Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **525,** 1535–1552 (2023) https://doi.org/10.1093/mnras/stad2315
Advance Access publication 2023 August 3

Energy balance SED modelling can be effective at high redshifts regardless of UV-FIR offsets

P. Haskell,[1]★ D. J. B. Smith [1], R. K. Cochrane [2], C. C. Hayward [2] and D. Anglés-Alcázar[2,3]

[1]*Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*
[2]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA*
[3]*Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA*

Accepted 2023 July 11. Received 2023 June 27; in original form 2023 January 4

ABSTRACT

Recent works have suggested that energy balance spectral energy distribution (SED) fitting codes may be of limited use for studying high-redshift galaxies for which the observed ultraviolet and far-infrared emission are offset (spatially 'decoupled'). It has been proposed that such offsets could lead energy balance codes to miscalculate the overall energetics, preventing them from recovering such galaxies' true properties. In this work, we test how well the SED fitting code MAGPHYS can recover the stellar mass, star formation rate (SFR), specific SFR, dust mass, and luminosity by fitting 6706 synthetic SEDs generated from four zoom-in simulations of dusty, high-redshift galaxies from the FIRE project via dust continuum radiative transfer. Comparing our panchromatic results (using wavelengths 0.4–500 μm, and spanning $1 < z < 8$) with fits based on either the starlight ($\lambda_{\rm eff} \leq 2.2\,\mu$m) or dust ($\geq 100\,\mu$m) alone, we highlight the power of considering the full range of multiwavelength data alongside an energy balance criterion. Overall, we obtain acceptable fits for 83 per cent of the synthetic SEDs, though the success rate falls rapidly beyond $z \approx 4$, in part due to the sparser sampling of the priors at earlier times since SFHs must be physically plausible (i.e. shorter than the age of the universe). We use the ground truth from the simulations to show that when the quality of fit is acceptable, the fidelity of MAGPHYS estimates is independent of the degree of UV/FIR offset, with performance very similar to that previously reported for local galaxies.

Key words: methods: data analysis – galaxies: fundamental parameters – galaxies: high-redshift.

1 INTRODUCTION

Spectral energy distribution (SED) fitting offers a powerful method of estimating galaxy physical properties from photometry. SED fitting programs take as input the available photometry, which can be >30 bands in the best studied fields to <10 elsewhere, then use models of varying complexity to infer the shape of the full SED and hence the underlying physical properties (for an introduction to SED fitting see for example Walcher et al. 2011; Conroy 2013).

The energy balance code MAGPHYS (da Cunha, Charlot & Elbaz 2008 – hereafter DC08) performs χ^2 fitting using two sets of pre-built libraries of model SEDs with a representative range of SFHs and dust models for star-forming galaxies. The energy balance criterion works in such a way that MAGPHYS considers only combinations of SFH and dust emission that are energetically consistent, in the sense that the energy absorbed by dust in the rest-frame UV is re-radiated in the FIR. During the fit, MAGPHYS finds the SFH and dust model that best fits the data, and calculates probability density functions (PDFs) for a variety of property values by marginalizing over all of the models which satisfy the energy balance criterion.

To determine the fidelity of the properties derived from SED fitting, three testing techniques have been used in previous studies. The first is to compare the derived physical parameters to those derived using simpler methods. DC08 tested how well MAGPHYS could fit observations from the *Spitzer* Infrared Nearby Galaxy Survey (SINGS; Kennicutt et al. 2003), producing acceptable best fit χ^2 results for 63 of the 66 galaxies. They also tested how well MAGPHYS could recover the properties of 100 of its own, randomly selected, models with noise added to the photometry. Here, $M_{\rm star}$, SFR, and $L_{\rm dust}$ were reported to be recovered to a high degree of accuracy. Similarly, Noll et al. (2009) tested the alternative energy balance SED fitting code CIGALE (Boquien et al. 2019) using the SINGS galaxies, replacing DC08's UBV observations with those from Muñoz-Mateos et al. (2009). Here, $L_{\rm dust}$ estimates compared well ($\pm$0.03 dex) with those derived by Draine et al. (2007), similarly the SFR estimates compared well (0.06 ± 0.05 dex) with those provided by Kennicutt (1998b) based on Hα emission (e.g. Kennicutt 1998a).

An alternative testing technique is to compare the results of different fitting programs when applied to the same data set. This will not provide evidence that the results are correct, but does give confidence that a given code performs similarly to others. Best et al. (2022 – in preparation) tested three energy balance based fitters – MAGPHYS, CIGALE, and BAGPIPES (Carnall et al. 2018) – together with AGNFITTER (Calistro Rivera et al. 2016). The four codes were each used to estimate $M_{\rm star}$ and SFR for galaxies in the Boötes, Lockman Hole, and ELAIS-N1 fields of the LOFAR Two Metre Sky

★ E-mail: paul.haskell@talk21.com

Survey (LoTSS Shimwell et al. 2017) deep fields first data release (Duncan et al. 2021, Kondapally et al. 2021, Sabater et al. 2021 and Tasse et al. 2021). The results of the runs were compared to determine how well they agreed with each other. For galaxies with no AGN, MAGPHYS, CIGALE, and BAGPIPES typically agreed to within 0.1 dex for stellar mass, with AGNFITTER differing by 0.3 dex. Similar levels of agreement were found for the SFRs of galaxies found not to contain an AGN. For galaxies with an AGN the situation was more mixed as neither MAGPHYS nor BAGPIPES are designed to handle AGN emission.

Hunt et al. (2019) compared the results of applying MAGPHYS, CIGALE, and GRASIL (Silva et al. 1998) to a sample of 61 galaxies from the Key Insights on Nearby Galaxies: a Far-Infrared Survey with *Herschel* (KINGFISH) survey (Kennicutt et al. 2011), including 57 of the SINGS galaxies. They found that stellar masses estimated using 3.6 μm luminosity agreed with all three codes to within 0.2 dex. Similarly, SED derived SFR estimates were within 0.2 dex of those derived using FUV + TIR luminosities and $H\alpha$ + 24 μm luminosities. The results for $M_{\rm dust}$ were more mixed, with GRASIL giving values 0.3 dex higher than MAGPHYS or CIGALE or the value determined using a single temperature modified blackbody. A similar approach with an even broader selection of fourteen SED fitting codes was taken by Pacifici et al. (2023), who found agreement on stellar mass estimates across the ensemble, but some discrepancies in their SFR and dust attenuation results. More recently, Cheng et al. (2023) used a modified version of MAGPHYS (MAGPHYS + photo-z; Battisti et al. 2019) to determine the photometric redshifts of 16 submillimetre galaxies (SMGs). The results were compared to the redshifts derived using EAZY (Brammer, van Dokkum & Coppi 2008), finding that for most sources the results were consistent.

The final, and perhaps most promising technique for validating SED fitting is to use simulated galaxies where the 'right' answer is known in advance. Wuyts et al. (2009) used the HYPERZ (Bolzonella, Miralles & Pelló 2000) SED fitting code on GADGET-2 (Springel 2005) simulations to recover mass, age, E(B-V), and A_V under a variety of conditions. They concluded that recovery of properties for ellipticals was generally good (residuals between 0.02 and 0.03 dex) with slightly poorer results for discs (residuals of 0.03 to 0.35 dex), with residuals increasing further to 0.02 to 0.54 dex during periods of merger-triggered star formation. Hayward & Smith (2015, hereafter HS15) used MAGPHYS on two GADGET-3 (Springel 2005) simulations of an isolated disc and a major merger of two disc galaxies at $z = 0.1$. Snapshots were taken at 10 Myr intervals and the radiative transfer code SUNRISE (Jonsson 2006) used to produce observations from seven different lines-of-sight around the simulation. In both scenarios, the attenuated SED was recovered with an acceptable fit (χ^2 within the 99 per cent confidence threshold; see Smith et al. 2012 for details) except for the time around the peak starburst/coalescence phase of the merger simulation. In both scenarios, $L_{\rm dust}$ was recovered well with $M_{\rm star}$ recovered to within 0.3 dex and SFR within 0.2 dex. $M_{\rm dust}$ was recovered less well, but still within 0.3 dex for the isolated galaxy and 0.5 dex for the merger. The conclusion from this study is that these properties of local galaxies can typically be recovered to within a factor of 1.5–3. Smith & Hayward (2018) studied a resolved simulated isolated disc, using spatial resolution as fine as 0.2 kpc. They found that MAGPHYS produced statistically acceptable results for $M_{\rm star}$, $L_{\rm dust}$, SFR, sSFR, and A_V for over 99 per cent of pixels within the r-band effective radius. At higher redshifts, Dudzevičiūtė et al. (2020, hereafter D20), used EAGLE (Schaye et al. 2015, Crain et al. 2015) simulations with SKIRT generated photometry (Baes et al. 2011, Camps & Baes 2020) to validate the performance of MAGPHYS for studying galaxies with redshifts up to 3.4. They found that MAGPHYS gave a remarkably linear correlation with the true (simulated) values, though with significant scatter (at the level of 10, 15, and 30 per cent for the dust mass, SFR and stellar masses, respectively) and significant systematic offsets (of up to 0.46 ± 0.10 dex for the recovered stellar mass).

These studies all provide evidence that SED fitting, particularly energy balance SED fitting, is working remarkably well and providing results often consistent with the ground truth once the uncertainties are accounted for.

However, several authors have questioned whether using an energy balance criterion is appropriate when viewing galaxies for which the UV and FIR are spatially offset from one another (e.g. Casey et al. 2017; Miettinen et al. 2017; Simpson et al. 2017; Buat et al. 2019). In such cases, while 'energy balance' is still expected overall (i.e. energy conservation is presumably not violated), significant spatial decoupling may lead to difficulties in recovering the true properties. Under such circumstances, the attenuation – and thus the intrinsic UV luminosity – may be underestimated because the UV-bright, relatively dust-free regions can result in a blue UV-optical slope even if the bulk of the young stars are heavily dust-obscured.

This concern has recently become testable with the sub-arcsecond resolution provided by the Atacama Large Millimetre/submillimetre Array (ALMA).[1], enabling direct observation of UV/optical and FIR offsets. There are now numerous papers reporting spatial offsets. Hodge et al. (2016), Rujopakarn et al. (2016), Gómez-Guijarro et al. (2018) and Rujopakarn et al. (2019) have discovered kpc offsets between star forming regions and centres of stellar mass while investigating the star formation and dust distributions in $2 < z < 4.5$ galaxies. Along these lines, Chen et al. (2017) found a significant offset in ALESS67.1, a SMG at $z = 2.12$, Cochrane et al. (2021) reported the same in the massive star-forming galaxy SHiZELS-14 at $z = 2.24$, and Bowler et al. (2018) detected a 3 kpc offset between the rest-frame FIR and UV emission in the Lyman-break galaxy ID65666 at $z \approx 7$.

The concern over the impact of decoupling between the dust and starlight is such that new SED fitting codes such as `MICHI2` (Liu 2020) and `Stardust` (Kokorev et al. 2021) mention the *absence* of energy balance as a key advantage in favour of using these codes for studying galaxies where spatial offsets are likely to be a factor. In Liu et al. (2021), `MICHI2` produced results very similar to MAGPHYS and CIGALE for a sample of high redshift galaxies, with stellar mass and dust luminosity estimates obtained to within 0.2 – 0.3 dex of those obtained using the two energy-balance codes. Similarly, Kokorev et al. (2021) used `Stardust` to fit 5000 IR bright galaxies in the GOODS-N and COSMOS fields, producing results which compared well with those derived using CIGALE with a mean $M_{\rm dust}$ residual of 0.09 dex, a mean $L_{\rm IR}$ residual of 0.2 dex and a mean $M_{\rm star}$ residual of 0.1 dex (albeit with a significant scatter of 0.3 dex).

An additional test of the likely impact of spatial offsets was conducted by Seillé et al. (2022), who used the CIGALE code to model the Antennae galaxy, Arp244, which is known to have very different UV and IR distributions (Zhang, Gao & Kong 2010). Seillé et al. (2022) found that the total stellar mass and SFR were consistent, whether they attempted to fit the integrated photometry of the galaxy or sum the results of fitting 58 different regions of Arp244 independently and summed the results (i.e. performance very similar to that found by Smith & Hayward 2018 for simulated galaxies without spatial offsets).

[1] http://www.alma.info

In this context, we now seek to further test the efficacy of energy balance SED fitting for these more challenging dusty, high redshift, star-forming galaxies by using high-resolution simulations with differing degrees of spatial offset between the apparent UV/FIR emission.

This paper is structured as follows. Section 2 describes the tools and methods used to create the observations and to fit the SEDs; Section 3 presents the results of the fitting including the derived values for several galaxy properties; Section 4 discusses these in the context of previous papers and Section 5 summarizes the conclusions. Throughout this work we adopt a standard cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2 METHOD

This section describes the simulation data and the creation of the synthetic observations. It also provides a brief introduction to MAGPHYS, details of the simulations, and how they were subsequently analysed.

2.1 Computing the SEDs of simulated galaxies

We analyse a set of 4 cosmological zoom-in simulations from the FIRE project[2] that were run using the FIRE-2 code (Hopkins et al. 2018) down to $z = 1$. The simulations use the code GIZMO (Hopkins 2015)[3], with hydrodynamics solved using the mesh-free Lagrangian Godunov 'MFM' method. Both hydrodynamic and gravitational (force-softening) spatial resolution are set in a fully-adaptive Lagrangian manner with fixed mass resolution. The simulations include cooling and heating from a meta-galactic background and local stellar sources from $T \approx 10 - 10^{10}$ K; star formation in locally self-gravitating, dense, self-shielding molecular, Jeans-unstable gas; and stellar feedback from OB and AGB mass–loss, SNe Ia and II, multiwavelength photo-heating and radiation pressure with inputs taken directly from stellar evolution models. The FIRE-2 physics, source code, and all numerical parameters are *exactly* identical to those in Hopkins et al. (2018).

The specific sample of simulations studied in this paper include the halos first presented in Feldmann et al. (2016). The FIRE-2 simulations for these halos were introduced, along with a novel on-the-fly treatment of black hole seeding and growth in Anglés-Alcázar et al. (2017). These halos were chosen because they are representative of the high-redshift, massive, dusty star-forming galaxies found in infrared-selected observational samples, Cochrane et al. (2019) showing that they present a clumpy dust distribution together with very different morphologies for stellar mass, dust, gas, and young stars. At $z = 2$, the galaxies central to the halos have half-light radii of 0.73, 0.98, 0.81, and 0.91 kpc; for additional information on these galaxies see Anglés-Alcázar et al. (2017) as well as Cochrane et al. (2019), Wellons et al. (2020), Parsotan et al. (2021) and Cochrane, Hayward & Anglés-Alcázar (2022).

To generate synthetic SEDs, Monte Carlo dust radiative transfer was performed on each time snapshot of the simulated galaxies in post-processing using the code SKIRT[4] SKIRT assigns single-age stellar population SEDs to star particles in the simulations according to their ages and metallicites. It then propagates photon packets through the simulated galaxies' ISM to compute the effects of dust absorption, scattering, and re-emission. Snapshots of the galaxies' evolution were taken at 15–25 Myr intervals with each galaxy 'observed' from 7 positions that uniformly sampled inclination angles from view 0 (aligned with the angular momentum vector) in steps of 30° to view 6 (anti-aligned). For full details of the SKIRT calculations, see Cochrane et al. (2019, 2022). This procedure yielded 6706 SEDs across the four simulated galaxies, spanning $1 < z < 8$.

Table 1. The filters used to create synthetic observations from the simulated photometry. The first column gives the telescope/survey, the second the instrument/filter name, and the third the effective wavelength of the filter.

Facility	Filter	$\lambda_{\rm eff}(\mu$m)
CFHT	Megacam u	0.39
PanSTARRS	g	0.48
PanSTARRS	r	0.61
PanSTARRS	i	0.75
PanSTARRS	z	0.87
PanSTARRS	y	0.96
UKIDSS	J	1.2
UKIDSS	K	2.2
Spitzer	IRAC ch1	3.4
Spitzer	IRAC ch2	4.5
Spitzer	IRAC ch3	5.6
Spitzer	IRAC ch4	8.0
Spitzer	MIPS 24 μm	24
Herschel	PACS green	100
Herschel	PACS red	160
Herschel	SPIRE	250
Herschel	SPIRE	350
Herschel	SPIRE	500

To compute photometry from the SEDs, we convolved the SEDs with appropriate filter response curves for the 18 bands listed in Table 1. These filters were chosen for similarity with previous work in the LoTSS deep fields (e.g. Smith et al. 2021), providing good coverage of the spectrum from the UV to the FIR with which to test how MAGPHYS performs in these idealized conditions. Fig. 1 shows the filter coverage for an example SED at $z = 1$, along with the emergent SED generated by SKIRT.

Fig. 2 examines the relationship between the properties of our simulated galaxies and those of high-redshift SMG populations in which spatial UV–FIR offsets have been observed. We compared four properties with observations, specifically the SFR relative to the galaxy main sequence (MS; upper left panel), the relationship between sub-mm flux density and $M_{\rm dust}$ (upper right), the degree of V band extinction (lower left), as well as the magnitude of the UV/IR offsets (lower right) in relation to studies in the literature. In the upper left panel we have compared the SFR in each snapshot with the MS parametrization from Schreiber et al. (2015) modified for our adopted Chabrier (2003) IMF using the method of Madau & Dickinson (2014), as a function of redshift. The magenta band indicates the typical ± 0.3 dex scatter associated with the MS (e.g. Tacchella et al. 2022). The simulated galaxies lie either on or above the MS in the vast majority of cases, and are therefore consistent with dusty, star-forming galaxies. The upper right panel of Fig. 2 shows the submillimetre flux density, S_{870}, as a function of the dust mass for the simulated galaxies and for the SMGs published in D20. While the simulations do not occupy the parameter space of the brightest SMGs, there is significant overlap, and they do lie along the same submm/dust mass relationship (see Hayward et al. 2011,

[2] http://fire.northwestern.edu
[3] http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html
[4] http://www.skirt.ugent.be/

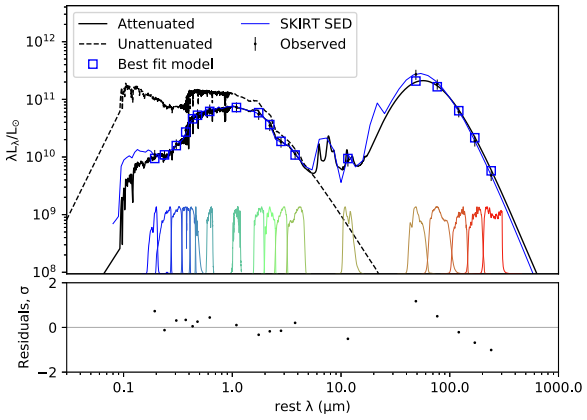


Figure 1. An example SED obtained using MAGPHYS, demonstrating the generally close agreement between the true and MAGPHYS-derived SEDs. In the upper panel, the solid black line shows the best fit MAGPHYS-derived SED, while the dashed black line indicates the MAGPHYS estimate of the unattenuated SED; the solid blue line represents the attenuated SED generated by SKIRT. The square markers represent the best fit photometry, with the SKIRT photometry shown as the points with error bars (as described in the legend). The coloured lines above the lower horizontal axis show the normalized filter curves used in this study. The lower panel shows the residual value in σ units between each observation and the best fit SED. The residual value is calculated as (observed flux – model flux)/observed error. This SED corresponds to simulated galaxy A1, snapshot 276, view 0, $z = 1.00$.

Cochrane et al. 2023). The lower left panel shows how the V-band extinction (A_V) for the simulations (the blue solid line indicates the median, with shading indicating the values enclosed by the 16th and 84th percentiles of the distribution at each redshift) compares with the corresponding values for the SMG samples from D20 (in purple) and Hainline et al. (2011, indicated by the red points with error bars). Although the D20 sample is on average more obscured than our simulations, similarity to the Hainline et al. (2011) SMGs is evident. The lower right panel shows the range of offsets between the UV and FIR emission in redshift bins. The solid lines indicate the mean simulated offset (blue for peak-to-peak, red for light-weighted mean), with shaded regions indicating the area enclosed by the 16th and 84th percentiles at each redshift. The black, red and green symbols indicate ALMA sources from Rujopakarn et al. (2016), and Rujopakarn et al. (2019) and Lang et al. (2019). Finally, the short green line marks the mean offset from Lang et al. (2019) over 20 SMGs with $1.6 < z < 2.5$.

To summarize, Fig. 2 demonstrates that the simulated sources are pre-dominantly dusty star-forming galaxies. While the D20 SMG sample is more extreme, the degree of extinction and the magnitude of the UV-FIR spatial offsets in the simulations show significant overlap with values published in the literature. The simulations are therefore a useful testing ground for determining the extent of our ability to recover the true properties of galaxies with plausible UV–FIR offsets using MAGPHYS.

2.2 MAGPHYS

MAGPHYS is an SED modelling code using Bayesian inference to derive best fit SEDs as well as estimates (best fit, median likelihood, and probability distribution functions) for a wide range of galaxy properties. A full description can be found in DC08 and da Cunha et al. (2015), but we include a brief overview. MAGPHYS uses two libraries of model galaxies: the first, the library of star-formation histories (SFH), consists of 50 000 models each comprising a UV/optical SED and associated galaxy properties; the second, the dust library, comprises 25 000 models each with an IR SED and associated properties.

The SFH library is built using the IMF of Chabrier (2003) and the stellar population synthesis (SPS) model of Bruzual & Charlot (2003). Exponentially declining star formation histories are superposed with random bursts, in such a way that a burst of star formation has occurred in half of the SFH library models within the last 2 Gyr.

Common to both libraries is the use of the Charlot & Fall (2000) two-component dust model. In this model, stellar populations younger than 10 Myr are attenuated by a greater amount than older stellar populations, under the assumption that these young stars are still embedded within their 'birth clouds'. These stellar populations are subject to a total optical depth $\tau_{\rm BC} + \tau_{\rm ISM}$, whereas older populations 'see' an optical depth of only $\tau_{\rm ISM}$, from the diffuse ISM. Charlot & Fall (2000) define the optical depth seen by stellar emission as

$$\hat{\tau_\lambda} = \begin{cases} \hat{\tau}_\lambda^{\rm BC} + \hat{\tau}_\lambda^{ISM} & \text{for stars <10 Myr,} \\ \hat{\tau}_\lambda^{ISM} & \text{for stars} \geq 10\,\text{Myr,} \end{cases}$$

where $\hat{\tau}_\lambda$ is the total optical depth for λ, $\hat{\tau}_\lambda^{\rm BC}$ is the optical depth of the birth clouds and $\hat{\tau}_\lambda^{ISM}$ is the optical depth of the ISM. These latter two are defined in MAGPHYS such that

$$\hat{\tau}_\lambda^{\rm BC} = (1-\mu)\hat{\tau}_V(\lambda/5500\text{Å})^{-1.3}, \text{ and} \quad (1)$$

$$\hat{\tau}_\lambda^{ISM} = \mu\hat{\tau}_V(\lambda/5500\text{Å})^{-0.7}, \quad (2)$$

where $\hat{\tau}_V$ is the mean V band optical depth and μ represents the fraction of $\hat{\tau}_V$ arising from the ISM.

The dust library is built from three main components: emission from very small grains ($< 0.01\,\mu$m) which can reach high temperatures if they absorb a UV photon; large grains (between $0.01 - 0.25\,\mu$m) in thermal equilibrium with the interstellar radiation field; and polycyclic aromatic hydrocarbons (PAHs) which are responsible for emission line features in the mid-infrared. The contribution of each component to the SEDs of the birth clouds and the ISM is chosen to broadly reproduce the range of SEDs found in nearby star-forming galaxies. The total IR SED is then modelled as the sum of the ISM and the birth cloud components.

The SFH and dust libraries are linked together in such a manner that the starlight absorbed by dust at short wavelengths is re-radiated at longer wavelengths, that is, the energy is balanced. During the fit, as well as ensuring that energy conservation (i.e. energy balance) is satisfied by construction (i.e. the luminosity absorbed by dust equals that emitted by dust), MAGPHYS combines those models in the optical library with those in the IR library that have similar contributions from dust in the ISM to the overall dust energy budget (the fraction of luminosity absorbed by the diffuse ISM component and that emitted by the diffuse ISM component, respectively). This is parametrized in MAGPHYS using the f_μ parameter; in the high-redshift version used in this work, values for the SFH and dust libraries must have $\Delta f_\mu < 0.2$ for the combination to be acceptable. In this way, each galaxy is fitted against a wide variety of 'empirical but physically motivated' (DC08) SFHs and dust content. By calculating the best fit χ^2 for each model combination that satisfies the conditions, a likelihood function is built for each galaxy property by assuming that $L \propto \exp(\frac{-\chi^2}{2})$. When all combinations of models in the libraries have been processed, a PDF is produced for each property by marginalizing the individual likelihoods. MAGPHYS outputs a pair of files for each fitted galaxy: one containing the best fit SED (an example of both the attenuated and

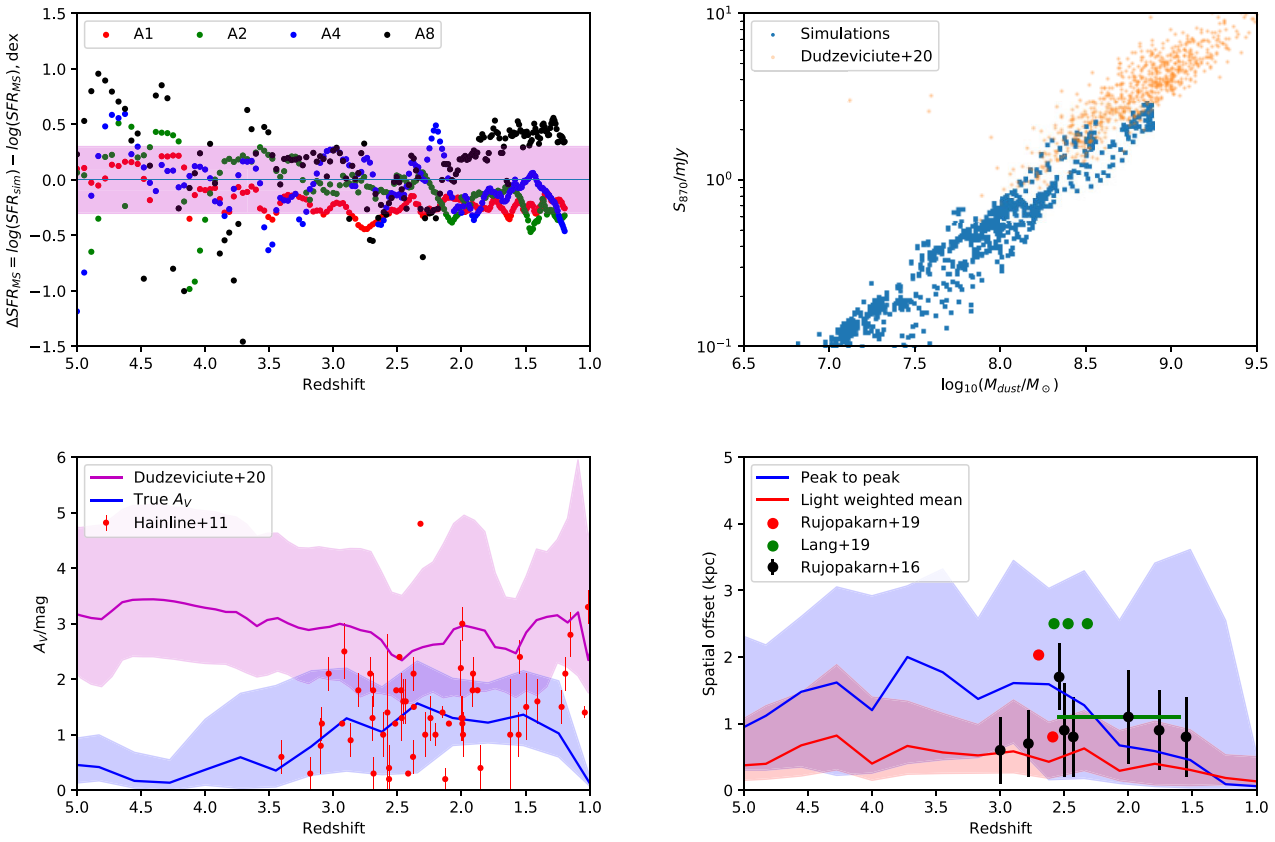


Figure 2. The properties of the simulated galaxies in their observational context. **Upper left:** the relationship between the simulated galaxies' SFRs and the galaxy main sequence (MS); for each snapshot, the *y*-axis shows the difference between the simulation SFR and the MS, with the magenta band indicating the typical ±0.3 dex scatter associated with the MS (e.g. Tacchella et al. 2022). **Upper right:** the relationship between the sub-mm flux density S_{870} and dust mass; the blue points represent the simulated data, while the orange points show galaxies from D20. **Lower left:** the variation in A_V as a function of redshift for the simulations (for which the median value at each z is shown by the solid line, within shading indicating values enclosed by the 16th and 84th percentiles of the distribution), along with a corresponding distribution from D20 (shown in purple). The SMG sample from Hainline et al. (2011) is shown by the red points. **Lower right:** the mean UV/FIR peak to peak (blue) and light-weighted mean (red) spatial offsets in redshift bins: the shading indicates the region enclosed by the 16th and 84th percentiles at each redshift, while the solid line indicates the median value. The red, green, and black circles are values for individual sources taken from the literature (as indicated in the legend), while the solid green line marks the reported average spatial offset across 20 SMGs from Lang et al. (2019).

unattenuated versions are shown, alongside the model photometry in Fig. 1), while the other contains the best fit model values and the PDFs.

This study uses the high-redshift version of MAGPHYS (da Cunha et al. 2015), which differs from the low-redshift version in two important ways: first, the prior distributions are modified to include higher dust optical depths, higher SFRs and younger ages; secondly, the effects of absorption in the intergalactic medium (IGM) at UV wavelengths are taken into account.

Some studies have sought to determine the extent to which AGN can influence the results of SED fitting (e.g. HS15, Best et al., in preparation). However, neither the simulations nor the SED fitting code used in this paper include AGN, and so this important aspect will not be discussed further.

2.3 Processing the data

To test how well MAGPHYS is able to recover the intrinsic properties of the simulated galaxies, we ran MAGPHYS four times on each synthetic SED, using different combinations of photometry and assumed redshift:

(i) run A – used all 18 filters;

(ii) run B – used all 18 filters, but with all SEDs shifted to a redshift of 2. This run was used as a comparison to detect any bias in the results due to redshift effects. This is discussed in Section 4.1;

(iii) Run C – used only the UV to near-IR filters (u – K);

(iv) Run D – used only the FIR filters (PACS 100 μm – SPIRE 500 μm).

Runs C and D are discussed in Section 3.2.3. We assumed a signal-to-noise ratio of 5 in every band, following Smith & Hayward (2018).

One of the key aims of this work is to determine how MAGPHYS performs when analysing galaxies for which the observed UV and FIR emission are spatially 'decoupled.' To do this, we characterize the offset between the UV and FIR emission in three different ways:

(i) the peak to peak offset: this is defined as the distance in parsecs between the points of maximum flux in the UV (0.3 μm) and FIR (100 μm) images;

(ii) the light-weighted mean offset: this is defined as the distance in parsecs between the light-weighted centres for the UV (0.3 μm) and FIR (100 μm) emission.

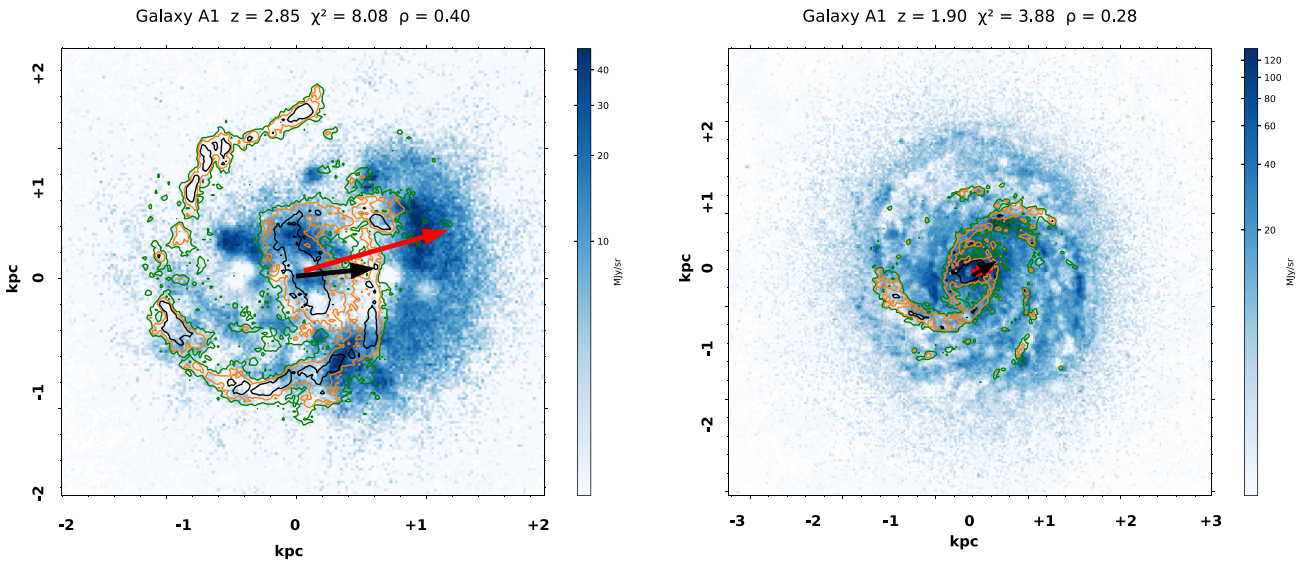


Figure 3. Visualizations of two views of galaxy A1, in the later stages of its evolution, showing differing degrees of UV–FIR offsets ranging from kpc-scale projected separation (left) to approximately co-spatial (right). In each panel, the image in blue shows the UV emission, the side colourbars showing the flux density of the emission in MJy/sr. The coloured contours show flux density for the FIR emission, ranging from green (3×10^4 MJy sr^{-1}), to orange (5×10^4 MJy sr^{-1}) to black (10^5 MJy sr^{-1}). In each panel, the base of the red vector is positioned at the peak FIR emission and the head at the peak UV emission, the base of the black vector is positioned at the light-weighted mean FIR emission and the head at the light-weighted mean UV emission. The title of each plot gives the galaxy name along with redshift, best fit χ^2 and Spearman ρ value.

(iii) the Spearman rank coefficient (Myers & Well 2003) comparing the degree of correlation between the UV (0.3 μm) image and the FIR (100 μm) image. A Spearman rank coefficient of $\rho > 0.8$ is considered necessary for a strong correlation. Spearman also returns a p value indicating a correlation confidence level, 99 per cent of our results returned p values indicating that the probability of the reported correlation being due to chance was <0.0001. The images were filtered to allow only the data points with intensity above the 80th percentile in either the UV or FIR images to be included in the analysis. This was done to avoid the comparatively very large number of low intensity pixels from unduly dominating the result. The 80th percentile was chosen as a reasonable value after comparing the results using different percentile values of the UV and FIR images by eye.

The three proxies were each calculated using the rest-frame UV and FIR maps for each snapshot and view to provide values that would be possible using real observational data with high enough spatial resolution and sensitivity. As an example, Fig. 3 shows two images of the simulated galaxy A1 in the later stages of its evolution, other examples can be seen in Cochrane et al. 2019. The image on the left shows a significant offset between the UV (shown as the blue image) and FIR (shown as contours) intensity, while in the right image (which has the same colour scheme) the UV and FIR appear almost coincident. In both panels the red vectors show the peak-to-peak offset, while the black vectors show the light-weighted offset. The Spearman ρ value is given in the title of each panel. We also calculated the offsets using the projected maps of the simulated young stars (age <10 Myr) and dust; however, there was no significant difference in the results and so the observed offsets are used throughout this paper.

In the following sections, where we compare derived values to true (simulated) values these are expressed as residuals in dex between the 50th percentile of the derived value's likelihood function and the true value

$$\Delta \log(\text{parameter}) = \log_{10}(\text{derived value}) - \log_{10}(\text{true value}). \quad (3)$$

It follows that positive offsets (Δ) represent MAGPHYS overestimates, and negative values indicate under-estimates. Throughout this work, where MAGPHYS results are shown averaged across the seven views of a snapshot, they are the mean of the individual median likelihood estimates.

3 RESULTS

In this section, we present results from the four runs described in Section 2.3. In all runs a successful fit was defined as one where the χ^2 value was equal to or below the 99 per cent confidence limit ($\chi^2_{\rm max}$), this was taken from standard χ^2 tables. The number of degrees of freedom was calculated as in Smith et al. (2012), which perturbed the output best fit SEDs from MAGPHYS with random samples from the standard normal distribution and found that it depended on the number of bands in the manner shown in Appendix B of that work. We are using the same MAGPHYS model and have assumed that the relation does not vary with the particular choice of bands or the redshifts of the sources being studied.

3.1 The fraction of mock observations with acceptable fits

From run A we find that MAGPHYS achieved a statistically acceptable fit (i.e. $\chi^2 \leq \chi^2_{\rm max}$) for 83 per cent (5567 out of 6706) of the snapshots. Note that the value of $\chi^2_{\rm max}$ varies with redshift because the SKIRT SEDs do not include wavelengths $< 0.08\,\mu$m, meaning that we are unable to generate synthetic photometry for the bluest filters at $z \gtrsim 3.9$.

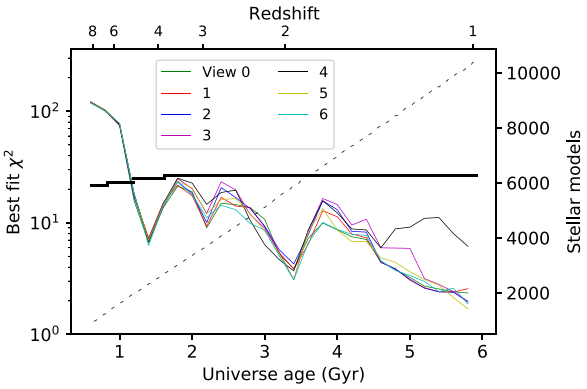


Figure 4. MAGPHYS produces statistically acceptable fits for virtually all snapshots at $z < 5$, irrespective of viewing angle. The best fit χ^2 as a function of universe age is shown for galaxy A1, colour-coded by view number. The χ^2 values have been averaged over bin widths of $\Delta z = 0.2$ (relative to the top horizontal axis) for clarity. The horizontal line indicates the χ^2 threshold below which a fit is deemed acceptable using the Smith et al. (2012) criterion, this value varies with redshift (see Table 2). The dashed line indicates the number of stellar models (relative to the right-hand y-axis) available to MAGPHYS at a given redshift with which to compare the input SED. Although not shown here, qualitatively similar results are obtained for the other simulations (A2, A4, and A8).

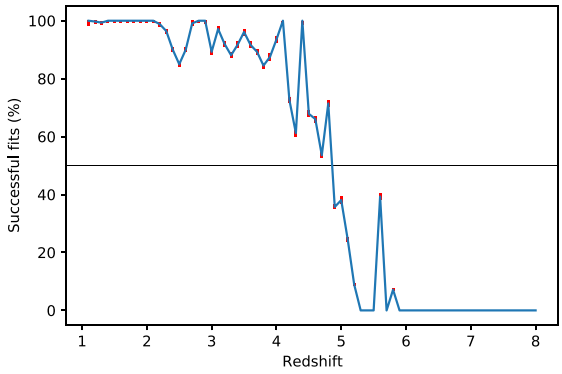


Figure 5. MAGPHYS success rate in fitting SEDs. The percentage of successful fits averaged across all views and snapshots of all galaxies as a function of redshift, note that standard Poisson errors are too small to be visible. The horizontal line marks a success rate of 50 per cent. The fraction of fits that are statistically acceptable decreases with increasing redshift due to the constraint that the SFH must be shorter than the age of the universe at that redshift, meaning that the size of the template library decreases with increasing redshift.

Table 2. The number of filters available and the value of $\chi^2_{\rm max}$ for different redshift ranges.

z	Filters	$\chi^2_{\rm max}$
$8.4 \leq z < 6.6$	15	21.67
$6.6 \leq z < 5.0$	16	23.01
$5.0 \leq z < 3.9$	17	24.75
$3.9 \leq z < 1.0$	18	26.72

The derived χ^2 values are broadly independent of viewing angle for all galaxies; as an example, Fig. 4 shows the χ^2 results for all snapshots and views for the galaxy A1. Fig. 5 shows how the fit success rate, averaged across all snapshots and views for all four galaxies, changes with redshift. We see from this that MAGPHYS can routinely produce acceptable fits to the synthetic photometry up to $z = 4$, but that the success rate drops to 50 per cent at $z \approx 4.85$ and to zero after $z \approx 5.9$. Different factors may be contributing to this effect. First, the number of SFHs from the MAGPHYS libraries that are compared with observations is a strong function of redshift. MAGPHYS does not consider SFHs longer than the age of the universe at a given redshift (the number of SFHs shorter than the Hubble time at each redshift is shown as the dashed line, relative to the right-hand axis in Fig. 4) and at $z \approx 5$ the number of such SFHs in the library is only 20 per cent of those available at $z \approx 1$. It is therefore clear that the prior is significantly more densely sampled at lower redshifts, leading to more acceptable fits in cases such as this, where the SFH itself is constrained only weakly by the photometry (e.g. Smith & Hayward 2018). Secondly, at these very early times in the simulations ($z > 5$), the model galaxies are low mass ($<10^9\,{\rm M}_\odot$) and bursts of star formation have a disproportionate influence on a galaxy's bolometric luminosity. This highly stochastic star formation is not well-modelled by the star formation histories included in the MAGPHYS libraries. It is possible that including additional bands of model photometry may provide better results, for example by an additional sub-millimetre data point providing an 'anchor' point to the Rayleigh–Jeans tail of the dust SED and in doing so enabling tighter constraints on the overall energy balance (though we note that the 500 μm band does sample this side of the dust SED out to $z \approx 4$). However, in this work we have chosen to focus on an example set of photometric data appropriate for studying dusty star-forming galaxies in general, and with an enforced SNR $= 5$ in every band we are not subject to some of the sensitivity (or resolution) limitations associated with using real *Herschel* data to study galaxies at the highest redshifts. We therefore defer testing our results with different photometric coverage for a future work. Throughout the remainder of this study, we follow the same approach used in previous MAGPHYS works both observational and numerical (e.g. HS15; Smith et al. 2012; Smith & Hayward 2018; Smith et al. 2021), and consider only those views for which an acceptable fit was obtained.

To investigate the influence of redshift on the MAGPHYS fit rate further, we used run B, in which the photometry is modified such that all SEDs were placed at $z = 2$. In this run, the size of the libraries and therefore the sampling of the priors used for SED fitting is the same for all snapshots. We find that the fit success rate increases to 93 per cent for the forced $z = 2$ runs, from 83 per cent for run A. Although it is tempting, we cannot attribute this change solely to the weakening of the SFH prior, since it is also possible that sampling different rest-frame wavelengths could impact the fit success rate (e.g. because of individual spectral features being redshifted into a particular observed bandpass; Smith et al. 2012). These effects are discussed further in Section 4.1.

3.2 Overall MAGPHYS performance

In studying the fidelity of the MAGPHYS parameter estimates, we have chosen to focus on five properties likely to be of the widest interest, namely SFR and sSFR (both averaged over the last 100 Myr), $M_{\rm star}$, $M_{\rm dust}$, and $L_{\rm dust}$. The true values for $M_{\rm star}$, SFR (averaged over the last 100 Myr), and $M_{\rm dust}$ were available from the simulation. The true values for $L_{\rm dust}$ were calculated by integrating under the SKIRT-

produced rest frame SED from $8\,\mu\mathrm{m} < \lambda < 1000\,\mu\mathrm{m}$, following Kennicutt (1998a).

3.2.1 *The fidelity of* MAGPHYS *results over time*

Fig. 6 shows the evolution in the true and derived physical properties of our simulated galaxies as a function of redshift (with a second horizontal axis at the top of each column showing the age of the universe at each redshift in our adopted cosmology). The different physical properties are shown along successive rows, while the different simulated galaxies are shown in successive columns, as indicated in the text at the top of each column. In each panel, the black line indicates the true values for each property, taken from the simulations, while the red line indicates the mean of the median-likelihood MAGPHYS estimates, where the averaging has been conducted over the seven different viewing angles. Similarly, the shaded red region in each panel indicates the area enclosed by the mean of the 16th and 84th percentiles of each parameter's MAGPHYS PDF (once more averaged over the seven views), to give the reader a feel for the typical error bar. Each lower panel shows the residual, for example $\Delta\log$ (SFR), as defined in equation (3).

In general, MAGPHYS-derived values show a significant degree of consistency, both in the temporal sense and by comparison to the true values. The temporal sense is a valuable test in its own right as, although MAGPHYS fits each snapshot independently, the true values shown in Fig. 6 mostly vary smoothly with time. That this is reflected in the MAGPHYS estimates once the error bars are taken in to account, offers broad encouragement for the use of MAGPHYS with observational data.

Further, we discuss the degree of fidelity in the MAGPHYS parameter estimates overall by comparing with the true (simulated) values. It is clear based on even a cursory inspection of the trends visible in Fig. 6 that the MAGPHYS estimates have broadly captured the behaviour visible in the true parameter values, such as increasing stellar mass and generally decreasing sSFR. Similar encouragement was found in the earlier work of HS15, though we now extend this to higher-redshift, dustier galaxies for the first time with a sample of very high-resolution simulations. The mean residuals, $\Delta\log$ (parameter), averaged over the full evolution of each simulated galaxy, are shown in Table 3.

Averaging the results across all views of all snapshots of all galaxies, we find that the stellar mass is typically underestimated by MAGPHYS, recovered with a mean residual of $\Delta\log(M_{\mathrm{star}}) = -0.29 \pm 0.09$. This 3.22σ result covers a wide range of simulated scenarios, ranging from the early stages of formation, through periods of starburst, tidal disruptions and merger events. By way of comparison, in HS15 the stellar mass was recovered to within 0.2 dex (which was also the typical uncertainty in that work) for the vast majority of snapshots, across both the isolated disc and major merger simulations. The principal exception to this excellent recovery being a 0.4 dex underestimate of the stellar mass during the peak period of AGN activity (which we do not simulate here). D20 also reported a larger systematic underestimation of stellar mass, with a deviation of -0.46 ± 0.10 dex; our results therefore fall between those of these two previous studies. We suggest two factors which may be contributing to this systematic underestimation of the stellar mass. First, a sub-optimal choice of SFH (such as we know we have made in this work, since we can see that the simulated galaxies do not have parametric SFHs in Fig. 6) has been shown to produce biased results (Carnall et al. 2019) and in particular an underestimate for stellar mass when applied to star-forming galaxies (Mitchell et al. 2013; Michałowski et al. 2014). Secondly, Mitchell et al. (2013) and Małek et al. (2018) have shown that the choice of attenuation law has an impact on the estimation on stellar mass (and it is also clear that the two-component geometry assumed by MAGPHYS is not consistent with the ground truth in the simulations where the radiative transfer calculates the attenuation due to ISM dust *in situ*).

In the second row of Fig. 6, we show that the MAGPHYS SFRs for our simulated galaxies are typically accurate to within $\Delta\log(\mathrm{SFR}) = -0.11 \pm 0.06$ of the true values (1.83σ). Of the five properties highlighted in this study, Fig. 6 shows SFR to be the one for which MAGPHYS produces perhaps the most accurate reflection of the true values once the uncertainties are considered. However, there are some points of disagreement that are worth mentioning. The first example of this is for galaxy A1 at $z \approx 1.7$: this deviation of $\approx -0.59 \pm 0.16$ dex (3.7σ) coincides with a local minimum of M_{dust}, perhaps resulting from a strong outflow, and is associated with a brief reduction in the SFR that is not apparent when averaging over 100 Myr. The second example is for galaxy A2 around $1.0 \leq z \leq 1.5$ at the point where the galaxy has the highest stellar mass ($M_{\mathrm{star}} > 10^{11}\,\mathrm{M}_{\odot}$), and is the most quiescent that we have simulated (sSFR $\approx 10^{-10}\,\mathrm{yr}^{-1}$). For comparison, HS15 found that SFR was typically recovered to around 0.2–0.3 dex accuracy.[5] D20 reported that SFR was typically underestimated by approximately 20 per cent – very similar to our value of $\Delta\log(\mathrm{SFR}) = -0.11 \pm 0.06$ dex – attributing this to differences in their adopted SFHs, dust model and geometry.

The observed effects in sSFR mirror those in stellar mass and SFR as expected. Averaging over all snapshots and views, we obtain a mean offset of $\Delta\log(\mathrm{sSFR}) = 0.18 \pm 0.13$, a 1.38σ result which is consistent with the findings of HS15.

Fig. 6 highlights the excellent recovery of the true dust mass; averaging over all snapshots reveals a mean residual of $\Delta\log(M_{\mathrm{dust}}) = -0.19 \pm 0.17$ (1.12σ), suggesting that the results are typically consistent with the true values once the uncertainties are taken into account, consistent with the findings of D20.

Overall L_{dust} is well recovered with a mean residual of $\Delta\log(L_{\mathrm{dust}}) = 0.09 \pm 0.04$; this 2.25σ result is again in line with the results of HS15. However, the fifth row of Fig. 6 may suggest a weak trend for a larger $|\Delta\log(L_{\mathrm{dust}})|$ in the sense that the MAGPHYS estimates increasingly underestimate the true values as the simulations progress and the galaxies develop lower sSFR (though note that the scale of the residual panel for L_{dust} is half as large as for the other parameters, which exaggerates the size of the effect). It is possible that the assumptions inherent in the two-component dust model used by MAGPHYS, originally optimized to reproduce the observations of local star-forming galaxies (DC08), are no longer appropriate for the high-mass ($M_{\mathrm{star}} \approx 10^{11}$), highly star-forming ($\mathrm{SFR} > 20\,\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$) galaxies that are simulated here.

Finally, while it is not always the case, A_V is in general underestimated, with a mean residual of $\Delta A_V = -0.22 \pm 0.07$ (3.14σ), similar to the overall fidelity of the stellar mass recovery. This underestimation of the degree of extinction at V band may be linked to the typical underestimation of the overall dust luminosity, though it is interesting to note this does not prevent excellent recovery of the star formation rate for the majority of snapshots.

[5]We note that HS15 compared MAGPHYS 100 Myr-averaged SFRs with instantaneous SFRs rather than values averaged over 100 Myr, as we do here. Due to the bursty SFHs of the simulated galaxies, these values can differ significantly (Sparre et al. 2017; Flores Velázquez et al. 2021). This topic is further discussed further in connection with A_V recovery.

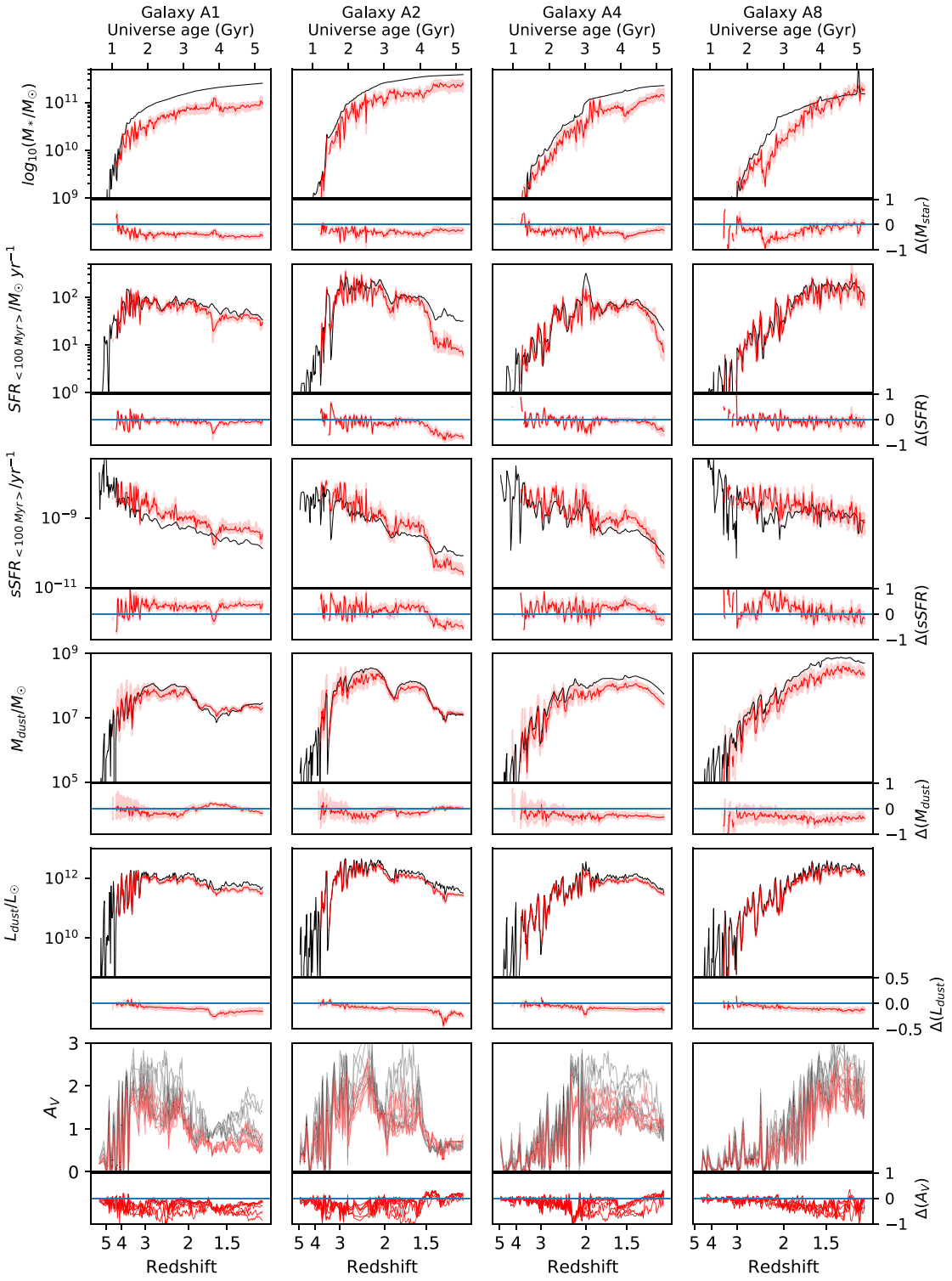


Figure 6. The overall MAGPHYS parameter estimation (red) compared with the true values from the simulation (black); MAGPHYS captures the overall true properties as a function of redshift. The columns refer to galaxies A1, A2, A4, and A8, respectively. In each row, the upper plot presents the evolution against universe age (upper x-axis) and redshift (lower x-axis), and the lower plot shows the residuals on the same x-axes (note that the range for $\Delta L_{\rm dust}$ is smaller than that for other properties). The top row presents the evolution of stellar mass, while the four subsequent rows present the corresponding evolution of SFR, sSFR, $M_{\rm dust}$, and $L_{\rm dust}$ respectively. In each main panel, the black line indicates the true values, the red line plots the mean across all views of the median recovered value, and the shaded area indicates the region enclosed by the typical error bar on each parameter (i.e. the mean difference between the 16th/84th percentile and the median, for the upper and lower bounds, respectively). In the final row, the black and red lines in the upper plot show the true and recovered values of A_V for the different views, while the lower plot shows the residuals for each view.

Table 3. Mean residuals – Δlog (parameter), as defined in equation (3) – for each property for each galaxy and the average across all galaxies; a negative value indicates an underestimate. The quoted uncertainties indicate the typical uncertainty that MAGPHYS derives on that galaxy parameter (equal to half the difference between the 16th and 84th percentiles of the derived PDF).

Galaxy	Δlog ($M_{\rm star}$)	Δlog (SFR)	Δlog (sSFR)	Δlog ($M_{\rm dust}$)	Δlog ($L_{\rm dust}$)	ΔA_V
A1	-0.37 ± 0.08	-0.10 ± 0.06	0.26 ± 0.12	-0.05 ± 0.14	-0.10 ± 0.04	-0.30 ± 0.07
A2	-0.28 ± 0.08	-0.21 ± 0.07	0.06 ± 0.13	-0.11 ± 0.15	-0.10 ± 0.04	-0.20 ± 0.07
A4	-0.27 ± 0.08	-0.08 ± 0.06	0.18 ± 0.13	-0.26 ± 0.19	-0.07 ± 0.04	-0.20 ± 0.07
A8	-0.24 ± 0.09	-0.05 ± 0.06	0.19 ± 0.14	-0.35 ± 0.21	-0.08 ± 0.03	-0.19 ± 0.07
Mean	-0.29 ± 0.09	-0.11 ± 0.06	0.18 ± 0.13	-0.19 ± 0.17	-0.09 ± 0.04	-0.22 ± 0.07

3.2.2 *Searching for systematic trends in the MAGPHYS fit results*

We used our simulations to determine the consistency of the MAGPHYS-derived galaxy properties across the range of values presented by the simulations. To do this, we binned the residuals defined using equation (3) across the full range of each property (stellar mass, SFR, sSFR, dust mass and dust luminosity) from the simulations and plotted the median bin residual. To gauge the significance of our results, we also averaged across all occupants of each bin to calculate the typical uncertainty associated with each MAGPHYS fit (although this is by no means constant in our results), and the scatter within each bin. The median residual, typical error bar, and the 16th and 84th percentile values for the scatter were plotted. Systematic trends might be expected to appear as deviations from horizontal lines in these figures; however, our results show that in all cases, the MAGPHYS results are remarkably consistent across the full range of values once the two sources of scatter are taken into account, and no further systematic trends can be identified. The plots are shown in Appendix A.

3.2.3 *The importance of panchromatic data in energy balance fitting*

We now discuss runs C and D, originally mentioned in Section 2.3. Run C used only the UV-NIR photometry from u to K band ($0.4\,\mu m < \lambda_{\rm eff} < 2.2\,\mu m$), while run D retained only the FIR data from the PACS and SPIRE instruments ($100\,\mu m < \lambda_{\rm eff} < 500\,\mu m$). While it is not possible to 'switch off' the energy balance criterion in MAGPHYS, runs C and D enable us to make a direct comparison of the results of 'traditional' SED fitting (i.e. attempting to recover the stellar mass or dust content of a galaxy from the optical/NIR data alone) with both the true values and the full panchromatic run. In both the starlight-only and FIR-only runs, MAGPHYS must rely on the physically motivated model and the energy balance assumption to estimate the properties usually associated with the missing observations (e.g. estimating the dust mass purely on the basis of the observed starlight, or the stellar mass using only FIR data).

Fig. 7 shows the results of these runs comparing the mean log Δ and typical uncertainty for the five properties for each of the three runs A, C, and D: full filter set, stellar-only and FIR-only.

The left panel of Fig. 7 shows the view and snapshot-averaged Δlog ($M_{\rm star}$) for the three runs. It is immediately clear that although the average Δlog ($M_{\rm star}$) is very similar for the stellar-only (0.31 dex) and all-filter (0.29 dex) runs, including the full set of data does reduce the typical uncertainty (shown by the error bars) from ±0.20 to ±0.09 dex. Unsurprisingly, attempting to estimate the stellar mass using only the FIR data leads not only to a large Δlog ($M_{\rm star}$) but also a significantly larger typical uncertainty (≈0.42 dex).

In the second panel, we show the corresponding results for Δlog (SFR). The power of panchromatic fitting is again clear, since

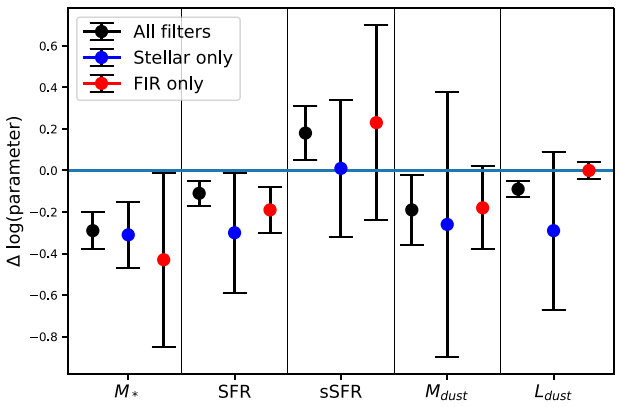


Figure 7. Using MAGPHYS to model panchromatic data gives better overall constraints on galaxy properties than sampling only a subset of the available wavelengths. Δlog (parameter) for each parameter of interest, averaged across all galaxies for three different MAGPHYS runs: (i) including all available photometry, (ii) stellar only – including only those bands that sample the starlight ($0.4\mu m < \lambda_{\rm eff} < 2.2\mu$m), and (iii) FIR only – including only the FIR data ($100\,\mu$m$< \lambda_{\rm eff} < 500\mu$m), with each set of results colour-coded as in the legend. The error bars on each data point represent the mean uncertainty for each MAGPHYS estimate, based on using the 16th and 84th percentiles of the estimated PDFs.

the largest Δlog (SFR) and typical uncertainty occur for the stellar-only fits, which can be influenced by the dominance of the lowest-attenuation sightlines (meaning that the amount of obscured star formation can be underestimated) as well as subject to the well-known age-dust degeneracy (e.g. Cimatti et al. 1997). Our results show that FIR-only SFR estimates are more reliable than those using the u to K-band photometry alone, since the FIR-only mean Δlog (SFR) $\approx 0.19 \pm 0.11$ is significantly closer to the true values than the corresponding stellar-only fits which have Δlog (SFR) $\approx 0.30 \pm 0.29$.

The situation is even more pronounced for the recovery of the sSFR, with Δlog (sSFR) for the three runs shown in the central panel of Fig. 7. Although the mean Δlog (sSFR) for the stellar-only run is closest to the true values, the typical uncertainties on the panchromatic run are more than a factor two smaller than the stellar-only estimates. The larger error bar represents a wide range of possible activity levels, making it impossible to unravel the age/dust degeneracy; by adding FIR data, the sSFR is better constrained. This, in turn, enables a constrained determination of the SFR and hence the cause of any observed reddening.

For $M_{\rm dust}$, Fig. 7 shows that the addition of stellar data makes very little difference to the mean Δlog ($M_{\rm dust}$) with FIR-only giving results within 0.18 dex and the full filter set 0.19 dex; this is comparable to the typical uncertainties (0.20 dex as opposed to 0.17 dex). Using

only the stellar data, the mean $\Delta\log(M_{dust})$ is 0.26 dex but the typical uncertainty is significantly increased to 0.64 dex, reflecting the difficulty associated with estimating the dust content of distant galaxies using data probing the starlight alone.

Finally, the right-hand panel of Fig. 7 shows the recovery of L_{dust} across the three runs. Interestingly, although the typical uncertainties are similar for the FIR-only and panchromatic runs, the inclusion of the UV/NIR data along with the energy balance criterion perhaps increases the mean $\Delta\log(L_{dust})$, although the significance of this difference is low.

3.3 Measuring the effect of UV/FIR 'decoupling' on the fidelity of MAGPHYS results

As discussed earlier, the primary goal of this work is to examine the fidelity of the MAGPHYS results as a function of the degree of correlation or apparent offset between UV and FIR emission using the three proxies for this 'decoupling' described in Section 2.3. The results are shown in Fig. 8, in which the mean Δ in dex for each parameter is plotted against the different measures for the degree of separation. Each of the five panels shows the residuals for one of the properties plotted against the degree of separation/correlation as measured by the three proxies. The coloured lines indicate the median residual in log-spaced bins, while the coloured shaded areas show the mean range enclosed by the 16th and 84th percentiles (i.e. the typical 1σ error in the limit of Gaussian statistics), and the grey shaded area shows the 16th and 84th percentile range of the scatter within each bin. The bin occupancy is shown by the grey background histogram relative to the right-hand axis. In many cases the scatter is larger than the typical uncertainties, this is likely to be the result of two effects. First, it reflects the fact that the MAGPHYS results contain a range of uncertainties that cannot be adequately summarized by a single error bar (the uncertainties show significant variation and contain outliers). Secondly, the uncertainties produced by MAGPHYS are likely to be underestimates. This is inevitably the case since the range of SEDs contained in any pre-computed library must by definition be smaller than the actual range of galaxy SEDs in the universe; for example neither real galaxies or those in our simulations have truly parametric SFHs. In addition, the MAGPHYS libraries may not be equally appropriate at all stages of our simulations.

The average performance of MAGPHYS is remarkably consistent, both as a function of the peak-to-peak distance between the UV and FIR images, and as a function of the light-weighted mean UV to FIR distance. In these cases, the mean Δ is less than ± 0.3 dex for all parameters, across the separations ranging from 0 to 10 kpc. In the lower plot of each panel we show the corresponding variation in Δ (in dex) as a function of the Spearman ρ calculated by comparing the UV and FIR images (recall that only the brightest 20 per cent of pixels were included in this calculation). Here again, the logarithmic difference between the derived and true properties appears independent of ρ once the mean uncertainties are taken in to account.

4 DISCUSSION

4.1 The redshift dependence of the MAGPHYS fit success rate

In Section 3.1 we showed that the fit success rate was a strong function of redshift, with 83 per cent of the mock observations having acceptable χ^2 overall, but no good fits being obtained at $z > 5.9$. Fixing each mock to be observed at $z = 2$ (Run B) resulted in an increase in the overall success rate to 93 per cent. A likely explanation for this is that the number of SFHs in the MAGPHYS library is a strong function of redshift (shown as the dashed line in Fig. 4, due to the requirement of considering only SFHs shorter than the Hubble time at the observed redshift), which results in significantly worse sampling of the priors at early epochs, particularly when the SFHs of galaxies are so weakly constrained by photometry (e.g. Smith & Hayward 2018).

In support of this idea, Fig. 9 shows the ratio of the best fit χ^2 obtained for our fiducial results (native redshift run A) to the corresponding value for the SEDs fixed to $z = 2$ (run B). It is clear that there is a systematic trend for the native χ^2 to be worse at $z > 2$ (corresponding to a universe age of $\leq$3.2 Gyr in our adopted cosmology) and better at $z < 2$. However this trend is by no means absolute, indicating that other effects such as the precise details of the rest-wavelengths being sampled and the number of available filters may also be playing a role.

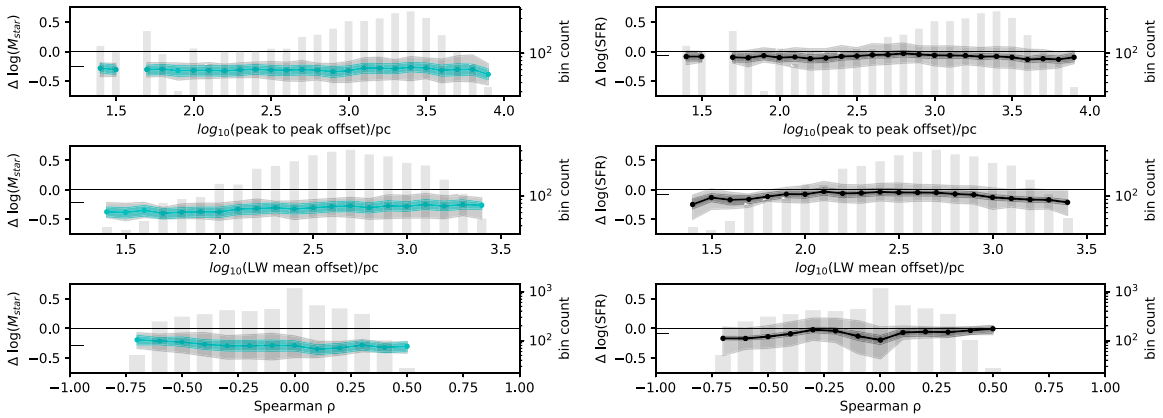
Interestingly, that the ratio of χ^2 for run A to that of run B does not converge on the right-hand side of this plot may indicate that the size of the MAGPHYS prior library still impacts the fit quality even at $z < 2$, though of course the difference is that at these comparatively late epochs the priors are sufficiently well-sampled to obtain statistically acceptable fits to the data.

4.2 The fidelity of MAGPHYS results for dusty, high-redshift galaxies

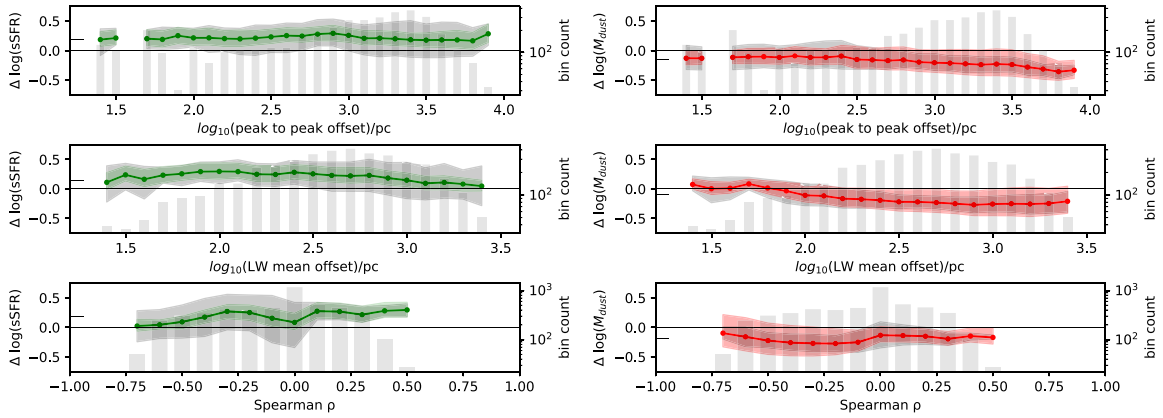
The principal aim of this study is to determine how the fidelity of the energy balance code MAGPHYS is impacted when it is applied to high-redshift galaxies for which the observed UV and FIR emission are offset, or spatially 'decoupled'. For such galaxies, the observed UV light potentially originates from young star clusters that are not spatially co-located with the young stars that dominate the dust heating and thus FIR emission. Consequently, it is possible that the relatively unobscured young stars could yield a blue UV-optical slope and cause SED modelling codes to underestimate the attenuation. It has been shown that the use of panchromatic data is important when fitting such galaxies (Roebuck et al. 2019), and fitters such as MAGPHYS use energy balance to produce physically motivated, panchromatic models that seek to minimize this underestimation. We determine the efficacy of this approach by analysing the logarithmic difference, Δ, between the true and median-likelihood estimates for stellar mass, SFR, specific SFR, dust mass and dust luminosity as a function of three proxies for the degree of 'decoupling' between the UV and FIR data.

In all cases, the performance of MAGPHYS appears independent of the degree of UV/FIR 'decoupling' as measured by all three proxies. We therefore conclude that energy balance SED fitting codes can perform just as well in the presence of such effects as they do when the dust and young stars are co-located within a galaxy.

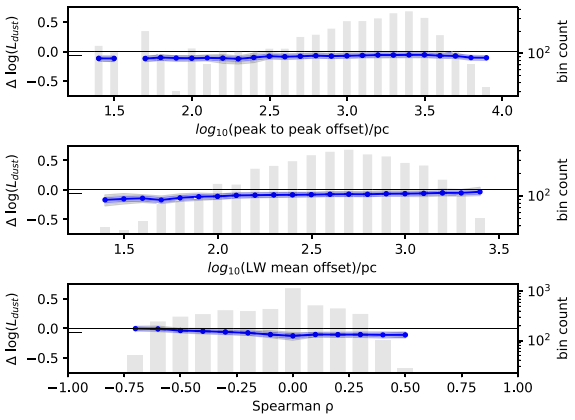
We suspect that the explanation for this success is that the Charlot & Fall (2000) dust attenuation model used by MAGPHYS is sufficiently flexible to handle this 'decoupling' in many cases and that the χ^2 algorithm is doing its job by identifying cases for which the model cannot yield a self-consistent solution (i.e. very low attenuation but high FIR luminosity). This has been shown to be the case for an unmodelled AGN contribution to the SED: Smith et al. (2021) noted that using the χ^2 threshold from Smith et al. (2012), which we have also implemented here, had the effect of flagging the vast majority of LOFAR-detected AGN as bad fits unless the AGN contribution to the emergent luminosity was very small. Of



(a) $M_{\rm star}$ (b) SFR



(c) sSFR (d) $M_{\rm dust}$



(e) $L_{\rm dust}$

Figure 8. The fidelity of MAGPHYS is largely independent of the extent of any UV/FIR offset, as measured by the three proxies, once the uncertainties are considered. Δlog (parameter) as a function of three proxies for the difference between the UV and FIR images – panel (a) presents the data for $M_{\rm star}$, (b) for SFR, (c) for sSFR, (d) for $M_{\rm dust}$, and (e) for $L_{\rm dust}$. For each property, the data points represent the mean over all views and snapshots in that bin. The shaded area of the same colour indicates area enclosed by the mean 16th and 84th percentile values within the bin. The grey shaded area shows area enclosed by the 16th and 84th percentile values for the scatter within each bin. The top plot in each panel shows the logarithmic difference Δ, as a function of the peak-to-peak distance between the UV and FIR images; the second and third panels show the corresponding log Δ as a function of the light-weighted mean UV-FIR offset and the Spearman rank correlation coefficient ρ between the 20 per cent brighter pixels in either the UV or FIR images. The short coloured lines adjacent to the left-hand y-axis represent the overall mean value. The grey histograms in each panel (a) to (e) show the bin occupancy relative to the right-hand axis.

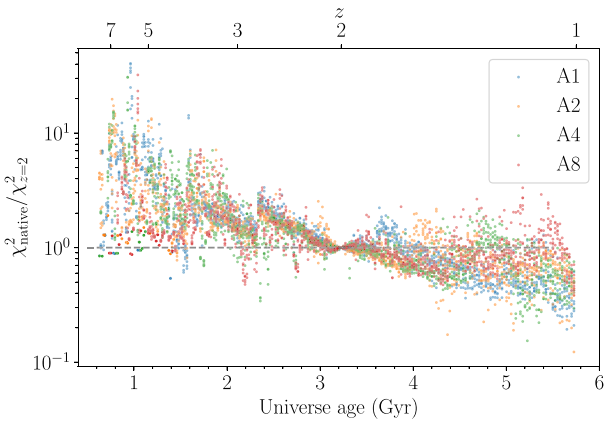


Figure 9. The best fit χ^2 depends on the size of the MAGPHYS library, which varies with the redshift assumed for the fit. This plot shows the ratio of best fit χ^2 obtained for run A (at the native redshift) to that obtained in run B (where all SEDs were fixed to $z = 2$). For galaxies on the left-hand side of this plot the prior gets larger in run B, while for galaxies viewed at later times, the opposite effect is apparent.

course, it is expected (e.g. Witt & Gordon 2000) and observed (e.g. Kriek & Conroy 2013; Boquien et al. 2022; Nagaraj et al. 2022) that the attenuation law is not universal and instead varies by galaxy type. Should additional flexibility be required in future, we note that other works have explored implementing modifications to the standard dust law, including Battisti et al. (2019) who added a 2175 Å feature to remove a systematic redshift effect, as well as Lo Faro et al. (2017) and Trayford et al. (2020) who allowed the power-law indices of equations (1) and (2) to vary. However, the fact that there is no scope to easily modify the dust parametrization assumed in MAGPHYS leaves us no option but to defer further investigation of this potentially important aspect for a future work.

The reason that some have claimed that energy balance should fail in galaxies with significant IR-UV offsets is that the unobscured lines of sight should dominate the UV emission, meaning that the attenuation that would be inferred from the observed UV-optical emission would be less than the total attenuation experienced by the stellar population as a whole. However, energy balance codes such as MAGPHYS use the FIR luminosity as a simultaneous constraint on the attenuation, and it would simply not be possible to obtain a satisfactory fit to both the UV-optical and FIR regions of the SED assuming low attenuation when the FIR luminosity is high.[6] Furthermore, we note that even in 'normal' galaxies that do not exhibit significant UV-FIR offsets, stars of a given age are not all subject to the same amount of attenuation (e.g. the Charlot & Fall 2000 dust model). Instead, even for a single age and line of sight, there is a distribution of dust optical depths, and this distribution varies with both the stellar age and line of sight considered. The Charlot & Fall (2000) model attempts to capture this complex age and line-of-sight dependence using only two effective optical depths. Though this underlying model is certainly very crude compared to both the simulations and real galaxies, HS15 have already shown that it is adequate to correct for the effects of dust attenuation in at least some low-redshift galaxies. There is no a priori reason to believe that it should 'break' above some offset threshold (which was the motivation for this study). Our results demonstrate that even when the width of the optical depth distribution experienced by young stars is very wide (i.e. in our simulations some young stars are almost completely unobscured, whereas others have line-of-sight UV optical depths >>1), the Charlot & Fall (2000) model can still adequately capture the overall effects of dust attenuation in most cases.

[6]It is tempting to investigate this by making a plot similar to Fig. 8 but including only those fits that exceed the χ^2 threshold we use to identify the bad fits. However, since the best fit model is statistically unacceptable, we cannot believe the parameter estimates produced by MAGPHYS in these cases, meaning that such a test is not meaningful.

5 CONCLUSIONS

Recent works (e.g. Hodge et al. 2016; Casey et al. 2017; Miettinen et al. 2017; Simpson et al. 2017; Buat et al. 2019) have questioned whether energy balance SED fitting algorithms are appropriate for studying high-redshift star-forming galaxies, due to observations of offsets between the UV and FIR emission (e.g. Hodge et al. 2016; Rujopakarn et al. 2016; Chen et al. 2017; Bowler et al. 2018; Gómez-Guijarro et al. 2018; Rujopakarn et al. 2019). Clumpy dust distributions within these galaxies may cause a small fraction of relatively unobscured young stars to influence the blue UV-optical slope and result in an underestimation of the attenuation even if the bulk of the young stars are completely dust-obscured. We have used four cosmological zoom-in simulations of dusty, high-redshift galaxies from the FIRE-2 project, together with the radiative transfer code SKIRT, to generate over 6700 synthetic galaxy SEDs spanning a redshift range $8 > z > 1$. We used these model data to test the fidelity of the galaxy properties recovered using the energy balance fitting code MAGPHYS with 18 bands of UV–FIR photometry, building on our previous related studies (HS15, Smith & Hayward 2015, 2018). Our principal findings are as follows:

(i) we find that the high-z version of MAGPHYS was able to produce statistically acceptable best fit SEDs for 83 per cent of the synthetic SEDs that we trialled. The fit success rate fell to 50 per cent for galaxies at $z > 4.85$ and zero for galaxies at $z > 5.9$. This reduction in fit success rate has two main contributing factors:

> (a) the fixed MAGPHYS libraries, combined with the requirement that model SFHs should be shorter than the age of the universe at any given redshift reduces the size of the MAGPHYS library available at higher redshifts, mean that the priors become increasingly poorly sampled at earlier times;
>
> (b) the evolution of the simulated galaxies is increasingly stochastic at the earliest times in our simulations due to their lower mass, causing bursts of star formation to have a disproportionate influence on a galaxy's bolometric luminosity that cannot be reconciled with the MAGPHYS prior libraries.

(ii) Where statistically acceptable best fits were obtained, we found that MAGPHYS fits are able to broadly capture the true evolution of the four zoom-in simulations that we studied (steady build-up of stellar mass, generally decreasing sSFR, evolution of dust mass), despite individual snapshots being fit independently. In addition, we find that the fidelity of this recovery is remarkably consistent across a broad range of galaxy properties sampled by the simulations, showing no evidence for strong systematics as a function of stellar mass, SFR, sSFR, dust mass, or dust luminosity.

(iii) Combining UV to FIR observations with an energy balance SED fitting code provides a powerful way to combine multiwavelength data, and obtain the most reliable estimates of the ground-

truth galaxy properties. The panchromatic results outperform those obtained by using either the stellar or dust emission alone.

(iv) We find no evidence that the performance of MAGPHYS depends on the degree of spatial 'decoupling' between the UV and FIR data, despite suggestions to the contrary by several other works. Indeed, our results show that the fidelity of the galaxy properties derived is very similar to that observed for local galaxies, for example in our previous work (Hayward & Smith 2015).

ACKNOWLEDGEMENTS

The authors thank the anonymous referee for their careful review of this paper. PH and DJBS would like to thank the Flatiron Institute for their hospitality. The Flatiron Institute is supported by the Simons Foundation. PH would also like to thank Dr Alyssa Drake for helpful input and suggestions. DJBS acknowledges support from the UK Science and Technology Facilities Council (STFC) under grant number ST/V000624/1. This research has made use of NASA's Astrophysics Data System Bibliographic Services. DAA acknowledges support by NSF grants AST-2009687 and AST-2108944, CXO grant TM2-23006X, and Simons Foundation award CCA-1018464.

DATA AVAILABILITY

The data presented in this paper will be shared on reasonable request to the first author. Selected snapshots of the four simulated galaxies analysed here are available as part of the FIRE-2 public data release, see Wetzel et al. (2023) and http://flathub.flatironinstitute.org/fire

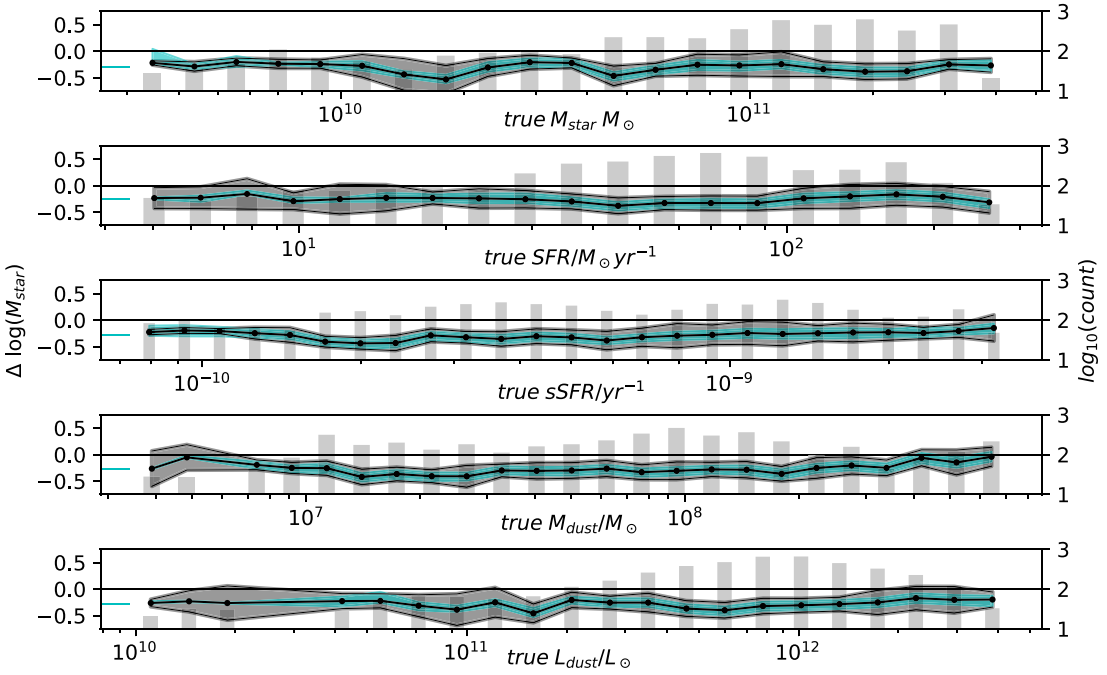
REFERENCES

Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017, MNRAS, 472, 109
Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, ApJS, 196, 22
Battisti A. J. et al., 2019, ApJ, 882, L61
Bolzonella M., Miralles J. M., Pelló R., 2000, A&A, 363, 476
Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, 103
Boquien M. et al., 2022, A&A, 663, 50
Bowler R. A. A., Bourne N., Dunlop J. S., McLure R. J., McLeod D. J., 2018, MNRAS, 481, 1631
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, L1503
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buat V., Ciesla L., Boquien M., Małek K., Burgarella D., 2019, A&A, 632, 79
Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W., 2016, ApJ, 833, L98
Camps P., Baes M., 2020, Astron. Comput., 31, 100381
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, L44
Casey C. M. et al., 2017, ApJ, 840, L101
Chabrier G., 2003, PASP, 115, 763
Charlot S., Fall S. M., 2000, ApJ, 539, L718
Chen C.-C. et al., 2017, ApJ, 846, L108
Cheng C. et al., 2023, ApJ, 942, L19
Cimatti A., Bianchi S., Ferrara A., Giovanardi C., 1997, MNRAS, 290, 43
Cochrane R. K. et al., 2019, MNRAS, 488, 1779
Cochrane R. K. et al., 2021, MNRAS, 503, 2622
Cochrane R. K., Hayward C. C., Anglés-Alcázar D., 2022, ApJ, 939, L27
Cochrane R. K., Hayward C. C., Anglés-Alcázar D., Somerville R. S., 2023, MNRAS, 518, 5522
Conroy C., 2013, ARA&A, 51, 393
Crain R. A. et al., 2015, MNRAS, 450, 1937
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595 (DC08)
da Cunha E. et al., 2015, ApJ, 806, L110
Draine B. T. et al., 2007, ApJ, 663, L866
Dudzevičiūtė U. et al., 2020, MNRAS, 494, 3828
Duncan K. J. et al., 2021, A&A, 648, 4
Feldmann R., Hopkins P. F., Quataert E., Faucher-Giguère C.-A., Kereš D., 2016, MNRAS, 458, 14
Flores Velázquez J. A. et al., 2021, MNRAS, 501, 4812
Gómez-Guijarro C. et al., 2018, ApJ, 856, L121
Hainline L. J., Blain A. W., Smail I., Alexander D. M., Armus L., Chapman S. C., Ivison R. J., 2011, ApJ, 740, L96
Hayward C. C., Smith D. J. B., 2015, MNRAS, 446, 1512 (HS15)
Hayward C. C., Kereš D., Jonsson P., Narayanan D., Cox T. J., Hernquist L., 2011, ApJ, 743, L159
Hodge J. A. et al., 2016, ApJ, 833, L103
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Hunt L. K. et al., 2019, A&A, 621, 51
Jonsson P., 2006, MNRAS, 372, 2
Kennicutt R. C. J., 1998a, ARA&A, 36, 189
Kennicutt R. C., 1998b, ApJ, 498, L541
Kennicutt R. C. J. et al., 2003, PASP, 115, 928
Kennicutt R. C. et al., 2011, PASP, 123, 1347
Kokorev V. I. et al., 2021, ApJ, 921, L40
Kondapally R. et al., 2021, A&A, 648, 3
Kriek M., Conroy C., 2013, ApJ, 775, L16
Lang P. et al., 2019, ApJ, 879, L54
Liu D., 2020, Astrophysics Source Code Library, record ascl:2005.002
Liu D. et al., 2021, ApJ, 909, L56
Lo Faro B., Buat V., Roehlly Y., Alvarez-Marquez J., Burgarella D., Silva L., Efstathiou A., 2017, MNRAS, 472, 1372
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Małek K. et al., 2018, A&A, 620, 50
Michałowski M. J., Hayward C. C., Dunlop J. S., Bruce V. A., Cirasuolo M., Cullen F., Hernquist L., 2014, A&A, 571, 75
Miettinen O. et al., 2017, A&A, 606, 17
Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, MNRAS, 435, 87
Muñoz-Mateos J. C. et al., 2009, ApJ, 701, L1965
Myers J. L., Well A. A., 2003, Research design and statistical analysis. Lawrence Erlbaum Associates, New Jersey
Nagaraj G., Forbes J. C., Leja J., Foreman-Mackey D., Hayward C. C., 2022, ApJ, 932, L54
Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793
Pacifici C. et al., 2023, ApJ, 944, L141
Parsotan T., Cochrane R. K., Hayward C. C., Anglés-Alcázar D., Feldmann R., Faucher-Giguère C. A., Wellons S., Hopkins P. F., 2021, MNRAS, 501, 1591
Roebuck E., Sajina A., Hayward C. C., Martis N., Marchesini D., Krefting N., Pope A., 2019, ApJ, 881, L18
Rujopakarn W. et al., 2016, ApJ, 833, L12
Rujopakarn W. et al., 2019, ApJ, 882, L107
Sabater J. et al., 2021, A&A, 648, 2
Schaye J. et al., 2015, MNRAS, 446, 521
Schreiber C. et al., 2015, A&A, 575, 74
Seillé L. M., Buat V., Haddad W., Boselli A., Boquien M., Ciesla L., Roehlly Y., Burgarella D., 2022, A&A, 665, 137
Shimwell T. W. et al., 2017, A&A, 598, 104
Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, L103
Simpson J. M. et al., 2017, ApJ, 839, L58
Smith D. J. B., Hayward C. C., 2015, MNRAS, 453, 1597
Smith D. J., Hayward C. C., 2018, MNRAS, 476, 1705
Smith D. J. B. et al., 2012, MNRAS, 427, 703
Smith D. J. B. et al., 2021, A&A, 648, 6
Sparre M., Hayward C. C., Feldmann R., Faucher-Giguère C.-A., Muratov A. L., Kereš D., Hopkins P. F., 2017, MNRAS, 466, 88

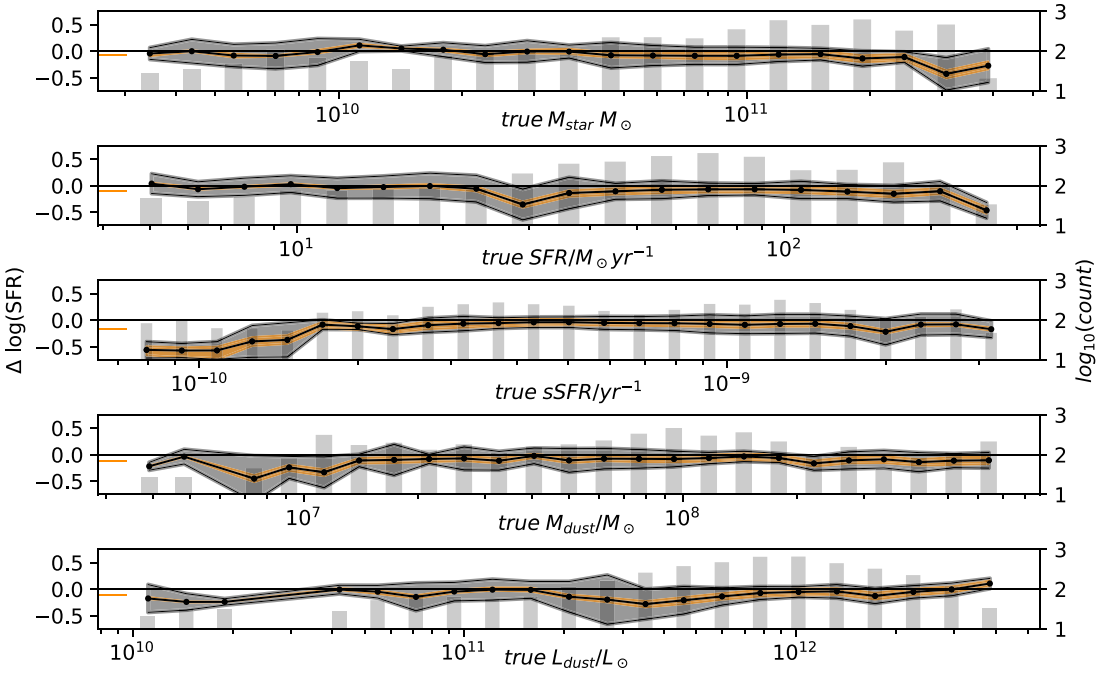
Springel V., 2005, MNRAS, 364, 1105
Tacchella S. et al., 2022, ApJ, 926, L134
Tasse C. et al., 2021, A&A, 648, 1
Trayford J. W., Lagos C. d. P., Robotham A. S. G., Obreschkow D., 2020, MNRAS, 491, 3937
Walcher J., Groves B., Budavári T., Dale D., 2011, Ap&SS, 331, 1
Wellons S., Faucher-Giguère C.-A., Anglés-Alcázar D., Hayward C. C., Feldmann R., Hopkins P. F., Kereš D., 2020, MNRAS, 497, 4051
Wetzel A. et al., 2023, ApJS, 265, 44
Witt A. N., Gordon K. D., 2000, ApJ, 528, L799
Wuyts S., Franx M., Cox T. J., Hernquist L., Hopkins P. F., Robertson B. E., van Dokkum P. G., 2009, ApJ, 696, L348
Zhang H.-X., Gao Y., Kong X., 2010, MNRAS, 401, 1839

APPENDIX: MAGPHYS FIDELITY ACROSS THE RANGE OF PROPERTY VALUES

As noted in the text, we analysed the systematic recovery of galaxy parameters in more detail by looking at the Δ for each parameter as a function of the true values of every other parameter. We do not detect any significant trends once the typical error bars (calculated as the mean of the difference between the 16th and 84th percentiles and the median likelihood value) and scatter on the derived values are accounted for. The fidelity of the MAGPHYS parameter recovery persists across a wide range of parameter space.

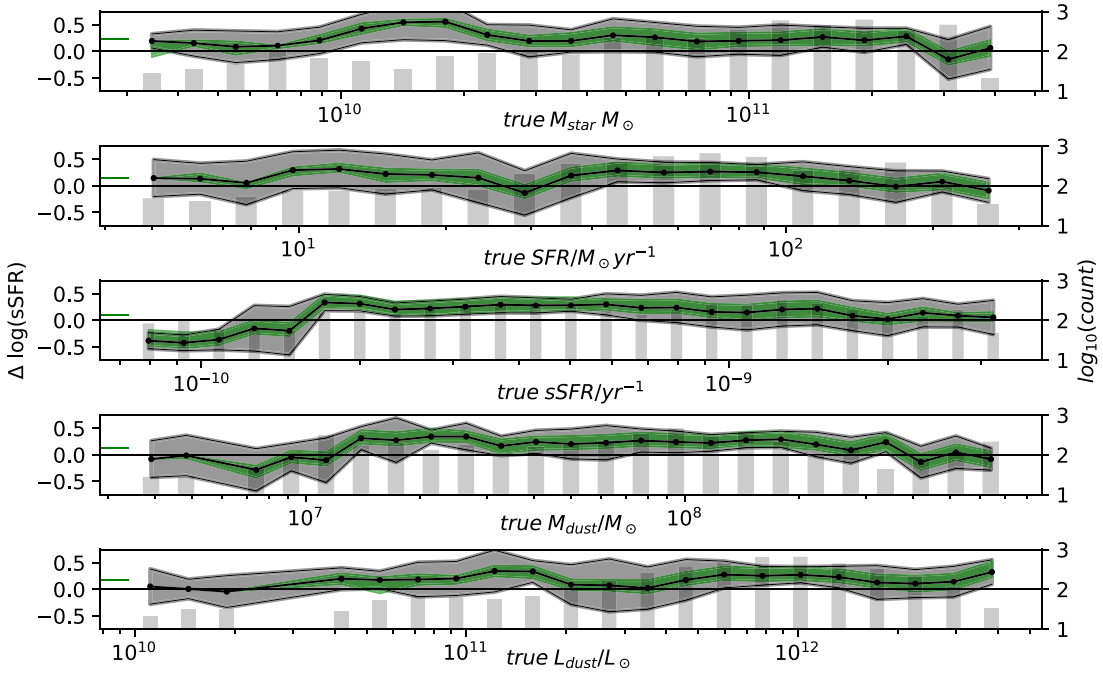


(a) $M_{\rm star}$

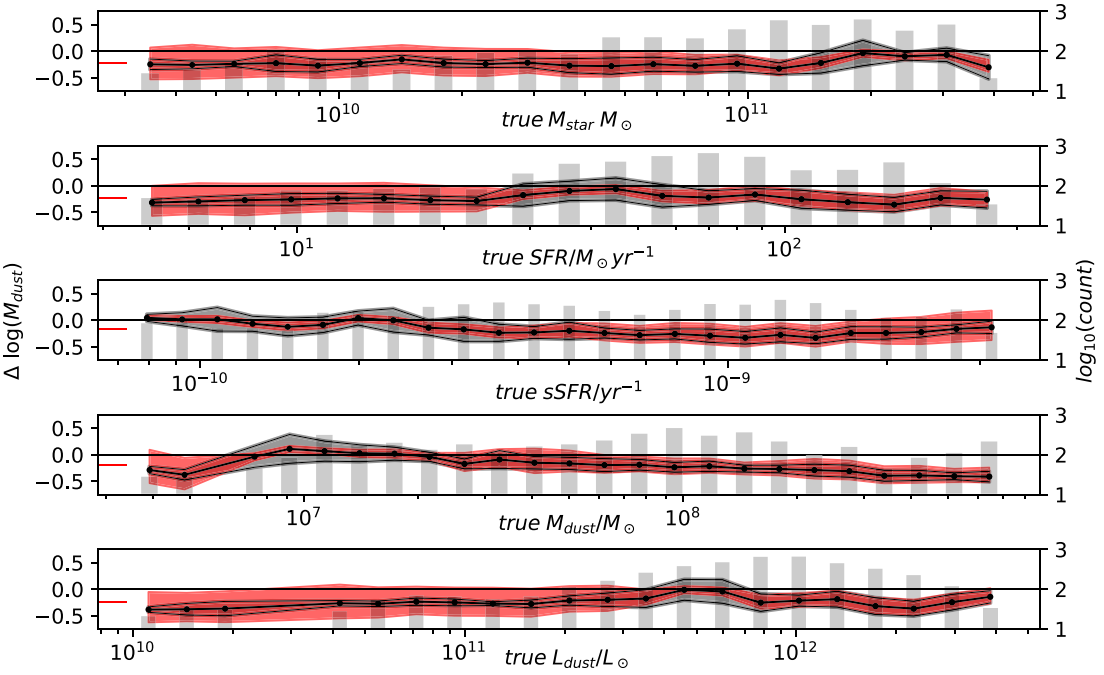


(b) SFR

Figure A1. The fidelity of MAGPHYS' recovery of $M_{\rm star}$ and SFR is remarkably consistent across the full range of true galaxy properties. The top five panels plot the relationship between the $M_{\rm star}$ residual and true value of the properties $M_{\rm star}$, SFR, sSFR, $M_{\rm dust}$, and $L_{\rm dust}$, respectively. The data points in black represent the median value for the residual in log-spaced bins; bin occupancy is shown by the background grey bar chart with log values read from the right-hand axis – note that bins with occupancy < 20 have been removed for clarity. In each case, the coloured band shows the median 16th and 84th percentile limits for the residuals within the bin and the bounded grey region shows the median 16th and 84th percentile limits for the scatter within the bin. The short coloured line on left-hand of each plot shows the average for the plotted value, residual values are read from the left-hand axis. The lower five panels show the same for the SFR residual.



(a) sSFR



(b) M_{dust}

Figure A2. Similar to Fig. A1, but showing the remarkably consistent recovery of sSFR and M_{dust} as a function of the true galaxy properties. The top five panels plot the relationship between the sSFR residual and true value of the properties M_{star}, SFR, sSFR, M_{dust}, and L_{dust}, respectively. The data points in black represent the median value for the residual in log-spaced bins; bin occupancy is shown by the background grey bar chart with log values read from the right-hand axis – note that bins with occupancy < 20 have been removed for clarity. In each case, the coloured band shows the median 16th and 84th percentile limits for the residuals within the bin and the bounded grey region shows the median 16th and 84th percentile limits for the scatter within the bin. The short coloured line on left-hand of each plot shows the average for the plotted value, residual values are read from the left-hand axis. The lower five panels show the same for the M_{dust} residual.

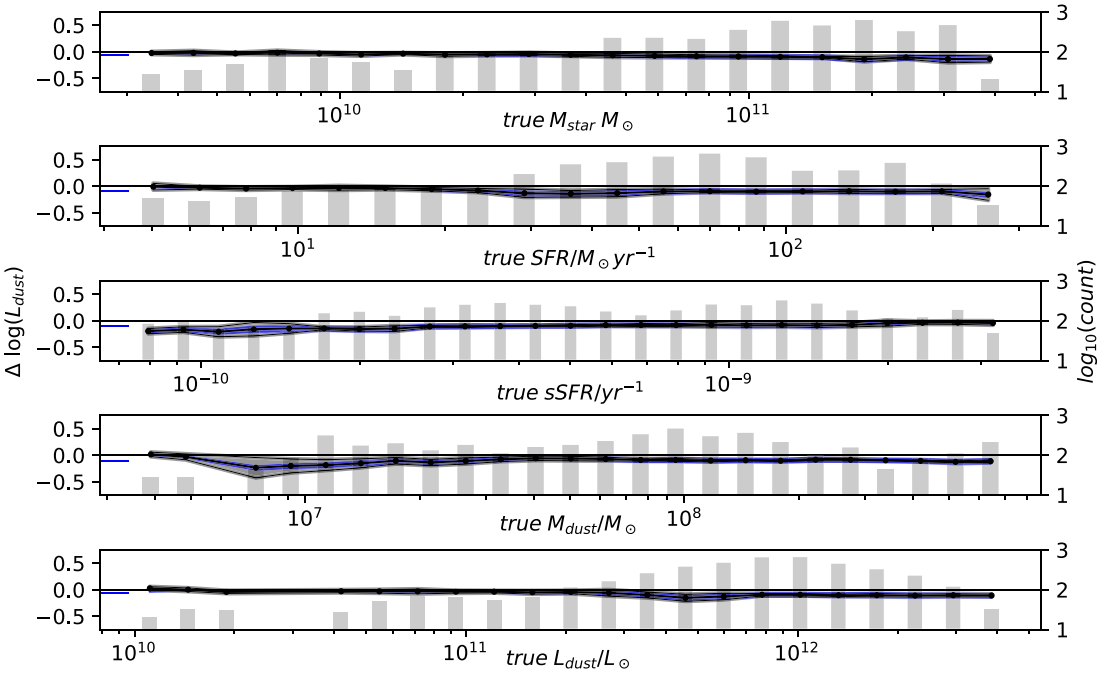


Figure A3. Similar to Fig. A1, but showing the remarkably consistent recovery of $L_{\rm dust}$. The five panels plot the relationship between the $L_{\rm dust}$ residual and true value of the properties $M_{\rm star}$, SFR, sSFR, $M_{\rm dust}$, and $L_{\rm dust}$, respectively. The data points in black represent the median value for the residual in log-spaced bins; bin occupancy is shown by the background grey bar chart with log values read from the right-hand axis – note that bins with occupancy < 20 have been removed for clarity. In each case, the coloured band shows the median 16th and 84th percentile limits for the residuals within the bin and the bounded grey region shows the median 16th and 84th percentile limits for the scatter within the bin. The short coloured line on left-hand of each plot shows the average for the plotted value, residual values are read from the left hand axis.

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