

Full spectrum fitting with photometry in PPF: stellar population versus dynamical masses, non-parametric star formation history and metallicity for 3200 LEGA-C galaxies at redshift $z \approx 0.8$

Michele Cappellari  

Sub-Department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Accepted 2023 August 17. Received 2023 July 30; in original form 2022 September 1

ABSTRACT

I introduce some improvements to the PPF method, which measures the stellar and gas kinematics, star formation history (SFH) and chemical composition of galaxies. I describe the new optimization algorithm that PPF uses and the changes I made to fit both spectra and photometry simultaneously. I apply the updated PPF method to a sample of 3200 galaxies at redshift $0.6 < z < 1$ (median $z = 0.76$, stellar mass $M_* \gtrsim 3 \times 10^{10} M_\odot$), using spectroscopy from the LEGA-C survey (DR3) and 28-bands photometry from two different sources. I compare the masses from new JAM dynamical models with the PPF stellar population M_* and show the latter are more reliable than previous estimates. I use three different stellar population synthesis (SPS) models in PPF and both photometric sources. I confirm the main trend of the galaxies' global ages and metallicity $[M/H]$ with stellar velocity dispersion σ_* (or central density), but I also find that $[M/H]$ depends on age at fixed σ_* . The SFHs reveal a sharp transition from star formation to quenching for galaxies with $\lg(\sigma_*/\text{km s}^{-1}) \gtrsim 2.3$ ($\sigma_* \gtrsim 200 \text{ km s}^{-1}$), or average mass density within 1 kpc $\lg(\Sigma_1^{\text{JAM}}/M_\odot \text{ kpc}^{-2}) \gtrsim 9.9$ ($\Sigma_1^{\text{JAM}} \gtrsim 7.9 \times 10^9 M_\odot \text{ kpc}^{-2}$), or with $[M/H] \gtrsim -0.1$, or with Sersic index $\lg n_{\text{Ser}} \gtrsim 0.5$ ($n_{\text{Ser}} \gtrsim 3.2$). However, the transition is smoother as a function of M_* . These results are consistent for two SPS models and both photometric sources, but they differ significantly from the third SPS model, which demonstrates the importance of comparing model assumptions.

Key words: techniques: photometric – techniques: spectroscopic – software: data analysis – galaxies: evolution – galaxies: formation – galaxies: high-redshift.

1 INTRODUCTION

The study of the stellar population of galaxies is an essential tool when trying to uncover how they have assembled. For this reason, a vast number of papers have tried to infer the galaxies' star formation history (SFH) and chemical composition from observations. The earliest results were based on simple galaxy colours as these were easier to obtain (see e.g. the lectures by Baade 1963). However, galaxy colours alone cannot strongly constrain both the galaxies' chemical composition and SFHs. Inferences from galaxy photometry alone are strongly affected for example by the age-metallicity (e.g. Worthey 1994) as well as SFH-dust degeneracies (e.g. Silva et al. 1998; Devriendt, Guiderdoni & Sadat 1999; Pozzetti & Mannucci 2000). For this reason, most of our knowledge on both the star formation and chemical composition of galaxies has been obtained from the numerous absorption features in their spectra.

1.1 Full spectrum fitting of nearby galaxies

Over the past two decades, libraries of high-resolution ($R \gtrsim 2000$) empirical stellar spectra were observed, which try to optimally sample all stages of stellar evolution. Prominent examples in the optical region include the STELIB (Le Borgne et al. 2003), ELODIE

(Prugniel & Soubiran 2001), MILES (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011), and recently, the MaStar (Yan et al. 2019) stellar libraries. An exception is the X-Shooter spectral library (XSL), which reaches $R \approx 10\,000$ and extends up to $2.5 \mu\text{m}$ (Chen et al. 2014; Verro et al. 2022a).

Stellar population synthesis (SPS) models based on empirical stellar spectra have been developed that can produce synthetic galaxy spectra at high resolution. Examples of these kind of models are the GALAXEV (Bruzual & Charlot 2003), the VAZDEKIS (Vazdekis et al. 2010, 2015), the FSPS (Conroy, Gunn & White 2009; Conroy & Gunn 2010), the MARASTON (Maraston 2005; Maraston & Strömbäck 2011; Maraston et al. 2020), and most recently, the XSL models (Verro et al. 2022b). Most current SPS models complement empirical stellar spectra with fully synthetic ones like BaSeL (Westera et al. 2002) or MARCS (Gustafsson et al. 2008). This allows one to cover stages of stellar evolution not well sampled by observations and also can extend the wavelength coverage, especially to the ultraviolet and infrared regions that are poorly covered by observations. Fully synthetic SPS were also developed like BPASS (Stanway & Eldridge 2018; Byrne et al. 2022) or the MARCS version of the Maraston models.

Initially, large spectroscopic studies of the stellar population in nearby galaxies were based on line indices of specific absorption features, generally using the LICK system (e.g. Worthey et al. 1994), but the availability of good-quality high-resolution SPS models (see review by Conroy 2013) motivated a shift to using

* E-mail: michele.cappellari@physics.ox.ac.uk

full-spectrum fitting (see review by Walcher et al. 2011). Various templates-fitting methods were developed for this task, like PPXF (Cappellari & Emsellem 2004; Cappellari 2017), STARLIGHT (Cid Fernandes et al. 2005), STECMAP (Ocvirk et al. 2006), VESPA (Tojeiro et al. 2007), FIT3D (Sánchez et al. 2016; Lacerda et al. 2022) and FIREFLY (Wilkinson et al. 2017). These methods and software were extensively used, e.g. to analyse the millions of spectra produced by integral-field spectroscopic surveys in the local Universe like ATLAS^{3D} (Cappellari et al. 2011), CALIFA (Sánchez et al. 2012), SAMI (Bryant et al. 2015) and MaNGA (Bundy et al. 2015).

1.2 Spectral fitting of high- z galaxies

At significant redshift (e.g. $z \gtrsim 1$), the cosmological surface brightness dimming (e.g. Hogg 1999) makes good-quality spectra more difficult to obtain as the contribution of the sky background starts to dominate. PPXF was used to measure the kinematic and stellar population of significant samples of galaxies at redshift $z \approx 1$ (e.g. Shetty & Cappellari 2015; Bezanson et al. 2018) but only of individual objects out to $z \approx 2$ (e.g. van de Sande et al. 2013; Belli, Newman & Ellis 2014, 2017) and $z \approx 3$ (e.g. Esdaile et al. 2021; Forrest et al. 2022). Most large studies of distant galaxies still had to rely on photometry alone.

The most essential parameters one wants to extract from high- z galaxies are their redshift and stellar mass (e.g. Muzzin et al. 2013; Weaver et al. 2022). Various template-fitting codes were developed to measure masses and redshift from photometric observations in multiple bands (I ignore here methods based on machine learning; see Salvato, Ilbert & Hoyle 2019 for a review). These include HYPERZ (Bolzonella, Miralles & Pelló 2000), BPZ (Benítez 2000), LEPHARE (Arnouts et al. 2002), ZEBRA (Feldmann et al. 2006), and EAZY (Brammer, van Dokkum & Coppi 2008). These methods are conceptually similar to the template-based spectral fitting ones used for nearby galaxies; however, they all adopt a Bayesian approach, instead of a least-squares fitting one. This makes the codes simpler and allows for easy inclusion of priors on galaxy parameters or non-Gaussian uncertainties; for example, one can assign a low probability to solutions where the galaxy has an unphysically large/small stellar mass.

Building on the photometric-redshift and full-spectrum fitting approaches, new software was later developed to fit spectra together with the photometry, while still retaining the same Bayesian approach of photometric-redshift codes. Examples of these are FAST (Kriek et al. 2009), BEAGLE (Chevallard & Charlot 2016), BAGPIPES (Carnall et al. 2018), the code described by Mendel et al. (2020) and PROSPECTOR (Johnson et al. 2021b).

Contrary to what is sometimes stated, *both* least-squares, or maximum-likelihood, and Bayesian methods can return model posteriors, when needed. The former uses bootstrapping (e.g. Efron & Tibshirani 1994) or Monte Carlo approaches. In fact, bootstrapping can be seen as an efficient way to compute the Bayesian posterior with non-informative priors (e.g. Rubin 1981; Efron 2011). Although bootstrapping is less flexible than general Bayesian methods, in many realistic situations, the uncertainties of model parameters are dominated by data systematic and model assumptions (as I also find later) rather than the details of the adopted statistical approach or by adopted priors.

1.3 This paper

In this paper, I proceed differently than most existing methods. Instead of adopting the standard Bayesian approach to fit photometry

and spectra, I present an extension of my PPXF least-squares full-spectrum fitting method to simultaneously fit photometry. A key difference in this approach is that it can be a few orders of magnitude faster than Bayesian methods. Apart from algorithmic differences, my PPXF approach to fitting photometry with spectra is analogue to the extension of the STARLIGHT least-squares full-spectrum fitting method (López Fernández et al. 2016; Werle et al. 2019). Least-squares methods appear complementary to existing ones as the extra speed allows for extra flexibility in the treatment of the stellar population, as shown later.

I illustrate the characteristics of the approach by fitting the VIMOS spectra and COSMOS photometry (Muzzin et al. 2013; Weaver et al. 2022) to study the joint SFH-metallicity distributions and the stellar population scaling relations of about 3200 galaxies from the LEGA-C survey (van der Wel et al. 2021) in the redshift range $0.6 < z < 1$.

Readers interested in the PPXF techniques should keep reading how to measure velocities in Section 2 and the PPXF updates in Section 3. While those only interested in the scientific results should skip the next two sections and go directly to the description of the data in Section 4 and results in Section 7. In this work, I adopt a standard cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

2 MEASURING VELOCITY AND REDSHIFT

In Cappellari (2017, section 2), I reviewed general and important facts that one should know before using any full spectral fitting method and PPXF in particular. Here I include only some updates, and I heavily refer the reader to my previous paper of this series to avoid duplicating material.

2.1 From measured velocity to observed redshift

The physical meaning of the velocity V returned by PPXF or any spectrum-fitting code is often a source of confusion. As discussed in Cappellari (2017, section 2.3), the reason for this is that V has no physical meaning. Even the recession velocity itself, for a distant galaxy, is an ill-defined concept with a debated interpretation (e.g. Bunn & Hogg 2009). It should never be used for quantitative work. What is well-defined empirically is the redshift z of a given spectrum:

$$1 + z \equiv \frac{\lambda_{\text{obsv}}}{\lambda_{\text{emit}}}, \quad (1)$$

where λ_{obsv} and λ_{emit} are the observed and rest-frame wavelength of a given spectral feature. The key formula that is needed to convert the V_{PPXF} returned by PPXF into redshift is (Cappellari 2017, eq. 8)

$$V_{\text{PPXF}} \equiv c \ln(1 + z), \quad (2)$$

with c the speed of light. This formula is exact by construction and it is the only one to use to attach a physical meaning to V_{PPXF} .

2.2 Separating peculiar velocities and cosmological redshift

When observing spectra of distant galaxies from a single aperture, redshift is all one can measure. However, when obtaining spatially resolved observations of galaxies, e.g. using integral-field spectroscopy (see review by Cappellari 2016), one needs to separate the cosmological redshift z_{cosm} , which only contains information on the galaxy distance, from the peculiar velocity V_{pec} . The latter is the one that satisfies, e.g. Newton's law of gravitation in a reference system that moves with the galaxy barycentre. It is the velocity that has to be used to construct dynamical models of the galaxy.

In Cappellari (2017, section 2.4), I suggested using the standard way of separating peculiar and cosmological redshift. However, there is a simpler and formally even more accurate way. In fact, the conversion of V_{PPXF} into redshift is unnecessary (see Baldry 2018). One can directly obtain V_{pec} using the velocities returned by PPXF as follows

$$V_{\text{pec}}(x, y) = V_{\text{PPXF}}(x, y) - V_{\text{PPXF}}(\text{bary}). \quad (3)$$

Here $V_{\text{PPXF}}(x, y)$ is the velocity returned by PPXF at the location (x, y) on the sky, $V_{\text{PPXF}}(\text{bary})$ is the velocity returned by PPXF for the galaxy (or cluster) barycentre, and $V_{\text{pec}}(x, y)$ is the peculiar velocity at location (x, y) . The latter is the only one with a clear physical meaning: it is the one to use in a dynamical model (e.g. Cappellari 2008) or to estimate the level of rotation in a galaxy (e.g. Emsellem et al. 2011). Importantly, equation (3) is always valid, regardless of whether the spectrum was de-redshifted to the rest-frame or not, before measuring V_{PPXF} . Note that this formula only works because of the way PPXF defines the relation between velocity and redshift in equation (2) and *cannot* be used with alternative definitions (e.g. $V \equiv cz$).

As an example of a practical application of these formulas, let's assume I am fitting a single spectrum of a high- z galaxy for which I have an estimate of the redshift z' (e.g. from photometry). It is generally convenient to de-redshift the spectrum by dividing each observed wavelength λ_{obs} to obtain an estimate of the rest-frame wavelength with

$$\lambda'_{\text{rest}} = \frac{\lambda_{\text{obs}}}{(1 + z')}. \quad (4)$$

I then fit the spectrum with PPXF to obtain V_{PPXF} . If the initial guess z' was perfect, I would obtain $V_{\text{PPXF}} = 0$, but in general, I will measure $V_{\text{PPXF}} \neq 0$, with uncertainty ΔV_{PPXF} . An improved estimate of the galaxy redshift z and its uncertainty Δz can be obtained using equations (2) and (3) as

$$V' = c \ln(1 + z'), \quad (5a)$$

$$V_{\text{tot}} = V_{\text{PPXF}} + V' \quad (5b)$$

$$1 + z = \exp\left(\frac{V_{\text{tot}}}{c}\right) = (1 + z') \exp\left(\frac{V_{\text{PPXF}}}{c}\right) \quad (5c)$$

$$\frac{\Delta z}{1 + z} \approx \Delta \ln(1 + z) \approx \frac{\Delta V_{\text{PPXF}}}{c}. \quad (5d)$$

Identical results are obtained without first bringing the spectrum to the rest frame and setting $z' = 0$ in equation (5). But one should remember to adjust the instrumental resolution as described in section 2.4 of Cappellari (2017).

3 UPDATES TO THE PPXF PACKAGE

I gave a detailed overview of the PPXF method in Cappellari (2017, section 3). I will not repeat that summary here, but instead, I will refer the reader to specific sections of that paper, while trying to keep a consistent notation. However, I substantially evolved PPXF since then, driven by the needs of my research and requests from colleagues. Here, I only describe the PPXF features that have changed. The description corresponds to the current version 8.2 of the public PPXF Python package.¹

¹ Available from <https://pypi.org/project/ppxf/>

3.1 Well-sampled variable- σ convolution

As discussed in Cappellari (2017, section 2.2), when fitting stellar templates to galaxy spectra, one generally needs to first match their resolution to that of the galaxy observations by convolving the templates with a Gaussian with dispersion σ that varies with wavelength λ . In a previous version of PPXF, I implemented this step as a direct summation of the spectrum, weighted by the Gaussian centred on every pixel (function `ppxf_util.gaussian_filter1d` in PPXF). Given that the Gaussian is typically non-zero only over $n_{\text{gau}} \sim 10$ pixels while the templates have $n_{\text{pix}} \gtrsim 1000$ spectral pixels, I limited the summation only to the non-zero pixels of the Gaussian kernel. In this way, the computation time of this summation scales as $t \propto n_{\text{pix}} \times n_{\text{gau}}$, which is comparable to that $t \propto n_{\text{pix}} \times \ln(n_{\text{pix}})$ achievable using Fourier convolution using the Fast Fourier Transform (FFT; Cooley & Tukey 1965).

A limitation of performing a convolution as a summation is that, when the σ of the Gaussian becomes comparable to the size of the sampled pixels, the convolution suffers from the same undersampling problems that motivated the use of an analytic Fourier Transform (FT) for the convolution in PPXF discussed in Cappellari (2017). For this reason, I now implemented an alternative procedure `ppxf_util.varsmooth`, which performs the variable- σ convolution using the FFT and uses the same approach of PPXF of using an analytic FT of the kernel, to avoid undersampling issues.

A natural idea to use the FFT for the convolution of a vector of values with a kernel with variable scale is to stretch the coordinate with the inverse of the scale, via interpolation, in such a way that the kernel has the same scale in the new coordinates. This idea was discussed, e.g. in 2014 on StackOverflow² and implemented in 2016 by Janez Kos on GitHub,³ in the procedure `varconvolve`. It was also discussed in Johnson et al. (2021b). The interpolation approach is also central for algorithms for the non-uniform FFT (NUFFT; e.g. Greengard & Lee 2004). Here I combine the interpolation approach with the use of an analytic FT as in Cappellari (2017) to produce Algorithm 1. The algorithm can also be used with a non-Gaussian kernel (e.g. Cappellari 2017, eq. 38) as long as only its scale changes with wavelength.

Algorithm 1 `varsmooth`: well-sampled variable- σ convolution

Given $\mathbf{x}, \mathbf{y}, \sigma_x, \mathbf{x}_{\text{out}}$ ▷ $\text{len}(\mathbf{x}) = \text{len}(\mathbf{y}) = \text{len}(\sigma_x) = p$
 $\sigma = \sigma_x / \nabla \mathbf{x}$ ▷ Convert σ_x into pixels⁴
 $\sigma_{\text{max}} = \max(\sigma) \times m$ ▷ Optional oversampling $m \geq 1$
 $s = \text{cumsum}(\sigma_{\text{max}} / \sigma)$ ▷ Cumulative sum
 $n = \text{ceil}(s_p - s_1)$ ▷ Ensure oversampling
 $\mathbf{x}_{\text{new}} = \text{subdivide}(s_1, s_p, n)$ ▷ n equispaced values
 $\mathbf{y}_{\text{new}} = \text{interpolate}(\mathbf{s}, \mathbf{y})(\mathbf{x}_{\text{new}})$
 $\mathcal{F}(\text{gauss}) = \exp(-\sigma_{\text{max}}^2 \omega^2 / 2) / \sqrt{2\pi}$ ▷ Gaussian Analytic FT
 $\mathbf{y}_{\text{conv}} = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{y}_{\text{new}}) \mathcal{F}(\text{gauss})]$ ▷ Convolution theorem & FFT
if \mathbf{x}_{out} is given **then**
 $\mathbf{s} = \text{interpolate}(\mathbf{x}, \mathbf{s})(\mathbf{x}_{\text{out}})$
 $\mathbf{y}_{\text{conv}} = \text{interpolate}(\mathbf{x}_{\text{new}}, \mathbf{y}_{\text{conv}})(\mathbf{s})$

This algorithm works well in practice, but one should be aware of its theoretical limitations. In fact, most interpolation methods can

² <https://stackoverflow.com/a/24186800>

³ <https://github.com/sheliak/varconvolve>

⁴ I use a centred finite-difference approximation of the gradient $\nabla \mathbf{x}$ instead of the spectral pixels size $\Delta \mathbf{x}$, which is generally not provided. The two quantities coincide in uniformly-sampled regions of the spectrum.

be described as a convolution with a specific kernel (e.g. Getreuer 2011), and one may think it would be better to remove the effect of this extra convolution as done, e.g. in the NUFFT methods. However, this situation is different as the spectra have noise, and one would have to perform the interpolation in a Bayesian framework (e.g. MacKay 1992, 2003). One should also consider that the spectra to fit generally already include additional interpolation and resampling, which would have to be modelled for rigorous results. But all this is unlikely to affect scientific results, and for this reason, it is beyond the scope of this paper.

3.2 CAPFIT non-linear least-squares with linear constraints

3.2.1 The problem

When fitting the kinematics of multiple kinematics components with PPXF, both for the stellar and gas emission components, it is often useful to be able to set constraints on some parameters as a function of other parameters. For example, when looking for spectra containing both narrow and broad gas emission lines to study active galactic nuclei (AGN; e.g. Oh et al. 2015; Fu et al. 2023), to avoid the degeneracy of fitting two similar lines, one may want to constrain the dispersion of the broad emission component, if present, to be significantly larger than that of the narrow one $\sigma_{\text{broad}} > \sigma_{\text{narrow}} + \Delta\sigma$, or as a fractional difference $\sigma_{\text{broad}} > f \times \sigma_{\text{narrow}}$. Or one may want to constrain the velocity of a possible broad emission component to differ less than a certain value from that of the narrow one $|V_{\text{broad}} - V_{\text{narrow}}| < \Delta V$. Efficiently setting this kind of constraint requires solving a constrained non-linear optimization problem.

For maximum computational efficiency and accuracy, one should exploit the special problem that PPXF has to solve (section 3.4 of Cappellari 2017). In particular, the function to minimize $f(\mathbf{x})$ is a sum of squares and the typical constraints are linear. The problem to solve can be expressed as

$$\begin{aligned} \text{minimize } & f(\mathbf{x}) = \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|^2 \\ \text{subject to } & \mathbf{A}_{\text{eq}} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}} \\ & \mathbf{A}_{\text{ineq}} \cdot \mathbf{x} \leq \mathbf{b}_{\text{ineq}}, \end{aligned} \quad (6)$$

where $\mathbf{r}(\mathbf{x})$ are the residual from the fit, and \mathbf{x} are the non-linear parameters, like the kinematics of different components. I have searched extensively for specialized software or an algorithm that I could easily use in PPXF to efficiently solve this specific problem but did not find any. For this reason, I developed my own.

One of the most effective ways of solving nonlinear problems with general constraints is the sequential quadratic programming (SQP) method, where at every iteration the algorithm solves a constrained quadratic problem that approximates the function at the current location (e.g. Nocedal & Wright 2006, chapter 18).

In least-squares problems, one can approximate the function near the current point \mathbf{x}_k as a second-order Taylor series, with $\mathbf{p} = \mathbf{x} - \mathbf{x}_k$, as follows (e.g. Nocedal & Wright 2006, section 10.2)

$$f(\mathbf{x}) \approx \frac{1}{2} \|\mathbf{J}_k \cdot \mathbf{p} + \mathbf{r}_k\|^2 \quad (7a)$$

$$= \frac{1}{2} \|\mathbf{r}_k\|^2 + \mathbf{p} \cdot (\mathbf{J}_k^T \cdot \mathbf{r}_k) + \frac{1}{2} \mathbf{p} \cdot (\mathbf{J}_k^T \cdot \mathbf{J}_k) \cdot \mathbf{p} \quad (7b)$$

$$\approx f(\mathbf{x}_k) + \mathbf{p} \cdot \nabla f(\mathbf{x}_k) + \frac{1}{2} \mathbf{p} \cdot \nabla^2 f(\mathbf{x}_k) \cdot \mathbf{p}, \quad (7c)$$

where \mathbf{J}_k is the Jacobian, which can be computed by finite differences, $\nabla f(\mathbf{x}_k) = \mathbf{J}_k^T \cdot \mathbf{r}_k$ is the gradient and $\nabla^2 f(\mathbf{x}_k) = \mathbf{J}_k^T \cdot \mathbf{J}_k$ is the *quasi-Newton* approximation of the Hessian matrix, whose full form is the following, but I ignored the second term (e.g. Nocedal & Wright

2006, eq. 10.5)

$$\nabla^2 f(\mathbf{x}) = \mathbf{J}^T \cdot \mathbf{J} + \sum_{j=1}^m r_j \nabla^2 r_j. \quad (8)$$

It is a characteristic of least-squares problems that one can approximate the Hessian ‘for free’ using \mathbf{J} and the reason it is important to adopt specialized methods for their solution. In the case of PPXF, the Hessian approximation is always especially good, even far from the solution, because the algorithm separates the linear and non-linear optimizations (sections 3.3–3.4 of Cappellari 2017) and ensures that $\sum_j r_j = 0$ at every step. This tends to cancel out the second term in the Hessian of equation (8).

In essence, a specialized SQP algorithm to solve equation (6) would consist of solving a sequence of quadratic sub-problems as follows

$$\begin{aligned} \text{minimize } & g(\mathbf{p}) = \|\mathbf{J}_k \cdot \mathbf{p} + \mathbf{r}_k\|^2 \\ \text{subject to } & \mathbf{A}_{\text{eq}} \cdot \mathbf{x} = \mathbf{b}_{\text{eq}} \\ & \mathbf{A}_{\text{ineq}} \cdot \mathbf{x} \leq \mathbf{b}_{\text{ineq}}. \end{aligned} \quad (9)$$

Algorithms for this type of problem are discussed e.g. by Fletcher (1987, section 11.2) or Gill, Murray & Wright (1981, chapter 5). I implemented those ideas into a trust-region algorithm (chapter 4 of Nocedal & Wright 2006), but I discovered that the approach is not sufficiently robust for my rather special situation.

The difficulty of the optimization problem I have to solve consists of the fact that the Jacobian can sometimes be completely degenerate. A common situation where this happens is when PPXF is fitting for the kinematics of emission lines or multiple stellar kinematic components. In this situation, the weights associated with a given emission line or stellar component may become exactly zero, because the line or component is simply not present in a certain galaxy spectrum. In other cases, the signal-to-noise ratio S/N may be too low to give any constraints to some parameters. In these cases, the gradient (column of \mathbf{J}) with respect to the parameters describing the kinematics of the missing component will be zero.

I tried using the Singular Value Decomposition (SVD; e.g. Press et al. 2007, section 15.4.2) during the iterations required to solve the quadratic sub-problem. But after extensive testing, e.g. during the development of the MaNGA Data Analysis Pipeline (Westfall et al. 2019), I was unable to find a robust criterion to decide which singular values must be edited and find the effective rank of my Jacobian.

3.2.2 The solution

To solve *unconstrained* least-squares optimization problems one of the most widely used techniques is the Levenberg–Marquardt (LM) method (Levenberg 1944; Marquardt 1963) and its state-of-the-art implementation in MINPACK (Moré 1978; Moré, Garbow & Hillstrome 1980). The success of the LM method comes from the fact that the method penalizes the \mathbf{J} matrix adaptively defining the quadratic sub-problem, in such a way that it always prevents degeneracy. This also makes the LM method a robust trust-region algorithm, as discussed in Fletcher (1987, section 5.2) or Nocedal & Wright (2006, section 10.3). Press et al. (2007, section 15.5.2) provides a less technical description.

In a previous version (<6.5) of PPXF, I used the LM algorithm, as modified in the MPFIT implementation (Markwardt 2009), which included a very useful but non-optimal treatment of box constraints (i.e. upper/lower limits on the parameters). Comparable box-constrained least-squares methods exist in SCIPY (Virtanen et al. 2020) as implemented in the trust-region reflective

algorithm (method='trf'; Branch, Coleman & Li 1999) and the dogleg algorithm (method='dogbox', Voglis & Lagaris 2004, Nocedal & Wright 2006, chapter 4) in `scipy.optimize.least_squares`. These methods are also available in PPXF but cannot support linear constraints.

After extensive experimentation with real-world cases, I implemented a novel hybrid between the SQP and LM methods, specialized for the non-linear least-squares with linear constraints (both equality and inequality). The algorithm consists of a trust-region quasi-Newton SQP method, with linear constraints, in which the matrix defining the quadratic sub-problem is penalized to avoid the risk of degeneracy, as in the LM method. I achieve this by replacing the quadratic sub-problem of equation (9) with the following (see Nocedal & Wright 2006, eq. 10.41)

$$\begin{aligned} \text{minimize } g(\mathbf{p}) &= \left\| \begin{pmatrix} \mathbf{J}_k \\ \sqrt{\lambda_k} \mathbf{D}_k \end{pmatrix} \cdot \mathbf{p} + \begin{pmatrix} \mathbf{r}_k \\ 0 \end{pmatrix} \right\|^2 \\ \text{subject to } \mathbf{A}_{\text{eq}} \cdot \mathbf{x} &= \mathbf{b}_{\text{eq}} \\ \mathbf{A}_{\text{ineq}} \cdot \mathbf{x} &\leq \mathbf{b}_{\text{ineq}}, \end{aligned} \quad (10)$$

where \mathbf{D}_k is a diagonal matrix, which makes the problem scale invariant. By default, the diagonal elements of \mathbf{D}_k are initialized with the norm $\| \cdot \|$ of the columns of \mathbf{J} and are updated during the iterations as suggested in Moré (1978, eq. 6.3). Close to the solution, when the quadratic model provides a good approximation of $f(\mathbf{x})$, then λ_k becomes small and equation (10) approximates equation (9). In this limit, the method behaves as an SQP method. When the quadratic approximation is inaccurate, λ_k becomes large and the method behaves as a trust-region LM method.

My resulting algorithm is rather simple because I did not worry about the efficiency of the solution of the quadratic programming sub-problem. The latter generally dominates the complexity of other state-of-the-art algorithms, which devise approximated matrix updates to save computation time (e.g. see the description of LM in Moré 1978). I also did not try to deal with large-scale problems and sparse matrices which also increase complexity and require specialized methods (e.g. Gill, Murray & Saunders 2005). Instead, I focused on the fitting of rather small nonlinear problems ($n \lesssim 50$ variables) in which computing the function $f(\mathbf{x})$ involves creating a complex model, as in PPXF. In this rather common situation, the time to solve the small quadratic programming sub-problem becomes negligible compared to that of evaluating $f(\mathbf{x})$. The solution is given in Algorithm 2, which I implemented in the `capfit` procedure in the PPXF package.

Except for the fact that the quadratic sub-problem is penalized and linearly constrained, the algorithm uses the standard trust-region framework (e.g. algorithm 5.2.7 of Fletcher 1987 or algorithm 4.1 of Nocedal & Wright 2006). For the convergence criteria, I follow the description in Moré et al. (1980, section 2.3).

3.2.3 Solving the quadratic sub-problem

I implemented two procedures to solve the quadratic programming sub-problem of equation (10). In both cases, I avoid explicitly constructing the Hessian $\mathbf{J}_k^T \cdot \mathbf{J}_k$ as this would degrade the conditioning of the system. The first procedure (`lsq_box`) is specialized for the common situation where only box constraints are present. It solves $\min \| \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \|^2$ with $\mathbf{lb} < \mathbf{x} < \mathbf{ub}$. For this, I use the active-set method adopted in the non-negative least-squares (NNLS) method (algorithm 23.10 of Lawson & Hanson 1995), which was generalized for box constraints with the Bounded-Variables Least-Squares (BVLS) procedure in the same book and in Stark & Parker (1995). My

Algorithm 2 CAPFIT: linearly-constrained nonlinear least-squares

```

Given  $\mathbf{x}_1, f, \lambda > 0, \eta \in [0, 1/4]$ 
 $\mathbf{r}_1 = f(\mathbf{x}_1)$ 
 $\mathbf{J} = \mathbf{J}(\mathbf{x}_1)$ 
Compute the scaling matrix  $\mathbf{D}$  ▷ See text
loop
  Obtain  $\mathbf{p}$  as solution of equation (10) with  $\mathbf{J}, \mathbf{D}, \mathbf{r}_1, \lambda$ 
   $\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{p}$ 
   $\mathbf{r}_2 = f(\mathbf{x}_2)$ 
   $r_{\text{act}} = \chi^2(\mathbf{r}_1) - \chi^2(\mathbf{r}_2)$  ▷ Using  $\chi^2(\mathbf{x}) \equiv \mathbf{x} \cdot \mathbf{x}$ 
   $r_{\text{pre}} = \chi^2(\mathbf{r}_1) - \chi^2(\mathbf{J} \cdot \mathbf{p} + \mathbf{r}_1)$ 
   $\rho = r_{\text{act}}/r_{\text{pre}}$  ▷ Actual vs predicted reduction
  if convergence test is satisfied then
    stop with solution  $\mathbf{x}_2$ 
  if  $\rho < 1/4$  then
     $\lambda = 4\lambda$ 
  else if  $\rho > 3/4$  then
     $\lambda = \lambda/2$ 
  if  $\rho > \eta$  then ▷ Successful step: move on
     $\mathbf{J} = \mathbf{J}(\mathbf{x}_2)$ 
    Adjust the scaling matrix  $\mathbf{D}$  ▷ See text
     $\mathbf{x}_1, \mathbf{r}_1 = \mathbf{x}_2, \mathbf{r}_2$ 

```

implementation closely follows Lawson & Hanson (1995) except for the important fact that (i) I allow for a starting guess and (ii) I include an initialization step (Algorithm 3), which generalizes to the box-constrained case the initialization loop in the FASTNNLS code⁵ by Andersson & Bro (2000). In realistic PPXF problems, using my new `lsq_box` with hundreds of spectral templates produced a typical speedup of a factor four compared to using the `scipy.optimize.nnls`, which is a wrapper to the Lawson & Hanson (1995) Fortran code. The procedure `scipy.optimize.lsq_linear` currently also does not support passing a starting guess.

Algorithm 3 Initialization `lsq_box` box-constrained least-squares

```

Given  $\mathbf{x}, \mathbf{A}, \mathbf{b}$ , lower  $\mathbf{lb}$  and upper  $\mathbf{ub}$  bounds,  $\mathcal{B} \neq \emptyset$ 
while  $\mathcal{B} \neq \emptyset$  do
   $\mathcal{B} = \{\mathbf{x} \mid (\mathbf{x} < \mathbf{lb}) \vee (\mathbf{ub} < \mathbf{x})\}$ 
  Set all  $\mathbf{x} \in \mathcal{B}$  to the nearest bound
   $\mathcal{F} = \{\mathbf{x} \mid \mathbf{lb} < \mathbf{x} < \mathbf{ub}\}$ 
   $\mathbf{x}' = \min \| \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \|^2$  for  $\mathbf{x} \in \mathcal{F}$ 
  Set  $\mathbf{x} = \mathbf{x}'$  for  $\mathbf{x} \in \mathcal{F}$ 

```

The second procedure (`lsq_lin`) solves general linearly constrained quadratic programming problems like in equation (9) using a modified version of the standard active-set technique described by Nocedal & Wright (2006, algorithm 16.3). The approach consists of solving a sequence of equality-constrained linear least-squares problems, for which I follow Golub & Van Loan (2013, algorithm 6.2.2), allowing for degenerate matrices using SVD. If the initial guess is unfeasible, I find a feasible point using the linear programming procedure `scipy.optimize.linprog` and the method='highs' by Huangfu & Hall (2017). Alternatively, the quadratic sub-problem can be solved within PPXF by minimizing the quadratic function in the form of equation (7b) using a general quadratic programming solver. In the current implementation, I use the interior-point solver `solvers.coneqp` from the CVXOPT

⁵Available from <https://ucphchemometrics.com/>

package⁶ by Andersen et al. (2011), which is much faster for large-scale problems. I have extensively tested the `lsq_box` and `lsq_lin` procedure described in this section by constructing batteries of tests using both exact analytic solutions and comparisons against `CVXOPT`.

The `capfit` procedure has been the default nonlinear optimization algorithm in PPXF for about 3 yr and has been used as a general optimizer independently of PPXF too. During this time, PPXF was used to fit millions of spectra from a variety of surveys (e.g. MaNGA, Bundy et al. 2015 and SAMI, Bryant et al. 2015). This allowed me to fix its handling of rare failures in degenerate situations, which are difficult to encounter in idealized examples. When applied to unconstrained non-linear least-squares problems `capfit` produces essentially the same iterates as the state-of-the-art LM implementations in `MINPACK` or `MPPFIT`, as expected. When used for box-constrained non-linear least-squares problems `capfit` is generally at least as efficient as the best algorithms in `scipy.optimize.least_squares`. However, `capfit` allows for the extra flexibility of using linear constraints, as well as for keeping variables tied to others or fixed.

3.3 Setting linear constraints on the template weights

In the previous section, I discussed the use of linear constraints on the kinematic parameters during the PPXF fit. Here, I note that linear constraints can also be used during the linear-fitting procedure in Cappellari (2017, section 3.3). These were used, for example, to constrain the sum of weights (e.g. the luminosity) of different template groups to constitute a certain fraction of the total light, e.g. to perform kinematic bulge/disks decompositions (e.g. Tabor et al. 2017, 2019; Oh et al. 2020) or to study stellar population of different kinematic components (e.g. Shetty et al. 2020b).

3.4 Multidimensional regularization

Considering without loss of generality that stellar population depends only on age, the fundamental equation used to model the spectrum of a composite stellar population is (e.g. Cid Fernandes et al. 2005; Ocvirk et al. 2006; Conroy 2013, section 2.3)

$$G_{\text{mod}}(\lambda) = \int_{t=0}^{t=T} \text{SSP}_{\lambda}(t, Z) \cdot \text{SFR}(T - t) dt, \quad (11)$$

where SFR is the star formation rate, SSP_{λ} is a Single Stellar Population spectrum per unit mass, with age t and metallicity Z , and T is the age of the Universe at the redshift of the galaxy. This expression is generalized in PPXF to study the distribution of more parameters, e.g. metallicity, α enhancement or initial mass function (IMF), in addition to the SFR.

I pointed out in Cappellari (2017, section 3.5) that equation (11) is an inhomogeneous Fredholm equation of the first kind, with kernel SSP_{λ} . And the recovery of the SFR(t) from the observed G_{mod} is a textbook example of ill-conditioned inverse problem (e.g. Hansen 1998; Kabanikhin 2011; Press et al. 2007, section 19.0). This means that one cannot find a unique solution from real data without further assumptions.

In Cappellari (2017, section 3.5), I discussed the implementation of linear regularization (e.g. Press et al. 2007, section 19.5) in PPXF to address this issue and study the stellar population in galaxies. I gave a formula for the second-order one-dimensional regularization. Given that there are alternative ways of generalizing a measure of smoothness of a function in dimension larger than one (e.g. Brady &

Horn 1983), I clarify here that the second-order regularization (`reg_ord=2`) in PPXF minimizes the total squared Laplacian $(\Delta w)^2$ of the weights w distribution, while the first-order one (`reg_ord=1`) minimizes the total squared gradient $(\nabla w)^2$. Both operators are implemented by standard finite differences.

Press et al. (2007, fig. 19.4.1) point out that, under some sensible conditions, the regularized solution has a simple Bayesian interpretation: it represents the most likely solution for the weights, given an adjustable prior on the amplitude of the fluctuations. However, the meaning of the fundamental degeneracy of the stellar population inversion, as well as of regularization, is best illustrated with an example.

I used a grid of 25 logarithmically spaced ages t and 6 metallicities $[M/H]$ from the SPS models by Vazdekis et al. (2015) to construct a synthetic spectrum in which the distribution of light contributed by each spectrum in the V band follows a bivariate Gaussian distribution $\mathcal{N}(t, [M/H])$ with mean age $t_0 = 1$ Gyr, mean $[M/H]_0 = -0.3$ and dispersion of 0.25 dex in both age and metallicity. I logarithmically sampled the spectrum at a velocity scale $\Delta V = c \Delta \ln \lambda = 50 \text{ km s}^{-1}$ per spectral pixel.

I show the resulting spectrum in the top panel of Fig. 1 and the input light-weights distribution in the second panel. The third panel shows a single PPXF fit, which is characterized by discrete sharp peaks as expected due to the ill-conditioning of the inversion problem. The fourth panel shows the result of averaging the weights obtained by fitting with PPXF 100 Monte Carlo realizations obtained by adding Gaussian noise on the same noiseless synthetic spectrum. Here, the average converges towards the true input distribution. Finally, in the bottom panel, I show the result of performing a single regularized PPXF fit (with `reg_order=2` and a typical `regul=30`). Here, the distribution looks comparable to that of the average of multiple realizations.

Regularization has its limitations, in fact, it is by construction a trade-off between agreement with the data and smoothness (e.g. Press et al. 2007, fig. 19.4.1), which may introduce biases. In general, when one is one is obtaining results by averaging many spectra, it may be better not to use regularization, or only use a minimal amount, to reduce possible biases, while allowing the differences in the noise between spectra to act as Monte Carlo realizations. But regularization is very useful when interpreting individual spectral fits and even to reduce noise in the SPS models themselves, which may introduce spurious features in the solutions (as I found later).

One can use bootstrapping of the residuals, while repeating the PPXF fits multiple times, to obtain averages as well as uncertainties in the distribution of the weights as done, e.g. by Kacharov et al. (2018, figs 8–13). In this case, it is important to perform the initial PPXF fit, from which the residuals are extracted, using some regularization, to obtain a less noisy and more representative best-fitting spectrum. I achieved good results perturbing the residuals using the easy-to-use wild bootstrap method (Davidson & Flachaire 2008).

3.5 Global non-linear fitting

In the most common situations, e.g. when fitting a single stellar kinematic component with emission lines, the spectral fitting problem has a single global minimum and the local optimization method of Section 3.2 is guaranteed to efficiently converge to it. However, in more complex situations, like when fitting multiple stellar or gas kinematic components, the fitting problem may present multiple minima and a local optimizer is not guaranteed to converge to the global minimum.

The standard way of dealing with multiple minima in PPXF is to perform the optimization of the variables in which the χ^2

⁶Available from <https://cvxopt.org/>

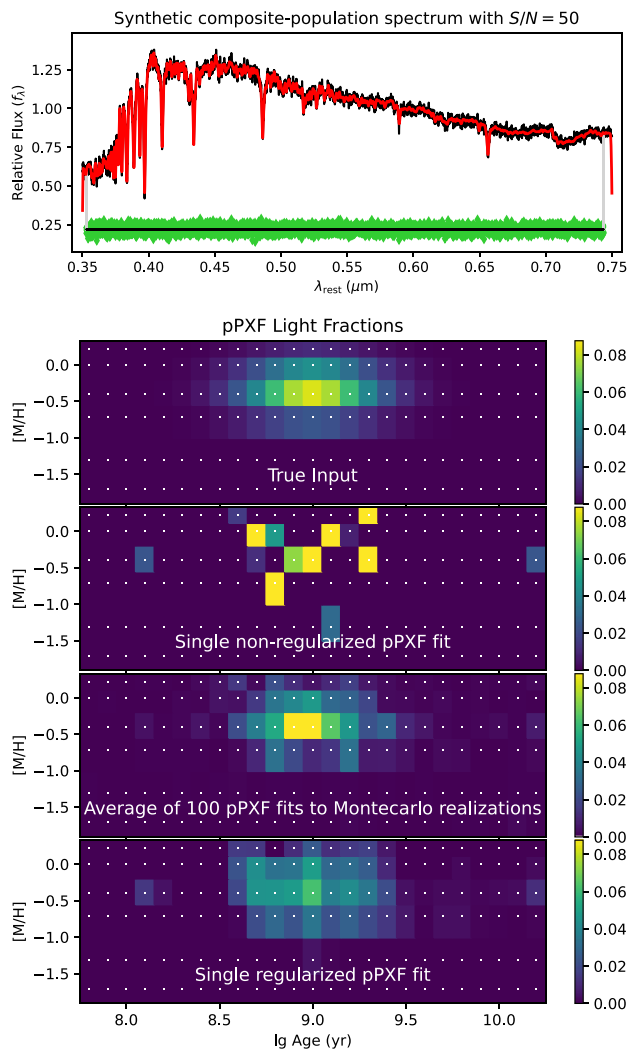


Figure 1. Top panel: synthetic input spectrum with Gaussian noise at $S/N = 50$. Second panel: input distribution of the V -band templates luminosity. Third panel: recovered weights from a single non-regularized PPXF fit to one Monte Carlo realization of the noise. Fourth panel: average of the weights recovered with PPXF by fitting 100 Monte Carlo realizations of the synthetic spectrum. Bottom panel: weights recovered with a single PPXF fit with regularization $regul=30$. This approximates the average distribution as expected.

function is multimodal outside of PPXF, while calling PPXF with those variables fixed, from inside a wrapper function. For example, when studying multiple kinematic components one may sample a grid of velocities and call PPXF with fixed velocities at every location (e.g. Mitzkus, Cappellari & Walcher 2017; Tabor et al. 2017; Bevacqua, Cappellari & Pellegrini 2022). If one is interested in the full posterior of certain parameters and computation time is not an issue, one may call PPXF with those parameters fixed from within a Bayesian method like MULTINEST (Feroz, Hobson & Bridges 2009), EMCEE (Foreman-Mackey et al. 2013), ADAMET (Cappellari et al. 2013a), or DYNesty (Speagle 2020), assuming the contribution of the non-fixed parameters to the posterior can be neglected.

In the current version of PPXF, one can also perform the global optimization within PPXF. This is currently implemented using the function `scipy.optimize.differential_evolution`, which uses the Differential Evolution algorithm by Storn & Price (1997). The Scipy function allows for linear constraints using the

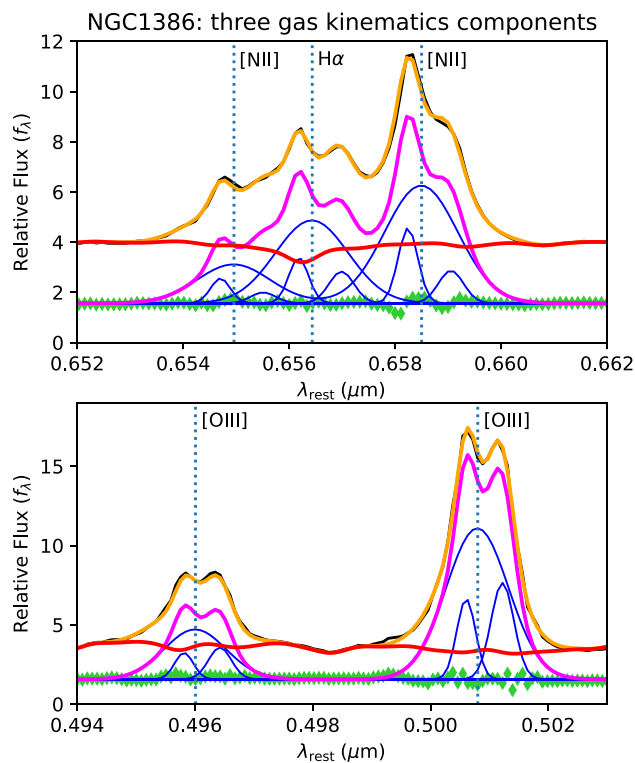


Figure 2. Fit of a MUSE spectrum of the active galaxy NGC 1386 (thin black line, mostly hidden by the fit), using the global-optimization option (`global_search=True`) and linear constraints (`constr_kinem`) in PPXF. Each of the five emission lines is modelled with three kinematic components (see text). The orange line is the PPXF total best fit, the red line is the best-fitting stellar spectrum alone, while the magenta is for the gas emissions alone, with individual components shown in blue. The fit residuals, arbitrarily offset, are shown with green diamonds. I only plot the region with the key emissions, but I fitted the full optical spectrum, which is needed to constrain the underlying stellar contribution.

method by Lampinen (2002). To save computation time, by default, I do not run the global optimization step until convergence, but I use it as starting point for the usual CAPFIT procedure.

An example of a situation where using both the global optimization and the linear constraints options can be useful, and the corresponding PPXF fit is shown in Fig. 2. The plots show the central spectrum of the active galaxy NGC 1386, extracted from the MUSE (Bacon et al. 2010) integral-field spectroscopic observations presented in Venturi et al. (2021). The emission line spectrum clearly requires at least three distinct kinematic components (see also Lena et al. 2015). The definition of the three kinematics components may appear ill-defined, due to the extensive blending of the lines. However, a meaningful decomposition can be obtained with some simple assumptions. Here I required the kinematics (V , σ) of all five emission lines to be the same within each of the three kinematic components. I additionally required the σ_{broad} of the broad component to be at least 200 km s^{-1} broader than either of the two narrow components as follows $\sigma_{\text{broad}} > \sigma_{\text{narrow}, 1}$ and $\sigma_{\text{broad}} > \sigma_{\text{narrow}, 2}$. These are linear constraints that I enforced using the `constr_kinem` keyword in PPXF. I also fix the ratios of the [OIII] and [NII] doublets to 1/3.

Linear constraints in PPXF were used extensively to produce the recent catalogue of broad and multiple gas emission line components for the full MaNGA galaxy survey (Fu et al. 2023).

3.6 Fitting spectra and photometry

Adding photometry to a full-spectrum fitting method is similar to adding a few extra pixels to the fit, which represent the fluxes measured in some observed photometric bands. The main differences are (i) that the photometric fluxes are independent of the line-of-sight velocity-distribution (LOSVD) $\mathcal{L}_n(v)$, unlike the spectroscopic ones and (ii) the photometry of a single galaxy is usually not enough to determine both the calibration errors and the template weights. This means that one cannot use polynomials as done for the spectroscopy.

I define a function that describes an individual template spectrum $T_n(\lambda)$ (either stars or gas), convolved $*$ with the LOSVD, which is allowed to be different for each of the N templates

$$g_n(\lambda) = T_n(\lambda) * \mathcal{L}_n(v). \quad (12)$$

With this notation, the model for the galaxy spectrum becomes

$$G_{\text{mod}}^{\text{spec}}(\lambda) = \sum_{n=1}^N w_n \left[g_n(\lambda) A_n(\lambda) \sum_{k=1}^K a_k \mathcal{P}_k(\lambda) \right] + \sum_{l=0}^L b_l \mathcal{P}_l(\lambda) + \sum_{j=1}^J c_j S_j(\lambda), \quad (13)$$

where the \mathcal{P}_k and \mathcal{P}_l are multiplicative and additive polynomials, respectively (of Legendre or trigonometric type), and the S_j are optional spectra of the sky. This model is similar to the one in Cappellari (2017, eq. 11), except that here each template spectrum can have a different attenuation function A_n . Moreover, both the attenuation and multiplicative polynomials can be used simultaneously rather than being alternatives. This is especially useful when including photometry in the fit.

The reason for this modification is that, when we have photometric bands that span a large wavelength range, we can infer the attenuation from the photometry itself, which is not modelled with polynomials. At the same time, we can use multiplicative polynomials to correct small errors in the spectral flux calibration. However, if we do not have photometry, we cannot tell apart reddening and multiplicative polynomials, because reddening is a special case of polynomials.

The model for the photometric measurements, in linear units which allow for negative fluxes, not magnitudes, is given by the following expression

$$G_{\text{mod}}^{\text{phot}}(\lambda_q) = \sum_{n=1}^N w_n \langle g_n(\lambda) A_n(\lambda) \rangle_q, \quad (14)$$

where $\langle \cdot \rangle_q$ represents the attenuated mean flux of the n -th template in the q -th photometric band with effective wavelength λ_q . Unlike the spectroscopic model of equation (13), the photometric model of equation (14) does not include polynomials or the sky spectrum.

In the common case of a photon-counting or energy-integrating detectors and assuming fluxes as f_λ (e.g. in units of $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$), the mean flux is given by (e.g. Bessell & Murphy 2012, eq. A11)

$$\langle f_\lambda \rangle_q = \frac{\int f_\lambda(\lambda) S_q(\lambda) \lambda d\lambda}{\int S_q(\lambda) \lambda d\lambda}, \quad (15)$$

where $S_q(\lambda)$ is the system photon response function and the integral extends over the region where S_q is non-zero. The definition of mean flux in equation (15) is the one used in the standard definition of magnitudes in the ultraviolet (e.g. for the GALEX spacecraft Martin et al. 2005), in the optical (e.g. for the SDSS optical survey York et al. 2000), or in the near-infrared (e.g. for the 2MASS survey Skrutskie et al. 2006). We can exactly convert mean fluxes in units of f_v (for

example, $\text{erg cm}^{-2} \text{s}^{-1} \text{Hz}^{-1}$) using this formula

$$\langle f_\lambda \rangle_q = \langle f_v \rangle_q \frac{c}{\lambda_p^2}, \quad (16)$$

where c is the speed of light and λ_p the source-independent pivot wavelength defined as (e.g. Koornneef et al. 1986; Bessell & Murphy 2012, eq. A16)

$$\lambda_p^2 = \frac{\int S(\lambda) \lambda d\lambda}{\int [S(\lambda)/\lambda] d\lambda}. \quad (17)$$

One could use different definitions of the observed mean fluxes by simply replacing equation (15).

In the common situation in which the covariance between the spectroscopic G^{spec} or photometric G^{phot} measurements are not known, or ignored, the residuals \mathbf{r} from the fit are as in Cappellari (2017, eq. 22)

$$r_p = \frac{G_{\text{mod}}^{\text{spec}}(\lambda_p) - G^{\text{spec}}(\lambda_p)}{\Delta G^{\text{spec}}(\lambda_p)}, \quad p = 1, \dots, P \quad (18a)$$

$$r_q = \frac{G_{\text{mod}}^{\text{phot}}(\lambda_q) - G^{\text{phot}}(\lambda_q)}{\Delta G^{\text{phot}}(\lambda_q)}, \quad q = P + 1, \dots, P + Q, \quad (18b)$$

with the difference that the vector of residuals now includes both the P spectroscopic and the Q photometric values. In other words, the total log-likelihood $\ln \mathcal{L}_{\text{total}}$ of a fit now becomes the sum of the spectroscopic and photometric ones

$$\ln \mathcal{L}_{\text{total}} = \ln \mathcal{L}_{\text{spec}} + \ln \mathcal{L}_{\text{phot}} = -\frac{\chi_{\text{spec}}^2 + \chi_{\text{phot}}^2}{2} + \text{const.} \quad (19)$$

Both the linear and non-linear fit and the regularization and the possible treatment of covariances proceed unchanged as already described in Cappellari (2017, sections 3.3–3.5). The only difference is one extra row in the matrix \mathbf{A} , defined in Cappellari (2017, section 3.3), for every photometric measurement.

According to the mean value theorem for integration, for every q -th band and n -th template, there exists a wavelength $\lambda_{q,n}$, which satisfies exactly

$$A_n(\lambda_{q,n}) \langle g_n(\lambda) \rangle_q = \langle g_n(\lambda) A_n(\lambda) \rangle_q. \quad (20)$$

When one has a good estimate of the galaxy redshift (e.g. from previous photometric redshift), or when performing a grid search for the best-fitting redshift with PPXF, the redshift of the spectrum does not change much during each PPXF fit. This makes the quantities $\langle g_n(\lambda) \rangle_q$ essentially independent of $\mathcal{L}_n(v)$. If I rewrite equation (14) as

$$G_{\text{mod}}^{\text{phot}}(\lambda_q) = \sum_{n=1}^N w_n A_n(\lambda_{q,n}) \langle g_n(\lambda) \rangle_q, \quad (21)$$

I can pre-compute the $\langle g_n(\lambda) \rangle_q$ and $\lambda_{q,n}$ for all templates before the fit using the initial redshift estimate. With this approach, adding photometry to a fit takes almost no extra time compared to fitting only the spectrum.

I found that the flux-weighted effective wavelength (e.g. Bessell & Murphy 2012, eq. A21)

$$\lambda_{q,n}^{\text{eff}} = \frac{\int g_n(\lambda) S_q(\lambda) \lambda^2 d\lambda}{\int g_n(\lambda) S_q(\lambda) \lambda d\lambda}, \quad (22)$$

well approximates the wavelength $\lambda_{q,n}$ defined by equation (20) for a range of attenuation parameters. In Fig. 3, I illustrate how accurately equation (20) is verified when approximating $\lambda_{q,n} \approx \lambda_{q,n}^{\text{eff}}$. I used all the 28 photometric bands described in Section 4.2 and all 387 FSPS SPS templates introduced in Section 4.3, which span extreme

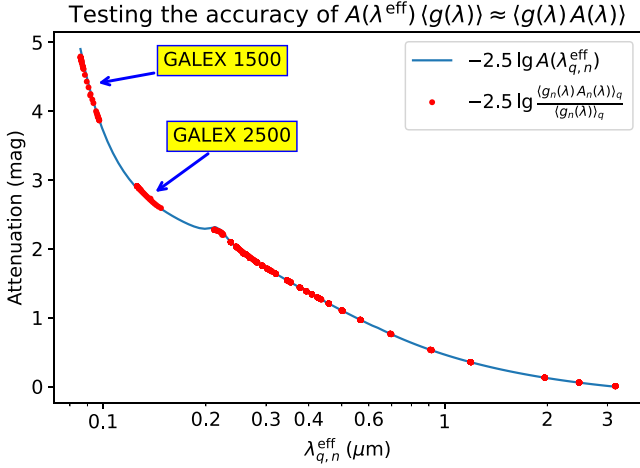


Figure 3. Comparison between the variation in the mean flux due to dust attenuation $\langle g_n(\lambda) A_n(\lambda) \rangle_q / \langle g_n(\lambda) \rangle_q$ (red circles) and the value of the attenuation curve at the rest-frame effective wavelength $A_n(\lambda_{q,n}^{\text{eff}})$ (blue solid line), plotted versus the rest-frame effective wavelength $\lambda_{q,n}^{\text{eff}}$ of each q -th filter and n -th template combination. This is computed for 28 filters at redshift $z = 0.8$ and a set of templates spanning extreme ages and metallicities as described in the text. The $\lambda_{q,n}^{\text{eff}}$ relative variation and the corresponding variation of the attenuation at that wavelength for a given filter are only significant in the far ultraviolet GALEX filters (indicated by the blue arrows). Even in that case, $A_n(\lambda_{q,n}^{\text{eff}})$ accurately predicts the true integrated attenuation.

ranges of age and metallicity. I adopt the attenuation function of equation (23), with realistic parameters $(A_V, \delta, E_b, f_{\text{nodust}}) = (1, 0, 0.25, 0)$ and the median redshift $z = 0.8$ of the LEGA-C sample. For every template-band combination, I compare the rigorous variation in mean flux due to the attenuation of each n -th template in the q -th band $\langle g_n(\lambda) A_n(\lambda) \rangle_q / \langle g_n(\lambda) \rangle_q$ with the value of the attenuation curve at the effective wavelength $A_n(\lambda_{q,n}^{\text{eff}})$. The two quantities must agree when the photometric band is narrow or the attenuation is approximately constant within the band, as is generally the case at optical or near-infrared wavelengths. However, in the far-ultraviolet, in the GALEX bands, $\lambda_{q,n}^{\text{eff}}$ and the corresponding $A_n(\lambda_{q,n}^{\text{eff}})$ vary significantly for different templates in the same band. Even so, as shown in Fig. 3 the approximation is still much better than our uncertainty of the attenuation curve itself. Moreover, for these large attenuations, generally, little flux is detected in the far-ultraviolet, making the observed uncertainties very large. For these reasons, in the analysis presented here, I use equation (21) to model the attenuation on the photometry. When higher accuracy is required, the full expression of equation (15) can be used.

3.7 Dust attenuation

As shown in equations (13) and (14), the new PPXF method allows each template to have a different attenuation curve. This feature can be used to vary the attenuation curve for specific groups of templates, based on the current understanding of dust attenuation in galaxies (see review by Salim & Narayanan 2020). Three groups of attenuation curves are expected to be useful: (i) for very young stars (with ages $t \lesssim 10^7$ yr), which are still embedded in their birth clouds (Charlot & Fall 2000; Granato et al. 2000); (ii) for the entire stellar population (both young and old), due to diffuse dust; and (iii) for the gas emission lines from star-forming regions.

In PPXF, one can adopt a generic function, which can be different for different templates and can have an arbitrary number of

parameters. The parameters can have values or can be kept fixed. By default, I currently implemented a four-parameters attenuation function in linear units $A(\lambda) = f(A_V, \delta, E_b, f_{\text{nodust}})$ defined by

$$D(\lambda) = \frac{E_b (\lambda \Delta\lambda)^2}{(\lambda^2 - \lambda_0^2)^2 + (\lambda \Delta\lambda)^2} \quad (23a)$$

$$k(\lambda) = \frac{A_V}{R_V} [k'(\lambda) + D(\lambda)] \left(\frac{\lambda}{\lambda_V} \right)^\delta \quad (23b)$$

$$A(\lambda) = f_{\text{nodust}} + (1 - f_{\text{nodust}}) 10^{-0.4k(\lambda)}. \quad (23c)$$

Here, equation (23a) is the Lorentzian-like Drude function adopted by Noll et al. (2009) to describe the UV bump around $\lambda_0 = 0.2175 \mu\text{m}$, with width $\Delta\lambda = 0.035 \mu\text{m}$. The equation (23b) is the expression adopted by Kriek & Conroy (2013), which includes the attenuation $k'(\lambda)$ and $R_V = 4.05$ from Calzetti et al. (2000, eqs. 4 and 5) and allows for a variable UV slope δ around the pivot V-band wavelength $\lambda_V = 0.55 \mu\text{m}$. Optionally, one can make E_b a function of δ (Kriek & Conroy 2013, eq. 3)

$$E_b = 0.85 - 1.9 \times \delta. \quad (24)$$

Finally, equation (23c) allows one to specify the fraction f_{nodust} of the stellar population (for the given template) that is unattenuated, as suggested by Lower et al. (2022). The resulting $A(\lambda)$ is the factor to multiply the template at the given wavelength to model the attenuation effect.

4 DATA AND SPS MODELS

In the rest of this paper, I present an analysis of the combined photometric and spectroscopic data, for a sample of about 3200 galaxies at $0.6 < z < 1$, making use of various of the new features of PPXF introduced in the first part of the paper.

4.1 Spectroscopy

I study a subset of the galaxy sample of the LEGA-C survey (van der Wel et al. 2016). It is an ESO ESO/Very Large Telescope (VLT) public spectroscopic survey targeting galaxies in the redshift range $0.6 < z < 1$, selected based on their observed K_s -band luminosity in the UltraVISTA/COSMOS catalogue by Muzzin et al. (2013), with a small variation of the K_s limit with z . In this work, I use the data from the LEGA-C third data release (DR3) presented in van der Wel et al. (2021). This redshift selection results in a mass-complete sample of 3445 galaxies in DR3 (90 per cent completeness above $\lg(M_*/M_\odot) \gtrsim 10.3$). The selection, the completeness level, the characteristics of the sample, and the data reduction are discussed extensively in van der Wel et al. (2016) and Straatman et al. (2018).

For my study, I focused only on the subsample of 3197 galaxies (including duplicates) in the DR3 catalogue with spectroscopic redshift $0.6 < z < 1$ and with measured stellar velocity dispersion σ_* . This a sample has redshift $z = [0.67, 0.76, 0.93]$ and average $S/N = [7, 14, 26]$ per \AA at the 16th ($-\sigma$), 50th (median) and 84th ($+\sigma$) percentiles. However, I verified that all my results are unchanged if I restrict the sample to the redshift range to $0.7 < z < 0.9$.

The survey data consist of spectra observed with the VLT/VIMOS multi-object spectrograph (Le Fèvre et al. 2003) covering the wavelength range $0.63\text{--}0.88 \mu\text{m}$ with a spectral resolution $R \approx 3500$, equivalent to an instrumental dispersion $\sigma_{\text{inst}} \approx 36 \text{ km s}^{-1}$ (van der Wel et al. 2021). For the PPXF fits, I logarithmically rebinned the spectra to a velocity scale $\Delta V = c \Delta \ln \lambda = \sigma_{\text{inst}}$ (Cappellari 2017, eq. 8) to make sure the spectrum is Nyquist sampled.

4.2 Photometry

I use two large collections of photometric measurements for the LEGA-C galaxies. The first is the UltraVISTA/COSMOS catalogue by Muzzin et al. (2013). It includes PSF-matched photometry in 30 bands from 0.15 to 24 μm . The catalogue is based on the *YJHK_s* NIR imaging data from UltraVISTA (McCracken et al. 2012). The optical data consist of broad-band Subaru/SuprimeCam data ($g^+r^+i^+z^+B_jV_j$), as well as u^* data from the CFHT/MegaCam (Capak et al. 2007; Taniguchi et al. 2007). It also includes the 12 optical medium bands (IA427-IA827) from Subaru/SuprimeCam (Capak et al. 2007). Also included are the GALEX FUV and NUV channels (Martin et al. 2005), and the 3.6, 4.5, 5.8, 8.0, and 24 μm channels from Spitzer’s IRAC + MIPS cameras (Sanders et al. 2007). The model predictions within the FUV GALEX band at $z \sim 0.8$ are quite uncertain, and few photons are generally expected to escape at those wavelengths. However, I still include this band in the fit to verify that this is indeed the case in the data. Moreover, significant detections are observed for the most star-forming galaxies.

The second photometric catalogue is the COSMOS2020 by Weaver et al. (2022). Highlights of this catalogue, compared to the one by Muzzin et al. (2013), are much deeper Subaru Hyper Suprime-Cam *grizy* broadband photometric measurements (Aihara et al. 2019) and deeper UltraVISTA DR4 observations *YJHK_s*. The extra depth is not an important feature for my study, as the LEGA-C galaxies were all well detected in Muzzin et al. (2013) by design. However, I use the COSMOS2020 to assess the sensitivity of my results to the use of independent data sets. For this work, I adopt the catalogue produced with the FARMER profile-fitting photometric extraction tool. For both photometric catalogues, I only included the typically 28 bands, which have a transmission full width at half-maximum (FWHM) fully contained in the wavelength range of the adopted stellar population templates (see later) at the redshift of each galaxy.

The current COSMOS2020 FARMER catalogue does not have the two GALEX bands, so I added them to compare it more closely with the UltraVISTA/COSMOS catalogue. I used the following steps for each galaxy: (i) I selected the bands that were common between the COSMOS2020 and the UltraVISTA/COSMOS catalogues. (ii) I used equation (34) to perform a linear least-squares fit and find a normalization factor κ that matches the two photometries for that galaxy. (iii) I applied the same factor κ to scale the two GALEX bands and included them in the COSMOS2020 bands.

4.3 Stellar population synthesis models

I used three independent SPS models to assess the sensitivity of the results to some of the adopted model assumptions. I selected the models based on two criteria (i) the ability to generate model spectra from the far UV at 0.1 μm to about 2 μm , to be able to constrain the decrease of f_λ towards the NIR region of the galaxy spectra and (ii) to include model spectra down to a young age of 1 Myr, to reproduce the many actively star-forming galaxies that are present in the sample. The age criterion forces me to exclude from this study the models by Vazdekis and Maraston, which I have extensively used in the past.

The three SPS models that satisfy my requirement and that I adopted are (i) the FSPS⁷ (Conroy et al. 2009; Conroy & Gunn 2010), (ii) the GALAXEV⁸ (Bruzual & Charlot 2003) and (iii) the BPASS⁹ SPS models (Stanway & Eldridge 2018; Byrne et al. 2022).

⁷ Available from <https://github.com/cconroy20/fps>

⁸ Available from <http://www.bruzual.org/bc03/>

⁹ Available from <https://bpass.auckland.ac.nz/>

For all three models, I tried to select a consistent set of templates. In all cases, I adopted the same set of 43 ages logarithmically spaced by 0.1 dex from 1 Myr to 15.85 Gyr, defined as

$$\lg(\text{Age}/\text{yr}) = 6, 6.1, 6.2, \dots, 10.2. \quad (25)$$

The oldest age is 2–3 \times older than the age of the Universe at the redshift of the sample, which varies in my standard cosmology from 5.75–7.75 Gyr between $z = 1$ and 0.6, but I did not truncate the models to physical ages, to check how well the data themselves can constrain the galaxy ages. For comparison, I additionally ran models where I constrained the maximum age in the fit to each galaxy to the age of the Universe at its redshift. I also excluded from all models the most extreme low metallicities $[Z/H] < -2$. Here is my other setup for the three SPS:

(i) **The FSPSmodels** allow one to compute SPS models for a specified set of parameters. I used the Python bindings¹⁰ (Johnson et al. 2021a) and the latest public v3.2 to compute a set of spectra with the above ages and 9 equally-spaced metallicities $[Z/H] = [-1.75, -1.5, -1.25, -1., -0.75, -0.5, -0.25, 0, 0.25]$, for a total of 387 SPS templates. I adopted a Salpeter (1955) IMF with a lower/upper mass cut of 0.08 and 100 M_\odot respectively, for consistency with BPASS and used the MIST isochrones (Choi et al. 2016). But I note that my results are virtually insensitive to the slope of the IMF at lower masses. I computed the SPS without including the effect of gas or dust and adopted default parameters for the other parameters. This returns SPS spectra computed using the MILES stellar library (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011) for the optical region, which is the one I fit in the LEGA-C spectra.

(ii) **The GALAXEVmodels** provide a Fortran code (version 2020) which I used to produce a set of SPS spectra with the same ages as above and computed at the provided 5 metallicities $[M/H] = [-1.74, -0.73, -0.42, 0, 0.47]$, for a total of 215 SPS templates. This SPS model also uses the MILES library to generate spectra of the optical region. Also here I adopted a Salpeter IMF. The models use the Padova isochrones (Bertelli et al. 1994; Girardi et al. 2000; Marigo et al. 2008).

(iii) **The BPASSmodels** v2.3 are provided as a set of files pre-computed at a given set of metallicity. I adopted the 10 metallicities¹¹ $[Z/H] = [-1.3, -1, -0.8, -0.7, -0.5, -0.4, -0.3, 0, 0.2, 0.3]$, for a total of 430 SPS templates. The models are provided for a single IMF having a power slope -2.35 (Salpeter slope) above $M > 0.5 M_\odot$ and a slope -1.3 at lower masses. I used the version of the SPS for single stars, ignoring binaries, with $[\alpha/Fe] = 0$, for consistency with the other two SPS models. These SPS models are fully synthetic. They use isochrones produced by a derivative of the Cambridge STARS code (Eggleton 1971) as described by Eldridge, Izzard & Tout (2008).

5 DYNAMICAL MASSES FROM SERSIC PHOTOMETRY

We do not know the true masses of galaxies, so we cannot tell how good the mass estimates from different stellar population codes are. One way to test the accuracy is to use good dynamical models. I show how I do this in this section. I use mass-follow-light axisymmetric JAM dynamical models that are as quick and simple to use as the usual virial estimator (e.g. Cappellari et al. 2006) and require the same data, but do not have its problems.

¹⁰ Available from <https://github.com/dfm/python-fps>

¹¹ The models are specified in metal mass fraction Z , and I converted it to $[Z/H]$ with ~ 10 per cent accuracy.

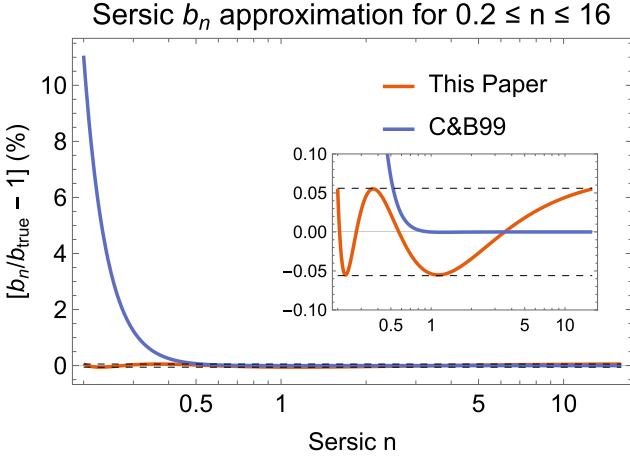


Figure 4. Fractional difference between the true coefficient $b(n_{\text{Ser}})$ in equation (26) and the minimax approximation (red line) for the interval $0.2 < n_{\text{Ser}} < 16$ of equation (28). I also show for comparison the approximation (blue line) by Ciotti & Bertin (1999), which is very accurate for $n_{\text{Ser}} > 0.5$, but starts deviating rapidly at smaller n_{Ser} . The inset shows the same thing using a different scale for the y-axis.

5.1 Updated coefficient for the Sersic profile

I assume a galaxy with surface brightness described by a Sersic profile with elliptical isophotes of constant axial ratio $q_{\text{obs}}^{\text{Ser}}$

$$I(m) = I(0) \exp \left[-b(n_{\text{Ser}}) \left(\frac{m}{R_{\text{e}}^{\text{Ser}}} \right)^{1/n_{\text{Ser}}} \right], \quad (26)$$

where the elliptical radius is

$$m^2 = x^2 + \left(\frac{y}{q_{\text{obs}}^{\text{Ser}}} \right)^2, \quad (27)$$

with x aligned along the galaxy photometric projected major axis.

The parameter $b(n_{\text{Ser}})$ is defined by the requirement that $R_{\text{e}}^{\text{Ser}}$ represents the semimajor axis of the isophote containing half of the total luminosity of the Sersic model. This implies that $b(n_{\text{Ser}})$ is the solution of $\Gamma(2n_{\text{Ser}}) = 2\Gamma(2n_{\text{Ser}}, b(n_{\text{Ser}}))$ (Ciotti 1991), where $\Gamma(a, z)$ is the incomplete gamma function (Olver et al. 2010, eq. 8.2.2) and $\Gamma(a) = \Gamma(a, 0)$ is the complete one (Olver et al. 2010, eq. 5.2.1). A useful approximation for $b(n_{\text{Ser}})$ was presented by Ciotti & Bertin (1999). This is very accurate for values of $n_{\text{Ser}} \gtrsim 0.5$ but starts becoming rapidly inaccurate for smaller n_{Ser} values. The LEGA-C catalogue contains values of the Sersic index down to $n_{\text{Ser}} = 0.2$, where I found that the Ciotti & Bertin (1999) approximation for $b(n_{\text{Ser}})$ reaches an error of 11 per cent (Fig. 4).

To overcome this limitation, I computed an alternative approximation for $b(n_{\text{Ser}})$. I adopted the same number of terms and the same mathematical form, but I adjusted the coefficients to obtain the minimax solution, minimizing the maximum absolute relative error of $b(n_{\text{Ser}})$ using non-linear optimization over an interval including the most extreme Sersic indices existing in the literature. I found the following expression

$$b(n_{\text{Ser}}) = \frac{0.000207}{n_{\text{Ser}}^2} + \frac{0.015987}{n_{\text{Ser}}} - 0.34025 + 2.0015 \times n_{\text{Ser}}, \quad (28)$$

which has a maximum absolute relative error of 5.6×10^{-4} in the whole interval $0.2 \leq n_{\text{Ser}} \leq 16$. The fact that the relative error reaches its maximum value, with alternating sign, at five values of n_{Ser} (Fig. 4)

confirms that this is indeed the minimax solution for the adopted function and interval (e.g. Press et al. 2007, section 5.15).

5.2 Jeans Anisotropic Models from Sersic photometry

I performed dynamical modelling of the full LEGA-C sample of 3197 galaxies with measured σ_* and $0.6 < z < 1$ using the Jeans Anisotropic Modelling (JAM) method¹² (Cappellari 2008, 2020).

Dynamical modelling of LEGA-C galaxies with JAM was previously done for a subset of 797 galaxies by van Houdt et al. (2021) using a Bayesian approach, but I extended it to all galaxies with measured σ_* . A simpler alternative to dynamical models would be to use a virial estimation of the galaxy masses (e.g. Cappellari et al. 2006) based on the fitted parameters of the (Sersic 1968) profiles provided in the LEGA-C DR3 catalogue (van der Wel et al. 2021). The virial estimates are also included in the DR3 catalogue.

The main advantage of virial estimators is that they are fast and easy. However, a common limitation is that they do not account for differences in the spectroscopic aperture and the instrumental point-spread function, which can be significant, especially in high-redshift observations like LEGA-C. Moreover, virial estimators do not allow one to explicitly assume an intrinsic shape or anisotropy for the galaxies under study.

Although virial estimators are still useful to study scaling relations (e.g. Cappellari et al. 2013a; Li et al. 2018; Zhu et al. 2023a), nowadays, there is no reason to use them to compute galaxy masses. In fact, using, e.g. the public JAM package, one can compute a more reliable dynamical mass for a galaxy approximated by a Sersic model with a similar time and effort as using the virial estimator, while allowing for intrinsic shape, anisotropy, aperture, and PSF effects, without introducing unnecessary approximations.

To simplify this task of computing accurate dynamical masses of galaxies with available fitted Sersic parameters and σ_* , I developed a simple procedure JAM_AXI_SERSIC_MASS, and I have made it publicly available in the updated version 7.2 of the JAM package. The procedure requires the following inputs from the users:

- (i) The parameters of the Sersic model for a galaxy:
 - (a) The semimajor axis of the half-light isophote $R_{\text{e}}^{\text{Ser}}$,
 - (b) The Sersic index n_{Ser} , and
 - (c) The observed axial ratio of the isophotes $q_{\text{obs}}^{\text{Ser}}$.
- (ii) The assumptions about the intrinsic properties of the galaxy:
 - (a) The intrinsic axial ratio $q_{\text{intr}}^{\text{Ser}}$ and
 - (b) The typical orbital anisotropy β .
- (iii) The parameters of the spectroscopic observations:
 - (a) The size and shape of the spectroscopic aperture, and
 - (b) The parameters of the PSF.
- (iv) The observed second moment σ_* (which includes both rotation and random motions) and uncertainty of the stellar line-of-sight velocity distribution, preferably measured at a similar wavelength as the imaging used to fit the Sersic model.

Given an angular diameter distance D_A , the procedure then returns the dynamical mass M_{JAM} of the Sersic model and its formal uncertainty in a fraction of a second.

The procedure uses the method and MGE_FIT_ID routine within the MGEFIT package¹³ (Cappellari 2002) to accurately fit the one-

¹²I used v7.2 of the JAMPY Python package <https://pypi.org/project/jampy/>

¹³V5.0 of Python MGEFIT from <https://pypi.org/project/mgefit>

dimensional Sersic profile of equation (26) with a one-dimensional Multi-Gaussian Expansion (MGE). Then assumes a fixed axial ratio $q_{\text{obs}}^{\text{Ser}}$ for all MGE Gaussians and an arbitrary reference total mass M_0 for the model. It uses the JAM_AXI.PROJ procedure in the JAMPY package (Cappellari 2008, 2020) to calculate a PSF-convolved prediction for the $V_{\text{rms},j}$ at a large set of discrete sky locations (x_j, y_j) finely sampling the adopted spectroscopic aperture. The luminosity-weighted second moment inside the whole aperture is computed as

$$V_{\text{rms}}^2 = \frac{\sum_j I_j V_{\text{rms},j}^2}{\sum_j I_j}, \quad (29)$$

where the summation extends to the pixels of flux I_j inside the aperture. Given the general scaling $M \propto V^2$ between the total mass and velocities in a dynamical model, the dynamical mass of the Sersic model is then given by

$$M_{\text{JAM}} = \frac{M_0 \sigma_*^2}{V_{\text{rms}}^2}. \quad (30)$$

The physical meaning of the dynamical mass M_{JAM} , as derived from mass-follow-light dynamical models or virial estimators, is often a source of confusion. This is because M_{JAM} is neither a total stellar mass, nor a total mass of a galaxy which includes its dark halo. Moreover, the value of M_{JAM} is highly dependent on the extrapolated outer profile. For example, a Sersic model with $n_{\text{Ser}} = 6$ contains 21 per cent of its total light outside $4R_e$, which is about the maximum radius one can observe in typical photometry. This strong dependence on extrapolation prevents comparison of masses when accuracies better than a few 20 per cent are desired.

The quantity that both dynamical and stellar population models are robustly measuring is the mass-to-light radius within the region covered by the spectroscopic (for the dynamics) or photometric (for the population) observations. Specifically, if one divides the Sersic dynamical mass M_{JAM} returned by the procedure by the analytic total luminosity of the same Sersic model (Ciotti 1991)

$$L_{\text{Ser}} = \pi I(0) \Gamma(2n_{\text{Ser}} + 1) \left(\frac{R_e^{\text{Ser}}}{b^{n_{\text{Ser}}}} \right)^2 q_{\text{obs}}^{\text{Ser}}, \quad (31)$$

the mass-to-light ratio

$$(M/L)_{\text{JAM}} = M_{\text{JAM}}/L_{\text{Ser}} \quad (32)$$

provides a very accurate approximation of the average M/L within a sphere of radius comparable to the size of the spectroscopic aperture.

I used JAM_AXI_SERSIC_MASS to compute M_{JAM} and $(M/L)_{\text{JAM}}$ for all the LEGA-C galaxies in my subsample. I assumed a spectroscopic aperture of 1×1 arcsec² and a characteristic PSF of 0.75 FWHM from van Houdt et al. (2021). I adopted as the intrinsic axial ratio for all galaxies the mean value $q_{\text{intr}}^{\text{Ser}} = 0.41$ of the Gaussian distribution inferred by van Houdt et al. (2021) by inverting the observed shape distribution of the LEGA-C sample. For the anisotropy, I used a typical value $\beta = 0.2$ expected for the assumed mean intrinsic shape (fig. 9, Cappellari 2016). I assumed a cylindrically oriented (`align='cyl'`) velocity ellipsoid for JAM (Cappellari 2008). The parameters of the Sersic profiles R_e^{Ser} , n_{Ser} , $q_{\text{obs}}^{\text{Ser}}$ and L_{Ser} come from the LEGA-C DR3 catalogue (van der Wel et al. 2021).

I also used the dynamical models to calculate the average projected density Σ_1^{JAM} within a circle of radius $R = 1$ kpc at the angular diameter distance of the galaxy. This quantity was shown to closely relate to galaxy quenching (Cheung et al. 2012; Fang et al. 2013), like σ_* . However, here instead of using the stellar mass as in previous studies, I used the dynamical mass from JAM. I obtained this value by circularizing and analytically integrating the best-fitting MGE as described in eq. (11) of Cappellari et al. (2013a).

6 SETUP FOR PPXF AND TESTS

In this study, fitting (twice) the VIMOS spectrum (≈ 2800 spectral pixels) and typically 28 photometric bands for a single galaxy with PPXF takes about 1 min. This compares with the ‘roughly 100 CPU hours’ reported by Tacchella et al. (2022) in a similar state-of-the-art study using DEIMOS spectroscopy and the PROSPECTOR Bayesian code (Johnson et al. 2021b). This is a computation time difference of nearly four orders of magnitude! Of course, the two methods perform quite different tasks and the large computational cost is a standard feature of Bayesian methods and not a weakness by itself. However, the execution time of a method affects the kind of tasks one can address and the variety of modelling choices one can explore as I outline in this section.

6.1 Non-parametric population model

A key feature made possible by least-squares methods is the ability to explore non-parametrically the joint distribution of SFH and chemical composition with high resolution. My setup uses 43 non-parametric age bins, and each age bin is allowed to have a different non-parametric metallicity (5–10 bins depending on the SPS code) for a total of up to 430 bins. This contrasts with the 10 non-parametric age bins and the single metallicity for the entire galaxy adopted by Tacchella et al. (2022). Crucially, even when using a non-parametric grid of a few hundred templates the least-squares method *guarantees* global convergence to the most likely weights distribution. This is because the constrained quadratic-programming problem being solved (Cappellari 2017, eq. 27) is known to possess a unique global minimum (e.g. Nocedal & Wright 2006). This contrast with Bayesian methods, where Tacchella et al. (2022, sec. 3.2) reported that ‘the fits do not converge within a reasonable amount of time’ with 14 non-parametric bins. The use of non-parametric models is important for a proper recovery of the stellar population in galaxies (Lower et al. 2020).

6.2 Polynomials

A least-squares method like PPXF allows for a quick exploration of different modelling assumptions. In the course of this study, I was able to easily test different options, each with all three SPS models, for all 3200 galaxies, to test how they affected the final results. To fit the spectrum, we can choose between additive or multiplicative polynomials (the photometry does not use any polynomials). Additive polynomials are specified by the PPXF keyword `degree` and can help with template mismatch, AGN modelling or sky subtraction errors. Multiplicative polynomials are specified by the PPXF keyword `mdegree` and can account for spectral flux calibration issues or reddening effects.

For my tests, I run models with both multiplicative and additive polynomials `degree` from `mdegree=degree=-1` (i.e. using only attenuation and no polynomials) to `degree=4` and found that the solution changes slightly without polynomials but quickly stabilizes as soon as one allows for a non-zero degree. Results were similar when only using additive or only multiplicative polynomials to adjust the spectral stellar continuum. This was a non-obvious result in the present analysis, given that the LEGA-C spectra were calibrated using SPS models (van der Wel et al. 2021) and this non-standard calibration may leave an influence on the results. The polynomials should effectively remove any memory of possible inaccuracies in spectral calibration. I adopted the PPXF keywords `mdegree=2`, `degree=-1` for my standard setup.

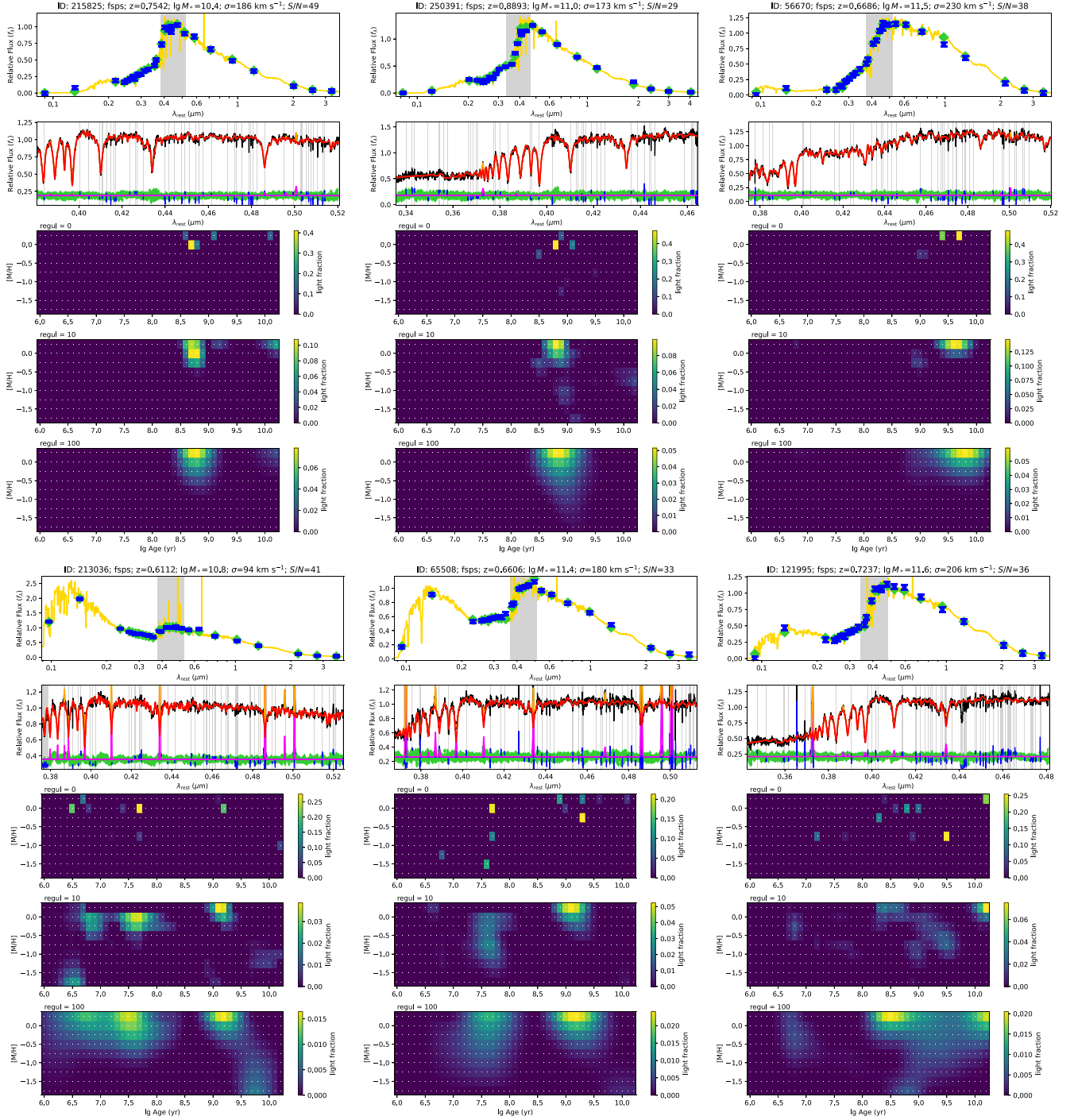


Figure 5. Examples of PPDF fits to the LEGA-C galaxy spectra and 28-bands photometry using the FSPS models. The top three galaxies require a model with a single short star formation event, while the bottom three galaxies require multiple discrete star formation events. For each galaxy, the top panel shows the photometric measurements (blue error bars) and the best fit (green diamonds), while the golden line shows the underlying best-fitting template with included emission lines. The grey vertical band indicates the range where spectroscopy was also fitted. The second panel shows the observed spectrum (black line) and the best-fitting total spectrum (orange line). The best-fitting stellar spectrum alone is shown in red and the gas emission one is in magenta. The residuals (arbitrarily offset) are indicated with green diamonds and the masked pixels with blue lines (and corresponding grey vertical bands). The last three panels show the distribution of the PPDF weights, indicating the bolometric luminosity L_{bol} of each stellar population of given age and metallicity. The weights are shown for (i) no regularization, (ii) regularization $\text{regul}=10$ and (iii) $\text{regul}=100$ as written in the plots.

6.3 Dust attenuation model

I truncated all three the SPS templates to $0.01 \mu\text{m} < \lambda < 5 \mu\text{m}$ (except for the `bpass` SPS which only extend to $\lambda < 2 \mu\text{m}$) to remove the influence of dust on the spectral shape (e.g. Conroy 2013,

fig. 1). This is because the modelling of dust from the energy balance of UV light reradiated to the IR requires several further assumptions and is not implemented in all modelling codes. Moreover, the currently available bands are included in the fitted range anyway.

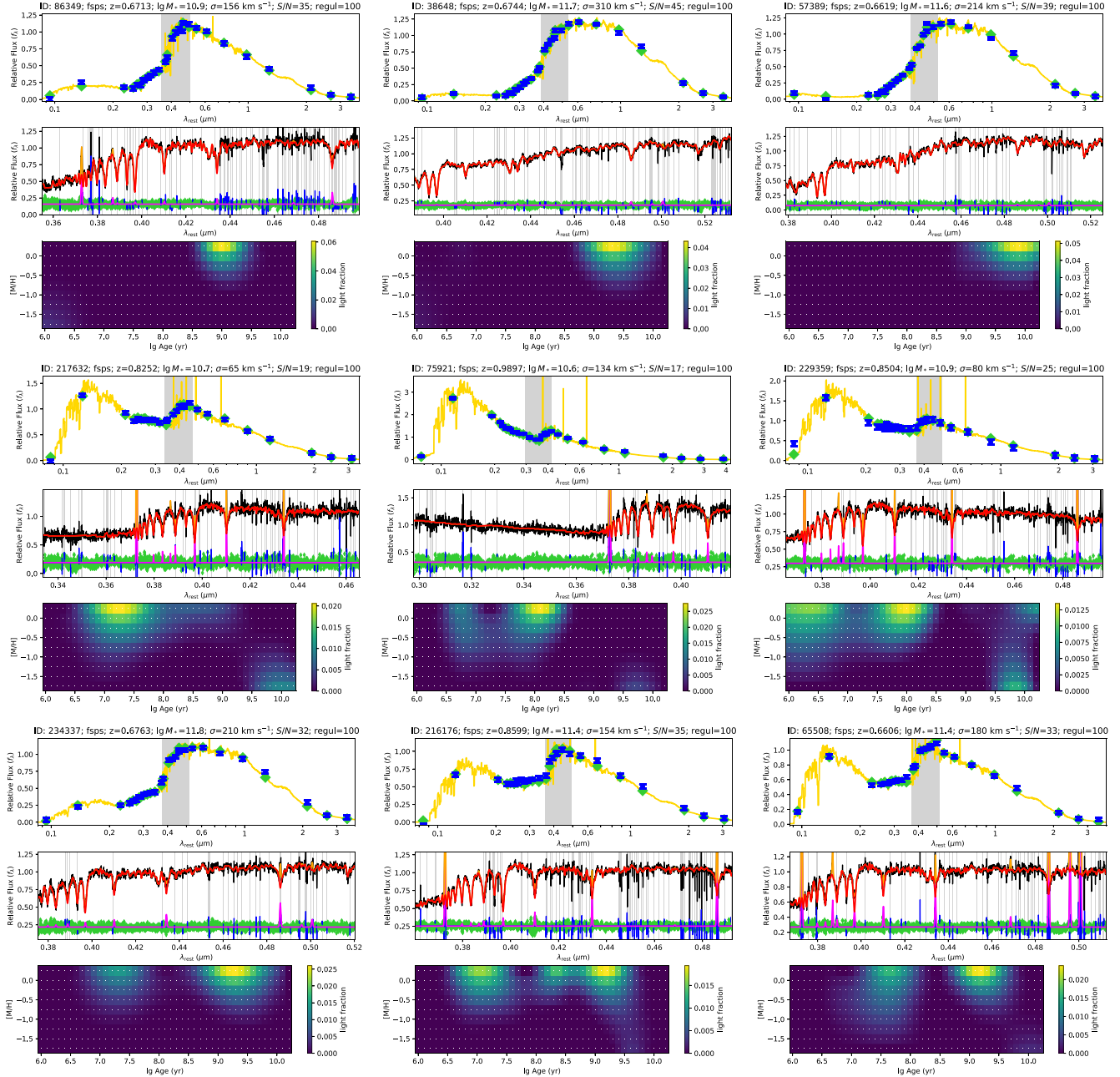


Figure 6. More examples of PPF fits to spectra and photometry. The meaning of the symbols is the same as in Fig. 5, but I show a single regularization ($\text{regul}=100$). The top three galaxies require a single star formation event, while the rest can only be modelled with multiple discrete star formation events.

This situation is changing rapidly with the JWST and will soon be revisited.

For the full set of 3200 galaxies, I experimented with three different assumptions for the attenuation: (i) I adopted a single four-parameter attenuation for all stellar templates as in equation (23); (ii) I reduced the attenuation curve to two parameters (A_V , δ) by assuming $f_{\text{nodust}} = 0$ and adopting the $E_b - \delta$ relation of equation (24); I still applied this attenuation for all stellar templates; (iii) I adopted the two-components attenuation model by Charlot & Fall (2000). For this, I used a birth-cloud attenuation of the form

$$k(\lambda) = A'_V \left(\frac{\lambda}{0.55 \mu\text{m}} \right)^{-1}. \quad (33)$$

I apply this only to the stellar templates younger than 10 Myr, while I used the same attenuation as in (ii) for the diffuse dust component affecting all stellar templates. In all three cases, I fit a different Calzetti et al. (2000) attenuation curve (equation 23 with $E_b = \delta = f_{\text{nogas}} = 0$) for the gas emission lines templates.

I found that options (i) and (ii) produce an insignificant difference in the final results, but in the first case, there is more degeneracy in the attenuation parameters, which makes any trend between dust parameters less obvious. Option (iii) generates a similar result for the older and intermediate populations, as expected. However, when allowing the youngest population to have its own attenuation, this becomes completely degenerate with the amount of star formation in that same young component. As a result, one can obtain good

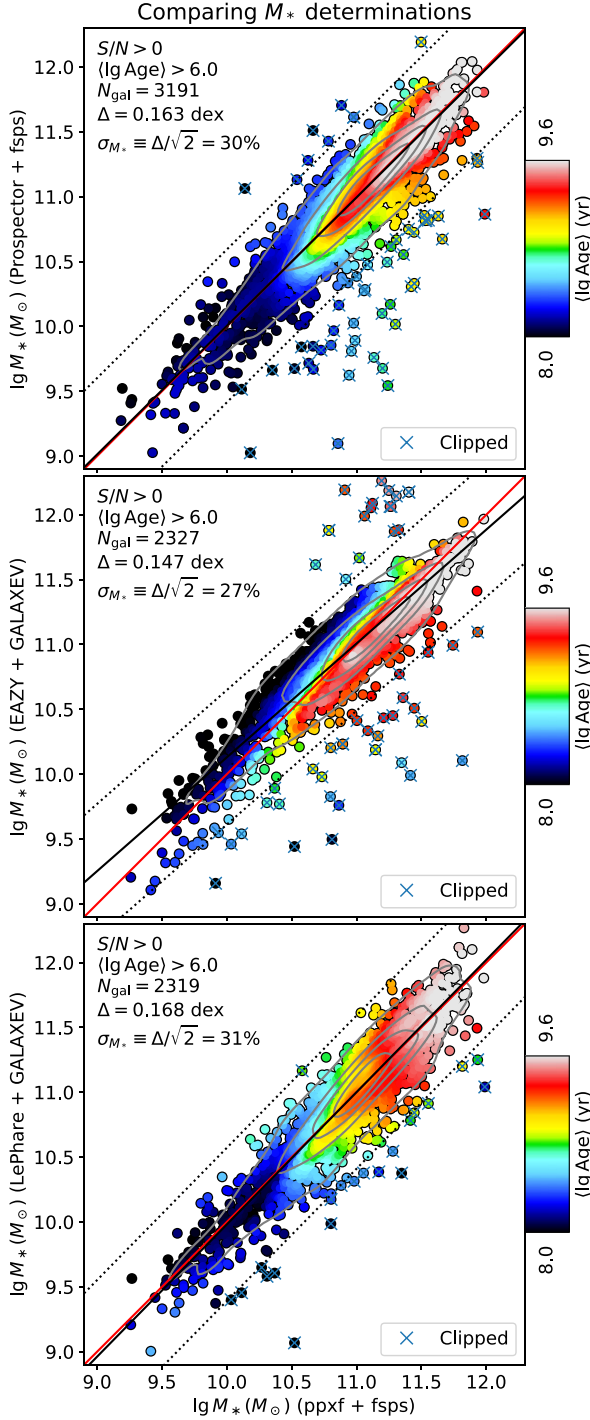


Figure 7. Comparison of PPXF stellar population mass M_* estimates against different methods. I use the M_* values from PROSPECTOR in van der Wel et al. (2021) or from EAZY and LEPHARE in Weaver et al. (2022), for the galaxies that are common in both catalogues. The three panels show the comparison for each method. The red line is the one-to-one relation, and the black line is the best fit using LTSFIT (Cappellari et al. 2013a). The x-symbols are the outliers identified by LTSFIT, and the dotted lines are the selection limits. The colours show the luminosity-weighted mean stellar age ($\lg \text{Age}$), smoothed by LOESS. The grey contours show the kernel density estimation of the galaxy distribution. I indicate the S/N and ($\lg \text{Age}$) selection criteria, the number of selected galaxies, the rms scatter Δ , and an approximate relative 1σ error σ_{M_*} in M_* for each panel. I estimated the latter as $\sigma_{M_*} \equiv \Delta/\sqrt{2}$ by assuming both M_* have the same uncertainty.

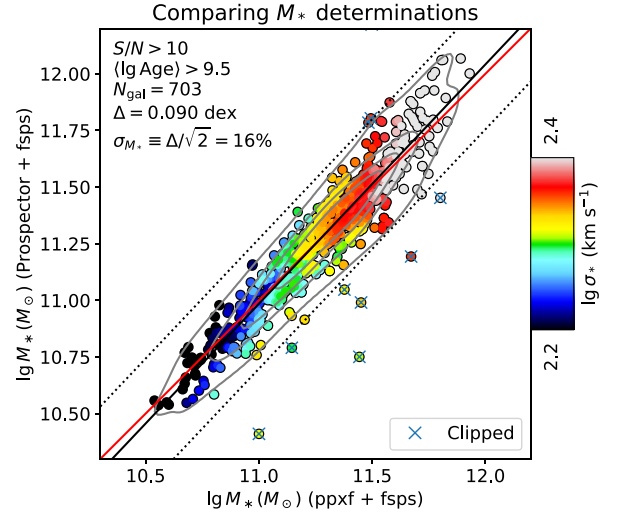


Figure 8. Same as in Fig. 7 for a subsample with high spectral $S/N > 10$ and old ages ($\lg \text{Age}$) > 9.5 . As the age range is limited, the colours show here the LOESS smoothed stellar velocity dispersion.

fits with an unlikely large attenuation associated with an equally large star formation, but with complete degeneracy between the two parameters. This is the well-known SFR-dust degeneracy mentioned in Section 1. It can be broken by introducing extra assumptions on the dust geometry and reradiated UV fraction, combined with rest-frame IR data, which I do not have.

In conclusion, I adopted as my standard choice the two-parameters attenuation function of option (ii), which applies to the whole population, as done, e.g. by Kriek & Conroy (2013). I enforced bounds on the parameters as $-1 < \delta < 0.4$ and $0 < A_V < 4$. As illustrated by Kriek & Conroy (2013, fig. 1), their two-parameters parametrization differs from both the Milky Way (Cardelli, Clayton & Mathis 1989) and Calzetti et al. (2000) attenuation laws but appear to better describe high- z galaxies. However, using the full four-parameter attenuation curve, which covers the Calzetti et al. (2000) curve and other curves from the Milky Way to the Large Magellanic Cloud (Gordon et al. 2003), does not alter my scientific conclusions.

6.4 Matching photometry and spectra

I have a spectroscopic redshift for every galaxy from the LEGA-C catalogue. I only included photometric bands in the fit if their FWHM is fully enclosed by the template's wavelength coverage at that redshift. Before the PPXF fits, I pre-computed the $\langle g_n(\lambda) \rangle_q$ and $\lambda_{q,n}$ in equation (21). I used the same photonic throughput file for all filters, produced for the EAZY code (Brammer et al. 2008) and available online.¹⁴

Spectroscopy was observed within 1 arcsec slits, while photometric observations were either measured within a $2''.1$ aperture (Muzzin et al. 2013) or are total magnitudes (Weaver et al. 2022). This means that calibration is needed to match the flux levels of photometry and spectroscopy. It's important to note that in some cases, the photometric fluxes may correspond to a different stellar population than that sampled by the spectra. With this in mind, I assumed that spectroscopy and photometry originate from a single spectral energy distribution and applied a constant scaling factor κ to the spectrum $G^{\text{spec}}(\lambda_p)$. This factor ensures that the synthetic photometry derived

¹⁴Available from <https://github.com/gbrammer/eazy-photos>

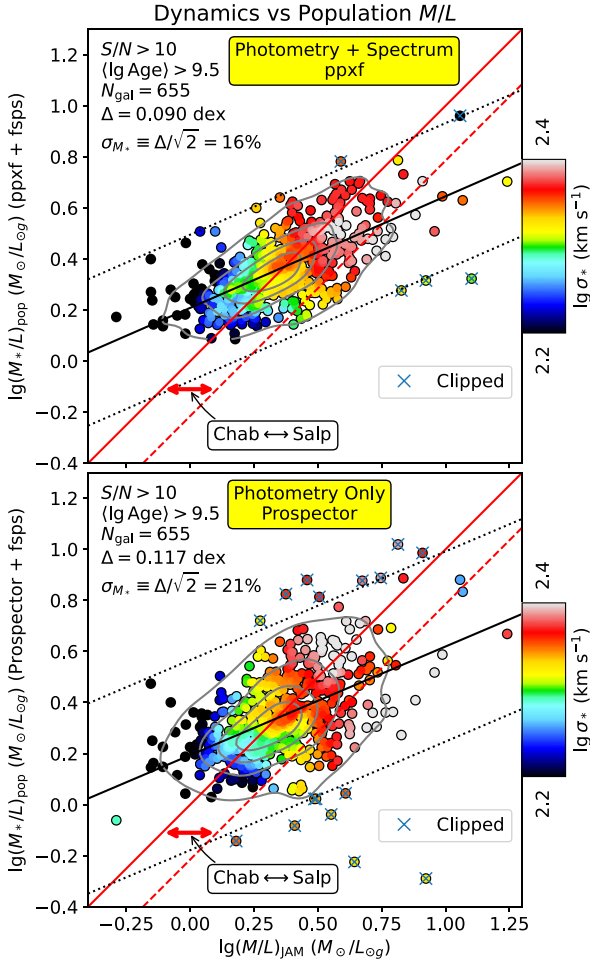


Figure 9. Comparison of dynamical and stellar population mass-to-light ratios. The dynamical mass-to-light ratio $(M/L)_{\text{JAM}}$ is derived from the JAM modelling, while the stellar population mass-to-light ratio $(M/L)_{\text{pop}}$ is derived from either PPXF (top panel) or PROSPECTOR (bottom panel). The red line shows the one-to-one relation, while the red dashed line indicates the shift that would be produced by changing the IMF from Chabrier to Salpeter. The black line shows the best linear fit using LTSFIT, which strongly deviates from the one-to-one relation. The x-symbols mark the outliers identified by LTSFIT, and the dotted lines mark the selection limits. The colours indicate the stellar velocity dispersion σ_* , smoothed by LOESS. The grey contours indicate a kernel density estimate of galaxies. For each panel, I show the signal-to-noise ratio S/N and the mean logarithmic age $\langle \text{lg Age} \rangle$ selection criteria, the number of selected galaxies N , the root mean square scatter Δ , and an approximate relative error σ_{M_*} in stellar mass M_* for both methods. I estimate σ_{M_*} as $\sigma_{M_*} \equiv \Delta/\sqrt{2}$ by assuming both methods have the same uncertainty.

from the spectrum using filter transmission curves matches the observed LEGA-C photometry in bands covered by the spectroscopy. This step calibrates the overall normalization of the spectrum flux using photometry before starting the PPXF fit.

To match the VIMOS spectra to the photometry, before calling PPXF, for every galaxy I first computed synthetic photometric fluxes $\langle f_\lambda \rangle_q$ from its VIMOS spectrum using equation (15), for the subset of photometric bands (typically 7) contained within the VIMOS wavelength range. I then multiplied the spectrum by a factor κ to minimise the χ^2 between the synthetic and observed photometry, only for the few bands in common. It can be computed with the

general analytic linear-fitting relation (e.g. Cappellari 2008, eq. 51)

$$\kappa = \frac{\mathbf{d} \cdot \mathbf{m}}{\mathbf{m} \cdot \mathbf{m}}, \quad (34)$$

where the ‘data’ vector \mathbf{d} has elements $d_q = p_q/\Delta p_q$, the observed photometric fluxes p_q , divided by their uncertainties Δp_q and the ‘model’ vector \mathbf{m} has elements $m_q = \langle f_\lambda \rangle_q/\Delta p_q$, the synthetic fluxes also divided by the data uncertainties.

Like Kriek & Conroy (2013), I did not use the catalogues’ formal photometric uncertainties in any of my fits, including when calculating κ and during PPXF fits. This is because the small error bars at longer wavelengths would have dominated the fits. Additionally, after many fits, it became apparent that systematic imperfections in the SPS model assumptions or data were the main source of uncertainty, rather than random noise. Instead, I use fixed linear uncertainties for all photometric bands of a given galaxy as explained in the next section.

6.5 Outliers removal

To remove outliers from the spectral fits, I follow a common practice (e.g. Westfall et al. 2019) that involves multiple PPXF fits and adjusting the uncertainties based on the fit residuals. My approach is designed for robustness as follows:

(i) I perform an initial PPXF fit assuming a reasonable fixed uncertainty $\Delta G^{\text{phot}}(\lambda_q)$ for all bands of 3 per cent of the maximum photometric flux $G^{\text{phot}}(\lambda_q)$ for that galaxy. Similarly, for the spectrum uncertainty $\Delta G^{\text{spec}}(\lambda_p)$, I adopt a constant value of 10 per cent of the median galaxy spectrum $G^{\text{spec}}(\lambda_p)$. A constant uncertainty is a good approximation for the VIMOS spectra and reduces the noise in the fit, with respect to adopting a more accurate but noisy error spectrum (e.g. as given by the reduction pipeline). The best fit is not sensitive to the scaling of uncertainties, which only affects the relative weight of the spectrum and photometry.

(ii) After the fit, I estimate the rms noise spectrum $\sigma_{\text{noise}}^{\text{spec}}$ per pixel from the fit residuals, in a statistically robust way, by computing for every spectral pixel the interval containing 68 per cent of the residuals, within a moving window of 100 pixels.

(iii) I mask the pixels deviating more than $3\sigma_{\text{noise}}^{\text{spec}}$ from the best fit. I repeat the masking in a loop, while iteratively adjusting the normalization of the best-fitting spectrum using the non-masked pixels from equation (34), until the mask does not change anymore.

(iv) I multiply $\Delta G^{\text{spec}}(\lambda_p)$ by $\sqrt{\chi_{\text{spec}}^2/P}$, where P is the number of non-masked spectral pixels, and I compute χ_{spec}^2 from the spectrum alone, in such a way that, after rescaling of the uncertainties, the resulting $\chi_{\text{spec}}^2 = P$.

(v) I do the same constant rescaling for the photometric uncertainties to enforce $\chi_{\text{phot}}^2 = Q$. Here, Q is the number of fitted photometric bands, and I compute χ_{phot}^2 from the photometry alone.

(vi) After the spectral masking and the rescaling of the photometric $\Delta G^{\text{phot}}(\lambda_q)$ and spectral $\Delta G^{\text{spec}}(\lambda_p)$ uncertainties, I perform a second PPXF fit, from which I extract the final results.

The rescaled uncertainties are generally of the same order of magnitude as the formal ones provided by the pipelines. However, there can be significant relative differences between different photometric bands. I tested the full LEGA-C sample and found that none of the results in this paper depended on whether I used the formal or rescaled uncertainties. However, the approach I adopted significantly reduced the number of cases where, after visual inspection, the formal best fit did not match the data well because of unrealistically small formal uncertainties that overemphasized certain photometric bands.

6.6 Gas model and kinematic constraints

For galaxies without AGNs, gas emission could be approximately predicted based on the galaxy SFH and could be included in the models, with some extra assumptions, based on photoionization models like CLOUDY (Ferland et al. 1998, 2013). This feature is implemented in FSPS and can be useful when fitting photometry alone where the gas emission are poorly constrained by the data. However, in my case, I have many good-quality spectra in addition to photometry and I want to be able to fit the gas more accurately than a model could predict. For this reason, I fit the gas emission lines in a model-independent way with PPXF.

With PPXF one can fit many gas emission lines simultaneously to the stellar continuum. This is especially important when studying the stellar population of star-forming galaxies or AGNs, where key absorption lines like the Balmer series are filled by emission. However, when fitting gas lines in relatively low S/N spectra, it is essential to set constraints on the parameters of the gas lines, to prevent possible degenerate situations. An example of a situation to avoid is when the spectrum does not have gas emission and the Gaussian describing an emission line becomes so wide as to become degenerate with the shape of the stellar continuum.

This is one of the types of situations for which I designed the linearly constrained algorithm of Section 3.2. For my fits to the LEGA-C spectra, after some experimentation focusing on the few problematic fits, I found it sufficient to require the dispersion of the gas emission lines to be smaller than the stellar one $\sigma_{\text{gas}} < \sigma_*$ and in addition I required the gas and mean stellar velocities to satisfy $|V_{\text{gas}} - V_*| < 500 \text{ km s}^{-1}$. I enforced these requirements as linear constraints in PPXF (keyword `constr_kinem`). My strict constraints on the gas dispersion is not always verified in galaxies and I would not recommend it when one is interested in the gas kinematics. However, it appears to work well in eliminating spurious solutions for the stellar population alone from the present type of spectra.

The emission lines that I included in the PPXF fits are all the lines listed in Belfiore et al. (2019, table 1). In particular, those falling within the LEGA-C wavelength range for $0.6 < z < 1$ are the Balmer series bluer than $H\beta$, the $[\text{OII}]\lambda\lambda 3726, 29$, $[\text{NeIII}]\lambda\lambda 3868, 69$ and $[\text{OIII}]\lambda\lambda 4959, 5007$ doublets and the $\text{HeII}\lambda 4687$. I force the kinematics of all the gas lines to be the same and I additionally fix the $[\text{OIII}]$ doublet to the 1/3 ratio. I fit the Balmer series as a single gas template with decrement for Case B recombination, for temperature $T = 10^4 \text{ K}$ and electron density $n_e = 100 \text{ cm}^{-3}$ from Storey & Hummer (1995). I allow the gas templates to have their own Calzetti et al. (2000) attenuation curve. Fixing the intrinsic ratios of the Balmer series allows me to provide a better extrapolation of the gas filling the weakest (higher-order) absorption lines of the series, even when the S/N of the spectrum is not high enough to constrain them.

Note that, although I do not include theoretical gas emission predictions in the SPS models, I do include the contribution of the emission lines that are spectroscopically constrained in the photometry. In particular, when Balmer lines are present in the spectrum, the line fluxes of the Balmer series and, in particular, $H\alpha$, which is outside the LEGA-C wavelength range, are included in the photometric fit. However, I checked that this inclusion has a minimal effect on the final results.

6.7 Velocity dispersion matching

The LEGA-C data have an instrumental dispersion of $\sigma_{\text{inst}} \approx 36 \text{ km s}^{-1}$ (Section 4.1), ignoring possible variations within the rather small wavelength range. The MILES stellar templates used in both the FSPS and GALAXEV models were observed with an instrumental

resolution of $\Delta\lambda \approx 2.50 \text{ \AA}$ FWHM (Falc3n-Barroso et al. 2011), equivalent to $R = \lambda/\Delta\lambda \approx 1600$ at the typical wavelength $\lambda \approx 0.4 \text{ \mu m}$ covered by LEGA-C. This corresponds to an instrumental dispersion

$$\sigma_{\text{inst}} = \frac{c}{R \sqrt{4 \ln 4}} \approx 80 \text{ km s}^{-1}. \quad (35)$$

Ideally, I would like to use SPS models based on stars with higher resolution than the galaxy spectra. However, assuming that the instrumental line spread functions are approximately Gaussian, one can still use a template with higher instrumental dispersion $\sigma_{\text{inst, tem}}$ than that $\sigma_{\text{inst, gal}}$ of the observed galaxy spectrum as long as $(\sigma_{\text{inst, gal}}^2 + \sigma_*^2) > \sigma_{\text{inst, tem}}^2$, where σ_* is the real ‘astrophysical’ dispersion of the galaxy stars. After the PPXF fit one can compute the corrected stellar dispersion with the standard expressions (section 2.2, Cappellari 2017)

$$\sigma_{\text{diff}}^2 = \sigma_{\text{inst, gal}}^2 - \sigma_{\text{inst, tem}}^2 \quad (36a)$$

$$\sigma_*^2 = \sigma_{\text{PPXF}}^2 - \sigma_{\text{diff}}^2. \quad (36b)$$

As σ_{diff}^2 is in this case a negative quantity, I can model the dispersion of galaxies down to $\sigma_* \gtrsim |\sigma_{\text{diff}}^2|^{1/2} = 71 \text{ km s}^{-1}$. I compared my fitted dispersions σ_* with the values in the LEGA-C DR3 catalogue, which were measured with PPXF using higher-resolution synthetic templates as described in Bezanson et al. (2018). I found a good agreement assuming $|\sigma_{\text{diff}}^2|^{1/2} \approx 48 \text{ km s}^{-1}$, which suggests possible inaccuracies in the quoted relative instrumental dispersion of the galaxies and the templates. Regardless of the reason for this discrepancy, only 40 of the 3197 galaxies in the catalogue with measured dispersion have $\sigma_* < 48 \text{ km s}^{-1}$, likely due to measurement uncertainties. This implies that I can safely use the SPS based on MILES models to study the stellar population of LEGA-C galaxies.

7 RESULTS

In this section, I describe the results of my stellar population modelling with PPXF. I also compare masses from stellar population and galaxy dynamics. The key quantities used in this paper are given in Table 1.

7.1 Spectral fit examples

In this paper, I focus on galaxy observable trends rather than on comparisons with models of galaxy formation. For this reason, instead of converting the SFH recovered by PPXF into stellar masses formed in a given time interval, I will always show the fraction of bolometric luminosity contributed by each template, as a function of their age and metallicity $[M/H]$. More precisely, I integrate the luminosity from the template spectra only within the region $0.1 < \lambda < 3 \text{ \mu m}$ covered by the data. This is to avoid the possibility of interpreting very young stars, which emit most of their luminosity for $\lambda < 0.1 \text{ \mu m}$, as contributing significantly to my observables, even when their flux is not detected in the data, but simply extrapolated. I still indicate my luminosity as L_{bol} because, except for extremely young stars, it still represents a very good approximation for it.

The advantage of using L_{bol} rather than SFH, is that one can get a direct sense of what the data actually show, without strongly non-linear conversions into masses, due to the large M/L differences of different stellar populations. In fact, I would argue that comparisons with models of galaxy formations are generally more meaningful when the models, for which all quantities are known accurately, are converted into luminous observables, rather than trying to do the reverse by extracting SFH in masses from the data.

Table 1. JAM dynamical masses and PPF stellar population results using SPS templates from FSPS.

ID LEGA-C	RA	DEC	$\lg M_{\text{JAM}}$	$\lg \Sigma_{\text{I}}^{\text{JAM}}$	$\lg M_{*}^{\text{PPXF}}$	$\langle \lg \text{Age} \rangle$	$\langle [M/H] \rangle$	A_V	δ
(1)	($^{\circ}$)	($^{\circ}$)	(M_{\odot})	($M_{\odot} \text{ kpc}^{-2}$)	(M_{\odot})	(yr)	(mag)	(mag)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	150.415222	1.758146	11.560	10.028	11.348	8.982	-0.271	1.122	0.111
4	150.412292	1.777617	11.405	9.548	11.350	9.107	-1.040	1.356	0.377
5	150.381104	1.780024	10.847	9.436	10.947	8.719	-0.475	0.304	-0.075
6	150.449783	1.783690	11.119	9.981	11.189	9.538	-0.140	0.000	0.400
8	150.378815	1.792945	11.653	9.919	11.445	9.088	-0.279	1.711	0.400
10	150.407440	1.803214	11.695	10.587	11.698	9.737	-0.094	0.097	0.239
11	150.423050	1.807328	11.601	9.649	11.579	9.032	-0.102	0.075	-1.000
12	150.405838	1.813150	10.817	9.137	11.082	8.554	-0.572	0.579	-0.165
13	150.385376	1.816353	11.113	10.036	11.190	9.285	-0.036	0.000	0.400
14	150.430817	1.820639	11.217	9.246	11.217	8.159	-0.740	0.770	-0.038

Columns (1), (2), and (3): ID, right ascension and declination J2000 in degrees from the LEGA-C catalogue of van der Wel et al. (2021). Column (4): JAM dynamical masses from Section 5.2 in solar masses. For accurate quantitative use, one should divide these masses by the luminosities of the Sersic models in the LEGA-C catalogues to obtain the total $(M/L)_{\text{JAM}}$ as described in Section 7.3; Column (5): dynamically-determined average mass density within a cylinder of radius $R = 1$ kpc along the line of sight. I computed this from the best-fitting JAM model; Column (6): Stellar masses from PPF using SPS templates from FSPS. These masses include living stars and stellar remnants, but exclude gas lost during stellar evolution. I assume a Salpeter IMF with a lower/upper mass cutoff of 0.08 and 100 M_{\odot} respectively; Columns (7) and (8): \lg Age and $[M/H]$ weighted by the bolometric luminosity (Section 7.4). Columns (9) and (10): V -band attenuation in mag and slope δ from equation (23), for the two-parameters attenuation described in Section 6.3. I show only the first ten rows of this table, while the full electronic table for 3197 galaxies (including duplicates) is available as Supporting Information from the MNRAS website.

In the course of this study, I fitted the 3197 galaxies of my subsample (Section 4.1) with PPF multiple times with different levels of regularization, or no regularization at all, to test the sensitivity of the results. In Fig. 5, I illustrate the effect of regularization on some high- S/N spectra. These figures, like Fig. 1, illustrate the ill-conditioning of the stellar population inversion, which prevents one from obtaining a unique solution, even from very good data. None the less, the figure also illustrates the ability of the method to distinguish the striking difference between (i) galaxies that can only be described, even at high regularization¹⁵ (`regul=100`) by a single star formation event at a very localized \lg Age (top three panels in Fig. 5) and (ii) galaxies that require multiple and separated star formation events to be described (bottom three panels in Fig. 5). The galaxies in the top panels are essentially described by a single SPS model, from 0.1 to 3 μm , for both spectra and photometry. This highlights the success of the SPS models in accurately predicting real galaxy spectra.

In Fig. 6, I show additional examples of PPF fits to good-quality spectra to give a sense of the variety of spectral morphologies and the corresponding variations in the L_{bol} weights distributions. I used in all these cases a high regularization (`regul=100`). Also, here, one can clearly see the striking difference between (i) the three galaxies in the top row, which can only be described as a single burst of star formation, which happened at different times and (ii) galaxies requiring multiple discrete star formation events. Star formation events appear to have a similar extent in \ln Age, which seems to imply that events in the past lasted longer than recent ones. This is likely an artefact of our general ability to more accurately detect age differences in recent events.

¹⁵A given value of the PPF keyword `regul` roughly implies that neighbouring weights w_{ij} can differ by $\Delta w_{ij} \sim 1/\text{regul}$. As I normalize all galaxy spectra to the same average flux (e.g. `average=1`), setting a given `regul` value roughly corresponds to requiring a similar level of smoothness in the distribution of the weight.

7.2 Comparing PPF stellar masses with other methods

The weights for different ages and metallicities produced by a fit with PPF can be converted into stellar masses. In this section, I compare the masses derived with PPF against the stellar masses produced by other codes.

For my comparisons, I used the published stellar masses M_{*} for the LEGA-C galaxies from the three stellar population codes: (i) LEPHARE (Arnouts et al. 2002), (ii) EAZY (Brammer et al. 2008), and (iii) PROSPECTOR (Johnson et al. 2021b). I extracted the values of stellar masses for the first two codes from the FARMER version of the COSMOS2020 catalogue (Weaver et al. 2022), while for the last code, I used the values described in the LEGA-C DR3 paper (appendix B of van der Wel et al. 2021) as kindly provided by Arjen van der Wel.

I tried to isolate the effect of the fitting methods from differences in the extrapolation of the galaxy's total luminosities. For this, with the M_{*} from the COSMOS2020 catalogues, I rescaled the masses by the difference in the total K_s band luminosity between the Weaver et al. (2022) and Muzzin et al. (2013) catalogues. No correction is needed to compare PPF and PROSPECTOR, given that for both I used the values based on the Muzzin et al. (2013) catalogue.

I show the comparison between the stellar masses derived with PPF and the other three codes in Fig. 7. I coloured the values with the stellar age derived by PPF. I LOESS smoothed the measured age values using the algorithm by Cleveland & Devlin (1988) as implemented in the LOESS package¹⁶ by Cappellari et al. (2013b) and using the keyword `rescale=True` to equalize the axes of maximum/minimum variance before smoothing. I used a small smoothing parameter `frac=0.1` in all plots of this paper. The LOESS-smoothed values are the two-dimensional equivalent of the average trend that is often shown in one-dimensional plots. The key difference is that the scatter cannot be easily shown in two dimensions

¹⁶I used LOESS v2.1 available from <https://pypi.org/project/loess/>

together with the average trend. The scatter is better visualized using a different projection.

To estimate the scatter between two pair of measurements, while removing outliers, I used the `LTSFIT` package¹⁷ described in Cappellari et al. (2013a, section 3.2), which combines the Least Trimmed Squares robust technique of Rousseeuw & Van Driessen (2006) into a least-squares fitting algorithm, which allows for errors in all variables and intrinsic scatter. Instead of using a fixed σ -clipping criterion with the ‘`clip`’ keyword in the `ltsfit` procedure, I used an adaptive clipping that depends on the sample size. This is the value that would produce on average one outlier in a Gaussian distribution of the given sample size. It can be computed using the `scipy` class `scipy.stats.norm` as `clip=abs(norm.ppf(p/2))`, with $p = 1/n$ and n the sample size. For reference, with $n = 100$, this gives `clip=2.58` (default for `ltsfit`), for $n = 500$, `clip=3.09` and for $n = 3000$, `clip=3.59`. The `ltsfit` procedure returns a robust estimate of the rms scatter Δ from the best-fitting relation. When the uncertainty of the two quantities I am comparing is the same, one can estimate it as $\sigma_{M_*} = \Delta/\sqrt{2}$.

In all my plots, I rescaled the masses provided by all other methods to have the same median as the PPXF values, which I did not modify. This is the reason why all plots follow the one-to-one relation without any overall offset. This is to remove the effect of differences in the assumed stellar IMF, gas loss or stellar remnants, whose investigation is outside the scope of this paper. I find that the observed scatter in all galaxies, when selected irrespective of their age or S/N , is in agreement with a 1σ uncertainty in the stellar mass of about 30 per cent for every method. This result is consistent across all six pairwise comparisons of the methods, with differences within the measurement uncertainties. However, the behaviour of the differences is markedly different as a function of mean ages. The comparison of PPXF against PROSPECTOR (Fig. 7), show that the scatter is smaller for older galaxies at given mass, but the younger ones generally scatter symmetrically around the one-to-one relation. The exception are the outliers, which have generally lower masses in PROSPECTOR than in PPXF. The comparison of PPXF and LEPHARE is similar to PROSPECTOR, but with less low-mass outliers. However, there is no evidence for a tightening of the correlation for older models. The comparison of PPXF and EAZY, unlike the other two models, shows a strong asymmetry as a function of age: older models tend to be less massive in EAZY than PPXF, while younger models are more massive in EAZY. This asymmetry is reminiscent of the difference between PROSPECTOR and EAZY reported in Leja et al. (2019). In fact, the same age asymmetry is seen when comparing EAZY with either PROSPECTOR or LEPHARE.

As suggested by Fig. 7, the scatter dramatically decreases (Fig. 8) if I compare PPXF and PROSPECTOR only for the galaxies with the oldest ages and largest spectral S/N (which also implies brightest photometry). Given the small age range, I coloured galaxies by their σ_* . For this subset of galaxies, the inferred scatter of about 16 per cent is half of that for the general population, without significant trends, except again for some outliers where PROSPECTOR gives lower masses than PPXF.

When comparing stellar mass estimates of real galaxies, it is often difficult to assess the real accuracy between different methods, because the true masses are unknown. In the next section, I will address this issue, for a subsample of the LEGA-C sample, using mass determinations from stellar dynamics.

7.3 Comparing JAM dynamical with stellar population M/L

One of the sources of confusion in comparing galaxy masses from stellar populations and dynamical modelling is the ambiguity of the so-called ‘dynamical mass’ of a galaxy. This term does not refer to a well-defined physical quantity, because in the standard cosmological model, the galaxy’s total mass is largely composed of dark matter, which is difficult to constrain with the available kinematic data of limited radial coverage. The quantity that the dynamical models reliably measure is the total density profile within the spatial extent of the kinematic tracer. However, this density profile cannot be easily converted into a mass, because it depends on the choice of the integration volume and on the assumptions about the galaxy shape and orientation (section 3.3.1 of Cappellari et al. 2013a). A more robust and convenient quantity for comparing population and dynamics is the total mass-to-light ratio $(M/L)_{\text{JAM}}$, within the inner regions of a galaxy. This quantity has a weak dependence on the integration volume and galaxy inclination. It should be always preferred for accurate comparisons.

In Section 5, I presented unbiased dynamical models of the stellar kinematics, based on the Sersic photometric models in the F814W/ACS band, for all the galaxies in the LEGA-C sample with available velocity dispersion. Previous studies of nearby galaxies using high-resolution integral-field stellar kinematics have demonstrated that this kind of models can reliably estimate the total dynamical mass-to-light ratios $(M/L)_{\text{JAM}}$ in the central regions of galaxies with uncertainties of about 5 per cent (Cappellari et al. 2006, 2013a; Shetty et al. 2020b; Zhu et al. 2023a). Importantly, these studies have also shown that precise and unbiased $(M/L)_{\text{JAM}}$ can be obtained using models where the total mass distribution follows the luminous one. In fact, these mass-follow-light models are more robust and precise than those that explicitly separate the luminous and dark matter, when the main goal is to measure the total $(M/L)_{\text{JAM}}$ (Cappellari et al. 2013a; Zhu et al. 2023b).

An additional complication is that the dynamics is sensitive to all mass components: stellar, gas, and dark matter, while the population only measures the stellar one. However, detailed nearby studies have shown that for the passive galaxy population, dark matter contributes only about ≈ 10 per cent of the total mass within $1R_e$ (Cappellari et al. 2013a; Zhu et al. 2023b), while gas mass has an even smaller contribution (e.g. Young et al. 2011). This allows us to assume that the dynamical $(M/L)_{\text{JAM}}$ accurately approximates the stellar one.

To compare the dynamical $(M/L)_{\text{JAM}}$ I need the same quantity from stellar population $(M/L)_{\text{pop}}$. Having the full spectra from the stellar population models, one can compute the $(M/L)_{\text{pop}}$ in any band. However, since I only have total stellar masses M_* from PROSPECTOR LEGA-C catalogue, I divide M_* by the total luminosity L_i in SUBARU i -filter from Muzzin et al. (2013) catalogues. This assumes that $(M/L)_{\text{pop}}$ is constant over the full galaxy, which is likely a decent approximation for passive galaxies. Using photometry consistent with mass derivation ensures no spurious differences in mass and luminosity extrapolation. However, differences between i -band and F814W filters may introduce some small systematic offset in the M/L comparison. However, I am interested in relative uncertainties more than absolute offsets. The M/L is usually reported in solar units. For this, I assume a solar luminosity $L_g = 5.11$ mag in AB system from Willmer (2018) and report M/L in units of M_\odot/L_\odot , given that F814W approximately corresponds to rest-frame SDSS g -band filter at median redshift $z \approx 0.8$ of my sample. This normalization is a constant and does not affect the comparison. The rest-frame wavelength of the filter varies by up to ≈ 10 per cent within the redshift range; however, this shift is the

¹⁷I used `LTSFIT` v6.0 available from <https://pypi.org/project/ltsfit/>

same for both dynamical and population M/L and does not affect the scatter.

I compare the $(M/L)_{\text{JAM}}$ from dynamical modelling and the $(M/L)_{\text{pop}}$ from SPS using a sample of old galaxies with high-quality spectra in Fig. 9. I only use the passive population for this comparison. I adjust the PROSPECTOR $(M/L)_{\text{pop}}$ by adding 0.19 dex to match the JAM median. I also convert the PPXF $(M/L)_{\text{pop}}$ from the Salpeter (1955) IMF to the Chabrier (2003) IMF by subtracting 0.215 dex (fig. 4 of Madau & Dickinson 2014). I do not change the PPXF value after this conversion.

The main findings from Fig. 9 are:

(i) The PPXF $(M/L)_{\text{pop}}$ values are more consistent with the $(M/L)_{\text{JAM}}$ values than the PROSPECTOR ones. The scatter is 0.090 dex for PPXF and 0.117 dex for PROSPECTOR. This suggests that adding spectra to PPXF improves the mass estimates.

(ii) The PPXF $(M/L)_{\text{pop}}$ comparison does not have the low- M/L outliers that appear in the PROSPECTOR comparison, indicating more reliable $(M/L)_{\text{pop}}$ or M_* estimates in PPXF with spectra than in PROSPECTOR with photometry only.

(iii) Both PPXF and PROSPECTOR show a similar trend in the $(M/L)_{\text{pop}} - (M/L)_{\text{JAM}}$ relation, which clearly deviates from a one-to-one relation. The trend implies that the galaxies with higher σ_* have more mass from dynamics than from population models at a fixed IMF. The variation is comparable to the mass difference between Chabrier and Salpeter IMF. This trend is consistent with previous studies that suggested a non-universal IMF based on dynamics and population of nearby (Cappellari et al. 2012; Li et al. 2017; Shetty et al. 2020a) and distant galaxies (Shetty & Cappellari 2014). Whatever the origin of this trend, this comparison shows that it is robust across different samples, redshift, and methods.

From the cross-comparisons between the scatter observed when comparing different estimates of the stellar masses, one can infer the accuracy of each individual technique, assuming as an approximation that it is constant. In fact, if we define σ_{method} the uncertainty of ‘method’, then the squared uncertainties between each pair of methods add linearly as follows

$$\begin{cases} \sigma_{\text{PPXF}}^2 + \sigma_{\text{Pros}}^2 = \Delta^2(\text{PPXF/Pros}) \\ \sigma_{\text{PPXF}}^2 + \sigma_{\text{JAM}}^2 = \Delta^2(\text{PPXF/JAM}) \\ \sigma_{\text{Pros}}^2 + \sigma_{\text{JAM}}^2 = \Delta^2(\text{Pros/JAM}) \end{cases} \quad (37)$$

where the scatter Δ was measured in Fig. 8 ($\Delta(\text{PPXF/Pros}) = 0.090$ dex) and Fig. 9 ($\Delta(\text{PPXF/JAM}) = 0.090$ and $\Delta(\text{Pros/JAM}) = 0.117$ dex). The positive solution of equation (37) gives the 1σ relative uncertainty of the three different methods on this data set:

$$\begin{cases} \sigma_{\text{PPXF}} = 0.035 \text{ dex} = 9 \text{ per cent} \\ \sigma_{\text{JAM}} = 0.083 \text{ dex} = 21 \text{ per cent} \\ \sigma_{\text{Pros}} = 0.083 \text{ dex} = 21 \text{ per cent} \end{cases} \quad (38)$$

This result shows that, at least for the limited case of the old population, where we can assume we know the ‘true’ stellar mass from galaxy dynamics, the inclusion of spectra in PPXF gives masses significantly more accurate than those using PROSPECTOR with photometry alone. This is encouraging, but of course, it should not be interpreted as PPXF being more accurate than PROSPECTOR, given that the latter could fit spectra as well and this would likely lead to comparable accuracy as PPXF. However, these extra comparisons are beyond the scope of this paper.

In Fig. 10, I also show the comparison between stellar dynamics and stellar population M/L for the full set of galaxies with high- S/N regardless of their age. This plot cannot be used to infer the accuracy

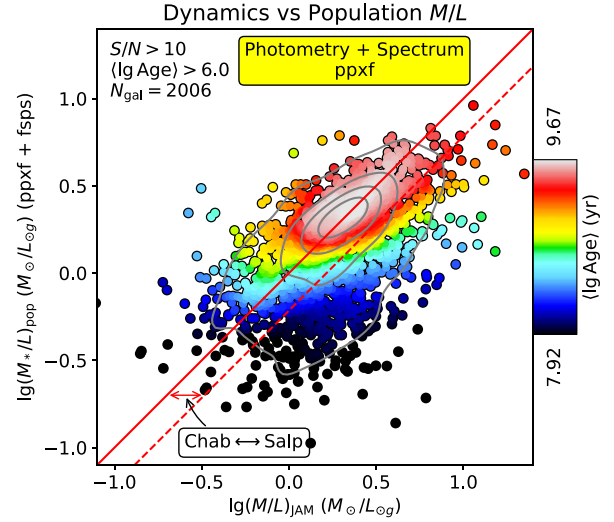


Figure 10. Same as in Fig. 9 but for all ages. The colours here indicate LOESS-smoothed mean luminosity-weighted ages (\lg Age). The contours of constant age are almost horizontal, meaning that the stellar population mass-to-light ratio $(M/L)_{\text{pop}}$ depends on age as expected, but the dynamical $(M/L)_{\text{JAM}}$ does not. Variations in the IMF can explain the $(M/L)_{\text{JAM}}$ differences for the old ages, while the mass fraction of dark matter and gas must drive the $(M/L)_{\text{JAM}}$ variations for the young ages.

of the mass estimates. In fact, detailed modelling of the MaNGA survey has shown that younger galaxies contain significant fractions of gas and dark matter (Zhu et al. 2023a), making the mass estimate from the stellar population significantly lower than the dynamical one, as observed.

7.4 Stellar population scaling relations

As I am focusing on observable trends, I define luminosity-weighted population quantities, summed over the template weights, as

$$\langle \lg \text{Age} \rangle = \frac{\sum_j L_{\text{bol},j} \times \lg \text{Age}_j}{\sum_j L_{\text{bol},j}} \quad (39a)$$

$$\langle [M/H] \rangle = \frac{\sum_j L_{\text{bol},j} \times [M/H]_j}{\sum_j L_{\text{bol},j}}. \quad (39b)$$

In Fig. 11, I show the distribution of ages, metallicities, and the Sersic (1968) index n_{Ser} on the $(M_{\text{JAM}}, R_e^{\text{maj}})$ plane, where the dynamical mass M_{JAM} closely approximates the total stellar masses M_* , and the half-light radius R_e^{maj} is the semimajor axis of the isophote containing half of the total light of the Sersic (1968) fitted to the galaxy photometry. Both R_e^{maj} and n_{Ser} are tabulated in the LEGA-C catalogue (van der Wel et al. 2021). See Cappellari et al. (2013a) for a discussion of why dynamical mass approximates the total stellar mass.

I LOESS smoothed the measured values in all coloured plots of this paper. The resulting LOESS-smoothed values represent the two-dimensional equivalent of the average values that are routinely shown in one-dimensional plots. However, in two-dimension, one cannot show the scatter together with the average values. To visualize the scatter, which is significant and not random, I use a projection along the σ_* axes later.

This figure shows the well-known fact, in the nearby Universe, that both ages and metallicities approximately follow lines of constant stellar velocity dispersion σ_* , or equivalently lines where $M_* \propto R_e^{\text{maj}}$ (compare this figure with MaNGA results in fig. 4 of Li et al. 2018

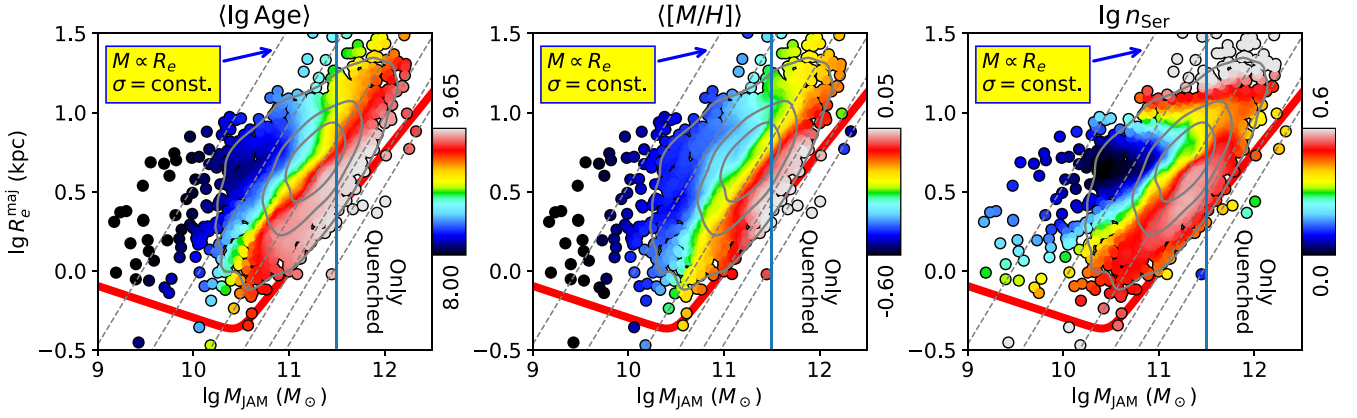


Figure 11. Galaxy properties on the dynamical mass versus size ($M_{\text{JAM}}, R_e^{\text{maj}}$) plane at $z \approx 0.8$. To good approximation $M_{\text{JAM}} \approx M_*$, see text. Galaxies are coloured by their luminosity-weighted ages ($\langle \text{lg Age} \rangle$) (left-hand panel), luminosity-weighted metallicities ($\langle [M/H] \rangle$) (middle panel) and by the Sersic index n_{Ser} of a fit to their photometry. I used the values from the FSPS models, but those from GALAXEV are very similar. All values are LOESS-smoothed to show average trends and I used `regul=1.0` in PPXF. The grey contours are a kernel density estimate of the galaxy distribution (using `scipy.stats.gaussian_kde`). Galaxy properties mainly follow lines of constant stellar velocity dispersion, which is indicated by the dashed grey lines for $\sigma_* = 50, 100, 200, 300, 400$, and 500 km s^{-1} from left to right, while mass is not a good predictor of their stellar population. However, above a stellar mass $\text{lg}(M_*/M_\odot) \gtrsim 11.5$ (blue vertical line), all galaxies are old (quenched), have a high metallicity and large Sersic index. The thick red line is the ‘zone of avoidance’ for nearby galaxies from Cappellari et al. (2013b), scaled down by a factor $1.6 \times$ to account for redshift evolution.

or Lu et al. 2023). I also show to guide the eye the local ‘zone of avoidance’ at high densities (Cappellari et al. 2013b, eq. 4), which I scaled down by a factor $1.6 \times$ in R_e^{maj} , roughly consistent with the general trends of decreasing galaxy sizes with redshift (e.g. van der Wel et al. 2014). The Sersic index n_{Ser} also approximately follows the distribution of ages and metallicity, in the sense that passive galaxies tend to have $\ln n_{\text{Ser}} \gtrsim 0.4$ or $n_{\text{Ser}} \gtrsim 2.5$ (red colour in the right-hand panel of Fig. 11). This n_{Ser} value is the one sometimes adopted to separate early-type from late-type galaxies (e.g. Bell et al. 2003; Shen et al. 2003). In the local Universe, below the stellar mass $\text{lg}(M_*/M_\odot) \lesssim 11.5$ the trend of n_{Ser} is due to a sequence of increasing bulge fraction, while above $\text{lg}(M_*/M_\odot) \gtrsim 11.5$ is the region of slow rotators with cores (see review by Cappellari 2016, fig. 23).

This result has a rather long history, both locally and at $z \sim 1$ (Chauke et al. 2018, 2019; Beverage et al. 2021; Barone et al. 2022; Hamadouche et al. 2022; Tacchella et al. 2022) but had not been seen so cleanly at this redshift before LEGA-C. For nearby galaxies, Kauffmann et al. (2003) clearly noted that galaxy population correlates better with mass surface density Σ than with M_* . It was later observed that Σ , or even better the virial predictor $\sigma_{\text{vir}} \propto M_*/R_e$ of the stellar velocity dispersion, inferred from photometry alone, remains a better predictor of galaxy colours out to $z \approx 3$ (Franx et al. 2008; Bell et al. 2012). However, it was still unclear at that time how accurately the photometric estimates were able to predict the actual stellar masses and the velocity dispersion of the stars. To address this issue, I used masses from dynamical models and σ_* from good-quality integral-field stellar kinematics rather than photometric estimates. In Cappellari (2011) I clearly concluded that ‘ σ_* (not Σ_e or M_*) is the best predictor of galaxy properties’ (see also Cappellari et al. 2013b). These early results were confirmed by several papers using larger samples and stellar kinematics of ever-increasing quality (e.g. Wake, van Dokkum & Franx 2012; McDermid et al. 2015; Scott et al. 2017; Li et al. 2018; Barone et al. 2018, 2020). In parallel, Cheung et al. (2012) and Fang et al. (2013) introduced the use of central surface density Σ_1 from photometry, within a fixed radius of 1 kpc, to predict quenching. A review is given in Cappellari (2016, see fig. 22).

Given that in Fig. 11 the main stellar population trends follow σ_* , in Fig. 12, I show how the luminosity-weighted ages and metallicity

depend on σ_* in the LEGA-C sample. The trends resemble quite closely the local results from the best integral-field spectroscopy from both SAMI (Scott et al. 2017) and MaNGA (Li et al. 2018). However, the top panels of Fig. 12 additionally illustrate the clear dependency between age and $[M/H]$ at fixed σ_* : the population of old galaxies at large σ_* is characterized by a larger metallicity than their younger counterpart at the same σ_* . Very clear is the bend in the (σ_*, Age) distribution around $\text{lg}(\sigma_*/\text{km s}^{-1}) \approx 2.3$ (also see Chauke et al. 2018). The results are very consistent between both the FSPS and GALAXEV SPS models. It is reassuring to see that the ridge of the age distribution in the top panels converges towards the age of the Universe at that redshift (grey horizontal band), while being slightly younger for GALAXEV vs FSPS. I also run models where I restricted the age of each galaxy to the Universe’s age at its redshift, as generally done for local studies. All results were qualitatively similar, except for the obvious truncation and corresponding clustering of the Ages values at the maximum Universe Age ≈ 6.6 Gyr at $z \approx 0.8$, which is indicated by a grey band in Fig. 12. The BPASS results are qualitatively in agreement but show substantial quantitative differences, especially in the $\sigma_* - [M/H]$ trend. Overall, this figure confirms the quality and consistency of these global results compared to local surveys.

Fig. 13 shows the distribution of the two dust attenuation parameters A_V and δ (Section 6.3) coloured by mean stellar age. One can see that at every UV slope δ the youngest galaxies have the strongest attenuation, except for the largest δ . Moreover, the largest attenuations in galaxies are only observed at large δ , close to the Calzetti et al. (2000) slope $\delta = 0$. Note, however, that there is a degeneracy between attenuation and continuum normalization near the upper limit of δ . Results are extremely consistent for the FSPS and GALAXEV SPS models, but again the BPASS results look quite different, although they all qualitatively agree.

7.5 Non-parametric star formation histories

Fig. 14 shows the non-parametric SFH of the galaxies in the LEGA-C sample as a function of key galaxy parameters. For this plot, I sorted the quantity of interest (e.g. σ_*) and constructed 30 bins in that quantity, each containing the same number of about 100 galaxies,

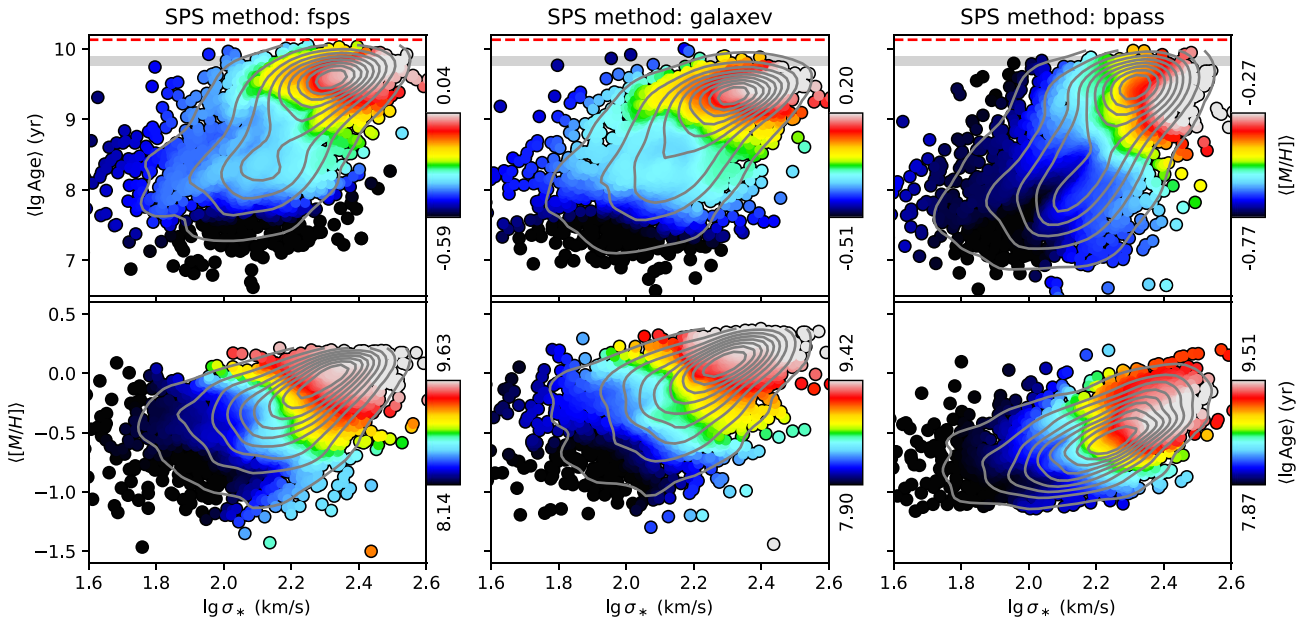


Figure 12. Luminosity-weighted ages ($\lg \text{Age}$) and metallicities ($[M/H]$) versus stellar velocity dispersion σ_* for the LEGA-C galaxies. The top panels show the age distribution, coloured by LOESS-smoothed metallicity. The grey horizontal band is the Universe age range between $0.6 < z < 1$ and the red dashed line is the present age. The bottom panels show the metallicity distribution coloured by LOESS-smoothed age. In all panels, the grey contours are the kernel density estimator of the galaxies' distribution. From left to right, I show results using the FSPS, GALAXEV and BPASS SPS models, all with `regul=10` in PPXF. There is a clear bend of the (Age, σ_*) trend around $\lg(\sigma_*/\text{km s}^{-1}) \approx 2.3$. At fixed σ_* metallicity depends on age with younger galaxies having lower $[M/H]$.

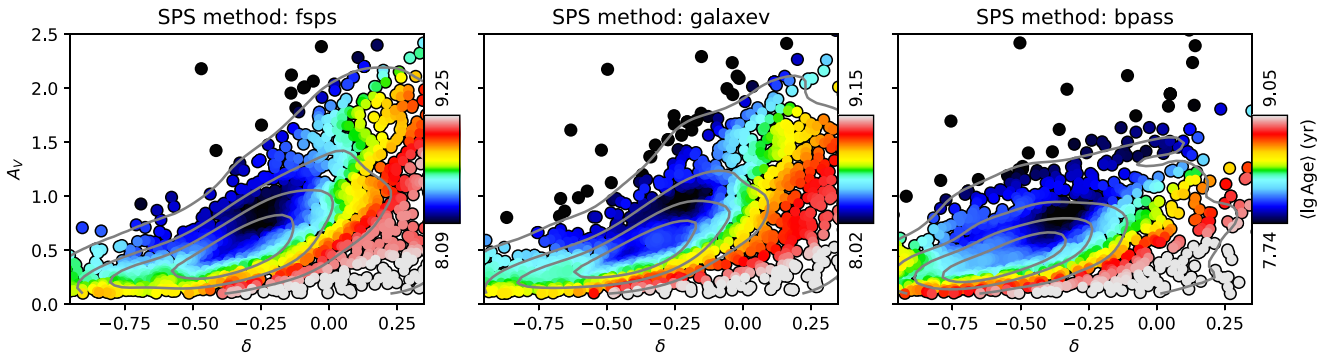


Figure 13. Dust properties vs galaxy age. The V band attenuation A_V in mag is plotted against the UV slope δ and coloured by LOESS-smoothed luminosity-weighted ages. The grey contours are the kernel density estimator of the galaxies' distribution. At every δ the youngest galaxies have larger attenuation, while the largest A_V are only observed around the Calzetti et al. (2000) attenuation curve, which has $\delta = 0$. From left to right, I show results using the FSPS, GALAXEV, and BPASS SPS models. The first two are highly consistent, but the latter, while still qualitatively similar is quantitatively very different.

in such a way that different bins have the same level of shot noise. I show the dependency of the SFH, parametrized as discussed by the light L_{bol} contributed in the spectrum by stellar populations of different ages, as a function of the following parameters:

(i) **Stellar velocity dispersion:** the plots show a clear trend of SFH with σ_* as expected from the trends between σ_* and age. What is new is the striking sharpness of the boundary between a regime $\lg(\sigma_*/\text{km s}^{-1}) \gtrsim 2.3$ (or $\sigma_* \gtrsim 200 \text{ km s}^{-1}$), above which the spectra are dominated by a population nearly as old as the Universe at that redshift, without evidence for subsequent star formation events, and below which suddenly galaxies have star formation at any time until the present time. Both the FSPS and GALAXEV SPS models indicate that galaxies still form the bulk of their stars at old times, but this age increases with σ_* by roughly a factor between 6 and 10 for a

variation in σ_* by a factor of 10. In the case of the GALAXEV models, the SFH indicate ongoing star formation at the lowest σ_* bins, while this is less so for the FSPS models.

(ii) **Density within 1 kpc:** this panel shows the same trend as the previous one, but with different units. This is because Σ_1^{JAM} is closely related to σ_* (see Fang et al. 2013), especially when using JAM dynamical masses instead of stellar population masses.

(iii) **Galaxy mass:** contrary to the dependency of SFH with σ_* , there is no sharp transition as a function of stellar mass, but rather a gradual trend. Only galaxies more massive than $\lg(M_*/M_\odot) \gtrsim 11.5$ are characterized by a single event of star formation at old times.

(iv) **Galaxy metallicity:** this panel shows that $[M/H]$ is as good as σ_* at predicting the boundary between the region of fully quenched galaxies and those that can have multiple star formation events. Here, it happens at $[M/H] \approx -0.2$ for the FSPS and $[M/H] \approx 0.0$

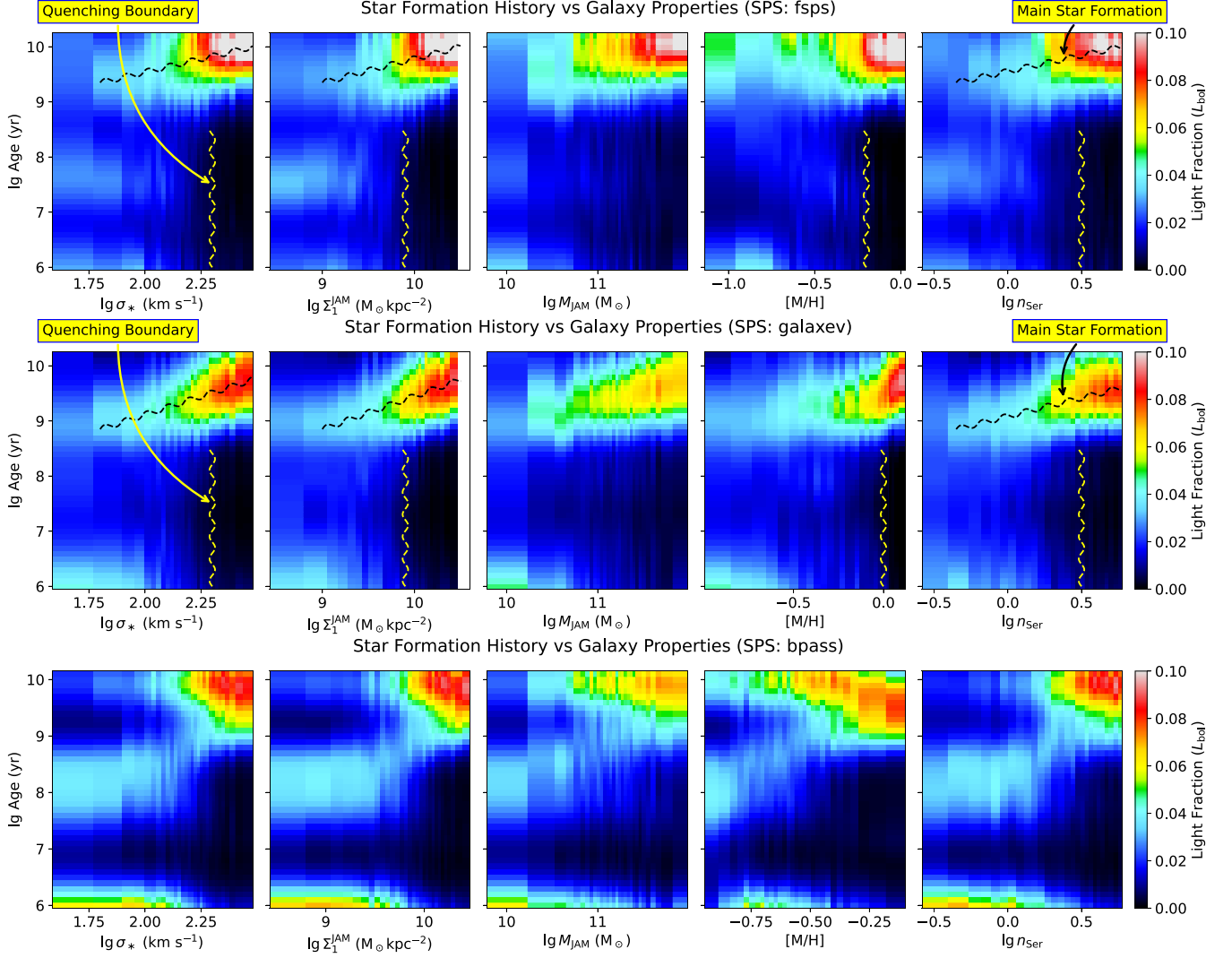


Figure 14. SFH versus galaxy properties. In all panels, the colours represent the SFH recovered with PPF (with `regul=100`) parametrized by the bolometric light L_{bol} fraction contributed by populations of different ages. The SFHs are shown as a function of key galaxy parameters: (1) the stellar velocity dispersion σ_* ; (2) the dynamical mass M_{JAM} , which well approximates M_* ; (3) the dynamically determined average density Σ_1^{JAM} inside a circle of radius 1 kpc centred on the galaxy; (4) the luminosity weighted metallicity ($[M/H]$); and (5) the exponent n_{Ser} of a Sersic profile fitted to the galaxy photometry (see text for definitions). Average values are computed for equal bins in σ_* (or the other parameters) of 100 galaxies each. Galaxies with largest σ_* , M_* , $[M/H]$ or n_{Ser} on average experienced their main star formation event long ago, but the typical age for the bulk of their star formation increases with σ_* (or the other parameters) as indicated by the slanted black dashed wavy lines in the left and right panels. For lower σ_* galaxies can form the stars at any time until the present. The plots show a beautifully clear and sharp quenching boundary at $\lg(\sigma_*/\text{km s}^{-1}) \approx 2.3$, or $\lg(\Sigma_1^{\text{JAM}}/\text{M}_\odot \text{kpc}^{-2}) \approx 9.9$, or $[M/H] \approx -0.2$ for FSPS and $[M/H] \approx 0.0$ for GALAXEV, or $\lg n_{\text{Ser}} \approx 0.5$, as indicated by the vertical yellow dashed wavy lines. There is no sharp boundary as a function of galaxy mass, but the transition is gradual and roughly happens around $\lg(M_*/M_\odot) \approx 11.5$. Note the generally good agreement between the results from the FSPS and GALAXEV SPS. The results using the BPASS models look problematic, with spurious structures at specific ages, which are most likely artefacts of the models.

for the GALAXEV, which are systematically shifted to larger values of metallicity.

(v) **Sersic index:** this panel show the SFH as a function of the Sersic (1968) exponent n_{Ser} . The boundary between fully quenched galaxies and galaxies that can have multiple star formation events happens here at $\lg n_{\text{Ser}} \approx 0.5$ or $n_{\text{Ser}} \approx 3.2$. Remarkably this boundary is here nearly as clean as that with σ_* .

In all panels, the results using the BPASS SPS models are again quite different from the other two. They show significant structure at specific ages and a less clear quenching boundary. The structure seen using the BPASS models is likely an artefact of the SPS rather than a real conspiracy in the star formation events.

The trends for different galaxy parameters and the overall consistency between the four panels, for both the FSPS and GALAXEV SPS models, can be understood by looking at Fig. 11 and noting that there is a region $\lg(M_*/M_\odot) \gtrsim 11.5$ above which all galaxies are quenched, high metallicity and have large Sersic index. Below that mass galaxies follow a trend of increasing bulge fraction, which increases σ_* , metallicity and makes galaxies more likely to quench. These results parallel those which have been extensively reported for local galaxies (see review by Cappellari 2016). What is new here is the clarity and sharpness of the empirical evidence of the boundary to quenching and the fact that this can be detected so well at a time when the Universe was half of its current age.

A rapid cessation of star formation for galaxies above a given critical value of σ_* , or of some other estimate of the central stellar density, varying with z , has often been invoked to explain the evolution of galaxy parameters over time (e.g. van Dokkum et al. 2015). An excellent review of the empirical evidence and models of a quenching boundary in galaxies is given in Chen et al. (2020, section 1). The physical mechanism for quenching is still under debate. Very briefly, one can group the main proposed theories into three broad classes: (i) ‘halo quenching’, where the gas gets shock-heated when falling into the gravitational potential of massive dark halos (e.g. Dekel & Birnboim 2006); (ii) ‘AGN feedback’, where a jet from the supermassive black hole either ejects the gas from its host galaxy (e.g. Silk & Rees 1998) or prevent it from infalling (e.g. Bower et al. 2006; Croton et al. 2006). See review by Somerville & Davé (2015). The panels in Fig. 14 provide a beautiful empirical confirmation of the theoretical assumptions that are made in many of those models.

7.6 Non-parametric joint SFH and metallicity distributions

Fig. 15 presents the non-parametric joint luminosity distribution of the age and metallicities of the stellar populations of galaxies in four different bins of σ_* . As in Fig. 14, also for this figure I sorted galaxies as a function of their σ_* and constructed four groups, of about 800 galaxies each, to ensure all panels have the same level of shot noise. Like before, I compare all three SPS models (FSPS, GALAXEV and BPASS). In addition, in the second row of Fig. 15 I show the result when using the FSPS model but adopting the photometric measurements from COSMOS2020 (Weaver et al. 2022) instead of the UltraVISTA catalogue (Muzzin et al. 2013). The first and second rows are barely distinguishable and this shows that any possible difference in the photometric calibration has a completely insignificant effect on the results. The distribution from both FSPS and GALAXEV is highly consistent, almost at the level of the individual ‘blobs’, except for slightly older younger ages and higher metallicities for the GALAXEV vs the FSPS models.

The plots indicate that even galaxies with low σ_* are still dominated by stars with high metallicity, but this is diluted by extra lower-metallicity populations acquired at different times. I should stress that the relatively smooth distribution in the maps are averages of many galaxies and should not be interpreted as the evolution of one individual galaxy, which is generally characterized by discrete star formation events. Moreover, not every feature of the maps is robust against variations in the data and SPS models.

The overall observed distribution could be interpreted in the context of the two-phases of galaxy formation (e.g. Oser et al. 2010). According to this scenario, the formation of galaxies has a ‘two-phase’ nature: a fast initial phase at $z \gtrsim 2$ where ‘*in situ*’ stars are created inside the galaxy from cold gas that falls in, and a longer phase since $z \lesssim 3$ where ‘*ex situ*’ already-formed stars are mainly acquired. In this phase, large systems increase their mass and radius by absorbing smaller stellar systems that were formed very early ($z \gtrsim 3$) outside of the central galaxy’s virial radius, or by smooth gas accretion from cosmological filaments (see Naab & Ostriker 2017, for a review).

Specifically, the old, high-metallicity component observed in Fig. 15 could be interpreted as the relic of the *in situ* formation, which was quickly metal enriched, while the lower metallicity would correspond to either acquired stars, previously formed in smaller stellar components, or to star formation due to accretion from low-metallicity cosmological filaments. The accreted component is only present below the critical ‘quenching boundary’ of $\sigma_* \lesssim 200 \text{ km s}^{-1}$.

Below that boundary, accretion can continue throughout the galaxies evolution. The high-metallicity old peak is visible for all four subsets of σ_* , but its age decreases with σ_* . This age trend in the old-age peak is the same already pointed out in Fig. 14.

The LEGA-C spectra I analysed are not spatially resolved, but a similar analysis of spatially resolved integral-field spectroscopic data for the MaNGA survey shows that, in low σ_* galaxies in the nearby Universe, the oldest higher-metallicity component is associated to the galaxy bulge, while low-metallicity gas accretion happens in the disk (e.g. Lu et al. 2023).

As expected, the BPASS models show again quite different results, with a markedly different metallicity distribution. As commented earlier, the results from this model should be treated with caution as they are likely dominated by spurious unknown effects in the models.

A caveat on these results on the metallicity distribution, which also affects other similar results on metallicity determinations from galaxy stellar spectra, is that the signature of metallicity variations becomes weaker at younger ages, where the S/N of the data also generally decreases. This can introduce possible systematic effects on metallicity trends. To exclude the effect of S/N , I verified that all results remain unchanged if I restrict the analysis to the 873 galaxies with $S/N > 20$ and even, at coarser resolution, for the subset of 126 galaxies with $S/N > 40$. It would still be valuable to compare the reported metallicity trends e.g. with those inferred from gas tracers from similar data.

8 SUMMARY

In the first half of this paper, I described some modifications to the PPXF method (Cappellari 2017), which is used to extract the stellar and gas kinematics, as well as the stellar population of galaxies. First, I described a novel constrained least-squares optimization algorithm that PPXF has been using for the past few years. Then I outlined the changes I made to PPXF to be able to fit photometric data together with the usual full-spectrum fitting. I also described some other minor changes.

In the second half of the paper, I presented an application of PPXF to the extraction of non-parametric SFHs and metallicity distributions for a sample of 3200 galaxies at redshift $0.6 < z < 1$ with spectroscopy from the LEGA-C survey DR3 (van der Wel et al. 2021), and with 28-bands photometric measurements covering from the far-ultraviolet (0.1 μm) to the near-infrared (3 μm) from either the UltraVISTA (Muzzin et al. 2013) or the COSMOS2020 catalogues (Weaver et al. 2022). I also constructed JAM dynamical models (Cappellari 2008, 2020) for all galaxies with measured stellar dispersion σ_* and available Sersic profile fits to the photometry.

For this study, I used and compared three spectral population synthesis (SPS) methods satisfying some criteria of age and wavelength coverage. This led to my selection of the FSPS (Conroy et al. 2009; Conroy & Gunn 2010), GALAXEV (Bruzual & Charlot 2003) and BPASS (Stanway & Eldridge 2018; Byrne et al. 2022) SPS methods.

I compared the dynamical masses from JAM against the stellar masses from the different stellar-population fitting methods. I found that PPXF with photometry and spectra provides more accurate masses than the other methods with photometry alone, as one would have expected.

I found that PPXF on these data reveals a striking difference between galaxies that are only consistent with a single star formation event from those that require multiple bursts of star formation.

I constructed scaling relations for the global stellar population parameters and found a remarkable similarity, but even clearer trends, between these results at $z \approx 0.8$ and those from the latest

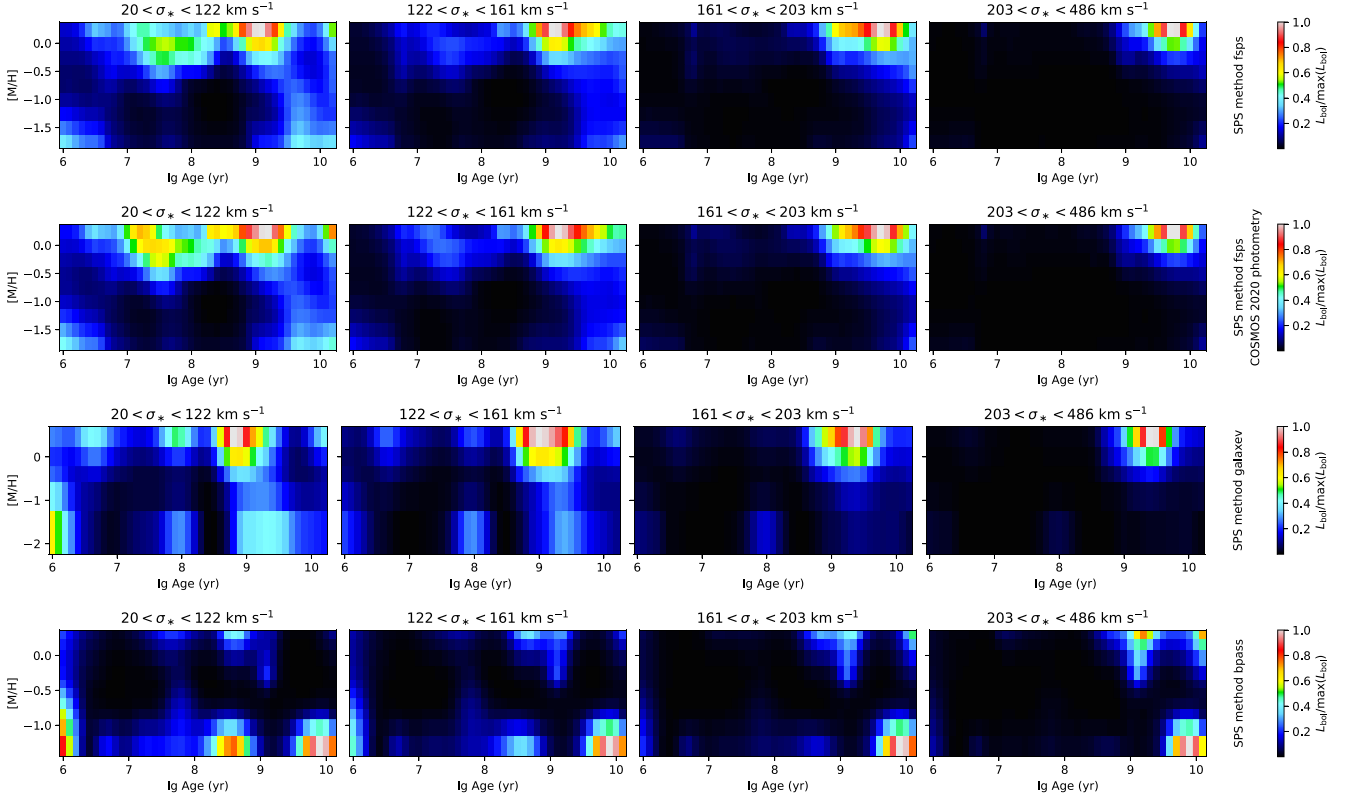


Figure 15. Joint star-formation history (SFH) and metallicity distributions. Each panel shows the average distribution of weights recovered with PPXF (with $r_{\text{regul}}=10$) for 4 bins containing 800 LEGA-C galaxies sorted by their stellar velocity dispersion σ_* (as indicated in the plot titles). The weights represent the bolometric luminosity L_{bol} contributed by populations of different ages and metallicities. The top row shows the results obtained with the FSPS SPS models. The second row still uses FSPS but adopts the COSMOS2020 (Weaver et al. 2022) instead of my default UltraVISTA photometric catalogue (Muzzin et al. 2013). The third row is derived using the GALAXEV SPS models and the bottom one with the BPASS models. One can see that the FSPS and GALAXEV SPS provide qualitatively consistent results even for some of the main ‘blobs’ in the distributions. Using an alternative photometric catalogue has virtually no effect on the results. At large σ_* galaxies on average quenched long ago and their stars have high metallicity. At progressively lower σ_* the age of the bulk of the star formation decreases, while still being dominated by high metallicity stars. However, the population is polluted by fresh accretion events of lower metallicity and a range of accretion times. As previously noted, the results using BPASS SPS models are significantly different and should not be trusted, without further analysis.

spectroscopic surveys in the nearby Universe. This gives some confidence in the meaningfulness of the results and highlights the quality of the spectro-photometric data.

Finally, I explored the non-parametric SFHs and the joint SFH and metallicity $[M/H]$ distributions. I found that the data indicate, on average over many galaxies, a remarkably sharp quenching boundary for the cessation of star formation, at a stellar velocity dispersion $\lg(\sigma_*/\text{km s}^{-1}) \approx 2.3$ ($\sigma_* \approx 200 \text{ km s}^{-1}$), or equivalently with average mass density within 1 kpc $\lg(\Sigma_1^{\text{JAM}}/\text{M}_\odot \text{kpc}^{-2}) \gtrsim 9.9$ ($\Sigma_1^{\text{JAM}} \gtrsim 7.9 \times 10^9 \text{ M}_\odot \text{kpc}^{-2}$), or at metallicity $[M/H] \approx -0.1$ (with some variation dependent on the adopted SPS model) or at Sersic (1968) index $\lg n_{\text{Ser}} \approx 0.5$ ($n_{\text{Ser}} \approx 3.2$). As expected, the transition is more gradual as a function of stellar mass. This abrupt quenching boundary has been invoked by several models of galaxy formation. These data provide one of the cleanest empirical evidence to date.

The joint age-metallicity distribution appears to support the two-phase scenario of galaxy evolution by revealing the relic of an old quickly-formed high-metallicity component and, below the quenching boundary $\sigma_* \lesssim 200$, multiple events of lower-metallicity accretion.

This paper only scratches the surface of what can be done with this data set and with similar ones that are being acquired at comparable and higher redshift. I have not explored e.g. obvious dependencies between SFH and stellar kinematics or environment (e.g. Cole et al.

2020; Sobral et al. 2022). Comparisons with galaxy formation models should be performed in the space of observable rather than using stellar masses which are empirically more uncertain. A similar analysis at higher redshift can reveal the onset and variation of the quenching boundary, which is a key but still quite uncertain parameter in galaxy formation models. JWST data are ideal to extend this kind of study to higher redshift.

ACKNOWLEDGEMENTS

I am grateful to the referee for an expert and very useful report. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under program IDs 194-A.2005 and 1100.A-0949 (The LEGA-C Public Spectroscopy Survey).

DATA AVAILABILITY

The LEGA-C DR3 spectra and catalogue are available [HERE](#), the UltraVISTA photometric catalogue [HERE](#), the COSMOS2020 catalogue from <https://cosmos2020.calet.org/>, the PPXF software from <https://pypi.org/project/ppxf/>, the JAM software from <https://pypi.org/project/jampy/>, the MGEFIT software from <https://pypi.org/project/mgefit/>, and the LTSFIT software from <https://pypi.org/project/ltsfit/>.

REFERENCES

- Aihara H. et al., 2019, *PASJ*, 71, 114
- Andersen M., Dahl J., Liu Z., Vandenbergh, 2011, in Lieven Sra S., Nowozin S., Wright S. J., eds, *Optimization for Machine Learning*. MIT Press, Cambridge, MA, p. 55
- Andersson C. A., Bro R., 2000, *Chemometr. Intell. Lab. Syst.*, 52, 1
- Arnouts S. et al., 2002, *MNRAS*, 329, 355
- Baade W., 1963, *Evolution of Stars and Galaxies*. Harvard Univ. Press, Cambridge, MA
- Bacon R. et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conference Series Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III*. SPIE, Bellingham, p. 8
- Baldry I. K., 2018, preprint (arXiv:1812.05135)
- Barone T. M. et al., 2018, *ApJ*, 856, 64
- Barone T. M., D'Eugenio F., Colless M., Scott N., 2020, *ApJ*, 898, 62
- Barone T. M. et al., 2022, *MNRAS*, 512, 3828
- Belfiore F. et al., 2019, *AJ*, 158, 160
- Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, *ApJS*, 149, 289
- Bell E. F. et al., 2012, *ApJ*, 753, 167
- Belli S., Newman A. B., Ellis R. S., 2014, *ApJ*, 783, 117
- Belli S., Newman A. B., Ellis R. S., 2017, *ApJ*, 834, 18
- Benítez N., 2000, *ApJ*, 536, 571
- Bertelli G., Bressan A., Chiosi C., Fagotto F., Nasi E., 1994, *A&AS*, 106, 275
- Bessell M., Murphy S., 2012, *PASP*, 124, 140
- Bevacqua D., Cappellari M., Pellegrini S., 2022, *MNRAS*, 511, 139
- Beverage A. G., Kriek M., Conroy C., Bezanson R., Franx M., van der Wel A., 2021, *ApJ*, 917, L1
- Bezanson R. et al., 2018, *ApJ*, 868, L36
- Bolzonella M., Miralles J. M., Pelló R., 2000, *A&A*, 363, 476
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Brady M., Horn B. K. P., 1983, *Comput. Vis. Graph. Image Process.*, 22, 70
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Branch M. A., Coleman T. F., Li Y., 1999, *SIAM J. Sci. Comput.*, 21, 1
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bryant J. J. et al., 2015, *MNRAS*, 447, 2857
- Bundy K. et al., 2015, *ApJ*, 798, 7
- Bunn E. F., Hogg D. W., 2009, *Am. J. Phys.*, 77, 688
- Byrne C. M., Stanway E. R., Eldridge J. J., McSwiney L., Townsend O. T., 2022, *MNRAS*, 512, 5329
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Capak P. et al., 2007, *ApJS*, 172, 99
- Cappellari M., 2002, *MNRAS*, 333, 400
- Cappellari M., 2008, *MNRAS*, 390, 71
- Cappellari M., 2011, in Paper presented at the conference on Galaxy Formation. Durham University, Durham, UK, available at: <http://astro.dur.ac.uk/Gal2011/talks.php>.
- Cappellari M., 2016, *ARA&A*, 54, 597
- Cappellari M., 2017, *MNRAS*, 466, 798
- Cappellari M., 2020, *MNRAS*, 494, 4819
- Cappellari M., Emsellem E., 2004, *PASP*, 116, 138
- Cappellari M. et al., 2006, *MNRAS*, 366, 1126
- Cappellari M. et al., 2011, *MNRAS*, 413, 813
- Cappellari M. et al., 2012, *Nature*, 484, 485
- Cappellari M. et al., 2013a, *MNRAS*, 432, 1709
- Cappellari M. et al., 2013b, *MNRAS*, 432, 1862
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018, *MNRAS*, 480, 4379
- Chabrier G., 2003, *PASP*, 115, 763
- Charlot S., Fall S. M., 2000, *ApJ*, 539, 718
- Chauke P. et al., 2018, *ApJ*, 861, 13
- Chauke P. et al., 2019, *ApJ*, 877, 48
- Chen Z. et al., 2020, *ApJ*, 897, 102
- Chen Y.-P., Trager S. C., Peletier R. F., Lançon A., Vazdekis A., Prugniel P., Silva D. R., Gonneau A., 2014, *A&A*, 565, A117
- Cheung E. et al., 2012, *ApJ*, 760, 131
- Chevallard J., Charlot S., 2016, *MNRAS*, 462, 1415
- Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, 823, 102
- Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, *MNRAS*, 358, 363
- Ciotti L., 1991, *A&A*, 249, 99
- Ciotti L., Bertin G., 1999, *A&A*, 352, 447
- Cleveland W. S., Devlin S. J., 1988, *J. Am. Stat. Assoc.*, 83, 596
- Cole J. et al., 2020, *ApJ*, 890, L25
- Conroy C., 2013, *ARA&A*, 51, 393
- Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
- Conroy C., Gunn J. E., White M., 2009, *ApJ*, 699, 486
- Cooley J. W., Tukey J. W., 1965, *Math. Comput.*, 19, 297
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Davidson R., Flachaire E., 2008, *J. Econometr.*, 146, 162
- Dekel A., Birnboim Y., 2006, *MNRAS*, 368, 2
- Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, *A&A*, 350, 381
- Efron B., 2011, *J. Biopharmaceutical Stat.*, 21, 1052
- Efron B., Tibshirani R. J., 1994, *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC
- Eggleton P. P., 1971, *MNRAS*, 151, 351
- Eldridge J. J., Izzard R. G., Tout C. A., 2008, *MNRAS*, 384, 1109
- Emsellem E. et al., 2011, *MNRAS*, 414, 888
- Esdaile J. et al., 2021, *ApJ*, 908, L35
- Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, *A&A*, 532, A95
- Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, *ApJ*, 776, 63
- Feldmann R. et al., 2006, *MNRAS*, 372, 565
- Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761
- Ferland G. J. et al., 2013, *RMxAA*, 49, 137
- Feroz F., Hobson M. P., Bridges M., 2009, *MNRAS*, 398, 1601
- Fletcher R., 1987, *Practical Methods of Optimization*. John Wiley & Sons, Ltd, Chichester, England
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Forrest B. et al., 2022, *ApJ*, 938, 109
- Franx M., van Dokkum P. G., Schreiber N. M. F., Wuyts S., Labbé I., Toft S., 2008, *ApJ*, 688, 770
- Fu Y., Cappellari M., Mao S., Lu S., Zhu K., Li R., 2023, *MNRAS*, 524, 5827
- Getreuer P., 2011, *Image Process. On Line*, 1, 238
- Gill P. E., Murray W., Wright M. H., 1981, *Practical Optimization*. SIAM, Philadelphia
- Gill P. E., Murray W., Saunders M. A., 2005, *SIAM Review*, 47, 99
- Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, *A&AS*, 141, 371
- Golub G. H., Van Loan C. F., 2013, *Matrix computations*, 4th edition. Johns Hopkins University Press, Baltimore
- Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, *ApJ*, 594, 279
- Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, *ApJ*, 542, 710
- Greengard L., Lee J.-Y., 2004, *SIAM Rev.*, 46, 443
- Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund Å., Plez B., 2008, *A&A*, 486, 951
- Hamadouche M. L. et al., 2022, *MNRAS*, 512, 1262
- Hansen P. C., 1998, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. Mathematical Modeling and Computation Vol. 4*. SIAM, Philadelphia
- Hogg D. W., 1999, preprint (arXiv:astro-ph/9905116)
- Huangfu Q., Hall J. A. J., 2017, *Math. Program. Comput.*, 10, 119
- Johnson B. et al., 2021a, *dfm/python-fsps: python-fsps v0.4.1rc1*, Zenodo, available at: <https://zenodo.org/record/4737461>
- Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021b, *ApJS*, 254, 22

- Kabanikhin S. I., 2011, *Inverse and Ill-posed Problems*. de Gruyter, Berlin
- Kacharov N., Neumayer N., Seth A. C., Cappellari M., McDermid R., Walcher C. J., Böker T., 2018, *MNRAS*, 480, 1973
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Koornneef J., Bohlin R., Buser R., Horne K., Turnshek D., 1986, *Highlights Astron.*, 7, 833
- Kriek M., Conroy C., 2013, *ApJ*, 775, L16
- Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, *ApJ*, 700, 221
- Lacerda E. A. D., Sánchez S. F., Mejía-Narváez A., Camps-Fariña A., Espinosa-Ponce C., Barrera-Ballesteros J. K., Ibarra-Medel H., Lugo-Aranda A. Z., 2022, *New Astron.*, 97, 101895
- Lampinen J., 2002, in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No.02TH8600)*. IEEE, New York, p. 1468
- Lawson C. L., Hanson R. J., 1995, *Solving Least Squares Problems (SIAM 1995 edition). Classics in Applied Mathematics, Vol. 15*. Prentice-Hall Inc., Englewood Cliffs, NJ
- Leja J. et al., 2019, *ApJ*, 877, 140
- Lena D. et al., 2015, *ApJ*, 806, 84
- Levenberg K., 1944, *Q. Appl. Math.*, 2, 164
- Le Borgne J. F. et al., 2003, *A&A*, 402, 433
- Le Fèvre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, *Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*. SPIE, Bellingham, p. 1670
- Li H. et al., 2017, *ApJ*, 838, 77
- Li H. et al., 2018, *MNRAS*, 476, 1765
- López Fernández R. et al., 2016, *MNRAS*, 458, 184
- Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, *ApJ*, 904, 33
- Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2022, *ApJ*, 931, 14
- Lu S., Zhu K., Cappellari M., Li R., Mao S., Xu D., 2023, preprint (arXiv:2304.11712)
- MacKay D. J. C., 1992, *Neural Comput.*, 4, 415
- MacKay D. J. C., 2003, *Information Theory, Inference & Learning Algorithms*. Cambridge Univ. Press, available at: <http://www.inference.org.uk/mackay/itila/>
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Maraston C., 2005, *MNRAS*, 362, 799
- Maraston C., Strömbäck G., 2011, *MNRAS*, 418, 2785
- Maraston C. et al., 2020, *MNRAS*, 496, 2962
- Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, *A&A*, 482, 883
- Markwardt C. B., 2009, in Bohlender D., Durand D., Dowler P., eds, *ASP Conf. Ser. Vol. 411, Astronomical Data Analysis Software and Systems XVIII*. Astron. Soc. Pac., San Francisco, p. 251
- Marquardt D. W., 1963, *J. Soc. Ind. Appl. Math.*, 11, 431
- Martin D. C. et al., 2005, *ApJ*, 619, L1
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- McDermid R. M. et al., 2015, *MNRAS*, 448, 3484
- Mendel J. T. et al., 2020, *ApJ*, 899, 87
- Mitzkus M., Cappellari M., Walcher C. J., 2017, *MNRAS*, 464, 4789
- Moré J. J., 1978, in *Lecture Notes in Mathematics*. Springer, Berlin, p. 105
- Moré J. J., Garbow B. S., Hillstom K. E., 1980, *User guide for MINPACK-1*. Argonne National Laboratory Argonne, IL, available at: <http://cds.cern.ch/record/126569>
- Muzzin A. et al., 2013, *ApJ*, 777, 18
- Naab T., Ostriker J. P., 2017, *ARA&A*, 55, 59
- Nocedal J., Wright S., 2006, *Numerical Optimization. Springer Series in Operations Research and Financial Engineering*. Springer, New York
- Noll S. et al., 2009, *A&A*, 499, 69
- Ocvirk P., Pichon C., Lançon A., Thiébaud E., 2006, *MNRAS*, 365, 46
- Oh K., Yi S. K., Schawinski K., Koss M., Trakhtenbrot B., Soto K., 2015, *ApJS*, 219, 1
- Oh S. et al., 2020, *MNRAS*, 495, 4638
- Olver F. W. J., Lozier D. W., Boisvert R. F., Clark C. W., 2010, *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge, available at: <http://dlmf.nist.gov/>
- Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, *ApJ*, 725, 2312
- Pozzetti L., Mannucci F., 2000, *MNRAS*, 317, L17
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 2007, *Numerical recipes: The art of scientific computing*, 3rd edn. Cambridge Univ. Press, Cambridge, available at: <http://numerical.recipes/book>
- Prugniel P., Soubiran C., 2001, *A&A*, 369, 1048
- Rousseeuw P., Van Driessen K., 2006, *Data Min. Knowl. Discovery*, 12, 29
- Rubin D. B., 1981, *The Annals of Statistics*, 9
- Salim S., Narayanan D., 2020, *ARA&A*, 58, 529
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Salvato M., Ilbert O., Hoyle B., 2019, *Nature Astron.*, 3, 212
- Sánchez S. F. et al., 2012, *A&A*, 538, A8
- Sánchez S. F. et al., 2016, *Rev. Mex. Astron. Astrofis.*, 52, 21
- Sánchez-Blázquez P. et al., 2006, *MNRAS*, 371, 703
- Sanders D. B. et al., 2007, *ApJS*, 172, 86
- Scott N. et al., 2017, *MNRAS*, 472, 2833
- Sersic J. L., 1968, *Atlas de galaxias australes*. Obs. Astron. Univ. Nacional de Córdoba, Córdoba
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, 343, 978
- Shetty S., Cappellari M., 2014, *ApJ*, 786, L10
- Shetty S., Cappellari M., 2015, *MNRAS*, 454, 1332
- Shetty S., Cappellari M., McDermid R. M., Krajnović D., de Zeeuw P. T., Davies R. L., Kobayashi C., 2020a, *MNRAS*, 494, 5619
- Shetty S., Bershady M. A., Westfall K. B., Cappellari M., Drory N., Law D. R., Yan R., Bundy K., 2020b, *ApJ*, 901, 101
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Silva L., Granato G. L., Bressan A., Danese L., 1998, *ApJ*, 509, 103
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Sobral D. et al., 2022, *ApJ*, 926, 117
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Speagle J. S., 2020, *MNRAS*, 493, 3132
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, 479, 75
- Stark P. B., Parker R. L., 1995, *Comput. Stat.*, 10, 129
- Storey P. J., Hummer D. G., 1995, *MNRAS*, 272, 41
- Storn R., Price K., 1997, *J. Global Optim.*, 11, 341
- Straatman C. M. S. et al., 2018, *ApJS*, 239, 27
- Tabor M., Merrifield M., Aragón-Salamanca A., Cappellari M., Bamford S. P., Johnston E., 2017, *MNRAS*, 466, 2024
- Tabor M., Merrifield M., Aragón-Salamanca A., Fraser-McKelvie A., Peterken T., Smethurst R., Drory N., Lane R. R., 2019, *MNRAS*, 485, 1546
- Tacchella S. et al., 2022, *ApJ*, 926, 134
- Taniguchi Y. et al., 2007, *ApJS*, 172, 9
- Tojeiro R., Heavens A. F., Jimenez R., Panter B., 2007, *MNRAS*, 381, 1252
- van Dokkum P. G. et al., 2015, *ApJ*, 813, 23
- van Houdt J. et al., 2021, *ApJ*, 923, 11
- van de Sande J. et al., 2013, *ApJ*, 771, 85
- van der Wel A. et al., 2014, *ApJ*, 788, 28
- van der Wel A. et al., 2016, *ApJS*, 223, 29
- van der Wel A. et al., 2021, *ApJS*, 256, 44
- Vazdekis A., Sánchez-Blázquez P., Falcón-Barroso J., Cenarro A. J., Beasley M. A., Cardiel N., Gorgas J., Peletier R. F., 2010, *MNRAS*, 404, 1639
- Vazdekis A. et al., 2015, *MNRAS*, 449, 1177
- Venturi G. et al., 2021, *A&A*, 648, A17
- Verro K. et al., 2022a, *A&A*, 660, A34
- Verro K. et al., 2022b, *A&A*, 661, A50
- Virtanen P. et al., 2020, *Nature Methods*, 17, 261
- Voglis C., Lagaris I. E., 2004, in *WSEAS International Conference on Applied Mathematics*. Corfu, Greece
- Wake D. A., van Dokkum P. G., Franx M., 2012, *ApJ*, 751, L44
- Walcher J., Groves B., Budavári T., Dale D., 2011, *Ap&SS*, 331, 1
- Weaver J. R. et al., 2022, *ApJS*, 258, 11
- Werle A., Cid Fernandes R., Vale Asari N., Bruzual G., Charlot S., Gonzalez Delgado R., Herpich F. R., 2019, *MNRAS*, 483, 2382
- Westera P., Lejeune T., Buser R., Cuisinier F., Bruzual G., 2002, *A&A*, 381, 524
- Westfall K. B. et al., 2019, *AJ*, 158, 231

- Wilkinson D. M., Maraston C., Goddard D., Thomas D., Parikh T., 2017, *MNRAS*, 472, 4297
- Willmer C. N. A., 2018, *ApJS*, 236, 47
- Worthey G., 1994, *ApJS*, 95, 107
- Worthey G., Faber S. M., Gonzalez J. J., Burstein D., 1994, *ApJS*, 94, 687
- Yan R. et al., 2019, *ApJ*, 883, 175
- York D. G. et al., 2000, *AJ*, 120, 1579
- Young L. M. et al., 2011, *MNRAS*, 414, 940
- Zhu K., Lu S., Cappellari M., Li R., Mao S., Gao L., 2023a, preprint (arXiv:2304.11714)
- Zhu K., Lu S., Cappellari M., Li R., Mao S., Gao L., 2023b, *MNRAS*, 522, 6326

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

cappellari2023_table1.txt

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

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.