where $\overline{S}$ are the observed band-averaged fluxes. Observations are compared against simulated fluxes obtained by means of Eq. (59) with a galaxy model sampled from some position θ in parameter space.

When observing a source with a given instrument, the signal to noise ratio (S/N) might be small due to instrumental or environmental noise of different origins. It is not rare that the measurement in some band is not considered a detection, due to the low value of S/N. Even though it should be kept in mind that also small values of the S/N are measurements with an associated error, and should therefore be treated as such, we allow for a different treatment of data-points and upper limits.

The χ^2 statistics appearing in Eq. (61) can be expanded as

$$\chi^2\left(\overline{S}\,\middle|\,\theta\right) \equiv \sum_{i=0}^{N_{\rm det}} \left(\frac{\overline{S}_i - \overline{S}_i(\theta)}{\sigma_i}\right)^2 + \sum_{j=0}^{N_{\text{up-lims}}} f\left[\overline{S}_j, \overline{S}_j(\theta), \sigma_j\right] \tag{62}$$

where the first sum is a simple χ^2 among all the $N_{\rm det}$ bands where the measurement has been classified as a detection. The second sum instead runs on the $N_{\text{up-lims}}$ functions f accounting for the probability that the upper-limit in band j is drawn from a given model θ.

We provide three different possible treatments for upper limits:

- χ^2 (default): non-detections are treated exactly as detections with a large error;
- naive: a simple step-wise function setting the log-likelihood to $-\infty$ (i.e. zero probability) when the model predicts a flux larger than observed and to 0 (i.e. probability equal to one) when the predicted flux is lower than the limit

$$f\left[\overline{S}_j, \overline{S}_j(\theta), \sigma_j\right] = \begin{cases} -\infty & \overline{S}_j(\theta) > \overline{S}_j \\ \\ 0 & \text{otherwise;} \end{cases} \tag{63}$$

- Sawicki (2012): the author proposes a modification of the χ^2 that consists of the integral of the probability of some observation up to the given proposed model. If the errors on data are Gaussian, this integral provides the following analytical expression for the corresponding log-likelihood:

$$f\left[\overline{S}_j, \overline{S}_j(\theta), \sigma_j\right] = -2\ln\left\{\sqrt{\frac{\pi}{2}}\sigma_j\left[1 + \text{erf}\left(\frac{\overline{S}_j - \overline{S}_j(\theta)}{\sqrt{2}\sigma_j}\right)\right]\right\}. \tag{64}$$

  Even though it can be argued that using the expression above is the most formally correct way of accounting for upper limits when errors are Gaussian, the combination of logarithm and error function is particularly risky in computational terms. Specifically, it tends to hit the numerical limit of floating point numbers representation accuracy really fast, leading to undefined behaviour. These problems, even though negligible in most of the occurrences, might lead to difficulties in the convergence of the posteriors for particularly complex posterior shapes.

As already mentioned though, in a Bayesian framework based on direct parameter-space sampling, even in the case of low S/N there is no real reason for using a different statistical treatment with respect to detections[9]. The large relative error already contains the information necessary to inform the χ^2 about the lack of flux in the specific band. We therefore set as default behaviour for upper-limits the usage of a standard χ^2. We strongly recommend to provide the actual measured flux, independently from its S/N, to the sampling algorithm. When such measurement is not available, a safe choice would be to set the flux measurement to the same value of the absolute error.

At the current state, GalaPy is thought for the study of individual galaxies and not for the study of the correlation between the parameters in a large sample of objects, therefore, a sophisticated treatment of noise and systematic uncertainties is not necessary (e.g. Kelly et al. 2012; Galliano 2018). Nevertheless, in preparation for future extensions of the library and for completeness, we have implemented a simplistic naive treatment of calibration errors and/or unknown systematic errors that might be present in the modelled datasets. To this purpose, we allow for the presence of a nuisance parameter, $f_{\rm sys}$, that modifies the measured uncertainties as

$$\tilde{\sigma_i}^2(\theta, f_{\rm sys}) \equiv \sigma_i^2 + f_{\rm sys}^2 \overline{S}_i^2(\theta). \tag{65}$$

This modified error depends on the model parameters θ through the predicted SED band flux $\overline{S}_i(\theta)$ and on the nuisance parameter $f_{\rm sys}$ as well as from the original measured error σ_i. By adding a positive value to the observed variance we are making the assumption it had been underestimated by a relative factor $f_{\rm sys}$.

We are not accounting for eventual correlations between observational bands thus our Gaussian log-likelihood is simply modified by an additional term accounting for the dependence of the variance on the model parameters. For the case of detections Eq. (61) becomes

$$\ln\mathcal{L}\left(\overline{S}\,\middle|\,\theta, f_{\rm sys}\right) \equiv -\frac{1}{2}\sum_i \left\{\frac{\left[\overline{S}_i - \overline{S}_i(\theta)\right]^2}{\tilde{\sigma_i}^2(\theta, f_{\rm sys})} + \ln\left[2\pi\,\tilde{\sigma}_i^2(\theta, f_{\rm sys})\right]\right\}; \tag{66}$$

we note that a similar modification is applied to the case of upper limits.

This simple noise model adds only one parameter to the multi-dimensional space that has to be sampled, therefore it does not particularly burden the sampling procedure. We have tested on multiple problem set-ups that $f_{\rm sys}$ is completely uncorrelated to the other free parameters: the only net effect on the final posterior is to make the constraints less tight, as it would be expected if errors in the observed dataset were larger. Nonetheless, we have also observed that the addition of this systematic error in some cases help in breaking the degeneracy between parameters, especially in the case of multi-modal posteriors such as when we are estimating photometric redshifts.

In closing this section, we highlight that the default behaviour of GalaPy currently only accounts for uniform uninformative priors, whose limits are set by the user. This choice is motivated by the argument that each galaxy should be considered as an independent object for which a-priori knowledge of the parameters can be hardly argued.

Accessing the GalaPy Python API, the aforementioned behaviours can be easily modified. Furthermore, more sophisticated statistical tools, such as non-Gaussian errors and non-uniform priors, are planned for future extensions of the library. We are already working at the implementation of a hierarchical Bayesian sampling scheme that, along with a more sophisticated treatment of the systematic errors, is intended for the application of GalaPy on large samples of galaxies, foreseeing upcoming data from future surveys.

[9] This argument also applies to the extreme case of a negative flux.

3.2. Samplers

The statistical framework of GalaPy comes with a `Sampler` object that provides a common interface for the parameter-space samplers we rely on, namely, `emcee` (Foreman-Mackey et al. 2013) and `dynesty` (Speagle 2020; Koposov et al. 2023). These two libraries provide different and complementary approaches to Monte Carlo sampling of a multi-dimensional space. We maintain both tools in order to provide a flexible machinery that can be adapted to different problems.

- `emcee`, provides an implementation of the Markov-chain Monte Carlo (MCMC) technique. Specifically, it implements an ensemble sampler with affine invariance (Goodman & Weare 2010) that, by instantiating many test particles (walkers) in the parameter space, builds first order Markov sequences of proposals that are tested against the likelihood. The dynamics of this system of particles is regulated by the requirement that, at each new step, a better estimate of the parameters is drawn.
- `dynesty` implements Dynamic Nested Sampling (Higson et al. 2019), a generalised version of nested sampling (Skilling 2004, 2006) where the number of test particles (here live points) is dynamically increased in regions of the posterior where a higher accuracy is required. The parameter space is modelled as a nested set of iso-likelihood regions that are sampled until the overall evidence reaches a stopping criterion set by the user. In our default hyper-parameters set-up we provide an 80%/20% posterior/evidence split and we model the posterior space with multiple ellipsoids (Feroz et al. 2009), as we expect to have multiple peaks and correlations when sampling high dimensional parameter spaces. We use the default stopping function

$$\mathcal{S}(f_p, s_p, s_\mathcal{Z}, n) \equiv f_p \times \frac{\mathcal{S}_p(n)}{s_p} + (1 - f_p) \times \frac{\mathcal{S}_\mathcal{Z}(n)}{s_\mathcal{Z}} < 1,$$

  where f_p is the fractional importance we place on posterior estimation (20%, as mentioned above), $\mathcal{S}_p$ is the posterior stopping function, $\mathcal{S}_\mathcal{Z}$ is the evidence stopping function, s_p is the posterior "error threshold", $s_\mathcal{Z}$ is the evidence error threshold, and n is the total number of Monte Carlo realisations, used to generate the posterior/evidence stopping values.

When sampling high-dimensional large volumes the degeneracy between parameters can easily generate a complex posterior topology, such as multiple peaks on some parameters or non-linear correlations. Our suggestion for an optimal usage of GalaPy is to rely on dynamic nested sampling in this case. As an empirical rule of thumb, we can recommend to rely on nested sampling when the number of free parameters is larger than 5 and when it is not necessary to include extremely complex priors (as this, even though feasible, is not trivial).

On the other hand, MCMC provides a more straightforward interface to the inclusion of sophisticated priors and proves to be efficient and to possibly converge faster when working on smaller and well-behaved volumes, i.e. when multiple peaks and complex correlations among parameters are not to be expected.

GalaPy comes with a default set-up for the hyper-parameters determining the behaviour of the two currently available samplers. The chosen values should work, and have been tested, on several common possible problems. For both the nested sampler and the MCMC sampler, a drawback of this default set-up is that it might not be the fastest to converge, nonetheless convergence should be guaranteed. We stress that it is not possible to provide a general set-up of the aforementioned hyper-parameters. Experienced users can access and modify the default values to better suit the specific needs of the problem at study.

As already mentioned, we plan to include additional samplers in future extensions of the library.

3.3. Results

We provide a `Results` class that collects all the information acquired during the sampling run and computes derived quantities for easy access and analysis. This includes all the full-SEDs computed for each position in the parameter space, all the derived quantities (masses, metallicities and temperatures) as well as all the coordinates in the parameter space and all the GalaPy objects built during the run.

`Results` objects tend to be particularly heavy in terms of both volatile and non-volatile memory. The typical size primarily depends on the number of samples that were needed to obtain a converged posterior and, secondarily, on the number of free-parameters and the other characteristics of the sampling run. Given the large amount of memory that could be necessary for computation and storage, we offer the possibility to store the results of a sampling run without computing the associated `Results` object, leaving this process for when the results have to be analysed.

The output formats available in GalaPy are:

- `pickle`: the standard Python serialisation protocol. `Results` object are computed at the end of a sampling run then serialised and stored in non-volatile memory. The typical size of the output file can reach up to ∼1 GB.
- `hdf5`: the Hierarchical Data Format (Folk et al. 2011), a widespread method for storing heterogeneous data. When using this format storage in non-volatile memory is possible in two flavours:
  - `light`: store only samples coordinates, likelihood values and weights along with minimal additional information to re-build the models used in the sampling (typical size 10 MB);
  - `heavy`: along with the information available also with the `light` option, all the additional derived quantities computed when building the `Results` object are stored (typical size up to ∼1 GB).

Once stored, results can be accessed and analysed by users in any moment. We note that, when choosing the HDF5 format in either its heavy or light version, results can be accessed even without having to instantiate a `Results` object and can be loaded in memory as simple dictionaries or accessed as regular HDF5 files. The drawback of choosing lightweight storage is an additional overhead when instantiating the `Results` object for the analysis.

By instantiating or de-serialising the `Results` class several functions for statistical analysis, TEX table formatting and plotting are made available. This should guarantee quick access to data and user-friendliness. All the plots and tables provided in the following Sect. 4 have been produced using these tools.

3.4. Analysis

We distinguish among two broad categories of quantities that are stored and/or that can be computed after a sampling run:

- free parameters, are all the parameters that define the behaviour of the emission model chosen. These parameters define the size of the parameter-space that is sampled by the Monte Carlo algorithm of choice. Parameters of this kind can

be e.g. the age of the galaxy, its redshift, the indexes of the extinction power-law in Eqs. (18) and (20). A complete list of all the possible free-parameters is provided in Table B.3.
– derived parameters, are all those parameters that do not directly define an additional dimension in the parameter-space inspected by the sampler but can be computed by choosing a given position in the parameter space.

GalaPy provides several different tools for analysing the results of a sampling run. These tools are primarily accessible as functions of the `Results` class described in Sect. 3.3 and by importing the sub-package `galapy.analysis`. The latter contains two modules: `plot` and `funcs`, which respectively provide interfaces for plotting and generating formatted tables of different statistics measured both on the free-parameters, θ, and on the derived parameters, δ.

For each sampled position in the free-parameters space we have an associated value of the log-likelihood, $\ln \mathcal{L}(\theta_i)$, and a weight, $\mathbf{w}_i$. Furthermore, we pre-compute at the end of the run several derived quantities automatically such as, the full SED in the whole wavelength grid (as it is defined by the SSP library of choice), temperatures of the two ISM components, masses of the different components (stars, dust and gas), metallicities, star formation rate.

In Bayesian inference we want to get to an estimate of the free-parameters posteriors, $P(\theta|D)$, given a dataset, D, a model of the data depending on the free-parameters, θ, and some priors, $P(\theta)$. From the sampled posterior one can derive an estimate of the true value of each parameter, free $\hat{\theta}$ or derived $\hat{\delta}$, using an estimator (such as e.g. the weighted mean of samples). Monte Carlo techniques allow to derive a sample of positions in the parameter space from which we can get to an approximate estimate of the posterior. It is therefore possible to weight each position in the parameters space by the likelihood and compute weighted summary statistics and estimators.

The two samplers currently available in GalaPy provide different philosophies to approximate the posterior. The Monte Carlo Markov chain (MCMC) method implemented in `emcee` generates samples proportional to the posterior, so that

$$w_i \equiv 1 \;\forall\; w_i \in \mathbf{w}. \tag{67}$$

On the other hand, the Dynamic Nested Sampling algorithm used in `dynesty`, generates samples in nested (possibly disjoint) "shells" of increasing likelihood. The associated estimate of the posterior is then obtained by combining the set of samples with weights defined as

$$w_i \equiv \frac{1}{2}\left[\mathcal{L}(\theta_{i-1}) + \mathcal{L}(\theta_i)\right] \times \left[X_{i-1} - X_i\right] \tag{68}$$

where $\mathcal{L}(\theta_i)$ is the likelihood of the i-th sample and X_i is its associated volume of the prior[10] where the likelihood $\mathcal{L}(\theta_i) \geq \lambda$ is above some threshold λ.

When a sampling run converges, as already mentioned, we provide users with all the samples, their associated log-likelihoods and weights, along with derived-parameter values in all these positions. In this way users can choose to use their custom estimators to get to an estimate of the true values of these parameters. Conveniently though, we also provide functions for computing some useful estimators, accessible either from the `galapy.analysis` sub-package or the `Results` class.

[10] The prior bounds the algorithm to inspect only a finite region of the multidimensional parameters space, which would otherwise belong to $\mathbb{R}^N$, where N is the number of free parameters.

Along with the weighted average and standard deviation, percentiles and best-fitting value (i.e. the position in the parameter space among all those sampled where the log-likelihood has assumed its maximum value) we also give the possibility to compute credible intervals around a given position of the parameter space. All of these quantities are weighted with values from Eqs. (67) and (68).

In particular, we define the central credible interval for a marginalised parameter θ as that region of the parameter space enclosed in an interval $[\theta_{\rm low}, \theta_{\rm upp}]$ defined around the best-fitting (i.e. maximum likelihood) value of the parameter, $\theta_{\rm best}$. The limits of this interval are defined by

$$\int_{\theta_{\rm low}}^{\theta_{\rm best}} P(\theta|D)\,\mathrm{d}\theta = \frac{\alpha}{2} \tag{69}$$

for the lower bound, and

$$\int_{\theta_{\rm best}}^{\theta_{\rm upp}} P(\theta|D)\mathrm{d}\theta = \frac{\alpha}{2} \tag{70}$$

for the upper bound. A value of, for instance, $\alpha = 0.68$ gives the 68% credible interval.

For highly asymmetric or multi-modal marginalised posteriors, one of the two half-integrals in Eqs. (69) and (70) might not encompass enough samples to embed the requested probability value. In these cases only upper/lower limits on the parameter value can be retrieved and the equations become

$$\int_{\theta_{\rm low}}^{+\infty} P(\theta|D)\,\mathrm{d}\theta = \alpha \tag{71}$$

for lower limits and

$$\int_{-\infty}^{\theta_{\rm upp}} P(\theta|D)\mathrm{d}\theta = \alpha \tag{72}$$

for upper limits.

4. Validation

In this section, we present a sanity check for GalaPy. We both verify that all the components of the library are behaving as expected as well as validate the scientific return of the physical models proposed. Even though we limit our presentation here to the aspects involving the science that can be performed with our tool, we provide some preliminary discussion on the computational side and on software performances in Appendix A. A thorough discussion on performances and a comparison with other codes goes beyond the scope of this manuscript and is left for a future work.

We test our model on both mock and real observations of star-forming and quiescent objects. Star-forming objects are complex structures that can host several, if not all, of the different components implemented in GalaPy, making them an excellent test-bench for investigating the interplay between the modules building up our library. We first test the constraining capabilities of our machinery by building a set of mock observations of simulated galaxies with different physical properties and perform the regression with GalaPy (Sect. 4.1). In Sect. 4.2, we use the In situ SFH model (Sect. 2.1.3) along with our dust model (Sect. 2.3.1) on a set of real sources, in order to validate the reliability of these models on estimating the astro-physical properties of sources.

Table 1. Number of generated sources for the different combinations of SFH models.

SFH model	Active	Passive
In situ	20+20	20+20
Delayed exponential	20+20	20+20
Constant	20+20	–

Notes. The X+Y notation marks that the sample is divided in two equivalent sub-samples. For X sources, the spectroscopic redshift is provided. For Y sources, we consider the photometric redshift a free parameter of the model.

4.1. Validation on mock sources

In order to verify and prove the efficacy of GalaPy, we first tested it against mock observations generated with the library itself.

As anticipated, we generated a set of mock observations of galaxies simulated using the GalaPy modelling framework. We generated different mock sources by randomly sampling a flat prior space for different models of SFH. For each of these sources we used the `parsec22.nt` SSP library (see Sect. 2.2; in Appendix B.1 we compare the BC03 libraries with the PARSEC22 libraries and show that the results obtained are consistent independently on the choice of SSP library).

In particular, for the sake of investigating the library reliability on a broad parameter space volume, we generated both actively star-forming and passively evolving galaxies. Table 1 summarises the number of mock sources generated for this test. In particular, for each SFH model we generated two sets of objects: 20 sources for which a spectroscopic estimate of redshift exists, and 20 sources for which redshift has to be estimated photo-metrically (i.e. $z_{\rm phot}$ is a free-parameter of the model). Furthermore, for the in situ and delayed exponential SFH models, we both generated sources that have still an active star-formation and sources that are passively evolving. On the other hand, for the Constant SFH model, we only selected actively star forming objects, as this particular model is intended for objects that are undergoing a secular evolution of their stellar content (e.g. late-type galaxies). We therefore run our test on a total of 200 mock sources, 100 of which are assumed to not have a spectroscopic determination of redshift.

In practical terms, actively star-forming sources are generated by setting the $\tau_{\rm quench}$ parameter to an arbitrarily large value, consistent with infinite[11]. On the other hand, for passively evolving objects, we sampled a random value for the quenching time, $\tau_{\rm quench}$ and imposed that the age should be sampled from an interval of values that is upper-limited by the sampled value of $\tau_{\rm quench}$. Each mock source has been generated by sampling uniformly a position in the parameter space defined by the priors summarised in Table 2. The overall SED of actively star-forming galaxies therefore consists of the dust-attenuated emission from stars including nebular thermal emission and non-thermal synchrotron as well as thermal emission coming from the two different components of our dust model. On the other hand, the SED of passively evolving mock sources is given by the un-attenuated stellar emission, possibly including some left-over nebular thermal emission and non-thermal synchrotron, if quenching had been particularly recent. The latter would anyways be extremely sub-dominant and un-investigated given the photometric system shown in Fig. 14.

In order to build a mock photometric observation, we need to assume a photometric system. This is graphically shown in Figs. 15 and 14, where, as a function of wavelength, we show the transmission corresponding to the 24 band-pass filters we used for actively star-forming mock galaxies and the 12 used for passively evolving mock sources, respectively. We selected filters from different well known experiments, covering a wide range of wavelengths. While for passive objects we select filters from the UV/optical bands to the near-infrared, using transmissions from SDSS, 2MASSm and *Spitzer*, for active objects we extend the spectral coverage up to the sub-mm/mm bands adding also filters from *Herschel* and ALMA.

To add errors to our mock observation we first associated to each different transmitted flux of each single mock source an error that has been randomly chosen to be between 10% and 50% of the flux. With this value set for all the fluxes, we then generated a random realisation of the mock measurement by extracting it from a Gaussian distribution with mean equal to the real value of the transmitted flux and standard deviation equal to the random error.

The free-parameters chosen for the sampling runs are the same free parameters we varied in generating the mock sources (Table 2). We allowed for nine free parameters, including the age and SFH and ISM defining parameters, in actively star-forming galaxies. For passively evolving sources we instead varied four free parameters, including age of the galaxy, age of quenching, and SFH defining parameters. In both cases, for 100 out of the 200 sources, redshift has been set as an additional free-parameter. Consistently, we modelled each source with the same SFH model and SSP library used for generating it.

For the sampling of the free-parameters of our models, we assumed a set of uninformative uniform priors whose limits correspond to those listed in Table 2, namely, the same intervals defining the parameter space volume sampled by the mock sources. For each source, we ran a dynamic nested sampling using `dynesty` with default GalaPy sampling hyper-parameters and stopping criterion[12].

Sampling runs take (on average) approximately 15 min per source to converge on eight physical cores of an Intel i9-10885H CPU @ 2.40 GHz with `x86_64` architecture. The time required for convergence strongly depends on the total number of samples extracted. When running with dynamic sampling, this number is not known in advance (more details in Speagle 2020). Our runs typically converge with a total number of valid samples between 10 to 20 thousands, each with a different weight. This does not reflect the actual number of likelihood calls for (given an average efficiency between 1% to 5% for these kinds of problems) a range spanning $5 \div 10 \times 10^5$.

4.1.1. Results for the whole sample

In Fig. 16, we show the distribution of the reduced χ^2 values for the best-fitting set of parameters obtained by means of the dynamic nested sampling runs detailed above. With a dashed

[11] A value $\tau_{\rm quench} \geq 2 \times 10^{10}$ Gyr is enough as it will always prove larger than the Age of the Universe at any epoch.

[12] Besides the posterior/evidence split and stopping function mentioned in Sect. 3.2, we set a higher-bound stopping criterion corresponding to the maximum effective number of likelihood calls $\max(N_{\rm eff}) = 5 \times 10^6$. Our initial tolerance is set to $\Delta \ln \hat{\mathcal{Z}} \lesssim 0.05$ with an initial maximum number of iterations `maxiter_init` $= 10^4$. We then add iteratively 10 batches of new live-points with a maximum number of iterations per batch corresponding to `maxiter_batch` $= 10^3$. We use the multiple-ellipsoidal decomposition (Feroz et al. 2009) as bounding criterion.

Table 2. Free parameters and priors for the mock galaxies of Sect. 4.1.

In situ		Delayed exponential		Constant	
Actively star-forming					
Parameter	Prior	Parameter	Prior	Parameter	Prior
log age (yr)	[7, 9]	log age (yr)	[7, 9]	log age (yr)	[7, 9]
redshift	[2, 8]	redshift	[2, 8]	redshift	[2, 8]
log $\psi_{\rm max}(\dot{M}_\odot)$	[0, 4]	log $\psi_{\rm norm}(\dot{M}_\odot$	[−2, 2]	log $\psi(\dot{M}_\odot)$	[−2, 2]
log $\tau_\star$ (yr)	[7, 9]	log $M_{\rm dust}(M_\odot)$	[7, 10]	log $M_{\rm dust}(M_\odot)$	[7, 9]
$f_{\rm MC}$	[0, 1]	$f_{\rm MC}$	[0, 1]	$f_{\rm MC}$	[0, 1]
log $N_{\rm MC}$	[0, 5]	log $N_{\rm MC}$	[0, 5]	log $N_{\rm MC}$	[0, 5]
log $R_{\rm MC}$/pc	[0, 5]	log $R_{\rm MC}$ (pc)	[0, 5]	log $R_{\rm MC}$ (pc)	[0, 5]
log $\tau_{\rm esc}$ (yr)	[4, 8]	log $\tau_{\rm esc}$/yr	[4, 8]	log $\tau_{\rm esc}$ (yr)	[4, 8]
log $R_{\rm DD}$ (pc)	[0, 5]	log $R_{\rm DD}$ (pc)	[0, 5]	log $R_{\rm DD}$ (pc)	[0, 5]
$f_{\rm PAH}$	[0, 1]	$f_{\rm PAH}$	[0, 1]	$f_{\rm PAH}$	[0, 1]
Passively evolving					
Parameter	Prior	Parameter	Prior	–	
log age (yr)	[9.2, 11]	log age (yr)	[9.2, 11]		
redshift	[0, 2]	redshift	[0, 2]		
log $\tau_{\rm quench}$ (yr)	[8, 9]	log $\tau_{\rm quench}$ (yr)	[8, 9]		
log $\psi_{\rm max}(\dot{M}_\odot)$	[2, 4]	log $\psi_{\rm norm}(\dot{M}_\odot)$	[−2, 2]		
log $\tau_\star$ (yr)	[7, 9]	log $\tau_\star$ (yr)	[7, 9]		

Notes. For each SFH model and both for actively star forming and passively evolving mock sources, we list the parameter symbols (normalised to their unit) and the prior upper and lower limits. The symbols used are consistent to those used in Sect. 2 and summarised in Table B.3.

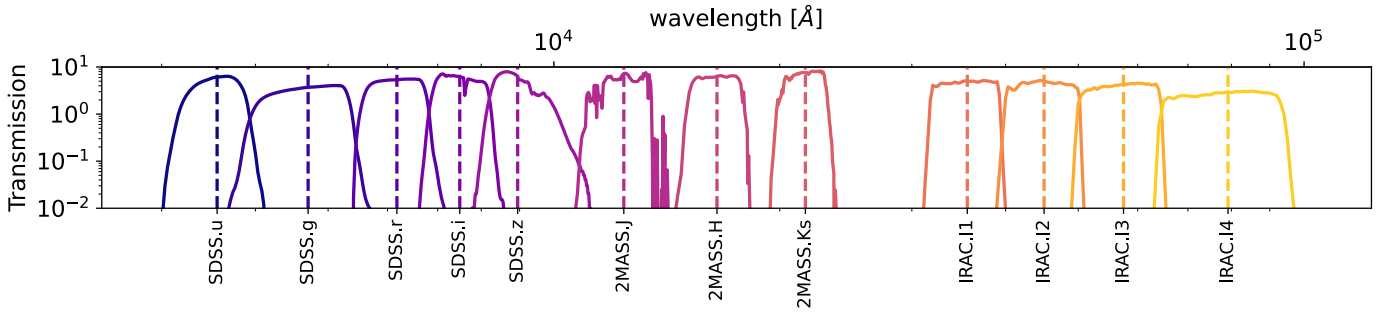

Fig. 14. Same as Fig. 15 for the mock observation of the passively evolving simulated galaxies of Sect. 4.1.

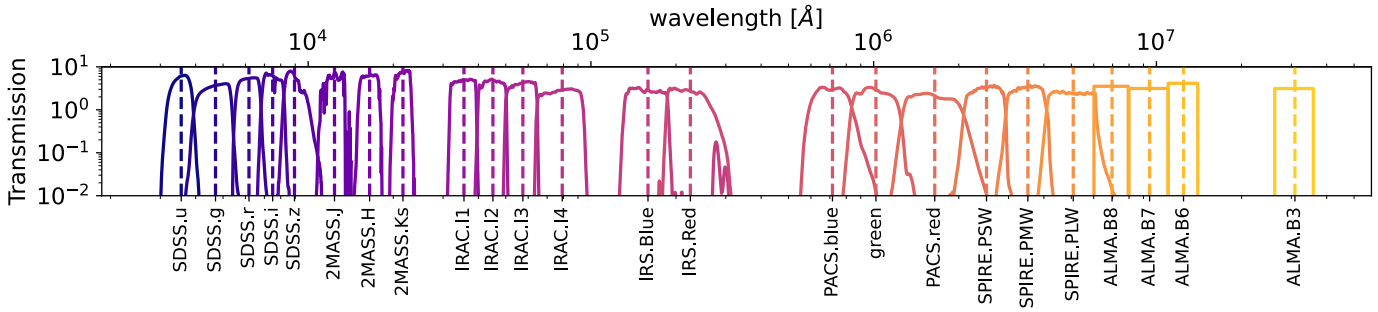

Fig. 15. Photometric system used to generate the mock observation for the actively star-forming simulated galaxies of Sect. 4.1. The lower x-axis shows the keyword name of the band-pass transmission while the upper x-axis shows the corresponding wavelength in angstroms. Transmissions are expressed in terms of photons and the dashed lines mark the position of the pivot wavelength for each band-pass filter. We note that this is just a possible set-up specific to the case of the mock galaxy of Sect. 4.1. It represents only a sub-set of the band-pass transmissions available in the GalaPy database.

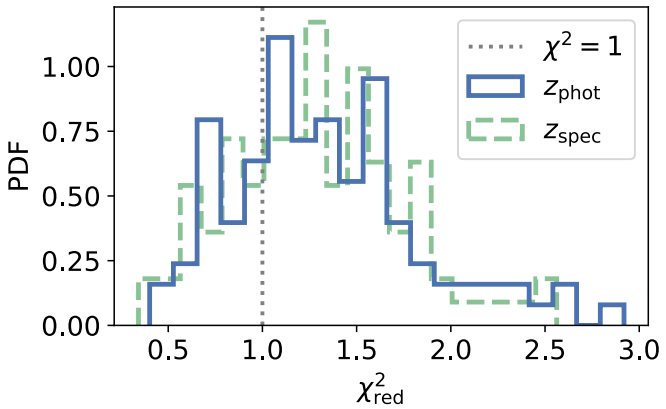

Fig. 16. Distribution of the reduced χ^2 for the best-fitting parameters of the two sets of sources (with and without spectroscopic redshift in dashed green and solid blue, respectively), as obtained by sampling the free-parameter space with `dynesty`.

green line we mark the distribution of the 100 sources for which we assumed a value of the spectroscopic redshift was available, while the solid blue histogram marks the distribution of the 100 sources for which redshift was a free-parameter. As a term of comparison, we show as a dotted vertical gray line the value corresponding to the expectanion value $\chi^2 = 1$. From the histograms, we can appreciate how, for almost all the sources, $0.5 \leq \chi^2_{\rm red} \leq 2.5$.

In Fig. 17, we show the values for a collection of relevant free and derived parameters obtained by computing the weighted median of the samples with errors given by the 16th and 84th percentiles, namely, the limits of a credible interval embedding a 68% probability. Different symbols and colours mark different combinations of SFH model and active/passive evolution, as detailed in the caption. Results are given in terms of the ratio between the measured and real value of the parameter. The top three panel show quantities available for all sources, actively star-forming and passively evolving (i.e. age of the galaxy, redshift and stellar mass), while the lower three panels show quantities that are defined only for the actively star-forming objects (i.e. dust mass, gas mass and current star formation rate). We highlight that, even though Fig. 17 collects only posterior values obtained for the 100 sources without spectroscopic redshift, equivalently consistent results have been found for the other set of sources.

The ratios in Fig. 17 show that the true value of each parameter is within the 68% credible interval for $\gtrsim 90\%$ of the mock observations, with this percentage increasing considerably if accounting for a 95% credible interval. In particular, photometric redshift, stellar mass and star formation rate show an exquisite agreement with the expected value.

It is also interesting to focus on the $M_{\rm dust}$ and $M_{\rm gas}$ parameters. As already discussed in Sect. 2.1, the method used to estimate these quantities in the In situ SFH model is different with respect to other empirical models. In particular, while for empirical models of SFH $M_{\rm dust}$ and $M_{\rm gas}$ are free-parameters, the in situ model predicts their value analytically, based on the SFH. It is therefore relevant that the estimates obtained by the In situ model (blue circles in Fig. 17) show a smaller error and a better agreement with the real value, with respect to the larger error-bars and scatter shown by the delayed exponential (green triangles) and constant (red squares) models.

We can conclude that the machinery we have built successfully retrieves the correct representation of data. We highlight that the collection of sources used for testing has been selected randomly from a considerably large parameter-space volume without any prescription for the mock observation to be representative of any real source population. Nonetheless the agreement of the results is almost perfect in all the dimensions, demonstrating how the tool is not limited to specific populations of objects and does not require a high level of fine tuning to get to a significant result. This is reflected on the small scatter of the marginalised posteriors of the parameters (as shown in Fig. 17).

4.1.2. Demonstration on a mock source of available tools

Before moving to the analysis of real sources, we select one out of the 200 mock observations generated in the previous section and we show more in detail the results inferred by the analysis of the posteriors on both the free and derived parameters. In particular, as it will be subject to a stress test on real sources on the following section, we pick one of the actively star-forming galaxies generated by sampling the priors defined for an in situ model of SFH (Table 2).

In Fig. 18, we compare the original mock observation (blue empty round markers with error-bars) with the model favoured by the free-parameters posterior distribution. The black solid line marks the best-fit model that results in a reduced $\chi^2_{\rm red} = 1.07$, also reported in the lower panel. With shades of grey we show the 1- and 2-σ confidence regions around the mean SED (in solid grey). Coloured solid lines show instead the contributions to the best-fitting SED, coming from the different components building up our galaxy model. It is worth to highlight the different contributions of molecular clouds and diffuse dust to the peak of dust emission, that naturally blend into the final SED to represent a wider distribution of emission in the MIR-FIR.

In the lower panel of Fig. 18, we show the standardised residuals between the best-fitting model and the observed fluxes. These values are defined as

$$\chi_i = \frac{S_{\rm O}(\lambda_i) - S_{\rm M}(\lambda_i)}{\sigma_{\rm O}(\lambda_i)}, \tag{73}$$

where $S_{\rm O}(\lambda_i)$ and $\sigma_{\rm O}(\lambda_i)$ are, respectively, the flux and error on of the datapoint at pivot wavelength λ_i and $S_{\rm M}(\lambda_i)$ is the corresponding modelled flux. We can see how the best-fit model correctly intercepts the mock-observation while being within 1-σ from the mean of the samples. This agreement is reflected on the marginalised posterior probability of the free-parameters.

Figure 19 shows the triangle plot for a sub-set of the free-parameters posteriors marginalised to 1D and 2D. These marginal posterior probabilities are given as histograms on the diagonal and as grey contours, for the 1D and 2D cases, respectively. We do not show all the parameters to avoid burdening the discussion, so we limit the demonstration to the overall galaxy parameters (i.e. age and redshift), the normalisation of SFH parameter (i.e. $\psi_{\rm max}$), and to the fraction of total emitted diffuse-dust energy contributed by PAH (i.e. $f_{\rm PAH}$). Specifically, the black dashed lines intercept the weighted median value of the samples, darker and lighter grey contours mark the 68% and 95% credible regions (note that, on top of each 1-dimensional marginalised posterior the median value for each parameter with the corresponding 68% credible interval is also reported). As a term of comparison we also show, with orange solid lines, the fiducial value of each parameter. All these fiducial values fall within the 68% credible interval. It is worth to highlight the

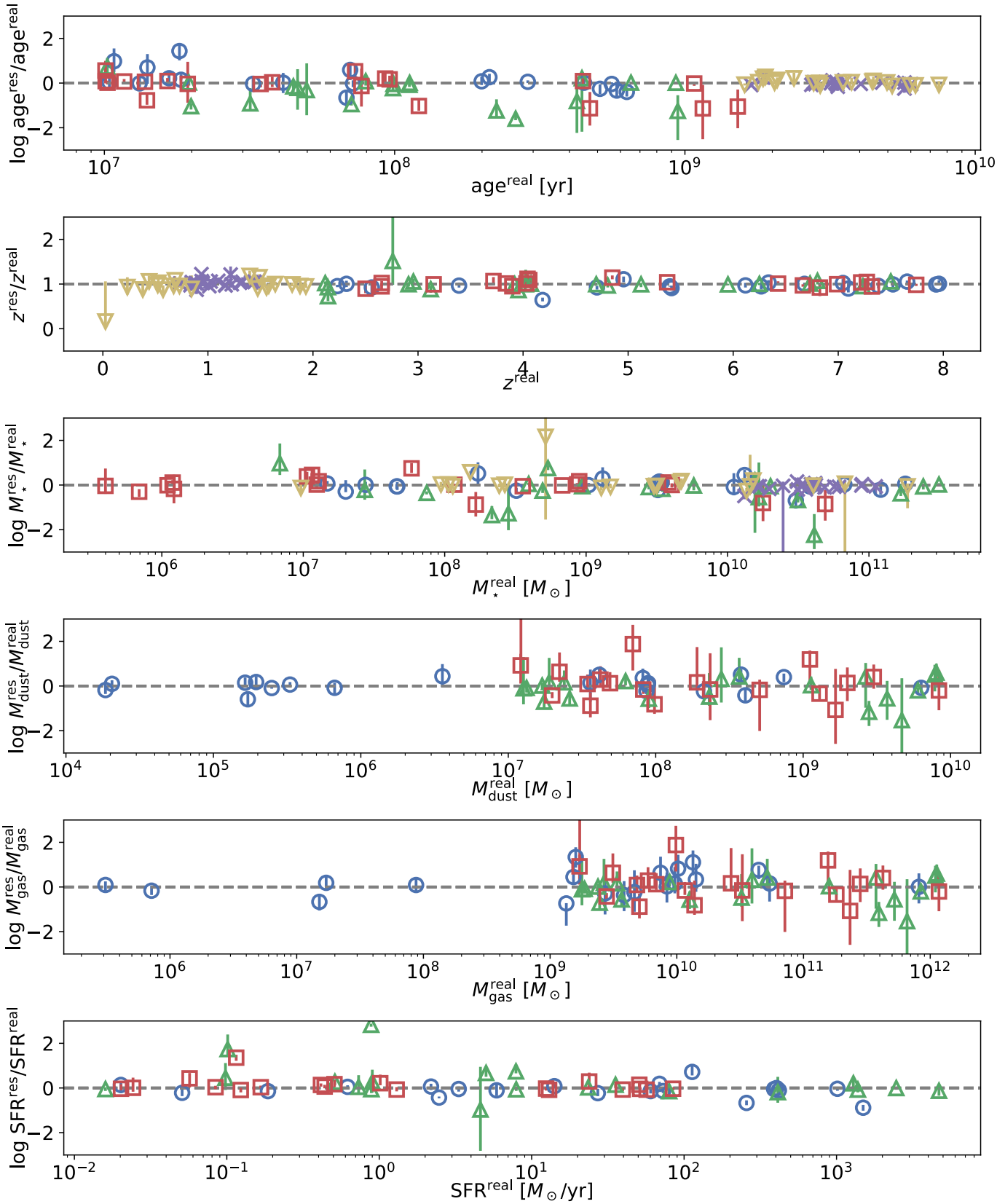

Fig. 17. Results for some of the free and derived parameters obtained by fitting with GalaPy the mock observation set generated without spectroscopic redshift detection (see Sect. 4.1). The results are presented in terms of the ratio between the median value of the weighted samples collected by running the `dynesty` sampler. Errors are given in terms of the 16th and 84th percentiles, defining a 68% credible interval around the median, as detailed in Sect. 3.4. Blue circles, green upward triangles and red squares are actively star-forming sources modelled with an in situ, delayed exponential and constant SFH model, respectively. Violet crosses and yellow downward triangles are instead passively evolving sources modelled with an in situ and delayed exponential SFH model, respectively.

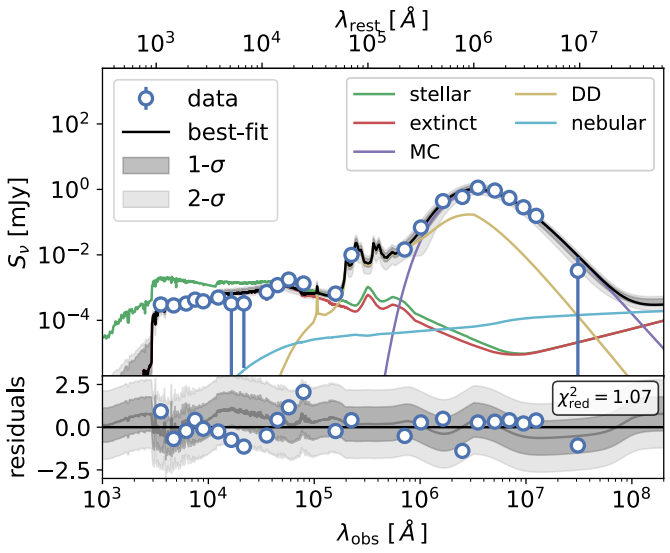

Fig. 18. Results of the parameters sampling for one of the mock observation of Sect. 4.1. Upper panel: best-fit SED (black solid line) and components (coloured lines) compared to the mock observation data (empty blue markers); 1-σ and 2-σ confidence around the mean of the samples is also shown with grey shaded areas. Lower panel: residuals with respect to the best-fit sample (black); mean and 1-/2-σ confidence regions are also shown with a grey solid line and shaded areas, respectively.

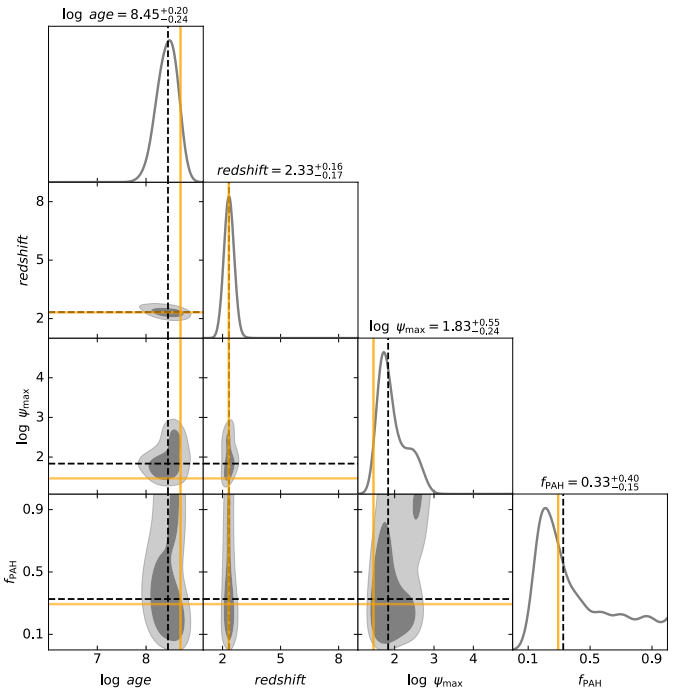

Fig. 19. Corner plot of the parameter posteriors obtained for the mock galaxy selected in Sect. 4.1.2, illustrating the one- and two-dimensional marginalised posteriors for a subset of the parameter space dimensions (for clarity and visualisation reasons). The shaded grey regions highlight (68%, 95%) confidence from darker to lighter, corresponding to (1-σ, 2-σ). The dashed black lines mark the position of the weighted median value of parameters while the values above the diagonal panels show the median and 68% percentile around the median (roughly corresponding to 1-σ confidence in a Gaussian approximation). As a term of comparison, we also show mark, with orange solid lines, the real value of each parameter.

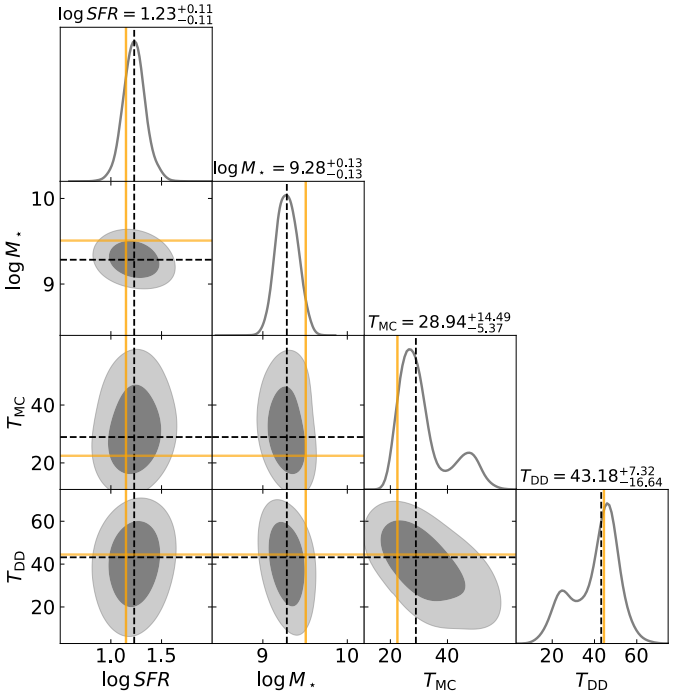

Fig. 20. Same as Fig. 19 for a sub-set of the derived parameters reported in Fig. 17.

goodness of our fit for the photometric redshift estimate, as it an extremely sought after quantity and the algorithm was able to correctly infer it with an error of ∼5%.

We can use the results obtained with the sampling algorithm to build probability densities also for the derived parameters, as it is shown in Fig. 20. We have selected four interesting derived parameters, namely the star formation rate, SFR, stellar mass, $M_\star$ and temperatures of the molecular, T_{MC}, and diffuse dust, T_{DD}, components. The probability densities in 2-dimensional and 1-dimensional space are then built by computing the derived parameter's values in each position of the free-parameters space, as defined by the galaxy emission model used to represent the mock dataset. Once again, we over-plot with orange solid lines the fiducial value of each parameter which, in all the cases, falls within the region encompassing 68% of the total probability. It is interesting to observe, on the 1D marginalised probabilities of the two temperatures, how in both cases there is a secondary peak that is symmetric in the two components. This is not a surprise as the two dust components compete in contributing both to the absorption at short wavelengths and to the re-emission at longer wavelengths. The modelling we have implemented it's nevertheless successful in distinguishing between the two, therefore favouring one of the two solution over the other.

4.2. Test on real sources

To validate the scientific throughput of the new models introduced in GalaPy, we compared their predictions with those performed with other SED-fitting codes and with different models for the SFH and for the dust-model. To this end, we selected a small set of sources with interesting properties from several different works. Most of the objects inspected in this section are either high-redshift dusty star-forming galaxies or their supposed descendants at low redshift, namely, quiescent galaxies.

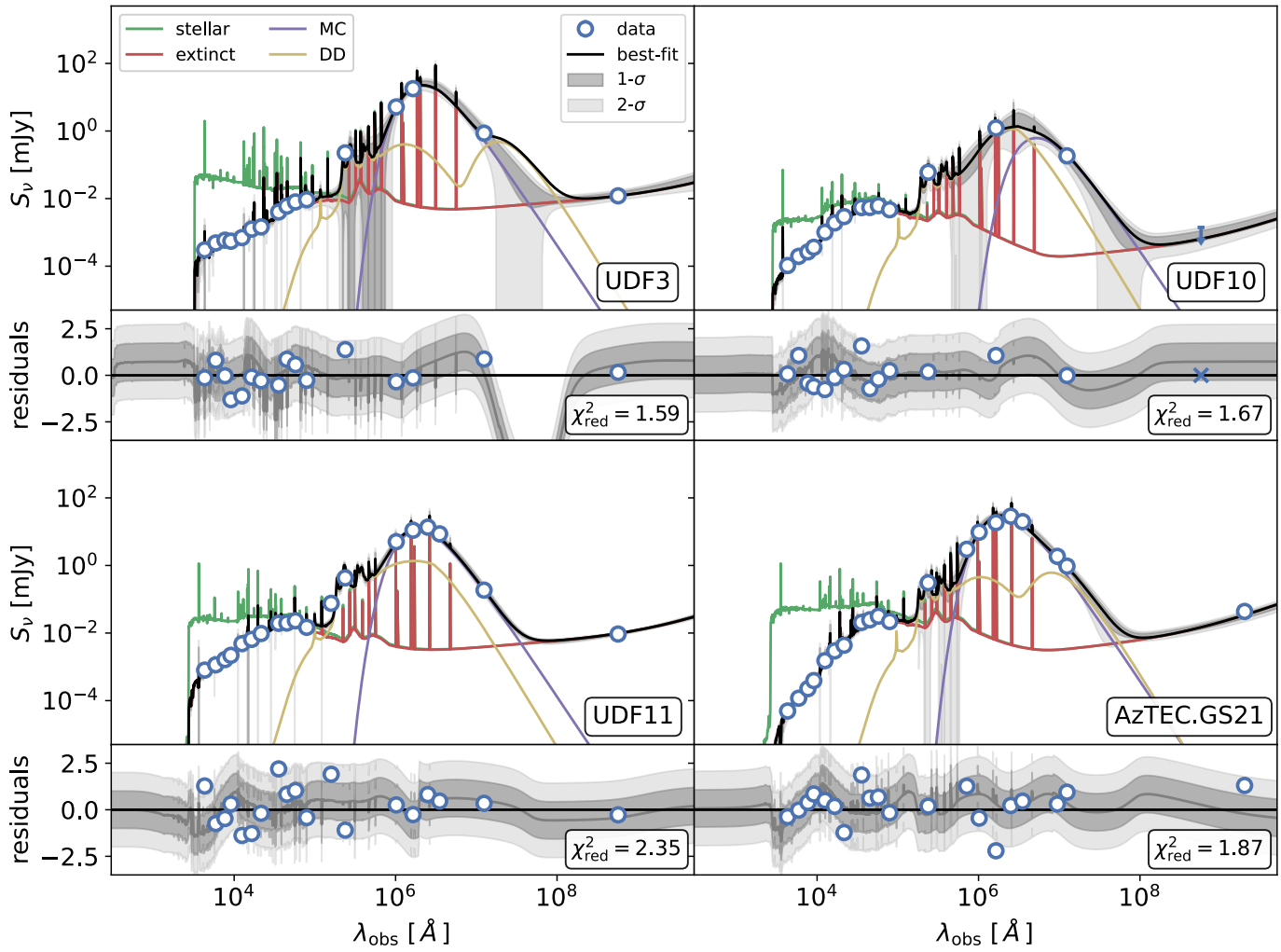

Fig. 21. Results for the four dusty star forming galaxies of Sect. 4.2.1 Upper panels: GalaPy fits to the photometric data of four galaxies selected from the Pantoni et al. (2021) sample (blue markers with error-bars). Best-fitting model (black solid line), different components (colour-coded as in legend), and 1- and 2-σ confidence levels (grey shaded regions) around the mean of the samples are also shown. Lower panels: Standardised residuals with respect to the best-fitting galaxy models, and 1-/2-σ confidence intervals around the mean of the samples (grey shaded regions) are shown; the reduced χ^2 of the fit is reported in each sub-panel.

We underline that a validation of other models of SFH has already been performed on mock sources in the previous section. Furthermore, as a thorough analysis of the sources would go beyond the scope of this validation section, we limit the discussion to a simple comparison of the results obtained fitting the SEDs with GalaPy with those obtained with other techniques, as they appear in the literature.

4.2.1. Dusty star-forming galaxies from Pantoni et al. (2021)

In the work from Pantoni et al. (2021, P21 hereafter), a set of 11 sources selected from 3 millimetre catalogues (ALMA data from Dunlop et al. 2017, reprocessed within the ARI-L project, Massardi et al. 2021; LABOCA, Yun et al. 2012; AzTEC, Targett et al. 2013) in the GOOD-S field and complemented with fluxes from several other bands available in the field. The selected sources are objects at the peak of cosmic star formation history, strongly attenuated by dust, with redshift spanning between $1.5 < z \lesssim 2.5$. The dataset comes with the advantage of having spectroscopic redshifts and a pan-chromatic coverage spanning from visible to radio bands.

In P21 SEDs have been fitted using the CIGALE code (Boquien et al. 2019), assuming a delayed exponential SFH and the BC03 SSP library. To account for the excess in NIR/MIR, the authors also included a power-law component that should ideally model the combination of diffuse dust, PAH and AGN. For our analysis, we instead used our in situ SFH model and the PARSEC22 SSPsm including lines from nebular regions. Our two-components dust-model automatically accounts for the NIR/MIR excess.

We selected four sources among the 11 available that are not classified as AGNs in the NIR and MIR regions of the spectrum. All the sources have at least 15 observations spanning from optical to radio. We fit the photometries letting ten model parameters vary, including age, SFH and dust properties.

The best-fitting SEDs obtained are shown as black solid lines in Fig. 21 along with the several components contributing to emission (coloured solid lines) and 1-/2-σ confidence around the mean. Band fluxes for each source are marked with empty blue circles. In UDF10, the flux measured with VLA in the radio band is not a detection but was classified as an upper limit, we mark it with a downward arrow in the upper part of the panel. In the

Table 3. Comparison of the quantities derived by fitting with GalaPy the several sources of Sect. 4.2 with the values found in the literature.

Source	GalaPy				Literature		
	log age (log yr)	SFR ($M_\odot\,\mathrm{yr}^{-1}$)	$T_{\rm MC}$ (K)	$T_{\rm DD}$ (K)	log age (log yr)	SFR ($M_\odot\,\mathrm{yr}^{-1}$)	$T_{\rm dust}$ (K)
Sect. 4.2.1: Dusty star-forming (1)							
UDF3	$7.58^{+0.72}_{-0.51}$ (7.35)	$555.80^{+164.87}_{-89.86}$ (702.42)	$53.44^{+6.56}_{-2.76}$ (55.47)	$18.49^{+19.23}_{-9.03}$ (5.54)	$8.37^{+0.08}_{-0.10}$	519 ± 38	73 ± 4
UDF10	$9.41^{+0.06}_{-0.08}$ (9.50)	$22.88^{+6.64}_{-5.96}$ (18.76)	$32.04^{+20.52}_{-9.95}$ (18.69)	$36.37^{+19.41}_{-10.11}$ (50.12)	$8.96^{+0.06}_{-0.07}$	41 ± 5	46 ± 7
UDF11	$8.85^{+0.14}_{-0.19}$ (8.80)	$185.70^{+25.00}_{-16.94}$ (182.34)	$55.10^{+6.36}_{-3.49}$ (52.61)	$39.53^{+11.86}_{-8.57}$ (46.84)	$8.58^{+0.08}_{-0.11}$	241 ± 19	69 ± 4
AzTEC.GS21	$8.35^{+0.20}_{-0.23}$ (7.99)	$335.05^{+56.67}_{-58.79}$ (241.24)	$52.61^{+3.68}_{-3.14}$ (50.22)	$16.26^{+5.16}_{-4.91}$ (10.55)	$8.87^{+0.06}_{-0.07}$	360 ± 18	63 ± 3
Sect. 4.2.2: Local late-type (2)							
NGC3364	$9.92^{+0.12}_{-0.18}$ (10.03)	$0.76^{+0.09}_{-0.08}$ (0.84)	$25.25^{+13.08}_{-2.92}$ (33.10)	$21.71^{+33.15}_{-2.54}$ (36.30)	–	1.64 ± 0.23	22.05 ± 1.14
NGC3898	$9.91^{+0.07}_{-0.11}$ (9.91)	$0.16^{+0.04}_{-0.03}$ (0.19)	$20.39^{+16.75}_{-3.80}$ (18.33)	$16.94^{+14.10}_{-2.94}$ (37.58)	–	1.22 ± 0.18	16.28 ± 4.87
NGC4254	$9.96^{+0.09}_{-0.12}$ (9.90)	$20.34^{+1.47}_{-1.43}$ (18.72)	$28.22^{+13.49}_{-3.61}$ (29.39)	$24.11^{+8.70}_{-3.50}$ (21.75)	–	5.15 ± 0.51	24.82 ± 0.65
NGC4351	$9.35^{+0.20}_{-0.13}$ (9.26)	$0.48^{+0.06}_{-0.07}$ (0.45)	$23.29^{+2.98}_{-1.48}$ (22.40)	$23.11^{+34.68}_{-7.08}$ (67.13)	–	0.11 ± 0.02	21.06 ± 1.42
Sect. 4.2.3: Lensed NIR-dark with upper-limits (3)							
J1135	$8.34^{+0.43}_{-0.60}$ (8.55)	$815.97^{+195.03}_{-159.43}$ (641.09)	$51.44^{+32.67}_{-22.38}$ (45.40)	$44.01^{+3.62}_{-6.58}$ (48.15)	– (8.0)	$933.25^{+345.77}_{-157.01}$	37.7 ± 1.5
Sect. 4.2.4: Stacked NIR-dark radio-selected (4)							
Median	$7.96^{+0.39}_{-0.28}$ (7.94)	$1475.12^{+442.71}_{-558.84}$ (1838.94)	$9.51^{+21.54}_{-6.58}$ (11.05)	$15.29^{+17.55}_{-4.65}$ (19.52)	$8.38^{+0.41}_{-0.22}$ (8.38)	$395.37^{+58.57}_{-145.91}$ (430)	$61.35^{+4.8}_{-20.4}$ (61.45)
Sect. 4.2.5: Quiescent (5)							
ETG1	$9.46^{+0.36}_{-0.03}$ (9.82)	0.0 –	– –	– –	9.45	10^{-7}	–
ETG2	$9.62^{+0.04}_{-0.07}$ (9.60)	0.0 –	– –	– –	9.92	10^{-5}	–
ETG3	$9.41^{+0.32}_{-0.08}$ (9.75)	0.0 –	– –	– –	9.80	$3 \cdot 10^{-5}$	–

Notes. The values have been obtained with combinations of SED fitting and post-processing. SED fits have been performed using MAGPHYS-PHOTOZ (Da Cunha et al. 2008; Battisti et al. 2019) for the median of Sect. 4.2.4 and CIGALE (Boquien et al. 2019) in all other cases. A long dash in a cell indicates that the corresponding value is either not available from literature or it does not have meaning in the model set-up used for the corresponding source.

References. (1) Pantoni et al. (2021); (2) Casasola et al. (2020); (3) Giulietti et al. (2023); (4) Behiri et al. (2023); (5) Donevski et al. (2023).

lower part of each panel we also plot the standardised residuals with respect to the best-fitting model and the reduced χ^2 value. We mark the location of the corresponding radio upper limit of UDF10 with a cross in the residuals plot.

We report the main best-fitting values of free and derived parameters obtained for the 4 galaxies in Tables 3 and 4. We also include the values for the same set of parameters obtained in the original P21 work as a term of comparison. Given that we implement models that are significantly different from the models used in P21, the values obtained with GalaPy do not match exactly those obtained in the original work. Nonetheless, the inferred properties (e.g. the dust and gas masses as reported in Table 4) are in good agreement with the ones obtained in the original work. It as to be further noted that, in the original work these quantities, along with the gas metallicity, were not derived directly from SED-fitting but using post-processing and ALMA-bands emission line analysis.

Focussing on the temperatures of the two dust components, we can first of all notice that, as expected, in most of the cases the MC component is the one having the highest temperature. On the other hand, it is interesting to notice how UDF10 shows a higher temperature in the DD component, while it is significantly older than the other three objects (i.e. age of $\sim 10^{9.5}$ yr). If we compare it to the right panel of Fig. 11, this result suggests that the object could be older than its characteristic escape time from molecular clouds and therefore be in a late stage of evolution. We infer that galaxy is significantly older than its characteristic escape time from molecular clouds, whose predicted value is

Table 4. Same as Table 3 for further properties found in the literature.

Source	GalaPy					Literature			
	log $M_{\rm dust}$ (log $M_\odot$)	log $M_{\rm gas}$ (log $M_\odot$)	log $M_\star$ (log $M_\odot$)	$Z_{\rm gas}$	$Z_\star$	log $M_{\rm dust}$ (log $M_\odot$)	log $M_{\rm gas}$ (log $M_\odot$)	log $M_\star$ (log $M_\odot$)	$Z_{\rm gas}$
Sect. 4.2.1: Dusty star-forming (1)									
UDF3	$8.52^{+0.60}_{-0.64}$	$10.24^{+0.91}_{-1.07}$	$10.19^{+0.50}_{-0.24}$	$0.03^{+0.04}_{-0.02}$	$0.02^{+0.02}_{-0.01}$	$8.30^{+0.15}_{-0.22}$	–	$10.95^{+0.05}_{-0.05}$	–
	(8.31)	(9.95)	(9.98)	(0.03)	(0.02)		(10.6)		(0.01)
UDF10	$8.66^{+0.27}_{-0.27}$	$10.28^{+0.49}_{-0.34}$	$10.60^{+0.05}_{-0.07}$	$0.03^{+0.01}_{-0.01}$	$0.02^{+0.00}_{-0.01}$	$7.98^{+0.23}_{-0.53}$	–	$10.40^{+0.05}_{-0.06}$	–
	(8.64)	(10.24)	(10.65)	(0.03)	(0.02)		(10.5)		(0.01)
UDF11	$9.25^{+0.09}_{-0.14}$	$11.50^{+0.27}_{-0.43}$	$10.71^{+0.08}_{-0.09}$	$0.01^{+0.01}_{-0.00}$	$0.01^{+0.00}_{-0.00}$	$7.86^{+0.16}_{-0.26}$	–	$10.81^{+0.06}_{-0.07}$	–
	(9.19)	(11.53)	(10.64)	(0.01)	(0.00)		(10.3)		(0.01)
AzTEC.GS21	$8.84^{+0.28}_{-0.31}$	$10.16^{+0.42}_{-0.40}$	$10.93^{+0.18}_{-0.17}$	$0.06^{+0.01}_{-0.02}$	$0.04^{+0.01}_{-0.01}$	$8.46^{+0.09}_{-0.12}$	–	$11.26^{+0.05}_{-0.05}$	–
	(8.22)	(9.41)	(10.81)	(0.08)	(0.05)		(10.8)		(0.02)
Sect. 4.2.2: Local late-type (2)									
NGC3364	$7.17^{+0.15}_{-0.22}$	$8.97^{+0.18}_{-0.27}$	$10.03^{+0.07}_{-0.08}$	$0.02^{+0.00}_{-0.00}$	$0.02^{+0.00}_{-0.00}$	$7.10^{+0.08}_{-0.10}$	–	$10.24^{+0.15}_{-0.22}$	–
	(7.37)	(9.21)	(10.05)	(0.02)	(0.02)				
NGC3898	$6.29^{+0.16}_{-0.14}$	$7.89^{+0.18}_{-0.15}$	$10.40^{+0.05}_{-0.07}$	$0.03^{+0.00}_{-0.00}$	$0.03^{+0.00}_{-0.00}$	$7.72^{+0.09}_{-0.12}$	–	$11.32^{+0.28}_{-0.03}$	–
	(6.38)	(7.99)	(10.39)	(0.03)	(0.03)		(9.14)		
NGC4254	$9.11^{+0.14}_{-0.14}$	$10.68^{+0.20}_{-0.17}$	$11.13^{+0.07}_{-0.08}$	$0.04^{+0.00}_{-0.00}$	$0.02^{+0.00}_{-0.00}$	$7.35^{+0.02}_{-0.02}$	–	$10.13^{+0.07}_{-0.08}$	–
	(8.95)	(10.48)	(11.15)	(0.04)	(0.03)		(9.53)		(0.01)
NGC4351	$6.30^{+0.25}_{-0.20}$	$7.98^{+0.28}_{-0.24}$	$9.77^{+0.09}_{-0.05}$	$0.03^{+0.00}_{-0.00}$	$0.02^{+0.00}_{-0.00}$	$5.98^{+0.08}_{-0.10}$	–	$9.16^{+0.00}_{-0.00}$	–
	(6.17)	(7.82)	(9.75)	(0.03)	(0.02)		(7.92)		(0.01)
Sect. 4.2.3: Lensed NIR-dark with upper limits (3)									
J1135	$9.28^{+0.47}_{-0.62}$	$10.70^{+0.65}_{-0.72}$	$11.15^{+0.39}_{-0.57}$	$0.06^{+0.02}_{-0.03}$	$0.03^{+0.01}_{-0.02}$	9.06 ± 0.04	11.04 ± 0.04	$\lesssim 11.73$	–
	(9.34)	(10.62)	(11.40)	(0.07)	(0.04)				
Sect. 4.2.4: Stacked NIR-dark radio-selected (4)									
Median	$9.26^{+0.39}_{-0.47}$	$11.01^{+0.59}_{-0.91}$	$10.94^{+0.24}_{-0.25}$	$0.03^{+0.04}_{-0.01}$	$0.02^{+0.02}_{-0.01}$	$8.91^{+0.08}_{-0.60}$	–	$11.08^{+0.19}_{-0.15}$	–
	(9.35)	(11.39)	(10.86)	(0.01)	(0.01)	(8.06)	–	(11.09)	–
Sect. 4.2.5: Quiescent (5)									
ETG1	–	–	$9.97^{+0.17}_{-0.07}$	–	$0.04^{+0.01}_{-0.02}$	–	–	9.90	–
	–	–	(10.15)	–	(0.02)				
ETG2	–	–	$10.82^{+0.10}_{-0.07}$	–	$0.05^{+0.00}_{-0.01}$	–	–	10.73	–
	–	–	(10.92)	–	(0.05)				
ETG3	–	–	$10.48^{+0.16}_{-0.94}$	–	$0.04^{+0.01}_{-0.02}$	–	–	10.85	–
	–	–	(9.35)	–	(0.01)				

Notes. The gas metallicity values from literature have all been converted to absolute value from line estimates by means of $\log Z/Z_\odot = 12 + \log(O/H) - \log(O/H)_\odot$ (Nagao et al. 2006) assuming $\log(O/H)_\odot = 8.69$ and $Z_\odot = 0.0153$. For each source, we provide in the first row the median and 68% credible interval and on the second row, within round brackets, the best fitting value of the given parameter. A long dash in a cell indicates that the corresponding value is either not available from literature or it does not have meaning in the model set-up used for the corresponding source.

References. (1) Pantoni et al. (2021); (2) Casasola et al. (2020); (3) Giulietti et al. (2023); (4) Behiri et al. (2023); (5) Donevski et al. (2023).

$\log \tau_{\rm esc}^{\rm UDF10}/{\rm yr} = 6.34^{+0.80}_{-0.25}$, suggesting that it might be in a late stage of evolution. If we trust the In situ evolution scenario for the formation of ETG galaxies, about to approach a quiescent phase of evolution. From both the plots in Fig. 21 and the values reported in Tables 3 and 4, it is clear that the assumption of a two-component dust model ensures that the dust peak is better modelled with respect to the case of a single component. In particular, the possibility of having two peaks along with PAH emission, allows the overall MIR/FIR model to have more freedom to adapt to the dataset. This is evident from the flex appearing at $\lambda \approx 10^7$ Å in UDF3 and from the broadening of the peak in UDF10, both effects due to the blending of the two grey-bodies. We can identify a common trend for all the 4 galaxies in the sample as in the single-component estimate of the temperature from the original work, the resulting measurement is systematically over-estimated with respect to the higher temperature dust component in the two-component model used in this work.

With GalaPy, we can also easily derive the characteristic attenuation curves of the modelled galaxies (as detailed in Sect. 2.3.1). The average attenuation curve for the 4 galaxies is shown in Fig. 22 with solid black lines. We also plot for reference Calzetti-like (Calzetti et al. 2000) attenuation at varying value of the R_V parameter. For wavelengths bluer than $\lambda_V \sim 5500$ Å, our attenuation is consistent with $R_V \sim 4$ for UDF10, UDF11 and AzTEC.GS21, while it is $R_V \gtrsim 10$ for UDF3. For wavelengths

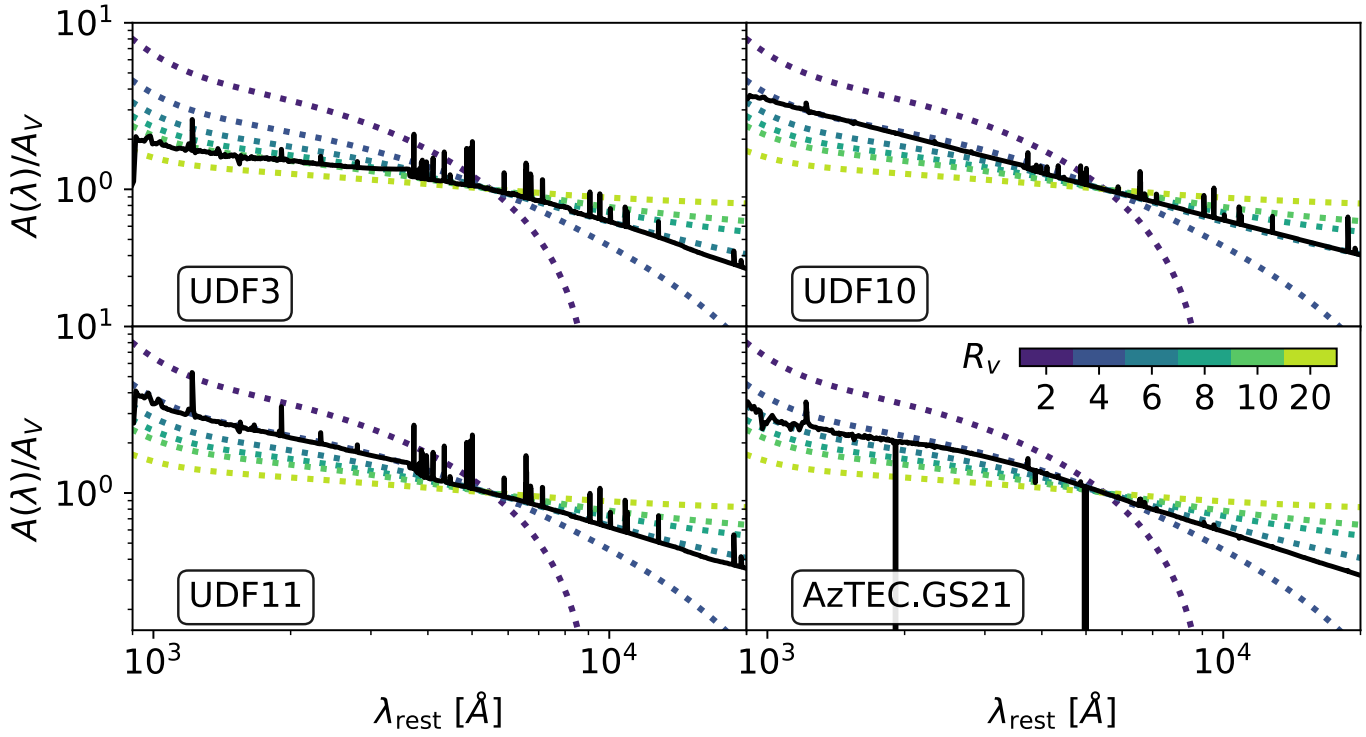

Fig. 22. Average attenuation curves (black solid line) predicted by GalaPy in the best-fitting models for the 4 galaxies of P21, normalised to the average value of attenuation in the $V-$band ($\approx$ 5500 Å). For reference, we also show Calzetti-like empirical attenuation curves for different values of the R_v parameter (colour-coded).

redder than λ_V all the galaxies have an attenuation that could be represented with $4 < R_V < 6$ Calzetti-shapes.

4.2.2. Local late-type galaxies from Casasola et al. (2020)

We extend the validation of our library to a small sub-sample of 4 local ($z < 0.01$) late-type galaxies from Casasola et al. (2020), extracted from the DustPedia database[13] (see also De Vis et al. 2019). The archive provides access to multi-wavelength imagery and photometry for 875 nearby galaxies as well as physical parameters for each galaxy (Davies et al. 2017; Clark et al. 2018) derived by means of the CIGALE code.

To further probe the reliability of GalaPy's results, we selected four galaxies that are not undergoing major interactions, do not show any nuclear activity in the X-ray, and are not classified as starburst. We once again selected our in situ model of SFH and the SSPs from the PARSEC22 library. We performed several exploration runs of the sampler on the four galaxies, and consequently decided for a model set-up completely equivalent to the one used for the sources of Sect. 4.2.1 for consistency and as the overall results were not differing substantially. Nonetheless, this is not intended to constitute a thorough analysis but just a sanity-check of GalaPy's validity. In terms of modelling set-up, we also allowed for an eventual systematic error parameterised as described in Sect. 3.1, marginalising the results over the nuisance parameter, f_{sys}.

The best-fitting model and 1-/2-σ confidence regions around the mean of the samples is compared against the multi-band photometry of the sources in Fig. 23 where, as usual, we also show the different contribution to the overall emission and the standardised residuals with respect to the best-fitting model and associated reduced χ^2. By inspecting the residuals in Fig. 23 we can see that the estimated best-fit model correctly intercepts the data-points even though, the extremely small errors on the optical flux measurements tend to make the nuisance systematic error parameter converge around values $f_{sys} \lesssim 0.1$.

A comparison of the derived astrophysical properties of the galaxies in the sample are provided in Tables 3 and 4. In particular, we show, for each source the median values with associated 68% credible interval, measured from the weighted posteriors, and the best fitting value (between parenthesis below each row of values). As expected, given the different shape of the delayed truncated SFH model used in the original work (Bianchi et al. 2018), the values for the SFR are in slight disagreement[14]. While the dust temperatures (Table 3) do agree within the error-bars, quantities that are more strictly related to the SFH model chosen (i.e. the age and SFR in Table 3, and the dust and gas masses in Table 4) deviate by more than 2-σ. On the other hand, for the stellar masses (also shown in Table 4) the agreement between the results is restored, even though this quantity also

[13] Available at dustpedia.astro.noa.gr. DustPedia is a collaborative focussed research project supported by the European Union under the Seventh Framework Programme (2007–2013) call (proposal no. 606847). The participating institutions are: Cardiff University, UK; National Observatory of Athens, Greece; Ghent University, Belgium; Université Paris Sud, France; National Institute for Astrophysics, Italy and CEA, France.

[14] A truncated SFH that does not drop to zero after truncation allows to both assume an early stage bulk of star formation and a late time constant stellar mass growth. This would in turn be similar to the approach we propose for ETGs in Sect. 4.2.5, if after truncation a null SFR is assumed.

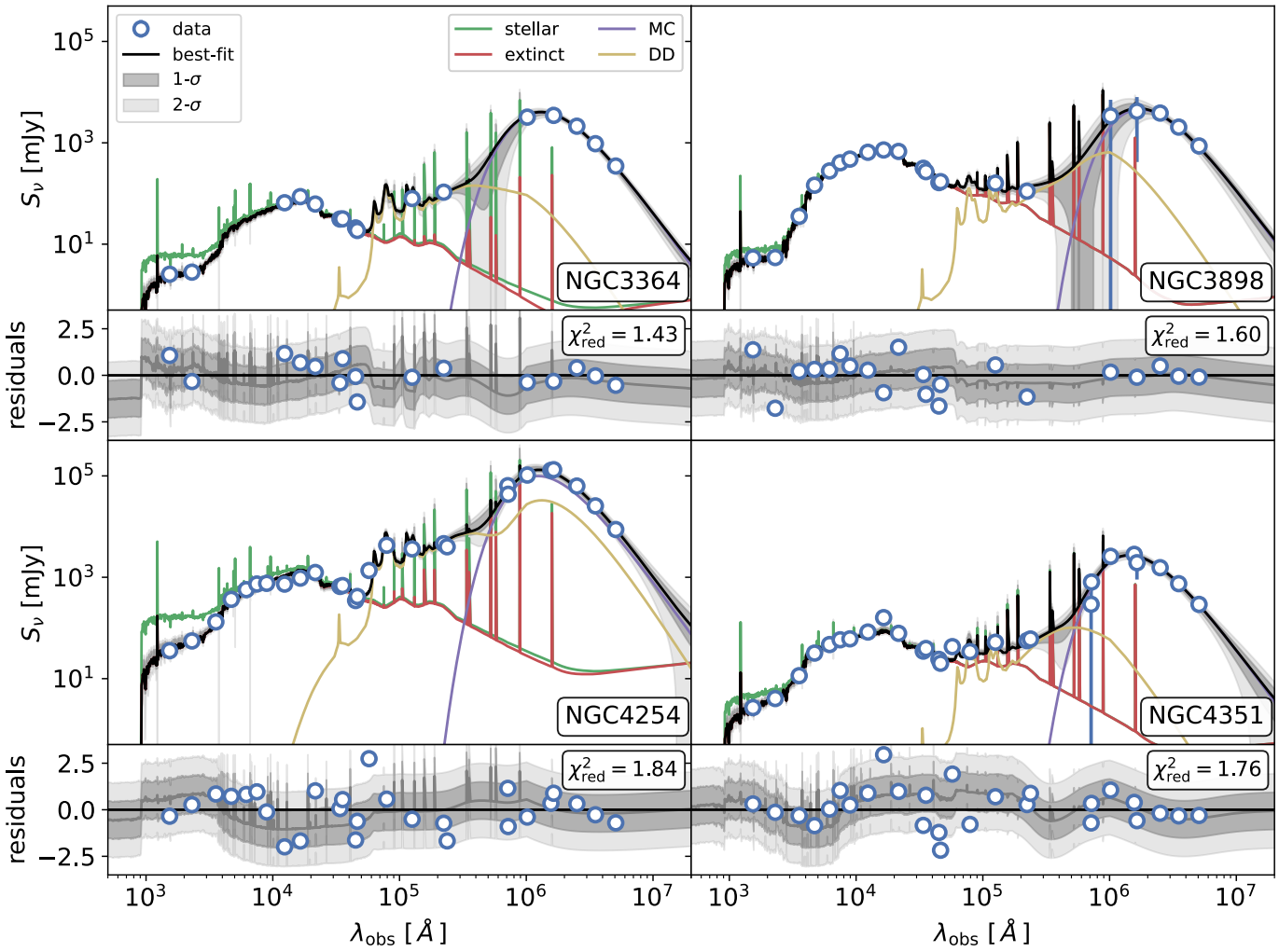

Fig. 23. Same as Fig. 21 for the Dustpedia sample of Sect. 4.2.2.

depends on the SFH and galaxy age. This last observation motivates us to advocate that GalaPy has found a different solution for the most probable properties of the objects.

A special mention must be made for NGC4254, where we measure the largest discrepancy with literature. In particular, while dust temperatures are still in agreement with the literature result, we measure values consistently higher by a factor $\approx 5 \div 10$ for the other examined quantities. In Hunt et al. (2019) the authors analyse a sample of objects, including NGC4254, with different SED fitting codes (i.e. MAGPHYS, CIGALE and GRASIL). With a photometric system similar to the one used here, the authors find values slightly larger to the ones reported here in the Literature columns of Tables 3 and 4. Nonetheless, these are still in disagreement with our estimates for the same parameters. We still checked that the SFR $\approx 20\ M_\odot$/yr we find is still allowed by the upper limits imposed with empirical relations (Lapi et al. 2011) connecting the object's flux in different bands with the SFR, with which we find $\mathrm{SFR}_{\mathrm{NGC4254}} \lesssim 50\ M_\odot\ \mathrm{yr}^{-1}$.

Finally, we also tried to run the analysis assuming a constant SFH model. The main differences worth to report are higher values for both the SFR and age, as a result of a SFH that, by construction, is more diluted in time. Nevertheless, the constant model is statistically disfavoured with respect to the In situ model as, with a larger number of parameters, produces values of the likelihood that are consistently lower for each of the four sources.

4.2.3. Lensed NIR-dark galaxy with upper limits from Giulietti et al. (2023)

As a further test-bench for GalaPy, we run the photometric analysis on a lensed, NIR-dark galaxy studied in Giulietti et al. (2023). HATLASJ113526.2-01460 (J1135 hereafter) was selected by Negrello et al. (2016) as a candidate lensed galaxy at redshift $z \sim 3.1$ (Harris et al. 2012) in the 12th Gama field of the *Herschel*-ATLAS survey, and then confirmed as a lensed NIR-dark galaxy by Giulietti et al. (2023).

The interest in testing GalaPy on this object resides on the large number of photometric points that are flagged as upper limits. As already mentioned in Sect. 3, the method of choice for treating upper limits in GalaPy is to consider them as regular points entering the same χ^2 likelihood used for detections. We assigned, to each of these points marked as non-detections, a flux equal to zero and an error equal to the noise value measured in the broadband photometry. We selected the same hyperparameters chosen for sources in Sect. 4.2.1: `parsec22.ntl` SSPs and the In situ SFH model, with ten free parameters.

We show the fitting results in Fig. 24, upper limits are marked as circles with arrows, while detections are round markers with error bars. In the lower panel, we show the standardised residuals and the χ^2 associated with the best-fitting model. As we associated a flux value of zero to non-detections, we mark their

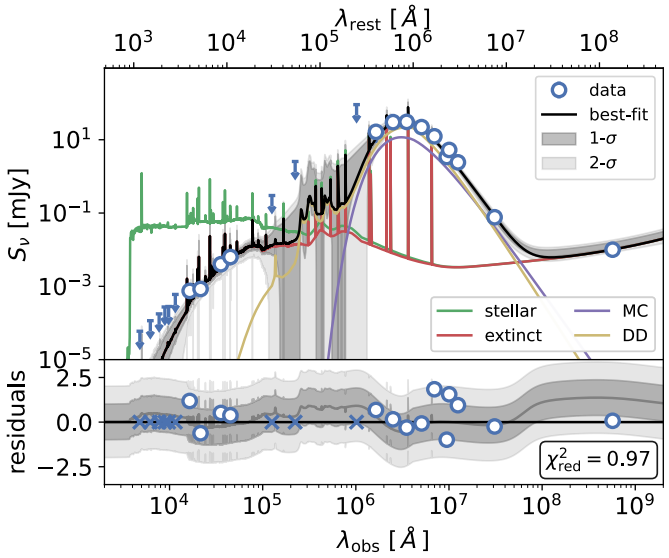

Fig. 24. Same as Fig. 21 for the J1135 lensed galaxy of Sect. 4.2.3.

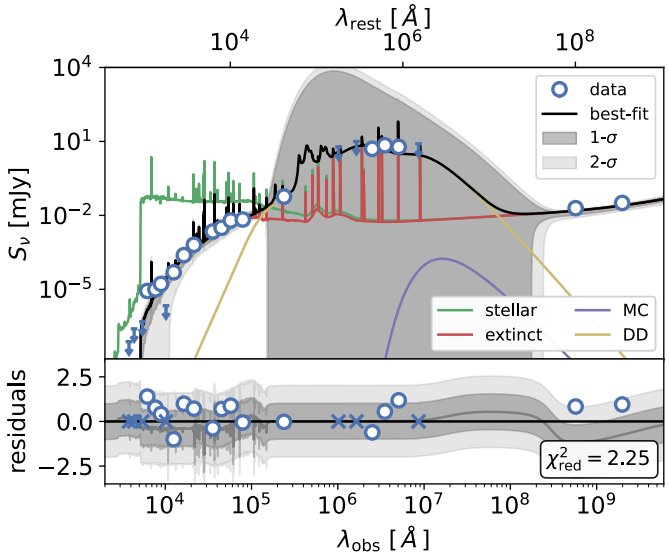

Fig. 25. Same as Fig. 21 for the median of Sect. 4.2.4.

corresponding positions with downward arrows at the 1-σ value measured for noise in the upper panel and as crosses in the lower panel.

Besides the solid value of the reduced χ^2 statistics, we can notice from the grey 2-σ confidence contour in the upper panel how the NIR-MIR region of the spectrum is just the upper limits, as a result of having just observed upper limits in that part. This is of course expected, as one of the free parameters of the model (i.e. the fraction of diffuse dust emission that is in PAH, $f_{\rm PAH}$) is completely and only determined by measurements in the NIR to MIR.

We conclude this section by highlighting the excellent agreement (Tables 3 and 4) between the parameter values derived with GalaPy, with respect to the values obtained in Giulietti et al. (2023), where the analysis has been performed with different, but compatible, methods (i.e. by analysing the line emission properties of the source in the ALMA bands). In particular, all the quantities do agree within the 1-σ credible interval to the measurements obtained in Giulietti et al. (2023), even though we find both a median and a best-fitting value for the SFR smaller than what found originally by the authors. We manage to also obtain a more precise measure the age of the object and for the stellar mass, consistent with the upper limit imposed in the original work (i.e. $\log M_\star^{\rm GalaPy}/M_\odot = 11.2^{+0.4}_{-0.6} < 11.7 = \log M_\star^{\rm Giulietti}/M_\odot$).

Even though not shown in the present manuscript, we have tested the reliability of our treatment of upper limits by both modifying the values used to mark non-detections and by using the other methods presented in Sect. 3.1 to include them in the likelihood. Concerning the former test, we have assumed both fluxes equal to the noise measurement and equal to three times this measurement (what is usually referred to as 1- and 3-σ upper limits). We observe that the result tends to be biased towards higher values for the predicted fluxes in the regions where we only have non-detections (i.e. optical and NIR/MIR). On the other hand, both running with the Sawicki method and with the naive method for the treatment of upper-limits guarantee convergence of the results but, besides requiring more samples (and therefore more time) to converge, the result tends to be statistically less solid, with values of the χ^2 statistics consistently higher. It is worth to mention that these considerations are not valid as a general test of the different possible approaches and serve solely has motivation for our final choice. Users of the library should tune their choices on the specific problem at hand.

4.2.4. Stacked NIR-dark radio selected galaxy with no spectroscopic redshift from Behiri et al. (2023)

We validate the library on a case that, to some extent, represents a more extreme case of dust obscured photometry with non-detections. In Behiri et al. (2023), the authors study a radio-selected sample from the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017) catalogue, based on a survey covering 2.6 deg^2 in the COSMOS field. Thanks to the extremely small value of its limiting flux density (12.6 μJy beam^{-1} at 5.5 σ), the survey has delivered one of the deepest samples ever obtained. Therefore, this dataset proves an ideal laboratory for estimating the contribution of galaxies at $z > 3$ to the cosmic SFRD.

In their work, the authors produce an ensemble analysis of the average sample properties, by performing SED fitting on the median photometric properties of the sample. The median SED (shown as round blue markers and downward arrows in Fig. 25), is obtained by stacking the individual sources maps for all the bands where a source is detected, and applying survival analysis (Isobe & Feigelson 1986) on all the bands where the presence of eventual upper limits has to be taken into account. This procedure results in 16 bands flagged as detections (empty blue markers with error bar) and 7 bands flagged as upper limits (downward arrows), on an overall wavelength range spanning from $\sim 5 \times 10^3$ Å to 2×10^9 Å. The redshift of this artificial median source is unknown, but is expected to be representative of the median redshift of the sample[15]. The resulting photometry is then fitted using the MAGPHYS+PHOTO-Z code (Da Cunha et al. 2008; Battisti et al. 2019), obtaining the results reported on the right half of Tables 3 and 4.

In running our fit with GalaPy, we select the usual set-up with In situ SFH and PARSEC22 SSPs including nebular lines, free-free and synchrotron and we let 12 physical parameters vary, including redshift, and an additional systematic error parameter

[15] The same argument also applies to the other physical properties.

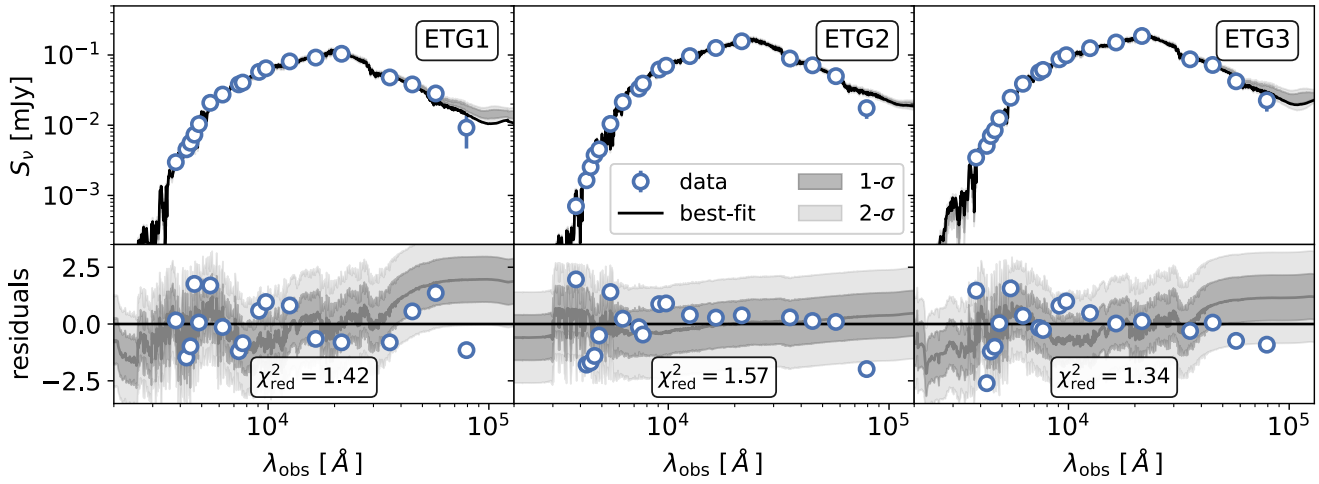

Fig. 26. Same as Fig. 21 for the quiescent sample of Sect. 4.2.5.

to account for eventual errors that might arise from the stacking procedure. The best-fitting SED with 1- and 2-σ confidence regions around the mean of the samples are pictured in Fig. 25 with a black solid line and shaded grey regions, respectively. The best-fit SED is decomposed into its different contributions, reported as coloured solid lines as listed in the legend. The reduced value of the $\chi^2_{\rm red} = 2.25$ is reported, along with the standardised residuals with respect to the best-fit model, in the lower panel of Fig. 25.

We observe a large uncertainty in the thermal dust emission peak, which is a symptom of the large number of upper-limits in the corresponding part of the observed spectrum. This level of uncertainty is reflected on the large uncertainty for the estimated temperatures of the two media, reported in Table 3 which, compared to the value obtained with MAGPHYS, have fairly small values. The largest difference between our results with those presented in the original work are though in the parameters mostly depending on the CSP, as this part is the most well sampled by the dataset at hand. In particular, we measure an age that is less than two times smaller than the one obtained with MAGPHYS, even though the two estimates are in agreement within the 68% credible interval. The combined effect of this small value of the age and our weighted median photometric redshift estimate of $z_{\rm GalaPy} = 4.41^{+0.34}_{-0.423}$, determine the extremely large value obtained for the SFR (i.e. $> 1400\ M_\odot\,{\rm yr}^{-1}$, see Table 3). Given the combination of age and redshift, this value has to be large to explain the 10 mJy flux measurement at the IR peak. As a term of comparison, the photometric redshift estimate obtained with MAGPHYS is $z_{\rm MAGPHYS} = 3.25^{+0.09}_{-0.11}$. Both values are consistent with the photometric redshift distribution of the sample from which the median fluxes have been obtained. Note though that, even though the median value of this redshift distribution would be more consistent with the MAGPHYS estimate on the median fluxes, also the photometric redshift estimate of the individual sources in the original work have been obtained using MAGPHYS. It would be interesting to apply our library on the whole sample but this, obviously, goes beyond the scope of this work.

Concerning the component masses instead, we are consistent with the values obtained by the authors of the original work. Both the total mass in dust and the total stellar mass are consistent within the 68% credible interval, even though the best-fitting values are different by some factors.

In closing this section, we stress that even though obtained from stacking the fluxes of different sources, this semi-mock observation should embed the characteristics of the NIR-dark radio-selected population of galaxies at high redshift, from which it has been built. Nonetheless note that, as already mentioned, a thorough comparison of the properties of the median fluxes with the median properties obtained from studying singular objects would require obtaining such individual properties with the same SED-fitting tool, which goes way beyond the objectives of this validation section.

4.2.5. Quiescent galaxies from Donevski et al. (2023)

To provide a consistency check that the in situ model delivered with GalaPy correctly describes the SFH of ETG progenitors and their evolution towards becoming quiescent, we ultimately validated our SED fitting tool on three quiescent galaxies extracted from the parent sample of spectroscopically selected massive quiescent galaxies in the COSMOS field, presented in Donevski et al. (2023). Here we fit the deep optical-to-NIR fluxes of representative sources at different redshifts, spanning in the range from $z = 0.1$ to $z = 0.6$. Testing on quiescent objects is a good test to check on the behaviour of the In situ model when the object is evolved, as this model has been derived to describe the progenitors of early type galaxies in all their evolution towards becoming quiescent.

Along with the in situ model of SFH, we once again select the PARSEC22 SSP libraries and allow for four free parameters: galaxy age, two free parameters of the SFH model, and an age corresponding to an eventual abrupt quenching, to simulate some violent feedback event cleaning up the galaxy from all its diffuse medium.

The best-fitting SEDs are shown in Fig. 26 as well as 1- and 2-σ confidence regions around the mean. In the lower panels, we show the standardised residuals with respect to the best-fitting model and associated reduced χ^2.

By inspecting the residuals in Fig. 26 we can see that the estimated best-fit model is correctly representing the dataset with $\chi^2_{\rm red}$ values between 1.3 and 1.6, although the extremely small errors on the optical flux measurements tend to make the nuisance systematic error parameter converge around values $f_{\rm sys} \lesssim 0.1$. Also the results for the derived parameters are extremely consistent with those found in the literature both in