THE IMPACT OF SYSTEMATIC UNCERTAINTIES IN STELLAR PARAMETERS ON INTEGRATED SPECTRA OF STELLAR POPULATIONS

SUSAN M. PERCIVAL AND MAURIZIO SALARIS
Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Birkenhead, CH41 1LD, UK; smp@astro.livjm.ac.uk, ms@astro.livjm.ac.uk
Received 2009 June 5; accepted 2009 August 7; published 2009 September 4

ABSTRACT

In this paper, we investigate a hitherto unexplored source of potentially significant error in stellar population synthesis (SPS) models, caused by systematic uncertainties associated with the three fundamental stellar atmospheric parameters: effective temperature \( T_{\text{eff}} \), surface gravity \( g \), and iron abundance \([\text{Fe}/H]\). All SPS models rely on calibrations of \( T_{\text{eff}} \), \( g \), and \([\text{Fe}/H]\) scales, which are implicit in stellar models, isochrones, and synthetic spectra, and are explicitly adopted for empirical spectral libraries. We assess the effect of a mismatch in scales between isochrones and spectral libraries (the two key components of SPS models) and quantify the effects on 23 commonly used diagnostic line indices. We find that typical systematic offsets of 100 K in \( T_{\text{eff}} \), 0.15 dex in \([\text{Fe}/H]\) and/or 0.25 dex in \( g \) significantly alter inferred absolute ages of simple stellar populations (SSPs) and that in some circumstances, relative ages also change. Offsets in \( T_{\text{eff}} \), \( g \), and \([\text{Fe}/H]\) scales for a scaled-solar SSP produce deviations from the model which can mimic the effects of altering abundance ratios to non-scaled-solar chemical compositions, and could also be spuriously interpreted as evidence for a more complex population, especially when multiple-index or full spectral energy distribution fitting methods are used. We stress that the behavior we find can potentially affect any SPS models, whether using full integrated spectra or fitting functions to determine line strengths. We present measured offsets in 23 diagnostic line indices and urge caution in the overinterpretation of line-index data for stellar populations.

Key words: galaxies: abundances – galaxies: evolution – galaxies: stellar content – stars: evolution – stars: fundamental parameters

1. INTRODUCTION

In recent years, stellar population synthesis (SPS) models have become a fundamental tool in the study of both Galactic (resolved) and extragalactic (unresolved) stellar populations. Predictions of both photometric and spectroscopic properties, in terms of broadband colors and line-index strengths, from various SPS models have yielded methods for tackling the well known age–metallicity degeneracy in simple (single-age, single-metallicity) stellar populations (SSPs; see, e.g., Percival et al. 2009; Kotulla et al. 2009; Coelho et al. 2007; Vázquez & Leitherer 2005; Maraston 2005; Bruzual & Charlot 2003 for some recent examples).

As observational data, and in particular spectroscopic data, are becoming more precise, the models are becoming more sophisticated and methods are being developed to determine more detailed information such as the level of \( \alpha \)-enhancement in SSPs, chemical evolution, and the general properties of the star formation histories of composite populations. However, results from different sets of models often disagree and there are currently many unresolved issues concerning their interpretation and implementation. Carter et al. (2009) investigated the ability of several “off the shelf” SPS models to reproduce optical and near infrared colors (from \( \mu \) through to \( K \)) of a small sample of well studied nearby elliptical galaxies. Fitting to SSPs, making the implicit assumption that giant ellipticals can be well represented by a simple population with no significant star formation history, they went on to compare the ages and metallicities predicted from the various models, using a simple reduced \( \chi^2 \) test. Carter et al. found that, for a similar \( \chi^2 \), best fits from different SPS models can give ages which differ by 5 Gyr or more whilst predicted \([\text{Fe}/H]\) can easily differ by 0.4 dex (a problem here being the \([\text{Fe}/H]\) sampling, which for most sets of SPS models is limited to \( \sim 0.3 \) dex). Predictions from spectroscopic data, in the form of diagnostic line indices, can have similar (if not worse) discrepancies, as demonstrated by the results of the IAU Symposium 241, Stellar Population Challenge.\(^1\) For the well studied Galactic Globular Cluster 47 Tuc, which was set as a test object with an observed spectrum provided from Schiavon et al. (2005), predicted ages from various models range from 5 Gyr to \( > 18 \) Gyr whilst predicted \([\text{Fe}/H]\) ranges from \( -0.1 \) to \( -0.8 \) dex.

Several recent studies have focused on discrepancies that can arise because of properties of the observed population, which may not be fully accounted for in the models. These include extended blue horizontal branches and blue stragglers which can both make a population look spuriously young (Schiavon et al. 2004; Cenarro et al. 2008, respectively, using diagnostic line indices), and thermally pulsing asymptotic giant branch (AGB) stars which strongly affect the near-infrared flux (and hence colors) at intermediate ages, significantly altering the inferred galaxy masses (Bruzual 2007). However, it is clear that there can be significant offsets between ages and metallicities determined from (or predicted by) different sets of SPS models which must, at least in part, be due to systematic differences in the model ingredients. This problem is clearly demonstrated by the “zero-point” offset in some models utilizing diagnostic line indices which means that inferred ages of Galactic globular clusters can be older than 14 Gyr. Although these problems have been noted by several groups (see, e.g., Vazdekis et al. 2001; Schiavon et al. 2002; Cenarro et al. 2008), little work has been done to determine the specific sources of these offsets.

\(^1\) http://www.astro.rug.nl/~sctrager/challenge/
or quantify their effects in terms of the true systematic errors which should be assigned to derived ages and metallicities.

Physical parameters of stars within a population, such as effective temperature, \( T_{\text{eff}} \), surface gravity, \( \log g \), and iron abundance, \([\text{Fe}/\text{H}]\), are key parameters in the construction of the SPS models (see Section 2) and one, as yet, unexplored aspect of SPS is the adopted temperature and metallicity scales used in the underlying stellar models and spectra. When generating stellar evolution models (which give rise to isochrones, a fundamental building block in SPS modeling), the temperature and metallicity scales effectively come out of the models themselves. This is also the case for synthetically generated stellar spectra, for which stellar atmosphere models and adopted line lists give rise to some intrinsic temperature and metallicity scale. For empirical spectra, a specific temperature and metallicity scale must be adopted in order to determine physical parameters of the stars. There can potentially be mismatches between the temperature and metallicity scales of the underlying stellar models and those of the stellar spectra (whether synthetic or empirical), but any impact this may have on the resulting SPS models has not yet been investigated systematically and quantified. The aim of this paper is to address this issue.

The structure of the paper is as follows: in Section 2, we review the ingredients required to construct SPS models whilst Section 3 describes some preliminary tests which motivated this work. Sections 4 and 5 describe the specific tests applied to the models and the results of these tests whilst Section 6 briefly assesses the significance of the results in the context of typical observational errors for extragalactic systems. In Section 7, we summarize our results and discuss their implications for the age and metallicity predictions of SPS models and their associated errors.

2. MODEL INGREDIENTS AND CONSTRUCTION

Our work focuses on the use of diagnostic line indices to determine ages and metallicities for SSPs—in particular we will utilize the commonly adopted technique of comparing plots of pairs of line indices with model grids in our investigation, although our results also have important implications for other methods, e.g., multiple-index fitting or whole-spectrum fitting.

Throughout this work, we will be using the high-resolution synthetic spectra from the BaSTI SPS models,2 described in Percival et al. (2009). Since we have control over all the ingredients of these models, we can separately investigate the effects of each input element on the resulting spectral energy distributions (SEDs) and line indices.

The two principal ingredients of any SPS model are the underlying stellar models, in the form of isochrones, which are “populated” according to some initial mass function (IMF) to create each stellar population, and the spectral library which is used to assign a spectrum to each data point in that population. The summing of these individual spectra, with appropriate weighting, results in the final integrated spectrum (SED) on which diagnostic line strengths, such as H\( \beta \) and various Fe and other metal lines, can be measured directly as equivalent widths (EWs). This was the method employed to produce the BaSTI database of SEDs, and is used here to test the models (see Percival et al. 2009, for more details). Alternatively, using methods pioneered by Worthey (1994), some groups use fitting functions to determine line strengths (see, e.g., Thomas et al. 2003; Schiavon 2007; Trager et al. 2008). Our study is also relevant to these methods since they also rely on calibrations from spectral libraries and are constructed in a similar way, i.e., by assigning the relevant quantity (EW or similar) to each point in a population and then summing along the isochrone.

The choice of isochrones used by SPS modelers is undoubtedly a contributory factor to the differing predictions from various groups. Underlying stellar models (and hence isochrones) from different sources can vary because of the specifics of assumed input physics and differences in stellar evolution codes. An in-depth discussion of these differences is beyond the scope of this paper, however, a detailed comparison of some of the most commonly used isochrone sets can be found in Pietrinferni et al. (2004). One practical difference between isochrones from different groups is the number of evolutionary points (EPs) that define each isochrone—effectively the sampling along the isochrone. BaSTI isochrones consist of 2250 discrete EPs covering all evolutionary stages whilst, for example, the commonly used Padova isochrones are typically defined by a few hundred EPs (e.g., those of Girardi et al. 2000).

Evolutionary points along isochrones are defined in terms of effective temperature \( T_{\text{eff}} \), luminosity \( L \), and mass \( M \), from which surface gravity, \( \log g \), is also derived. \( T_{\text{eff}} \) and \( \log g \), along with \([\text{Fe}/\text{H}]\) (and/or total \( Z \) and degree of \( \alpha \)-enhancement) are used to match, or create by interpolation, an appropriate spectrum (or fitting function) to each EP, since spectra in libraries are usually parameterized by these three quantities. \( T_{\text{eff}} \) and \([\text{Fe}/\text{H}]\) (and by implication \( \log g \)) must be defined on some scale, and there can potentially be mismatches between the scales arising from the underlying stellar models/isochrones and those adopted by the spectral library used. This is especially relevant when empirical spectral libraries are employed as this requires an evaluation of the physical parameters of the observed stars, however, it is also an issue when using synthetic spectral libraries if the stellar models and the atmosphere models are on different scales. Any mismatches in these scales will equally affect SPS models using either full SEDs or fitting functions.

3. PRELIMINARY TESTS

As a preliminary evaluation of the typical systematic offsets to be expected in effective temperature scales, we compared the listed \( T_{\text{eff}} \) values for EPs along BaSTI isochrones with those that would be “predicted” from the empirical \( T_{\text{eff}} \)-color calibrations of Alonso et al. (1996b; for dwarfs and subdwarfs) and Alonso et al. (1999b; for giants). The Alonso et al. (1996b, 1999b) calibrations give equations to calculate \( T_{\text{eff}} \) from various colors, in combination with \([\text{Fe}/\text{H}]\)—for this test the \((V-K)\) color calibration was used, since this is the one upon which the atmospheric parameters for the MILES empirical spectral library (Sánchez-Blázquez et al. 2006b, used in a later test) are based. Taking the \((V-K)\) color for each EP along an isochrone (see Pietrinferni et al. 2004, for details of isochrone colors), along with \([\text{Fe}/\text{H}]\), \( T_{\text{eff}} \)'s were calculated from the Alonso et al. calibrations and then compared with the listed isochrone \( T_{\text{eff}} \)’s. Systematic differences were found, which have some metallicity dependence, and are typically around 70 K at \([\text{Fe}/\text{H}]\) \( \sim -2 \), in the sense that the listed isochrone \( T_{\text{eff}} \)'s are hotter. These differences are similar to the typical quoted systematic uncertainties on \( T_{\text{eff}} \), which are of the order 50–100 K (see, e.g., Alonso et al. 1996b).

Ramírez & Meléndez (2005) determined temperatures of 135 dwarf and 36 giant FGK stars using a method very similar to the Alonso et al. studies and found that, in general, the agreement in temperature scales was very good, and the mean

---

2 http://albione.oa-teramo.inaf.it/
uncertainty in derived temperatures is 75 K for dwarfs and 60 K for giants. However, Casagrande et al. (2006) found that for 18 stars in common between their study and that of Ramírez & Meléndez (2005), there was an average difference of 105 ± 72 K, with the Casagrande et al. (2006) scale being hotter. Casagrande et al. (2006) go on to say “though not negligible, such differences are within the error bars of current temperature determinations.” More recently, González Hernández & Bonifacio (2009) rederived $T_{\text{eff}}$ for all stars in the Alonso et al. samples (Alonso et al. 1996a, 1999a) using the infrared flux method, and found differences in temperature scales of $\sim$64 K and $\sim$54 K for dwarfs and giants, respectively, in the sense that the González Hernández & Bonifacio (2009) scale is hotter. However, comparing with the Ramírez & Meléndez (2005) calibration they found that for low-metallicity dwarfs, the González Hernández & Bonifacio (2009) scale is cooler by $\sim$87 K.

[Fe/H] and log $g$ calibrations are, of course, intrinsically linked to $T_{\text{eff}}$ when determining atmospheric parameters for empirical spectra and so systematic uncertainties in $T_{\text{eff}}$ imply similar uncertainties on these parameters also. Quoted uncertainties are typically of the order 0.25 dex in log $g$ and 0.15 dex in [Fe/H] (see, e.g., Soubiran et al. 1998).

In the early stages of creating integrated spectra for the BaSTI SPS database, we performed a preliminary comparison between synthetic spectra from the Munari et al. (2005) spectral library (as used for the high-resolution BaSTI SEDs) and the MILES empirical spectral library (Sánchez-Blázquez et al. 2006b; as used by Cenarro et al. 2008). For this exercise, we took a subset of the MILES spectra with zero reddening ($E(B - V) = 0$) and with the designation SKC, indicating that the atmospheric parameters are from Soubiran et al. (1998; these stars being the “gold standard” for the MILES library). To avoid the ambiguity of whether or not to assume scaled-solar or $\alpha$-enhanced synthetic spectra, we used only stars with either [Fe/H] $\geq$ −0.3 (assumed to be scaled solar) or [Fe/H] $\leq$ −1.0 (assumed to be $\alpha$-enhanced—note that in the Munari library, the level of $\alpha$-enhancement is fixed at $\alpha$/Fe = 0.4). By interpolating in [Fe/H], $T_{\text{eff}}$, and log $g$ amongst spectra in the Munari et al. (2005) library, a matching synthetic spectrum was created for each star in the MILES subsample. Line strengths were then measured directly on the two sets of spectra using the LECTOR each star in the MILES subsample. Line strengths were then measured directly on the two sets of spectra using the LECTOR.

In the following tests, we examine the effects of altering each of the atmospheric parameters in the BaSTI SPS models separately, within the typical uncertainties, whilst holding all other elements constant—these uncertainties are taken to be ±100 K in $T_{\text{eff}}$, ±0.25 dex in log $g$, and ±0.15 dex in [Fe/H]. We stress that these tests are exploring the effects of a mismatch in scales between the underlying stellar models (and isochrones) and the adopted spectral library in the SPS models, which potentially give rise to systematic errors in derived ages and metallicities for SSPs. When presenting our results we are making the implicit assumption that the differential behavior we find would be quantitatively the same, or very similar, for other SPS models, whether based on full integrated spectra or on fitting functions methods to determine line strengths.

4. THE TESTS

All the tests described below were performed on two test-case SSPs with ages $t = 14$ Gyr, to represent a typical old elliptical galaxy, and $t = 4$ Gyr, which is representative of the intermediate ages found in the sample of ellipticals studied by Trager et al. (2000). In both cases, the scaled-solar abundance, solar metallicity models were used ($Z = 0.0198$ and [Fe/H] = +0.06) with no convective overshooting, and with the Reimers mass-loss parameter, $\eta$, set at 0.2. High-resolution (1 Å pixel$^{-1}$) integrated spectra for the reference models were created, as described in Percival et al. (2009), where the interested reader can also find more details on the underlying stellar models. We note here that we have also applied all the following tests to low-metallicity $\alpha$-enhanced SSPs, typical of Galactic globular clusters, and obtain quantitatively very similar results.

When constructing the integrated spectrum for each test, the underlying isochrone is left unchanged, but individual spectra are assigned to each EP along the isochrone with the appropriate offset in each atmospheric parameter, as listed below. Each atmospheric parameter was tested for two instances—one in which the whole isochrone is affected (i.e., stars at all evolutionary phases) and one in which only the giant stars are affected. For these purposes, we took all EPs with, simultaneously, $T_{\text{eff}} < 5000$ K and $g < 3$ to represent giants. The justification for treating giants only as a separate case is that, for empirical $T_{\text{eff}}$ calibrations, dwarfs and giants are often treated separately and consequently have separate $T_{\text{eff}}$-color relationships. It is also generally the case that cool stars are harder to model theoretically due to the effects of phenomena such as molecular opacities, convection, and mass loss which become increasingly important for cooler stars, hence there is more likely to be a discrepancy in temperature scales for cool red giant branch (RGB) and AGB stars. Tests were performed on the two reference SSPs as follows.

Test 1. Spectra assigned with $T_{\text{eff}}$ increased by 100 K compared to isochrone $T_{\text{eff}}$'s (whole isochrone).

Test 2. Spectra assigned with $T_{\text{eff}}$ increased by 100 K compared to isochrone $T_{\text{eff}}$'s, for giants only (i.e., $T_{\text{eff}} < 5000$ K and $g < 3$ on isochrone).

Note that we could not perform fully equivalent tests in which $T_{\text{eff}}$ was decreased by 100 K. This is because there is a low-temperature cutoff in the spectral library used for the BaSTI SPS models (Munari et al. 2005; see Percival et al. 2009 for more details) which means that the coolest stars could not be consistently included in the test as an offset of $\sim$100 K would take them below this threshold. However, we did perform this

---

3 See http://www.iac.es/galeria/vazdekis/.
test using a truncated version of the 4 Gyr SSP, i.e., with a cut at 100 K above the threshold $T_{\text{eff}}$ so that this test could be done by assigning spectra with both increased and decreased $T_{\text{eff}}$ values.

Tests 3a, b. log g increased/decreased by 0.25 dex (whole isochrone).

Tests 4a, b. log g increased/decreased by 0.25 dex (giants only).

Tests 5a, b. [Fe/H] increased/decreased by 0.15 dex (whole isochrone).

Tests 6a, b. [Fe/H] increased/decreased by 0.15 dex (giants only).

Two further tests were also performed in which two “worst-case” scenarios were created. For the first of these, two parameters were altered simultaneously over the whole isochrone, namely, $T_{\text{eff}}$ increased by 100 K and log g decreased by 0.25 dex. For the second of these tests, all three parameters were altered—$T_{\text{eff}}$ was increased, log g decreased, and [Fe/H] increased by 0.15 dex. The purpose of these tests was to check whether the effects on line indices are additive when two or three atmospheric parameters are offset simultaneously.

Line strengths for 23 diagnostic indices were then measured for the reference cases and test cases in the same way as before, i.e., by measuring EWS directly on the integrated spectra themselves. Indices used are the 21 indices defined by the bandpasses in Trager et al. (1998; noting that, by convention, CN1, CN2, Mg1, Mg2, TiO1, and TiO2 are quoted in magnitudes rather than as EWSs in Angstroms) plus H$\beta$ and H$\gamma$ as defined in Worthey & Ottaviani (1997). We stress that all numbers quoted are line strengths as measured directly on the integrated spectra and are not transformed onto the Lick system.
C24668, but for many indices offsets go in the opposite sense to those for the whole-isochrone case, e.g., CN1, CN2, and most of the Fe lines.

Results of the decreased $T_{\text{eff}}$ test (using the truncated 4 Gyr SSP) are not listed, as the measured values of offsets would not be applicable to the full SSP. However, we note that, for all 23 measured indices, offsets for the decreased $T_{\text{eff}}$ truncated SSP were found to be symmetrically opposite with respect to the increased $T_{\text{eff}}$ truncated SSP case, implying that this would also be the case for the full SSP.

To demonstrate the effect of these offsets on estimated ages and metallicities of SSPs, we have constructed representative index–index grids from the BaSTI models for a range of ages and metallicities and overplotted the test-case SSPs—Figures 1, 2, and 3 show grids of $H\beta$ versus Mg$b$, Fe5406, and C24668, respectively. In each figure, the left-hand panel shows the results of the whole-isochrone tests and the right-hand panel shows the results of the giants-only tests. The two test reference SSPs (solar metallicity, $3\times10^{-3}$, and $1.5\times10^{-2}$, and $14$ Gyr) are marked as solid squares. Results of various tests are labeled on the diagrams (see the text for details).

Assessing the effect of temperature changes on derived metallicities is more complicated, since it depends on which index is used and whether metallicity is defined in terms of $[\text{Fe/H}]$ or total metallicity, $Z$. The three grids displayed in Figures 1–3 demonstrate the problem. For a 100 K increase in $T_{\text{eff}}$, the inferred $[\text{Fe/H}]$ from the $H\beta$–Mg$b$ plane is $\sim$0.1 dex lower than the reference SSP, from the $H\beta$–Fe5406 plane it is virtually unchanged, and from the $H\beta$–C24668 plane $[\text{Fe/H}]$ apparently increases by $\sim$0.05 dex. Offsets are similar for both the “whole-isochrone” and “giants-only” cases. The $H\beta$–C24668 grid also demonstrates another important point—for the 14 Gyr SSP, the offset for the increased 100 K case decreases the absolute value of C24668 slightly, however, the inferred $[\text{Fe/H}]$ increases as this grid is not orthogonal (as is also the case for all other commonly used age/metallicity diagnostic grids). Similarly constructed grids (not displayed here) show that, with increasing $T_{\text{eff}}$, Ca4227 behaves in the same way as Mg$b$, [MgFe] behaves as Fe5406 (but traces total $Z$ rather than $[\text{Fe/H}]$—see Percival et al. 2009), and CN1 and CN2 behave similarly to C24668. The issue of the contradictory behavior of various diagnostic metal lines and the impact on inferred chemical composition will be discussed further in Section 7.

5.2. Effects of Altering log $g$

It was found that, for the log $g$ tests, offsets in all indices were found to be symmetrical for the increased/decreased cases with respect to the reference case, hence the results are listed in Tables 1 and 2 as single values for each pairing of tests 3a/3b and 4a/4b, with the appropriate $\pm$ or $\mp$ sign. For both test-case SSPs, it can be seen that $H\beta$, $H\delta_F$, and $H\gamma_F$ increase significantly when log $g$ is decreased by 0.25 dex (and decrease when log $g$ is increased), affecting derived ages in a quantitatively similar way to a 100 K increase in $T_{\text{eff}}$. Mg$b$, Ca4227, Fe5406, [MgFe], and C24668 also behave in a quantitatively very similar way with decreasing log $g$ to the increased $T_{\text{eff}}$ case. Figures 1–3 show that, with increasing log $g$, [Fe/H] inferred from the $H\beta$–Mg$b$ plane increases by $\sim$0.1 dex, whilst that from the $H\beta$–Fe5406 plane is virtually unchanged and from the $H\beta$–C24668 plane [Fe/H] apparently decreases by $\sim$0.1 dex.

5.3. Effects of Altering $[\text{Fe/H}]$

As for the log $g$ and $T_{\text{eff}}$ tests, the $[\text{Fe/H}]$ tests produced offsets in all indices that were found to be symmetrical for the increased/decreased cases with respect to the reference
Results are listed in Tables 1 and 2 as single values for each pairing of tests 5a/5b and 6a/6b, with the appropriate ± or ‡ sign. As would be expected, all metal lines increase in strength when [Fe/H] is increased and decrease when [Fe/H] is decreased. For the whole-isochrone tests, the inferred change in [Fe/H] from the diagnostic diagrams is consistent with the modeled change in [Fe/H], i.e., ∼ 0.15 dex, whilst for the giants-only test the inferred change in [Fe/H] is generally about half this level.

An important point to notice is that Hβ increases significantly when [Fe/H] is increased over the whole isochrone (and decreases when [Fe/H] is decreased) inducing ∼ 12% change in derived ages for a 0.15 dex change in [Fe/H], however, for the giants-only case Hβ is virtually unchanged. Rather puzzlingly, Hβ and Hγ behave in the opposite sense to Hβ in that they both decrease significantly with increased [Fe/H], and vice versa. Hδ and Hγ also display quantitatively similar offsets for the whole-isochrone and giants-only cases, unlike the situation for Hβ. The implications of this contradictory behavior will be discussed further in Section 7.

5.4. “Worst-case” Scenarios

The results from the “worst-case tests,” in which two or three parameters were altered simultaneously, are not listed—however, in both cases the offsets in all indices were found to be additive. For the two-parameter test (altering $T_{\text{eff}}$ and log g simultaneously) resultant offsets are simply the sum of the offsets found for tests 1 and 3b, whilst for the three-parameter test ($T_{\text{eff}}$, log g, and [Fe/H]) offsets are the sum of those from tests 1, 3b, and 5a. Results of the two “worst-case” tests are also illustrated in Figures 1–3.

6. SIGNIFICANCE OF RESULTS: COMPARISON WITH TYPICAL OBSERVATIONAL ERRORS

In order to assess the significance of our results for observed stellar populations, we briefly compare the line-index offsets detailed above with typical observational errors for extragalactic systems. Table 3 shows the mean observational errors on various line-index strengths from three sources. Sánchez-Blázquez et al. (2006a) present a study of 98 early-type (E and S0) galaxies in the local field, small groups, and some Virgo and Coma cluster members, whilst Trager et al. (2008) study 12 elliptical and 50 galaxies in the Coma cluster. For higher redshift systems, we used data from Sánchez-Blázquez et al. (2009), which comprises a catalog of 215 red sequence galaxies in clusters and groups, with redshifts between $z \sim 0.45$ and $z \sim 0.75$. For the Sánchez-Blázquez et al. (2009) sample, observational errors on all quoted indices are slightly larger than the largest offsets found in our work, so that the systematic effects that we find would be a significant, but not the dominant, source of error for these systems. However, for the more local systems observational errors for all the key diagnostic indices are significantly less than the offsets caused by the mismatches in stellar parameters detailed above. Most importantly, the observational errors on Hβ, Hγ, and Hδ are smaller by a factor of ∼ 2 than the offset induced by a 100 K mismatch on temperature scales, which would lead to a systematic shift in the inferred ages for these systems. Similarly, observational errors on Ca4227, Mg b, and most of the Fe lines are significantly smaller than the offsets in indices caused by a mismatch in log g or [Fe/H] scales. It should be borne in mind that any offsets resulting from stellar parameter mismatches induce systematic rather than random errors.

We remind the reader that all our results discussed so far pertain to intermediate age and old stellar populations. At the suggestion of the referee, we also performed several tests on a 500 Myr, solar metallicity population. At this age (and younger), stellar populations do not have an RGB as such, and so we performed the basic tests of increasing $T_{\text{eff}}$ by 100 K, decreasing log g by 0.25 dex, and altering [Fe/H] by 0.15 dex, for the whole-isochrone case only. The offsets in indices found for this younger age population are generally a factor ∼ 2 smaller than those found for the 4 Gyr and 14 Gyr SSPs and so in most cases are less significant than the observational errors quoted in Table 3. Also, the strength of the Balmer lines increases much more rapidly with decreasing age for ages below ∼ 1 Gyr and so even a substantial change in Hβ of 0.2 dex (larger than any of the offsets found here) only changes the inferred age by ∼ 10–20 Myr.

6.1. Effect on Broadband Colors

The aim of this paper is primarily to investigate the effect of mismatches in stellar parameters on diagnostic line indices, however, we have also made a brief study of the effect on broadband colors for the main tests outlined in Section 4. In general, a mismatch in the log g scale has a negligible effect on colors whilst the effects of a 0.15 dex mismatch in [Fe/H] are small and are typically within the likely photometric errors, i.e., ≲ 0.02 mag in all the optical and near-infrared bands. As might be expected, the effect of increasing $T_{\text{eff}}$ by 100 K (test 1 above) is to make all the broadband colors bluer. For the intermediate age and old populations studied here, the effect is small in the optical bands, ∼ 0.01 mag in $(B - V)$ and $(B - V)$, and 0.02 mag in $(R - I)$, implying a minimal change in the inferred age, however, the effect becomes greater at longer wavelengths, making the $(J - K)$ color bluer by ∼ 0.04 mag. Hence, broader baseline colors such as $(V - K)$ and $(B - K)$ are affected more strongly by an increase in $T_{\text{eff}}$, the colors becoming bluer by at least 0.1 mag. As a result, a 100 K increase in $T_{\text{eff}}$ can significantly decrease the inferred metallicity, rather than making a population look younger, if a color–color diagram such as $(V - I)/(V - K)$ or $(B - K)/(J - K)$ is used (see Figure 1 of, respectively, Salaris & Cassisi 2007 and James et al. 2006).
7. SUMMARY AND DISCUSSION

We have investigated a potentially significant source of systematic error which can affect the ages and metallicities of stellar populations derived from SPS modeling, caused by systematic uncertainties in the three principal stellar atmospheric parameters \( T_{\text{eff}}, \log g, \) and \([\text{Fe}/\text{H}]\). In practice, we have tested the effects of a mismatch in temperature, metallicity, and gravity scales between the underlying isochrones and the spectra (or fitting functions) used to construct SPS models. We have done this by constructing integrated spectra for two solar metallicity SSP models aged 4 Gyr and 14 Gyr, and applying offsets of 100 K in \( T_{\text{eff}}, \) 0.25 dex in \( \log g, \) and 0.15 dex in \([\text{Fe}/\text{H}]\), which we take as typical zero-point uncertainties in these key atmospheric parameters. We have then quantified the effect on various diagnostic line indices by measuring EWs on the resultant spectra and calculating offsets between the unaltered reference SSPs and the altered versions.

We note here that the magnitude of the offsets we find as a result of altering the stellar parameters for SSPs, as described, is quantitatively very similar to the differences in line-index strengths found between the various spectral libraries, detailed in Section 3. Hence, our results are consistent with the differences between the spectral libraries being largely caused by zero-point differences in their \( T_{\text{eff}}, \log g, \) or \([\text{Fe}/\text{H}]\) scales. We take this as an indication that our results give a good estimate of the likely systematic errors in line-index strengths inherent in any population synthesis model, due to possible zero-point mismatches in stellar atmospheric parameter scales.

Using simple index–index diagrams to make a preliminary assessment of the impact on stellar population parameters inferred from the models, we find that absolute ages derived from the \( \text{H}\beta \) index can easily be affected at the 20% level for both old and intermediate age populations. Relative ages can also be affected, albeit at a lower level. Inferred ages are more complicated to interpret if the \( \text{H}\beta \), \( \text{H}\gamma \), \( \text{H}\delta \) and \( \text{H}\gamma \) indices are also considered because of their behavior in response to a systematic shift in \([\text{Fe}/\text{H}]\), which goes in the opposite sense to that of \( \text{H}\beta \). The inferred systematic errors in \([\text{Fe}/\text{H}]\) and/or \( Z \) are also hard to quantify because of the opposite behavior of certain key metal indicators in response to systematic offsets in \( T_{\text{eff}} \) and \( \log g \).

This behavior has implications for methods which fit simultaneously to several indices (or perform a full SED fit) to derive ages and metallicities of stellar populations, since a failure to fit several indices simultaneously could, spuriously, be interpreted as an indication of non-solar abundance ratios. Also, mismatches between observational data and model SSPs are often taken as evidence for the presence of components which are not fully accounted for in the models, such as missing (or extreme) stellar evolutionary stages or a composite population.

Percival et al. (2009) demonstrated that, for SSPs, the \( \text{Fe}5406 \) index traces Fe only, whilst \([\text{MgFe}]\) traces total metallicity, \( Z \) (as first noted by Thomas et al. 2003). These two indices, in combination, therefore provide an estimate of the level of \( \alpha \)-enhancement, or simply whether a population has non-solar abundance ratios. Figure 4 shows the \( \text{Fe}5406-[\text{MgFe}] \) plane with lines from the BaSTI 14 Gyr (constant age) scaled-solar and \( \alpha \)-enhanced models, joined at approximately equal \([\text{Fe}/\text{H}]\) (note that at constant \( Z, [\text{Fe}/\text{H}] \) is lower for \( \alpha \)-enhanced models than the corresponding scaled-solar ones). Overplotted are the results from the whole-isochrone tests for increased \( T_{\text{eff}} \) (100 K), increased \( \log g \) (0.25 dex), and increased/decreased \([\text{Fe}/\text{H}]\) (0.15 dex). In this diagram, one expects that any deviation in abundance ratios will move points horizontally, i.e., any degree of \( \alpha \)-enhancement moves points from the scaled-solar line on the left, toward the \( \alpha \)-enhanced line on the right (see Figure 9 of Percival et al. 2009—note that lines of different age are completely degenerate in this diagram). Here, it can be seen that altering any of the atmospheric parameters simply moves the scaled-solar SSP points along the scaled-solar line, changing the inferred \([\text{Fe}/\text{H}]\) but not altering the inferred abundance ratios for other elements to non-solar ratios. This is an important point to notice since individual element abundances, including several \( \alpha \) elements, appear to alter substantially, as demonstrated by the measured offsets in the various line indices listed in Tables 1 and 2.

In conclusion, we urge caution against the overinterpretation of stellar population parameters from line-index data, in terms of the inferred scaled-solar or non-scaled-solar abundance ratios and also the inferred presence of a composite population, especially when multiple-index or full-SED fitting methods are employed. We find that, for SSPs, \( \text{Fe}5406 \) in combination with \([\text{MgFe}]\) provides the most robust indication of non-solar abundance ratios. We remind the reader that our results potentially impact on all SPS methods, whether fitting functions or full SEDs are employed. Measured offsets for 23 commonly used diagnostic line indices are provided, and we encourage the user to determine the overall impact on their observational data and preferred fitting method.

Figure 4. \( \text{Fe}5406 \) vs. \([\text{MgFe}]\) diagram for 14 Gyr SSP scaled-solar (solid line) and \( \alpha \)-enhanced (dashed line) models, joined at approximately equal \([\text{Fe}/\text{H}]\) (dotted lines) for \([\text{Fe}/\text{H}] \sim +0.06, -0.3, -0.7, \) and -1.0 (decreasing from the top downward). Test reference SSPs are marked as squares, and other symbols are as for Figure 1, for the 4 Gyr (open symbols) and 14 Gyr (solid symbols) cases.

We thank the anonymous referee for a constructive report and some useful suggestions which helped to put our results in context. S.M.P. expresses heartfelt thanks to Elaine Smith-Freeman for many useful discussions and for providing the initial motivation to do this work. S.M.P. acknowledges financial support from the Science & Technology Facilities Council (STFC) through a Postdoctoral Research Fellowship.
REFERENCES

Alonso, A., Arribas, S., & Martínez-Roger, C. 1996a, A&AS, 117, 227
Alonso, A., Arribas, S., & Martínez-Roger, C. 1996b, A&A, 313, 873
Alonso, A., Arribas, S., & Martínez-Roger, C. 1999a, A&AS, 139, 335
Alonso, A., Arribas, S., & Martínez-Roger, C. 1999b, A&AS, 140, 261
Bruzual, A. G. 2007, in IAU Symp. 241, ed. A. Vazdekis & R. F. Peletier (Dordrecht: Kluwer), 125
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Carter, D., et al. 2009, MNRAS, 397, 695
Casagrande, L., Portinari, L., & Flynn, C. 2006, MNRAS, 373, 13
Cenarro, A. J., Cervantes, J. L., Beasley, M. A., Marín-Franch, A., & Vazdekis, A. 2008, ApJ, 689, L29
Coelho, P., Bruzual, G., Charlot, S., Weiss, A., Barbuy, B., & Ferguson, J. W. 2007, MNRAS, 382, 498
Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371
González Hernández, J. I., & Bonifacio, P. 2009, A&A, 497, 497
James, P. A., Salaris, M., Davies, J. I., Phillipps, S., & Cassisi, S. 2006, MNRAS, 367, 339
Kotulla, R., Fritz, U., Weilbacher, P., & Anders, P. 2009, MNRAS, 396, 462
Maraston, C. 2005, MNRAS, 362, 799
Martins, L. P., & Coelho, P. 2007, MNRAS, 381, 1329
Munari, U., Sordo, R., Castelli, F., & Zwitter, T. 2005, A&A, 442, 1127
Percival, S. M., Salaris, M., Cassisi, S., & Pietrinferni, A. 2009, ApJ, 690, 427
Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168
Ramírez, I., & Meléndez, J. 2005, ApJ, 626, 446
Salaris, M., & Cassisi, S. 2007, A&A, 461, 493
Sánchez-Blázquez, P., Gorgas, J., Cardiel, N., & González, J. J. 2006a, A&A, 457, 787
Sánchez-Blázquez, P., et al. 2006b, MNRAS, 371, 703
Sánchez-Blázquez, P., et al. 2009, A&A, 499, 47
Schiavon, R. P. 2007, ApJS, 171, 146
Schiavon, R. P., Faber, S. M., Rose, J. A., & Castilho, B. V. 2002, ApJ, 580, 873
Schiavon, R. P., Rose, J. A., Courteau, S., & MacArthur, L. A. 2004, ApJ, 608, L33
Schiavon, R. P., Rose, J. A., Courteau, S., & MacArthur, L. A. 2005, ApJS, 160, 163
Soubiran, C., Katz, D., & Cayrel, R. 1998, A&AS, 133, 221
Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 339, 897
Trager, S. C., Faber, S. M., & Dressler, A. 2008, MNRAS, 386, 715
Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, AJ, 119, 1645
Trager, S. C., Worthey, G., Faber, S. M., Burstein, D., & Gonzalez, J. J. 1998, ApJS, 116, 1
Vazdekis, A., Salaris, M., Ariimoto, N., & Rose, J. A. 2001, ApJ, 549, 274
Vázquez, G. A., & Leitherer, C. 2005, ApJ, 621, 695
Worthey, G. 1994, ApJS, 95, 107
Worthey, G., & Ottaviani, D. L. 1997, ApJS, 111, 377