

SDSS-IV MaNGA: Excavating the fossil record of stellar populations in spiral galaxies

Thomas Peterken,^{1*} Michael Merrifield,¹ Alfonso Aragón-Salamanca,¹
 Amelia Fraser-McKelvie,¹ Vladimir Avila-Reese,² Rogério Riffel,^{3,4} Johan Knapen,^{5,6}
 Niv Drory⁷

¹*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

²*Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, 04510 CDMX, México*

³*Instituto de Física (IF), Universidade Federal do Rio Grande do Sul, CP 15051, Porto Alegre 91501-970, RS, Brazil*

⁴*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil*

⁵*Instituto de Astrofísica de Canarias, Vía Láctea S/N, E-38205 La Laguna, Spain*

⁶*Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Spain*

⁷*McDonald Observatory, The University of Texas at Austin, 1 University Station, Austin, TX 78712, USA*

Draft copy

ABSTRACT

We perform a “fossil record” analysis for ≈ 800 low-redshift spiral galaxies, using STARLIGHT applied to integral field spectroscopic observations from the SDSS-IV MaNGA survey to obtain fully spatially-resolved high-resolution star formation histories (SFHs). From the SFHs, we are able to build maps indicating the present-day distribution of stellar populations of different ages in each galaxy. We find small negative mean age gradients in most spiral galaxies, especially at high stellar mass, which reflects the formation times of stellar populations at different galactocentric radii. We show that the youngest ($< 10^{8.5}$ years) populations exhibit significantly more extended distributions than the oldest ($> 10^{9.5}$ years), again with a strong dependence on stellar mass. By interpreting the radial profiles of “time slices” as indicative of the size of the galaxy at the time those populations had formed, we are able to trace the simultaneous growth in mass and size of the spiral galaxies over the last 10 Gyr. Despite finding that the evolution of the measured light-weighted radius is consistent with inside-out growth in the majority of spiral galaxies, the evolution of an equivalent mass-weighted radius has changed little over the same time period. Since radial migration effects are likely to be small, we conclude that the growth of disks in spiral galaxies has occurred predominantly through an inside-out mode (with the effect greatest in high-mass galaxies), but this has not had anywhere near as much impact on the distribution of mass within spiral galaxies.

Key words: Galaxies:disc – Galaxies:evolution – Galaxies:formation – Galaxies:spiral – Galaxies:structure

1 INTRODUCTION

Understanding how, when, and where galaxies built their mass is key to cosmology and astronomy. Analysis of the evolution of the masses and sizes of galaxies has generally been limited to comparisons of different galaxy populations at different redshifts. Studies done in this manner have shown that galaxies have grown in radius whilst building their mass (e.g. ???????), giving rise to the concept of “inside out” formation. It is thought that such growth in the most massive galaxies has been due to some combination of multiple minor mergers (??), gas accretion (?), and quasar feedback (?). These approaches have given us a good insight into how the

average properties of galaxies have evolved over cosmic time, but because we cannot track the evolution of any individual system in this way, it is difficult to go beyond such global properties. Although some studies of galaxies at different redshifts have managed to show inside-out growth in disk-like galaxies (e.g. ??), most are restricted to the highest mass galaxies, so this picture of inside-out growth is normally limited to early-type galaxies.

An alternative approach which is more suited to late-type galaxies is to explore the stellar populations in different regions of a galaxy, particularly through studying how the mean stellar age varies with radius. This method requires high-quality spectral data at multiple locations across the face of a galaxy, and so has only been undertaken in detail for large numbers of galaxies since the advent of integral-field

* E-mail: Thomas.Peterken@nottingham.ac.uk

spectroscopic surveys such as the Calar-Alto Legacy Integral Field Array (CALIFA; ?), Sydney-AAO Multi-object Integral field spectrograph (SAMi; ?), and Mapping Nearby Galaxies at APO (MaNGA; ?) surveys. Using such a “fossil record” approach applied to integral-field spectroscopic data has revealed that most galaxies exhibit negative age gradients (e.g. ????) — with younger outskirts than centres — or earlier formation times of the central regions (e.g. ?) providing more evidence for a dominant “inside out” growth mode occurring in galaxies of all Hubble types. This is also backed up by ? for the case of NGC 7793, who find that broad-band observations of resolved stellar populations in this nearby spiral galaxy indicate a clear gradient in stellar age.

By applying stellar population modelling methods to integral-field spectroscopic data from the CALIFA survey, ??, ?, ?, and ? have shown that it is possible to reveal much more about a galaxy’s history by deriving full star-formation histories rather than mean ages. We have shown previously that such analyses of the spatial variation in stellar populations of spiral galaxies can help us understand the structure of the spiral arms and bars, but here we investigate how such approaches can also help us study the evolution and growth of populations of galaxies.

Comparative studies of the masses and sizes of galaxies at different lookback times are most effective to measure the growth of early-type galaxies since these are typically the most massive and luminous objects at any given redshift so are easy to identify. By contrast, a fossil record analysis acts as a complementary approach best suited to — but by no means limited to; see e.g. ? — studying the growth of late-type galaxies, as such galaxies have in general had continued growth over the last several Gyr. This extended star-formation in late-type galaxies can be traced using fossil record methods, providing that care is taken to ensure that older populations can be detected when the flux may be dominated by the younger and brighter populations. Of course, there exists a population of spiral galaxies which are passive (see for example ??) — contrary to the well-known relation between the morphology and star formation rate (??) — so a morphological classification does not always define the extent of the star-formation history of each galaxy. However, for consistency, we have chosen to study a galaxy population selected on their morphology rather than colour, to better understand how this well-defined galaxy class have evolved over time.

Here, we perform full spectral fitting of spiral galaxies from the MaNGA survey (?) and measure spatially-resolved star formation histories, to uncover their formation sequences.

This paper is structured as follows. In §2, we outline the data we use from the MaNGA survey. In §3 we describe how a sample of spiral galaxies from the MaNGA target list was selected, and in §4 we detail the spectral fitting method employed (with some tests of this method outlined in Appendix A). We then describe how the derived star-formation histories are processed in §5. The mean age and metallicity gradients are derived in §6. In §7 and §8 we analyse the star-formation histories and spatially-resolved stellar populations in more detail, and infer the evolution of the mass-size rela-

tion in §9. Finally, we discuss the interpretation and context of the results in §10.

2 DATA

2.1 MaNGA

MaNGA (?) is part of the fourth generation of the Sloan Digital Sky Survey (SDSS-IV; ?). By the survey’s completion in 2020, MaNGA will have acquired 2.5 arcsec resolution integral-field spectroscopic observations of more than 10,000 galaxies in the redshift range $0.01 < z < 0.15$ (?). The survey makes use of the BOSS spectrograph (?) on the 2.5-metre SDSS telescope (?) at the Apache Point Observatory, which has a spectral resolution of $R \approx 2000$ and covers a large wavelength range of 3600-10300 Å. The raw data’s calibration is described by ?, with the datacubes then reduced using MaNGA’s data reduction pipeline (DRP; ?). For each target galaxy, observations are taken out to at least either $1.5 R_e$ or $2.5 R_e$ (to form the “Primary” and “Secondary” samples respectively; ?), where R_e is the elliptical half-light radius measured photometrically by the NASA-Sloan Atlas (?). This is achieved using integral-field units of five different sizes, from the 127-fibre IFUs with a diameter of 32 arcsec, to the 19-fibre IFUs of 12 arcsec diameter (?).

The MaNGA target selection was chosen to obtain a flat distribution in $\log(\text{stellar mass})$ (?). Neither the Primary nor Secondary samples are therefore volume-limited; instead the high-mass galaxies are over-represented while the low-mass galaxies are under-represented. The Primary+ (“colour-enhanced”) sample is an extended Primary sample but with an oversampling of the “green valley” galaxies (?) so is therefore also unrepresentative in this way too. However, since the sample selection in all cases is well-defined (?), a weighting has been determined for each galaxy to correct for these selection biases and form a representative volume-limited sample (referred to as the “Primary+ sample weighting” throughout this paper).

In this work, we make use of some of the analysis outputs of MaNGA’s data analysis pipeline (DAP; ?). Specifically, we use the measured stellar velocities v_* , deprojected radii R , and emission line spectra [which are themselves described in detail by ?], all of which are derived using full spectral modelling. The data we use here is from the internal MaNGA product launch 8 (MPL-8) data release, which contains completed observations of 6778 galaxies.

2.2 Galaxy Zoo

We also make use of the morphological classifications of each MaNGA galaxy provided by volunteer “citizen scientists” as part of Galaxy Zoo (??). The second phase of the project (Galaxy Zoo 2, hereafter GZ2; ?) includes publicly-available detailed classifications of galaxies based on SDSS DR7 imaging. The users’ classifications are weighted and combined to obtain a consensus fraction for each answer to each question for each galaxy, using methods described by ? and ?. We use the redshift-debiased and user-weighted probabilities — which we denote as $p_{\text{classification}}$ — from the ? catalogue.

3 SAMPLE SELECTION

A sample of spiral galaxies was drawn from the MPL-8 data release using the recommendations of ?, Table 3; see also ? for another recent implementation. We first remove the 45 galaxies in the matched MPL-8/GZ2 catalogues that more than 50% of GZ2 users have classified as having some form of star or artifact in the image. To filter out elliptical galaxy morphologies, we select the 4201 galaxies with $p_{\text{features or disk}} > 0.43$ and at least 20 classifications in this question, as recommended by ?.

Since we are interested in the variation in stellar population properties across the face of each spiral galaxy, we remove edge-on galaxies from this sample. This cut can be made with either the GZ2 classifications — specifying $p_{\text{not edge-on}} > 0.8$ — following ?, or using an axis ratio cut — requiring $\frac{b}{a} \geq 0.4$ — following ?. To select only face-on galaxies, we choose galaxies that satisfy the ? criterion and have an axis ratio of $\frac{b}{a} \geq 0.5$ (corresponding to an inclination of $i \leq 60^\circ$ assuming the galaxies can be modelled as a thin intrinsically circular disks). We used this higher axis ratio cut compared to that used by ? to ensure that we have selected only galaxies for which the radial structure is clearly resolvable with MaNGA. Of the 5902 MPL-8 galaxies for which GZ2 classifications are available, this leaves a sample of 1686 close-to-face-on disky galaxies. Of these, 1314 galaxies satisfy the ? requirement for spiral galaxies of $p_{\text{spiral}} > 0.8$ and 20 individual classifications in this question.

We then remove 109 galaxies which have flags for bad or questionable-standard data in the MaNGA DRP, or for which the MaNGA MPL-8 DAP dataproducs are unavailable. To ensure consistency in the spatial resolution relative to the galaxy size, we remove galaxies which are part of MaNGA’s Secondary sample. For the final sample of spiral galaxies, we therefore select only those 795 which are in the Primary+ MaNGA sample, for which MaNGA observations extend to at least $1.5 R_e$. The median redshift of galaxies in our sample (weighted by the MaNGA Primary+ sample weighting) is $z = 0.026$, and 75% of the (weighted) sample are at redshifts $z < 0.03$.

4 SPECTRAL FITTING

Using a similar technique to that employed in ? and ?, we fit each spectrum in each galaxy using STARLIGHT (?). We first de-redshift and subtract the emission-line spectrum using the MaNGA DAP (??), and then fit using E-MILES (?) single stellar population (SSP) templates. The de-redshifted and emission-subtracted MaNGA spectra are rebinned onto a linear wavelength scale (as required by STARLIGHT) before fitting. STARLIGHT then uses an iterative method to find the best-fit linear combination of the input templates, and returns the relative weights given to each SSP template in the fit, along with line-of-sight velocity v_* , velocity dispersion σ_* , and the amount of dust reddening A_V .

4.1 Template stellar population spectra

We use a combination of 9 ages ($\log(\text{age}/\text{years}) = 7.85, 8.15, 8.45, 8.75, 9.05, 9.35, 9.65, 9.95, 10.25$) and 6 metallicities ($[M/H] = -1.71, -1.31, -0.71, -0.40, +0.00, +0.22$)

from the standard E-MILES library (?), based on the earlier MILES library of ?), assuming a ? IMF, ?, “Padova” isochrones, and Milky-Way $[\alpha/\text{Fe}]$ (“baseFe”). To sample the full star-formation histories, we also include the younger templates of ? covering 6 ages ($\log(\text{age}/\text{years}) = 6.8, 6.9, 7.0, 7.2, 7.4, 7.6$) and the two recommended metallicities ($[M/H] = -0.41, +0.00$), which are generated using the same method as the E-MILES set of ?, but with the earlier ? version of the Padova isochrones. Combining these libraries allows us to exploit the high spectral resolution of both MaNGA and E-MILES templates, while still being able to fully fit the whole of the star-formation histories of star-forming regions without combining different libraries produced in completely different ways.

4.2 Starlight configuration

We use STARLIGHT in a “long fit” mode to prioritise robustness over computation time, based on the recommendations from extensive testing of STARLIGHT by ? and ?. We limit the fit to the wavelength range of 3541.4 to 8950.4 Å, where the raw E-MILES templates have a constant FWHM of 2.51 Å. To ensure that the model and measured spectra have consistent resolution, we degraded each of the SSP templates to the wavelength-dependent resolution of the median spaxel spectrum from all galaxies in our sample, using the line spread function measured by the DAP (??).

Since the DAP robustly models the emission lines (?), we use STARLIGHT in its “NOCLIP” mode to ensure that all of the diagnostic absorption lines are fully fitted. To ensure that the star-formation history of each spaxel is measured as fully as possible (defined here as the mass weights assigned to each SSP template divided by the time interval between that template and the next-youngest one), we require STARLIGHT to retain at least 97% of its fit’s total light during the “EX0” phase of reducing the number of templates used in the final fit (i.e. `EX0s_method_option = CUMUL`, `EX0s_Threshold = 0.03`). This configuration helps to recover the presence of older stellar populations even when their flux has been obscured by the presence of a younger population, for example. The light weights we use from the fits are those contributed by each template at 4020 Å (as used by ?), and we allow the sum of weights at this wavelength to be between 50% and 150% of the input spectrum¹. In all subsequent analysis, we use either the mass weights (using the implicit mass-to-light ratios included in the E-MILES SSP models and assuming a flat Λ CDM cosmology with $H_0 = 70$ km/s) or only compare the spatial variation of flux weights of a specific age, rendering the exact choice of reference wavelength irrelevant.

From the STARLIGHT mass weights, a measure of the total stellar mass within the MaNGA FOV can be readily calculated. Reassuringly, we find that these stellar masses agree well with those measured by the NASA-Sloan Atlas

¹ Any given model may represent the entire spectrum well at all wavelengths *except* that at which we measure the weights. It is therefore possible for a good fit’s light weights at any chosen wavelength not to sum to exactly 100% of the input spectrum, so allowing this large discrepancy in the weights’ sum ensures that the best fit of the whole spectrum is used.

(?), but with a small offset due likely due to the difference in FOV limitations. We discuss this comparison further in §9. Although it is possible that the (?) stellar mass measurements are more robust than the STARLIGHT-derived measurements, the consistency between the two measurements is close enough to allow us to use either one. However, in measuring the mass growth in §9, we are limited to using the STARLIGHT measurements. Therefore, for consistency, any quoted galaxy stellar mass measurements are those measured by STARLIGHT unless stated otherwise. The E-MILES library contain stellar mass loss predictions for each of the SSP templates, allowing a measurement of the current mass and an initial mass at time of formation for each population contained within each spectrum. Unless otherwise stated, the mass weightings used in this work are the present-day masses of each template, to avoid reliance on the mass loss predictions. In any case, we find that all results presented here are entirely unaffected by this distinction.

4.3 Treatment of dust extinction

STARLIGHT has the capacity to fit a general dust law with extinction A_V , and also include an extra extinction YA_V which is applied only to specified templates in the fit. This could, for example, allow for the possibility that the youngest stellar components are affected by dust extinction to a greater extent than those populations which would be expected to be free of their birth clouds. The exact values of YA_V measured by STARLIGHT would in that case be an interesting parameter to model and investigate. However, in practice, we found that this extra degree of freedom caused STARLIGHT’s fits to be drawn towards negative extinctions when we included a YA_V term for all populations younger than $10^{7.05}$ years. This is likely due to the combination of the limited wavelength range for which these youngest templates dominate the spectrum due to their extreme colours, and the lack of any significant spectral information beyond their continuum shape. To the best of our knowledge, the YA_V parameter in STARLIGHT has not successfully been applied to any real spectral fitting to date.

We therefore include a single ? dust law in the fit, which has the same A_V for all templates. We allow A_V to vary in the range of $-1 \leq A_V \leq 8$, and we find that over 90% of the spaxel fits are within the range $0.1 \leq A_V \leq 0.8$.

4.4 Kinematics

We use the stellar velocity dispersion σ_* measured by the MaNGA DAP (?) as an initial kinematic guess for the de-redshifted input spectrum’s STARLIGHT fits, but allow this to vary as a free parameter in the range of $\sigma_* = 20$ to 900 km/s. Unlike other spectral fitting tools such as pPXF (?), STARLIGHT is not fine-tuned for measuring stellar kinematics, and we do not expect STARLIGHT’s exact measurements of σ_* to impact the measured SFHs. Similarly, despite de-redshifting each spectrum individually using the DAP’s stellar velocity v_* measurements, this is a free parameter in the range of $v_* = -600$ to 600 km/s to allow STARLIGHT to find its best possible fit, using $v_* = 0$ as the initial guess. In practice, we find that STARLIGHT’s fits are consistent with $v_* = 0$, with little deviation in σ_* from the DAP measurements.

In the tests outlined in §4.6 and described in more detail in Appendix A, we find that setting these kinematic parameters to be fixed or variable has no effect on STARLIGHT’s ability to measure stellar populations. Therefore, to accommodate for any wavelength calibration offsets between the DAP and the SSP templates — and for any uncertainties in the measurement of the spectral resolution — we allow the values of σ_* and v_* in the fit to vary.

4.5 Ignoring the youngest stellar populations

We do not expect the star formation rate or chemical evolution to vary significantly within the last 10^8 years in the majority of cases (see e.g. ?), but initial tests with STARLIGHT revealed that there was a significant correlation between the weights assigned to templates of $\gtrsim 10^{9.5}$ years and those of $\lesssim 10^{7.2}$ years, resulting in a sharp peak in the SFH at $\sim 10^7$ years. This effect was found to be present in all locations of all galaxies regardless of signal-to-noise or the strength of dust extinction, and often resulted in an implied SFR of the galaxy to be at least an order of magnitude greater than at any previous time in its history, of up to $\sim 25 M_\odot/\text{yr}$. ? showed that this phenomenon seems to be related to the known “UV upturn” seen in old stellar populations, which is normally attributed to horizontal branch stars in the planetary nebula phase; see ? for a review. The cause and presence of this excess of blue light is not accounted for in the old SSP template spectra, so STARLIGHT is forced to attribute it to another population.

We first attempted to mitigate this effect by fitting only from 3700 Å instead of 3541.4 Å, but found this had no effect on the derived star-formation histories, and we chose not to increase this lower wavelength limit further to avoid impacting the valuable Balmer absorption series. We then performed another fit with STARLIGHT but where we had combined all of the templates younger than $10^{7.5}$ years for each metallicity into a single template representing a flat SFR over that time interval, and used these two templates in the fit instead of the original eight over this time interval. Comparing the STARLIGHT results of the two approaches shows that enforcing a flat SFR in the youngest templates has no noticeable effect on the SFH in ages $\geq 10^{7.5}$ years at all (< 0.1 dex change in the measured SFR at any look-back time $t \geq 10^{7.5}$ years), but the weights assigned to these new templates still exhibited correlation with those assigned to older stellar populations. Similarly, when we compared STARLIGHT fits using only those SSP templates of ages older than $10^{7.5}$ years, we found that the excess of hot stars was simply assigned to whichever stellar population was youngest. The rest of the star-formation histories were unaffected, indicating that *older stellar populations are reliably measured regardless of how the youngest populations are treated in the fit*. We concluded that the youngest stellar population available in the STARLIGHT fit would always have a “cross-talk” effect with populations $\gtrsim 10^{9.5}$ years. The flux assigned to the youngest populations will always be a combination of the “true” flux from stars of that age, as well as a spurious contribution from the hot stars present but not modelled in older populations.

It may be possible to effectively separate these two effects when stellar population models are able to fully model the hot stellar remnants or other factors responsible for

the UV upturn. However, for the purposes of this work, the weights and fluxes of stellar populations younger than 30 Myr are fundamentally unreliable, so we include these SSP templates in the fit but then *we ignore these populations entirely* and do not use their weights in deriving the STARLIGHT-measured star formation histories. The SFHs are not likely to have varied over this time period (?), but such young stellar populations are clearly present in many galaxies, so by including these SSPs in the fit but ignoring their weights in subsequent analysis allows the spectrum to be fully modelled. Limiting the measurements of the derived SFHs to exclude the region younger than 30 Myr does not limit the results from our analyses.

Based on this, we advise users of STARLIGHT and other stellar population fitting software to carefully consider the effects of attempting to measure star-formation histories to young ages without accounting for the limitations of SSP models to include the UV upturn. Cautious interpretation of all derived SFHs is essential to determine which parts of a SFH are likely to be correctly measured. However, we do see that the older populations are almost entirely unaffected however the youngest populations are modelled, so are robustly reliable.

4.6 Effects of low signal-to-noise

Many authors spatially bin neighbouring spaxels of integral-field spectroscopic data to create regions with approximately constant signal-to-noise ratio (SNR) before fitting. However, since we wish to retain the full spatial information of the stellar populations — and therefore fit each spectrum independently instead of binning — we must ensure that the STARLIGHT fits in regions with low SNR are reliable. ? showed that STARLIGHT may exhibit bias in the fitting of spectra with low SNR, but ? contend that these effects are not significant in most physical applications and with the robust STARLIGHT configuration used here.

In Appendix A, we outline a series of tests to measure the effect of low signal-to-noise ratio on STARLIGHT’s recovered fits, and its ability to recover a stellar population of known age or a star-formation history of known shape using the configuration described above. We find that combining spectra with a given signal-to-noise ratio and comparing the fit of this combined spectrum with the fits of the spectra it contains, STARLIGHT is consistent for the low signal-to-noise regions. We also find that STARLIGHT is able to recover the age of a known stellar population with a signal-to-noise ratio as low as 5. Similarly, with the configuration outlined in §4.2, we find that STARLIGHT can reliably measure the shape of a known SFH in such low signal-to-noise conditions, indicating that we are able to detect the presence of older stellar populations when obscured by brighter younger populations.

These tests imply that, assuming the E-MILES model spectra are accurate representations of the stellar populations they represent, we expect STARLIGHT to be able to recover the true SFHs under all the conditions analysed in the remainder of this paper. Notwithstanding this robustness, to ensure that the low signal-to-noise regions of the galaxy are not affecting our results in ways we don’t anticipate, in all stages of our analysis we ensure that we weight spaxels by their flux or mass, ensuring that the central re-

gions with good fits are up-weighted, and low signal-to-noise regions are down-weighted.

5 TIME-SLICING

From the SSP template weights obtained in the STARLIGHT fits, we are able to reconstruct the star formation history (SFH) and metallicity distributions at every location in each galaxy in the spiral sample. From the SFHs, it is straightforward to reconstruct an image of the total flux (or mass) emitted by (or contained in) stars of any given age. To ensure that we are not over-interpreting small-scale noise in the age-distributions of weights assigned to individual templates, we first smooth the SFHs before any analysis is done on these images. We have smoothed by 0.3 dex in age, but smoothing by any factor between 0.2 and 0.5 dex does not affect results significantly².

As an illustration, Figure 1 shows an animation of a single galaxy (MaNGA plate-IFU 8329-12701) from the spiral sample, stepping through stellar population ages from 17 Gyr down to 30 Myr, highlighting the wealth of information contained in the spatially-resolved SFHs available using STARLIGHT and MaNGA. Such animations can be made for any of the galaxies in the sample, but here we show an example of a galaxy observed using the largest-sized (127-fibre) IFU to demonstrate the amount of information potentially available through such time slicing.

It is worth emphasising that we can only measure the current location of stars in the galaxy, so that we can only treat a “time slice” at any given stellar age as an approximation of the structure of the galaxy at that time, since we cannot undo the effects of dynamical heating or radial mixing and migration. However, ? showed that the shape of vertical colour gradients seen in edge-on disk galaxies imply that radial migration occurs at a slower rate than the intrinsic growth of the galactic disk. Simulations of galactic disks also suggest that stellar populations are in general equally likely to migrate inwards or outwards (?), and only by sufficiently small distances that this effect has only minor effects on the radial distribution of populations (???), so here we assume that the current distribution of a given stellar population is — to a first approximation — representative of the distribution of star formation in the galaxy at the corresponding lookback time.

Clearly this assumption does not hold true for spiral structures, since such distributions will become diluted rapidly with the disk’s rotation. However, for the youngest stellar populations, we showed in ? that interpreting spatially-resolved star-formation histories in this “time-slicing” approach can help to understand spiral arms and bars. ? also showed that a similar approach can be used to understand the properties of AGN, and other studies with CALIFA showed that this approach can offer clues to the history of a galaxy’s radial profile (????).

² We find that the measured uncertainties in the SFHs in the tests described in Appendix A are typically lower than 0.2 dex, but those measurements don’t account for any error or bias in the SSPs themselves, so we have conservatively included 0.3 dex smoothing on temporal scales here.

Figure 1. *Top:* Animation of a MaNGA spiral (plate-IFU 8329-12701) showing the spatially resolved flux (colour-coded by the metallicity) of stars as a function of age, from 10 to 0.03 Gyr. *Middle:* Weighting function used. The STARLIGHT output’s temporal information is smoothed to 0.3 dex. Red points indicate the SSP ages used. *Bottom:* Colour map indicating the flux (in units of 10^{-14} erg s $^{-1}$ cm $^{-2}$ Å $^{-1}$ spaxel $^{-1}$) and metallicity (in units of $\log(Z/Z_{\odot})$) of the stellar population. Dashed vertical lines indicate the SSP metallicities used. *Requires Adobe Reader version ≥ 9 (and $\leq 9.4.1$ on Linux) or similar. A high-resolution version is available online as supplementary material.*

6 MEAN AGES AND METALLICITIES

A first-order measurement of the SFHs resolved across the face of a galaxy is that of the mean age or — with a similar calculation — of spatially-resolved metallicity. Using the mass weights assigned to each SSP template by STARLIGHT in the fits for each spaxel spectrum, we derive mass-weighted mean age and metallicity (specifically $\langle \log(\text{age}/\text{yr}) \rangle_{\text{mass}}$ and $\langle \log(Z/Z_{\odot}) \rangle_{\text{mass}}$ respectively) maps. We then plot the light-weighted median of all spaxels’ mean $\log(\text{age})$ and $\log(\text{metallicity})$ within radial bins of width $0.045 R_e$ (where R_e is the elliptical Petrosian effective radius measurements from the NSA) against the elliptical galactocentric radius R (in units of R_e), and find a best-fit straight line to these data using a least-squares fit. The fitting is only performed out to $1.2 R_e$ to avoid the edges of the hexagonal-shaped IFU FOVs and to ensure consistency between galaxies. From these best-fit lines, we obtain a mean age and

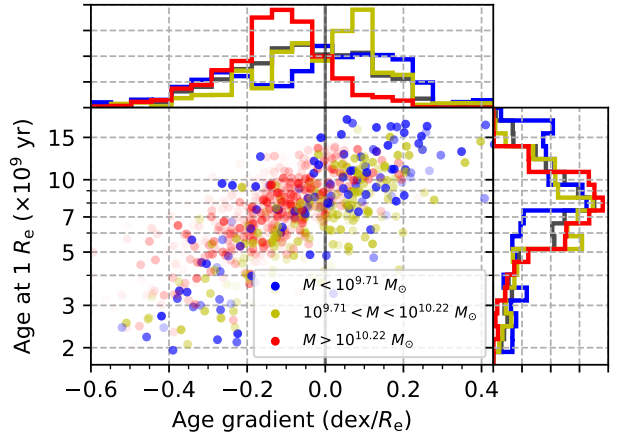


Figure 2. Mass-weighted mean ages at $1 R_e$ (vertical axis, where R_e is the elliptical Petrosian radius measurements from the NSA) and mean age gradients (horizontal axis) for each galaxy, coloured by the galaxy’s total stellar mass. Each data point’s transparency is defined by the MaNGA Primary+ sample weighting. The weighted histograms at the top and right indicate the distributions of ages and their gradients respectively, where the grey line indicates the distribution of the whole sample and the coloured lines indicate those for each of the three mass bins.

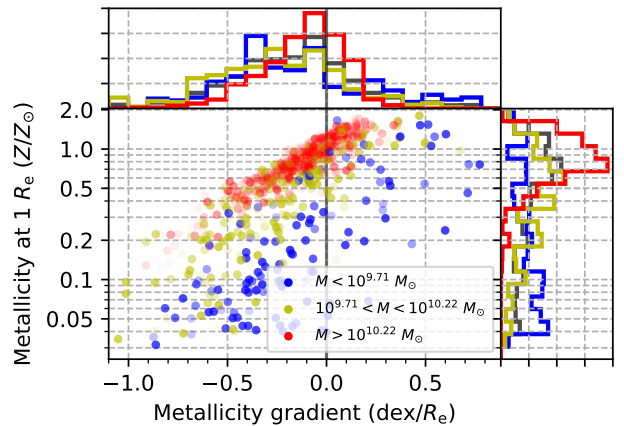


Figure 3. As Figure 2, but for mass-weighted mean metallicities and their gradients.

metallicity gradient, and a characteristic age and metallicity value of the stellar populations located at $1 R_e$, a measure which was shown to be representative of the galaxy as a whole.

The distributions of age gradients and ages at $1 R_e$ are shown in Figure 2, and equivalent metallicity measurement in Figure 3. We find that, on average, a majority (approximately 60%) of the spiral sample exhibit slight negative age gradients, implying younger outskirts. This agrees with the general picture found by others (???) and is usually taken to be evidence for inside-out formation being dominant in the most massive galaxies. When the sample is split into three mass bins (of $M < 10^{9.71} M_{\odot}$, $10^{9.71} < M <$

$10^{10.22} M_{\odot}$, and $M > 10^{10.22} M_{\odot}$ ³), we find that the approximately 80% of the highest-mass galaxies exhibit negative age gradients while only 50% of the lowest-mass galaxies do. This difference suggests that inside-out formation is more dominant in high-mass galaxies.

We find that most (≈ 60 – 80%) galaxies in all mass bins exhibit slight negative metallicity gradients, and Figure 3 highlights a strong mass-metallicity correlation too, as first suggested by ?.

7 MASS BUILDUP TIMES

Measuring only a mass-weighted mean age or metallicity does not make use of all of the available information in the age distribution of SSP template weights. From a full spectrum fitting approach, it is also possible to use the width of the distribution in stellar age, as well as its mean value. To this end, from a given smoothed SFH, we define the time T_{95} by which 95% the total stellar mass of that spectrum was built up. We measure a T_{95} for all light within $R < 1.2 R_e$ of each galaxy. We find that T_{95} correlates with the total stellar mass of the galaxy, as shown in Figure 4: all galaxies with present-day stellar masses within $1.2 R_e$ of $M_{\star} \geq 2 \times 10^{10} M_{\odot}$ formed the bulk of their mass at least 5 Gyr ago, while most of those with stellar masses $M_{\star} \leq 10^{10} M_{\odot}$ were still building their mass as recently as ≈ 2 Gyr ago. This effect is reflected in the known relation between the stellar mass and star formation rates in galaxies, and the results shown here agree well with other fossil record studies (??), empirical modelling (??), and theoretical modelling (??) including previous analysis of MaNGA galaxies (?). There is a population of low-mass spiral galaxies with large values of T_{95} , but no equivalent population of high-mass galaxies with small build-up times, highlighting that low-mass spiral galaxies have had more varied histories than their high-mass counterparts, as found by ?.

Using the spatial information available with MaNGA, we are also able to measure how the local value of T_{95} varies with galactic radius R in galaxies of different masses, using the same total stellar mass bins as in Figures 2 and 3. In Figure 5, T_{95} for each spaxel in the sample of spiral galaxies plotted against the galactocentric radius shows that the stellar populations currently at the centres of high-mass galaxies formed on average significantly earlier (by ≈ 0.7 dex or a factor of 5) than those in low-mass galaxies. By contrast, the galaxy’s outskirts built up at approximately the same time regardless of the mass of the host galaxy. At $\approx 1 R_e$, the discrepancy in T_{95} is much less, at ≈ 0.3 dex (or a factor of 2).

To quantify this effect, we obtained the radial profiles of T_{95} for each individual galaxy. We find that these profiles are well-described by a straight line in radius vs. $\log(T_{95})$, so we calculate a best-fit straight line, weighting spaxels by their flux. We find that the majority of galaxies ($> 80\%$) in each mass bin show a negative gradient, as we show in

³ The mass bin thresholds used here were chosen such that a volume-limited sample of spiral galaxies selected in the method described in §3 would contain equal numbers of galaxies in each bin, determined using the “EWEIGHT” sample weighting for the Primary+ MaNGA sample.

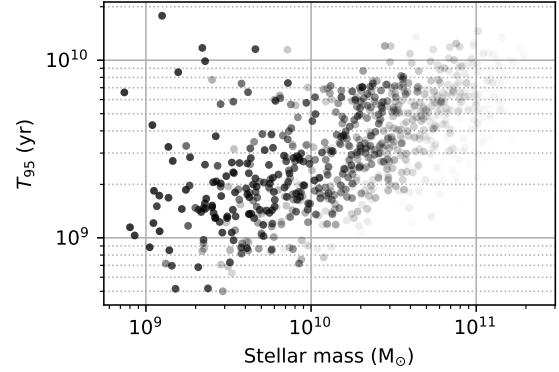


Figure 4. Time since 95% of the total stellar mass within $1.2 R_e$ had been assigned in the STARLIGHT fits (T_{95}) for galaxies of different present-day stellar mass. All spiral galaxies with high present-day mass built the bulk of their mass at early times, but most low-mass galaxies were building their mass more recently. The transparency of each point is defined by the galaxy’s MaNGA Primary+ sample described in §2.1.

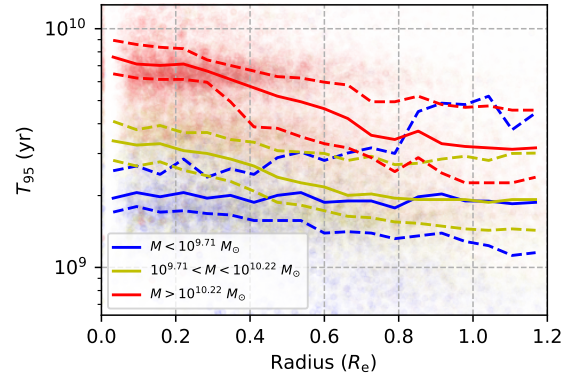


Figure 5. Time since 95% of the stellar mass built up (T_{95}) in each spaxel of each galaxy in the spiral sample. Each spaxel in each galaxy is shown as a point, with the opacity defined by the product of the total spaxel flux and the galaxy’s MaNGA Primary+ sample weight. Colours denote the galaxy’s total present-day mass. Solid lines represent a weighted running median, and dashed lines are one-third and two-third weighted percentiles. The outskirts of galaxies of all masses built up at approximately similar times, but the centres of massive galaxies formed significantly earlier than those of low-mass galaxies. The apparent horizontal feature in the high-mass data points at ~ 7 Gyr is an artefact: there’s a large number of spaxels which have not reached T_{95} by 8.9 Gyr but have by the next oldest SSP at 4.5 Gyr, causing an apparent cluster between these two ages.

Figure 6, implying younger outskirts than galactic centres. Assuming that the stellar populations of any given age have not significantly migrated since their birth, this is evidence for inside-out growth occurring in the great majority of spiral galaxies. We find strongest evidence in the highest-mass galaxies, for which $> 90\%$ exhibit negative gradients in T_{95} . These results are consistent with the mean age gradient analysis of §6, which is not surprising since both approaches are measures of the age distributions contained within the derived SFHs. However, directly determining a quantity such

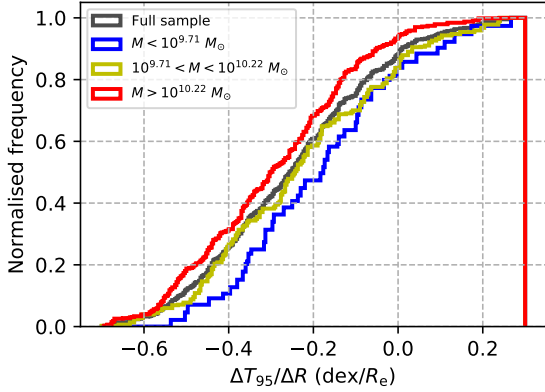


Figure 6. Distribution of gradients of T_{95} vs. galactic radius R for galaxies of different masses. Most galaxies show evidence for inside-out formation, and the effect is strongest in high-mass galaxies.

as T_{95} is returning something much closer to a physical measurement of how the mass of the galaxy has built up over time.

When a 90%, 75% or 50% threshold was used instead of the 95% threshold results shown here, we found no change to the qualitative conclusions. The higher 95% threshold was used to ensure that the buildup time of more galaxies and spaxels was within the range $0.8 \gtrsim T_{95} \gtrsim 5$ Gyr where spectral fitting methods are most sensitive, and avoids saturation at either extreme of the stellar age range we are able to measure.

8 CONCENTRATION OF STELLAR COMPONENTS

Another more physically-motivated way to expand beyond measuring mean age gradients to infer the radial build-up of spiral galaxies is to analyse the spatial extent of individual stellar populations of different ages. We showed in ? that it is possible to measure such distributions directly using time-slicing techniques. The animation in Figure 1 suggests systematic variation in how concentrated the stellar populations are in one particular spiral galaxy. Older populations are most centrally-concentrated in the bulge regions of the galaxy while the younger populations make up the more extended disk. This illustrates the general consensus of the cores of galaxies having younger ages than the surrounding disks.

To quantify the variation in spatial extent of different stellar populations in the full galaxy sample, we choose to measure a concentration of each stellar population in each spiral galaxy. A concentration can be defined in a number of ways [for example as defined by (?)] which often require a larger FOV than MaNGA offers in order to measure a background flux. Here we define the concentration c of a population of stellar age t as

$$c(t) = \frac{\langle m \rangle_{r \leq 0.5 R_e}(t)}{\langle m \rangle_{r \leq 1.2 R_e}(t)} \quad (1)$$

where $\langle m \rangle_{r \leq k R_e}(t)$ is the mean mass contained in all spaxels within $k \times R_e$ using the R_e elliptical Petrosian radius

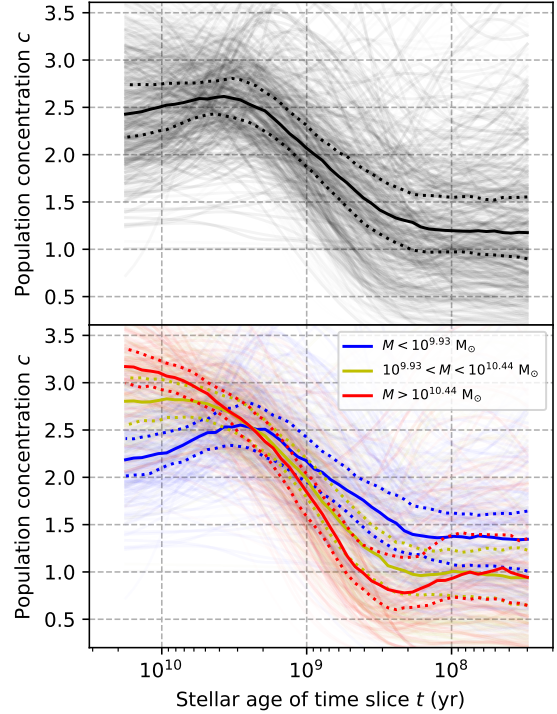


Figure 7. Population concentration c of each galaxy's stellar populations in the spiral sample as a function of stellar age, where c at each time slice is defined by Equation 1. Each galaxy's line is weighted by its MaNGA Primary+ weighting. The heavy line shows the weighted median of all galaxies, and the dashed lines indicate the weighted one-third and two-third percentiles. *Top:* All galaxies. *Bottom:* The same, but with galaxies coloured by their total (present-day) stellar mass. The youngest stellar populations are more spatially extended than the oldest populations in all galaxies, with the effect strongest in higher-mass galaxies.

values of each galaxy from the NASA-Sloan Atlas (?). This measure ensures that the extent of each population is scaled by the size of the present-day galaxy, and only requires data from within the MaNGA footprint. This definition of $c(t)$ also means that a completely uniform (i.e. radially flat) distribution has a value of $c = 1$, with $c < 1$ indicating a distribution which rises with radius in the inner region.

The concentration c of stellar populations of different ages in each galaxy in the full sample is shown in Figure 7. There is a clear trend of older populations being most centrally-concentrated (with typical values of $c \approx 2.5$ at $t \geq 2$ Gyr), and the younger stars in all galaxies exhibiting the most spatially extended distributions (with $c \approx 1$ at $t \leq 0.1$ Gyr). This is unsurprising since this is simply a different way of presenting and interpreting the same effects as in §7, but in a manner that utilises more of the temporal information available to illustrate how radial gradients in mass-to-light ratios (e.g. ?) are created.

We find that there is a strong dependence of $c(t)$ on total (current) galactic stellar mass. Using the same mass bins as in Figures 2 and 3, we find that in the highest-mass galaxies, the oldest ($\gtrsim 6$ Gyr) stellar populations are almost three times more concentrated than the youngest populations ($\lesssim 0.1$ Gyr), while in the lowest-mass galaxies this ratio is less than two.

By repeating this analysis using the mean 4020 Å flux mass in the definition of $c(t)$ (i.e. replacing $m_{r \leq k R_e}(t)$ with $f_{r \leq k R_e}(t)$ in Equation 1), the results are unchanged. This is unsurprising since the radial variation in mass-to-light ratio is unlikely to be significant for any single time slice t .

This analysis reinforces the conclusion that inside-out growth is the primary formation mode in the majority of spiral galaxies, and that the effect is strongest in higher-mass galaxies.

9 MASS-SIZE DISTRIBUTION

Although the mass buildup times in §7 and the variation in concentration in §8 both show evidence for inside-out formation being the dominant growth mechanism in spiral galaxies, these analyses are still not directly comparable measurements to those used in most studies over different redshifts. Previously, observational evidence for inside-out formation in galaxies has come from analysing how the masses and sizes of galaxies increase simultaneously over time, by measuring these properties of different populations at different redshifts (e.g. ?????). This comparison is something that can be directly made using time-slicing methods with integral field spectroscopy for a single galaxy population, to understand how the total mass and size growth has occurred over time.

9.1 Deriving half-light and mass measurements

At each stellar age t , we define the stellar mass to be the sum of the masses in all populations with ages $\geq t$ within $1.2 R_e$, using the temporally-smoothed distribution of weights from STARLIGHT. We can also define a measurement $r_1(t)$ of the light size of a time slice t as being the radius of half the light contained within $1.2 R_e$ (using R_e elliptical Petrosian radius measurements from the NSA) of all of the light emitted by stars older than t . This definition is used since the MaNGA observations are limited in their fields of view. This limitation prohibits us from reliably measuring a sky background, forbidding a direct half-light radius measurement in a normal approach.

To ensure that the radius and stellar mass measurements defined here using the STARLIGHT fits are reliable, we compare these measurements for the present-day galaxy (i.e. $t = 0$) with the known size and mass measurements of the galaxies in the NSA (see Figure 8). We find that r_1 is a good proxy for the NSA elliptical Petrosian half-light measurements, with an offset of ≈ 0.2 dex which is a consequence of both the limited MaNGA FOV and the difference in wavelengths used. (The NSA radii are measured in the r band imagery, but the measurements for the STARLIGHT outputs are done on a model 4020 Å image, which would be located in the g band.) We also find that the total stellar masses determined by STARLIGHT are consistent with the photometry-derived masses in the NSA. Both mass measurements assume the same IMF, so a small observed offset is likely due to MaNGA's limited FOV.

?, ?, ? and others have shown that, unlike the early-type galaxies, the mass-size relation for spiral galaxies is weak. However, using the STARLIGHT-derived measurements

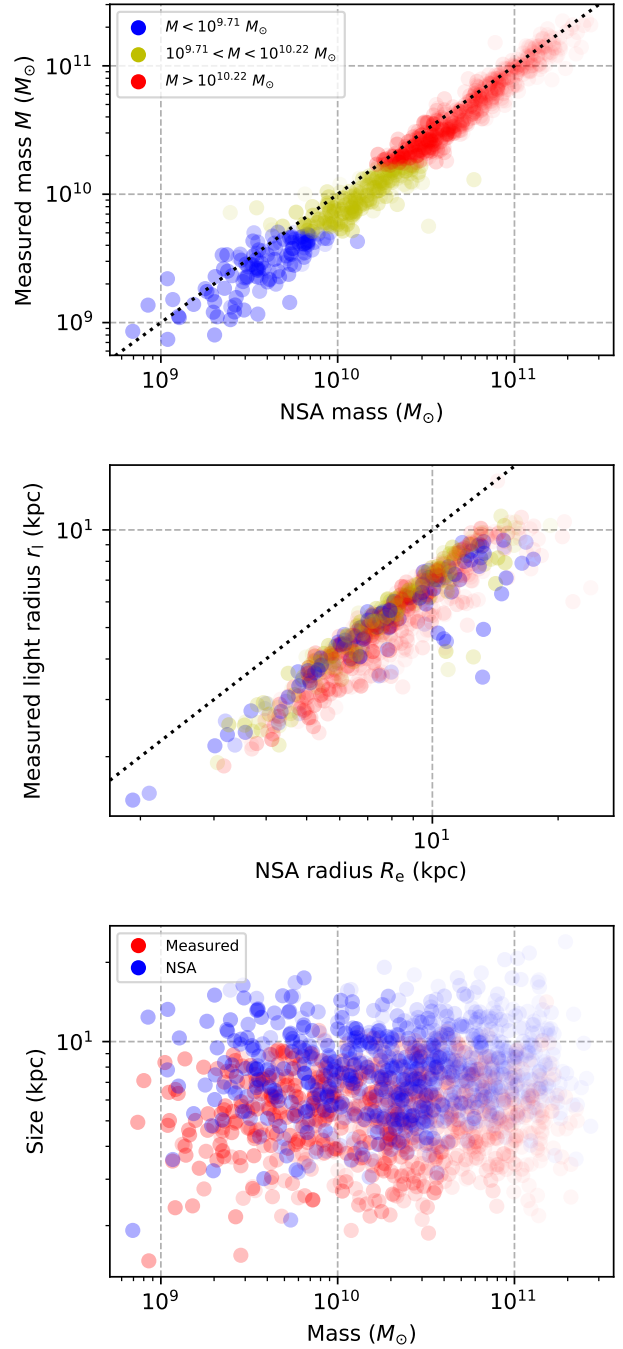


Figure 8. Comparison of the masses (*top*) and radii r_1 (*middle*) of the spiral sample taken from the NASA-Sloan Atlas (?) and measured by STARLIGHT within $1.2 R_e$. The dashed lines indicate equality. Both measurements are consistent with the NSA values, with the offset in measured radius r_1 attributable to the limitations of the MaNGA FOV. *Bottom:* We find no mass-size relation for the sample of spiral galaxies at the present day when using the STARLIGHT- or NSA-measured parameters (red and blue respectively). The transparency of the points indicate the relative Primary+ MaNGA sample weighting for each galaxy.

of the galaxies’ masses and sizes, we find no strong mass–size trend in the present day sample of spiral galaxies at all; a Spearman rank test results in a correlation p -value of only $p = 0.84$ for the measured data, and similar for the NSA values. This lack of a relation may indicate that the Galaxy Zoo classifications for low-mass galaxies under the conservative selection criteria used here may be slightly biased so that the smaller low-mass galaxies are less likely to be classified as spirals.

9.2 Evolution of the mass–half-light-radius plane

Having reassured ourselves that our mass and r_1 radius measurements are appropriate proxies for the photometric measurements in the present-day galaxies, we can now explore how the mass–size plane changes over time. The upper panels of the animation in Figure 9 shows the evolution of the mass– r_1 plane over the last ≈ 10 Gyr. Figure 10 also shows the distribution of galaxies in the mass– r_1 plane at four different redshifts. The measurements shown are using STARLIGHT’s current mass measurements of each SSP template (see §4.2 for the distinction between current and initial mass weights). In reality, the mass loss of each population will have been gradual over the galaxy’s evolution rather than instantaneous as this approach implies. However, as stated in §4.2, by instead adopting the initial mass — and therefore assume that no mass loss occurs at all — we find no significant change to these results. The “reality” would of course be between these two extremes. However, since the two cases reach near-identical results, we present here only the results for the current mass template weightings to avoid uncertainties in modelling time-dependent mass loss estimates separately for each SSP at each time-step.

Assuming an absence of significant systematic radial migration effects, we find that the growth in r_1 of these galaxies has only occurred over the last ≈ 3 Gyr, while the bulk of the growth in mass occurred before this. We also find that galaxies generally have not changed their relative mass group, instead growing in mass and size at the same rates as those of similar masses and sizes. This cohort behaviour implies that, although every galaxy has had a unique formation history, tracing the average evolution of a galaxy population (e.g. by measuring galaxy properties at different redshifts) is representative of how most galaxies have evolved over the same time period.

9.2.1 Mass dependence

By splitting the galaxy into the subsamples of different mass bins as before, we find that over the last 10 Gyr, the low-mass galaxies have grown significantly more in mass (≈ 0.17 dex) but less in light radius r_1 (≈ 0.05 dex) than the high-mass galaxies (≈ 0.14 dex growth in mass, ≈ 0.1 dex in r_1). The small ≈ 0.03 dex difference in mass growth rates between the samples combined with the mass dependence of T_{95} seen in §7 indicates that the low-mass galaxies have only built up slightly more mass relative to the high-mass galaxies, but that this growth occurred later. We also see this downsizing effect in the “turnup” time — at which galaxies stop growing significantly in mass and start growing in light radius r_1 — which occurred earlier in high-mass than low-mass galaxies (≈ 3.5 Gyr ago compared to ≈ 1 Gyr ago).

9.3 Evolution of the mass–half-mass-radius plane

While the light-weighted radius measurements are directly comparable to the size evolution of galaxy populations observed at different redshifts, the mass distribution of a galaxy is more fundamental to its build-up. In Figures 9 and 10, we therefore also show the evolution of the mass–size plane but using a half-mass size r_m (equivalently defined as the radius containing half of the stellar mass within $1.2 R_e$ due to the limitations of the MaNGA FOV) using the SSP template mass-to-light ratios. We find that despite increases in the observed light size r_1 of the galaxy population, the corresponding increase in the mass size r_m of the same galaxies is minimal; we find an increase of $\lesssim 0.05$ dex in size for almost all galaxies, even in those with low present-day stellar masses. This weak evolution is in agreement with the results presented by ?? using entirely independent approach to show that the half mass radius does not evolve significantly compared to the evolution of the half light radius.

The physical size growth of spiral galaxies over the last 10 Gyr has therefore been extremely small, at typically only 10% growth. Such an increase in mass radius — however slight — requires a radial increase in the regions of ongoing star formation. Since younger stellar populations dominate the light of a spiral galaxy at any time slice or lookback time, the increase in measured radius in observations of the same galaxies becomes significant. A small amount of star formation in the outskirts of the galaxies will contribute a large amount to the light while contributing comparatively little to the bulk of the galaxy, causing a strong mass-to-light gradient. Direct measurements of the growth of galaxies from observations therefore produce an overestimate of the underlying mass growth rate. This effect has also been recently quantified for cosmological galaxy catalogs from CANDELS (??), who showed that the half-light radius growth of galaxies, both star-forming and quiescent, previously reported in many works is significantly weaker for the half-mass radius.

Interestingly, it has been reported by ? that the structure of stellar populations seen in the Milky Way provide evidence for a slower growth in half-mass radius than in the half-light radius, and the evidence presented here — as well as from high-redshift surveys (see above) — suggests that this feature is common in the growth of spiral galaxies. This slow size growth of spiral galaxies seems to be in tension with predictions from semi-analytical models and hydrodynamics simulations of galaxy evolution in the context of the Λ CDM cosmology (see for a discussion ? and more references therein).

10 POTENTIAL LIMITATIONS

10.1 Limitations of the data

Due to the limited FOV of the MaNGA observations, we are unable to measure a true half-light (or half-mass) radius for any given “time slice”, since we do not have any background in the images. We are able to confirm in Figure 8 that the radius of half of the light (or mass) contained within $1.2 R_e$ is a good proxy for the present-day galaxy, but we have no way of confirming this at other stellar population ages. However, since we find that the oldest populations are most concentrated, the measured sizes in the earlier age-steps in

Figure 9. Evolution of the mass–size plane of the spiral sample over time. The lookback time and corresponding redshift are indicated at the top of the figure. The left column shows the mass–size plane for the light radius (r_1 , *top*) and mass radius (r_m , *bottom*) measurements. The right column indicates the overall change in each galaxy’s mass and size (in dex) from the first frame of the animation. The redshift of each galaxy is accounted for, such that in any given frame the star-formation histories of each galaxy is sampled at the difference between the frame age and the lookback time implied by the galaxy’s redshift.

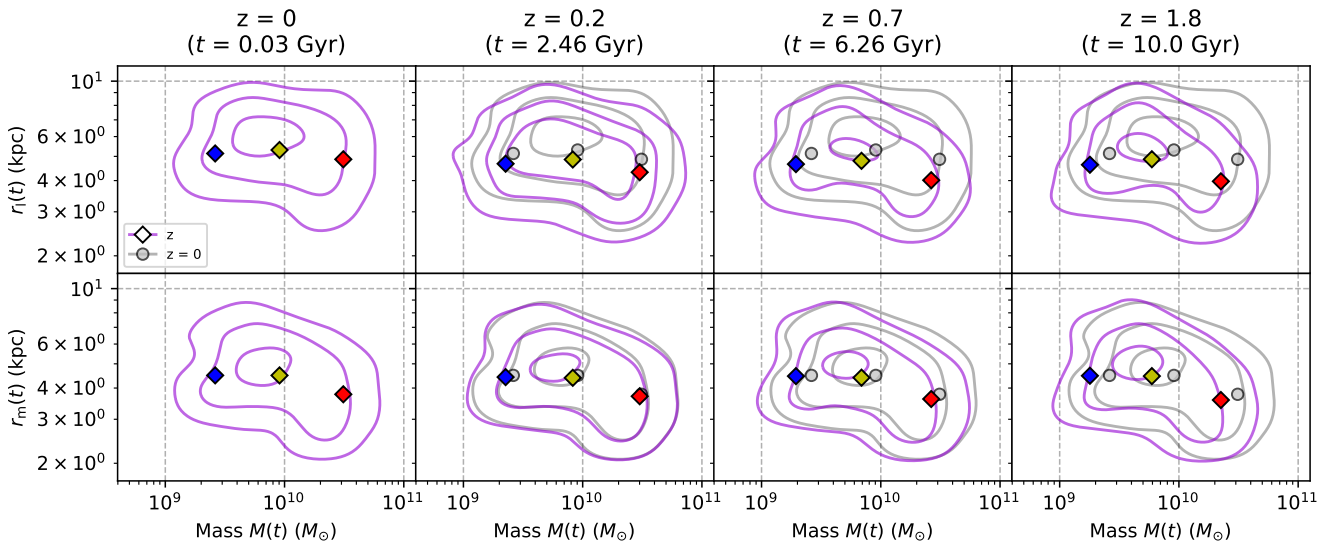


Figure 10. The distributions of galaxies at individual snapshots in Figure 9 showing the mass– r_1 (*upper panels*) and mass– r_m (*lower panels*) planes at selected redshifts z . The corresponding lookback times t are also indicated. As fiducials, the grey contours and circle markers indicate the distribution of galaxies and the mean positions of each mass bin at $z = 0$, while the magenta contours and coloured diamond points indicate the distribution and mean positions at each redshift’s time slice.

the mass–size evolution are likely to be closer to the true sizes. The observed increase in size is therefore a conservative estimate of the real change. Since little of a galaxy’s mass is located outside $1 R_e$ (e.g. ?), we expect that the mass radius r_m measurements are likely to be close to true half-mass radii.

10.2 Stellar population models and spectral fitting

Although we show in Appendix A that STARLIGHT can measure stellar populations if the models used to do so are correct, this work assumes that the model spectrum templates of the E-MILES (?) and ? libraries are representative of the true observed stellar populations. There are a number of unresolved problems in the field of stellar population modelling; see ? for a comprehensive review. For example, in §4.5, we described a correlation between weights assigned to populations $\gtrsim 10^{9.5}$ years and those of $\lesssim 10^{7.2}$ years due to a deficiency in the SSP templates. This is likely to be related to the UV upturn problem, due to the presence of hot stars in old stellar populations (?) which is not accounted for in the SSP models. There is also uncertainty surrounding the shape of the IMF and ongoing debate on whether it varies between and within galaxies (????). In principle, any variation of the IMF over cosmic time is likely to affect our analysis too.

We also assume here that stellar metallicity is a one-dimensional parameter. In reality, the individual elemental abundances can vary from star to star. Further time-slicing work can be done to measure the simultaneous change in star-formation histories and metallicity evolution, including variation in α -enhanced metals, but this is beyond the scope of this project. Although we are confident that the fitting methods used here can recover the distributions of stellar population ages and metallicities, the degeneracy between metallicity and $[\alpha/\text{Fe}]$ is harder to assess. However, in ?, we found that removing the extra metallicity dimension appears to have little effect on the derived star-formation history.

This work has also assumed a single ? extinction model which affects every stellar population contained within a single spectrum equally. As we state in §4, we expect that younger stellar populations are instead likely to be affected by a greater amount of extinction, but we are unable to resolve this difference in non-parametric fitting using STARLIGHT. How this deficiency affects the measured star-formation histories is not known.

Notwithstanding these shortcomings and assumptions used in the fitting process, the resulting star-formation histories tell a consistent story of inside-out formation in spiral galaxies with no noticeable artifacts, and the coherent structures visible in the time-slicing of galaxy 8329-12701 shown in Figure 1 gives confidence in the fitting method for the purposes described here. The clear inside-out formation reported here might even be underestimated: ? have recently shown that any intrinsic signature of inside-out growth is diminished by the instrumental/observational setting and the stellar population modelling, mainly the age resolution of the SSP templates.

10.3 Effects of radial mixing and mergers

Time-slicing methods can only reveal the current locations of different stellar populations in a galaxy. In this work, we have interpreted these present-day distributions to be indicative of the radial distributions of star formation at the age of the stellar population, and make no attempt to correct for the effects of radial migration or mergers. Fortunately, simulations suggest that the radial distributions of stellar populations in a galactic disk are not significantly altered by radial migration (???), indicating that the assumptions made here are at least approximately valid.

High-resolution simulations of Milky Way-like galaxies show that radial migration has no preferential direction, with most stars being scattered similarly inwards and outwards, by typically no more than 1–2 kpc (?). Instead, stars are equally likely to move in either direction over their lives (??), with observations implying that any resulting observed growth as a result of migration occurs slower than the intrinsic growth of the disk (?). Any radial migration of an initially centrally-concentrated distribution of stars is likely to become slightly less concentrated over time, an effect which is observed in the stellar metallicity distributions of the solar neighbourhood in the Milky Way (e.g. ??). A galaxy with a radial distribution of star formation that is not varying over time would be observed using time-slicing methods to have been slightly decreasing in measured radius over the same time frame, since the oldest populations will have more time to disperse and would therefore appear at larger radii. The measured variations of spatial distribution of stellar populations of different ages in §8 are also therefore likely to be a close lower limit on the true variation of the sizes of spiral galaxies over the same time period. Similarly, the recovered change in light size r_1 in §9 is therefore a slightly conservative but representative estimate of how the galaxy evolved over the same time period.

11 CONCLUSIONS

We have derived spatially-resolved star formation histories for a sample of 795 low-redshift spiral galaxies using STARLIGHT applied to integral-field spectroscopic observations from SDSS-IV MaNGA. From this fossil record analysis, we have built maps indicating the regions in which stellar populations of different ages are located in any given galaxy. We analysed the radial profiles of these “time slices” to extract the historical growth of the population of spiral galaxies. The main findings are:

- Using E-MILES single stellar population template spectra, the star formation histories measured by STARLIGHT are unreliable for the youngest populations used in the fit (in this case those younger than 3×10^7 years). We found evidence that this is related to the UV upturn (?) and a solution to this problem requires population models to include the presence of hot old stars (whatever their nature) in the oldest population templates. However, despite this degeneracy between the oldest and youngest template weights, the derived star-formation histories of the stellar populations older than 3×10^7 years are trustworthy.
- We have quantified evidence for inside-out galaxy growth in three different ways, which all indicate that such a

growth mode is dominant in the majority of spiral galaxies, and is most significant in high-mass galaxies:

- The mass-weighted mean age gradient of spiral galaxies tends to be slightly negative; the outskirts are younger than the centres in $\approx 60\%$ of all spiral galaxies. This fraction rises to 80% for galaxies with stellar mass $M > 10^{10.22} M_{\odot}$.
 - By measuring a time T_{95} by which 95% of the stellar mass had built up in each location of the galaxy, we find that T_{95} decreases with radius in the majority galaxies. Gradients in T_{95} are steepest in the highest-mass galaxies.
 - The concentration c of each “time slice” was found for each galaxy. The youngest stellar populations (younger than $\approx 10^{8.5}$ years) are more radially extended than the oldest ($\approx 10^{10}$ years old) populations in all cases, and this effect is most significant in high-mass galaxies.
- By considering the simultaneous increase in stellar mass and the increase in light radius with the addition of ever-younger stellar populations, we found that the mass–size distribution of spiral galaxies evolves with very little change in rank; galaxies grow in mass and size at similar rates to other galaxies with similar masses and sizes. This suggests that a “like for like” approach when comparing the sizes and masses of distributions galaxies at different redshifts is representative of how the individual galaxies themselves have evolved.
 - We found that over the last 10 Gyr, galaxies with high present-day stellar masses have grown their half-light size by approximately twice the amount that low-mass galaxies have, although low-mass galaxies have grown slightly more in mass.
 - However, when the half-mass radius of the galaxies was used instead, we found that spiral galaxies have barely altered their radial mass distributions over the same time period. Although galaxies appear to grow in (light) size over cosmic time, we show that this is an overestimate of their actual physical growth. This apparent discrepancy is due to a small amount of star formation occurring in the outskirts being able to dominate a galaxy’s light while contributing very little to the physical bulk of the galaxy.

ACKNOWLEDGEMENTS

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für As-

tronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

We are grateful for access to the University of Nottingham’s **Augusta** high performance computing facility.

J.H.K. acknowledges financial support from the European Union’s Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 721463 to the SUNDIAL ITN network, from the State Research Agency (AEI) of the Spanish Ministry of Science, Innovation and Universities (MCIU) and the European Regional Development Fund (FEDER) under the grant with reference AYA2016-76219-P, from IAC project P/300724, financed by the Ministry of Science, Innovation and Universities, through the State Budget and by the Canary Islands Department of Economy, Knowledge and Employment, through the Regional Budget of the Autonomous Community, and from the Fundación BBVA under its 2017 programme of assistance to scientific research groups, for the project “Using machine-learning techniques to drag galaxies from the noise in deep imaging”.

APPENDIX A: TESTING FULL-SPECTRUM FITTING

It is common to spatially bin neighbouring spaxels of integral-field spectroscopic data to create regions with a minimum signal-to-noise ratio (SNR) before fitting. However, since we wish to retain and measure the full spatial information of the stellar populations — and therefore fit each spectrum independently instead of binning — we require the STARLIGHT fits in regions with low SNR to be reliable. ? showed that STARLIGHT may exhibit bias in the fitting of spectra with low SNR, but ? contends that these effects are not significant in most physical applications and with the robust STARLIGHT configuration used here. To assess this conclusion, we have performed a series of tests laid out here.

A1 Average fits of regions with low signal-to-noise

To test how the signal-to-noise ratio (SNR) of a spectrum in a MaNGA datacube may affect the fitting results from STARLIGHT, we combined spectra of a single galaxy (plate-IFU 8329-12701) within different SNR bins to form a single integrated spectrum for each bin. In combining spectra from the MaNGA datacube, emission lines were removed and the

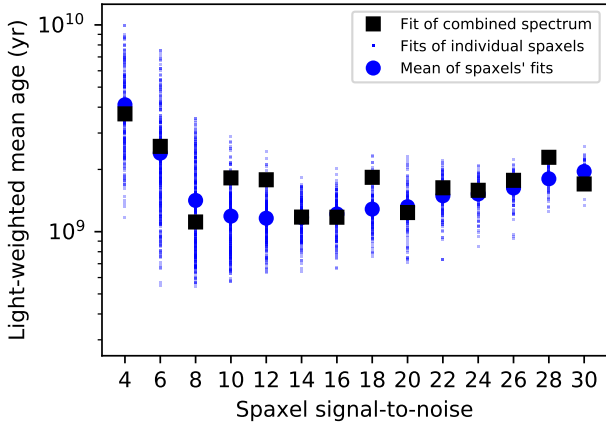


Figure A1. The mean age of spaxels in different signal-to-noise ratio bins (*blue*) in a spiral galaxy compared to the mean age of the spectrum of all spaxels combined (*black squares*). There is no significant bias in the average age measured by STARLIGHT compared with varying SNR.

spaxel spectra de-redshifted and interpolated onto a common wavelength base before summing. Each single spaxel’s SNR was then defined as the median value over the fitting wavelength range of the ratio between the spaxel’s flux spectrum and the reciprocal of the square root of the inverse variance spectrum (as measured by the MaNGA DRP).

We chose to combine spaxel spectra in SNR bins of width 2, centred on every even value. A single spectrum was created by combining all spectra from spaxels with SNR between 3 and 5, another from spaxels with SNR between 5 and 7, etc., up to a spectrum comprising the sum of all spaxels with a SNR between 29 and 31. Each of the combined spectra’s signal-to-noise ratio is greater than 60 and most are greater than ~ 200 . These combined spectra were then fit using STARLIGHT with an identical configuration to that of the science fitting to see how their measured star-formation histories varied from the average of their constituent parts. In the absence of any systematic bias in STARLIGHT, it would be expected that the average SFH measured in all individually-fitted spaxels in a given SNR bin should be the same as the SFH measured from the average spectrum of those spaxels, regardless of the actual variation in SFH shape between spaxels in a given bin.

We find that the light-weighted mean age of the summed spectrum is always within ~ 0.2 dex of the mean age of all individual spaxels (separately fitted) in a given SNR bin, indicating that signal-to-noise effects do not significantly bias the average results (see Figure A1). However, we find that the full star-formation histories of the summed spectra are most discrepant for the bins of larger SNR, as shown in Figure A2. This is likely due to an effect of small systematics (e.g. sky subtraction or flux calibration) dominating over random noise when summing spectra of already-high signal-to-noise ratios. In summing high-SNR spectra, the modest reduction in combined SNR is outweighed by the increase in systematic errors when considering the fine detail required to measure a SFH. The fact that these effects have less significance in measuring the average properties highlights the

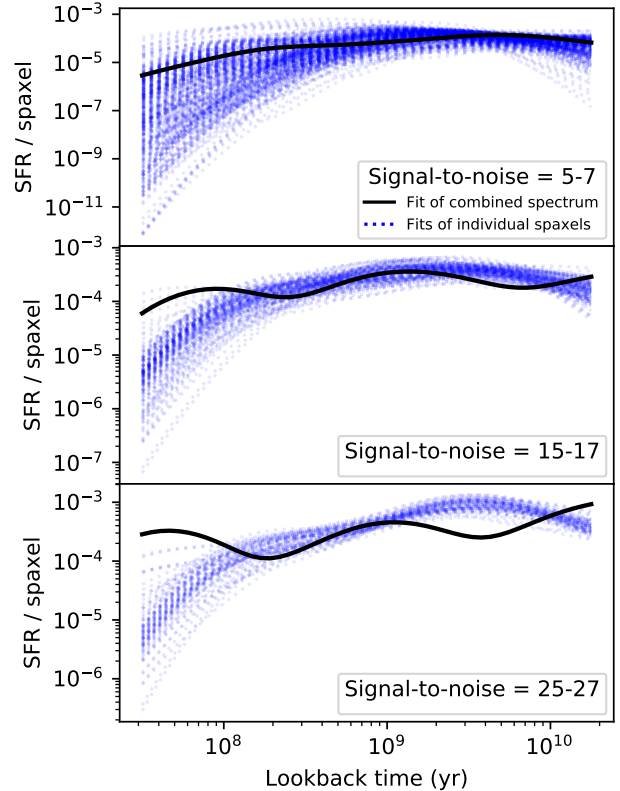


Figure A2. The star-formation histories of spaxels (*blue*) in three different signal-to-noise ratio bins compared to the SFH of the spectrum of all spaxels combined (*black*). STARLIGHT shows worse performance at higher signal-to-noise ratios.

level of extra complexity involved in measuring SFHs over mean ages.

A2 Recovery of a single stellar population of known age

To further examine whether the above effects are due to systematics in the spectra rather than in STARLIGHT, we tested how well STARLIGHT is able to return the age and metallicity of a single stellar population with known parameters. We can create spectra representing single stellar populations of any age and metallicity by interpolating over the grid of E-MILES SSP template spectra. We produced spectra representing 200 ages and three metallicities using a bilinear interpolation in 2D log space of the 66 SSPs used in the fitting. We then degraded these spectra to signal-to-noise ratios of 5, 10, 15, and 20 by adding Gaussian noise, and also applied a dust extinction with $A_V = 0.2$ using a τ extinction curve. We blur these 2400 individual spectra to the MaNGA LSF, and then applied STARLIGHT using the same SSPs and STARLIGHT configuration as described in the main text to compare how well the populations are recovered under different circumstances.

The 0.3-dex temporally-smoothed distributions of mass weights measured from the STARLIGHT output of these known SSPs is shown in Figure A3 for a signal-to-noise ratios of 5 and $Z = Z_\odot$. Results for a signal-to-noise ratio of

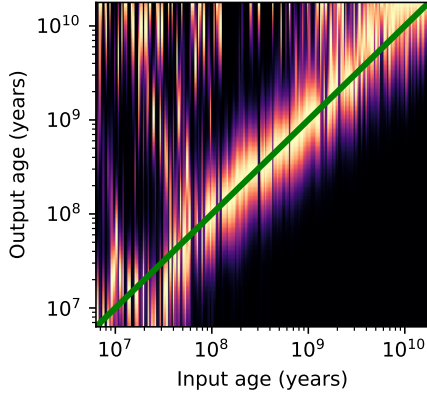


Figure A3. Distributions of the measured mass weights given by STARLIGHT in fits of stellar populations of known age, for a signal-to-noise ratio of 5 and $A_V = 0.2$. Input spectra are interpolated from the grid of E-MILES SSPs at $Z = Z_\odot$. The recovered weight distributions are smoothed by 0.3 dex in age, and the green line indicates equality between input and output ages. The noise in the weights assigned to templates older than the input SSP age is due to the STARLIGHT configuration used: specifically, we force 30% of all templates to be assigned a *flux* weight for consistency with the main text. This small amount of noise in flux weights is then amplified in the conversion to mass weights.

10, 15, or 20 were not noticeably improved, and providing an input metallicity of $Z = 10^{-0.625} Z_\odot$ or $Z = 10^{-1.25} Z_\odot$ did not change the results. STARLIGHT is able to recover the mean age of input stellar populations of all ages older than $\approx 10^8$ years even with an input spectrum signal-to-noise ratio of 5, highlighting the diagnostic power of using such a long wavelength range to model the large-scale continuum in fitting.

A tendency for STARLIGHT to also include small levels of older populations is highlighted. This is a consequence of the combination of a strong trend in mass-to-light ratio with stellar population age and the robust STARLIGHT configuration used. As STARLIGHT fits the input spectrum (i.e. in light space), we force it to assign a weight to at least 30% of all SSP spectra to ensure full SFH recovery in science cases. In the case of a single input stellar population, this will result in small spurious weights being given to other templates, and this noise becomes amplified in old populations when considering mass weights.

It is not necessarily clear how much the boundary between the two template libraries affects STARLIGHT’s inability to recover stellar populations younger than $\approx 10^{7.5}$ years, or whether it is purely due to the lack of diagnostic spectral information at these ages.

A2.1 Effects of kinematics and dust

To test whether these results are improved when STARLIGHT does not also have to model the kinematics of the input spectrum, we repeated these test, but with the velocity and dispersion fixed to the known input values. The results were entirely unchanged.

We also performed these same tests with $A_V = 0$ and $A_V = 0.8$ (instead of $A_V = 0.2$ as used above) to check that STARLIGHT is still able to recover populations in low- and

high-extinction environments, and found the results to be unchanged here too.

A3 Recovery of a known star-formation history

Finally, to simulate the effects of low SNR on a spectrum comprising multiple stellar populations (such as is found in real galaxies), we created spectra of three different star-formation histories (SFHs) using the template SSPs. The different SFHs reflect different cases:

- A: a flat SFH, where the star-formation rate is defined as $SFR(t) = 0.1 M_\odot \text{yr}^{-1}$ for all lookback times t ;
- B: a peaked and then declining SFH (as seen in many galaxies) with a star-formation rate represented by $SFR(t) = [0.2 + \mathcal{N}_{9.2}^{1.5}(\log(t))] M_\odot \text{yr}^{-1}$, where $\mathcal{N}_\mu^\sigma(x)$ denotes a Gaussian function of x centred on $x = \mu$ with standard deviation of σ dex;
- C: a declining and rejuvenating SFH with a star-formation history represented by $SFR(t) = [1.2 - \mathcal{N}_{8.1}^{0.8}(\log(t))] M_\odot \text{yr}^{-1}$.

In building these SFHs, we assign weights to each SSP assuming that they represent all star-formation between their nominal age and the next-youngest SSP.

We also included two different metallicity distributions, neither of which vary with stellar population age:

- X: a flat distribution over the range of metallicities in the E-MILES templates, where the relative flux of each SSP of any given age is defined by $F_i = 1$;
- Y: a peaked distribution where the relative flux of each SSP is defined at each age t by $F_i = \mathcal{N}_{-0.71}^{0.3}(\log(Z_i/Z_\odot))$.

These spectra were then degraded to different SNRs of $S/N = 3, 5, 7, 10, 15, 20$, and 30. When each of these 42 spectra of known SFHs were fit using STARLIGHT in the same configuration used in the main text, we found the general shape of $SFR(t)$ is recovered in all cases, as shown in Figure A4. For a SNR greater than 5 (the lowest typically found in the outskirts of MaNGA galaxies in the Primary or Primary+ samples; see ?), the derived SFHs show very good agreement (with a value of $SFR(t)$ within 0.15 dex of the known value at all lookback times t) to the input SFHs. This is also shown in Figure A5, where a simple χ^2 goodness-of-fit measurement is obtained between the input SFH and each of the recovered SFHs. STARLIGHT is able to recover SFHs A and B for all cases with $S/N \geq 10$, and while the measurement of SFH C is worse, it is recovered equally well for $S/N \geq 7$ in metallicity distribution X and $S/N \geq 15$ for metallicity distribution Y.

A4 Implications

We assume throughout this work that the E-MILES model spectra are accurate representations of the stellar populations they represent. A full test of whether this is indeed the case (as conducted by e.g. ?) is beyond this study, but the tests shown here imply that if this is true, we expect STARLIGHT to be able to recover the true SFHs under all the conditions analysed in this work. In fact, we find evidence that fitting spaxels individually — rather than summing spectra from neighbouring spaxels — may be the most

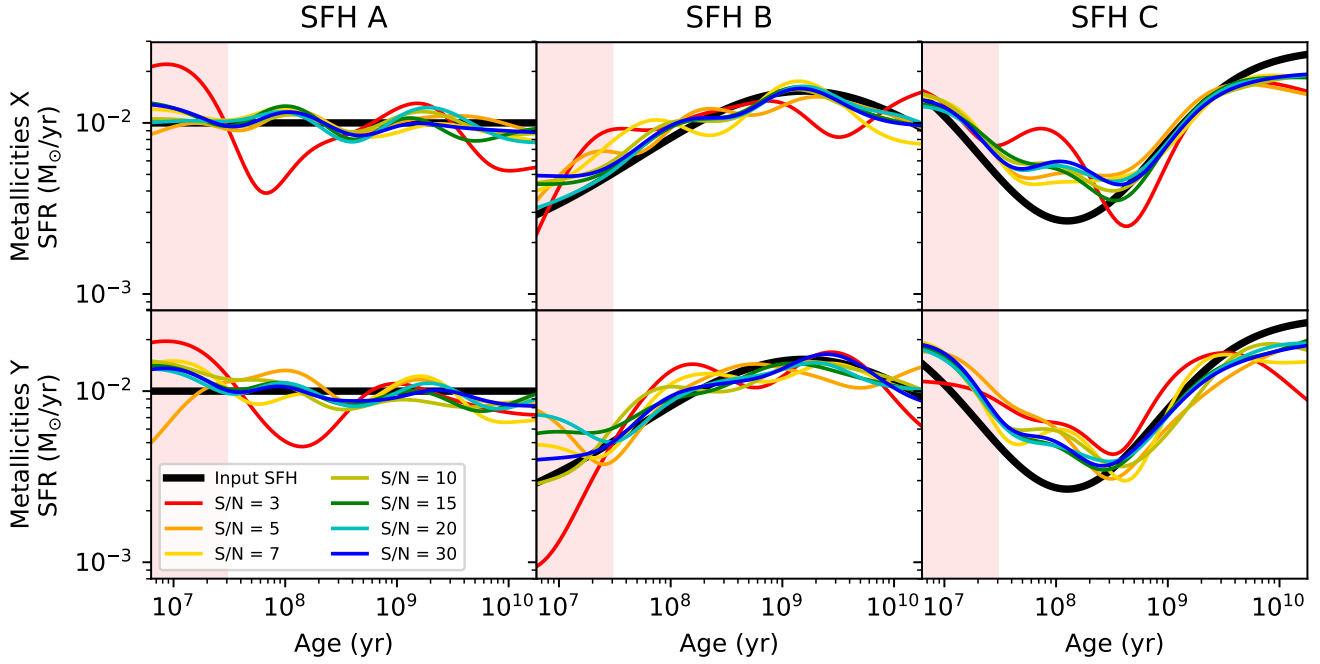


Figure A4. The measured SFHs for different input spectrum signal-to-noise ratios (coloured lines) compared to the input SFH shape (black line) for each SFH (*left to right*) and metallicity distribution (*top and bottom*). Recovered SFHs are smoothed by 0.3 dex in age. The red-shaded region indicates that for which SSP weights are ignored in science cases (see main text). The general shape is recovered well in all cases, particularly for signal-to-noise ratios greater than 5.

robust approach to avoid the dominance of systematics from compromising the ability to measure a SFH. Notwithstanding this robustness, to ensure that the low signal-to-noise regions of the galaxy are not affecting our results in ways we don't anticipate, in all stages of the science analysis shown in this work, we ensure that we weight spaxels by their flux or mass, ensuring that the central regions (with higher SNR and therefore with probably good fits) are emphasised, and low signal-to-noise regions are down-weighted.

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

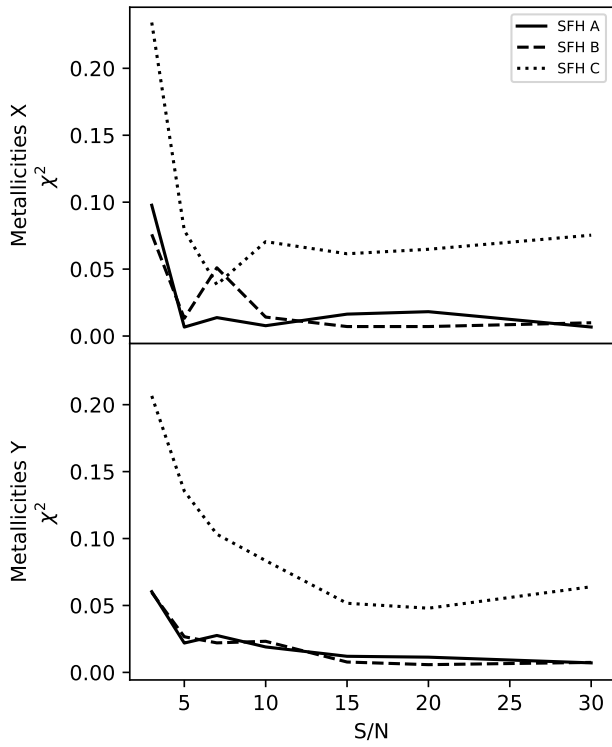


Figure A5. A χ^2 goodness-of-fit measurement of the input to the measured SFH in Figure A4, for each of the three SFH shapes (line styles) and metallicity distributions (*top* and *bottom*) for different input signal-to-noise ratios. Increasing S/N above 7 does very little to improve STARLIGHT's ability to recover the SFH.