Bayesian hierarchical modelling of the M_*–SFR relation from $1 \lesssim z \lesssim 6$ in ASTRODEEP

L. Sandles,[1,2]★ E. Curtis-Lake,[3] S. Charlot,[4] J. Chevallard [5] and R. Maiolino[1,2,6]

[1]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[2]*Cavendish Laboratory Astrophysics Group, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[3]*Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK*
[4] *Institut d'Astrophysique de Paris, Sorbonne Université, CNRS, UMR7095, F-75014 Paris, France*
[5]*Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[6]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

ABSTRACT

The Hubble Frontier Fields represent the opportunity to probe the high-redshift evolution of the main sequence of star-forming galaxies to lower masses than possible in blank fields thanks to foreground lensing of massive galaxy clusters. We use the BEAGLE SED-fitting code to derive stellar masses, $M_\star = \log(M/M_\odot)$, SFRs, $\Psi = \log(\psi/M_\odot \, \mathrm{yr}^{-1})$, and redshifts from galaxies within the ASTRODEEP catalogue. We fit a fully Bayesian hierarchical model of the main sequence over $1.25 < z < 6$ of the form $\Psi = \alpha_{9.7}(z) + \beta(M_\star - 9.7) + \mathcal{N}(0, \sigma^2)$ while explicitly modelling the outlier distribution. The redshift-dependent intercept at $M_\star = 9.7$ is parametrized as $\alpha_{9.7}(z) = \log[N(1+z)^\gamma] + 0.7$. Our results agree with an increase in normalization of the main sequence to high redshifts that follows the redshift-dependent rate of accretion of gas on to dark matter haloes with $\gamma = 2.40^{+0.18}_{-0.18}$. We measure a slope and intrinsic scatter of $\beta = 0.79^{+0.03}_{-0.04}$ and $\sigma = 0.26^{+0.02}_{-0.02}$. We find that the sampling of the SED provided by the combination of filters (*Hubble* + ground-based Ks-band + *Spitzer* 3.6 and 4.5 μm) is insufficient to constrain $M_\star$ and Ψ over the full dynamic range of the observed main sequence, even at the lowest redshifts studied. While this filter set represents the best current sampling of high-redshift galaxy SEDs out to $z > 3$, measurements of the main sequence to low masses and high redshifts still strongly depend on priors employed in SED fitting (as well as other fitting assumptions). Future data sets with *JWST* should improve this.

Key words: methods: data analysis – methods: statistical – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: star formation.

1 INTRODUCTION

The relationship between star formation rate (SFR) and stellar mass of 'normal' star-forming galaxies has been well-studied and is often referred to as the 'star-forming main sequence' (originally labelled as such by Noeske et al. 2007). For masses less than $\log(M/M_\odot) \lesssim 10.1$, the main sequence is commonly parametrized as a straight line while at higher masses there is evidence for a redshift-dependent turn-over (Whitaker et al. 2014; Lee et al. 2015; Schreiber et al. 2015; Tasca et al. 2015; Tomczak et al. 2016; Leslie et al. 2020; Leja et al. 2021). ALMA observations suggest that the resolved main sequence is a by-product of two more physically connected relations; that between stellar mass and molecular gas densities, and that between the molecular gas and SFR densities (Lin et al. 2019; Baker et al. 2022). However, direct measurements of the molecular gas reservoir are unfeasible for large samples at high redshifts, and measurements of the main sequence remain relevant as we move into the *James Webb Space Telescope* (*JWST*) era.

Speagle et al. (2014) provide a thorough review of a compilation of 25 studies of the star-forming main sequence. They show that many of the discrepancies between measurements of slope and normalization can be resolved once two primary issues have been corrected for: the method chosen to select star-forming galaxies and the method used to calculate SFR (e.g. from emission lines, rest-frame ultra-violet continuum, spectral-energy distribution fitting). Having calibrated the results within the literature, Speagle et al. (2014) report that both the slope ($\sim 0.4-0.8$) and normalization (~ 2 orders of magnitude) increase from redshift 0 to 4, whilst the intrinsic scatter remains relatively constant (~ 0.2 dex).

In recent years, much work has been done to constrain the star-forming main sequence at higher redshifts (Steinhardt et al. 2014; Salmon et al. 2015; Santini et al. 2017; Pearson et al. 2018; Thorne et al. 2021; Bhatawdekar & Conselice 2021). Steinhardt et al. (2014) show that for massive galaxies ($> 10^{10} \, M_\odot$) the main sequence extends to at least $z = 6$. Salmon et al. (2015) use multiwavelength photometry to determine an almost constant main-sequence relation, though with mildly increasing normalization, between $3.5 < z < 6.5$. They study samples chosen at constant number density spanning the redshift range to link progenitor galaxies, finding evidence for rising star formation histories (SFHs) in these objects. Bhatawdekar &

Conselice ([2021](#)) push the redshift boundary even further providing evidence of a main sequence between $6 < z < 9$.

There has also been significant efforts to constrain the lower mass end of the main sequence. Tasca et al. ([2015](#)) analyse a sample of star-forming galaxies from the VIMOS (VIsible Multi-Object Spectrograph) Ultra-Deep Survey (VUDS; Le Fèvre et al. [2015](#)) with confirmed spectroscopic redshifts ranging from $0 < z < 6$. Their results confirm that the main sequence extends to masses as low as $10^7 \, M_\odot$ for $0.0 < z < 0.7$. Boogaard et al. ([2018](#)) use the deepest MUSE (Multi Unit Spectroscopic Explorer) observations of the Hubble Ultra Deep Field and the Hubble Deep Field South to similarly constrain the low mass end of the main sequence for redshifts $0.11 < z < 0.91$. Santini et al. ([2017](#)) exploit the gravitational lensing of large foreground clusters to probe the main sequence to masses as low as $10^{7.5} \, M_\odot$ for $z < 4$ and $10^{8.0} \, M_\odot$ for $4 < z < 6$.

The specific SFR (sSFR) is defined as SFR divided by stellar mass and gives a measure of the current star formation activity compared to the integrated past history. At a fixed mass, sSFR is analogous to the normalization of the main sequence. If the SFR closely follows the evolution of the mass accretion rate on to parent haloes, the sSFR will be expected to vary with redshift as $\propto (1 + z)^{2.25}$ (Birnboim, Dekel & Neistein [2007](#); Neistein & Dekel [2008](#); Dekel et al. [2009](#); Fakhouri, Ma & Boylan-Kolchin [2010](#)). The semi-analytic model of Dutton, van den Bosch & Dekel ([2010](#)) predicts such evolution in sSFR, as do hydrodynamical simulations (Furlong et al. [2015](#); Donnari et al. [2019a](#)). For $z \gtrsim 3$, observational studies appear to agree with the predictions (Koprowski et al. [2014](#), [2016](#); Mármol-Queraltó et al. [2016](#); Santini et al. [2017](#)).

The recent work of Leja et al. ([2021](#)) provides a new framework to derive the main sequence from the density of objects in the mass-SFR plane. They fit to objects in the 3D-*HST* (Skelton et al. [2014](#)) and COSMOS-2015 (Laigle et al. [2016](#)) catalogues with a non-parametric star formation history (Leja et al. [2019a](#)), finding lower normalization of the main sequence by $\sim 0.2-0.5$ dex over $0.2 < z < 3$. This lower normalization resolves a tension between observations and cosmological simulations such as EAGLE (Furlong et al. [2015](#)) and Illustris-TNG (Donnari et al. [2019a](#)).

Speagle et al. ([2014](#)) and Katsianis et al. ([2020](#)) have demonstrated how sensitive the determination of the main sequence is to the measurement of SFR, while Leja et al. ([2021](#)) demonstrates how sensitive it can be to the chosen SFH. The latest SED fitting codes (e.g. MAGPHYS, da Cunha, Charlot & Elbaz [2008](#); BEAGLE, Chevallard & Charlot [2016](#); PROSPECTOR, Leja et al. [2017](#); Johnson et al. [2021](#); CIGALE, Boquien et al. [2019](#); Yang et al. [2020](#); BAGPIPES, Carnall et al. [2018](#); BAYESED, Han & Han [2012](#), [2014](#), [2019](#); Dense Basis Iyer & Gawiser [2017](#); Iyer et al. [2019](#); PROSPECT, Robotham et al. [2020](#))[1] are able to constrain a variety of physical parameters per galaxy including SFHs, dust attenuation, metallicities, and nebular emission, all of which can have a large impact on the derived masses and SFRs. However, the level at which certain properties can be constrained is sensitively dependent on the available data set, as demonstrated in Curtis-Lake et al. ([2021](#), hereafter CL21). They find that the emission-line contribution to rest-frame optical broad-band photometry at high redshifts ($z \sim 5$ in that study) leads to poorly constrained, biased SFR and stellar mass estimates whereas medium-band filters can significantly improve the constraints. Current data sets probing high redshifts do not have access to medium-band filters

spanning the rest-frame optical. In fact, beyond $z \sim 4$, there are only two main filters probing the rest-frame optical; the 3.6 and 4.5-μm bands of the *Spitzer* space telescope.

One primary advantage of the latest SED fitting codes is the derivation of robust uncertainties on the derived parameters. However, incorporating these complex, often co-varying uncertainties in population-wide studies require methods beyond standard linear regression which some studies have been addressing. Kurczynski et al. ([2016](#)) perform sigma-clipping to determine what objects are on the main sequence. They account for co-varying uncertainties by modelling the mass-SFR constraints as single, bivariate Gaussians while fitting to the main sequence in redshift bins. Boogaard et al. ([2018](#)) fit a hyperplane in stellar mass, SFR and redshift, self-consistently taking account of the uncertainties using the method of Robotham & Obreschkow ([2015](#)), which models a Gaussian scatter perpendicular to the main sequence. Their sample consists of emission-line selected galaxies from a MUSE survey, so star-forming galaxy selection is based on emission-line properties. Santini et al. ([2017](#)) and Pearson et al. ([2018](#)) forward model the main sequence before comparing to observations within redshift bins. Leja et al. ([2021](#)) use an innovative normalizing flow to measure the density in mass-SFR-redshift, defining the main sequence as the ridge in this space in order to avoid parametrizing the main sequence and outlier distributions separately. They sample from the individual object posterior probability distributions to marginalize over the uncertainties in mass and SFR.

CL21 suggest a Bayesian hierarchical method to model the main sequence. With this work, we extend their approach to include redshift dependence as well as an explicit model to account for outliers. We re-visit the Hubble Frontier Fields, studied by Santini et al. ([2017](#)), using the ASTRODEEP catalogues to probe to lower masses and higher redshifts than achievable in blank fields, in order to provide constraints on the low-mass end of the main sequence over a wide redshift range from a consistent data set. We re-visit this data set with self-consistent SFR and mass constraints derived with BEAGLE and fit a fully Bayesian hierarchical, redshift-dependent model of the low-mass, linear portion of the main sequence. We investigate the limitations of this data set with respect to constraining mass and SFR of individual galaxies with BEAGLE, demonstrating how to determine when these constraints are robust and how poor constraints can impact the measurements of the main sequence. This data set represents the best achievable sampling of galaxy SEDs at very high redshifts ($z \gtrsim 3$) before we have data from *JWST*. In this sense, it provides a representative view of the limitations of what we can measure currently. This study will aid in the understanding of any differing constraints derived with *JWST* at very high redshifts.

The layout of this paper is as follows: Section [2](#) describes the data and our SED fitting method; Section [3](#) outlines our model of the star-forming main sequence; in Section [4](#), we present our results; Section [5](#) discusses the potential biases and limitations of the data set for constraining the main sequence, as well as of our method and in Section [6](#), we summarize our conclusions.

Throughout this work, we have assumed a Chabrier ([2003](#)) IMF with an upper mass cutoff of $100 \, M_\odot$. We employ a flat ΛCDM cosmology with $\Omega_\Lambda = 0.7$, $\Omega_M = 0.3$, and $H_0 = 70 \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$. Magnitudes are in the AB system.

2 DATA AND SED FITTING

The ASTRODEEP catalogue (Merlin et al. [2016a](#); Castellano et al. [2016](#); Di Criscienzo et al. [2017](#)) includes four of the six Frontier Fields: Abell 2744, MACS0416, MACS0717, and MACS1149, as

---

[1]see http://www.sedfitting.org/Fitting.html for more codes, as well as Pacifici et al. (in preparation).
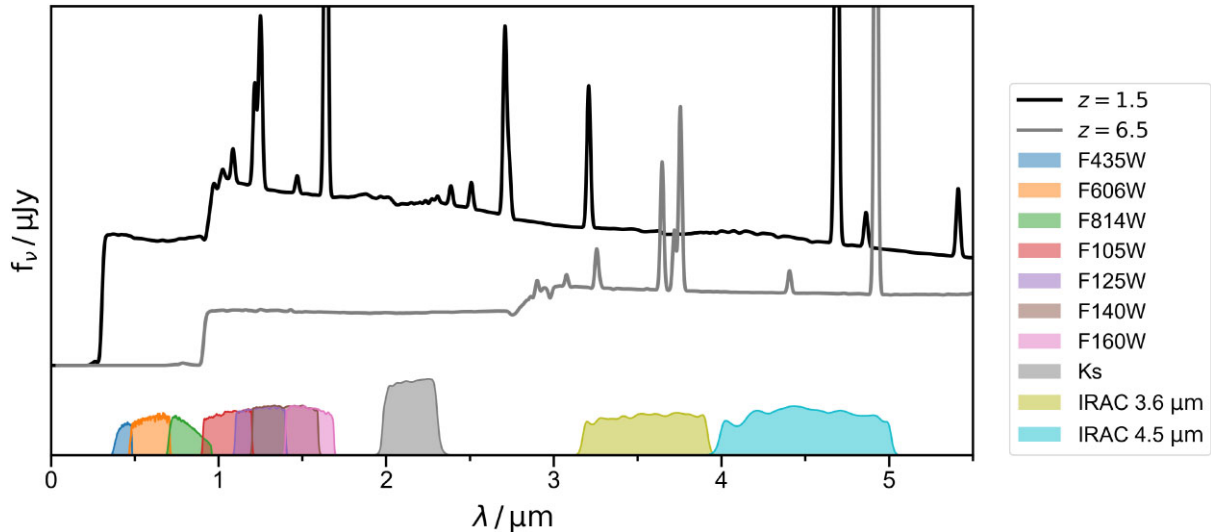
Figure 1. The black and grey lines show example spectra of mock galaxies at redshifts $z = 1.5$ and $z = 6.5$, respectively. Below the spectra, we show the profiles of the 10 broad-band filters included in the ASTRODEEP catalogue (see legend). The profiles are plotted with arbitrary normalization and offset from the spectra for clarity. We see that at $z = 6.5$ only the IRAC 3.6- and 4.5-μm filters sample the rest-frame optical.

well as their corresponding parallel fields. The *HST* Advanced Camera for Surveys (ACS) provides optical imaging while *HST* Wide-field Camera 3 (WFC3), ground-based HAWK-I (High Acuity Wide field K-band Imager) and *Spitzer* IRAC (Infrared Array Camera) provide imaging over the near-infrared. These provide a total of 10 filters that are displayed in Fig. 1.

Merlin et al. (2016a) and Di Criscienzo et al. (2017) describe how the catalogues are produced but we summarize the main points here. The F160W image is used for primary object detection and provides the base of the catalogue. New objects detected in a stacked IR image (F105W, F125W, F140W, and F160W-band) are added to the catalogue. The total ASTRODEEP catalogue contains 29 373 objects. For the purpose of this work we only use the cluster fields, containing 15 379 objects. The HAWK-I *K*s-band imaging and *Spitzer* IRAC imaging has significantly poorer resolution than the *HST* data. Therefore the ASTRODEEP team use a deconfusion method, using the software TPHOT (Merlin et al. 2015, 2016b), to perform photometry in these longer wavelength images, taking the high-resolution *HST* detection image as a prior of the source shapes and positions.

The catalogue includes quality flags that we use to run a first pass selection of objects to analyse. We discarded all objects with RELFLAG = 0[2] leaving 11 818 objects.

2.1 SED fitting

We wish to exploit the full form of the posterior distribution in stellar mass, $M_{\star}$ [$= \log(M/M_{\odot})$], SFR, Ψ [$= \log(\psi/M_{\odot} \text{ yr}^{-1})$], and redshift, z to derive constraints on the main sequence and its evolution. Although the ASTRODEEP team supplied photometric redshifts and derived physical parameters, to achieve our goal, we re-fit to the photometry using BEAGLE (BayEsian Analysis of GaLaxy sEds), a Bayesian SED fitting code (Chevallard & Charlot 2016). A

Table 1. Parameters and associated priors set in BEAGLE for fitting to the ASTRODEEP catalogue.

Parameter	Prior
$\log(t/\text{yr})$	$\mathcal{N}(8.0, 2.0^2)$, truncated $\in [6.0, 10.0]$
$\log(\tau_{\text{SFR}}/\text{yr})$	Uniform $\in [7.0, 10.5]$
$\log(M_{\text{tot}}/M_{\odot})$	Uniform $\in [5.0, 12.0]$
$\log(Z/Z_{\odot})$	Uniform $\in [-2.1, 0.3]$
z	Uniform $\in [0.0, 15.0]$
$\hat{\tau}_{\text{V}}$	$\exp(-\hat{\tau}_{\text{V}})$, for $\hat{\tau}_{\text{V}} \in [0.0, 6.0]$
$\log U_{\text{S}}$	Dependent
ξ_{d}	Fixed 0.3
μ_{d}	Fixed 0.4
$n_{\text{H}}/\text{cm}^{-3}$	Fixed 100
$(\text{C/O})/(\text{C/O})_{\odot}$	Fixed to solar, where $(\text{C/O})_{\odot} = 0.44$
$M_{\text{up}}/M_{\odot}$	Fixed 100

detailed description of the BEAGLE parameters which can be adjusted is given in table 2 of Chevallard & Charlot (2016). We do not use the full flexibility of BEAGLE and limit our exploration to the parameters listed in Table 1, which we describe briefly in this section.

BEAGLE was written to incorporate physically consistent models of nebular plus stellar emission. For this work, we model the stellar emission using the version of the Bruzual & Charlot (2003) stellar population synthesis models described in Vidal-García et al. (2017, see their paper for more details). For the nebular emission (line and continuum), we adopt the ionization-bounded nebulae models of Gutkin, Charlot & Bruzual (2016) that self-consistently trace the production and transmission through the interstellar medium (ISM) of the light from the youngest stars (< 10 Myr).

We characterize the nebular emission using galaxy-wide ionized gas parameters: the interstellar metallicity Z_{ISM}, which we set equal to the metallicity of the young ionizing stars Z; the typical ionization parameter of a newly ionized H II region, U_{s},[3] which characterizes

---

[2]This flag value implies unreliable photometry due to either a flagged error from SEXTRACTOR (Bertin & Arnouts 1996), unpyhsical flux in the detection band, less than five reliable *HST* measurements or close proximity in the image to foreground clusters, stellar spikes or the frame edge.

[3]Note that U_{S} differs from the volume-averaged ionization parameter, $\langle U \rangle$ according to $\langle U \rangle = 9/4 \, U_{\text{S}}$.

the ratio of the photon density to hydrogen density at the inner edge of the Strömgren sphere; and the mass fraction of interstellar metals in the galaxy locked into dust grains ξ_d. CL21 demonstrates that $\log U_S$ and ξ_d are poorly constrained from broad-band photometric data and can bias main-sequence determinations. We, thus, fix ξ_d to 0.3, and impose a relation between $\log U_S$ and $\log(Z_{ISM}/Z_\odot)$ taken from observations

$$\log U_S = -3.638 + 0.055 \log(Z/Z_\odot) + 0.68 \log^2(Z/Z_\odot). \quad (1)$$

This relation is taken from the observational data presented in Carton et al. (2017; private communication). Similarly to CL21, within the H II region, we fix the carbon to oxygen abundance ratio (C/O) to the solar value of $(C/O)_\odot = 0.44$, the hydrogen density to $n_H = 100 \, \text{cm}^{-3}$, and model the intergalactic absorption as prescribed by Inoue et al. (2014).

To maintain consistency with previous observational studies (e.g. Kurczynski et al. 2016; Santini et al. 2017), we adopt a delayed exponentially declining (DE) SFH of the form $\psi(t) \propto t \, \exp(-t/\tau_{SFR})$, where $\psi(t)$ is the SFR, t is the time since the formation of the oldest stars, and τ_{SFR} is the time between the onset of star formation and the peak of the SFH. This SFH allows for very low SFR at a given stellar mass (not allowed by a constant SFR), while also describing a rising SFH when $t < \tau_{SFR}$, which has been suggested to be suitable for high redshifts (e.g. Salmon et al. 2015). The integral of the SFH with respect to time gives the total amount of stellar mass formed, M_{tot}, and is the parameter sampled over within BEAGLE. M gives the stellar mass in stars at a given time after accounting for the return fraction to the ISM after stars die, and is the parameter used for our measurements of the main sequence.

CL21 show that fitting to *JWST* broad-band fluxes of $z \sim 5$ simulated galaxies with a DE SFH results in poorly constrained physical parameters which in turn biases the measurement of the main sequence. This is due to the unknown contribution of emission line fluxes to the broad-band filters, and the effects are mitigated when medium-band filters are available. Our data set does not include medium-band filters so where emission lines contribute a significant fraction of the broad-band flux at high redshifts, we may still derive biased stellar masses and SFRs. At lower redshifts, however, the line equivalent widths are lower and hence the relative contribution of emission lines compared to the stellar continuum is much smaller. We investigate the effect of poor constraints on our derived main-sequence parameters in Section 5.1.

We incorporate dust attenuation using the physically motivated two-component model of Charlot & Fall (2000). The components of this model are the diffuse dust distributed uniformly throughout a galaxy's ISM, and the dust within denser stellar birth clouds. Within this model, stars older than 10 Myr only see the effects of diffuse dust within the ISM, having a *V*-band optical depth equal to that of the ISM, $\hat{\tau}_V^{ISM}$. The birth clouds enshrouding stars younger than 10 Myr have an optical depth $\hat{\tau}_V^{BC}$, giving a total optical depth to young stars of $\hat{\tau}_V = \hat{\tau}_V^{ISM} + \hat{\tau}_V^{BC}$. The fractional attenuation of stars residing in the ISM compared to those residing in stellar birth clouds is given by

$$\mu_d = \frac{\hat{\tau}_V^{ISM}}{\hat{\tau}_V^{ISM} + \hat{\tau}_V^{BC}}. \quad (2)$$

We use the updated treatment of dust in BEAGLE[4] which accounts for the effects of dust within the Gutkin et al. (2016) nebular models themselves, as described in CL21, Section 2.

We fit to 11 818 objects within the four cluster fields, with six free parameters: $(\log(t/\text{yr}), \log(\tau_{SFR}/\text{yr}), M_{tot}, Z, z,$ and $\hat{\tau}_V)$. Table 1 shows the prior distributions configured within BEAGLE.

We do not include HAWK-I and *Spitzer* photometry in the fitting if the ASTRODEEP COVMAX flag indicates that an object suffers severe blending with another source during the TPHOT extraction process (COVMAX$_{filt}$ > 1). The COVMAX flag is publicly available for the Abell 2744, MACS0416, and their parallels while the ASTRODEEP team provided the flags for MACS0717 and MACS1149 (private communication).

When fitting to the observed photometry, we allow for a minimum relative error which is added in quadrature to the measurement uncertainties. This minimum error accounts for the possible calibration differences from photometry derived with different telescopes, as well as the uncertainties in the models. For the *HST* photometry, we allow a minimum relative error of 0.04. This is higher than previously suggested (e.g. Chevallard & Charlot 2016) after finding that the brightest galaxies had poor-fitting χ^2 values due to the small measurement uncertainties. Values of 0.05 and 0.1 are applied to the HAWK-I and IRAC photometry, respectively. HAWK-I and IRAC images require deconfusion, and hence likely suffer systematic uncertainties that are not accounted for in the supplied photometric errors.

Fig. 2 is an example (Abell 2744 cluster, ID 331) of the BEAGLE output available for each fitted object.

2.2 Photometric redshift analysis

The ASTRODEEP collaboration provide photometric redshift estimates which are the median values taken from six independent methods as described in Castellano et al. (2016; Abell 2744 and MACS0416) and Di Criscienzo et al. (2017; MACS0717 and MACS1149).

Prior to analysing BEAGLE-derived photometric redshifts, we discard objects with a F160W AB magnitude fainter than 27.5. This cut was employed by Santini et al. (2017), and based on simulations by Merlin et al. (2016a) designed to determine the detection completeness of the images. The limit corresponds to 90–95 per cent completeness for point-like sources and 50–80 per cent for extended discs with a 0.2 arcsec half-light radius. In addition, we reject objects with a poor fit by BEAGLE defined as having a minimum $\chi^2 > 13.28$.[5] We also impose a lower mass cut as described in Section 2.3.

In Fig. 3, we compare the BEAGLE-derived (posterior median) photometric redshifts to those in the ASTRODEEP catalogue (see Castellano et al. 2016, section 3). The plot shows objects from the four cluster fields which satisfy the above criteria. Whilst the majority of objects lie close to the identity relation, there are many which BEAGLE has identified as $z \sim 4$ in contrast to an ASTRODEEP redshift of $z \sim 0.5$. Photometric redshifts are primarily determined by the detection of a break in the observed SED. In this scenario, ASTRODEEP has assigned a Balmer break (at rest-frame 3646 Å) to the observed break while BEAGLE has assigned a Lyman break (at rest-frame 1216 Å).

For this filter-set, the Lyman break is not reliably bracketed by two filters until $z \sim 4.5$. At redshifts lower than this, reliable determination of Lyman versus Balmer break will be improved with the Ks and IRAC bands sampling red-ward of the Balmer

---

[4] Available from BEAGLE v0.27.1.

[5] Fits with a minimum $\chi^2 = 13.28$ are consistent with our model 99 per cent of the time, under the assumption of 10 available ASTRODEEP filters, with BEAGLE fitting for six independent parameters.
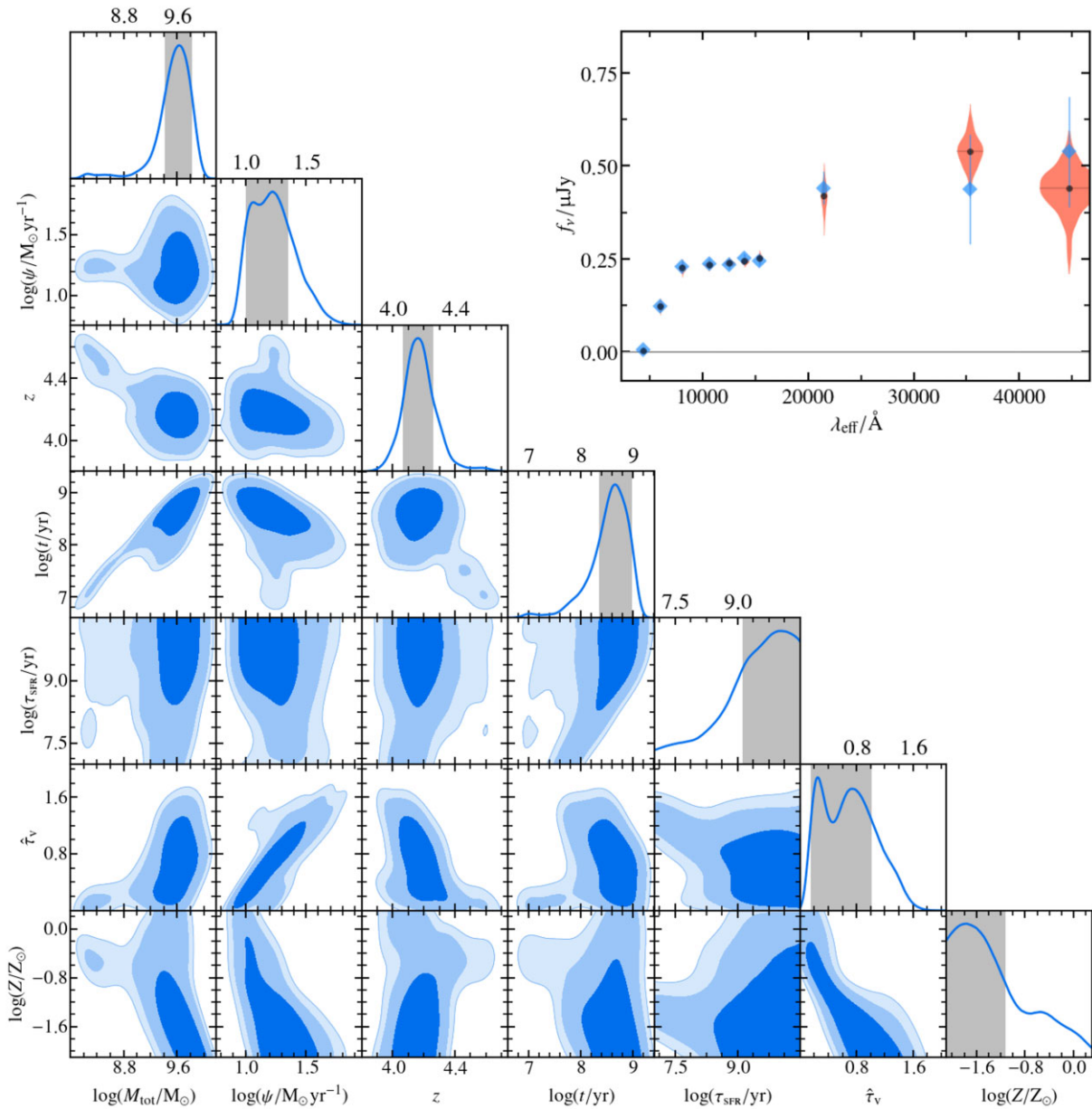
Figure 2. Abell 2744 cluster, ID 331. Bottom left-hand panels: The diagonal panels show the marginal probability distributions for each of the six fitted parameters ($\log(M_{tot}/M_\odot)$, z, $\log(t/\mathrm{yr})$, $\log(\tau_{SFR}/\mathrm{yr})$, $\hat{\tau}_V$, and $\log(Z/Z_\odot)$) as well as $\log(\psi/M_\odot\,\mathrm{yr}^{-1})$. The other panels show the joint posterior distributions for every pair of parameters. Top right-hand panel: Blue diamonds represent the observed SED. Orange violins show the predicted model fluxes as determined by the posterior probability distributions of the fitted parameters.

break. The majority of objects with disagreement between BEAGLE and ASTRODEEP lack robust IRAC photometry (as shown by blue points on the plot), and therefore only show one observed break in the SED. In this situation, for any given object and photometric redshift code, there is some probability that the observed break is incorrectly assigned. For different codes, this probability will vary depending on template set and priors. Castellano et al. (2016) takes the median value of multiple (six) codes, thus mitigating this issue if at least 50 per cent of the codes choose the correct value. Since we require rest-frame optical photometry for our $M_\star$ constraints, those objects with poor photometric redshift estimates would be rejected at the next stage, even if BEAGLE *had* agreed with the ASTRODEEP determinations.

By $z \sim 4.5$, the Balmer break falls red-wards of the Ks-band. We therefore require objects above $z > 3.5$ to have at least one robust photometric point from the three longest wavelength filters (Ks, 3.6 μm, 4.5 μm), while above $z > 4.5$, we require at least one IRAC flux point. Those objects that lack good IRAC/Ks photometry at these redshifts tend to be due to significant confusion in the *Spitzer* images. This is not dependent on the intrinsic properties of the objects themselves, rather the projected distribution of sources on the sky. We therefore do not expect this cut to significantly bias our main sequence determination. Furthermore, we apply a lower redshift limit of 1.25 as below this the F435W-band no longer probes the rest-frame far ultra-violet required for secure SFR determination.
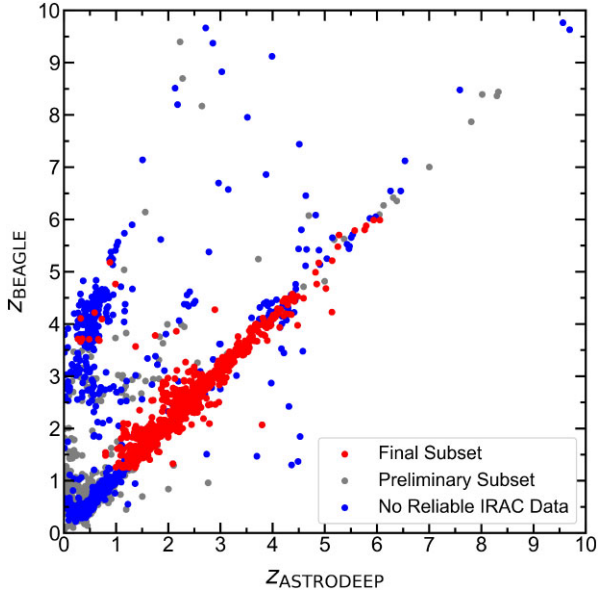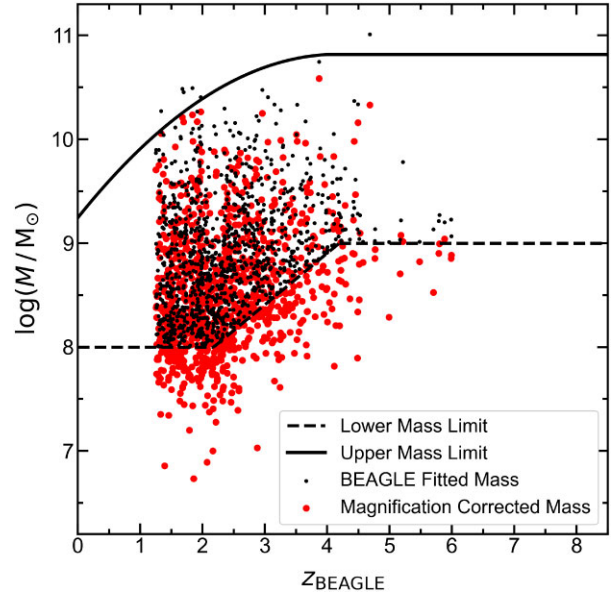
Figure 3. BEAGLE-derived photometric redshifts (posterior medians) plotted against ASTRODEEP redshifts. All objects with RELFLAG = 1, F160W magnitude < 27.5, BEAGLE-fitted minimum $\chi^2 < 13.28$ and a redshift-dependent lower mass cut applied prior to correcting for gravitational lensing (see text and Fig. 4) are plotted. The red points mark the 1038 galaxies chosen as our final subset (see text). Blue points show the objects which have no reliable IRAC data and that do not make it into our sample. Grey points have good IRAC photometry but do not make it into our sample.

Figure 4. Shown in blue are the BEAGLE-derived posterior median stellar mass and redshift estimates plotted against each other for our final sample of 1038 objects. The dashed black line shows the lower limit imposed upon the BEAGLE-derived stellar masses based on 95-per cent mass completeness for F160W magnitude < 27.5 (see text for details). The cuts are imposed prior to correcting the derived properties for the effects of gravitational lensing. Magnification-corrected stellar masses are shown in red. The solid black line shows the redshift-dependent turnover mass as fitted by Tomczak et al. (2016), fixed as a constant for $z > 4$. This upper limit is applied after magnification corrections.

We visually inspected the images and SEDs for all objects with either a BEAGLE redshift (z_{BEAGLE}) or an ASTRODEEP redshift ($z_{\mathrm{ASTRODEEP}}$) of greater than 3.5. We leverage the better accuracy of the ASTRODEEP photometric redshifts by discarding remaining objects (with both z_{BEAGLE} and $z_{\mathrm{ASTRODEEP}} < 3.5$) if $|z_{\mathrm{BEAGLE}} - z_{\mathrm{ASTRODEEP}}| > 1$.

2.3 Sample selection for main-sequence analysis

For analysing the main sequence, we need a sample that is complete in stellar mass. We therefore impose a redshift-dependent mass cut in our samples. This mass limit was calculated using the JAGUAR mock catalogue (Williams et al. 2018), which was produced with the same stellar and nebular models, making the limits self-consistent with the BEAGLE fits performed here. We calculate the mass limit, as a function of redshift, at which the sample is 95-per cent complete in stellar mass for F160W magnitude < 27.5. The limit is displayed as the dashed black line in Fig. 4. The limit is a function of the position of the main sequence in the $M_\star - \Psi$ plane, how well the F160W limit approximates a stellar mass limit, and how the brightness of the objects vary with redshift. At low redshifts, the F160W cut approximates a stellar mass cut, whereas at high redshifts, it approximates a cut in SFR, where the transition between these two limits causes an increase in the lower mass limit between $z \sim 2.5-4$. Where the mass limit is approximately flat, the change in position of the objects in the $M_\star - \Psi$ plane must be compensating the reduction in flux with increasing redshift. We apply the cuts based on $M_\star$ estimates prior to correcting them for the effects of gravitational lensing (as the F160W is a limit of the image, not the intrinsic galactic properties). These are shown as blue points in Fig. 4. The red points show the masses after lensing is accounted for, demonstrating that we probe below the $M_\star$ limits of standard blank fields. We correct the BEAGLE-

derived masses and SFRs using the magnification value supplied in the ASTRODEEP catalogues (see Castellano et al. 2016, for details).

A redshift-dependent upper mass limit is also imposed on the magnification-corrected values to ensure that we are not including objects in the regime where the main sequence has been observed to flatten. Between $0 < z < 4$, we take the parametrization of the turnover mass from Tomczak et al. (2016) and for $z > 4$, we choose a fixed turnover mass of $\sim 10^{10.8} M_\odot$. This limit is shown as the thick black line in Fig. 4.

In summary, the full set of selection criteria includes selecting objects with reliable photometry identified by RELFLAG = 1 with F160W magnitude <27.5 and BEAGLE fits with $\chi^2 < 13.28$. We require agreement with ASTRODEEP redshift within $|\Delta z| < 1$ for $z < 3.5$ and visual inspection above $z > 3.5$. We ensure objects have photometry sampling the rest-frame optical, enabling stellar mass determination. Finally we apply the upper and lower mass limits described here. Our final sample spans $1.25 < z < 6$ and includes 1038 objects which are shown as red points in Figs 3 and 4.

3 MODELLING THE MAIN SEQUENCE

In this section, we detail the steps that we have taken to model the star-forming main sequence spanning redshifts $1.25 < z < 6$.

At a single redshift, ordinary linear regression applied to the star-forming main sequence fails to fully account for heteroskedastic, co-varying errors, and the non-uniform distribution of $M_\star$. Kelly (2007, hereafter K07) proposes a Bayesian hierarchical method to address these concerns, which has been extended by CL21 to work with the output joint posteriors of $M_\star$ and Ψ derived from SED fitting with BEAGLE. This approach allows for the self-consistent propagation of

measurement uncertainties on to the parameters which describe the main-sequence relation: the slope, intercept, and intrinsic scatter.

Throughout this section, we refer to Bayesian terms such as prior probability, likelihood, and posterior probability. It is therefore informative to recap Bayes' theorem, which states that the posterior probability distribution of the model parameters, $P(\Theta \mid \mathbf{D}, H)$, can be expressed as

$$P(\Theta \mid \mathbf{D}, H) \propto P(\mathbf{D} \mid \Theta, H) P(\Theta \mid H), \tag{3}$$

where $P(\mathbf{D} \mid \Theta, H)$ is the likelihood of the data $\mathbf{D}$ given a model (or hypothesis), H, with associated parameters, Θ. The prior probability, $P(\Theta \mid H)$, describes the knowledge we have of the model before analysis of the data.

CL21 apply their model to a mock photometric sample of main-sequence galaxies at $z \sim 5$. They model the main sequence as a linear relation with Gaussian scatter, which we re-write, subtly, to find the normalization of the relation at $\log(M/M_\odot) = 9.7$

$$\Psi = \alpha_{9.7} + \beta(M_\star - 9.7) + \mathcal{N}(0, \sigma^2), \tag{4}$$

$\alpha_{9.7}$ is the normalization of the main sequence at a stellar mass of $\log(M/M_\odot) = 9.7$, β is the slope, and $\mathcal{N}(0, \sigma^2)$ denotes a Gaussian distribution centred on zero with a variance of σ^2 and describes the intrinsic scatter about the relation. Throughout this paper, when describing SFR and stellar mass in log space, we use Ψ [$= \log(\psi/M_\odot \, \mathrm{yr}^{-1})$] and $M_\star$ [$= \log(M/M_\odot)$].

The three levels of the K07 Bayesian hierarchical model are: the distribution of stellar masses, which is not assumed to be uniform; the distribution of Ψ given $M_\star$ (equation 4); and the lowest level describes data given the unknown true $M_\star$ and Ψ values. In our case the data consists of photometric fluxes and uncertainties (see CL21, section 3.4 for more details). The K07 model is designed to marginalize over the unknown, true values of $M_\star$ and Ψ for each object when deriving the parameters of interest, namely $\alpha_{9.7}$, β, and σ.

In this work, we extend the model of CL21 by including the redshift evolution of the main sequence. It is also important to account for objects which do not belong to the star-forming main sequence, which we shall refer to as 'outliers'. We explicitly model these outliers to ensure that the uncertainty of which objects belong to the main sequence is fully accounted for in our analysis. To determine what form of redshift evolution to include in the model, we first measure the main sequence in a series of redshift bins (Section 3.1). We describe our model for the redshift evolution of the main sequence in Section 3.2.

3.1 Redshift bins

3.1.1 Modelling outliers

Not all galaxies belong to the star-forming main sequence. Quiescent galaxies will lie significantly below the main sequence while starbursting galaxies, which may be experiencing a recent or ongoing merger, can lie significantly above the relation. In order to investigate how to appropriately model the outliers in our sample, we initially divide our subset of 1038 objects based upon their BEAGLE-derived posterior medians into redshift bins of $1.25 < z < 2$, $2 < z < 3$, $3 < z < 4$, $4 < z < 5$, and $5 < z < 6$.

Hogg, Bovy & Lang (2010) suggest a simple model for incorporating outliers. We therefore investigate the possibility of extending the work of CL21 using this model which allows objects to either reside on the main sequence or within a separate outlier distribution

which is described as a simple Gaussian

$$\Psi \sim \mathcal{N}(\mu_{\mathrm{OL}}, \sigma_{\mathrm{OL}}^2), \tag{5}$$

where $\mathcal{N}(\mu_{\mathrm{OL}}, \sigma_{\mathrm{OL}}^2)$ is a normal distribution with mean μ_{OL} and standard deviation σ_{OL}. This model is implemented as follows

$$\begin{aligned} P(\Psi \mid M_\star) &= P_{\mathrm{MS}} + P_{\mathrm{OL}}, \\ P_{\mathrm{MS}} &= (1 - p_{\mathrm{OL}}) \, P(\Psi \mid \alpha_{9.7} + \beta(M_\star - 9.7) + \mathcal{N}(0, \sigma^2)), \\ P_{\mathrm{OL}} &= p_{\mathrm{OL}} \, P(\Psi \mid \mathcal{N}(\mu_{\mathrm{OL}}, \sigma_{\mathrm{OL}}^2)), \end{aligned} \tag{6}$$

where P_{MS} is the probability that the object belongs to the main sequence and P_{OL} is the probability that the object is an outlier. The parameter p_{OL} defines the ratio of the integrals of the functions describing the outlier and main-sequence distributions at $M_\star$, respectively. We restrict $p_{\mathrm{OL}} < 0.5$, and $\sigma_{\mathrm{OL}} > 1$ ensuring that within the main sequence the relative probability of any given object being an outlier is very small. However, where the probability that an object belongs to the main sequence becomes negligible, there is a higher probability that the object belongs to the outlier distribution. When implementing this outlier model within redshift bins, we have to make a decision about how we treat the mass distribution. We make the assumption that the distribution of $M_\star$ in the outlier population is the same as that of the objects on the main sequence.

During our preliminary tests, it became clear that the majority of the quiescent outliers sitting below the main sequence were highly unconstrained in Ψ. This freedom allowed these objects to be modelled as belonging to the main sequence, effectively lowering the measured normalization, biasing the slope and increasing the intrinsic scatter. We decided to remove such poorly constrained outliers beforehand by sampling from each object's posterior distributions of $M_\star$, Ψ, and z, rejecting objects with standard deviation in $\Psi > 2$ for the samples within the redshift bin. This method removed 38, 26, 8, 0, and 0 outliers below the main sequence from the bins $1.25 < z < 2$, $2 < z < 3$, $3 < z < 4$, $4 < z < 5$, and $5 < z < 6$, respectively.

Having accounted for the majority of the quiescent outliers below the main sequence, we test the proposed outlier model (we label this method OL-Gauss) and compare it to two other methods. The first of these methods calculates the main sequence without any outlier rejection beyond the objects with poorly constrained Ψ (we label this method OL-Minimal). The second method identifies outliers from iterative 3σ-clipping, where we iteratively remove objects further away than three standard deviations from the best linear fit to posterior medians in $M_\star$ and Ψ. This is implemented prior to the removal of the poorly constrained objects below the main sequence. The method of clipping outliers prior to fitting the main sequence is comparable to approaches within the literature (e.g. Kurczynski et al. 2016; Santini et al. 2017). We label this method OL-Clipped.

Once outliers are removed for the OL-Minimal and OL-Clipped methods, the main sequence is measured from the remaining objects using the CL21 Bayesian hierarchical model. The OL-Gauss method measures the main sequence using an updated version of the model, adapted to include the outlier model described in equation 6.

3.1.2 Redshift bin results

Fig. 5 displays Ψ versus $M_\star$ for the objects in each of the five redshift bins. The points, with various symbols, display the BEAGLE-derived posterior medians in $M_\star$ and Ψ, while the errors show the marginalized 68-per cent credible intervals. Objects that are removed because they have poor constraints in Ψ are coloured green and objects that are removed by the OL-Clipped method are displayed as
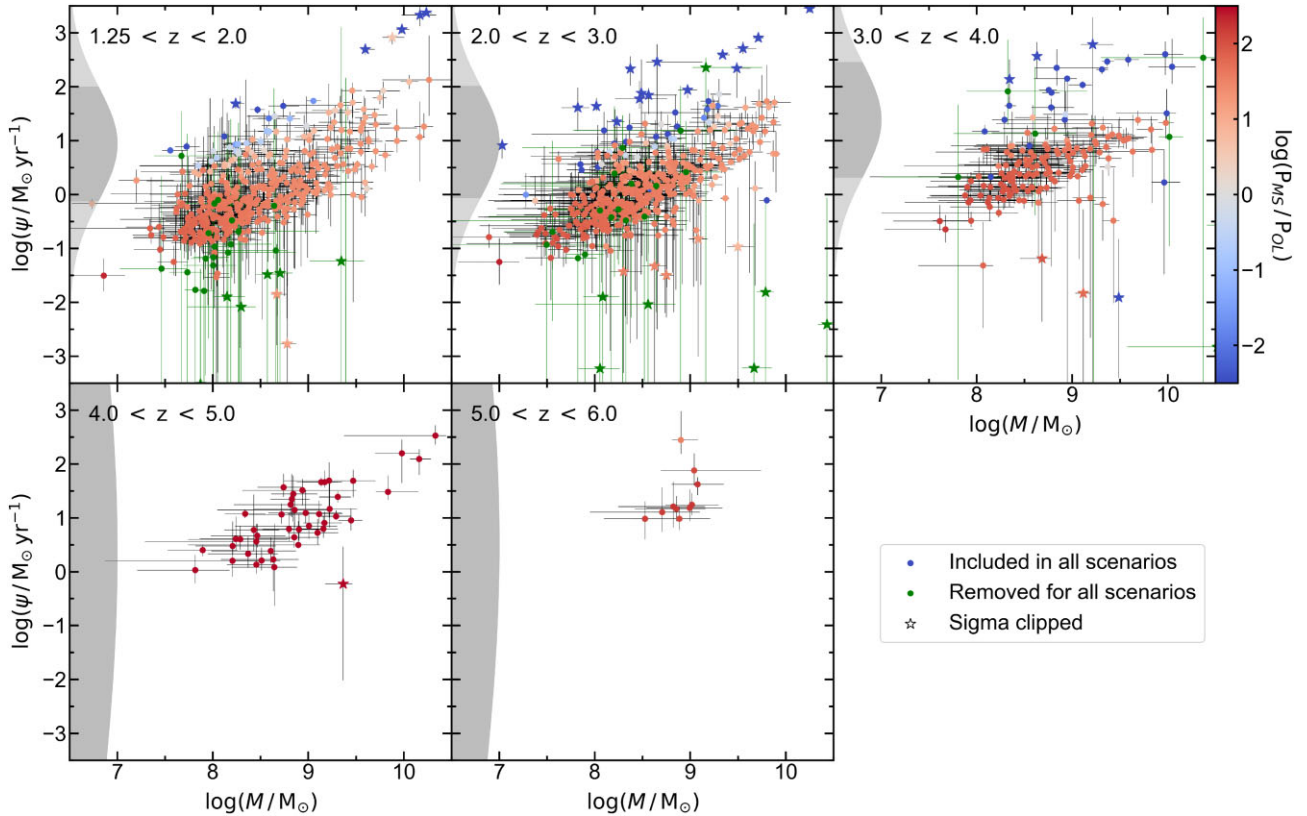
Figure 5. BEAGLE-derived posterior median $\log(\psi/\mathrm{M_\odot\,yr^{-1}})$ plotted against $\log(M/\mathrm{M_\odot})$ in redshift bins. The error bars show marginalized 68-per cent credible intervals in these two parameters. The red-blue colour-coding represents logarithm of the ratio of probability that a given object belongs on the main sequence to the probability that it is an outlier (see equation 6). Green symbols represent the objects removed, regardless of outlier treatment, due to poorly constrained Ψ. Stars of any colour show the objects removed during the 3σ-clipping procedure for the OL-Clipped method. The grey histograms on the left of each panel represent the best-fitting outlier distribution, showing 68-per cent credible regions with dark grey. In the bottom two panels, the outlier distribution is shown as broad, since the distribution is unconstrained due to lack of obvious outliers in redshift bins $4 < z < 5$ and $5 < z < 6$.

stars. The remaining points are coloured by $\log(P_{MS}/P_{OL})$, therefore showing the relative probability of being on the main sequence or within the outlier distribution when using the OL-Gauss method.

The left three panels of Fig. 6 show the derived posterior median values of the main-sequence parameters, $\alpha_{9.7}$, β, and σ, in redshift bins spanning $1.25 < z < 6$. The derived values are also reported in Table 2. The blue shaded rectangles and solid lines show the 68-per cent credible regions and posterior medians, respectively, for the constraints derived with the OL-Gauss method. The dashed and dotted lines show the posterior medians for the parameters derived with the OL-Minimal and OL-Clipped methods, respectively.

Fig. 6 (top left-hand panel) shows that all methods measure an increasing normalization with redshift. For the lowest two redshift bins, the OL-Minimal method returns a higher normalization ($\sim 1.2-1.3$) than the other two methods ($\sim 1.0-1.2$). This is to be expected as the OL-Clipped and OL-Gauss methods both identify a fraction of the objects above the main sequence as outliers, lowering the measured normalization. Within the $3 < z < 4$ bin, however, very few objects are rejected above the relation with the OL-Clipped method, making the results of the OL-Minimal and OL-Clipped methods very similar. The OL-Gauss method, however, ends up assigning a lot of the objects above the relation a high probability of belonging to the outlier distribution, returning a lower normalization. This demonstrates how sensitive the results are to the chosen method to account for outliers.

Fig. 6 (middle left-hand panel) shows that overall there is no strong evidence of varying slope with any of the three methods. OL-Gauss measures a steeper slope than the OL-Clipped and OL-Minimal methods in the $2 < z < 3$ and $3 < z < 4$ redshift bins, because some objects slightly above the relation at masses $M_\star \sim 8 - 9$ have non-zero probability of belonging to the OL-Gauss outlier distribution (Fig. 5, top middle and top right-hand panels). In the lowest redshift bin, however, the OL-Clipped method measures the shallowest slope (~ 0.76), but OL-Minimal measured the steepest slope (~ 0.89). This is because the OL-Minimal run includes two objects below the relation at $M_\star \sim 8.5 - 8.7$ that are identified as having poor constraints on Ψ, yet are clearly below the main sequence (Fig. 5, top left-hand panel). The steeper slope from the OL-Minimal method is therefore less reliable.

The results for the intrinsic scatter in the three lowest redshift bins (Fig. 6, bottom left-hand panel) show that OL-Minimal measures the largest scatter, while the OL-Gauss method shows the lowest measurements with a trend of decreasing scatter with increasing redshift. With further investigation of the $3 < z < 4$ redshift bin (that with lowest OL-Gauss scatter measurement), we find that many of the objects are drawn to a tight main-sequence relation with the OL-Gauss method. This is because the co-varying uncertainties in $M_\star$ and Ψ (for the objects with high probability of being on the relation) approach the expected magnitude of the intrinsic scatter within the underlying relation. K07 demonstrate that in this
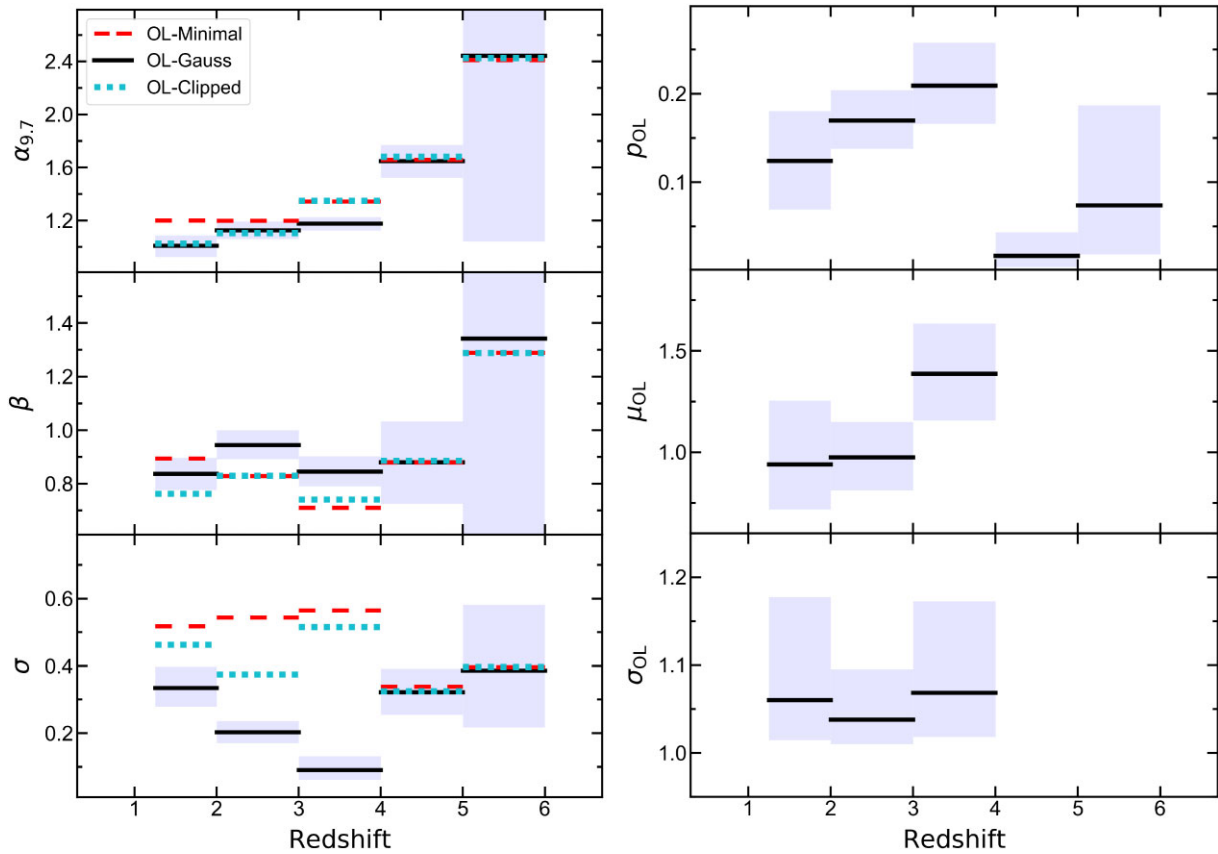
Figure 6. Redshift bin results showing the posterior median values for the best-fitting OL-Minimal, OL-Gauss, and OL-Clipped methods (as shown in the legend). The shaded blue regions show the 68-per cent credible intervals for the OL-Gauss method. Left-hand panel: From top to bottom shows the normalization, slope, and intrinsic scatter of the main sequence as a function of redshift. Right panel: From top to bottom shows $p_{\rm OL}$, $\mu_{\rm OL}$, and $\sigma_{\rm OL}$ of the outlier distribution as a function of redshift. The redshift bin $4 < z < 5$ and $5 < z < 6$ results have been omitted from the bottom two panels for clarity, as in this scenario, $p_{\rm OL}$ approaches 0.

Table 2. Posterior median values and 68 per cent credible intervals for the fitted main sequence and outlier parameters per redshift bin.

Parameter	$1.25 < z < 2$	$2 < z < 3$	$3 < z < 4$	$4 < z < 5$	$5 < z < 6$
$\alpha_{9.7}$	$1.01^{+0.08}_{-0.09}$	$1.13^{+0.07}_{-0.07}$	$1.18^{+0.05}_{-0.05}$	$1.65^{+0.12}_{-0.13}$	$2.44^{+1.51}_{-1.40}$
β	$0.84^{+0.06}_{-0.06}$	$0.94^{+0.06}_{-0.05}$	$0.85^{+0.06}_{-0.06}$	$0.88^{+0.15}_{-0.16}$	$1.34^{+2.09}_{-1.82}$
σ	$0.33^{+0.06}_{-0.06}$	$0.20^{+0.03}_{-0.03}$	$0.09^{+0.04}_{-0.03}$	$0.32^{+0.07}_{-0.03}$	$0.39^{+0.20}_{-0.17}$
$p_{\rm OL}$	$0.12^{+0.06}_{-0.06}$	$0.17^{+0.03}_{-0.03}$	$0.21^{+0.05}_{-0.04}$	$0.02^{+0.03}_{-0.01}$	$0.07^{+0.11}_{-0.06}$
$\mu_{\rm OL}$	$0.94^{+0.31}_{-0.22}$	$0.97^{+0.17}_{-0.16}$	$1.39^{+0.25}_{-0.23}$	$0.35^{+6.39}_{-6.73}$	$0.52^{+5.94}_{-6.85}$
$\sigma_{\rm OL}$	$1.06^{+0.12}_{-0.05}$	$1.04^{+0.06}_{-0.03}$	$1.07^{+0.10}_{-0.05}$	$5.35^{+3.06}_{-2.98}$	$5.26^{+3.19}_{-2.98}$

regime, the method will underestimate the intrinsic scatter. This demonstrates a potentially problematic feature of the OL-Gauss model. In Fig. 7, we show a simplified example where essentially, objects can be identified as belonging to the main sequence (red points) if they have posterior distributions that overlap in $M_\star - \Psi$ space (effectively if the uncertainties are broad, as shown by red ellipses) while objects without overlapping posteriors (blue points) can be assigned to the outlier model. This will lead to shrinkage in the scatter about the derived main sequence (black arrows) by the amount allowed by the overlap. This behaviour may be particularly problematic if constraints on $M_\star$ and Ψ are poor, but led to occupy a similar region in $M_\star - \Psi$ space by informative priors at the SED-fitting stage. We discuss this case further in Section 5.1.

Understanding this behaviour allows us to mitigate its effect when deriving the full redshift evolution model with the OL-Gauss method (Section 3.2).

The three right-hand panels of Fig. 6 show the measured posterior medians and 68-per cent credible intervals for the parameters describing the fitted outlier distribution in the OL-Gauss method: $p_{\rm OL}$, $\mu_{\rm OL}$, and $\sigma_{\rm OL}$. The increase in $p_{\rm OL}$ with redshift mirrors the decrease in σ with redshift for the lowest three redshift bins. This demonstrates a degeneracy between these two parameters, and the importance of clearly identifying outlier galaxies for constraining σ of the main sequence. For clarity, we have not plotted the redshift bins $4 < z < 5$ and $5 < z < 6$ in the panels displaying $\mu_{\rm OL}$ and $\sigma_{\rm OL}$, as the probability of an object belonging to the outlier distribution,
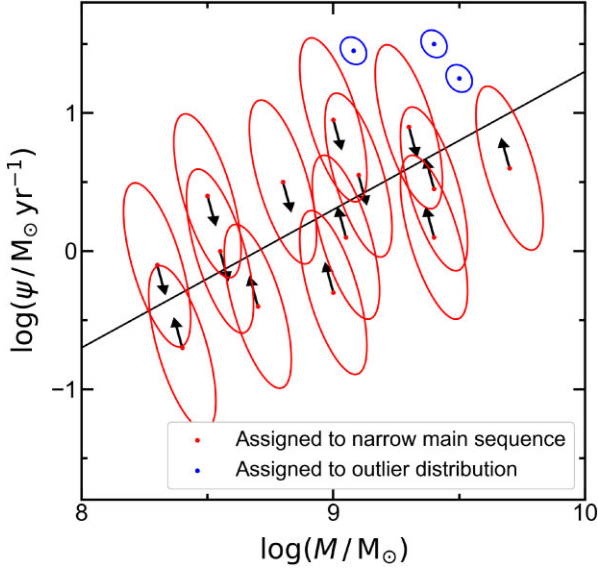
Figure 7. A simplified model of the main sequence (relation shown as a solid black line) showing $\log(\psi/M_\odot \mathrm{yr}^{-1})$ plotted against $\log(M/M_\odot)$. This cartoon displays the behaviour of the model when some objects on the main sequence have poor constraints that overlap (red points) while some have good constraints (blue points). Even though all objects belong to the main sequence, in this case, with our model, the red points would be assigned to a main sequence with small intrinsic scatter while the objects with better constraints (blue points) would be assigned to the outlier model. In this scenario, the large overlap between the red objects will lead to shrinkage in the derived intrinsic scatter, as demonstrated by the black arrows.

p_{OL}, approaches zero in the highest redshift bins, leaving μ_{OL} and σ_{OL} unconstrained. This, in turn, leads to all three methods measuring a similar main-sequence slope ($\sim$0.9, $\sim$1.3), intercept ($\sim$1.6, $\sim$2.4), and intrinsic scatter ($\sim$0.3, $\sim$0.4) in the $4 < z < 5$ and $5 < z < 6$ bins, respectively. We note that this may be due to the small sample sizes making outlier identification less secure, rather than there being fewer outliers in the underlying sample.

We have shown that the derived normalization and intrinsic scatter of the main sequence are highly sensitive to the presence and treatment of outliers in the data. In Section 3.2, where we model the full redshift evolution of the main sequence, we choose to model the outlier population with the OL-Gauss method as it is the *only* method that can propagate the uncertainties on the treatment of outliers on to the parameters of interest. We note that the σ_{OL} posterior probability is close to the prior lower limit. The reason we impose a lower limit of $\sigma_{\mathrm{OL}} > 1$ (as well as the upper limit in p_{OL}) is to ensure that the outlier distribution does not account for objects primarily on the main sequence. As we are fitting two Gaussians to the population, a narrow outlier distribution introduces degeneracies. A wide outlier distribution also ensures that it is accounting for objects far from the main sequence. We will proceed with the current model as there are so few objects within the outlier population that adding further free parameters is unlikely to considerably improve the main-sequence constraints.

The results in this section also demonstrate that we cannot simultaneously constrain the intrinsic scatter and the outlier model within each redshift bin (especially within the $2 < z < 3$ and $3 < z < 4$ redshift bins), and so we account for this when constructing our full redshift-dependent model, as described in the following section.

3.2 Redshift evolution of the main sequence

Our Bayesian hierarchical model of the full redshift-dependent main sequence is composed of three levels. The first describes the distribution of stellar masses and redshifts

$$M_\star \mid \boldsymbol{\eta} \sim \mathrm{P}(M_\star \mid \boldsymbol{\eta}),$$
$$z \mid \boldsymbol{\theta} \sim \mathrm{P}(z \mid \boldsymbol{\theta}), \tag{7}$$

where we model $\mathrm{P}(M_\star \mid \boldsymbol{\eta})$ as a weighted linear combination of three Gaussians, called a Gaussian Mixture Model (GMM). The corresponding set of three means, standard deviations, and relative weightings are denoted as $\boldsymbol{\eta}$. We note that the mass distribution is not modelled as a function of redshift. This does not mean that the $M_\star$ distribution within each redshift bin must look identical, but simply that the model learns the collapsed mass distribution of all objects regardless of their redshifts. We model $\mathrm{P}(z \mid \boldsymbol{\theta})$ as a uniform distribution $\mathcal{U}(1.25, 6)$ between our redshift limits. One would ideally also model the redshift distribution as non-uniform, potentially with another GMM. Modelling the redshift distribution explicitly would potentially help for deriving the relative likelihoods of peaks in probability that are separated significantly in redshift. However, in implementing this model, we found that we had to handle these objects separately (as described later in this section). We therefore chose to not add more free parameters to the model at this level.

Given $M_\star$ and z, the second level of the model describes the probability distribution of Ψ as a linear combination of a main-sequence distribution and an outlier distribution

$$\Psi \mid M_\star, z \sim \mathrm{P}_{\mathrm{MS}} + \mathrm{P}_{\mathrm{OL}},$$
$$\mathrm{P}_{\mathrm{MS}} = (1 - p_{\mathrm{OL}}(z))\left[\alpha_{9.7}(z) + \beta(z)(M_\star - 9.7) + \mathcal{N}(0, \sigma(z)^2)\right],$$
$$\mathrm{P}_{\mathrm{OL}} = p_{\mathrm{OL}}(z)\,\mathcal{N}(\mu_{\mathrm{OL}}(z), \sigma_{\mathrm{OL}}(z)^2), \tag{8}$$

where $\alpha_{9.7}$, β, σ, p_{OL}, μ_{OL}, and σ_{OL} are now functions of redshift.

We determined suitable parametrizations for slope, intercept, and scatter based upon the OL-Gauss redshift bin measurements shown in Fig. 6. The measurements of slope are consistent with being constant within the 68-per cent credible regions so we therefore choose to model it as constant

$$\beta(z) = \beta. \tag{9}$$

The main-sequence normalization, $\alpha_{9.7}$, is shown to increase with redshift from $z \sim 1.25$ to $z \sim 6$. We have shown that the normalization can be very strongly dependent on the modelling of the outliers and our redshift bin results are likely affected by this. We see strong evolution in $\alpha_{9.7}$ to higher redshifts that is not well-reproduced by the parametrization of Speagle et al. (2014), and we do not trust the relatively low normalization of the $3 < z < 4$ bin compared to the $4 < z < 5$ bin for reasons described in Section 3.1.2. We therefore chose to proceed with the physically motivated parametrization that follows the redshift evolution of the rate of accretion of gas on to dark matter haloes (Birnboim et al. 2007). We discuss the implications of this choice in Section 5

$$\alpha_{9.7}(z) = \log(N(1 + z)^\gamma) + 0.7, \tag{10}$$

where N and γ are the free parameters of our redshift-dependent model. The 0.7 is added for simplicity when plotting sSFR (at $M_\star = 9.7$) against redshift.

As discussed in Section 3.1.2, the measured value of the intrinsic scatter is strongly dependent on the treatment of outliers. We have demonstrated that, at least for redshift bins $2 < z < 3$ and $3 < z <$

4, the intrinsic scatter and the outlier model cannot be constrained independently. We therefore choose to parametrize the intrinsic scatter as a constant across all redshifts

$$\sigma(z) = \sigma. \tag{11}$$

This parametrization effectively allows the better $M_\star$ and Ψ constraints in the lowest redshift bin to constrain σ across all redshift bins. This choice is due to the limitations of the current data, which we discuss further in Section 5.1.

Fig. 6 (top right-hand panel) shows a lack of obvious outliers at redshifts $z > 4$. As guided by our OL-Gauss method results, we opt to fit with constant p_{OL}, μ_{OL}, and σ_{OL} for $1.25 < z < 4$, while at $z > 4$, we make the assumption that all of our remaining objects are galaxies belonging to the star-forming main sequence. We implement this by setting p_{OL} to zero at the highest redshifts

$$\mu_{OL}(z) = \mu_{OL},$$
$$\sigma_{OL}(z) = \sigma_{OL},$$
$$p_{OL}(z) = \begin{cases} p_{OL} & \text{for } 1.25 < z < 4 \\ 0 & \text{for } z > 4 \end{cases}. \tag{12}$$

It is worth noting here that our implementation of the outlier model is not the same as sigma-clipping. Each object has a probability of belonging to either the main sequence or the outlier distribution, in contrast to sigma-clipping which permanently removes outliers from the subset. However, an outlier model of this sort is not without potential risks. For example, at a given mass, as the star-forming main sequence evolves with redshift, it will cross directly through the fixed outlier model. If the main sequence and outlier distributions were of comparable probability and width, the outlier distribution could have a similar (but not identical) effect to sigma-clipping of the relation (reducing the derived intrinsic scatter). We have attempted to mitigate this risk by constraining the relative probability of the outlier distribution to $p_{OL} < 0.5$ and its width to $\sigma_{OL} > 1$. Our complete model is then better described as a high probability main sequence, with a low probability distribution of outliers. To ensure that our derived main sequence is not dependent on our implementation of the outlier model, we additionally investigate the effect of using a uniform outlier distribution. The results of this test suggest that our measurement of the star-forming main sequence is robust, as further discussed in Section 4.

Originally, the third level of the K07 model accounts for the data and associated uncertainties, which are assumed to be point-wise estimates. Our data is one step further removed, being instead fluxes and flux uncertainties, rather than direct measurements of $M_\star$, Ψ, and z. We follow the approach of CL21 (section 3.4) for dealing with this extra level of complexity. For simplicity of implementation, the individual object joint posteriors on $M_\star$, Ψ, and z are modelled as a linear combination of three tri-variate Gaussians.

We found when implementing this model that the Gibbs sampler does not efficiently sample between peaks in posterior probability for a given object that are very far apart. It is beyond the scope of this work to re-visit the sampling method. Instead, we use the information provided by the full sample to determine the most likely redshift peak in objects with peaks separated by $\Delta z \geq 2$. Each separate probability peak (determined from the Gaussians fitted to the $M_\star–\Psi–z$ posterior probability space) is multiplied by the probability that the object lies on the main sequence as measured within redshift bins (Fig. 6), and integrated. The peak with higher probability overall is kept. Where two of the three Gaussians overlap within 1.5σ in redshift, their probabilities are summed together and joint probability compared to the final peak. Where all three Gaussians overlap, no peak is rejected.

Table 3. Parameters, fitted results (with 68-per cent credible intervals) and associated priors for the redshift-dependent main-sequence model including the outlier distribution.

Parameter	$1.25 < z < 6$	Prior
N	$0.12^{+0.04}_{-0.03}$	Uniform $\log N \in [-3.0, 2.3]$
γ	$2.40^{+0.18}_{-0.18}$	Uniform $\in [0.0, 5.0]$
β	$0.79^{+0.03}_{-0.04}$	Uniform $\in [-5.0, 5.0]$
σ	$0.26^{+0.02}_{-0.02}$	Uniform $\in [0.05, 5.0]$
$p_{OL}{}^a$	$0.19^{+0.03}_{-0.03}$	Uniform $\in [0.0, 0.5]$
μ_{OL}	$0.98^{+0.11}_{-0.10}$	Uniform $\in [-10.0, 10.0]$
σ_{OL}	$1.01^{+0.02}_{-0.01}$	Uniform $\in [1.0, 10.0]$

Note. [a]Fixed to 0 for $z > 4$.

Although only ∼6 per cent of galaxies had peaks removed, this was necessary because a small fraction of low-redshift objects with a peak at high redshift can significantly affect the results since there are far fewer high-redshift galaxies.

The implementation of the full model is based initially on the J. Meyers PYTHON implementation[6] of the K07 Gibbs sampler updated by CL21 to accept GMM fits to posterior $M_\star–\Psi–z$ distributions derived from BEAGLE fitting. We release the code and input values used for this work.[7]

4 RESULTS

In summary, our final model includes the following free parameters that we wish to constrain: the redshift evolution of the normalization at $\log(M/M_\odot) = 9.7$, parametrized by N and γ; the (redshift-independent) slope, β; the (redshift-independent) scatter, σ; and the parameters describing the outlier distribution, p_{OL}, μ_{OL}, and σ_{OL}. Table 3 gives the parameters and priors used in this work (including our results as described in the following section).

To constrain main-sequence parameters (N, γ, β, σ, p_{OL}, μ_{OL}, and σ_{OL}) for our subset of 1038 objects, we ran the full Bayesian hierarchical model for 20 000 iterations, four separate times. We checked for convergence between and within chains following the method described in chapter 11.4 of Gelman et al. (2013), ensuring an $\hat{R}$ value of <1.1. We use the second half of each chain (rejecting any burn-in phase) and combine them to determine the constraints on the parameters of interest. The results are given in Table 3.

We display the results in Fig. 8. The left-hand panel shows the BEAGLE-derived posterior median $M_\star$ and Ψ plotted in the $M_\star–\Psi$ plane colour-coded by redshift. We see a clear sign of an increase in the normalization with redshift, consistent with previous literature results. The dashed black line passing through $M_\star = 9.7$ indicates the mass at which we define the normalization of the main sequence, $\alpha_{9.7}$. The right-hand panel shows the offsets from the fitted redshift-evolving main sequence versus $M_\star$, colour-coded by the logarithm of the ratio of the probability that objects are on or off the main sequence. A value of zero corresponds to an equal likelihood that the object is on or off the main sequence. In our model, objects with a posterior median redshift value of $z > 4$ are assigned a zero probability of belonging to the outlier distribution and are shown as dark red circles with artificially assigned values of $\log(P_{MS}/P_{OL}) = 3$.

[6]https://github.com/jmeyers314/linmix
[7]https://github.com/ls861/M-SFR-Sandles2022
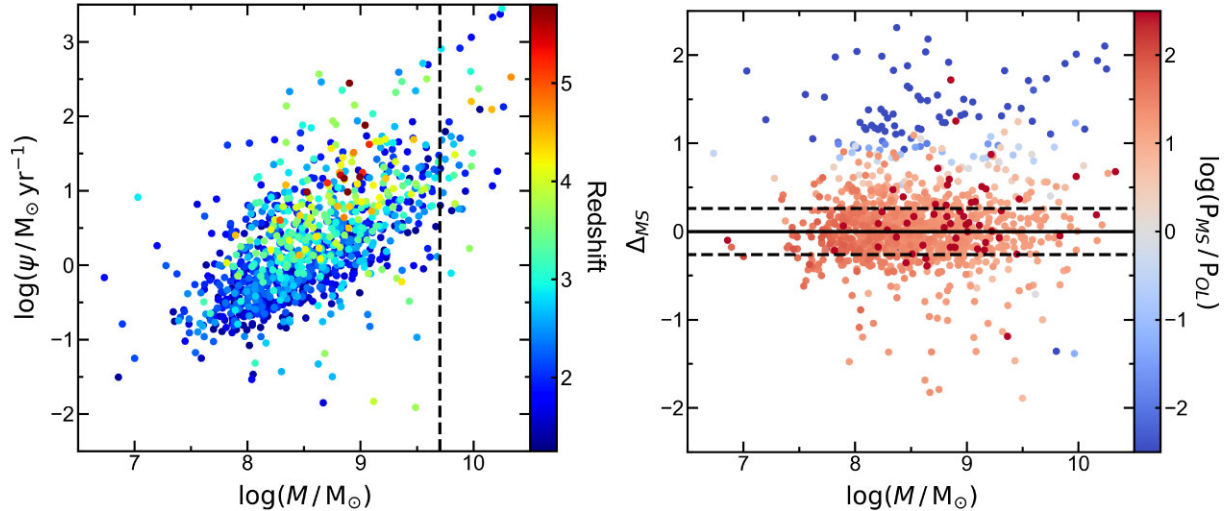
Figure 8. Left-hand panel: BEAGLE-derived posterior median values of $\log(M/M_\odot)$ and $\log(\psi/M_\odot \, yr^{-1})$ colour-coded by posterior median redshift for the objects in our sample. The vertical dashed line at $\log(M/M_\odot) = 9.7$ indicates the fixed mass at which the normalization in the main sequence is fitted. Right-hand panel: The same objects as in the left-hand panel, showing the residual between $\log(\psi/M_\odot \, yr^{-1})$ and the fitted redshift-evolving main sequence, plotted against $\log(M/M_\odot)$. The colour-coding shows the relative probability that the objects are on or off the main sequence.

The objects with poorly constrained Ψ that are rejected prior to fitting are not plotted, leaving more objects significantly above the relation than below. The outlier model primarily accounts for objects above the main sequence but is broad enough to also encompass those below the relation.

Fig. 9 shows the derived posterior median values and 68-per cent credible intervals (red line and shaded regions, respectively) of the main-sequence slope, redshift-dependent normalization and intrinsic scatter spanning $1.25 < z < 6$. The OL-Gauss redshift bin values are shown as solid black lines and shaded blue regions. The panels additionally include data obtained from the literature[8] and simulations. The top panel shows the redshift evolution of $\alpha_{9.7}$. The $1.25 < z < 2$ redshift bin results are higher than those of the full model. This is a known problem with our full model parametrization. It is based on the measured accretion rate of gas on to parent haloes, but at low redshift many studies measure higher SFRs than accounted for by this evolution. The recent work of Leja et al. (2021) find a much lower normalization for the main sequence at low-to-intermediate redshifts, agreeing well with the predictions from hydrodynamic simulations of galaxy formation (e.g. Illustris-TNG Donnari et al. 2019a). Our measured normalization is still higher than that measured by Leja et al. (2021), which can be attributed to the different SFHs employed (their non-parametric histories show older, more massive galaxies than estimates derived with simple analytic forms like the DE used here). We discuss further the limitations of our results with respect to SFH in Section 5.1. The high-redshift bins are driving the fit of the redshift evolution of $\alpha_{9.7}$, showing that our results are inconsistent with a flatter evolution at high redshifts as measured by e.g. Speagle et al. (2014) and Pearson et al. (2018).

The middle panel of Fig. 9 shows the measurements of main-sequence slope, β, across the full redshift range. Our measurements agree well with those of Kurczynski et al. (2016), Speagle et al. (2014), and Pearson et al. (2018) above $z > 2$, but are somewhat shallower than those measured by Santini et al. (2017) and Leja et al.
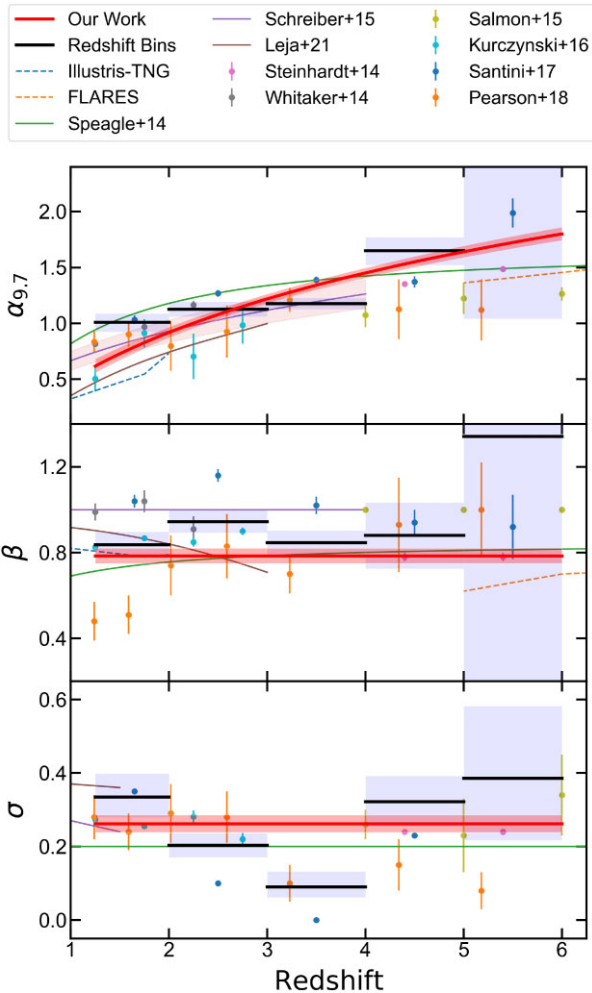
(2021). The Schreiber et al. (2015) and Salmon et al. (2015) slope values are fixed (where we take the low mass slope of the curved relation fitted in Schreiber et al. 2015, and we have chosen to plot the results from Salmon et al. 2015 fitted with a fixed slope). We discuss in Section 5.1 the effects of the priors employed in BEAGLE, which will strongly affect the measured slope.

The bottom panel of Fig. 9 shows measurements of the scatter about the main sequence. Our constant, intrinsic scatter estimate, σ, agrees well with Pearson et al. (2018) up to $z \sim 3$, Kurczynski et al. (2016), Steinhardt et al. (2014), and Salmon et al. (2015). We note that the value of scatter plotted for Steinhardt et al. (2014) is an observed scatter, rather than the intrinsic value. Interestingly, some of the studies show decreasing scatter with particularly low estimates at $z \gtrsim 3$ (e.g. Santini et al. 2017; Pearson et al. 2018) which agree better with our $z \sim 3$ redshift bin results. However, at these redshifts the Pearson et al. (2018) main sequence has a very strong lower limit that appears to be biasing the scatter to low values (see their figure 8). Additionally, Santini et al. (2017) used the same data set as us, and so are likely inhibited by the same limitations in constraints on $M_\star$ and Ψ, which would underestimate the σ at $z \sim 3$.

As discussed in Section 3.1.1, it was important to ensure that our derived main sequence was not strongly dependent on our implementation of the outlier model. As a simple check, we decided to also fit the full $1.25 < z < 6$ subset with an adjusted outlier model: a truncated Gaussian between $-2.0 < \Psi < 3.75$ with fixed $\mu_{OL} = 0.0$ and $\sigma_{OL} = 9.0$ (effectively a uniform outlier distribution between $-2.0 < \Psi < 3.75$ for $z < 4$). The redshift evolution of the fitted main-sequence intercept ($N = 0.16^{+0.05}_{-0.04}$ and $\gamma = 2.29^{+0.19}_{-0.19}$) remained consistent with the original model ($N = 0.12^{+0.04}_{-0.03}$ and $\gamma = 2.40^{+0.18}_{-0.18}$), whilst the slope was measured to be only slightly lower at $\beta = 0.71^{+0.04}_{-0.04}$. The intrinsic scatter was determined to be significantly higher: $\sigma = 0.46^{+0.03}_{-0.04}$. We note that this was primarily due to the uniform outlier model being assigned a low probability ($p_{OL} = 0.04^{+0.02}_{-0.02}$), with most of the objects having a high probability of being on the main sequence. This suggests that a uniform outlier model is inadequate for describing the outlier population.

Fig. 10 shows the bi-variate posterior distributions between each pair of parameters. The main diagonal shows the marginalized

---

[8]We ensure consistency of IMF using conversion factors -0.21 for Salpeter to Chabrier $M_\star$, -0.20 for Salpeter to Chabrier Ψ and -0.03 for Kroupa to Chabrier $M_\star$ and Ψ.

Figure 9. Redshift evolution of the normalization, $\alpha_{9.7}$ (upper panel), slope, β (middle panel) and scatter, σ (bottom panel) of the main sequence. Red lines show the relations derived from the posterior medians of our fitted parameters. The red shaded regions show the 68-per cent credible intervals. For the case of $\alpha_{9.7}$, we sample from the joint posterior distribution of N and γ, before calculating the distribution of $\alpha_{9.7}$ at each given redshift. The redshift bin results from the OL-Gauss method (Fig. 6) are shown as solid black lines with 68-per cent credible intervals shaded blue. Results from the literature are also overplotted following the legend. Illustris-TNG results come from Donnari et al. (2019b) and FLARES results are taken from Lovell et al. (2021) (values obtained via private communication). All data in the top two panels are plotted for $M_* = 9.7$. Uncertainties in literature $\alpha_{9.7}$ values are calculated in quadrature if the original work measured the intercept at a different mass. Where the fitted main sequence allows for curvature at high masses, we plot the low mass linear slope (e.g. Schreiber et al. 2015). For Leja et al. (2021), we use broken power-law parametrization fitted to the ridge in density in $M_* - \Psi$ space (their table 1). For Whitaker et al. (2014), we use the broken power-law fit results. The scatter is sometimes measured as a function of mass. For Schreiber et al. (2015), Santini et al. (2017), and Leja et al. (2021), we plot the values of scatter taken at masses $M_* = 10.2$, 9.2, and 9.7, respectively. We have converted the values where necessary to be consistent with a Chabrier IMF.

posterior distributions of each individual parameter, with vertical lines representing the median and 68-per cent credible intervals. This helps to understand where degeneracies between different parameters may impact our results. We see a negative correlation between σ and p_{OL}, p_{OL} and μ_{OL}. This further demonstrates the sensitivity of our σ estimates to the details of the outlier model. We also see strong

degeneracies between γ, N, and β, such that low N requires a low β, but high N. This might explain why we measure shallower slope in the full model compared to in the redshift bins.

Fig. 11 shows the redshift evolution of the specific star formation rate (sSFR) at $\log(M/M_\odot) = 9.7$. By definition, at a fixed stellar mass, sSFR follows the redshift evolution of the main-sequence normalization at that mass. Measurements of sSFR are not always derived from measuring the main sequence, and so we can compare to more results in the literature. Our measured evolution is clearly more consistent with the data that show significantly higher sSFR at high redshifts, compared to the data that suggest a flatter evolution to high redshift. Our measurement of γ ($2.40^{+0.18}_{-0.18}$) is consistent with the evolution of the accretion rate of gas on to parent haloes [shown as the dashed blue line, with evolution $\sim (1 + z)^{2.25}$].

5 DISCUSSION

5.1 Choice of star formation history

#### 5.1.1 Determining constraints on M_* and Ψ

In Section 4, we have presented our measurement of the slope, intercept and scatter of the star-forming main sequence between redshifts $1.25 < z < 6$. We performed the BEAGLE fits with a delayed exponentially declining SFH of the form $\psi(t) \propto t \exp(-t/\tau_{SFR})$. CL21 demonstrated that the constraints on τ_{SFR} can be poor, and subsequently lead to significantly biased estimates on main-sequence parameters. The CL21 study was based on simulated data at $z \sim 5$, using a set of *JWST* Near-infrared camera (NIRCam) filters. Our data set spans a wide redshift range and uses a different set of filters, but we can still evaluate the possible impact of poorly constrained SFH parameters on the derived main sequence by comparing fits performed using different priors on τ_{SFR}. We therefore re-ran BEAGLE using a uniform prior on $1/\tau_{SFR}$ (a prior suggested by Carnall et al. 2019), within the same limits as our fiducial prior, which was uniform on $\log(\tau_{SFR})$ (see Table 1).

The results for the $M_* - \Psi$ plane for the $1.25 < z < 2$ redshift bin are shown in Fig. 12, where we plot the values prior to correcting for any magnification correction to more clearly display any effects from the priors. The original posterior medians are shown as blue points, while the new posterior medians are coloured red, and their 68-per cent credible regions in M_* and Ψ are displayed as grey error bars (for clarity, we do not plot uncertainties for the original estimates). We see a large excess of red points significantly below the relation with very large uncertainties in Ψ. Objects with $\log(\psi/M_\odot \text{ yr}^{-1}) < -1.8$ are shown as large circles, as are the corresponding objects fitted with the original prior. As a comparison, we applied the method described in Section 3.1 to derive a main-sequence slope, intercept, and scatter for the red objects. As one may expect from visually inspecting Fig. 12, we measured a steeper slope of $0.93^{+0.09}_{-0.09}$ (compared with $0.84^{+0.06}_{-0.06}$) and a larger intrinsic scatter of $0.67^{+0.04}_{-0.04}$ (compared with $0.33^{+0.06}_{-0.06}$). The newly fitted main-sequence intercept ($1.15^{+0.10}_{-0.11}$) was consistent with that of the original run ($1.01^{+0.08}_{-0.09}$). However, the presence of an outlier distribution was significantly down-weighted with $p_{OL} = 0.00^{+0.01}_{-0.00}$ (compared with $0.12^{+0.06}_{-0.06}$).

To understand the behaviour of the fits with uniform prior on $\log(\tau_{SFR})$, we need to visualize the resulting prior in $M_* - \Psi$ space. This is shown in Fig. 13. We see a 'ridge' in the original prior, which is very close to the values of Ψ measured for objects that are subsequently measured to have very low Ψ with the new prior (we have plotted the posterior medians, so they are slightly above the ridge which would represent the extent of the 95-per cent credible
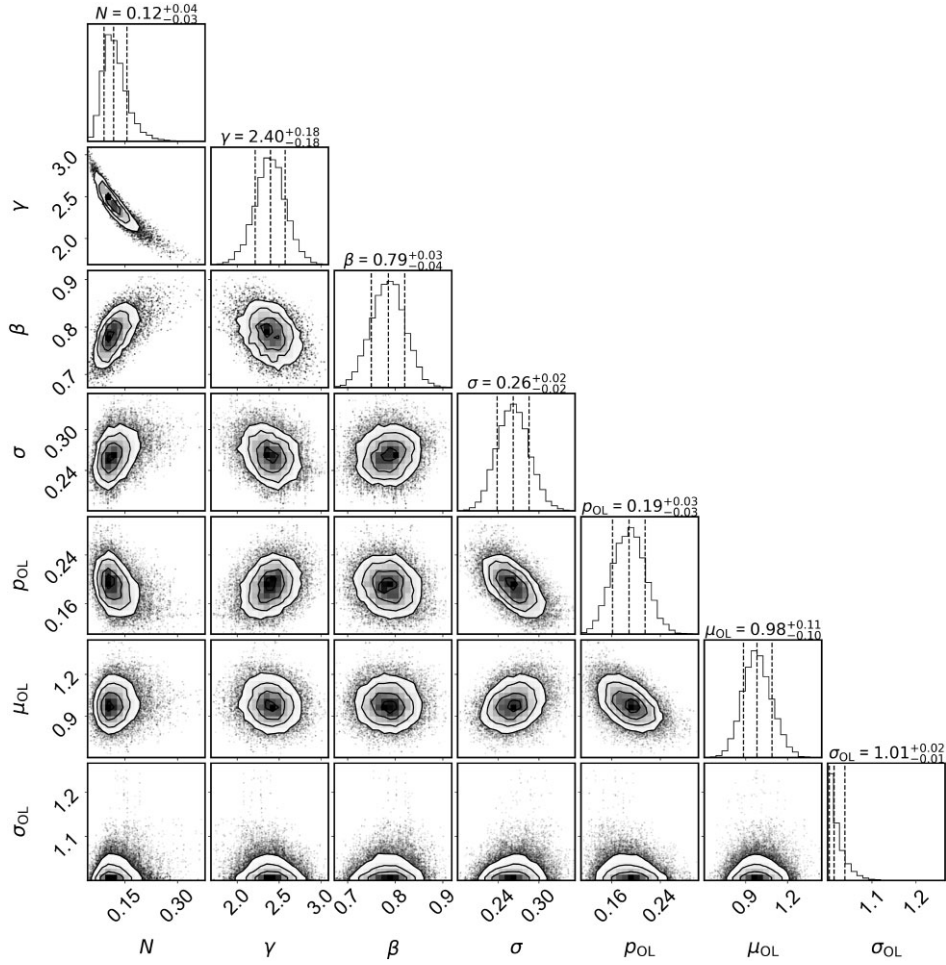
Figure 10. Main diagonal: Marginal probability distributions for each of the seven fitted parameters (N, γ, β, σ, p_{OL}, μ_{OL}, and σ_{OL}) in our full redshift dependent run. Vertical dashed lines show the median and 68-per cent credible intervals. Off-centre: Joint posterior distributions for every pair of parameters with solid black lines to show the 1, 2, and 3σ contours.
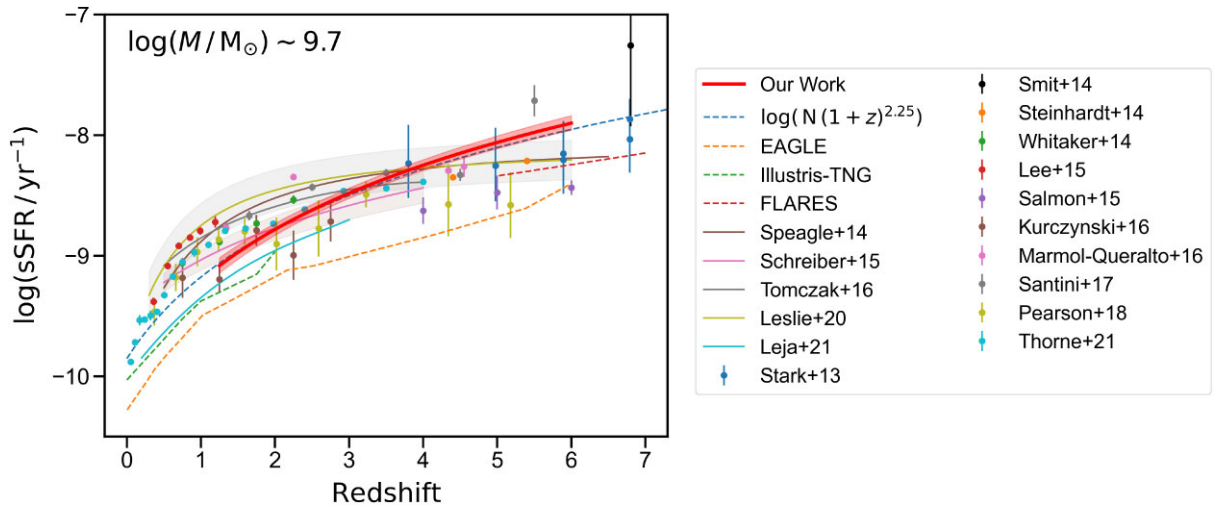
Figure 11. The evolution of log(sSFR) plotted as a function of redshift for galaxies with a stellar mass of $\sim 10^{9.7}\,\mathrm{M_\odot}$. Our work is shown by the solid red line with shaded 68-per cent credible intervals (as described in caption of Fig. 9). Coloured circles show previous observational results from the literature. The results for Leslie et al. (2020) and Tomczak et al. (2016) are taken from the relations fitted only to star-forming galaxies. Results from the EAGLE simulations (Ref-L100N1504 model) are taken from Furlong et al. (2015), Illustris-TNG from Donnari et al. (2019b), and FLARES from Lovell et al. (2021). The dashed blue line represents a simple functional form consistent with the evolution of the accretion rate of gas on to parent haloes, normalized to our work at $z = 2$.
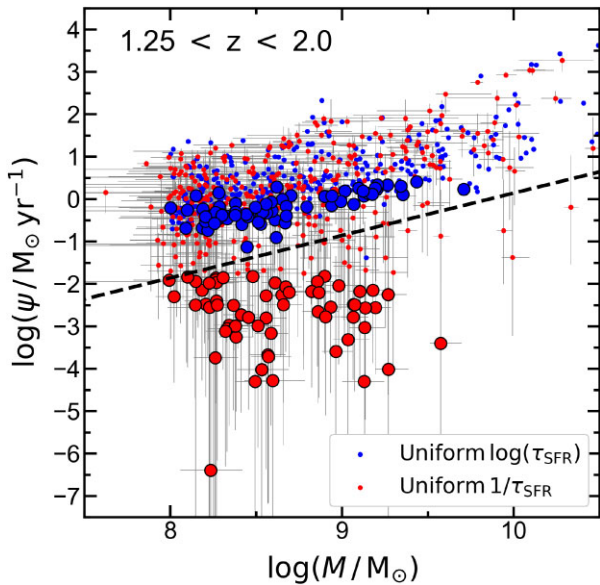
Figure 12. SFR versus stellar mass for objects within the $1.25 < z < 2$ bin measured with different priors on τ_{SFR}. We display the SFR and mass constraints before correcting for magnification to more clearly show the effects of the priors. The blue points show the measurements used in our fiducial model, fitted with a uniform prior on $\log(\tau_{SFR})$, while the red points show the results when fitted with a uniform prior on $1/\tau_{SFR}$. Grey error bars showing 68-per cent credible intervals in $\log(M/M_{\odot})$ and $\log(\psi/M_{\odot}\mathrm{yr}^{-1})$ are shown for the red points (for clarity, we do not plot uncertainties for the original estimates). Large circles show the objects that, when fit with uniform prior on $1/\tau_{SFR}$, give $\log(\psi/M_{\odot}\mathrm{yr}^{-1}) < -1.8$. The black dashed line shows the limit at which the prior density falls off quickly for the prior used in our fiducial model (see Fig. 13).

intervals). The prior in $M_{\star}-\Psi$ space is, in fact, a combination of how the priors on τ_{SFR} and t interact. When $t < \tau_{SFR}$, the SFH is in the rising portion, prior to the exponential decline. This rising history actually has a hard lower limit in $M_{\star}-\Psi$ shown by the dashed line in Figs 12 and 13 (middle and right-hand panels). The uniform prior on $\log(\tau_{SFR})$ has larger weight in high τ_{SFR} values (left-hand panel, Fig. 13), which puts higher weighting into the rising portion of the SFH. This can lead to uncertainties on Ψ that are small, suggesting that Ψ has been constrained by the data, when in fact the small uncertainties are caused by the informative prior on τ_{SFR}.

We used a redshift-dependent mass completeness cut (Fig. 4) to determine which objects we would use to measure the main sequence. This simple exercise demonstrates that for low-mass objects (still above the mass completeness cut), the signal-to-noise ratio is insufficient to constrain Ψ. To obtain results of the main sequence that are not dominated by the priors on SFH parameters, one needs to determine the mass limit at which $M_{\star}$ and Ψ are both constrained to a certain level of accuracy. Within this redshift bin, a by-eye assessment (from Fig. 12) suggests that a lower mass limit of $\log(M/M_{\odot}) \gtrsim 9.3 - 9.5$ would be appropriate, more than an order of magnitude higher than our mass-completeness limit.

An alternative approach that would allow deriving constraints on the main sequence to lower stellar mass, would involve censoring the data with poor constraints on Ψ. This might take the form of retaining only objects with Ψ above a lower limit which is defined by how well Ψ is constrained. This type of modelling is explicitly accounted for in the package LEO-PY, (Feldmann 2019), but the underlying model does not explicitly account for outliers. This is mitigated by Feldmann (2019) when fitting to the main sequence at

$0.01 < z < 0.05$ by modelling the distribution as asymmetrically distributed about the main sequence, allowing a tail to low Ψ to account for quiescent galaxies and those in transition. However, we have demonstrated that objects above the relation also need to be accounted for in an outlier model. It is beyond the scope of this paper to include modelling of censored data, which we defer to future work.

It is possible that the lower-limit in the prior for a rising SFH is reducing the slope we measure from the main sequence, since the region of low Ψ at low mass required to produce a steeper slope is relatively inaccessible thanks to the effective prior. This is likely why our study and that of Kurczynski et al. (2016), who use the same SFH and similar prior (priv. communication) on τ_{SFR}, measure a lower slope than Santini et al. (2017). Santini et al. (2017) measure $M_{\star}$ with a similar SFH to ours (but constrained to the rising portion at $z > 4$), but estimate Ψ indirectly using assumptions of a constant Ψ (via the Kennicutt & Evans 2012 UV-luminosity to Ψ calibration). This breaking of the dependence of Ψ and $M_{\star}$ on the same SFH can reduce the impact, somewhat, of hard lower- and upper-limits in the $M_{\star}-\Psi$ plane imposed by priors. However, the individual $M_{\star}$ and Ψ estimates are still limited by their respective priors and assumptions.

5.1.2 Form of SFH

Our chosen SFH is still very constraining in form; it ties the current SFR directly to the past star formation. There is an argument, often used, that the rest-frame ultra-violet varies with SFR on a timescale of ~ 100 Myr, and so broad-band photometry is not sensitive to short timescales of star formation. By this argument, short timescales do not have to be represented in the SFH when fitting only to broad-band photometry. This assumption is clearly incorrect at high redshift where emission lines (sensitive to Ψ on timescales ~ 10 Myr) have been demonstrated to significantly affect broad-band fluxes (Curtis-Lake et al. 2013; Stark et al. 2013; de Barros, Schaerer & Stark 2014; Smit et al. 2014; Curtis-Lake et al. 2021). New studies with more complex SFHs demonstrate how measurements made with simpler SFH prescriptions can be biased (Carnall et al. 2019; Leja et al. 2019a). Leja et al. (2019b) derive older ages and lower SFRs when fitting to multi-band photometry from the 3D-*HST* catalogues (Skelton et al. 2014) with a SFH described by discrete bins of star formation. This leads to a lower measurement of the normalization of the main sequence (Leja et al. 2021).

We have shown that our SFR estimates for objects with $\log(\psi/M_{\odot}\mathrm{yr}^{-1}) \lesssim 0$ in the $1.25 < z < 2$ bin are highly dependent on SFH priors. However, for objects with firm $M_{\star}$ and Ψ constraints with the DE SFH, it can be instructive to fit with a SFH with more freedom. We fit with a simple history that completely decouples the present SFR with the previous SFH (and hence with the accumulated stellar mass). This SFH describes the current star formation with a constant history over the last 10 Myr (with a uniform prior between $-4 < \log(\psi/M_{\odot}\mathrm{yr}^{-1}) < 4$) while earlier times are described by a DE. We label it DE SFH + BURST. The results are displayed in Fig. 14 as red points, with the 68-per cent credible regions in $M_{\star}$ and Ψ shown as grey error bars. We plot the constraints prior to correcting for magnification. The blue points show the original posterior median constraints. As expected, given more freedom in the SFH, a very large fraction of the objects have very poorly constrained Ψ (low SFR objects with large uncertainties in SFR), while some objects also have very poorly constrained masses (those with high Ψ that also have large uncertainties on $M_{\star}$). When further analysing the DE SFH + BURST sample in order to derive a main-sequence slope, intercept, and scatter, we determined that a cut of 1.0 dex in Ψ
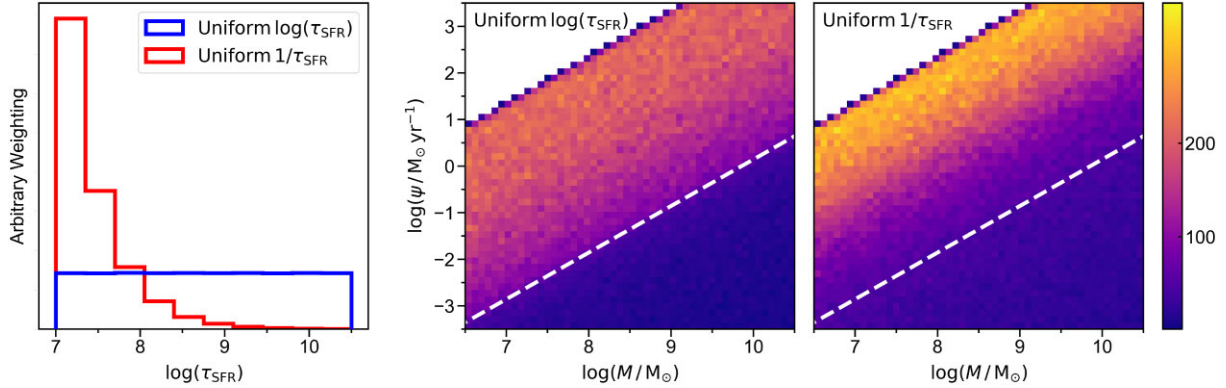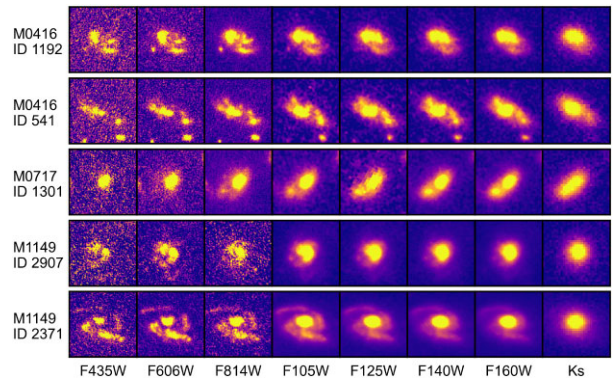
Figure 13. The two priors on τ_{SFR} employed when testing the dependence of the results in the $1.25 < z < 2$ bin on the priors employed in the fits. The fiducial prior employed for our fits is uniform on $\log(\tau_{\text{SFR}})$, and we also fit with a prior that is uniform on $1/\tau_{\text{SFR}}$. The two priors are plotted in the left-hand panel. The middle panel shows the prior probability weighting in the $M_{\star} - \Psi$ plane for our fiducial prior on τ_{SFR}, while the uniform prior on $1/\tau_{\text{SFR}}$ is shown in the right-hand panel. The colour-coding shows the weighting in the prior with arbitrary normalization. The white dashed line shows the lower limit in the prior space imposed by the rising portion of the DE SFH (see text for details). As can be seen, the weighting in our fiducial prior (middle) falls significantly below this line.
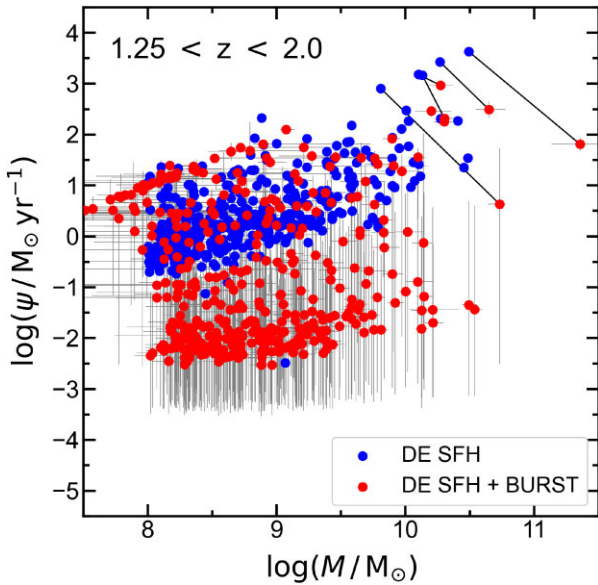
Figure 14. SFR versus stellar mass for objects in the $1.25 < z < 2$ bin for objects fitted with two different SFHs. As for Fig. 12, we display the SFR and mass constraints before correcting for magnification as it more clearly shows the effects of unconstrained parameters. The blue points show posterior medians measured when fitting with a delayed exponential (DE) SFH, while the red points show the posterior medians measured with a DE SFH where the SFR within the most recent 10 Myr is constant, and allowed to vary independently of the previous history (DE SFH + BURST). For clarity, we show the 68 per cent credibility regions for the DE+BURST measurements. The black lines connect the measurements for five object originally identified as residing above the main-sequence relation with the DE SFH, that sit on or below the main sequence when fitted with a DE + BURST SFH.


Figure 15. Postage stamps of the five objects originally identified as outliers at high $M_{\star}$ and Ψ in the $1.25 < z < 6$ bin (see Fig. 14). The objects increase in BEAGLE-derived $M_{\star}$ (prior to magnification correction) with the DE SFH from top to bottom.

uncertainty would be necessary to remove the highly unconstrained objects. For the DE SFH + BURST sample in the $1.25 < z < 2$ redshift bin, this flagged 83 per cent of the objects. We therefore did not proceed to fit a main sequence to this subset. For comparison, only 13 per cent of the original DE SFH sample had uncertainties in Ψ greater than 1.0 dex.

For certain high SFR objects that are originally identified as outliers above the main sequence, the DE SFH could not allow

for a recent burst of star formation without forcing the galaxy to be very young ($t \sim 10^{7.2}$). We demonstrate the trajectory of five such outliers when fitting with the more flexible SFH by black lines connecting the original posterior medians (blue points) with the new estimates (red points). When the SFH allows the current SFR to be decoupled from previous SFH, these objects are either found to lie in regions consistent with the measured main sequence extrapolated to higher $M_{\star}$ (4/5), or have very poorly constrained SFR (1/5). These results suggest that the DE SFH did not encompass the true SFR of the objects, leading to the objects being incorrectly interpreted as outliers from the main sequence. We show the postage stamp cut-outs for these objects in Fig. 15, which show no strong evidence for the very young ages derived with the DE SFH.

From so few objects with firm constraints with the more flexible SFH, we cannot comment on the likely bias on the measurements of the main sequence due to our original choice. However, we have demonstrated that fitting with a more complex SFH would have been unfeasible for our sample given the current broad-band constraints without firm priors. Leja et al. (2019b) used more flexible SFHs in their analysis, but demonstrated that the results are dependent on the priors on their SFH in Leja et al. (2019a). They chose a 'continuity' prior that down-weights sharp transitions between bins of star
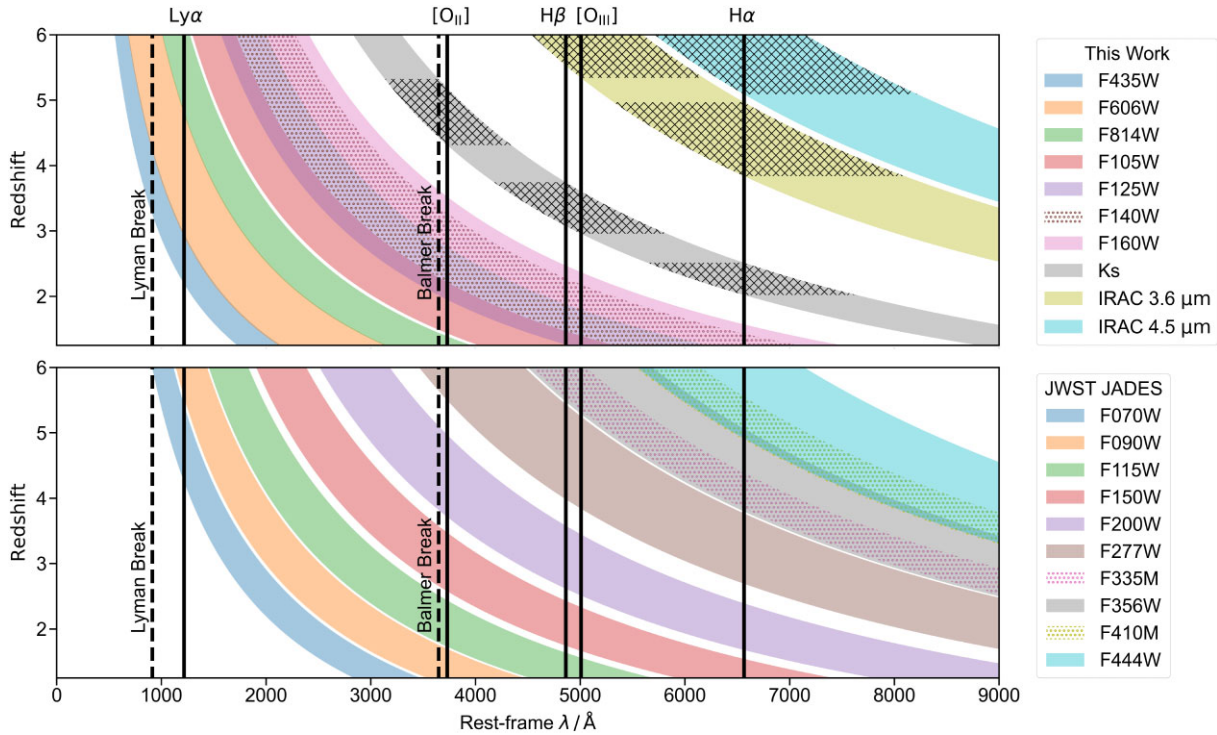
Figure 16. Redshift versus rest-frame wavelength showing the filter coverage over key spectral features. The dashed black lines from left to right show the Lyman and Balmer breaks. The solid black lines from left to right represent Ly α, [OII]λλ3726Å, 3729Å, H β, [OIII]λ5007Å, and H α. Top panel: The shaded regions represent the ASTRODEEP filter-set. The dotted area shows the F140W filter which overlaps the F125W and F160W filters. The black hatched regions show the redshifts at which the *K*s and IRAC filters are contaminated by strong emission lines. Bottom panel: The shaded regions represent the JADES broad-band filter-set. The dotted regions show the two JADES medium-band filters, F335M, and F410M.

formation. This prior is somewhat justified by the demonstration that the chosen SFH+priors brings the observed cosmic SFR density and stellar mass growth into agreement for the first time. This means that on average the prior is appropriate at the redshift studied ($z \lesssim 2.5$). However, it is not clear that these priors are suitable for determining the impact of bursty star formation on the form of the main sequence at higher redshifts ($z \gtrsim 2.5$), where short time-scale star formation can significantly affect the broad-band fluxes via the emission lines. Ideally we should first use data to determine how bursty the star formation is in systems at high redshift to settle on a suitable prior. Our demonstration here shows that the data set used in this study is not appropriate for this purpose for most galaxies in the sample. We must await *JWST* data sets, with medium-band filters spanning the rest-frame continuum, and rest-frame optical spectroscopy constraining the emission lines, and even continuum emission with the lower resolution mode.

5.2 The limitations of the filter-set and prospects for main-sequence measurements in the future

The filter-set used in this work is close to the optimum available data to study the low-mass end of the main sequence at high redshifts before the advent of *JWST*. The addition of HAWK-I *K*s-band provides a vital data point between the reddest *HST* filter and bluest *Spitzer* filter.

Fig. 16 displays the filter coverage over key spectral features as a function of redshift for the ASTRODEEP (top panel) and the *JWST* Advanced Deep Extra-galactic Survey (JADES) filter-set (bottom panel). The Lyman and Balmer breaks are shown as

dashed lines, while key emission lines are shown as solid lines with corresponding labels. Regions where the *K*s-band or IRAC 3.6- and 4.5-μm filters are contaminated by bright emission lines (either [OII]λλ3726Å, 3729Å, [OIII]λ5007Å, H β or H α) are shown as hatched regions. To derive firm M⋆ and Ψ estimates requires a firm photometric redshift estimate as well as filters sampling the rest-frame ultra-violet to rest-frame optical. In particular, the rest-frame optical should provide constraints of the stellar continuum free from contamination by bright emission lines. One would ideally also have reasonable constraints on the shape of the Balmer break. The firm photometric redshift can come from filters bracketing either the Lyman or Balmer breaks.

For the ASTRODEEP filter-set, the $3 < z < 4$ bin has two *Spitzer* filters probing the rest-frame optical free from emission-line contamination, but the Lyman break is passing through the lowest wavelength filter (F435W) and the Balmer break strength and position is muddied by the contamination of [OIII]λ5007Å and H β to the *K*s-band. Both these effects significantly reduce the accuracy of the photometric redshift constraints, which impacts the constraints on M⋆ and Ψ, and also the constraints on σ, as explained in Section 3.1.2. By $z > 4$, the Lyman break is securely bracketed by two filters, thereby improving the photometric redshift constraints, and the 4.5-μm filter provides constraints of the stellar continuum in the rest-frame optical. This provides firmer M⋆ and Ψ constraints, and explains why our scatter estimate at $4 < z < 5$ is no-longer significantly underestimated in the redshift bin analysis (see Fig. 6).

The bottom panel of Fig. 16 shows the coverage from an example *JWST* NIRCam filter-set, that was chosen for JADES. The coverage from the two longest wavelength broad-band filters (F356W and

F444W) appears similar to that of IRAC 3.5 and 4.5 μm. However, imaging with the F356W and F444W filters will have a far greater depth and resolution, which in turn minimizes the uncertainties arising from the deconfusion process. The F200W and F277W filters provide wavelength coverage across the gap between the *HST* and *Spitzer* filters in the ASTRODEEP catalogue. Additionally there are two medium-band filters (F335M and F410M), which in total provides six filters red-ward of the Balmer break, significantly mitigating the issue of emission line contamination. There are considerably more medium and narrow band filters to choose from, and part of the area covered by JADES (specifically the Hubble Ultra Deep Field) will also be visited by proposal 1963 (Williams et al. 2021). This imaging will provide additional medium band filters (F182M, F210M, F430M, F460M, and F480M), further sampling the rest-frame optical of high redshift galaxies. This extra imaging will provide multiple anchors probing the stellar continuum free from emission line contamination.

It is worth noting that the addition of optical *HST* ACS photometry is still beneficial in determining robust photometric redshifts where the JADES filter-set does not bracket the Lyman break, as well as for constraining the rest-frame UV continuum at low redshifts.

6 CONCLUSIONS

We used BEAGLE to fit to photometry in the ASTRODEEP catalogue for the first four Frontier Field clusters: Abell 2744, MACS0416, MACS0717, and MACS1149. Gravitational lensing due to the large foreground clusters has enabled us to probe masses as low as $10^7 - 10^8\,\mathrm{M_\odot}$, between redshifts $1.25 < z < 6.0$.

We have presented a Bayesian hierarchical model of the star-forming main sequence which accounts for the heteroskedastic, co-varying errors on stellar mass, SFR, and redshift, as well as the presence of outliers.

To determine a suitable parametrization for our full model, we initially fitted the main-sequence relation within different redshift bins. Our initial analysis demonstrated that the sampling of galaxy SEDs provided by the filter-set used (representing the best filter-set currently available for probing faint galaxies at high redshift) provides $M_\star$ and Ψ estimates that are too poorly constrained to warrant fitting with a fully flexible model. We describe here the decisions made and results for the redshift-dependent model.

(i) We fit with a slope that is constant with redshift, measuring $\beta = 0.79\pm^{0.03}_{0.04}$.

(ii) We choose a physically motivated parametrization for the evolution in the normalization of the main sequence, based on the expected evolution of accretion rate of gas on to the parent haloes, with the form $\alpha_{9.7}(z) = \log(N(1 + z)^\gamma) + 0.7$. We measure $N = 0.12^{+0.04}_{-0.03}$ and $\gamma = 2.40^{+0.18}_{-0.18}$. The value of γ is consistent with the value expected if sSFR scales with accretion on to dark matter haloes (a value of 2.25) and the data is consistent with a rising sSFR to high redshifts.

(iii) Having removed the majority of outliers located below the main sequence due to their highly unconstrained measurements of SFR, we account for outliers at $z < 4$ by modelling them simply as belonging to a broad Gaussian distribution in Ψ with mean and standard deviation constant with redshift, as well as the probability of an object being an outlier.

(iv) For $z > 4$, we set the probability of outliers to zero finding no strong evidence for them from the redshift bin results.

(v) We find that intrinsic scatter about the main sequence is highly degenerate with the outlier model parameters, and cannot be accurately determined separately within the $3 < z < 4$ bin. For the full model, we resort to fitting a scatter that is constant with redshift, and measure an intrinsic scatter (deconvolved from uncertainties on $M_\star$ and Ψ) of $\sigma = 0.26^{+0.02}_{-0.02}$.

We have explored the limitations of the data and demonstrated how to diagnose when the data may be insufficient to constrain the star-forming main sequence without significant biases. We re-fitted the galaxies in the $1.25 < z < 2$ bin (those galaxies in our sample with the most complete sampling of their SEDs, and therefore likely the best physical parameter constraints) in two ways. First, with a different prior on τ_{SFR}, which describes the time-scale of decay in our delayed exponentially declining SFH. Our results show that with our fiducial prior, the $M_\star$ and Ψ estimates appeared well-constrained, yet when the prior is changed, it shows that objects which were originally fitted with $\log(\psi/\mathrm{M_\odot\,yr^{-1}}) \lesssim 0.0$ give much lower SFR estimates. The fiducial prior was therefore somewhat informative and veiling which objects had poorly constrained SFR. We also re-fit galaxies in the $1.25 < z < 2$ bin with a less constraining SFH that allowed the recent 10 Myr of constant star formation to vary independently of the previous SFH. We demonstrate how few objects had well-constrained $M_\star$ and Ψ estimates with this history, meaning that in order to fit more complex and realistic SFHs, we first require an improved data set with better constraints.

The improved sampling of the SED that can be achieved with *JWST* NIRCam broad and medium-band filters, as well as the consistent depth that can be achieved will significantly improve the constraints on the main sequence at high redshifts.

DATA AVAILABILITY

The data and code underlying this article are available at https://github.com/ls861/M-SFR-Sandles2022. The full ASTRODEEP catalogue (Merlin et al. 2016a; Castellano et al. 2016; Di Criscienzo et al. 2017) is publicly available at http://astrodeep.u-strasbg.fr/ff/.

REFERENCES

Baker W. M., Maiolino R., Bluck A. F. L., Lin L., Ellison S. L., Belfiore F., Pan H.-A., Thorp M., 2022, MNRAS, 510, 3622
de Barros S., Schaerer D., Stark D. P., 2014, A&A, 563, A81
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bhatawdekar R., Conselice C. J., 2021, ApJ, 909, 144
Birnboim Y., Dekel A., Neistein E., 2007, MNRAS, 380, 339
Boogaard L. A. et al., 2018, A&A, 619, A27
Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018, MNRAS, 480, 4379
Carnall A. C., Leja J., Johnson B. D., McLure R. J., Dunlop J. S., Conroy C., 2019, ApJ, 873, 44

Carton D. et al., 2017, MNRAS, 468, 2140

Castellano M. et al., 2016, A&A, 590, A31

Chabrier G., 2003, PASP, 115, 763

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Chevallard J., Charlot S., 2016, MNRAS, 462, 1415

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

Curtis-Lake E. et al., 2013, MNRAS, 429, 302

Curtis-Lake E., Chevallard J., Charlot S., Sandles L., 2021, MNRAS, 503, 4855 (CL21)

Dekel A. et al., 2009, Nature, 457, 451

Di Criscienzo M. et al., 2017, A&A, 607, A30

Donnari M. et al., 2019a, MNRAS, 485, 4817

Donnari M. et al., 2019b, MNRAS, 489, 3036

Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690

Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267

Feldmann R., 2019, A&C, 29, 100331

Furlong M. et al., 2015, MNRAS, 450, 4486

Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, Bayesian Data Analysis, Third Edition. CRC Press, Boca Raton

Gutkin J., Charlot S., Bruzual G., 2016, MNRAS, 462, 1757

Han Y., Han Z., 2012, ApJ, 749, 123

Han Y., Han Z., 2014, ApJS, 215, 2

Han Y., Han Z., 2019, ApJS, 240, 3

Hogg D. W., Bovy J., Lang D., 2010, preprint (arXiv:1008.4686)

Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, MNRAS, 442, 1805

Iyer K., Gawiser E., 2017, ApJ, 838, 127

Iyer K. G., Gawiser E., Faber S. M., Ferguson H. C., Kartaltepe J., Koekemoer A. M., Pacifici C., Somerville R. S., 2019, ApJ, 879, 116

Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021, ApJS, 254, 22

Katsianis A. et al., 2020, MNRAS, 492, 5592

Kelly B. C., 2007, ApJ, 665, 1489 (K07)

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Koprowski M. P., Dunlop J. S., Michalowski M. J., Cirasuolo M., Bowler R. A. A., 2014, MNRAS, 444, 117

Koprowski M. P. et al., 2016, MNRAS, 458, 4321

Kurczynski P. et al., 2016, ApJ, 820, L1

Laigle C. et al., 2016, ApJS, 224, 24

Le Fèvre O. et al., 2015, A&A, 576, A79

Lee N. et al., 2015, ApJ, 801, 80

Leja J., Johnson B. D., Conroy C., van Dokkum P. G., Byler N., 2017, ApJ, 837, 170

Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019a, ApJ, 876, 3

Leja J. et al., 2019b, ApJ, 877, 140

Leja J. et al., 2021, preprint (arXiv:2110.04314)

Leslie S. K. et al., 2020, ApJ, 899, 58

Lin L. et al., 2019, ApJ, 884, L33

Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2021, MNRAS, 500, 2127

Merlin E. et al., 2015, A&A, 582, A15

Merlin E. et al., 2016a, A&A, 590, A30

Merlin E. et al., 2016b, A&A, 595, A97

Mármol-Queraltó E., McLure R. J., Cullen F., Dunlop J. S., Fontana A., McLeod D. J., 2016, MNRAS, 460, 3587

Neistein E., Dekel A., 2008, MNRAS, 383, 615

Noeske K. G. et al., 2007, ApJ, 660, L43

Pearson W. J. et al., 2018, A&A, 615, A146

Robotham A. S. G., Obreschkow D., 2015, PASA, 32, e033

Robotham A. S. G., Bellstedt S., Lagos C. D. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, MNRAS, 495, 905

Salmon B. et al., 2015, ApJ, 799, 183

Santini P. et al., 2017, ApJ, 847, 76

Schreiber C. et al., 2015, A&A, 575, A74

Skelton R. E. et al., 2014, ApJS, 214, 24

Smit R. et al., 2014, ApJ, 784, 58

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, ApJ, 763, 129

Steinhardt C. L. et al., 2014, ApJ, 791, L25

Tasca L. a. M. et al., 2015, A&A, 581, A54

Thorne J. E. et al., 2021, MNRAS, 505, 540

Tomczak A. R. et al., 2016, ApJ, 817, 118

Vidal-García A., Charlot S., Bruzual G., Hubeny I., 2017, MNRAS, 470, 3532

Whitaker K. E. et al., 2014, ApJ, 795, 104

Williams C. C. et al., 2018, ApJS, 236, 33

Williams C. C. et al., 2021, JWST Proposal, Cycle 1. p. 1963

Yang G. et al., 2020, MNRAS, 491, 740

This paper has been typeset from a T$_E$X/L^{A}T$_E$X file prepared by the author.