

Stellar Populations and Dynamics of Nearby Early-Type Galaxies

By

Christina Baldwin

A thesis submitted to Macquarie University
for the degree of Doctor of Philosophy
Department of Physics & Astronomy
October 2019



MACQUARIE
University
SYDNEY • AUSTRALIA

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Christina Baldwin

Acknowledgements

Firstly, thanks to Richard, without whom this PhD wouldn't have happened, and I'd probably be off swimming in money somewhere. Thanks to Lee Spitler for the (wise?) advice, and to Adrian and Bron for always being there to "discuss science". To all my fellow PhD ~~sufferers~~ comrades - I probably couldn't have done it without you. Thanks to Dylan for cooking me food on those rare occasions I was actually working, Guff for insisting on being my best friend despite the position already being filled, and RFL for being an RFL without ever really R-ing. To Tiffany and Stefania for things I'm not going to explicitly mention in my acknowledgements. To New Shane for lighting up my days with his frown. To Old Shane, for finally vacating the office. To Reece and Glen and Woody, because I feel like they should get a mention. Most especially, thanks to Dean, who will probably insist on being acknowledged despite leaving the country and never getting in touch again.

Thanks also to my family, for supporting my blind quest for those extra two letters in front of my name. Thanks especially to Rob and Lisa for letting me intrude these last few months. To The Legion - now that I'm finished, we can finally get back to the Timeo! Thanks to Andrew, for giving me an excuse to finally go to the Maldives, and Vicky for not being pregnant yet. Alex and My, you're still ok.

Serious thanks go to the following people, who have worked with me over the last four years. Thanks to Charlie Conroy and Meng Gu for discussions on spectral fitting (and to Charlie for providing updated FSPS models to test). Thanks to Gary Mamon for welcoming me at the IAP and sharing with me the crazy amount of science you're doing. Extra special thanks go to Michele Cappellari for teaching me all about dynamical modelling and how to eat pizza correctly, and for insisting I switch to Python; to Claudia Maraston for being so warm, and enlightening me on everything stellar populations; and to Nic Scott for that time we almost worked together on MGEs.

Abstract

In this thesis, I use stellar population and dynamical modelling techniques to study the star formation histories and mass distributions of early-type galaxies. First, I make a detailed comparison of four commonly used stellar population synthesis codes, which have been shown to give discrepant results when applied to near-infrared data. This has commonly been attributed to different prescriptions for the thermally-pulsing asymptotic giant branch phase in the various models. In this work, however, I show that the discrepancies extend beyond this: applying the four models to high quality optical and near-infrared spectroscopy of 12 fast-rotating early-type galaxies, I find the largest differences between models is due instead to the choice of stellar spectral library. Models including updated, high resolution stellar libraries are the most self-consistent when comparing optically-derived properties with near-infrared ones. Second, I study the effects of incorporating stellar population information into dynamical models, using high signal-to-noise MUSE spectroscopy of three fast-rotating early type galaxies. Specifically, I include spatial variation in the stellar mass-to-light ratio (M/L) due to gradients in age, metallicity and stellar initial mass function (IMF). I show that age and metallicity gradients have little effect on dynamically derived quantities, however inclusion of IMF gradients of the strength observed by van Dokkum et al. (2017) increase the derived dark matter fraction within an effective radius by 50%. Finally, I perform a completely self-consistent stellar population and dynamical modelling analysis using high quality MUSE data. I measure gradients in age, metallicity and IMF directly from the same data from which I extract kinematics. I find similar M/L gradients to other stellar population-based IMF studies, however dynamical models which include these measured M/L values cannot reproduce the observed kinematics. If the stellar mass is allowed to be rescaled, it is consistently scaled downward, and the fits to the kinematics become acceptable, indicating that the M/L gradient is not incompatible with the data, however the absolute normalisation of the M/L from the stellar population models is higher than allowed by the kinematics. Overall, this thesis highlights the importance of self-consistent studies, which accurately account for systematic errors with a thorough exploration of the techniques and models used.

Contents

Acknowledgements	v
Abstract	vii
Contents	ix
1 Introduction	1
1.1 Early-type galaxies	2
1.2 Stellar population synthesis	6
1.2.1 Stellar initial mass function	6
1.2.2 Stellar evolution calculations	7
1.2.3 Stellar Libraries	8
1.3 Outline and aims of this thesis	8
2 Comparison of Stellar Population Model Predictions Using Optical and Infrared Spectroscopy	11
2.1 Introduction	11
2.2 Sample	13
2.3 Observations and Data Reduction	14
2.3.1 Near-Infrared Spectroscopy	14
2.3.2 Optical Spectroscopy	17
2.4 Stellar Population Synthesis	17
2.4.1 Stellar Population Models	17
2.4.2 Star formation histories from full spectral fitting	19
2.4.3 Line strengths	21
2.5 Results	22
2.5.1 Quality of the fits	22
2.5.2 Comparison of Star Formation Histories	24
2.5.3 Model Ingredients	29
2.5.4 Line strengths	32
2.6 Discussion	34
2.6.1 Should optical and near-infrared properties agree?	34
2.6.2 Why does the near-infrared give such inconsistent results?	36
2.6.3 Can we constrain treatment of the TP-AGB?	37
2.6.4 Comparison with previous studies	37
2.7 Conclusion	40
3 Probing Variations of the Initial Mass Function with MUSE: Observations, Data Reduction and Initial Analysis	43
3.1 Introduction	43
3.2 Sample, Observations and Data Reduction	44

3.3	Analysis: Stellar Kinematics	49
3.3.1	Voronoi Binning	49
3.3.2	Kinematic Extraction	50
3.3.3	Comparison with ATLAS ^{3D} kinematics	52
3.4	Analysis: Stellar Populations	53
3.4.1	Full spectral fitting	54
3.5	Conclusions	56
4	Combining stellar population M/L variations and dynamical modelling	57
4.1	Introduction	57
4.2	Jeans Anisotropic Modelling	58
4.3	Building the mass model	59
4.3.1	Multi-Gaussian Expansion (MGE) Method	59
4.3.2	Luminosity density model	59
4.3.3	Conversion from luminosity to mass	60
4.3.4	Dark matter mass model	64
4.4	Summary of models used	64
4.4.1	Optimisation	65
4.5	Results: Stellar Mass Profiles	66
4.6	Results: JAM Dynamical Models	67
4.6.1	Benchmark : constant M/L models	70
4.6.2	Varying M/L cases	73
4.7	Discussion	78
4.7.1	Effect of M/L gradients on dynamical properties	78
4.7.2	Constraining IMF variation and normalisation with kinematics	79
4.8	Conclusion	80
5	Testing spatial IMF variations with self-consistent dynamical modelling	81
5.1	Introduction	81
5.2	Full spectral fitting using PyStaff	82
5.2.1	Monte Carlo Simulations	84
5.3	Recovering the IMF from spatially resolved IFU data	85
5.3.1	Radial profiles	88
5.4	Utilising our measured M/L in dynamical models	89
5.4.1	Building self-consistent dynamical models	90
5.4.2	Rescaling the stellar mass	91
5.5	Discussion	98
5.6	Conclusion	99
6	Conclusion	101
6.1	Near-infrared stellar population synthesis models	101
6.2	IMF variation in early-type galaxies: incorporating stellar population gradients into dynamical models	102
6.3	IMF variation in early-type galaxies: fully self consistent stellar population and dynamical modelling	103
A	Appendix A: Tabulated optical and near-infrared star formation histories, and near-infrared line strength indices from Chapter 2	105
B	Appendix B: Inferences on MUSE galaxy sample for the various dynamical models from Chapter 4	109

C Appendix C: Inferences on the MUSE galaxy sample using PyStaff	121
References	133

1

Introduction

The standard cosmological paradigm of hierarchical growth of dark matter halos has been very successful at explaining the large scale structure of the Universe. In the Λ -CDM cosmology, following the Big Bang, quantum density fluctuations expanded to macroscopic size during the period of rapid expansion known as “inflation” (Guth, 1981). Within these over-dense regions, cold dark matter condensed into bound structures (Figure 1.1). Inside these immense structures, gas, acting under gravity, collapsed to form stars, making up the galaxies we see scattered throughout the Universe (White & Rees, 1978). Over time, the dark matter halos merged, forming the large scale filaments and voids which are the scaffold upon which galaxy formation takes place (Saunders et al., 1991, Shectman et al., 1996).

Numerical simulations of dark matter alone are able to well reproduce the large scale structure of the Universe, as observed through the distribution of galaxies. However, tensions arise on smaller scales due to the complications of baryonic physics. For example, the apparent “downsizing” of galaxy populations (Cowie et al., 1996), whereby the most massive galaxies are the oldest, is in conflict with the naive expectation of a hierarchical universe. In the Λ -CDM framework, dark matter halos are built bottom-up, such that the most massive halos assemble last. Within this context, one would expect the most massive galaxies to form from the mergers of smaller objects, leading to young massive objects at low redshift. To resolve this issue, simulations typically invoke

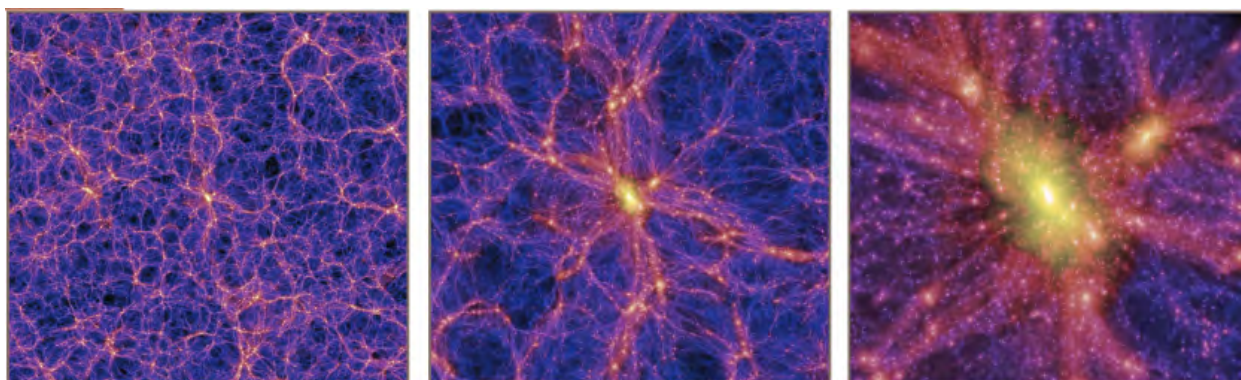


Figure 1.1: The large scale structure of the Universe, as predicted by numerical simulations. The brighter the colour, the higher the density of dark matter. The panels zoom into the structure, with the right-most panel showing an individual dark matter halo (Lemson & Consortium, 2006).

some form of “feedback” which prevents galaxies from forming stars efficiently. However this is typically done in a formulaic way using sub-grid recipes, without fully understanding the physics of the diverse processes that drive galaxy evolution. Differentiating, for example, between AGN and stellar feedback requires further observational constraints which can only be obtained from detailed studies of local galaxies.

These and other outstanding issues in galaxy evolution can be studied using different approaches. One common method is to measure the evolution of various galaxy properties across a wide redshift range. This is a direct way to study galaxy evolution, but it is limited by the extreme faintness of distant galaxies, and the lower data quality obtainable at high redshift means that, in general, only global galaxy properties can be compared across cosmic time. A complementary approach is to study the fossil record encoded in the stellar populations and dynamics of nearby galaxies. The high signal-to-noise and spatial resolution achievable when targeting nearby galaxies allows very detailed studies of individual objects. Comparing the observed stellar populations with the predictions of stellar population models gives a wealth of information about the formation history of galaxies, and is the method we utilise in this thesis. As detailed below, we specifically study early-type galaxies, which are particularly suited to these types of studies, as they have simpler star formation histories and their spectra are typically not complicated by dust, or emission from ongoing star formation. The formation history of early-type galaxies is particularly interesting, as these massive objects contain a significant fraction of the stellar mass in the current day Universe (Bell et al., 2003), and are thought to be the endpoints of galaxy evolution.

1.1 Early-type galaxies

Galaxies can be loosely divided into two groups based on their morphologies: the so-called late-type and early-type (elliptical and lenticular) galaxies. While different classification systems exist, a common criterion to differentiate between the two groups is the presence or absence of spiral arms. In Hubble’s classic tuning fork (Hubble, 1926, shown in Figure 1.2), ellipticals are ordered by their apparent ellipticity, while spirals are classified by the prominence of the spiral arms, and the presence of bars. In this scheme, the two branches are joined by the lenticular class, which was thought to be a transition stage between the two. This simple morphological classification relies on imaging, and as such, depends strongly on image depth, the wavelength band of the observation, and the galaxy orientation. Despite these difficulties, it turns out to be surprisingly robust, in that it correlates with many stellar population properties. Late-type galaxies are typically rotationally supported (disky), blue, actively star forming, and contain young stellar populations, as well as large amounts of gas and dust. Early-type galaxies (ETGs) are typically pressure supported (spheroidal), red, passive, and contain old stellar populations, with much less gas and dust (Blanton & Moustakas, 2009, Roberts & Haynes, 1994).

The clear distinction between the two groups was upended by the SAURON (de Zeeuw et al., 2002) and ATLAS^{3D} (Cappellari et al., 2011) surveys, who obtained integral field spectroscopy of a volume-limited sample of 260 nearby ETGs. These groundbreaking surveys found that ETGs could be divided into 2 groups: the slow rotators, which tend to be massive and spheroidal, and the fast rotators, which make up the majority of ETGs (Emsellem et al., 2011, Krajnović et al., 2011). These fast rotators have disks, and resemble spirals which have had their gas and dust removed. This led to a new classification system based on galaxy kinematics, which is thus insensitive to projection effects. The ATLAS^{3D} comb (shown in the right hand panel of Figure 1.2) emphasises the similarity between the fast rotator ETGs, and the spirals, which form parallel sequences in this new system, implying a close evolutionary link between the two groups. Subsequent large IFU surveys of all galaxy types such as CALIFA (Sánchez et al., 2012), SAMI (Croom et al., 2012), and MaNGA (Bundy et al., 2014), have confirmed this: finding, for example that late-type galaxies

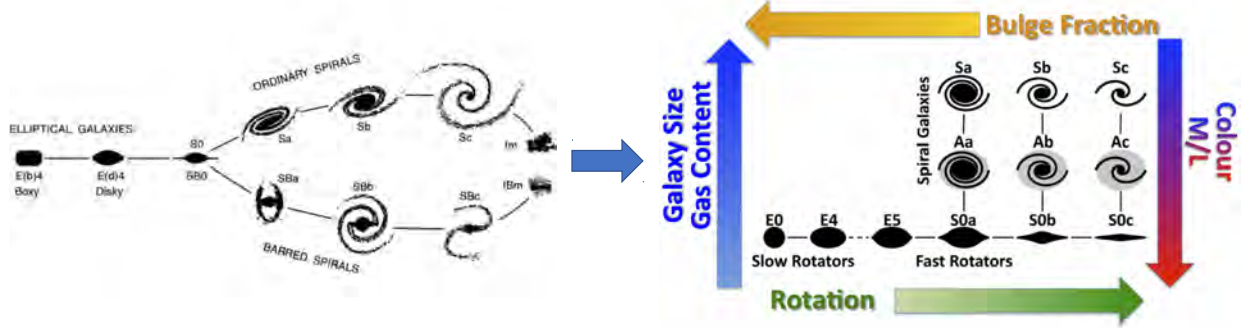


Figure 1.2: Comparison between the Hubble tuning fork and the ATLAS^{3D}comb. The Hubble tuning fork (left; Hubble, 1926) casts the early-type galaxies and late-type galaxies as physically distinct classes, joined by the intermediate lenticular class, while the ATLAS^{3D}comb (right; Cappellari et al., 2011) shows a close relationship between the two groups, which lie on parallel sequences.

and early-type fast rotators form a continuous class of objects in terms of their kinematic properties (Cortese et al., 2016).

ETGs are known to obey remarkably tight scaling relations between their velocity dispersion (σ) and other population properties, for example, the Fundamental Plane (Djorgovski & Davis, 1987, Dressler et al., 1987) which links their velocity dispersion, effective radius and luminosity. The tightness of these scaling relations and their evolution as a function of redshift is successfully explained by the traditional picture of monolithic collapse at high redshift (Eggen et al., 1962), followed by passive evolution until the present (Larson, 1975). This classic picture is evolving, however, with studies over the past decade finding ETGs contain young populations (e.g. Kaviraj et al., 2008, Kuntschner et al., 2010, Trager et al., 2000a), gas and dust (Young et al., 2011), and even ongoing star formation (Davis et al., 2014). Revisiting the classic scaling relations in light of these updated observations have led to new insights into the process of ETG formation. The existence of the Fundamental Plane has long been understood as due to the fact that galaxies obey the virial theorem. However the exponents of the Fundamental Plane were repeatedly shown not to match the predictions of the virial theorem (e.g. Bernardi et al., 2003, Colless et al., 2001, Hudson et al., 1997). Variations in either the stellar mass-to-light ratio (M/L), surface brightness profiles, or dark matter fractions across the galaxy population were proposed as candidates to explain this so-called “tilt” of the Fundamental Plane. Cappellari et al. (2006) fit general dynamical models to 2D kinematics of 25 ETGs, and showed that the tilt of the Fundamental Plane was due to variations of M/L with velocity dispersion.

However the total M/L derived from dynamical models was found to vary systematically compared to the M/L derived from stellar populations, implying a variation in either the stellar initial mass function (IMF) or the dark matter fraction. Distinguishing between these two required accurate modelling of the stars and the dark matter, for a larger sample of galaxies. The application of general axisymmetric JAM models to the ATLAS^{3D} survey, with its large sample size and 2D stellar kinematics was able to break this degeneracy, and showed conclusively that the IMF varied systematically across the ETG population, regardless of the chosen dark matter prescription (Cappellari et al., 2012). Cappellari et al. (2012, 2013b) introduced a mismatch parameter α , defined as the ratio between the dynamically determined and the stellar population estimates of M/L . They

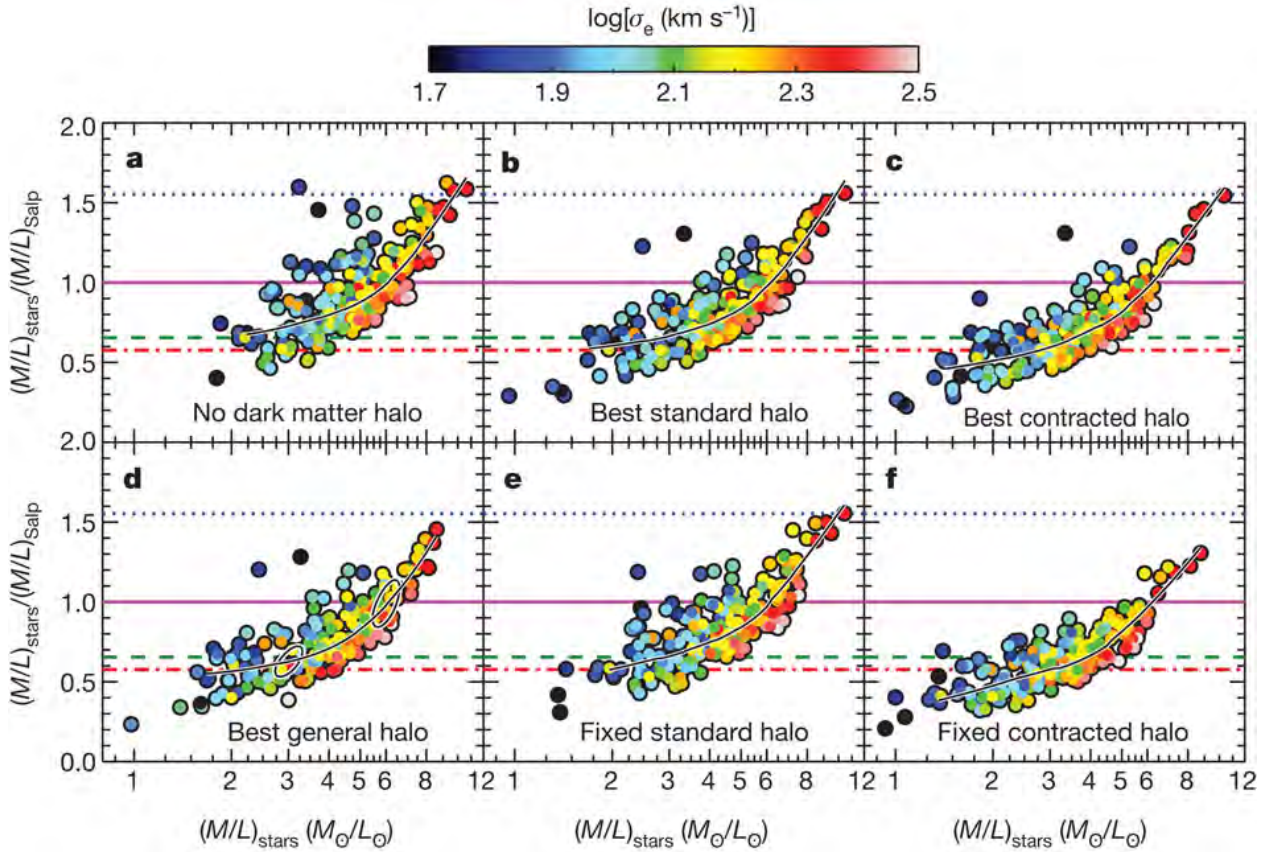


Figure 1.3: The IMF mismatch parameter as a function of the M/L_{stars} determined using dynamical models. The mismatch parameter is defined as the ratio of the stellar M/L s determined using dynamical modelling and stellar population modelling. The horizontal lines indicate the expected values of the ratio for various IMFs, e.g. a Kroupa IMF is shown by the green dashed line, while a galaxy with a Salpeter IMF would lie on the solid magenta line. Colours indicate the galaxies' stellar velocity dispersion. The six panels show different prescriptions for dark matter included in the dynamical models, however all show the same qualitative behaviour. Taken from Cappellari et al. (2012).

found that more massive galaxies have a larger α , indicating an excess of high- or low-mass stars compared to a Milky Way IMF (see Figure 1.3). This result was supported by other studies which use gravitational probes to constrain the total mass (e.g. Auger et al., 2010, Thomas et al., 2011, Tortora et al., 2013, Treu et al., 2010), and was also consistent with studies of the strengths of specific gravity sensitive spectral features, which imply that this mismatch is due to an excess of low-mass stars (e.g. Conroy & van Dokkum, 2012, Ferreras et al., 2013, La Barbera et al., 2013, Rosani et al., 2018, Spiniello et al., 2014, van Dokkum & Conroy, 2010, see Figure 1.4). While the global trends found by these independent techniques are in qualitative agreement, direct comparisons of the two methods on the same data yield little to no correlation (Smith, 2014), implying errors in one or both techniques. Furthermore, recent studies using spatially-resolved data have found strong radial gradients in the IMF to be common throughout the early-type galaxy population, leading to an emerging picture whereby the observed bottom-heavy IMFs may be limited only to the central regions of massive early type galaxies (e.g. La Barbera et al., 2016, Parikh et al., 2018, Sarzi et al., 2018, van Dokkum et al., 2017, see Figure 1.5.).

All taken together, current studies imply a more complicated star formation history for ETGs than traditionally thought. Multiple stellar population properties including age, metallicity, M/L , and IMF have been shown to vary as a function of velocity dispersion in fast rotator ETGs (McDermid et al., 2015). Cappellari et al. (2013b) show that velocity dispersion traces bulge mass fraction,

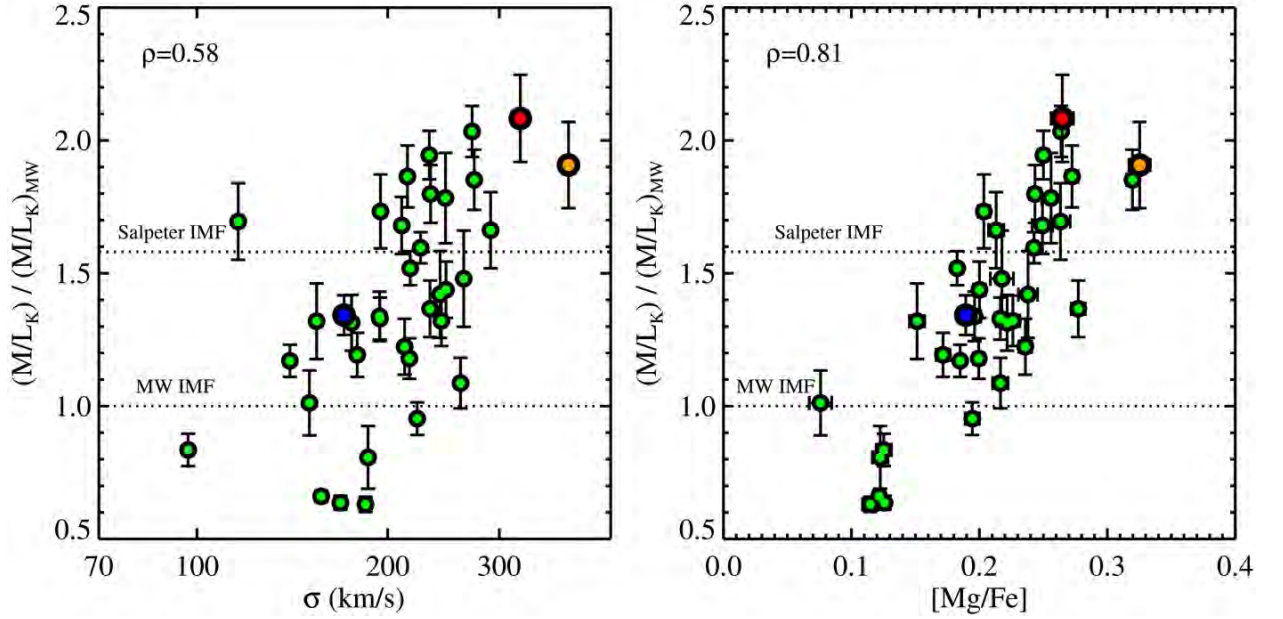


Figure 1.4: The IMF mismatch parameter as a function of velocity dispersion, and of $[Mg/Fe]$. The mismatch parameter is defined as the ratio of two stellar M/L s determined using stellar population modelling: M/L_K is the M/L from full spectral fitting where age, metallicity and IMF are free to vary, while $(M/L_K)_{MW}$ is the M/L obtained from full spectral fitting assuming a fixed Milky Way-like (Kroupa) IMF. The horizontal lines indicate the expected values for the ratio for a Kroupa or a Salpeter IMF. Taken from Conroy & van Dokkum (2012).

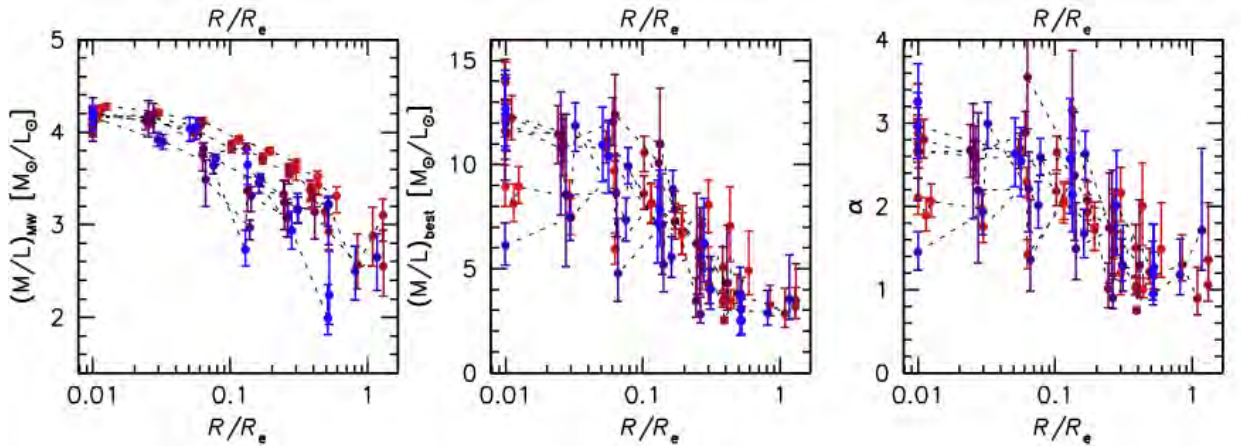


Figure 1.5: Stellar population gradients derived from full spectrum fitting by van Dokkum et al. (2017). The first panel shows the stellar mass-to-light ratio (M/L) obtained assuming a Milky Way-like (or Kroupa) IMF; the middle panel shows the M/L obtained when fitting for the IMF directly from the data; the last panel shows the IMF mismatch parameter $\alpha = (M/L_{best})/(M/L_{MW})$, giving, in effect, the extra variation in M/L which is due to IMF. Different colours indicate the six different galaxies in the sample, all of which show the same qualitative behaviour.

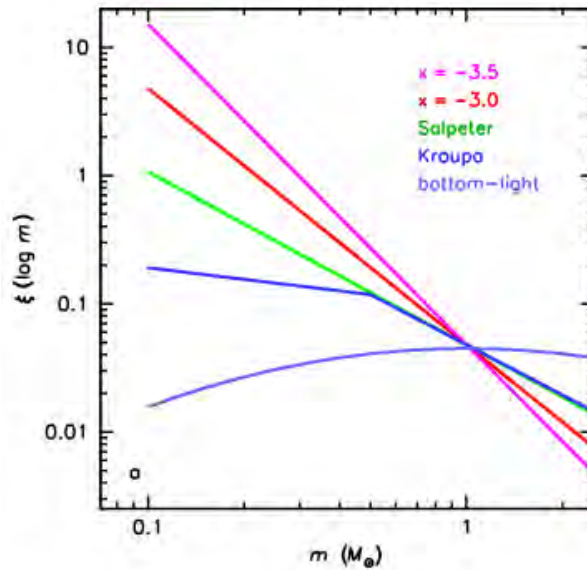


Figure 1.6: Various stellar IMFs, ranging from bottom-light (lacking low-mass dwarfs compared to a canonical, Milky Way-like IMF) in blue, to extremely bottom-heavy (containing an excess of low-mass dwarfs) in pink. The IMFs are normalised at $1M_{\odot}$. Taken from van Dokkum & Conroy (2010).

implying a link between the growth of the bulge and the quenching of star formation. This, plus the smooth extension of the sequence of ETG properties to spirals, leads to a picture where fast rotator ETGs evolve from spirals, gradually growing their bulges and shutting down star formation. Further detailed studies of local objects would add clarity to this active and intriguing field.

1.2 Stellar population synthesis

Understanding the star formation histories of unresolved galaxies requires stellar population synthesis (SPS) modelling, which uses stellar evolutionary theory to predict the spectral energy distributions (SEDs) one would expect from a particular population of stars. Pioneered by Tinsley (1968), the field of stellar population synthesis is a dynamic one, with many detailed models under active development from multiple groups (Bruzual & Charlot, 2003, Conroy & Gunn, 2010, Conroy et al., 2009, 2018, Larson & Tinsley, 1978, Maraston & Strömbäck, 2011, Röck et al., 2016, Schiavon, 2007, Searle et al., 1973, Vazdekis, 1999, Worthey, 1994).

Most SPS models use the isochrone synthesis method to generate SED predictions. This technique has three main ingredients: stellar evolutionary tracks, stellar spectral libraries, and an assumed form of the stellar initial mass function (IMF). The spectrum of a population of a particular age, metallicity and abundance is generated by integrating the relevant stellar spectra along an isochrone, weighted by the assumed IMF. This requires tracks covering all relevant stellar evolutionary phases, and stellar spectral libraries spanning a wide range of surface gravity g and effective temperature T_{eff} . A great deal of work has gone into advancing all three of the main ingredients, however there remain large uncertainties in all, which are briefly described below.

1.2.1 Stellar initial mass function

The stellar initial mass function (IMF) is an empirical function describing the distribution of stellar masses formed in a single burst of star formation. Low mass stars are much more numerous than

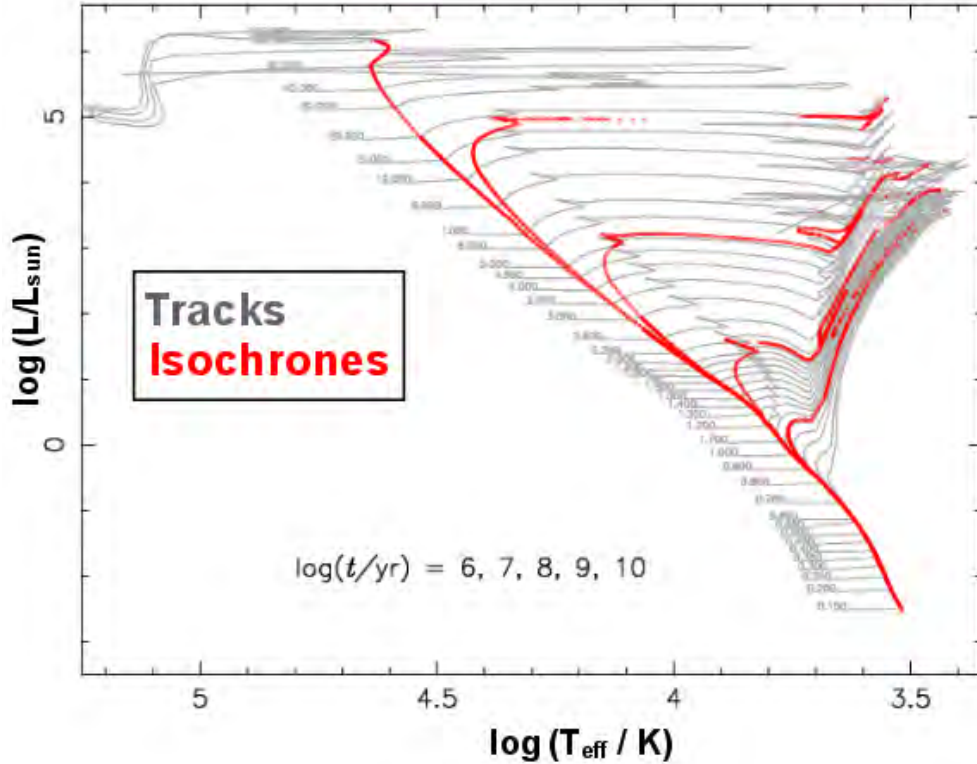


Figure 1.7: Stellar evolutionary tracks (grey lines) and isochrones (red lines), taken from Zhang et al. (2013), showing the locations of stars on the Hertzsprung-Russell diagram. The masses of stars range from $0.1 - 100M_{\odot}$ at logarithmic intervals of 0.05. The ages of the isochrones are $\log(t/\text{yr}) = 6, 7, 8, 9$ and 10 .

high mass stars, and as such, they dominate the mass budget of a population without contributing much to the light. The form of the IMF is thus of paramount importance in studies of unresolved populations, as it sets constraints on the numbers of low mass stars present. Resolved star counts in nearby systems imply the IMF is universal across time and space (see review by Bastian et al., 2010), having the form of a broken power law (Kroupa, 2001) or following a log-normal distribution (Chabrier, 2003). However, recent integrated-light studies of external galaxies have found the IMF to vary systematically across the ETG population as a function of galaxy mass, having an excess of low mass stars in massive ellipticals, compared to this canonical IMF, as described above. Figure 1.6 shows some commonly used IMF shapes.

Constraining the shape of the IMF is of interest for a number of reasons. As well as constraining the physics of star formation, the form of the IMF dictates a galaxy's chemical evolution and determines the amount of feedback after a burst of star formation. As the derivation of physical galaxy properties from stellar population modelling is based on an assumed IMF, the IMF is perhaps the largest systematic uncertainty in these properties, and as such, has a strong impact on our understanding of galaxy evolution.

1.2.2 Stellar evolution calculations

Stellar evolutionary tracks describe the evolution of a single star with time. These determine where a star with specific stellar parameters lies on the Hertzsprung-Russell (H-R) diagram. The stellar tracks are used in order to build up an isochrone¹, which is the input to stellar population models (Figure 1.7). Various stellar tracks exist (e.g. Padova: Marigo & Girardi (2007), Marigo et al.

¹a curve on the H-R diagram representing a population of stars of the same age and metallicity

(2008); BaSTI: Pietrinferni et al. (2009); Geneva: Lejeune & Schaerer (2001)), covering a wide range of initial masses and metallicities. No single set of isochrones spans the necessary range of ages, metallicities, and evolutionary phases required for stellar population modelling, meaning different sets (which make different assumptions regarding stellar physics) are often spliced together. Particular stellar evolutionary phases such as the thermally-pulsing asymptotic giant branch phase are difficult to model, due to the complex mass loss occurring in this phase, leading to differences between different sets of tracks. SPS models such as Maraston (Maraston, 1998, 2005, Maraston & Strömbäck, 2011) have chosen to forgo the isochrone approach for post-Main Sequence phases, instead making the luminosity of these phases proportional to the amount of nuclear fuel burned in each phase, and thus independent of the details of stellar evolution. This leads to even larger discrepancies between models which use this fuel consumption theorem versus the standard isochrone synthesis approach, which is particularly noticeable in the near-infrared, where these cool, post-Main Sequence giants dominate the integrated light.

1.2.3 Stellar Libraries

The stellar population technique is also limited by the quality of the stellar spectral libraries used. In order to generate accurate predictions, the models require high quality libraries covering all relevant phases. However observational constraints mean that empirical libraries are typically limited to the composition of the solar neighbourhood, as well as being limited in terms of resolution and wavelength coverage. High quality stellar spectral libraries are currently available for a wide range of parameter space at optical wavelengths (e.g. STELIB (Le Borgne et al., 2003), MILES (Cenarro et al., 2007, Falcón-Barroso et al., 2011, Sánchez-Blázquez et al., 2006), ELODIE (Prugniel et al., 2007)), however spectral libraries outside this wavelength range are typically lacking either resolution, coverage or both. Recent advances in instrumentation have led to the creation of high quality empirical near-infrared libraries such as the IRTF Spectral Library (Rayner et al., 2009), and X-Shooter Spectral Library (Chen et al., 2014), opening the near-infrared up to the kinds of detailed stellar population studies that have long been carried out in the optical. Theoretical libraries, in principle, are available at any resolution and wavelength, and covering a wide range of parameter space. However they rely critically on atomic and molecular line lists, which are highly incomplete. To understand real galaxies, it is important to use both approaches, for example, by applying theoretical abundance corrections to empirical spectra in order to replicate the spectra of massive α -enhanced ellipticals (e.g. Conroy et al., 2018)

1.3 Outline and aims of this thesis

This thesis studies the stellar populations and dynamics of nearby ETGs in order to better understand their formation and evolution. In Chapter 2, we make a detailed comparison of near-infrared stellar population synthesis models, using a well studied sample of nearby ETGs. The near-infrared wavelength range contains signatures of post-main sequence phases which may be important contributors to the light of intermediate age populations, as well as signatures of the low-mass dwarfs which can be used to constrain the initial mass function. Recent advances in instrumentation have opened up the near-infrared for routine study, and this wavelength range will provide valuable insights in the age of JWST. It is thus of vital importance to test the accuracy of the stellar population models in this regime, which currently make widely divergent predictions for the same stellar population parameters.

We then move on to a study of IMF variation in ETGs, using both dynamical modelling, and optical-NIR IMF-sensitive spectral features in combination with stellar population models. The inconsistency between stellar population and dynamical determinations of the IMF described above

is still not understood. Radial gradients in stellar population properties, coupled with the different aperture sizes typically used in the two methods may be a contributing factor. The study of IMF variation is a rapidly evolving field, however past studies have focused mostly on the *global* variation between galaxies. Only recently have spatially resolved IMF measurements become possible, as they require very high quality integral field spectroscopy, which must be compared to accurate stellar population models. We add to this burgeoning field a detailed analysis of three nearby ETGs, for which we have obtained high signal-to-noise IFU data from the Multi Unit Spectroscopic Explorer (MUSE). MUSE offers both the extended wavelength range and fine spatial sampling we require to accurately constrain the IMF using both stellar population and dynamical techniques.

In Chapter 3 we lay out the goals of this study, and present the MUSE data we use for our IMF analysis. We also describe the techniques we use to extract the various data products which go into our modelling, such as maps of the galaxy kinematics and the stellar population properties of interest.

In Chapter 4 we use axisymmetric dynamical models to model the kinematics of our sample. A key difference between this and previous studies is that we relax the assumption of a constant stellar mass-to-light ratio (M/L), which is typically used in dynamical modelling. In this work we measure the M/L variation in our sample due to gradients in the stellar age and metallicity, and test the effects of including these gradients on our dynamically-derived properties. We also test the effects of including IMF variation as observed recently in nearby galaxies.

In Chapter 5 we carry out a completely self-consistent stellar population and dynamical study of these data. We perform full spectral fitting of the same MUSE data from which we extract our kinematics. Using a Markov Chain Monte Carlo approach, we are able to simultaneously constrain IMF and various individual element abundances, allowing us to map IMF variation within our objects as a function of radius. We also derive the spatial M/L variation in our galaxies due to gradients in age, metallicity and IMF, which we incorporate into our dynamical models. We are then in a position to study the effects of including self-consistent stellar population methods on the dynamically derived properties, and draw conclusions about the reliability of the various modelling techniques.

Comparison of Stellar Population Model Predictions Using Optical and Infrared Spectroscopy

The material presented in this chapter is based on material that has been previously published in the journal Monthly Notices of the Royal Astronomical Society (Baldwin et al., 2017)

Abstract

We present Gemini/GNIRS cross-dispersed near-infrared spectra of 12 nearby early-type galaxies, with the aim of testing commonly used stellar population synthesis models. We select a subset of galaxies from the ATLAS^{3D} sample which span a wide range of ages (SSP-equivalent ages of 1–15 Gyr) at approximately solar metallicity. We derive star formation histories using four different stellar population synthesis models, namely those of Bruzual & Charlot (2003), Conroy, Gunn & White (2009; 2010), Maraston & Strömbäck (2011) and Vazdekis et al. (2016). We compare star formation histories derived from near-infrared spectra with those derived from optical spectra using the same models. We find that while all models agree in the optical, the derived star formation histories vary dramatically from model to model in the near-infrared. We find that this variation is largely driven by the choice of stellar spectral library, such that models including high quality spectral libraries provide the best fits to the data, and are the most self-consistent when comparing optically-derived properties with near-infrared ones. We also find the impact of age variation in the near-infrared to be subtle, and largely encoded in the shape of the continuum, meaning that the common approach of removing continuum information with a high-order polynomial greatly reduces our ability to constrain ages in the near-infrared.

2.1 Introduction

The near-infrared is an important wavelength range which traces populations of a range of ages (Maraston et al., 2006, Origlia & Oliva, 2000a,b), suffers lower dust obscuration than the optical (Peletier et al., 1995), and will be the wavelength range probed by facilities like the James Webb Space Telescope (JWST) and adaptive-optics corrected Extremely Large Telescopes (ELTs). In

spite of this, the near-infrared remains a poorly explored wavelength range, largely due to the difficulty in observing and calibrating ground-based near-infrared data, compared to optical data.

In the integrated light of galaxies, the near-infrared is dominated by cool stars, which may be either luminous evolved giants, or slow-burning low-mass dwarfs. A particular phase of interest is the thermally-pulsing asymptotic giant branch (TP-AGB) phase. TP-AGBs are cool, intermediate mass giants, which are thought to dominate the light of stellar populations with ages between $\sim 0.2 - 2$ Gyr (Maraston, 2005). TP-AGB stars may contribute up to 40 % of the bolometric light of populations in this age range, and up to 80 % in K band (Maraston et al., 2006). The complex physics of this phase (Marigo et al., 2008) make it difficult to model theoretically, and its short lifetime makes it difficult to calibrate observationally (Maraston, 2005). Consequently, current near-infrared stellar population synthesis models make widely divergent predictions for the age range in which the TP-AGB dominates.

The Maraston models (Maraston, 1998, 2005, Maraston & Strömbäck, 2011, hereafter M98, M05 and M11), for example, which use the fuel consumption theorem of Renzini (1981), include a large contribution from stars in the TP-AGB phase compared to traditional models such as Bruzual & Charlot (2003, hereafter referred to as BC03), which use the isochrone synthesis technique. This large contribution in the Maraston models leads to the presence of strong molecular carbon and oxygen absorption features throughout the near-infrared spectra of intermediate age populations, which could in principle be used as a tracer of these populations. These strong absorption features are not found in other popular SPS models such as BC03, which give much less weight to the TP-AGB phase.

Use of the Maraston models to derive properties of high redshift galaxies leads to estimates of age and mass that are systematically lower than those derived using the BC03 models (by $\sim 60\%$ on average), highlighting the importance of a correct treatment of this phase (Maraston et al., 2006). Since 2005, studies on this topic have produced conflicting results. Observationally, the molecular bands predicted by the M05 models have been detected in a number of studies (Lyubenova et al., 2012, Riffel et al., 2015, and others), however, authors were typically unable to strongly link these features to intermediate-age populations. These results are in contrast to the study by Zibetti et al. (2013) who specifically targeted post-starburst galaxies. These galaxies fall in the exact age range at which the TP-AGB phase should dominate, and yet the authors did not observe the strong molecular features predicted by the Maraston models. Photometric studies show similar conflicts, with studies providing support both for (MacArthur et al., 2010, Maraston et al., 2006) and against (Conroy & Gunn, 2010, Kriek et al., 2010) a large TP-AGB contribution. Newer models have since been created which use their own prescription for the TP-AGB phase (i.e. the Flexible Stellar Population Synthesis models by Conroy et al. (2009)), or use new high resolution near-infrared libraries (Röck et al., 2016), adding to the variety of model predictions for this wavelength range.

As well as features relevant to intermediate age populations, the near-infrared contains spectral lines which, in old populations, can be used to probe the initial mass function - a fundamental galaxy property which has recently been found to vary systematically in early-type galaxies as a function of stellar mass. Most of the evidence for a systematically varying IMF has come from features blueward of $1 \mu\text{m}$ (Na 8200, FeH, CaT, etc.), however features which are differentially sensitive to giant or dwarf stars are found throughout the entire near-infrared range. Some are longer wavelength transitions of well-known IMF tracers (e.g. NaI doublets at $1.14 \mu\text{m}$ and $2.21 \mu\text{m}$), while others are new (KI doublets at $1.17 \mu\text{m}$ and $1.25 \mu\text{m}$, CO bandhead at $2.3 \mu\text{m}$).

In this chapter, we explore this spectral region and various contemporary models for a well-defined sample of galaxies spanning a range of optical ages (from post-starburst to ancient and passive) to compare the quality of fits of the different models, examine the impact of model choice on the recovered star formation history, and compare the derived properties with those measured at optical wavelengths. We use high signal-to-noise near-infrared spectroscopy of a sample of well-studied galaxies, so we are able to compare the properties derived using various near-infrared

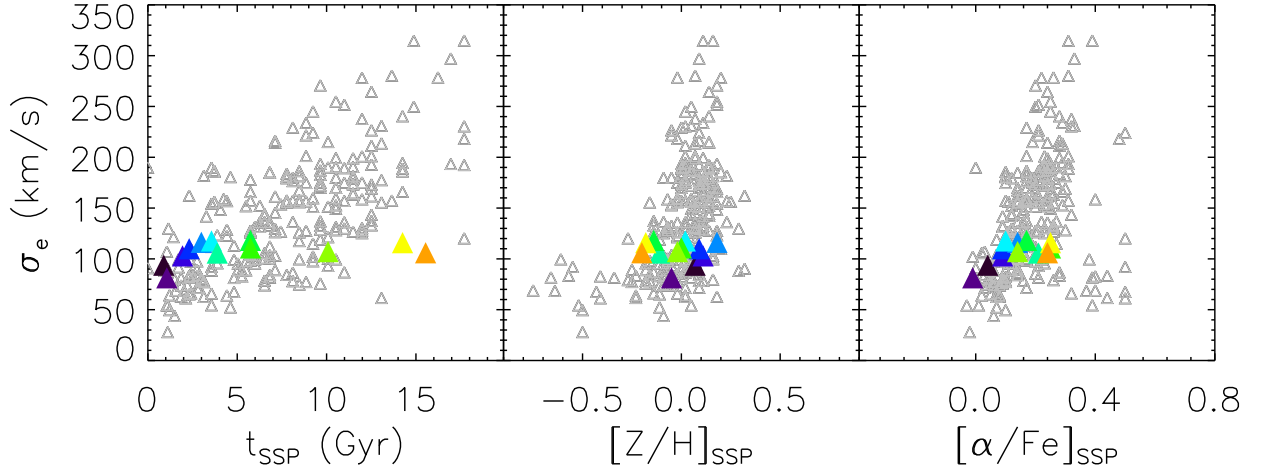


Figure 2.1: Sample selection. Panels show (from left to right) mean central stellar age, metallicity and abundance ratio varying against effective stellar velocity dispersion, all derived from ATLAS^{3D} optical spectroscopy, calculated in an $R_c/8$ aperture. The data and references are given in Table 2.1. Coloured triangles show our sample selection, taking a narrow range in velocity dispersion ($80 - 120 \text{ km s}^{-1}$) and a broad range in age. The colours are the same as in Figure 2.2.

models with those derived for the same populations using a range of other wavelength regimes, including *aperture-matched* optical spectroscopy from the ATLAS^{3D} survey (Cappellari et al., 2011).

The chapter is organized as follows. In Section 2.2 we describe the sample. Section 2.3 details the observations and the data reduction. In Section 2.4 we give a description of the four models we intend to test, namely, BC03, M11, Flexible Stellar Population Synthesis models (Conroy & Gunn, 2010, Conroy et al., 2009, hereafter FSPS) and the Vazdekis models, which have been recently extended to the near-infrared by Röck et al. (2016, hereafter V16). Our results are found in Section 2.5, followed by our discussion and conclusions.

2.2 Sample

The early-type galaxies selected for this work were drawn from the ATLAS^{3D} survey (Cappellari et al., 2011). The selection criterion was that the galaxies span a narrow range of velocity dispersion (σ), as shown in Figure 2.1. Following the tight scaling relations of stellar populations with σ (McDermid et al., 2015, Thomas et al., 2005, Trager et al., 2000b), this narrow range in σ ensures that the galaxies are all of approximately solar metallicity and abundances, while spanning a large range of optically-derived SSP-equivalent ages ($1 - 15 \text{ Gyr}$). Isolating age as the main variable allows us to focus on which features drive the age-sensitivity in this spectral range, and empirically test the impact of the predicted TP-AGB independently of specific models. The narrow range in velocity dispersion also ensures that intrinsic line width can be compared within the sample without being strongly affected by velocity broadening (Kuntschner, 2004).

The sample galaxies have excellent supporting data: mean ages, metallicities and α -abundances were calculated for all galaxies in the sample by comparing the predictions of the Schiavon (2007) single stellar population models to optical Lick indices, and complete star formation histories were also calculated from full spectral fitting of their optical spectra (McDermid et al., 2015). IC 0719, NGC 3032, NGC 3156, NGC 3182, NGC 3489 and NGC 4710 were shown to contain molecular gas (Combes et al., 2007, Young et al., 2014), corresponding to regions of ongoing star formation (Davis et al., 2014), and have young mean stellar ages. NGC 3032 and NGC 3156 are consistent with post-starburst galaxies such as those targeted in Zibetti et al. (2013). Other objects show no direct evidence of molecular gas, and have old ages, consistent with having experienced a single

Table 2.1: Properties of observed sample.

Galaxy (1)	Right Ascension (2)	Declination (3)	Velocity (4)	σ_e (5)	M_K (6)	t_{SSP} (7)	t_m (8)	$[Z/H]_{\text{SSP}}$ (9)	$[\alpha/\text{Fe}]_{\text{SSP}}$ (10)
IC0719	11 40 18.5	09 00 36	1833	118	-22.7	3.6	9.5	0.02	0.10
NGC3032	09 52 08.1	29 14 10	1562	94	-22.0	0.9	2.4	0.07	0.04
NGC3098	10 02 16.7	24 42 40	1397	118	-22.7	5.7	9.8	-0.14	0.17
NGC3156	10 12 41.2	03 07 46	1338	81	-22.2	1.1	2.3	-0.05	-0.01
NGC3182	10 19 33.0	58 12 21	2118	111	-23.2	5.7	10.8	-0.00	0.25
NGC3301	10 36 56.0	21 52 56	1339	110	-23.3	2.3	6.3	0.09	0.09
NGC3489	11 00 18.6	13 54 04	695	103	-23.0	1.9	4.4	0.11	0.09
NGC4379	12 25 14.7	15 36 27	1074	106	-22.2	15.5	12.6	-0.20	0.24
NGC4578	12 37 30.5	09 33 18	2292	107	-22.7	10.1	11.8	-0.02	0.14
NGC4608	12 41 13.3	10 09 20	1850	116	-22.9	14.3	12.7	-0.18	0.25
NGC4710	12 49 38.8	15 09 56	1102	106	-23.5	3.9	8.9	-0.11	0.21
NGC5475	14 05 12.4	55 44 31	1671	116	-22.9	3.0	7.0	0.18	0.14

Columns : (1) Galaxy name. (2) Right ascension (J2000.0) (3) Declination (J2000.0). (4) Heliocentric velocity in km s^{-1} . (5) Effective velocity dispersion σ_e measured within $1 R_e$ (km s^{-1}). (6) Absolute K-band magnitude. (7) SSP-equivalent age in Gyr. (8) mass-weighted mean age from optical full spectral fitting (Gyr) (9) SSP-equivalent total metallicity. (10) SSP-equivalent α -abundance. Columns (4)-(6) were taken from Table 3 in Cappellari et al. (2011, columns 7,9 respectively). Columns (6)-(9) were taken from Table 1 in McDermid et al. (2015), and are values calculated on the galaxies' optical spectra within an $R_e/8$ aperture.

burst of star formation at high-redshift. Our sample therefore covers a very broad range of star formation histories, thus, if there are specific age-dependent features, such as those arising from a short-lived TP-AGB phase, we expect to trace them with this sample. The basic galaxy parameters are summarised in Table 2.1.

2.3 Observations and Data Reduction

2.3.1 Near-Infrared Spectroscopy

High signal-to-noise ratio near-infrared spectra were obtained over six nights between 1 February, 2012 and 1 May, 2012, using the Gemini Near-Infrared Spectrograph (GNIRS) on the Gemini North 8 m telescope in Hawaii. They were taken in queue mode through observing program GN-2012A-Q-22, with observing conditions constraints of 85th percentile for seeing ($1''.1$ or better in V-band), 50th percentile for cloud cover (photometric), and unconstrained water vapour and sky background. GNIRS cross-dispersed (XD) spectroscopy mode was utilised, providing a complete spectrum from $0.8\text{--}2.5 \mu\text{m}$ across five orders. Use of the short ($0.15''/\text{pixel}$) blue camera, a 32 l/mm grating, and a slit of width $0.3''$ gives an instrumental resolution of $R \sim 1700$.

For each observation, the slit was oriented at the mean parallactic angle at the time of observation to minimise the impact of losses due to atmospheric refraction. As the galaxies are extended objects and the slit used in XD mode is short ($7''$), the telescope was offset $50''$ from the object to observe a blank region of sky in order to facilitate the removal of night sky emission lines and dark current. Object-sky pairings were observed in an ABA pattern, giving a neighbouring sky frame to every galaxy exposure. Immediately after each sequence of exposures, a set of calibration images was taken with the telescope at the position of the galaxy, consisting of continuum and emission-line lamps for flat fielding and wavelength calibration purposes, respectively. This ensured minimal instrument flexure between the science and calibration exposures. An observing log is given in Table 2.2.

The data were reduced using version 2.0 of the XDGNIRS pipeline made public by Rachel

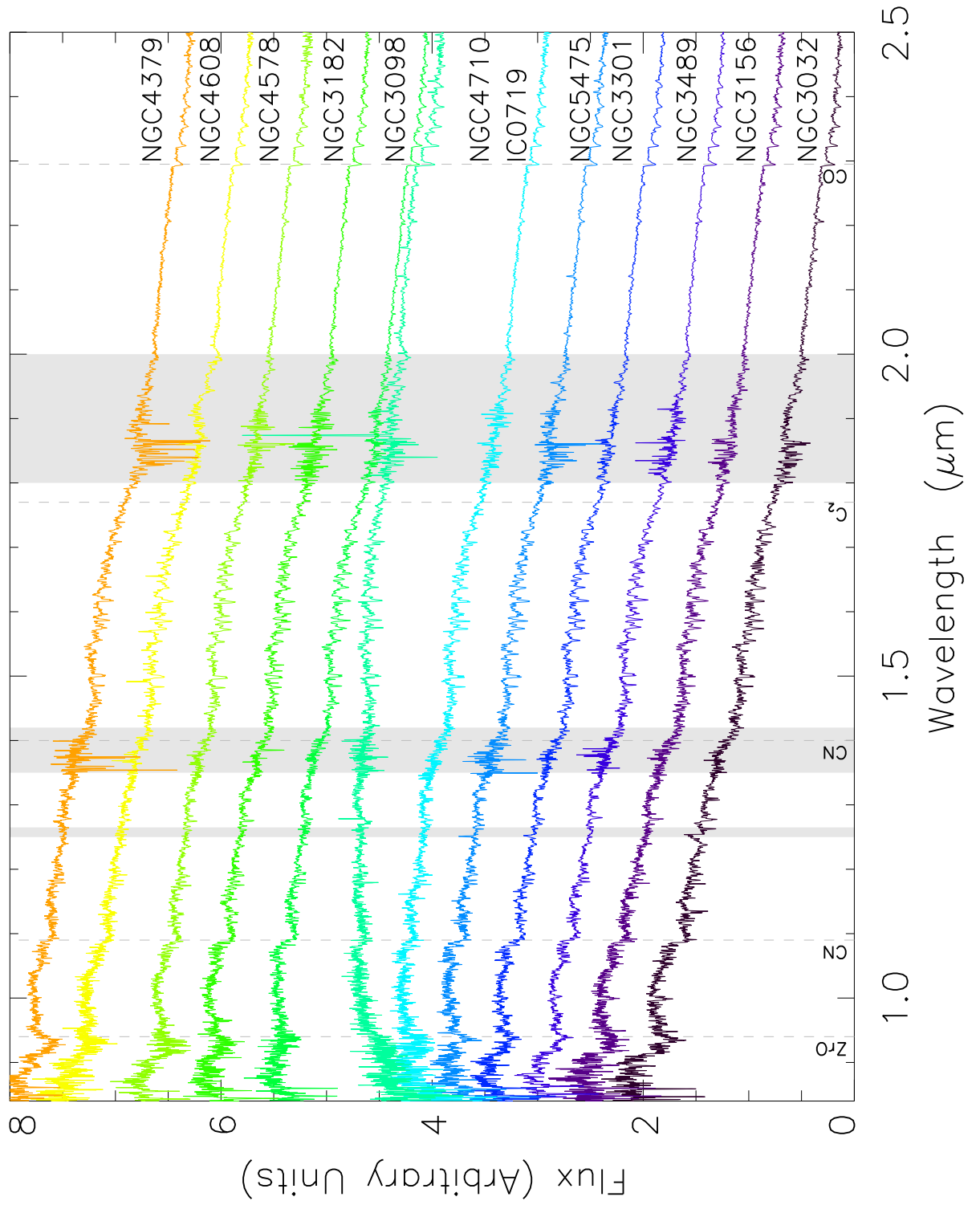


Figure 2.2: Rest frame spectra of the galaxy sample, sorted by increasing ATLAS^{3D} derived SSP age (age increases upwards). The grey shaded boxes mark regions of poor atmospheric transmission as defined in the text. Important molecular features are marked with dashed lines.

Table 2.2: Observing Log

Galaxy	Date (dd-mm-yyyy)	$N_{\text{exp}} \times T_{\text{exp}}$ (s)	Slit Angle °E of N	Airmass	S/N_K ¹
IC0719	04-02-2012	8x120	30	1.1	67
NGC3032	10-02-2012	6x120	267	1.3	78
NGC3098	05-02-2012	6x120	147	1.0	89
NGC3156	04-02-2012	10x120	310	1.0	101
NGC3182	04-02-2012	8x120	160	1.2	99
NGC3301	04-02-2012	4x120	265	1.0	80
NGC3489	01-02-2012	4x120	75	1.4	198
NGC4379	03-02-2012	6x120	75	1.1	52
NGC4578	04-02-2012	6x120	50	1.1	102
NGC4608	01-05-2012	8x120	56	1.1	92
NGC4710	10-02-2012	6x120	284	1.2	104
NGC5475	10-02-2012	7x120	222	1.4	68

¹ Average signal-to-noise between $2.1 - 2.2\mu\text{m}$ calculated as the ratio of the median of the galaxy spectra and the standard deviations of the residuals from the pPXF best fit template.

Mason via the Gemini DRForum². The pipeline calls a variety of tasks, mostly from the Gemini IRAF package (Cooke & Rodgers, 2005) to convert the original 2D data into a flux-calibrated 1D spectrum. Firstly the spectra were cleaned of pattern noise caused by the detector controller, using python code available online via the Gemini website³. The code uses the periodicity of the noise to establish a correction, which is subtracted from the frames. Radiation events caused by radioactive lens coatings on the GNIRS short camera were identified by comparison with a clean minimum frame at each nod position, and interpolated over using IRAF's *fixpix* task. Cleaned files were divided by a flat field created from a combination of quartz-halogen and infrared flats taken after each science observation. Subtraction of a sky frame removed night sky emission lines as well as other static artefacts in the detector, such as stable hot pixels. The GNIRS detector is tilted with respect to the spectral and spatial axes of each order, so the orders were rectified using daytime pinhole flats. The spectra were then wavelength calibrated using an argon arc frame and 1D spectra were extracted from each order using IRAF task *apall*. Spectra were extracted within an aperture of $\pm R_e/8$, although in one galaxy (NGC 4710) the slit length did not allow this and we extracted a spectrum using the full slit length instead ($\pm R_e/12$). The extraction was carried out such that the sum was weighted by the signal to noise in each pixel (Horne, 1986) and the noise statistics of the CCD were used to detect deviant pixels, which were cleaned and and replaced.

An airmass-matched A-type star of magnitude $m_K = 6.5 - 8$ was observed before or after each science object to correct for telluric absorption and to approximately flux calibrate the science data. Before applying the telluric correction, the star's intrinsic hydrogen lines were removed by interactively shifting and scaling a model spectrum of the star Vega (Kurucz, 1993). IRAF's *telluric* task was then used to interactively shift and scale the (hydrogen line-free) telluric star spectra to achieve the optimal telluric removal in the science target.

For the purposes of flux calibration, the telluric star was approximated as a black-body continuum having the effective temperature of the star's spectral type (Pecaut & Mamajek, 2013). This black-body spectrum was scaled to the value of the star's K band magnitude as taken from 2MASS (Skrutskie et al., 2006). Each order of the science target was multiplied by this function. Finally, the orders were joined together using the IRAF task *odcombine*, which ensures a smooth overlap

²<http://drforum.gemini.edu/topic/gnirs-xd-reduction-script/>

³<http://www.gemini.edu/sciops/instruments/gnirs/data-format-and-reduction/cleanir-removing-electronic-pattern->

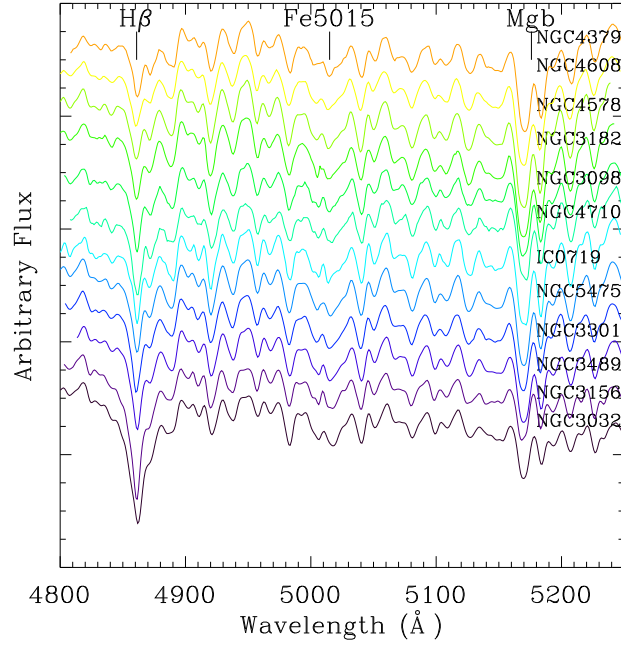


Figure 2.3: Rest frame SAURON optical spectra, ordered by decreasing SSP-equivalent age from top to bottom. The colours are the same as in Figure 2.2.

region between the different orders. The resulting spectra⁴ are shown in Figure 2.2, ordered by the SSP-equivalent age derived by ATLAS^{3D}.

2.3.2 Optical Spectroscopy

We also utilise the optical spectroscopy obtained as part of the ATLAS^{3D} survey in order to test the various models for self-consistency. SAURON spectra cover the wavelength range 4800 – 5380 Å at a spectral resolution of 4.2 Å full width at half-maximum (FWHM). Observations are described in Cappellari et al. (2011) and the data reduction is detailed in Bacon et al. (2001). To test for self-consistency, we derived star formation histories for all models using both wavelength ranges. To accurately compare properties derived using our new GNIRS near-infrared spectroscopy with those derived from SAURON optical integral-field spectroscopy, we re-extract optical spectra integrated along an effective slit of width and length equal to that of the GNIRS aperture, centred on the galaxy nuclei and aligned along the parallactic angle. The optically derived parameters were thus obtained on spectra extracted from the same region as covered by our near-infrared data. The spectra are shown in Figure 2.3, ordered by the SSP-equivalent age derived by ATLAS^{3D}.

2.4 Stellar Population Synthesis

2.4.1 Stellar Population Models

There are nowadays a large number of stellar population models available for use in extragalactic work. These models differ in their choice of stellar tracks and spectral libraries, as well as the parameter space they cover (age, metallicity, wavelength range, initial mass function) and their methods used for including various post-Main Sequence phases (e.g. blue horizontal branch, TP-AGB). Here we describe the salient features of the models used in this work.

⁴available on github <https://github.com/cbaldwin1/Reduced-GNIRS-Spectra>

Maraston & Stromback (2011)

The M11 models use the isochrones and stellar tracks from Cassisi et al. (1997) up to the main sequence turn-off, after which time they use the fuel consumption theorem of Renzini (1981) to calculate the energetics of post-Main Sequence (PMS) stages of stellar evolution such as the TP-AGB phase. With this approach, the luminosity of PMS stars is proportional to the amount of nuclear fuel burned in each PMS phase, making the contribution of each phase independent of the details of stellar evolution. The specific contribution of the TP-AGB phase at each age was calibrated against Magellanic Cloud clusters.

For the stellar spectra, the M11 models offer the option of four different libraries of flux-calibrated empirical stellar spectra, namely Pickles (Pickles, 1998), ELODIE (Prugniel et al., 2007), STELIB (Le Borgne et al., 2003) and MILES (Cenarro et al., 2007, Sánchez-Blázquez et al., 2006), as well as the MARCS (Gustafsson et al., 2008) library of very high resolution theoretical stellar spectra. Each library is supplemented with empirical stellar spectra of O- and C-rich TP-AGB stars from Lançon & Mouhcine (2002). We used the Pickles library to fit the near-infrared data, as it is the only library in the M11 model set to cover the near-infrared at the intermediate ages which are of interest here. The library covers a wavelength range of 1150 Å to 2.5 μm at a resolving power $R \approx 500$. Using this library, model SEDs are available for 58 ages, logarithmically spaced from 2.5 Myr to 15 Gyr at solar metallicity only, for three choices of IMF (Salpeter, Chabrier and Kroupa). It should be noted that approximately half of the stars in the Pickles library lack spectroscopic observations above 1 μm . At these wavelengths, the “spectra” of these stars consist of a smooth energy distribution formed from broad-band photometry, hence some near-infrared absorption features may be poorly resolved. For the optical fits, we used the MILES library, which covers 3500 – 7430 Å at a resolution of 2.51 Å (FWHM), which is the same optical library used by the majority of other models tested.

Bruzual & Charlot (2003)

The BC03 models use the Padova 1994 tracks (Alongi et al., 1993, Bressan et al., 1993, Fagotto et al., 1994) supplemented with the Vassiliadis & Wood (1993) models for TP-AGB stars and the Vassiliadis & Wood (1994) models for post-AGB stars. The default stellar spectral libraries are the empirical STELIB library for the optical ($R \approx 2000$; Le Borgne et al., 2003), extended with the semi-empirical BaSeL library ($R = 200\text{--}500$; Westera et al., 2002) for ultraviolet and near-infrared predictions. BC03 also offer the option of generating solar metallicity SEDs extending to the near-infrared using the Pickles library, which has a resolution of $R \sim 500$. This is higher resolution than the BaSeL library, and corresponds to the same library used by the M11 models (see above). We therefore use STELIB to fit the optical data, and Pickles to fit the near-infrared. Model predictions are available for the same age and metallicity ranges as the M11 models described above, and for Salpeter and Chabrier IMFs.

Conroy (2009)

The FSPS models by Conroy & Gunn (2010), Conroy et al. (2009) use stellar evolution tracks from the Padova group (Marigo & Girardi, 2007, Marigo et al., 2008). The spectral libraries are MILES in the optical, extended with the semi-empirical BaSeL3.1 library (Lejeune et al., 1997, 1998, Westera et al., 2002), supplemented with empirical TP-AGB spectra from the library of Lançon & Mouhcine (2002). These models introduce parameters ΔL and ΔT , which allow the user to modify the location of the TP-AGB track on the HR diagram. Originally these shifts were with respect to the Padova tracks, however the current models set the default values of L and T to the best-fit values calculated in Conroy & Gunn (2010). We used the default model values for this study. This model

Table 2.3: Overview of the default basis sets for each stellar population model used in this study.

SPS Model	Stellar Tracks	Optical Library	Optical Resolution	Near-Infrared Library	Near-infrared Resolution	Metallicities (Z_{\odot})	IMF
M11	Cassisi	MILES	2.5 Å	Pickles	500	0.02	Salpeter
BC03	Padova (1994)	STELIB	3 Å	Pickles	500	0.02	Salpeter
FSPS	Padova (2008)	MILES	2.5 Å	BaSeL	200	0.02	Salpeter
V16	BaSTI	MILES	2.5 Å	IRTF	2000	0.02	Salpeter

set provides templates for 188 ages from 0.3 Myr – 15 Gyr and 22 metallicities at the resolution of the BaSeL spectral library ($R \approx 200$). Models can be computed for a range of standard as well as non-standard IMFs (Salpeter, Chabrier, Kroupa, piece-wise power-law IMF, top-heavy IMF). The FSPS models allow multiple aspects of the model creation to be modified, however for the purpose of comparison, we use these models in their default state. Unless explicitly stated otherwise, from now on ‘FSPS’ refers to the model set described above.

Vazdekis (2016)

The current Vazdekis models (hereafter V16) incorporate a number of recent updates (Röck et al., 2015, 2016, Vazdekis et al., 2010, 2012, 2015), which extend the original Vazdekis models (2003) from 1,680–50,000 Å. Models are available for two sets of stellar tracks: Pietrinferni et al. (2004, hereafter BaSTI) and Girardi et al. (2000). We utilise the BaSTI tracks, which use the synthetic AGB technique to cover the full AGB phase, including up to the end of the thermal pulses. The spectral libraries used in the optical and near-infrared ranges are MILES and IRTF (Rayner et al., 2009) respectively. IRTF is an empirical spectral library covering the J, H and K bands (9500–24,000 Å) at medium resolution ($R \sim 2000$). The library contains ~ 200 stars, mostly of solar metallicity, including late-type stars, AGB, carbon and S stars. Models are available for ages 1–15 Gyr only, for metallicities between -0.40 and 0.26, for single and double power-law IMFs with slopes between 0.3 and 3.3.

Model summary

Table 2.3 summarises the default basis set we have chosen for each group of models. The basis sets were chosen so as to cover approximately the same parameter space for all models, as well as span the range of properties derived for our sample using optical spectroscopy. As the IRTF spectral library can generate models only for $t \geq 1$ Gyr, and the Pickles library for solar metallicity only, this is the range we use for all models. We use a Salpeter IMF for all fits.

2.4.2 Star formation histories from full spectral fitting

We use the penalised pixel fitting code (pPXF)⁵ by Cappellari & Emsellem (2004) to fit a linear combination of basis spectra to the data for each model set. The code fits the entire input spectrum, with the exception of regions strongly affected by telluric absorption, which are masked out. We note that, while there are many different spectral fitting codes and techniques available (e.g. STECKMAP Ocvirk et al. 2006, MOPED Heavens et al. 2000, NBURSTS Chilingarian et al. 2007, STARTLIGHT Cid Fernandes et al. 2005, FIREFLY Wilkinson et al. 2015), cross-comparison of these various techniques have been shown to yield rather consistent results (e.g. Ferré-Mateu et al., 2012, Koleva et al., 2008, Maksym et al., 2014, Mentz et al., 2016, Wilkinson et al., 2015). In

⁵<http://purl.org/cappellari/idl>

particular, Goddard et al. (2017) explore the effect of changing fitting code and models, using Firefly and STARLIGHT, and conclude that in order to understand population model effects one should keep the fitting approach the same. We therefore keep the fitting approach fixed, in order to investigate the differences introduced by the choice of models, where detailed comparisons are less well documented.

pPXF convolves the model templates with a Gaussian line-of-sight velocity distribution, which broadens the model SSPs in order to best fit the data. The convolved templates are then fit to the data as the code searches for the minimum of the equation

$$\chi^2 = \sum_{n=1}^N \frac{(M - O)^2}{\sigma^2} \quad (2.1)$$

where M is the model spectrum, O is the observed spectrum, and σ is the error spectrum. The fractional weights assigned to each template give the mass-weighted contribution of that population to the observed spectrum.

We use regularisation, which is a standard way to solve ill-posed problems (Press et al., 1992). This is implemented in pPXF via the `REGUL` keyword, which penalises the χ^2 value when neighbouring weights do not vary smoothly. A larger `REGUL` value leads to a smoother output, by effectively reducing the error associated with the linear regularisation constraints. We use a fixed regularisation value of 50, which is consistent with the typical weights returned by pPXF. This gives a unique solution, i.e. it returns the smoothest solution out of the many degenerate solutions that are equally consistent with the data. The mass-weighted mean age of each galaxy is calculated as the average of the template ages weighted by the template weights, and does not depend strongly on the chosen regularisation value.

Preparatory Steps

Prior to the fit, models were resampled onto pixels of constant linear wavelength where necessary. To account for differences in resolution between data and models, we convolved them using a Gaussian of FWHM given by:

$$\text{FWHM}_{\text{Gaussian}} = \sqrt{(\text{FWHM}_{\text{max}}^2 - \text{FWHM}_{\text{min}}^2)} \quad (2.2)$$

In every case except V16, the data were higher resolution than the models, and had to be degraded to carry out the fit. The observed galaxy spectrum and the model of interest were then rebinned to a log scale so as to have a constant velocity step. The log rebinning was done first on the model/data with the coarser velocity step size and then the data/model was rebinned to the same velocity step. As such, the size of the velocity step varied for the different data-model combinations (from ~ 45 km/s for V16 and BC, for which the step size was set by the data, 98.5 km/s for M11 and 787 km/s for FSPS). Regions of poor atmospheric transmission between the J H & K bands were masked⁶, as well as other regions where telluric removal was particularly difficult.

An initial fit was carried out using all templates simultaneously, including high order additive polynomials. Additive polynomials change the strengths of absorption features, so are not recommended for stellar population analyses, but are useful at reducing template mismatch when deriving kinematics. We thus obtained the best numerical fit possible and secured the correct velocity and velocity dispersion values, which were then fixed in the subsequent fits, reducing the freedom pPXF has to compensate the template choice with varying kinematic parameters.

⁶Poor atmospheric transmission is defined as where the atmospheric transmission model of Lord (1992) is $\leq 50\%$, assuming a water vapour column of 1.6mm and an airmass of 1.0.

Treatment of the Continuum

There are two common strategies when carrying out spectral fitting: one uses a high order polynomial to fit the continuum, such that derived properties are based primarily on the strengths of strong, narrow absorption features in the spectrum. The other is to preserve the continuum, taking advantage of the extra information it encodes.

We implemented both approaches to determine the effect of each on the derived star formation histories (SFHs). In theory, keeping the continuum shape provides more information on the stellar populations present in the galaxies, however it assumes that both models and data are accurately flux calibrated, and that extinction is understood. In this case, we use the well-known extinction curve of Calzetti (Calzetti et al., 2000), where dust is treated as a simple foreground screen. This option is implemented in pPXF, and characterised by an output colour excess $E(B - V)$.

We also tested the effect of including a multiplicative polynomial. Multiplicative polynomials compensate for large-scale model and data deficiencies driving the fit without affecting the relative strength of absorption lines. We use order 10, as this was found to remove large-scale discrepancies without fitting intrinsic galaxy features.

Noise estimate and χ^2

The χ^2 calculated by pPXF depends on an estimate of the noise, as seen in Equation (2.1). To be able to meaningfully rank the models, all of which are rebinned and interpolated in different ways, is not straightforward. Initially we calculated the noise in each galaxy as the standard deviation of the residuals obtained from a good fit. However, the χ^2 calculated by pPXF was found to depend on the log rebinning and resolution correction convolution, as both of these processes introduce covariance between pixels, causing the noise estimate to be underestimated for particular models.

As these corrections are model dependent, we estimated the noise independently for every model set. As before, we carried out a good quality fit using the model with the highest quality near-infrared spectral library (V16) and high order polynomials to ensure a good fit to all the intrinsic galaxy features seen in the data. However, we matched the resolution and velocity step size used in the fit to that of the model of interest, giving a unique noise estimate for every model set, and χ^2 values that can be compared between models.

Age estimate uncertainties

We used a Monte Carlo approach to estimate uncertainties in the derived mean ages. Random noise was added to the galaxy spectrum which was consistent with the derived error spectrum. 500 realisations were generated for each object, and we again performed full spectral fitting on each of these new realisations, without regularisation. The error in the mean ages was taken to be the standard deviation of the output age array, divided by $\sqrt{2}$, to account for the fact we added noise onto an already noisy spectrum. This does not take into account systematic errors in the models themselves.

2.4.3 Line strengths

The measurement of line strength indices, when compared with the predictions of SSP templates, provides an alternative comparison of different stellar population models, which is largely insensitive to uncertainties in flux calibration. To measure the strength of atomic absorption features we follow a procedure similar to that used for the measurement of the well-known Lick indices in the optical (Worthey et al., 1994). Line strength indices are typically defined by a central bandpass covering the feature of interest, flanked by two pseudocontinuum bandpasses, one to the red side

Table 2.4: Index Definitions.

Index	Blue Pseudocontinuum (μm)	Central Bandpass (μm)	Red Pseudocontinuum ^a (μm)	Units	Source ^b	Notes
CN 1.1	1.0770 - 1.0830	1.0970 - 1.1030		mag	1	ratio
Na 1.14	1.1337 - 1.1367	1.1367 - 1.1420	1.1420 - 1.1450	\AA	2	
CN 1.4	1.3800 - 1.3900	1.4130 - 1.4230		mag	3	ratio
C ₂ 1.8	1.7520 - 1.7620	1.7680 - 1.7820		mag	1	ratio
Na 2.21	2.1934 - 2.1996	2.2000 - 2.2140	2.2150 - 2.2190	\AA	4	
Fe a 2.2	2.2133 - 2.2176	2.2250 - 2.2299	2.2437 - 2.2497	\AA	5	
Fe b 2.2	2.2133 - 2.2176	2.2368 - 2.2414	2.2437 - 2.24971	\AA	5	
Ca 2.2	2.2450 - 2.2560	2.2577 - 2.2692	2.2700 - 2.2720	\AA	5	
Mg 2.2	2.2700 - 2.2720	2.2795 - 2.2845	2.2850 - 2.2874	\AA	5	
CO 2.3	2.2460 - 2.2550	2.2880 - 2.3010	2.2710 - 2.2770	mag	6	

^aIn the case of CO, ‘Red Pseudocontinuum’ indicates the most redward pseudocontinuum band, which is still blueward of the feature itself. ^b**Sources.** (1) Gonneau et al. (2016) ; (2) Smith et al. (2015) ; (3) newly defined in this work; (4) Riffel et al. (2011) ; (5) Silva et al. (2008) ; (6) Mármol-Queraltó et al. (2008)

of the feature and one to the blue. The strength of the feature is then calculated as an equivalent width:

$$EW = \int_{\lambda_1}^{\lambda_2} \left(1 - \frac{S(\lambda)}{C(\lambda)} \right) d\lambda \quad (2.3)$$

where $S(\lambda)$ is the observed spectrum and $C(\lambda)$ is the local pseudocontinuum, which is given by the linear function connecting the average flux in the pseudocontinuum bands. For one sided molecular ‘break’ features such as CN and C₂, we calculate index strength as in Gonneau et al. (2016):

$$I(X) = -2.5 \log_{10}[F_b(X)/F_c(X)], \quad (2.4)$$

where $F_b(X)$ and $F_c(X)$ are the fluxes in the absorption band region and the pseudo-continuum region of index X.

The two main contributions to the statistical uncertainties in measuring line indices come from the statistical Poisson and detector noise in the index pass bands, and from errors in the measured radial velocity. The errors on the strengths of the features were calculated using Monte Carlo simulations, by shifting the bandpasses randomly according to the velocity error obtained from pPXF, after adding random noise at a level consistent with the observed error spectrum.

The bandpasses and pseudo-continua we use for each index are listed in Table 2.4. The models and data each have a different native resolution, thus, we degrade to the lowest common resolution before measuring indices. We did not correct for the effect of differing velocity dispersion broadening, as, compared to the difference in instrumental resolution, the difference in index strengths due to σ was negligible.

2.5 Results

2.5.1 Quality of the fits

First we focus on how well each model is able to fit the observed near-infrared data. Figure 2.4 shows the fits obtained for the first half of the sample using a Calzetti extinction curve, and Table A.1 gives the reduced χ^2 values (χ^2_{red}) for these fits, as well as the mass-weighted mean age and the $E(B - V)$ value.

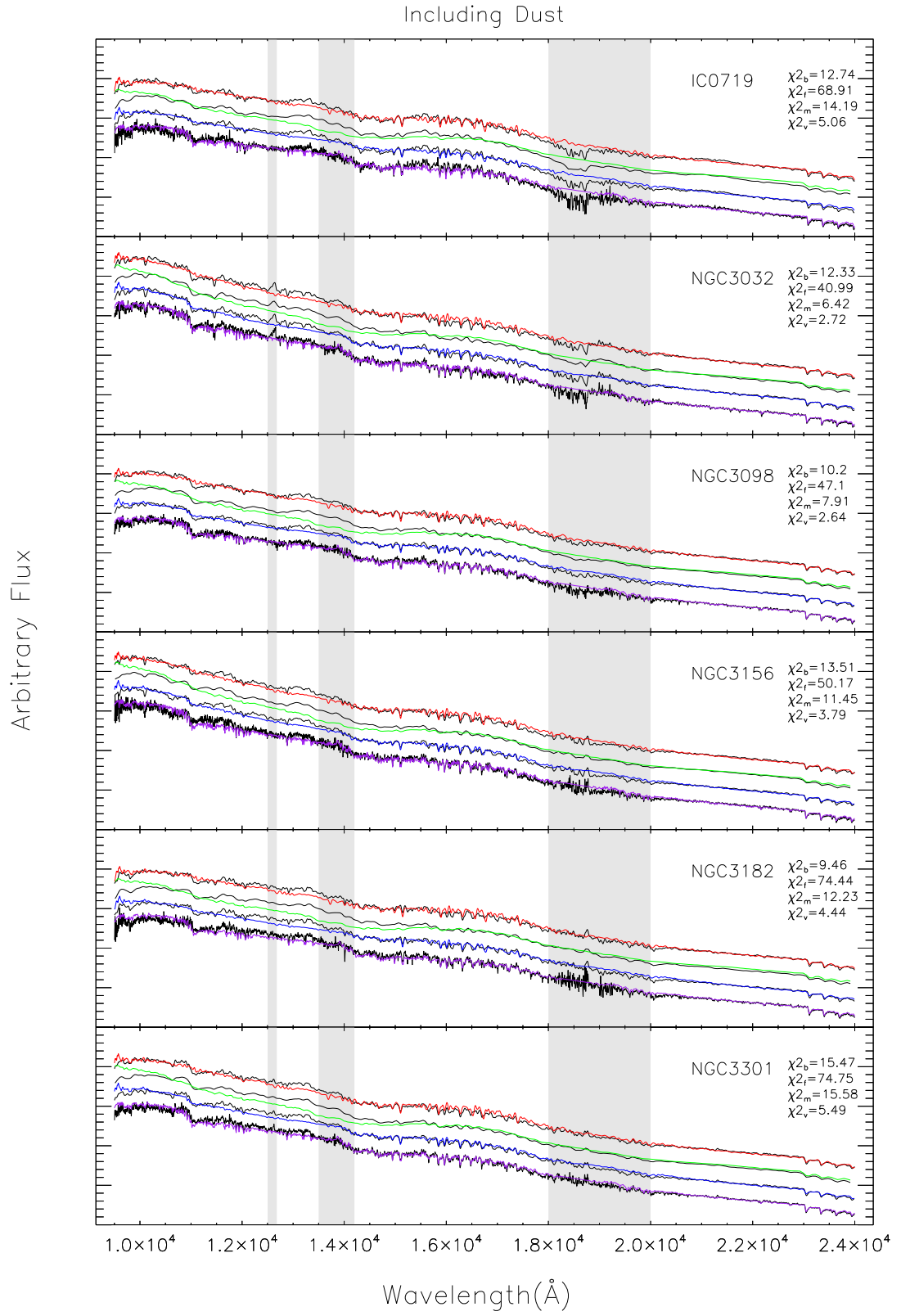


Figure 2.4: Comparison of spectral fits using the various models for 6 galaxies in the sample. Fits are carried out with a Calzetti extinction curve. The BC03 best fitting template is plotted in red, FSPS in green, M11 in blue and V16 in purple. Regions masked due to strong atmospheric absorption are shown in grey.

As discussed in section Section 2.4.2, the χ^2_{red} values can be used to rank the models in terms of how well they are able to fit the data. Table A.1 shows that when using an extinction curve the V16 models give the best fits (mean $\chi^2_{\text{red}} = 4.9 \pm 2.8$), followed by M11 and BC03 (mean $\chi^2_{\text{red}} = 14.2$ and 14.3 respectively), while FSPS fit the worst (mean $\chi^2_{\text{red}} = 83.2$). Inspection of Figure 2.4 shows that the V16 models generally fit very well across the entire near-infrared range. M11 and BC03 also do a good job at reproducing the overall shape of the spectra, although there are particular areas where the match is poorer (typically around $1.3 \mu\text{m}$ in both models). This could be due to the Pickles stellar library, which is input to both models. As previously mentioned, the Pickles library does not have complete spectral coverage for all stellar types over the entire near-infrared wavelength range. The FSPS models are the most different to the data in terms of overall spectral shape, which is reflected in the poor values of χ^2_{red} obtained with these models. FSPS also consistently requires the largest amount of reddening to match the data. This is inconsistent with SDSS imaging, which shows many of the galaxies have no obvious sign of dust. In many cases, the V16 and BC03 models can match observed shape without invoking any reddening at all, with the notable exception of NGC 4710, which is visibly dusty at both optical and near-infrared wavelengths. This mismatch is attributable to the low resolution BaSeL library used as input to the FSPS models, as we show in Section 2.5.3. The fact that BC03, V16 and M11 are able to well reproduce the overall shape of the data indicates that the flux calibration is likely not significantly wrong in either the data or these models, or at least is not responsible for all the apparent differences between the models.

Fitting with a polynomial alleviates the problem of shape mismatch between models and data, and the quality of fit becomes driven by the ability of the models to match the strengths of specific absorption features. Figure 2.5 shows the fits obtained for the first half of the sample using a multiplicative polynomial of order 10, and Table A.3 gives the χ^2_{red} values and mean ages. The ranking of the models in terms of their χ^2_{red} values now becomes V16 (mean $\chi^2_{\text{red}} = 1.5 \pm 0.3$), followed by M11 (mean $\chi^2_{\text{red}} = 3.6 \pm 0.8$), FSPS (mean $\chi^2_{\text{red}} = 4.5 \pm 1.0$), BC03 (mean $\chi^2_{\text{red}} = 5.0 \pm 1.0$). Inspection of Figure 2.5 and Figure 2.6, which shows the residuals for each fit, allows us to pinpoint which specific features are poorly fit by the various model sets. The BC03 models are unable to fit the CN feature at $1.1 \mu\text{m}$, as these models do not contain this feature. The fit to the numerous CO features in H- and K-band, as well as a number of atomic features in K band are also not perfectly matched. FSPS are also unable to fit CN1.1 well, despite containing a CN feature in the models, as the shape of the models in this region does not look similar to the data. V16 fit very well across the entire spectral range, but slightly overpredict CN1.1 strength at all ages. M11 also typically fit most features across the spectral range well, slightly underpredicting the strength of CN $1.1 \mu\text{m}$ and the CO bandheads at $1.6 \mu\text{m}$ and $2.3 \mu\text{m}$. There is a significant residual visible in all model fits of the two post-starburst galaxies (NGC 3032 and NGC 3156) at $1.005 \mu\text{m}$. This is most likely Pa- δ absorption from A-type stars, which have lifetimes of ~ 100 Myr, and thus are not included in our selected set of models, for which the youngest template is 1 Gyr. This is further evidence that these two galaxies experienced a recent burst of star formation, supporting the strong H β absorption visible in the optical spectra of these galaxies shown in Figure 2.3.

2.5.2 Comparison of Star Formation Histories

We next compare the galaxy ages and star formation histories that are derived using the various models. Figure 2.7 shows representative optical and near-infrared fits, and the SFHs derived from them, for NGC 4578, using a multiplicative polynomial. The optical fits typically have lower χ^2_{red} than the near-infrared fits. SFHs derived from optical spectroscopy of all models typically agree very well, and correlate well with the SSP ages from ATLAS^{3D} (McDermid et al., 2015). In particular, all models derive young ages for our two post-starburst galaxies (NGC 3032 and NGC 3156) when fitting their optical spectra. When fitting near-infrared spectra alone, however,

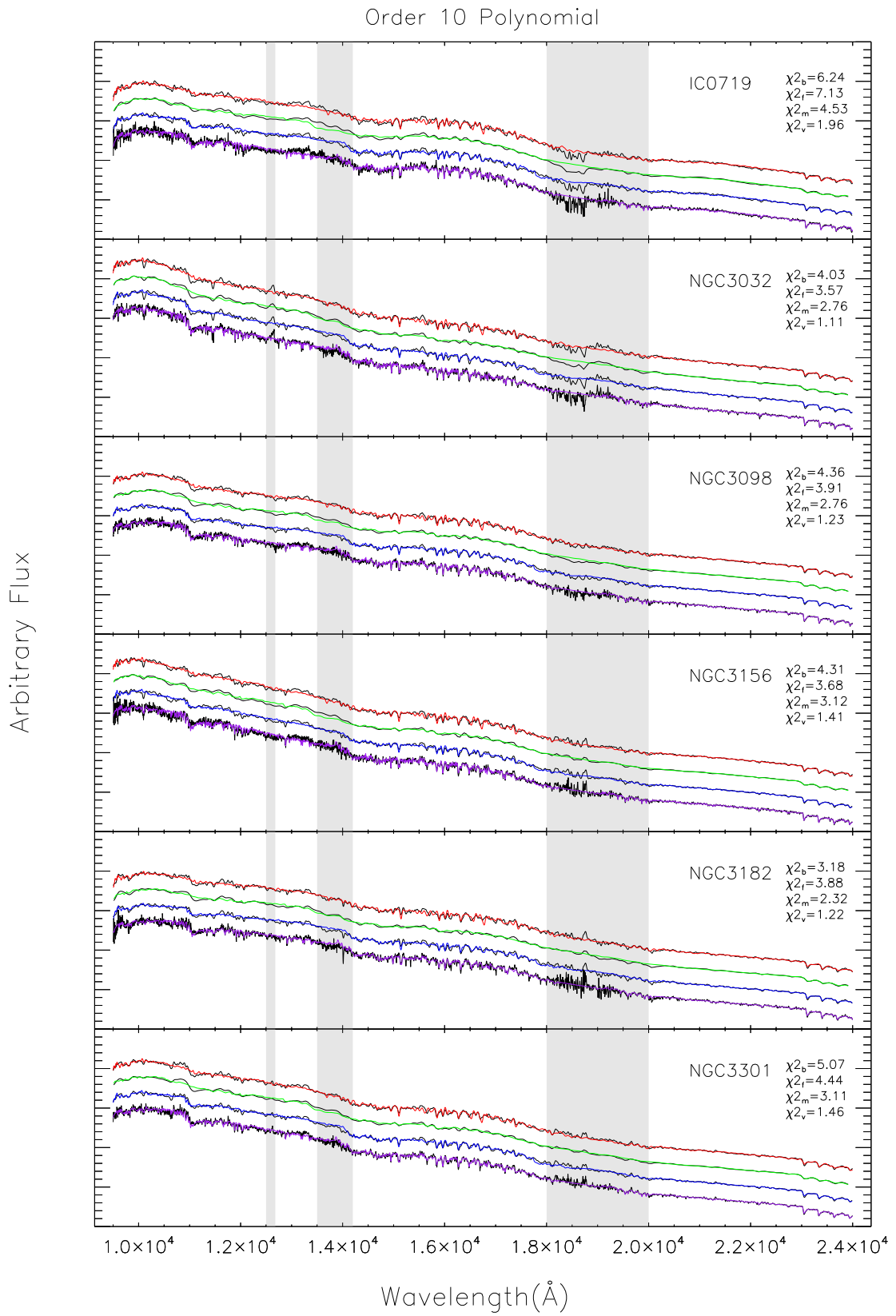


Figure 2.5: As in Figure 2.4, but with a multiplicative polynomial instead of extinction curve.

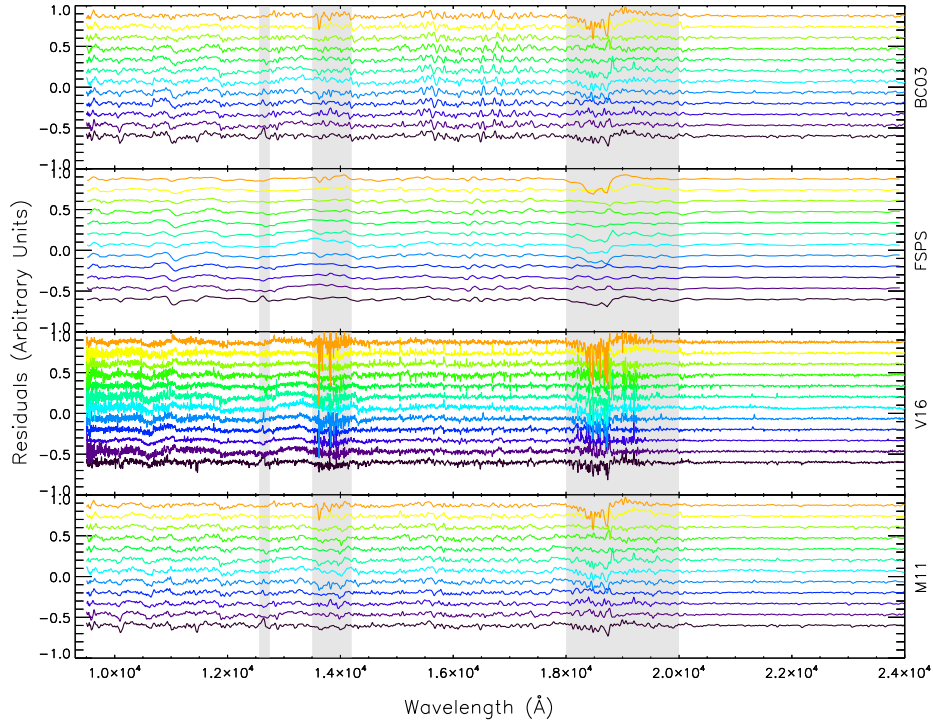


Figure 2.6: Fit residuals for each model. Fits were carried out with an order 10 multiplicative polynomial. Each panel shows residuals for a particular model set (top to bottom : BC03, FSPS, V16, M11). Galaxies are ordered by age, with age increasing upwards. Masked regions are shown in grey. Note that the spectral resolution varies strongly between models.

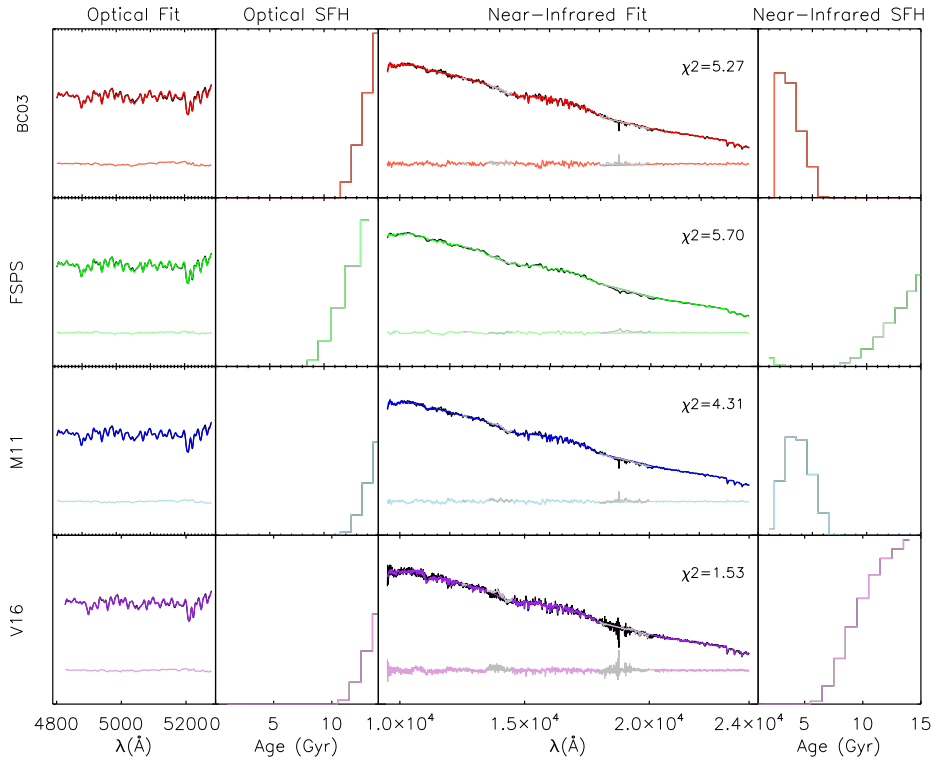


Figure 2.7: NGC 4578: Comparison of spectral fits and star formation histories obtained from the four models, as labelled on the left axis, using a multiplicative polynomial. Panels show (from left to right) the optical spectra: observed (black), best fit and residuals (coloured); the star formation history derived from the optical fit; the near-infrared spectra (coloured as before); the star formation history derived from the near-infrared fit.

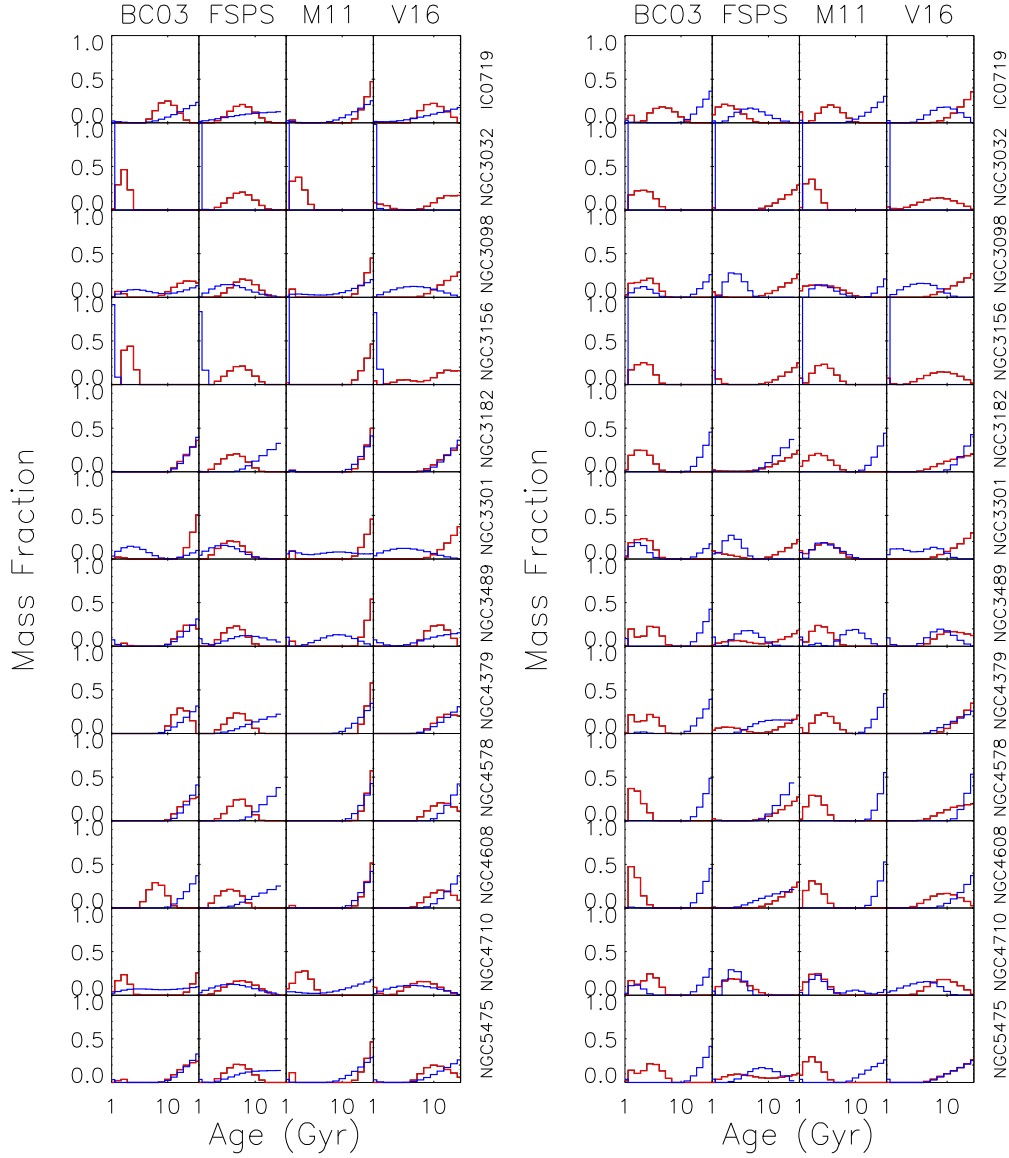


Figure 2.8: Star formation histories derived from the near-infrared spectra (red) compared with the optical spectra (blue). On the left we show fits carried out using a Calzetti extinction curve, and on the right using a multiplicative polynomial.

the derived SFHs are model dependent, and also depend strongly on treatment of the continuum. Figure 2.8 shows the SFHs derived for the entire sample from both wavelength ranges. The optical SFHs are plotted in blue and the near-infrared SFHs in red. As the results depend on the treatment of the continuum, we analyse the two cases separately.

Reddening case

Regardless of wavelength range, BC03 identifies post-starbursts NGC 3032 and NGC 3156 as young, and NGC 4710 as containing a substantial fraction of young stars. Both wavelengths also require smaller fractions of young stars in NGC 3098, NGC 3301, NGC 3489 and NGC 5475. The remainder of the galaxies are solely old in both wavelength ranges, with the curious exception of NGC 4608, which appears younger in the near-infrared than it does in the optical. Use of the M11 models in the near-infrared also leads to young ages for NGC 3032 and NGC 4710, in agreement with the optical, however the post-starburst galaxy NGC 3156 appears old in this case. As with BC03, M11 models require small fractions of young stars in NGC 3098, NGC 3301, NGC 3489,

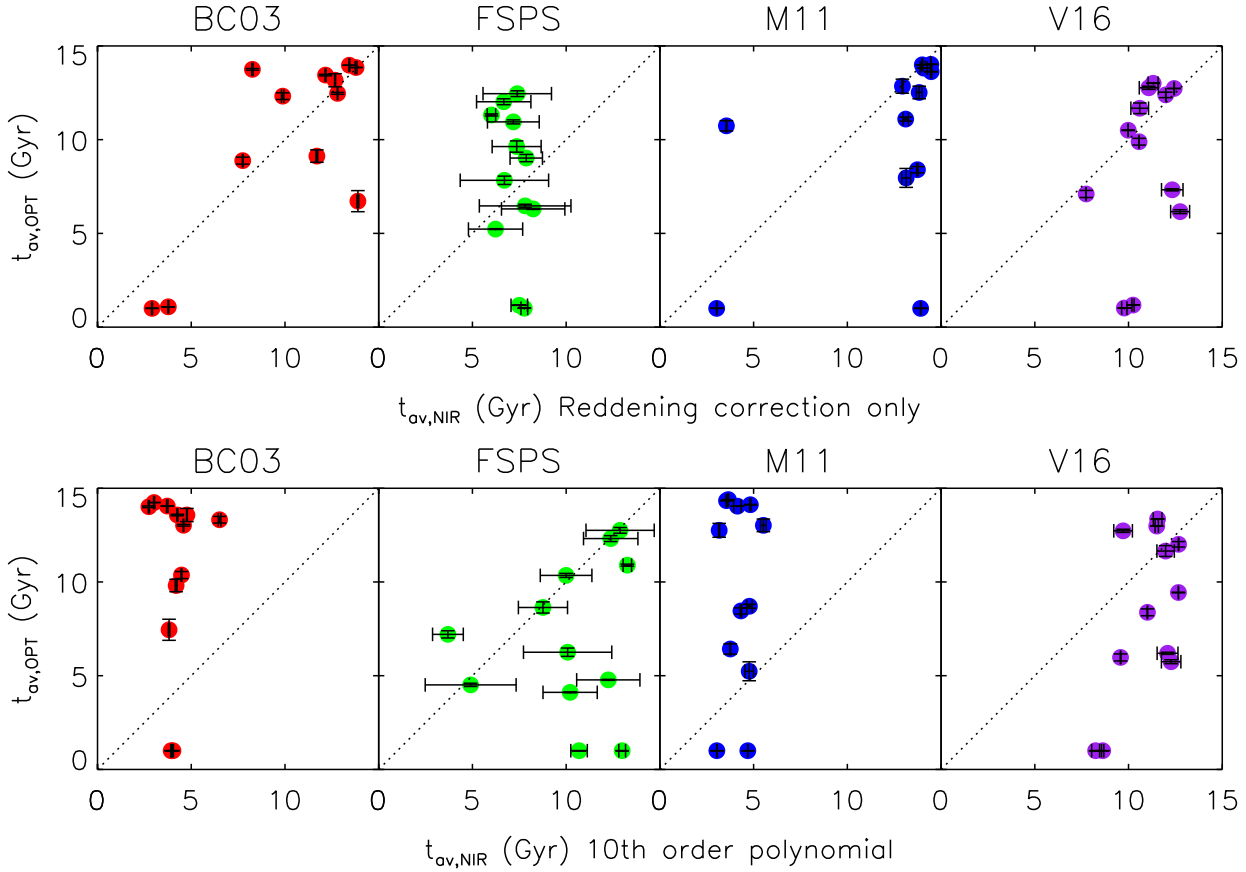


Figure 2.9: Comparison of mass-weighted mean ages derived using the optical and near-infrared spectra for each model set. Fits are carried out with a Calzetti extinction curve (top) and multiplicative polynomial (bottom). Note that the spectral library used in FSPS here is BaSel, compared with Figure 2.13, which uses an updated spectral library.

NGC 4608 and NGC 5475. The remainder of the galaxies appear old at both wavelengths. The V16 near-infrared fits imply a small burst in NGC 3032, atop a pre-existing old population, which is the most physically likely scenario. NGC 3156 and NGC 4710 also have extended SFHs, however the mass-weighted mean ages derived using V16 are all old ($t \gtrsim 8$ yr), contrary to the optical SFHs, which span a large range of mean ages. Galaxies such as NGC 3098 and NGC 3301 have rather inconsistent SFHs when comparing optical and near-infrared results. The near-infrared FSPS fits do not appear to be very strongly age sensitive, with most galaxies appearing ‘middle-aged’ ($t \approx 6 - 8$ Gyr), regardless of the optical ages (but see Section 2.5.3, where we show that this is likely due to the BaSel library used in these models). In Figure 2.9, we summarise these results by comparing, for each model, the mass-weighted ages of our sample derived from both wavelength ranges.

Polynomial case

While use of a reddening law with the M11 and BC03 models led to SFHs which bore some resemblance to the optical ones, modelling the continuum with a multiplicative polynomial causes these models to effectively lose all sensitivity to age in the near-infrared. In this case, the BC03 and M11 SFHs all look incredibly similar, giving mass-weighted mean ages which all lie between $\approx 3 - 7$ Gyr despite the sample spanning ≈ 13 Gyr in optically derived mass-weighted age (see Figure 2.9). The V16 near-infrared SFHs are similar to those derived with reddening: all galaxies appear largely old, but NGC 3032 and NGC 3156 have somewhat more extended SFHs, resulting in mean ages of ≈ 8 Gyr for these two, which, while still old, are the youngest in the sample. The FSPS SFHs display the opposite behaviour to the other three models: in this case the polynomial fits result in SFHs that vary strongly from galaxy to galaxy, however they do not display a clear relationship with the optically derived SFHs.

2.5.3 Model Ingredients

This section explores the effects of various model ingredients on derived physical properties by carrying out spectral fits using different model permutations. From now on, all fits are done with a Calzetti extinction curve rather than a polynomial.

Non-solar metallicities

M11 and BC03 predictions using the Pickles library are available only for solar metallicity. So as to compare the various model sets on an equal basis, we fit using only solar metallicity templates for every model. The sample was selected (based on optical SSP-equivalent parameters) to be approximately solar metallicity, so we expect this basis set to be adequate. We test this assumption by extending the metallicity coverage using the V16 models, which are available for multiple metallicities.

Figure 2.10 shows the SFHs derived using the default V16 basis set (black, solar metallicity only) and a basis set including five different metallicities from $-0.35 \leq [Z/Z_{\odot}] \leq 0.26$ (blue). For most galaxies, the difference is minimal. The difference in mean ages is typically less than 1 Gyr, with the addition of multiple metallicities usually leading to a marginally younger mean age. The SFHs themselves are very similar. The largest differences are seen in IC 0719 NGC 3032, NGC 3156 and NGC 4710, which look burstier when fitting with multiple metallicities. These galaxies all have clear signs of recent star formation at other wavelengths. As such, they are expected to contain young, super-solar populations. Even so, the results obtained using the extended basis set are not substantially different to those obtained using solar metallicity only.

Spectral libraries and stellar tracks

Most current model sets are available for a variety of input stellar tracks and spectral libraries which the user can choose from. Here we include a brief analysis of the effect of these choices on the derived stellar population properties.

First, we study the impact of stellar tracks. Figure 2.11 shows the SFHs derived with the V16 models using our default choice of the BaSTI tracks (black), as well as when using the Padova tracks (blue). There is a small but noticeable change in the derived SFHs, with the Padova tracks resulting in more extreme SFHs i.e. when using the Padova tracks, the post-starburst galaxies appear younger while the rest of the sample appear older. The change in the mass-weighted mean ages is typically less than 2 Gyr. The older ages are due to a known effect in the Padova (2000) tracks, whereby a hotter RGB at old ages leads to a bluer colour in these tracks compared to previous works by the Padova group, as well as other popular stellar evolution calculations. This

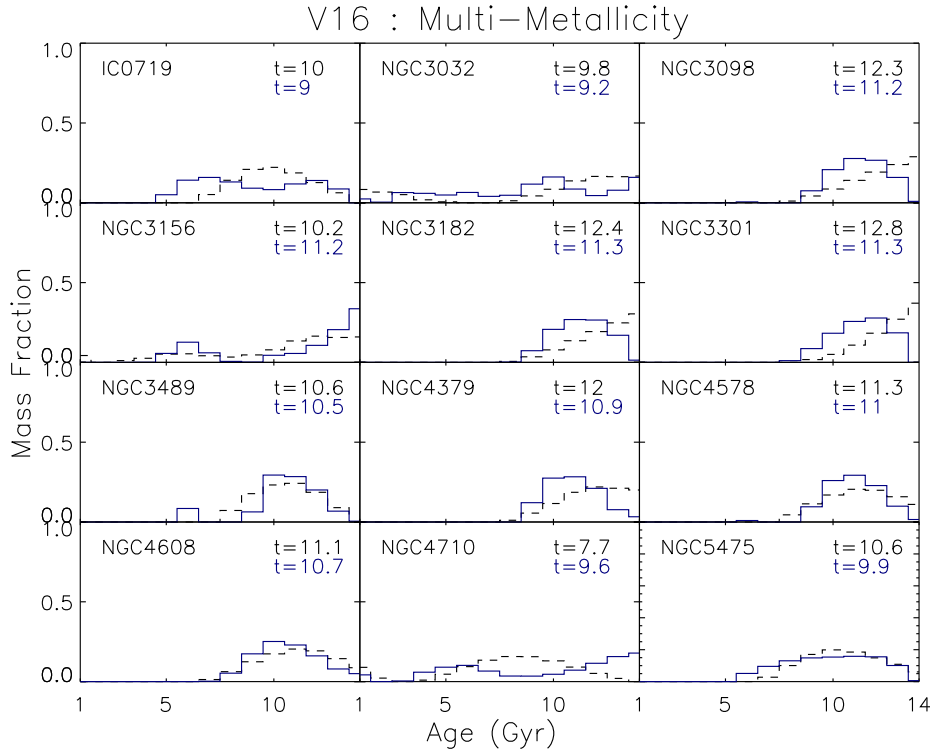


Figure 2.10: Effect of extending metallicity coverage using V16 models. The SFHs derived using our default basis set (solar metallicity only) are shown as black dashed lines, and the SFHs derived allowing metallicity to vary at each age from -0.35 to 0.26 are shown in blue. The derived mass-weighted mean ages are printed in each panel in Gyr. SFHs are derived using a reddening law.

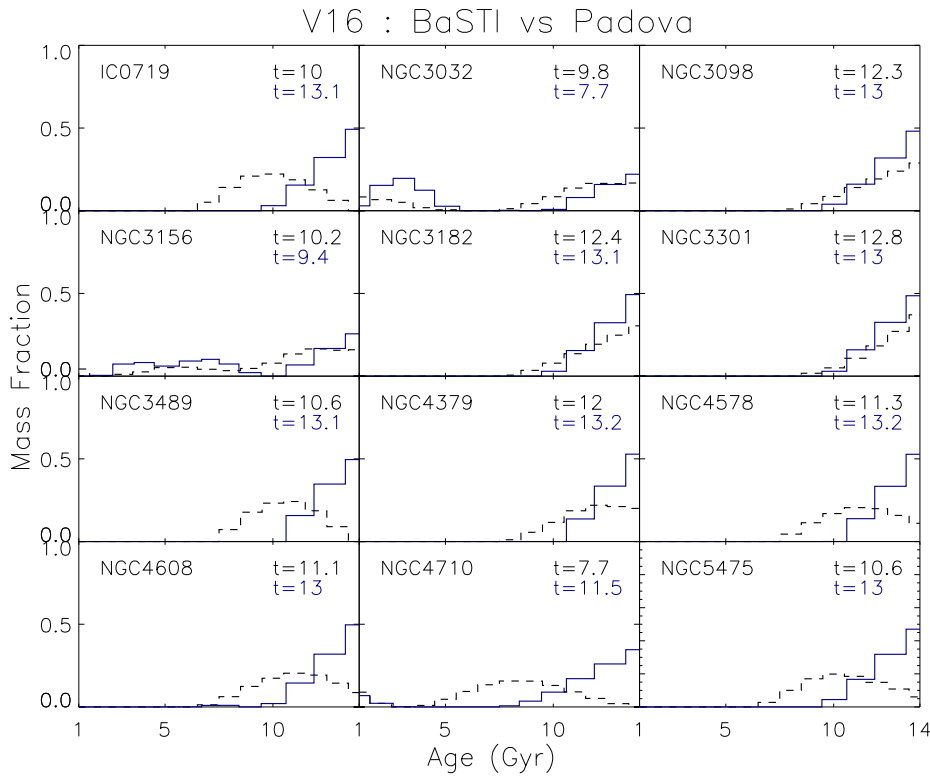


Figure 2.11: Effect of varying tracks using V16 models. The SFHs derived using our default tracks (BaSTI) are shown as black dashed lines, and the SFHs derived using the Padova tracks are shown in blue. The mass-weighted mean ages are printed in each panel in Gyr. SFHs are derived using a reddening law.

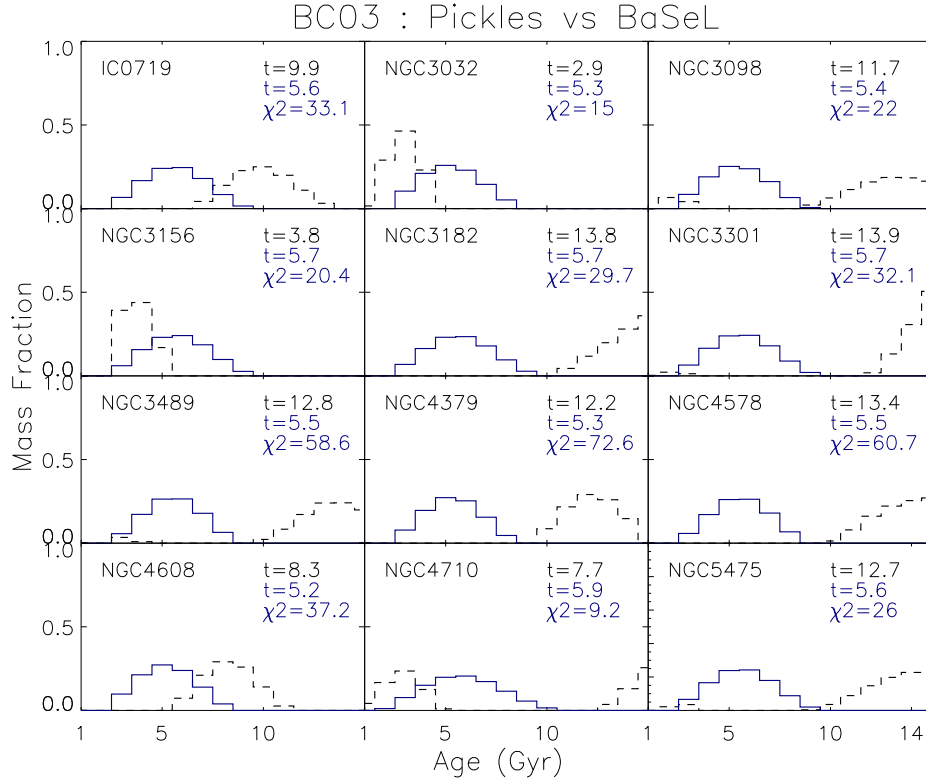


Figure 2.12: Effect of varying spectral libraries using BC03 models. The SFHs derived using our default library (Pickles) are shown as black dashed lines, and the SFHs derived using the BaSeL library are shown in blue. SFHs are derived using a reddening law. The χ^2_{red} of the BaSeL fits are printed in each panel, as well as the mass-weighted mean ages calculated for each library.

means that for the same colour, use of the Padova (2000) tracks will result in an older age (see e.g. Bruzual & Charlot (2003)).

Turning next to the choice of spectral libraries: we originally used BC03 with the Pickles empirical library (the same library used by M11), however the default BC03 library at near-infrared wavelengths is the Basel theoretical library (the same library used by the default FSPS). The FSPS fits had the largest χ^2_{red} and the resultant SFHs displayed the lowest correlation with the optical SFHs. To test whether this was due to choice of spectral library, we carried out full spectral fitting using BC03 with BaSeL. Figure 2.12 shows the SFHs derived with an extinction law using BC03-Pickles (black) and BC03-BaSeL (blue). There is a significant discrepancy between the two. When using BC03 with the BaSeL library, the derived SFHs are almost identical to those found previously using FSPS. The χ^2_{red} (plotted in each panel) are high ($\chi^2_{\text{red}} = 34.7 \pm 20$) and the SFHs are essentially insensitive to age, indicating that the BaSeL library was likely the primary cause of the of the poor behaviour seen when using the FSPS models.

The FSPS models have recently been updated (see Hutchinson et al., 2016) to include a new, high-resolution theoretical spectral library (C3K) as well as new MESA Isochrones and Stellar Tracks (MIST; Choi et al., 2016), and as a final test, we fit the sample using both of these updates. Figure 2.13 shows the mean ages derived from both optical and near-infrared spectroscopy using various FSPS setups with an extinction law. While the current publicly available FSPS models have no age sensitivity in the near-infrared, the situation drastically improves with use of the updated high resolution C3K library. Using FSPS-C3K-Padova on both wavelength ranges, one obtains better agreement between mass-weighted mean ages for the oldest galaxies. Use of FSPS-C3K-MIST brings the youngest galaxies into better agreement as well. The SFHs derived using FSPS-C3K-Padova are shown in Figure 2.14. The agreement between optical and near-infrared SFHs is

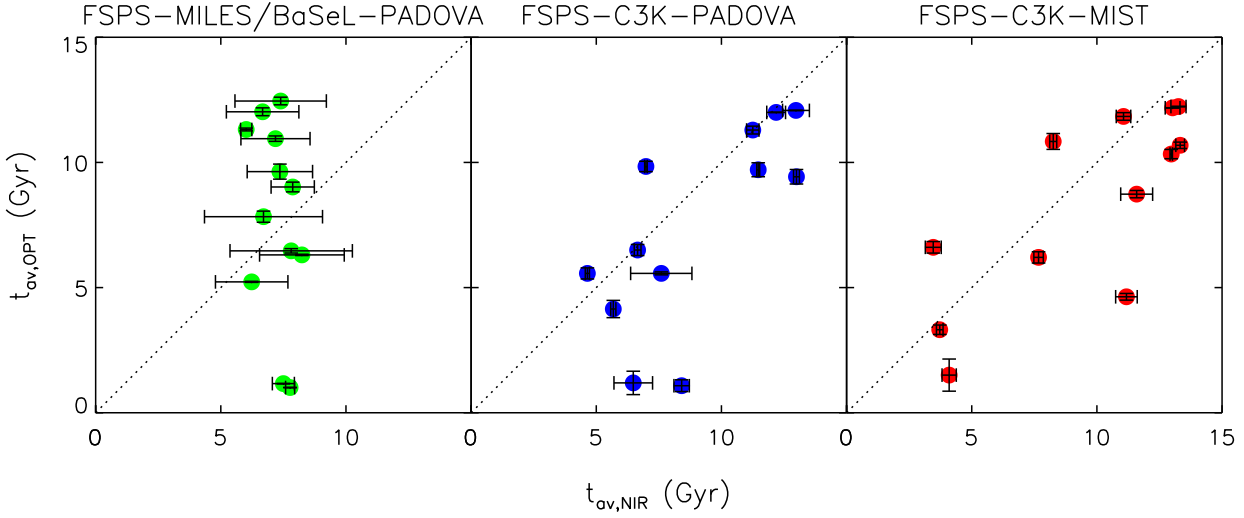


Figure 2.13: Comparison of mass-weighted mean ages derived from optical and near-infrared fitting with various FSPS setups, using a Calzetti extinction curve.

much better than seen previously. The largest discrepancies are seen in the post-starburst galaxies. Optically, these galaxies were formed entirely in the last 2 Gyr. The near-infrared SFHs contain young bursts on top of a substantial old component. This is a more physically realistic scenario than that implied by the optical - the optical result is likely due to the old population being outshone by the young stars at these wavelengths. The updated FSPS-C3K models give χ^2_{red} values comparable to V16 (mean $\chi^2_{\text{red}} = 4.8 \pm 2.4$ with an extinction curve). An example fit is shown in Figure 2.15 for each of the model versions and the derived properties are given Table A.5.

To summarise: the largest effect on the derived SFHs comes from the choice of stellar spectral library. Choice of stellar tracks, and metallicity coverage have only secondary effects on the SFHs derived from our near-infrared spectroscopy.

2.5.4 Line strengths

Figure 2.16 shows the strengths of our chosen indices compared with the predictions of the four models. Model predictions are plotted as coloured lines, and the galaxies (filled circles) are plotted against their optically-derived mass-weighted ages from ATLAS^{3D} (McDermid et al., 2015). These mass-weighted ages are less dominated by young stars, and thus more representative of the ages of the dominant population in the galaxies. All indices were measured at the resolution of the Pickles library (BC03/M11). The FSPS predictions plotted here as those using the C3K library. We chose to use this rather than the default FSPS-BaSeL, as those models are not line-by-line radiative transfer models, and are thus not suitable for detailed spectral predictions. As mentioned earlier, the Pickles library used in M11 and BC03 lacks spectroscopic observations for many of its stars, and as such, some near-infrared absorption features may be poorly resolved. We still include the measured Pickles line strengths, but for comparison, we also include index strengths measured in the M11 model in their version based on the MARCS library. This comprises a comprehensive grid of cool stars (2500-8000 K) at very high resolution ($R=20,000$) over a large wavelength range (1300-20,000 Å), allowing the calculations of SSP models with ages larger than 3 Gyr, for various chemical compositions (see www.maraston.eu/M11). As already noted, creation of model atmospheres in the near-infrared is not a trivial task, due to the multiplicity of lines and molecules to be taken into account when the temperature is low. This comparison can give us a qualitative idea on which indices allow a fair comparison with data.

Figure 2.16 shows that molecular indices CN1.1, CN1.4, and C_2 are predicted by the M11

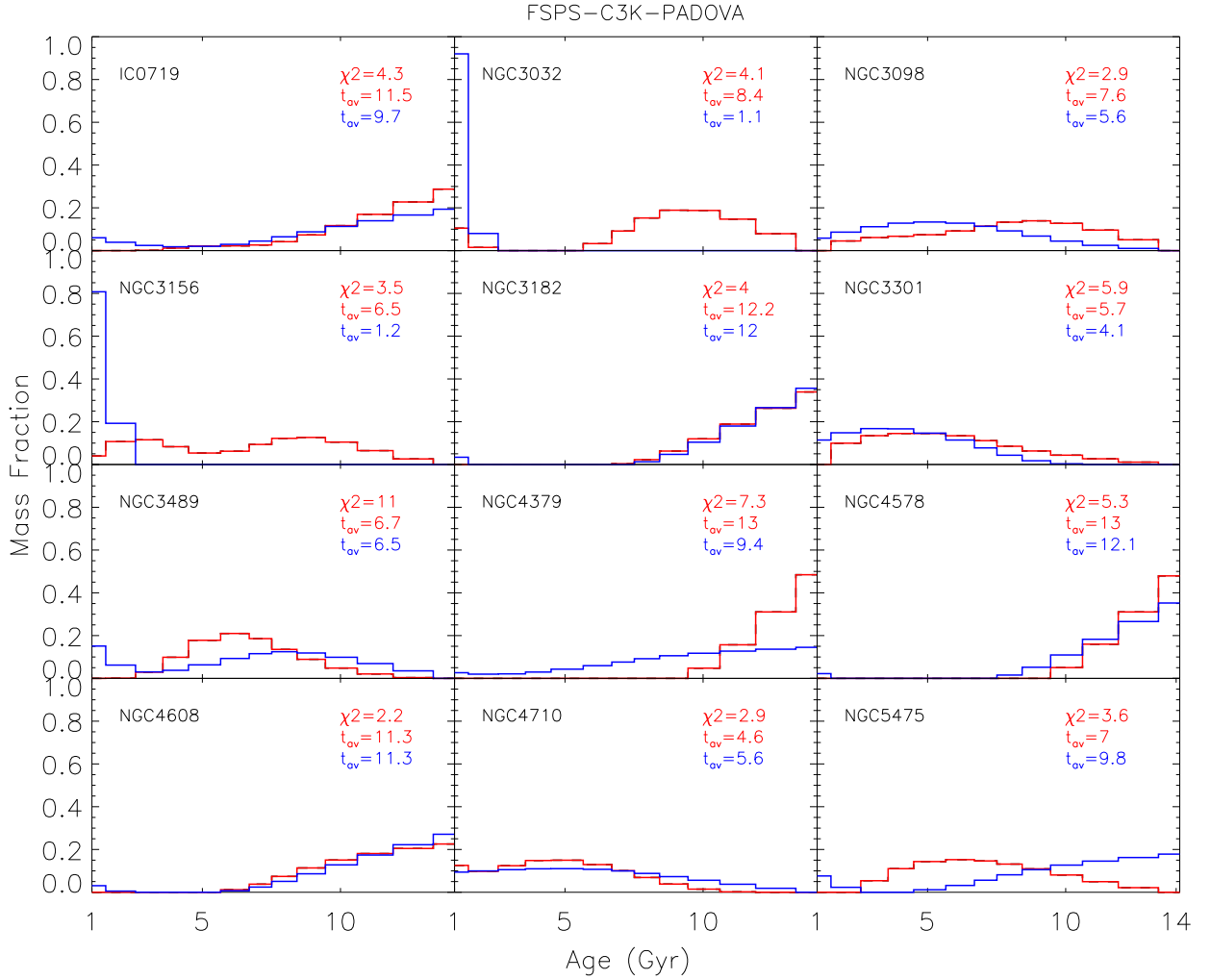


Figure 2.14: SFHs derived from optical (blue) and near-infrared (red) spectroscopy using FSPS-C3K-PADOVA models. The mass-weighted mean age derived in the near-infrared (optical) is given in each panel in red (blue) and the χ^2_{red} of the near-infrared fit is printed in red.

models to be strong in intermediate age populations (≤ 2 Gyr) which are affected by the TP-AGB phase. Other models predict no such upturn (or a much milder upturn) at these ages. The data do not show a strong increase in these indices at the youngest ages, rather the measured index strengths are approximately constant with age (though recall the data show mass-weighted average ages of integrated populations, not true SSPs). The strength of CN1.4 measured in the data is less than that predicted by V16 at all ages, and greater than predicted by all other model sets at most ages. CN1.1 displays similar behaviour, but the strength is accurately modelled by FSPS-C3K. The underprediction of CN by most model sets at old ages is likely due to deficiencies in the spectral libraries used. Use of the M11 models with the theoretical MARCS library rather than Pickles predicts CN1.1 strengths which are in better agreement with FSPS-C3K and the data. The strength of C_2 in the data is consistent with all models for old ages, despite a range in mass-weighted and SSP-equivalent ages that spans 1-15 Gyr i.e. there is no evidence of a significant contribution from TP-AGB stars in any of the galaxies in the sample. The molecular CO strength measured in the MARCS library stands out as being systematically lower than the data. Our CO index definition measures the strength of the first CO feature in the bandhead. In the MARCS models, this first CO feature is visibly weaker than the following features, contrary to the behaviour of other models.

Most atomic features are not predicted by the models to vary strongly as a function of age in the near-infrared range. Measurements support this lack of age variation, however the strengths

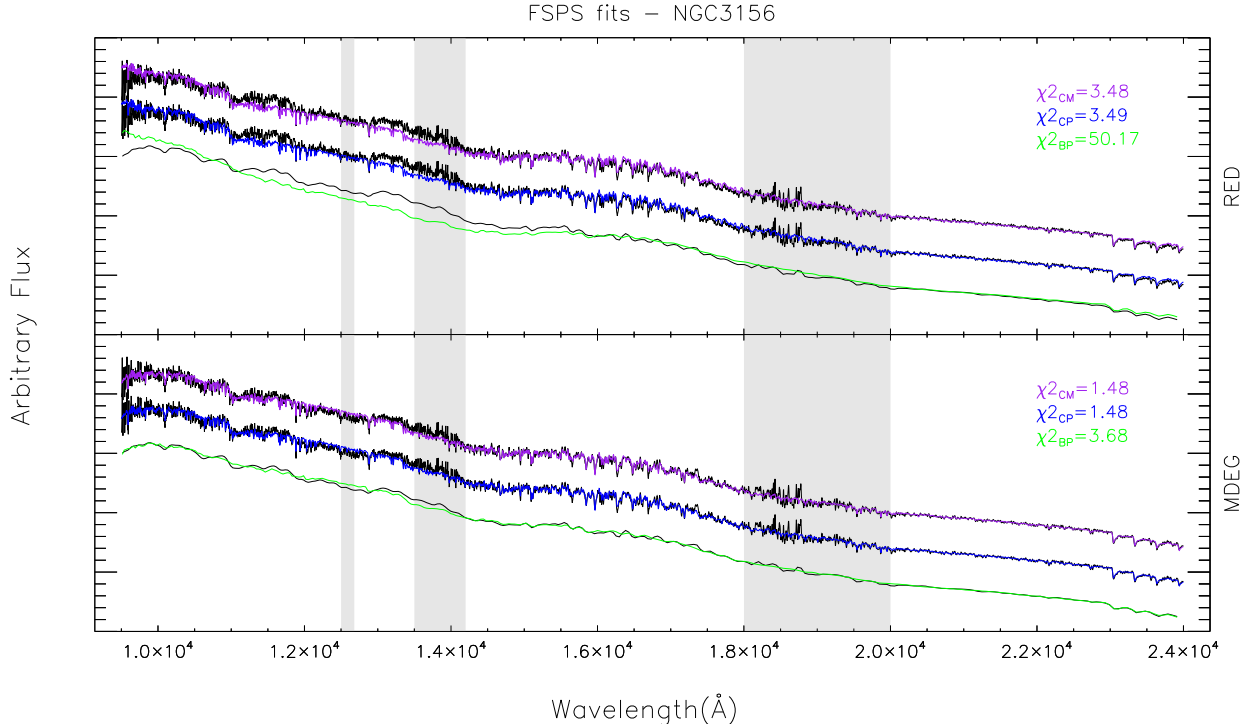


Figure 2.15: Comparison of spectral fits with the various FSPS versions using a Calzetti extinction curve (top panel) and an order 10 polynomial (bottom panel). From top to bottom: FSPS-C3K-MIST (update; purple), FSPS-C3K-PADOVA (update; blue), FSPS-BASEL-PADOVA (default; green). Masked regions are shown in grey.

of particular features are not fit equally well by all models. The BC03 and M11 models based on the Pickles library are outliers in the measured Na2.21 strength compared to other models. Both sodium features (Na1.14 and Na2.21) are somewhat enhanced in the data compared to model predictions (although approximately half the sample is consistent with V16 for Na1.14). An enhanced sodium contribution compared to model predictions has been seen in a number of recent works (e.g. Smith et al., 2015). Possible explanations for this discrepancy include enhanced $[\text{Na}/\text{Fe}]$ or a bottom heavy IMF (i.e. an IMF containing more dwarf stars than a canonical Kroupa IMF). Multiple studies have found evidence for a bottom heavy IMF in the centres of massive ellipticals (e.g. Conroy & van Dokkum, 2012, Ferreras et al., 2013, Spiniello et al., 2014, and others), however this is not expected in this sample based on the low σ_e of these galaxies. NGC 3032 ($t_{\text{SSP}} = 0.9$ Gyr) has unusually high Na and Ca compared to all other galaxies in the sample, which could be due to populations < 1 Gyr present in the galaxy, although the same enhancement is not seen in NGC 3156 ($t_{\text{SSP}} = 1.1$ Gyr).

2.6 Discussion

In the sections above, we compared the SFHs derived from optical and near-infrared spectroscopy using various models, as well as comparing the strengths of various near-infrared indices with the predictions of these models. Here we discuss some of the questions raised by this analysis.

2.6.1 Should optical and near-infrared properties agree?

One of the checks we carried out was to test the various models for self-consistency, which we did by deriving SFHs from both optical and near-infrared wavelength ranges using the same model

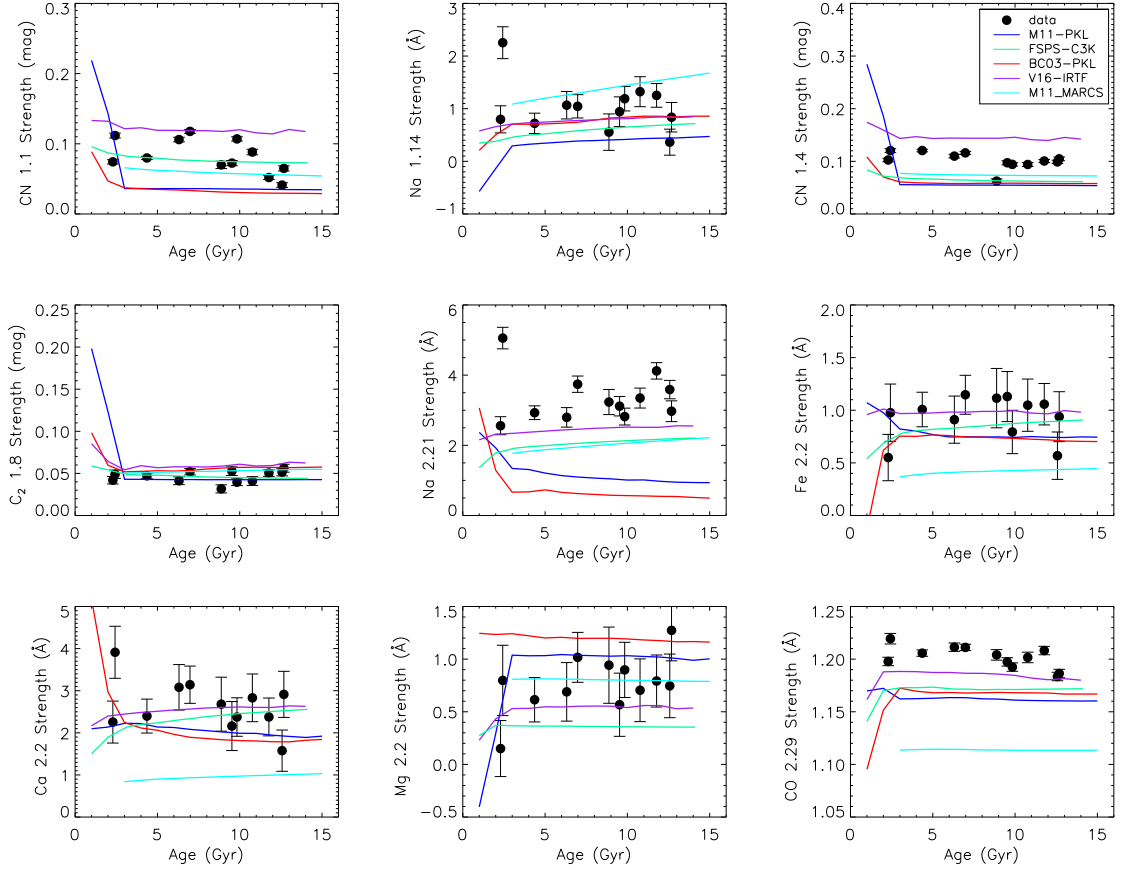


Figure 2.16: Strengths of chosen indices measured in the data, compared with the predictions of SPS models. The model predictions are plotted with coloured lines as a function of SSP age. The sample measurements are plotted against the galaxies’ optically-derived mass-weighted ages.

set. However, the question may be asked, should the properties derived using different wavelength ranges actually agree with one another?

In theory, if the models are correct, they should give consistent results at all wavelengths. In practice, however, this is not necessarily the case. Optical SSP-equivalent properties are known to be heavily biased towards the age of the youngest population present, as young stars are more luminous and blue, thus contributing more to the light-weighted values (Serra & Trager, 2007, Trager et al., 2000b). The near-infrared, on the other hands, is dominated by stars on the red and asymptotic giant branches, so SSP-equivalent ages derived in the two wavelength regimes may not necessarily agree in galaxies containing young stars.

The problem is somewhat alleviated, however, by the use of full spectral fitting to derive non-parametric SFHs. Fitting a linear combination of SSPs (compared to finding the single model template that most closely matches the strengths of particular absorption features) allows one to also include old populations which do not contribute substantially to the light. The situation is improved even further by enforcing regularisation in the fits. Regularisation effectively imposes a penalty on the χ^2 when the derived SFH is not smooth. While the requisite young populations will still be included in the SFH, the smoothness criterion ensures that the SFH includes as large a fraction of old templates as possible without significantly degrading the fit. In practice, the optical is still somewhat biased to young ages, as the old fraction is difficult to constrain given that the majority of the optical information is provided by the young populations. Inspection of Figure 2.8 shows that the agreement between the optical and near-infrared SFHs within a model set is much worse than the agreement between the optical SFHs of different models. This may result from the different SFH-constraining power of the two spectral regions, at least with the quality of data and

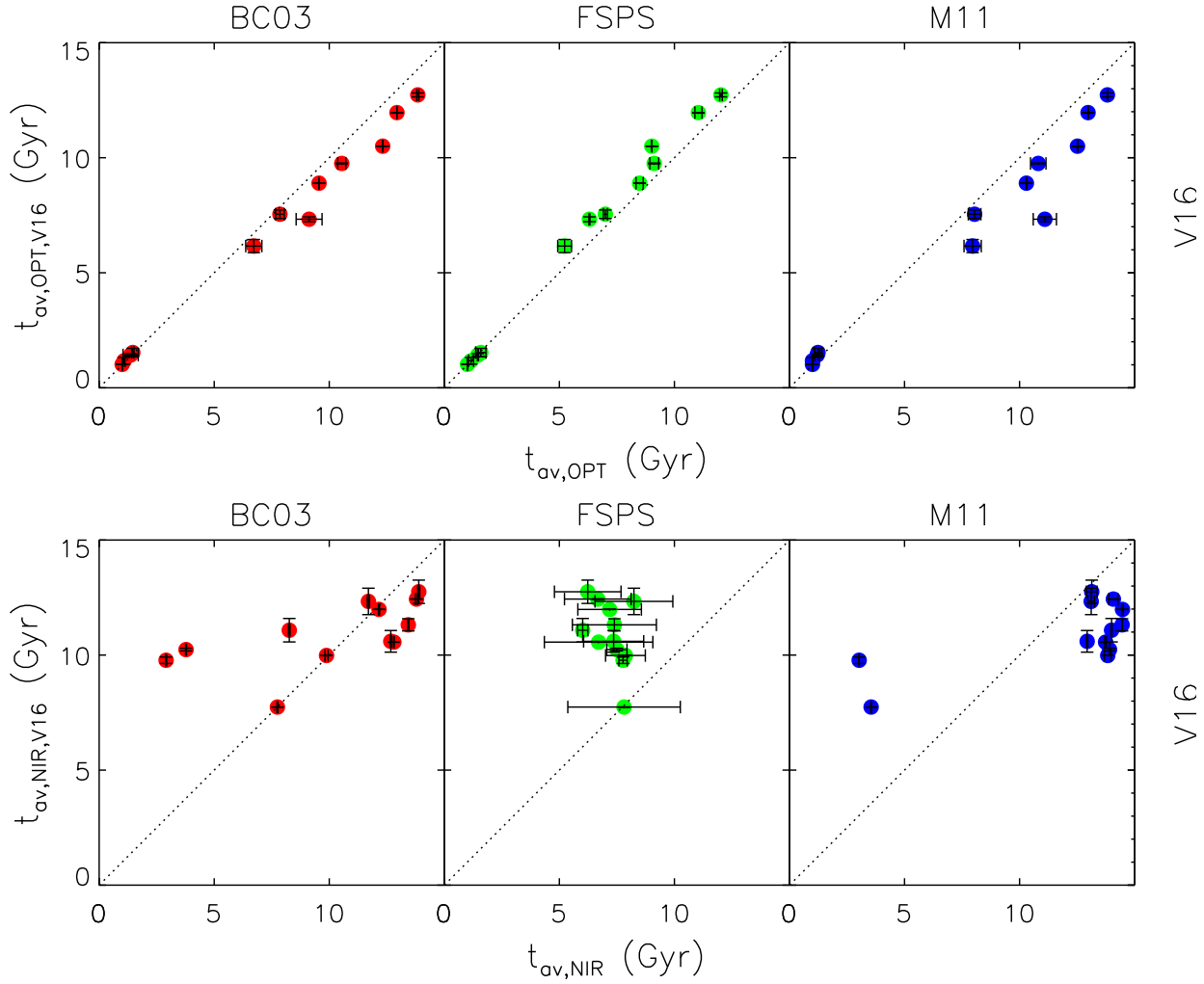


Figure 2.17: Mass-weighted mean ages derived from optical (top) and NIR (bottom) spectroscopy for the various models plotted against the ages derived by V16, using a Calzetti extinction curve.

models used here.

2.6.2 Why does the near-infrared give such inconsistent results?

In Figure 2.17 we plot the mass-weighted mean ages derived by BC03, FSPS and M11 against the mass-weighted mean ages derived by V16. These figures illustrate that all models give consistent results in the optical, and yet differ substantially in the near-infrared. Why is this the case? At optical wavelengths, the change due to age is strong, with $H\beta$ showing a linear decrease as a function of log age (the change in $H\beta$ over the relevant age range is a factor of ≈ 25 greater than the typical measurement error). As shown in Figures 2.2 and 2.16, the near-infrared shows no such strong age variation. Rather, the response to age is very subtle. CN1.1 is predicted by the M11 models to vary by a factor 50 greater than the typical measurement error over the entire range of interest. However, almost all of this variation is at ages $t < 3$ Gyr. Looking only at ages $t \geq 3$ Gyr, differences are within the errors ($\delta EW = 0.002$ mag vs measurement error = 0.003 mag). Other models display even milder variation. For CO, the predicted variation is $\approx 3\text{--}18\times$ the measurement error, depending on the model. The consequence of this generally subtle variation with age is that the derived properties become driven by systematic effects, with the largest effect being due to the choice of spectral library, as is seen in Figures 2.12 and 2.13. The choice of isochrones has much less effect, but is still not-negligible. Furthermore, the subtle age information is likely

encoded in the continuum shape, rather than the strengths of specific features, which are observed not to vary strongly across the sample. This means that the common technique of using a high order polynomial to remove the continuum when deriving SFHs gives unreliable results in this wavelength range. This is also shown in Figure 2.8 where the models lose constraining power when including a polynomial rather than an extinction law.

2.6.3 Can we constrain treatment of the TP-AGB?

The lack of age variation in the molecular absorption measured in the data implies that either the TP-AGB phase has less effect on the integrated light of galaxies than is indicated by the M11 models, or that TP-AGB stars are not present in large numbers in any galaxies in our sample. We are, of course, dealing with composite populations, rather than SSPs, meaning that the effect of the TP-AGB phase should be diluted in the data compared with the models. The question is whether this alone could account for the observed lack of variation in TP-AGB sensitive features despite the broad range in the age-sensitive $H\beta$ index.

We test this using the M11 models to construct simple two-component mock populations, where the age of the old population is 15 Gyr and the age of the young population varies from 0.3 – 3 Gyr. We vary the fraction of young stars from 0 – 100% and study the effect on the strength of CN1.1 and $H\beta$.

Figure 2.18 shows the strengths of $H\beta$ and CN1.1 measured in our data as filled circles, while the predictions of the various M11 composite populations are shown as coloured tracks. The mass fraction of the young population varies along the tracks. Different colours show composite populations with different ages for the young population. The M11 composite models can reproduce all the indices measured in the sample. The very strong $H\beta$ values but only moderate CN values seen in NGC 3032 and NGC 3156 can be reproduced by the M11 models assuming a 0.3 Gyr burst making up $\sim 20\%$ of the mass, on top of a dominant, old population. The $H\beta$ and CN values measured in the rest of the sample are also able to be reproduced in this scenario. In other words, the observed large range of measured $H\beta$ strength but only mild variation in CN1.1 strength may be explained by the galaxies being composite populations, and thus, we cannot strongly exclude strong TP-AGB contributions to younger populations in our sample without better constraining the shape of the underlying SFH.

Regardless, treatment of the TP-AGB phase does not have a substantial effect on our derived properties. The M11 models, which are unique in their treatment of the TP-AGB phase, give fits which are equally as good as other models (M11 χ^2_{red} values are approximately the same as BC03, which uses the same spectral library), and give SFHs which are equally as good as other models. The M11 and BC03 SFHs are almost identical for most galaxies, again indicating that the derived SFH is largely driven by spectral library rather than specifics of the inclusion of the TP-AGB phase for our particular sample. The situation may be different at high redshift, because the fraction of stars around the peak TP-AGB period is significantly higher. (see, for example, Capozzi et al. (2016), Kriek et al. (2010)). As the metallicity dependence of TP-AGB features is not well-known (Maraston, 2005), different results may also be obtained for non-solar metallicities.

2.6.4 Comparison with previous studies

This result is in contrast to some other studies, which have found TP-AGB treatment has a large effect on derived properties, with use of the Maraston models being either inconsistent with observations (Conroy & Gunn, 2010, Kriek et al., 2010, Zibetti et al., 2013) or more consistent than TP-AGB light models (Capozzi et al., 2016, MacArthur et al., 2010, Maraston et al., 2006, Rif-fer et al., 2015). Most studies to date have been based on photometry, and as such are not directly

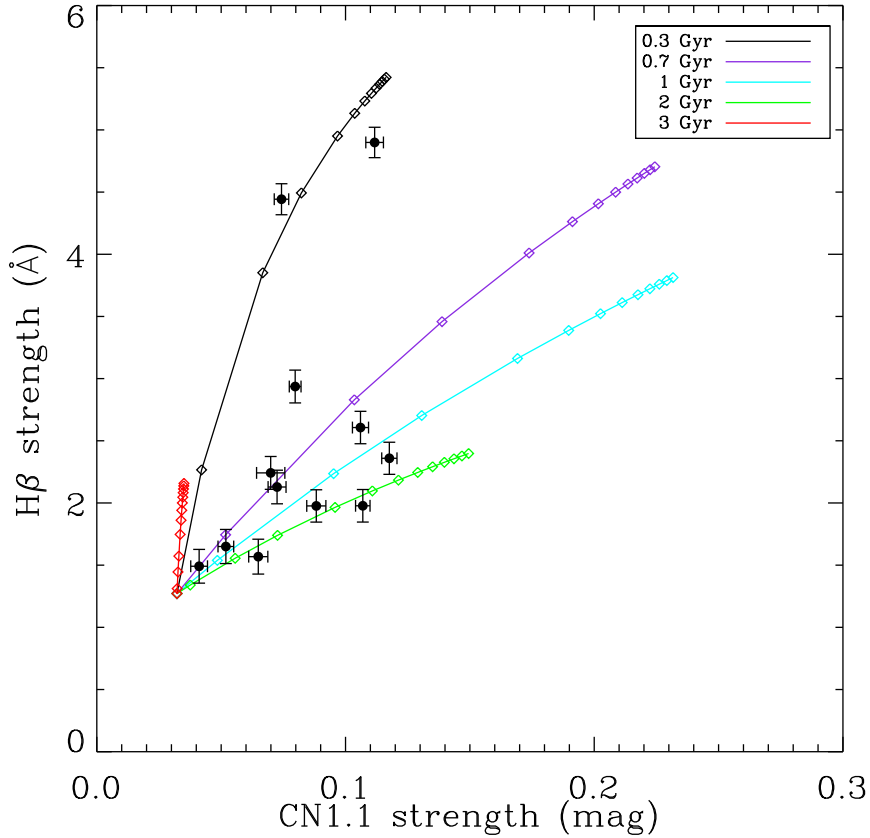


Figure 2.18: Toy model showing a comparison between the observed strengths of CN1.1 and $H\beta$ in our sample and the M11 composite models. The coloured lines show composite populations with a young population of various ages, as given in the legend, added onto a population of age 15 Gyr. The fraction of the young population increases along the tracks from left to right, and values are [0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1] .

comparable, but two recent studies have been based on near-infrared spectroscopy - those of Zibetti et al. (2013, hereafter Z13), and Riffel et al. (2015, hereafter R15).

R15 suggests TP-AGB stars contribute noticeably to a stacked near-infrared spectrum made up largely of Seyfert galaxies from the Palomar survey. This result was based on fitting near-infrared spectra of individual IRTF stars to the stacked galaxy spectrum. However, the authors also find that other evolved stars can reproduce most of the observed absorption features, and the intermediate age galaxies in their sample do not display stronger ‘TP-AGB features’ than the old galaxies. This is consistent with what we find in our sample.

Our results seem to contradict Z13, who used optical and near-infrared spectrophotometry of post-starburst galaxies to study the contribution of the TP-AGB phase in galaxies for which it should be at a maximum. The authors looked for the strong molecular features at 1.4 and 1.8 μm , predicted by the Maraston models to be present at intermediate ages. They did not observe these strong features, which is consistent with what we find (see Figure 2.2 and Figure 2.16). However, they also found that the intermediate age Maraston models, which reproduce the optical age-sensitive indices measured in their sample, produce near-infrared colours which are too red compared to the data. They thus conclude that the Maraston models have too strong a TP-AGB contribution, as they cannot reproduce the spectrophotometry of post-starburst galaxies under a range of simple or two-component composite population assumptions.

We have two galaxies in our sample which are very similar to the post-starburst galaxies targeted by Z13 (NGC 3032 and NGC 3156). Both have strong Balmer lines but little or no [OII] or $H\alpha$ emission and have light-weighted mean ages between 0.5-1.5 Gyr, as calculated from their

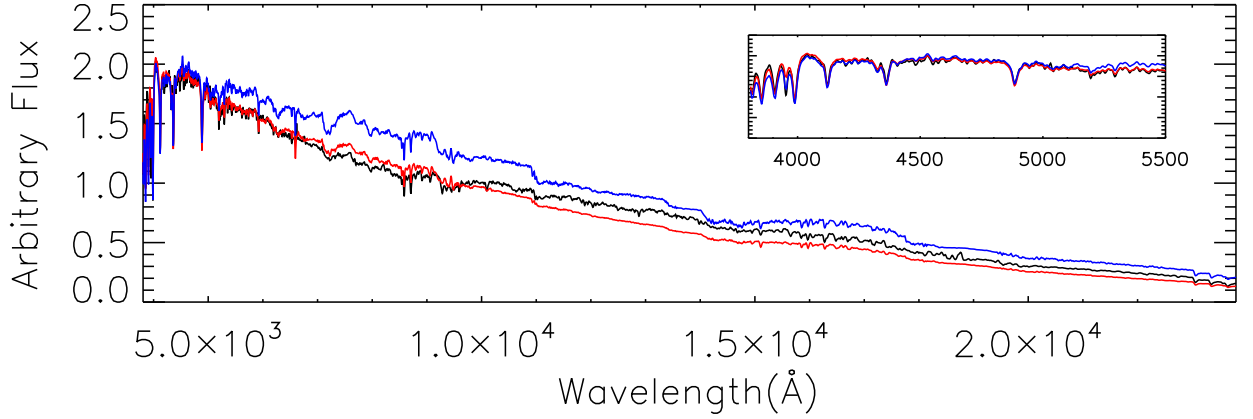


Figure 2.19: Results of the fit to post-starburst NGC3156’s SDSS+GNIRS spectrum when assuming a two-burst population composed of a 10 Gyr population plus a burst making up 20% of the mass. The ages of the young populations are 0.3 Gyr (red) and 0.5 Gyr (blue). Both provide excellent fits to the optical, as shown in the inset.

optical spectra. NGC3156 also has SDSS spectra available, which overlaps with the GNIRS wavelength range, meaning that we are able to create a spectrum spanning the entire optical-NIR range, thus avoiding issues of absolutely flux calibrating data from two different instruments. The SDSS spectrum covers the wavelength range 3808–9208 Å, while our GNIRS spectrum spans 8500–25,300 Å. We normalise the spectra in the overlapping wavelength range and stitch the GNIRS spectrum on to the end of the SDSS one. The SDSS spectrum was obtained in a 3” fibre, centred on the galaxy nucleus, while GNIRS uses a 7x0.3” slit, from which we extract a spectrum in an $R_e/8$ aperture, meaning that the apertures are not exactly matched, however both spectra are dominated by the bright galaxy nucleus. This situation is similar to the Z13 study, which also compares 3” SDSS spectra to slit spectra taken with ISAAC.

With our continuous optical-IR spectrum of a post-starburst galaxy, we tested whether we could reproduce the Z13 result. First we carried out fits using the composite populations tested by Z13. We constructed simple (two-burst) composite populations, consisting of an old, 10 Gyr component plus a burst. The age of the burst varied, starting from 50 Myr, and the burst fraction varied from 10% to 100%. We used the same ages as in Z13, (50 Myr, 0.3, 0.5, 1, 1.5, 2, 10 Gyr). Similarly to Z13, we fit the five optical indices from Gallazzi et al. (2005; $D4000_n$, $H\delta_A + H\gamma_A$, $H\beta$, $[MgFe]'$ and $[Mg_2Fe]$) and extrapolated the fits to the near-infrared. Using these simple composite population templates we were unable to fit the entire spectrum well with any combination tested by Z13. Figure 2.19 shows two of the resulting best fitting composite templates. We see the same behaviour as is shown in Figure 10 of Z13, which is that a good fit to the optical can overpredict (or underpredict) the near-infrared flux when fitting a simple two-burst population.

However, this problem is alleviated with our technique of fitting the entire spectral range and allowing for a more general SFH. When we carry out full spectral fitting using multiple templates spanning a large range of ages (Figure 2.20), we see that the M11 models are able to simultaneously match the strong Balmer absorption seen in the optical part of the spectrum, while not overpredicting the near-infrared flux. We fit the spectrum using a Calzetti extinction curve, which provides some low-order correction for slight flux calibration errors, but cannot correct for any near-infrared colour excess which may be present in the Maraston models - in fact it can only make the spectrum redder. Directly fitting the spectrum with templates spanning a range of ages rather than trying to match individual indices using simple stellar population models or simple two-component populations allows us to better constrain the stellar populations present, and in this case, we find again that the Maraston models are able to match the data with a plausible SFH. We therefore interpret the apparent differences between our findings and those of Z13 as deriving

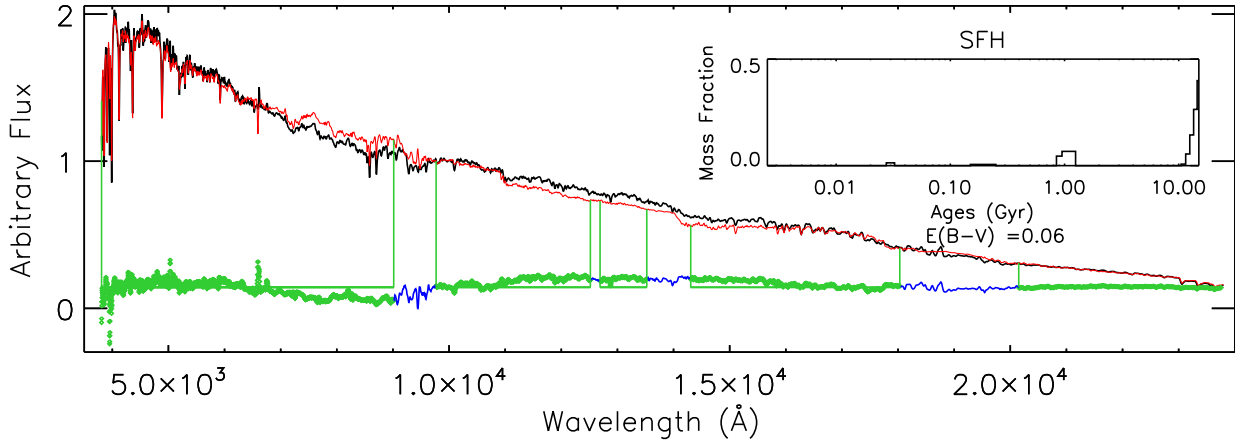


Figure 2.20: Results of the fit to post-starburst NGC3156’s SDSS+GNIRS spectrum. The pPXF fit is shown in red, and the derived star formation history from this fit is shown in the inset.

from our approach of fitting the full spectrum with a more comprehensive star formation history.

2.7 Conclusion

We presented in this chapter high quality near-infrared spectra of a sample of galaxies from the ATLAS^{3D} survey. These galaxies were chosen to span a large range of optically-derived ages, from post-starburst galaxies to galaxies with ages close to the age of the universe. This sample was chosen in order to study the effect of age on near-infrared wavelengths, which is currently poorly understood, as various stellar population models make conflicting predictions in this wavelength range.

We fit linear combinations of SSP templates to our data, to derive star formation histories using four different stellar population models, namely Bruzual & Charlot (2003), Conroy, Gunn & White (2009; 2010), Maraston & Strömbäck (2011) and Vazdekis et al. (2016). We find the V16 models typically give the best fits, owing to their high quality near-infrared stellar spectral library. We find that derived star formation histories typically differ between the various models, but also depend on the fitting method chosen, even within the same model set. Star formation histories derived using an extinction curve to model the continuum are typically more consistent with SFHs derived from optical spectroscopy. Use of a multiplicative polynomial to model the continuum causes models to lose age sensitivity, indicating that in the near-infrared regime, age sensitive information is encoded largely in the continuum shape rather than in the strengths of specific features. This is supported by our measurement of the strengths of specific indices, most of which do not vary strongly as a function of optical age.

We find that the largest differences in derived SFHs are caused by the choice of stellar spectral library. While the low-resolution near-infrared BaSeL library in FSPS results in poor agreement between optical and near-infrared derived SFHs, the use of the high resolution theoretical C3K library results in the most self-consistent SFHs across the two wavelength ranges. We therefore suggest that the inclusion of high quality near-infrared stellar spectral libraries (such as IRTF, XSL, C3K) into current near-infrared stellar population models should be a top priority for modellers.

Inclusion of the TP-AGB phase appears to have only a second-order effect in this sample, with “TP-AGB heavy” and “TP-AGB light” models resulting in comparable fits and SFHs. In particular, the M11 and BC03 models, when using the same spectral library, produce very similar SFHs from full spectral fitting, despite a very different TP-AGB treatment. We do not observe strong variation in TP-AGB sensitive indices as a function of age, which is predicted by the single-burst

M11 models; however optical and near-infrared indices $H\beta$ and CN can be reproduced by a two-burst population constructed with M11, including TP-AGB stars, so our galaxies cannot strongly constrain the contribution of the TP-AGB. Finally, we compare to previous works addressing the TP-AGB fraction in galaxies. We find general agreement with our findings where comparable data were used, and apparent disagreements with studies using lower quality data may be explained through the combination of higher-quality data and the use of a more general spectral fitting approach. This demonstrates the importance of applying this type of analysis to high quality data with a large, continuous wavelength coverage for a range of galaxy types to provide more stringent tests on stellar population models in the future.

