bellazzini M., Ferraro F. R., Pancino E., 2003, ApJ, 597, L25
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moultaka J., Boisson C., Joly M., Pelat D., 2004, A&A, 420, 459
Moultaka J., Pelat D., 2000, MNRAS, 314, 409
Mowla L., van Dokkum P., Merritt A., Abraham R., Yagi M., Koda J., 2017, ApJ, 851, 27
Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793
Ocvirk P., 2010, ApJ, 709, 88
Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006a, MNRAS, 365, 46
Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006b, MNRAS, 365, 74
Pandya V. et al., 2018, ApJ, 858, 29
Papastergis E., Adams E. a. K., Romanowsky A. J., 2017, A&A, 601, L10
Papovich C., Dickinson M., Ferguson H. C., 2001, ApJ, 559, 620
Paxton B. et al., 2013, ApJS, 208, 4
Paxton B. et al., 2015, ApJS, 220, 15
Paxton B. et al., 2018, ApJS, 234, 34

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3

Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21

Péroux C., Howk J. C., 2020, ARA&A, 58, 363

Polzin A., van Dokkum P., Danieli S., Greco J. P., Romanowsky A. J., 2021, ApJ, 914, L23

Renzini A., 2006, ARA&A, 44, 141

Rinaldi P., Caputi K. I., van Mierlo S., Ashby M. L. N., Caminha G. B., Iani E., 2022, ApJ, 930, 128

Rong Y., Guo Q., Gao L., Liao S., Xie L., Puzia T. H., Sun S., Pan J., 2017, MNRAS, 470, 4231

Ruiz-Lara T. et al., 2015, A&A, 583, A60

Ruiz-Lara T. et al., 2018, MNRAS, 478, 2034

Saifollahi T., Trujillo I., Beasley M. A., Peletier R. F., Knapen J. H., 2021, MNRAS, 502, 5921

Saifollahi T., Zaritsky D., Trujillo I., Peletier R. F., Knapen J. H., Amorisco N., Beasley M. A., Donnerstein R., 2022, MNRAS, 511, 4633

Sales L. V., Navarro J. F., Peñafiel L., Peng E. W., Lim S., Hernquist L., 2020, MNRAS, 494, 1848

Salvador-Rusiñol N., Vazdekis A., La Barbera F., Beasley M. A., Ferreras I., Negri A., Dalla Vecchia C., 2020, Nat. Astron., 4, 252

Sánchez-Blázquez P. et al., 2006, MNRAS, 371, 703

Sánchez-Blázquez P., Ocvirk P., Gibson B. K., Pérez I., Peletier R. F., 2011, MNRAS, 415, 709

Schaye J. et al., 2010, MNRAS, 402, 1536

Schiavon R. P., 2007, ApJS, 171, 146

Schiavon R. P., Rose J. A., Courteau S., MacArthur L. A., 2004, ApJ, 608, L33

Schiavon R. P., Rose J. A., Courteau S., MacArthur L. A., 2005, ApJS, 160, 163

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Serra P., Trager S. C., 2007, MNRAS, 374, 769

Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, AIP Conf. Ser. Vol. 735, Bayesian Inference and Maximum Entropy Methods inScience and Engineering. Am. Inst. Phys., New York, p. 395

Speagle J. S., 2020, MNRAS, 493, 3132

Tacchella S. et al., 2021, ApJ, 926, 36

The Astropy Collaboration et al., 2013, A&A, 558, A33

The Astropy Collaboration et al., 2018, AJ, 156, 123

Thomas D., Maraston C., Bender R., de Oliveira C. M., 2005, ApJ, 621, 673

Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS, 404, 1775

Trager S. C., Faber S. M., Worthey G., González J. J., 2000, AJ, 120, 165

Tremmel M., Wright A. C., Brooks A. M., Munshi F., Nagai D., Quinn T. R., 2020, MNRAS, 497, 2786

Trujillo-Gomez S., Reina-Campos M., Kruijssen J. M. D., 2019, MNRAS, 488, 3972

van de Schoot R. et al., 2021, Nat. Rev. Methods Primers, 1, 1

van der Hulst J. M., Terlouw J. P., Begeman K. G., Zwitser W., Roelfsema P. R., 1992, in Worrall D. M., Biemesderfer C., Barnes J., eds, ASP Conf. Ser. Vol. 25, Astronomical Data Analysis Software and Systems I. Astron. Soc. Pac., San Francisco, p. 131

van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22

van Dokkum P. et al., 2016, ApJ, 828, L6

van Dokkum P. et al., 2017, ApJ, 844, L11

van Dokkum P. et al., 2019, ApJ, 880, 91

van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, ApJ, 798, L45

Van Nest J. D., Munshi F., Wright A. C., Tremmel M., Brooks A. M., Nagai D., Quinn T., 2022, ApJ, 926, 92

Vazdekis A., 1999, ApJ, 513, 224

Villaume A. et al., 2022, ApJ, 924, 32

Virtanen P. et al., 2020, Nat. Methods, 17, 261

Wasserman A. et al., 2019, ApJ, 885, 155

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Weisz D. R. et al., 2011, ApJ, 739, 5

Werner M. W. et al., 2004, ApJS, 154, 1

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

Worthey G., 1994, ApJS, 95, 107

Wright A. C., Tremmel M., Brooks A. M., Munshi F., Nagai D., Sharma R. S., Quinn T. R., 2021, MNRAS, 502, 5370

Yagi M., Koda J., Komiyama Y., Yamanoi H., 2016, ApJS, 225, 11

Yozin C., Bekki K., 2015, MNRAS, 452, 937

APPENDIX A: THE SFH OF DF44

For comparison with future works, in Table A1, we provide the fraction of SF, and cumulative fraction of stellar mass formed, within the time bins of the non-parametric models. We list the 16th, 50th, and 84th percentiles of the distributions, where we note that the 50th percentiles of the fractional SFHs do not necessarily sum to unity.

Table A1. Summary of SFH results.

Time bin (Gyr)	Extended SFH prior						Concentrated SFH prior					
	SF fraction			Cumulative fraction of M_*			SF fraction			Cumulative fraction of M_*		
	16	50	84	16	50	84	16	50	84	16	50	84
10^{-9}–0.03	0.0703	0.0795	0.0967	1.0000	1.0000	1.0000	0.0042	0.0082	0.0099	1.0000	1.0000	1.0000
0.03–0.10	0.0050	0.0152	0.0240	0.9986	0.9988	0.9989	0.0160	0.0181	0.0221	0.9995	0.9996	0.9998
0.10–0.50	0.0007	0.0014	0.0061	0.9980	0.9983	0.9986	0.0000	0.0000	0.0004	0.9974	0.9977	0.9980
0.50–1.00	0.0014	0.0031	0.0049	0.9972	0.9979	0.9983	0.0000	0.0001	0.0002	0.9974	0.9975	0.9979
1.00–2.00	0.0003	0.0015	0.0039	0.9964	0.9970	0.9975	0.0000	0.0000	0.0001	0.9973	0.9974	0.9978
2.00–3.00	0.0014	0.0046	0.0086	0.9949	0.9962	0.9970	0.0000	0.0000	0.0001	0.9971	0.9974	0.9978
3.00–4.01	0.0011	0.0089	0.0158	0.9917	0.9932	0.9955	0.0000	0.0000	0.0005	0.9969	0.9973	0.9977
4.01–5.36	0.0091	0.0145	0.0435	0.9833	0.9886	0.9925	0.0000	0.0001	0.0004	0.9964	0.9971	0.9973
5.36–7.16	0.0166	0.0702	0.1351	0.9591	0.9790	0.9847	0.0000	0.0002	0.0016	0.9943	0.9969	0.9971
7.16–9.57	0.0722	0.1548	0.3310	0.8543	0.8993	0.9606	0.0000	0.0000	0.0022	0.9917	0.9950	0.9969
9.57–12.80	0.3036	0.3726	0.5208	0.5596	0.6872	0.8085	0.0000	0.0000	0.0008	0.9851	0.9938	0.9958
12.80–13.47	0.0583	0.1918	0.3111	0.0179	0.0618	0.1116	0.9678	0.9714	0.9737	0.9822	0.9916	0.9947

Notes. The fraction of SF and the cumulative fraction of stellar mass formed are listed for each time bin of the non-parametric SFH model. The 16th, 50th, and 84th percentiles of the posterior (i.e. the 68 per cent CR) are listed. We note that the 50th percentiles of the fractional SFH do not necessarily sum to unity. The SF time-scales listed in Table 3 are interpolated from these step functions.

APPENDIX B: SYSTEMATIC BIASES IN MEASURING SFHS

B1 SFH biased by blue horizontal branch stars

The use of integrated light to reconstruct stellar populations has the caveat that multiple types of stars can share spectral signatures. This is the case for young, massive main-sequence stars and old, metal-poor stars on the blue side of the horizontal branch (HB); both act to amplify the equivalent width of the Balmer lines. A population of blue HB stars produces a flux shortward of 3000 Å, which increases with decreasing metallicity due to a hotter main-sequence turn-off. Neglecting to include a blue HB population in

models can lead to predictions of unrealistically young ages (e.g. Worthey 1994; Schiavon et al. 2004; Thomas et al. 2005; Schiavon 2007). The difficulty of distinguishing between these two stellar populations has been noted in GCs and dwarf galaxies (e.g. Monaco et al. 2003; Schiavon et al. 2004; Conroy et al. 2018; Cabrera-Ziri & Conroy 2022), as well as in elliptical galaxies (Maraston & Thomas 2000).

In the SFH fit to DF44 (for which the primary age indicator is the H β absorption line), we see a rise in the SFR in the two most recent time bins corresponding to the last 100 Myr (by $1.8^{+0.2}_{-0.4}$ dex for the extended SFH, $2.5^{+1.5}_{-1.1}$ dex for the concentrated SFH) – yet, there are no corresponding emission lines to suggest the presence of a young stellar population. Given the low metallicity of this UDG
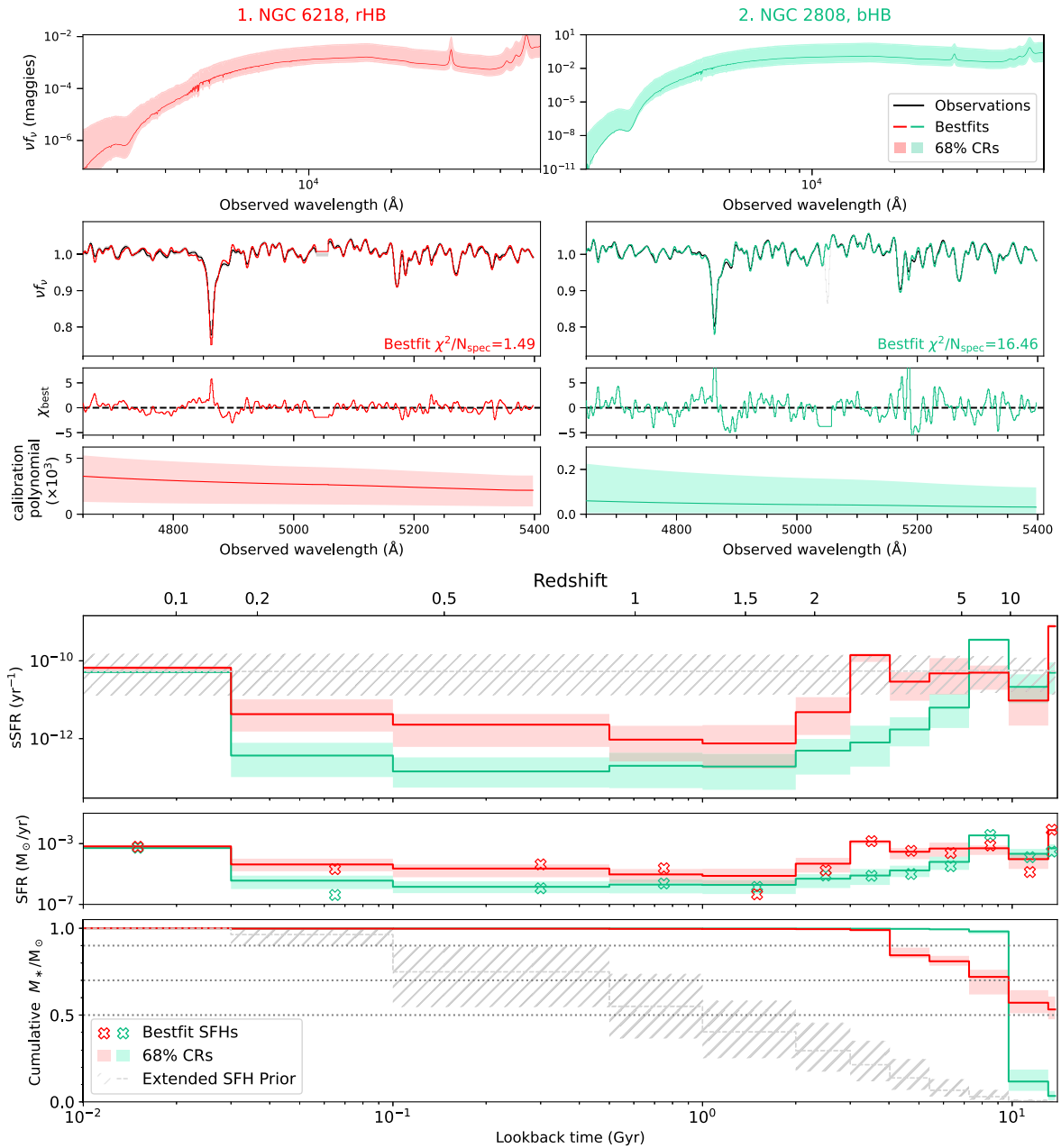

Figure B1. Summary of the fitting results for two Milky Way globular clusters selected to have a similar metallicity as DF44: NGC 6362 (redder HB) and NGC 2808 (bluer HB). Top panel: fits to observations, similar to Fig. 3 for DF44. Bottom panel: posterior distributions for SF and mass growth, similar to Fig. 4 for DF44. Both clusters are fitted with a fixed mass, $\log(M_*/M_\odot) = 8$.
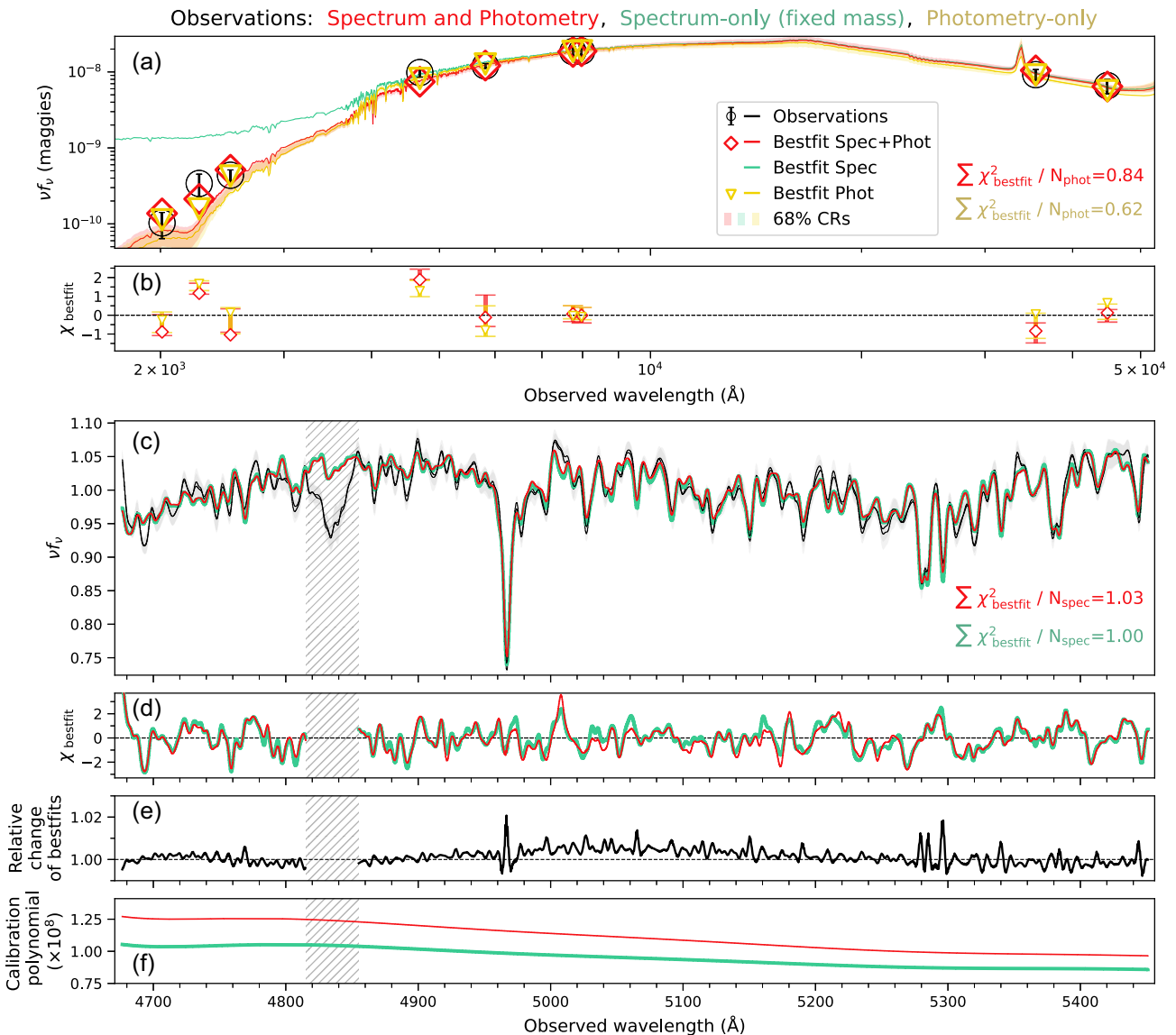
Figure B2. Comparison of fits with the spectrum and photometry (red), spectrum only (green, with mass fixed to the value from van Dokkum et al. 2016), and photometry only (yellow), assuming an extended SFH prior. The observed data (black) are compared to the best-fitting models (coloured lines) and the 68 per cent CR of 500 randomly drawn models from the posteriors (shaded coloured regions). The corresponding posteriors are shown in Fig. B3. Panel (a): the observed (circles) and best-fitting (diamonds and triangles) photometric points, where the reduced χ^2/N_{data} of the best-fitting SED are listed. Panel (b): the χ ([data − model]/σ) of the best-fitting photometric points. Panel (c): the observed spectrum (uncertainties shown in grey) and best-fitting spectra (multiplied by the spectrophotometric calibration polynomial). The hatched grey region indicates the spectral region masked throughout the fitting process. Panel (d): the χ of the best-fitting spectra as a function of wavelength. Panel (e): the relative change of the best-fitting models, i.e. the ratio of the two best-fitting spectra. Panel (f): the spectrophotometric calibration models, with 68 per cent CRs shown as shaded regions.

($\log(Z_*/Z_\odot) \sim -1.2$), the presence of a blue HB population would not be unexpected.

While the bias between the age and the blue HB stars is well known when fitting SSPs (Conroy et al. 2018), it is not yet well studied for non-parametric SFHs. Ocvirk (2010) provided a first look at the impact of blue HB stars on linear combinations of SSP models, finding that the presence of blue HB stars can be inferred as a recent burst of star formation (SF) at ∼100 Myr, contributing less than around 10 per cent of the total stellar mass. While this provides a promising explanation for the apparent SF bursts we observe in the SFH of DF44, we follow a similar test using the non-parametric SFH described in Section 3.1.

In order to investigate if our SFHs are affected by the presence of blue HB stars that mimic a burst of SF within the last 100 Myr, we fit the SFHs of two Galactic GCs, one with a known blue HB and the other without. We select the GCs from Schiavon et al. (2005), with metallicities similar to DF44: NGC 2808 has [Fe/H] =−1.29 and $(B-R)/(B+V+R) = -0.49$ (bluer HB), and NGC 6218 (M12) has [Fe/H] =−1.32 and $(B-R)/(B+V+R) = 0.97$ (redder HB). Given that GCs are reasonable approximations of SSPs, we expect an early single burst of SF only. Fits to the spectra[21] of NGC 2808

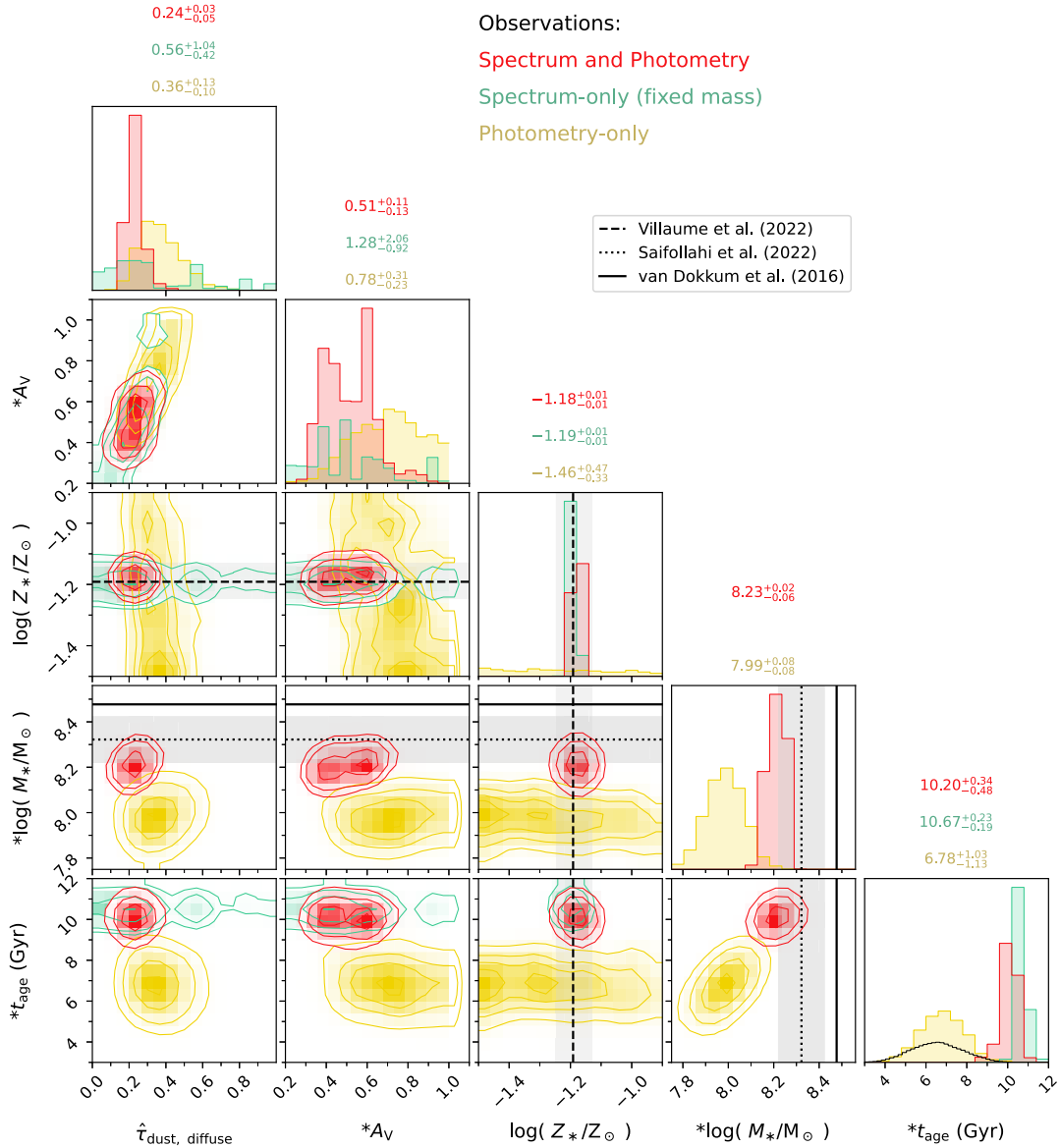---

[21]Downloaded from http://www.noao.edu/ggclib.

Figure B3. Comparison of posterior pdfs derived with the photometry-only (yellow), spectrum-only (green, with mass fixed to the value from van Dokkum et al. 2016), and both spectrum and photometry (red), assuming an extended SFH prior. Fits are shown in Fig. B2. Posteriors of selected fitted and derived (marked with asterisks) parameters are shown. Contours are shown smoothed with a $n = 1$ Gaussian kernel. Black lines denote the expected results from the literature: stellar metallicity from Villaume et al. (2022), and stellar mass from van Dokkum et al. (2016) and Saifollahi et al. (2022). The median and uncertainties from the 68 per cent CR are listed along the top of the one-dimensional histograms. The implicit age prior is shown as a black histogram for reference.

and NGC 6218, over the same wavelength range as DF44, following the procedure described in Section 3, are shown in Fig. B1. The top panels summarize the comparison between the observations (black) lines, and models (coloured lines). Similar to Fig. 4, the bottom panels show the sSFR, SFR, and mass assembly histories. An extended SFH was assumed, and the total stellar mass was fixed to $10^8\,M_\odot$.

In both cases, we find an increase in the SFR within the last 100 Myr, although to a larger extent for the GC with the blue HB stars (by $1.5^{+0.5}_{-0.3}$ dex for NGC 6218 and $2.6^{+0.4}_{-0.4}$ dex for NGC 2808). In addition, we see that both SFHs are early and short-lived, although there are modest levels of SF at >2 Gyr, which likely results from the models being unable to precisely match the high S/N spectra ($\sim$180 and $\sim$480, respectively). We conclude from this comparison that some component of the recent SF burst we measure for DF44 could plausibly be related to a population of blue HB stars.

B2 Fitting the spectroscopy and photometry together versus separately

Figs B2 and B3 show the results of fitting the models to observations of DF44, where we include the following input: (i) using only the photometry (yellow), (ii) using only the spectroscopy (green), and (iii) using both the photometry and spectroscopy (red), and assuming an extended SFH prior. We note that the stellar mass is fixed (to the value reported by van Dokkum et al. 2016) for the spectrum-only fit as the continuum was subtracted from the spectrum.

Similar to Fig. 3 discussed in Section 4, in Fig. B2, the observations (black lines and markers) are shown relative to the best-fitting models (coloured lines and markers, where the colours denote which observations were fit). Shaded coloured regions indicate the
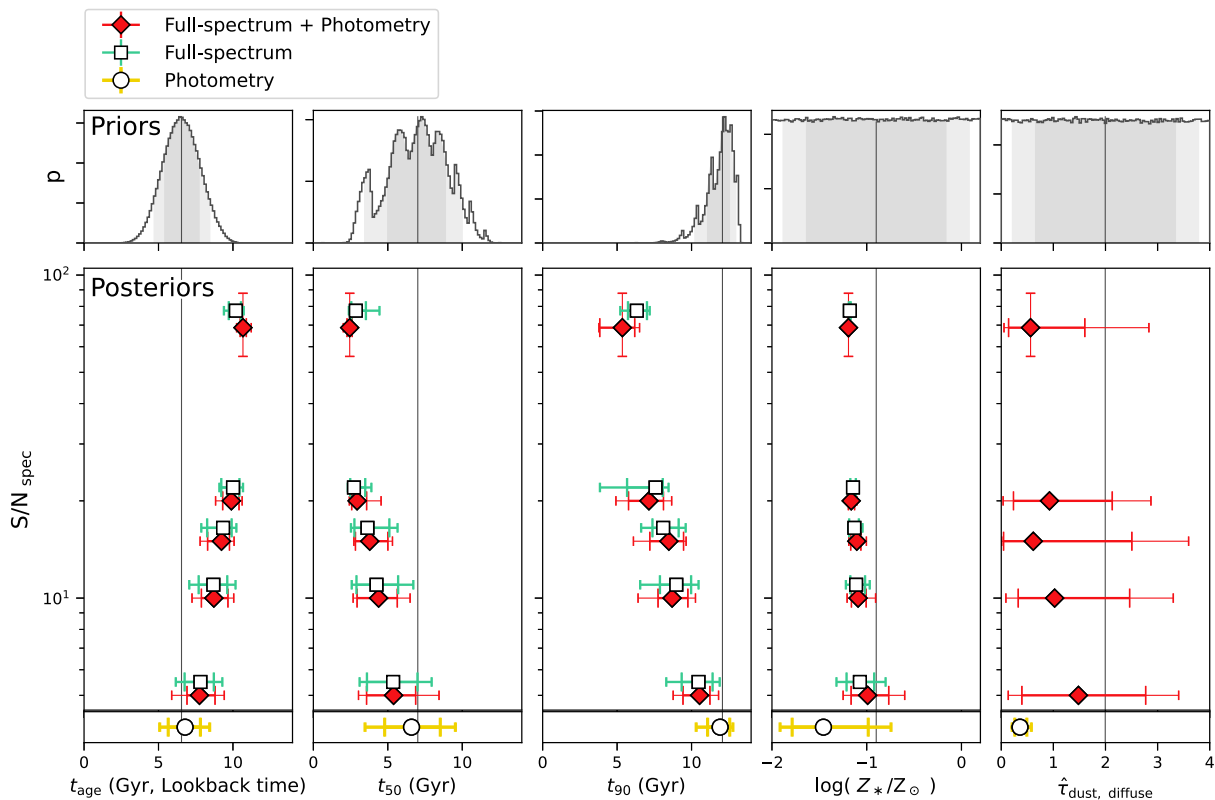
Figure B4. Using the extended SFH prior, we compare the priors (top panels) and posterior parameter estimates (bottom panels) as a function of S/N. Values determined from full-spectrum fits with (diamonds) and without (squares; vertically offset for clarity) including the photometry are shown, as well as from photometry alone (circles). Points are shown with error bars corresponding to the 68 (thick and wide) and 95 per cent (thin and narrow) percentiles of the posteriors. Similarly, for the prior, we show the median with a grey line, and 68 and 95 per cent CRs with shaded regions. While mass-weighted ages are shown in units of lookback time, the time-scales t_{50} and t_{90} have units of time since the big bang ($t = 0$ is the big bang).

68 per cent CRs from sampling the posterior pdfs, where the grey shaded region indicates the uncertainties in the spectrum.

Both best-fitting SED models match the photometry with reasonable $\chi_{\rm bestfit}$. In comparison, the UV flux is significantly overestimated when fitting only the spectroscopy. Since the UV provides information about recent SF, and the UV to optical colours constrain the dust attenuation, we do not expect to constrain these properties from the spectrum alone.

A comparison of the observed spectrum with the best-fitting models is also shown in Fig. B2, with the $\chi_{\rm bestfit}$ as a function of wavelength, and the spectrophotometric calibration polynomial (see Section 3.3). The ratio of the two best-fitting models, shown flattened by dividing through by a polynomial, shows that the fits are similar at the 2 per cent level. The only notable differences between the two best-fitting models are around the Hβ line and Mg II features at $\sim$5285–5305 Å (observed-frame). The positive ratio of the Hβ line between the spectrum-and-photometry fit over the spectrum-only fit is consistent with the UV flux being constrained for the former, such that the absorption line is preferentially shallower. The difference in the Mg II lines reflects the difference in metallicities predicted for each fit, as well as the inability of the (fixed scaled-solar abundance) models to be flexible to such features.

Fig. B3 compares the basic stellar properties (normalization of the diffuse dust attenuation curve, *V*-band extinction, stellar metallicity, stellar mass, and mass-weighted age) for the fits to the three sets of observations. This figure is akin to Fig. 5, discussed in Section 4. For comparison, black lines indicate values measured in the literature:

dashed lines indicate the stellar isochrone metallicity measured by Villaume et al. (2022), while dotted and solid lines indicate the stellar mass measured by van Dokkum et al. (2016) and Saifollahi et al. (2021), respectively. For reference, the prior on the age (which is implicit, as the age is determined by the time bin widths and SFH) is shown as a black histogram.

The broad-band NUV to NIR photometry (yellow) and continuum normalized spectroscopy (green) carry different information about the galaxy properties. The broad (yet coarse) photometry provides a tighter constraint on the dust attenuation, while the spectroscopy constrains the metallicity. The dust attenuation cannot be determined from the spectroscopy alone because of the lack of continuum information; the spectrophotometric calibration marginalizes over the continuum shape, and is degenerate with both the stellar mass and dust attenuation. On the other hand, the metallicity is tightly constrained by the spectroscopy as there is detailed information among the numerous absorption lines.

Despite the formal consistency of the dust and metallicity parameters between these two fits (given the large uncertainties), the age posteriors are significantly different. The age posterior from the photometry largely traces the (implicit) prior. A tighter pdf for the stellar metallicity provides a more precise estimate of the age, as expected given the degeneracy between these two parameters.

Simultaneously fitting the photometry and spectroscopy (shown in red) constrains the full set of parameters. In the particular case of DF44, the results are largely informed by the spectroscopy, which covers a broad range in metallicity and age features – the inclusion
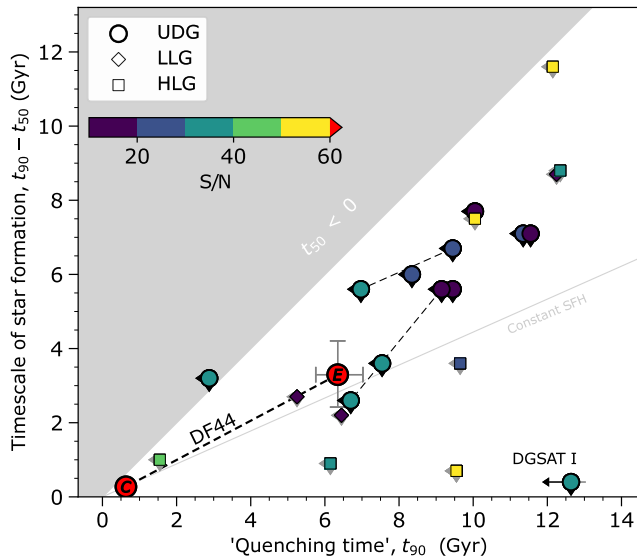
Figure B5. SF time-scales of UDGs (circles) and dwarfs (LLGs and HLGs; squares and diamonds, respectively) for observations from the literature. We approximate the quenching time as when 90 per cent of the stellar mass is in place (t_{90}), while the time-scale $t_{90} - t_{50}$ gives a sense of the duration of SF, i.e. how concentrated/extended the SFH is. Other than DF44, we show the points from observations with arrows indicating that they are upper limits (see text). Points are coloured according to their S/N, where DF44 has a mean S/N of 96 Å^{-1}. Dashed lines connect points measured for the same object, but from different studies. Data shown are taken from the published values of: Ferré-Mateu et al. (2018), Ruiz-Lara et al. (2018), and Martín-Navarro et al. (2019). The points from Ruiz-Lara et al. (2018) are shown with S/N = 32 Å^{-1}, the median of the reported range in values.

of the photometry only modestly affects the posteriors. The stellar mass derived from the combined data sets is consistent with that of Saifollahi et al. (2022), while the photometry-derived posterior is skewed lower by ∼0.23 dex, which is likely also related to the lower estimate for the stellar metallicity. The combined result shows DF44 to be very old, metal-poor, and perhaps with some small amount of dust.

B3 SFH biased by choice of prior

Fig. B4 demonstrates the S/N dependence of the bias imposed by the choice of SFH prior, which, in this case, is an extended SFH in describing a very old stellar population. We refit the KCWI spectrum of DF44 with the extended SFH prior ($\alpha_D = 1$), successively increasing the uncertainties of the spectrum such that the S/N$_{spec}$ = 5, 10, 15, and 20. The medians of the recovered pdfs are shown for the mass-weighted age (in lookback time), t_{50} and t_{90} (in time since the big bang), $\log(Z_*/Z_\odot)$, and diffuse dust, with error bars corresponding to the 68 (thick and wide) and 95 per cent (thin and narrow) CRs. Points mark the results from fitting the spectrum and photometry simultaneously (diamonds), the spectrum alone (squares, offset vertically for clarity), and the photometry alone (circles). The prior distributions are shown in the top panels. Note that because the implicit priors for the SFH time-scales depend on the widths of the SFH time bins (a step function), the distributions are not necessarily smooth.

The SFH time-scales are more heavily weighted by the SFH prior at low S/N. This is particularly true for t_{90}, which we use as a proxy of the quenching time. In contrast, neither the stellar metallicity nor the dust is significantly biased, or at least the offsets are well within

the (large) uncertainties. While having a complete set of observations informs many of the galaxy properties, the choice of a 'good' SFH prior is important.

B4 Comparing results between studies – prior and data dependence

Fig. B5 shows a comparison of the SF time-scales of UDGs (circles) and dwarfs (squares and diamonds) for observations from the literature (for Coma galaxies in almost all cases). We compare the time at which we consider the galaxy quenched, t_{90}, with how extended the SFH is, $t_{50} - t_{90}$. The grey shaded region denotes the parameter space where ages (t_{50}) are older than the Universe (e.g. OGS1 from Ruiz-Lara et al. 2018). We show the results from the literature as upper limits given the possible biases in SFH time-scales discussed above related to the S/N, and choice of SFH priors.

Except for DF44, all the literature values were measured using the full-spectrum fitting code STECKMAP. Notably, STECKMAP smooths the SFHs via (tuneable) regularization akin to Gaussian priors on the SFH and age–metallicity relations (see the discussion in Section 3.1.1). The details of the regularization differ between all studies, where, for example, Ruiz-Lara et al. (2015) present the outcome of averaging several results with various smoothing parameters. Martín-Navarro et al. (2019) show in their appendix A the difference in their regularized and unregularized results to be ∼1 Gyr in t_{50} and ≲ 0.4 Gyr in t_{90}.

Ferré-Mateu et al. (2018) compared their SFH time-scales derived from STECKMAP with those from an alternative fitting code, STARLIGHT, which does not impose regularization but does require relative-flux-calibrated spectra. Between the two fitting approaches, Ferré-Mateu et al. (2018) found consistent results in that the SFHs are extended and had similar quenching times. That said, STARLIGHT preferred starting SF ∼2 Gyr later, such that the ages were younger and star-forming time-scales were shorter. In contrast, the 'burstier' prior used in this work produced earlier SF and quenching.

Because of the difficulties in determining the ages of old stellar populations, even subtle differences in data or analysis can impact results beyond the expected uncertainties. As an example, we can compare measurements for two UDGs, DF26/Yagi93 and Yagi418, both studied by Ferré-Mateu et al. (2018) and Ruiz-Lara et al. (2018); the values are connected with dashed lines in Fig. B5. Each author used rest-frame optical spectroscopy (where Ruiz-Lara et al. 2018 reported a higher S/N and had a wider wavelength coverage) and used the same code (STECKMAP). However, the median mass-weighted ages differ by ∼1 Gyr (uncertainties were not reported, but the luminosity-weighted ages are formally consistent). In both cases, the higher S/N data provided a solution shifted in the expected direction (i.e. towards older and less-extended SFHs).

While DF44 appears to have (one of) the shortest SFHs and earliest quenching times, we caution that a detailed comparison should consider priors and the S/N. A poorly chosen SFH prior will have a stronger bias at a low S/N. For example, in using an extended SFH prior with the DF44 KCWI spectrum degraded to S/N = 20, we recover $t_{50} \sim 2.9 \pm 0.5$ Gyr and $t_{90} \sim 7.1 \pm 1.2$ Gyr (see Fig. B4 in Appendix B3), which overlaps with the lower end of UDGs in Fig. B5. This suggests that some of these objects could be older, and have less-extended SFHs.

Along the same lines, we do not include photometry-derived results in Fig. B5 as the comparison can be misleading, given the different choices (and relative contributions) of SFH priors. In the preceding sections, we have shown that the photometry-derived ages are younger than the spectroscopy- or combined-derived ages. There

is a similar difference between the results of Pandya et al. (2018, with optical to NIR photometry; not shown in Fig. 8) and Martín-Navarro et al. (2019, with rest-frame optical spectroscopy, S/N ∼ 10 Å⁻¹). Both studied the UDG DGSAT I, although using different fitting methods and assuming different SFHs. Pandya et al. (2018) fitted their photometry (via MCMC) to a delayed-exponential model, while Martín-Navarro et al. (2019) fitted their spectroscopy with STECKMAP. We note that in this example the priors are considerably different. For a delayed exponential model with linearly uniform priors with $\tau = 0.1$–10 and $t_{age} = 0.1$–14 Gyr, the implicit prior on the mass-weighted age has a median of 3.2 Gyr. In comparison, a constant SFH has a median age of half the age of the Universe, ∼6.8 Gyr (see also the discussion in Johnson et al. 2021b). While the luminosity-weighted ages are similar (∼3 Gyr), their mass-weighted ages are discrepant by >1 Gyr (t_{age} in the delayed-exponential model is the onset of SF, where for a $\tau > 3$, this corresponds to ages considerably younger than t_{age}). The metallicities are also discrepant by >1 dex, although Martín-Navarro et al. (2019) found that DGSAT I is unusually α-enhanced. Several other studies have studied UDGs from photometry alone (e.g. Greco et al. 2018; Barbosa et al. 2020), and have similarly noted younger ages than spectroscopy-derived results.

We additionally note that Martín-Navarro et al. (2019) uses a set of SSP models different than used in both this work and the other UDGs studies discussed here. Neither the choice of SSP models

or application of regularization would explain the significant offset between the SFHs of DGSAT I and the other UDGs, however.

APPENDIX C: DEGENERACY BETWEEN DUST ATTENUATION AND FLUX FROM OLD STELLAR POPULATIONS IN THE NUV

The normalization of the dust attenuation curve ($\hat{\tau}_{dust,\ diffuse}$) and the fraction of old stars, both parameters of our physical model, are degenerate at optical and UV wavelengths. As a brief example of this degeneracy, Fig. C1 shows the photometry for DF44 (black points) relative to three model SEDs with SSPs (i.e. not the results of fitting the physical model described in Section 3). Taking the grey model as the 'fiducial' model, slight variations in the age and dust are shown by the purple and cyan models, respectively. While the 2.8-Gyr age increase or 0.2-dex increase in diffuse dust produces an equivalent effect in the NUV, they have opposing effects at wavelengths >1 μm. Coloured markers show the expected photometry in two *JWST* filters in the mid-infrared, with S/N ∼ 5 to reflect the average uncertainty of the IRAC data. In this example, the 'old' and 'dusty' models are slightly distinguishable in *F560W* ($\Delta m_{AB} \sim 0.6\sigma_m$) but very different in *F770W* ($\Delta m_{AB} \sim 3\sigma_m$). The inclusion of mid-infrared data to our data set would allow us to assess whether DF44 is as dusty as our results suggest or a product of the complex degeneracies between physical parameters (see Section 4.1).
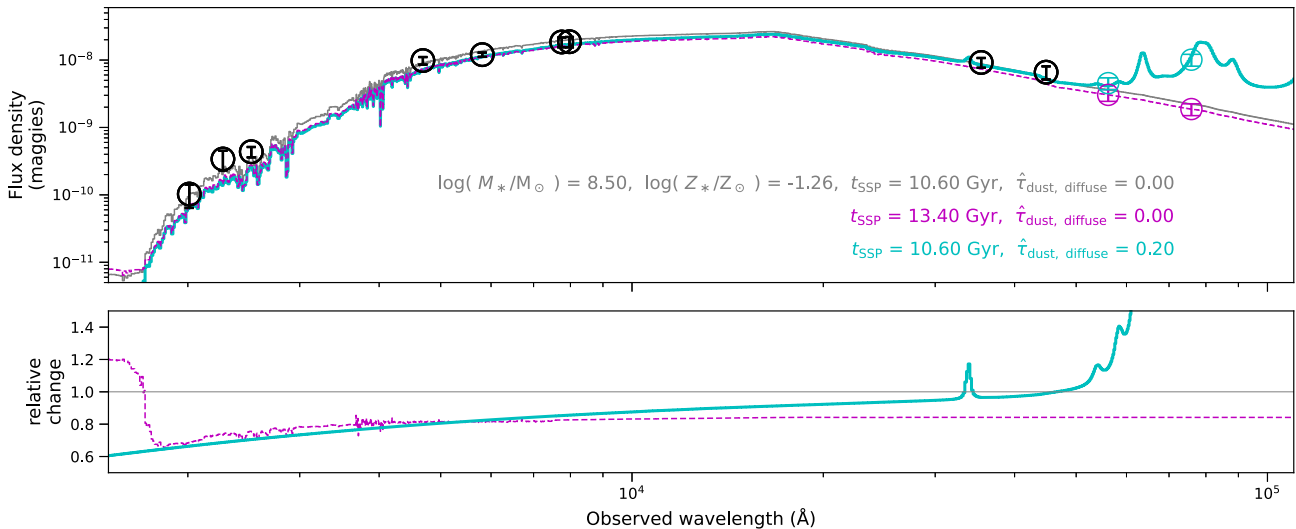

Figure C1. A brief demonstration of the degeneracy between dust attenuation and age on the shape of SEDs. Top panel: photometry of DF44 (black markers) and models (coloured lines) for three SSP populations. Photometric points corresponding to the 'old' (purple dashed) and 'dusty' (solid cyan) models are shown as measured by the *JWST F560W* and *F770W* filters (coloured markers), with S/N ∼ 5. Bottom panel: the relative change between the fiducial (grey) and older or dustier models. While the effect of either increasing the age or dust acts similarly at wavelengths <1 μm, the effect acts in the opposite sense in the mid-infrared.

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