The Astrophysical Journal, 868:84 (13pp), 2018 December 1 https://doi.org/10.3847/1538-4357/aae829

Setting the Stage for Cosmic Chronometers. I. Assessing the Impact of Young Stellar Populations on Hubble Parameter Measurements

Michele Moresco[1,2,7], Raul Jimenez[3,4], Licia Verde[3,4], Lucia Pozzetti[2], Andrea Cimatti[1,5], and Annalisa Citro[1,2,6]
[1] Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, I-40129, Bologna, Italy; michele.moresco@unibo.it
[2] INAF—Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129 Bologna, Italy
[3] ICC, Instituto de Ciencias del Cosmos, University de Barcelona, UB, Marti i Franques 1, E-08028, Barcelona, Spain
[4] ICREA, Pg. Lluis Companys 23, E-08010 Barcelona, Spain
[5] INAF—Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy
[6] Center for Gravitation, Cosmology and Astrophysics, Department of Physics, University of Wisconsin–Milwaukee, 3135 N. Maryland Avenue, Milwaukee, WI 53211, USA
Received 2018 April 16; revised 2018 October 10; accepted 2018 October 10; published 2018 November 26

Abstract

The expansion history of the universe can be constrained in a cosmology-independent way by measuring the differential age evolution of cosmic chronometers. This yields a measurement of the Hubble parameter $H(z)$ as a function of redshift. The most reliable cosmic chronometers known so far are extremely massive and passively evolving galaxies. Age-dating these galaxies is, however, a difficult task, and even a small contribution of an underlying young stellar population could, in principle, affect the age estimate and its cosmological interpretation. We present several spectral indicators to detect, quantify, and constrain such contamination in old galaxies and study how their combination can be used to maximize the purity of cosmic chronometers selection. In particular, we analyze the Ca II H/K ratio, the presence (or absence) of Hα and [O II] emission lines, higher-order Balmer absorption lines, and UV flux; each indicator is especially sensitive to a particular age range, allowing us to detect young components ranging between 10 Myr and 1 Gyr. The combination of these indicators minimizes the contamination to a level below 1% in the case of ideal data. More importantly, it offers a way to control the systematic error on $H(z)$ as a function of the contamination by young stellar populations. We show that for our previous measurements of the Hubble parameter, the possible bias induced by the presence of a younger component is well below the current errors. We envision that these indicators will be instrumental in paving the road for a robust and reliable dating of the old population and its cosmological interpretation.

Key words: cosmological parameters – cosmology: observations – galaxies: evolution – galaxies: stellar content

1. Introduction

The cosmic chronometers (CC) approach, initially proposed by Jimenez & Loeb (2002), can provide cosmology-independent measurements of the expansion rate of the universe: the Hubble parameter $H(z)$.

It is difficult to overstate the importance of having independent methods to measure the same cosmological parameter. First of all, different methods are affected by different systematics, therefore offering a test for robustness of the measurement. At a more general level, different methods that rely on different physics offer powerful consistency checks of the underlying model and can even be used to constrain or explore new physics. A case in point is the measurement of the Hubble constant H_0. There is a current discrepancy (at the $\sim 3\sigma$–4σ level) between the expansion history measured by the local cosmic ladder method at $z \sim 0$ (Riess et al. 2018) and the one inferred from cosmic microwave background (CMB) observations by assuming the Λ cold dark matter (ΛCDM) model (Planck Collaboration et al. 2016). While the discrepancy level does not yet reach the "golden" discovery threshold, it has attracted attention and been scrutinized closely. There are two solutions to this discrepancy: either it is due to systematic errors in either method (or both methods), or it could be the signature of new physics and the first sign that the standard minimal ΛCDM cosmological model is not adequate to describe the universe (for a detailed analysis of the tension, see, e.g., Verde et al. 2013; Bernal et al. 2016 and references therein).

The basic idea of the CC method is that the Hubble parameter is related to the scale factor a and the differential redshift–time relation ($dt/dz \simeq \Delta t/\Delta z$) as

$$H(z) = \frac{\dot{a}}{a} = \frac{-1}{1+z}\frac{\Delta z}{\Delta t}, \quad (1)$$

assuming a Friedman–Robertson–Walker (FRW) metric. With this minimal assumption, if the differential redshift–time relation could be measured, then one could measure the universe's expansion history in a cosmology-independent fashion, therefore directly testing the cosmological model. Measuring redshifts, and thus Δz (e.g., between two galaxies), is straightforward with spectroscopic observations. However, measuring Δt is much more challenging. It requires that standard clocks exist throughout the universe and that they can be read. The requirements on this are less stringent than one could naively imagine, since what is needed are differential ages, not absolute ages: systematic effects that result in a constant offset of the age determination get canceled out in the differential measurement. This is the origin of the idea of CCs and that stellar evolution could provide such standard clocks (Jimenez & Loeb 2002). If an old and passively evolving stellar population could be found across a range of redshifts, such as its stars had all formed synchronously and evolved passively since, that would be suitable to be our CC (others clocks have been proposed—e.g., Daly et al. 2008; Diaferio et al. 2011; Lavaux & Wandelt 2012—but will not be discussed here).

[7] Corresponding author.

Main-sequence stars burn H into He via the *pp*-chain. Empirical evidence for this process has been provided by the solar neutrino experiments that measure the neutrino flux from the Sun. Once the chemical composition of an ensemble of stars composing a simple stellar population is known, one can predict its age with high accuracy (in particular, if it is possible to resolve them). Indeed, stellar dating has been and is used to calculate, independently of cosmology, the age of the universe by dating globular clusters (e.g., Jimenez et al. 1995; Marín-Franch et al. 2009), being this determination is one of the first hints for a cosmological constant. Relative ages of globular clusters are computed accurately to the % level (Marín-Franch et al. 2009). More recently (Moresco et al. 2012a), a different approach has been suggested that does not directly rely on the estimate of the age of a stellar population but instead uses a direct spectroscopic observable (the 4000 Å break) known to be linearly related (at fixed metallicity) to the age of the stellar population. In this way, it is possible to rewrite Equation (1) in the form

$$H(z) = \frac{-A}{(1+z)}\frac{dz}{dD4000}, \quad (2)$$

with the advantage of decoupling systematic and statistical effects and being more robust against the models assumed to calibrate the relation.

Unfortunately, we do not see individual stars or resolved stellar populations like globular clusters beyond the Local Group; to find CCs, we have to rely on dating unresolved stellar populations (Tinsley 1968). With the advent of high-resolution spectroscopy in 10 m class telescopes and more accurate stellar models and fitting methods (see, e.g., Reichardt et al. 2001; Tojeiro et al. 2007; Choi et al. 2014; Chevallard & Charlot 2016; Citro et al. 2016), it is now possible to estimate the ages of the stellar populations of galaxies with increasing accuracy. This has been done for large samples like the Sloan Digital Sky Survey (SDSS), VIMOS, and others or targeted small samples of galaxies. If the old stellar population of these galaxies (or an identifiable sample of them across different redshifts) were synchronized, these could be CCs.[8] There is considerable empirical evidence (Dunlop et al. 1996; Spinrad et al. 1997; Cowie et al. 1999; Heavens et al. 2004; Thomas et al. 2005; Treu et al. 2005; Panter et al. 2007; Citro et al. 2016) for the existence of a population of galaxies that formed its stellar population very early (∼1 Gyr after the Big Bang), finished its major star formation episode by $z \gtrsim 2$, and since then has been evolving passively with only very minor "frosting," if any at all. These galaxies tend to be very massive and harbored in the highest-density regions of galaxy clusters. In theory, a young stellar component, even if not dominant in terms of the stellar mass of the global population, can influence the spectrum and potentially bias the measurements. Hence, these objects have to be selected accurately to be CCs and to estimate their old stars' age correctly. In the last decade, the CC approach has been applied to several data sets of increasing size to cover a wider range in redshift, thus providing a more detailed picture of how H changes as a function of redshift (Jimenez et al. 2003; Simon et al. 2005; Stern et al. 2010; Moresco et al. 2012a; Zhang et al. 2014; Moresco 2015; Moresco et al. 2016b). These measurements yield the expansion history directly and independently of the adopted cosmological model; they do not rely on a CMB-calibrated "ruler" as, e.g., baryon acoustic oscillations (BAO) do. This makes the CC method so attractive for model testing.

The main aim of this paper is twofold: (i) quantify the impact on the Hubble parameter, estimated with the CC approach, of a possible contamination from a young component and provide its corresponding covariance matrix as a component of the total error budget; and (ii) provide a clear set of indicators and a selection workflow that can be used to select a pure sample of passively evolving galaxies, minimizing (with respect to other approaches adopting only one or a few criteria) the possible contamination from a young component. This paper is organized as follows. In Section 2, we pedagogically explain the challenges of the CC approach and review approaches in the literature. In Section 3, we describe the method used to explore the presence of a young component underlying the spectrum of an older, passively evolving population. In Section 4, we explore where the signal of a young contribution is localized in the observed spectrum, identifying different tracers to discriminate such a component at different ages, namely the Ca II H/K ratio, the presence of [O II] or Hα emission lines, the UV flux, and the presence of strong-absorption Balmer lines. In Section 5, we dissect how such contributions could potentially affect the measurement of $D4000$ and hence of the Hubble parameter $H(z)$. In Section 6, we present the recipe for an optimal selection of CCs based on the results of the previous sections. Finally, in Section 7, we draw our conclusions.

This paper is the first of a series of papers in which we will revisit the main systematic effects related to the CC method. In Paper II, we will address the dependence of the results on the assumed stellar population synthesis (SPS) models using state-of-the-art models, and in Paper III, we will discuss in further detail the issue of metallicity and how to improve its measurement.

2. The Problem

The CC method relies on accurate differential dating of integrated stellar populations, which is very challenging, and one needs to be very aware of the systematics that can impact the parameter extraction from the integrated stellar light. First, the integrated light is the convolution of many individual bursts of star formation, each with its own age and metallicity. These get combined to produce the integrated light of a galaxy. While the direct problem, to predict the integrated light given a star formation law, is straightforward, the inverse problem is much harder.

In mathematical terms, single stellar populations (SSPs) are the building blocks of any arbitrarily complex population, since the latter can be computed as a sum of SSPs once the star formation rate is provided. In other words, the luminosity of a stellar population of age t_0 (since the beginning of star formation) can be written as

$$L_\lambda(t_0) = \int_0^{t_0}\int_{Z_i}^{Z_f} L_\lambda^{\rm SSP}(Z, t_0 - t)\,dZdt, \quad (3)$$

where the luminosity of the SSP is

$$L_\lambda^{\rm SSP}(Z, t_0 - t) = \int_{M_u}^{M_d} {\rm SF}(Z, M, t) l_\lambda(Z, M, t_0 - t)\,dM \quad (4)$$

[8] In principle, any galaxy could be used as a CC if the old stellar population is dominant and can be disentangled from the young one.

and $l_\lambda(Z, M, t_0 - t)$ is the luminosity of a star of mass M, metallicity Z, and age $t_0 - t$; Z_i and Z_f are the initial and final metallicities; M_d and M_u are the smallest and largest stellar masses in the population, and $\mathrm{SF}(Z, M, t)$ is the star formation rate at time t. It is worth emphasizing that $l_\lambda(Z, M, t_0 - t)$ can be computed accurately from the first principles of stellar evolution. This is the direct problem, which is well under control.

But it is easy to see from Equation (3) that the inversion process needed to recover the age of underlying SSPs of a galaxy may fail. It might happen that the integrated light will not have enough information to permit a full inversion of Equation (3). Fortunately, this inversion problem was addressed formally in Tojeiro et al. (2007; but see also Connolly et al. 1995; Yip et al. 2004; Wild et al. 2007; Chen et al. 2012; Yip et al. 2014) by effectively letting the data decide how much information could actually be recovered. Given an amount of data and signal-to-noise ratio (S/N) over a given spectral range, the above procedure returns a coarser or finer reconstructed star formation history (SFH), with a corresponding uncertainty, depending on the information content of the data. Of course, other approaches can be exploited to estimate the uncertainty on the SFH reconstruction, such as regularization (Cappellari & Emsellem 2004) and bootstrap (Walcher et al. 2015). The specific choice of technique is not central to our argument here.

The main systematics that could potentially affect the CC method have been extensively studied. These are (i) the dependence on the stellar metallicity estimate, (ii) the reliance on SPS models, (iii) the progenitor bias, and (iv) the presence of an underlying young component. For points (i), (ii), and (iii), we refer to Moresco et al. (2012a, 2016b), where a detailed analysis of these issues has been presented. The robustness and consistency of CC relative-ages estimation has been studied before, e.g., by Crawford et al. (2010) and Liu et al. (2016). These authors found that $H(z)$ can be recovered with a 3% error at intermediate redshifts, keeping systematic errors safely below the statistical ones. Here we concentrate on point (iv).

It is worth recalling the different approaches taken in the literature, which have evolved with the size and quality of the data sets. In Jimenez et al. (2003), the age determination was done by reconstructing the so-called red envelope (Butcher et al. 1983; Lilly & Gunn 1985; O'Connell 1988). Full spectral fitting assuming a single age was used to age-date a sample of simulated galaxies where a young population was superimposed on an old one. It is shown (see Figure 5 of Jimenez et al. 2003) that even without an accurate preselection of target galaxies, a random burst of recent star formation (a young population) in a mostly dominant old population does not affect the recovered age–redshift relation obtained from the red envelope, provided there are enough galaxies to populate the red envelope. This technique was successfully applied to a densely sampled galaxy survey (SDSS). Alternatively, it is possible to proceed with a more accurate selection of passively evolving galaxies, studying their mean evolution as a function of cosmic time. Both approaches are viable, each one having its own advantages and drawbacks. In Moresco et al. (2012a), both have been explored, and it was found that the limit of the red envelope approach is that it is highly dependent on statistics, needing a large data set to properly sample the extreme of the distribution (e.g., the 5% or 1% oldest galaxies); on the contrary, the other approach needs a better control in the sample selection, but it is more stable, providing smaller errors at the end (as shown in Moresco et al. 2012a).

In Stern et al. (2010), full spectral fitting with an extended SFH was used, and the wavelength coverage required to recover the age–redshift relation from simulations was explored (see Figures 1–3) as well. In Moresco et al. (2016b), a different approach was taken. A new estimator ($D4000$) for the galaxy age was adopted that is robust, provided it is applied to a preselected sample of passively evolving galaxies (CC). It is also shown how to recover the metallicity and SFHs of galaxies and how this information can be used to robustly select such a sample of galaxies suitable to be CCs.

We live in an era of massive spectroscopic surveys that will bring a treasure trove of galaxies and could further increase the precision and accuracy of the CC method. This offers the prospect of providing more precise cosmology-independent measurements of $H(z)$ with the potential of new physics discoveries. Passively evolving galaxies will not be specifically targeted, and spectra will cover the range in the rest-frame near-UV to near-IR. Here we study how to select CCs from features in the optical rest-frame spectrum and eliminate the contamination of an underlying star-forming population.

In particular, we present several spectral indicators that facilitate the disentangling and discrimination of the presence of a star-forming population contribution contaminating passively evolving galaxies, further strengthening the selection criteria of optimal CCs and thus paving the way for a robust and reliable dating of the old population. We then present a complete "recipe" to select and age-date CCs.

3. Method

To study the impact of a young component on the spectrum of an older, passively evolving galaxy, we build synthetic spectra composed of a mixture of a younger and an older population. We explore a variety of different models to assess the dependence of our results on the SPS model assumed. We create the library of synthetic spectra using the Maraston & Strömbäck (2011; hereafter M11), updated 2016 Bruzual & Charlot (2003; hereafter BC16), and Vazdekis et al. (2016) models. All models are based on the latest MILES stellar library (Falcón-Barroso et al. 2011) but encompass different recipes and ingredients to construct the synthetic spectra (for further discussion, see e.g., Maraston et al. 2006; Maraston & Strömbäck 2011). We also choose the Chabrier initial mass function (IMF; Chabrier 2003), but as demonstrated in Moresco et al. (2012a), the impact on the $D4000$ (the main feature analyzed in this work) of the adopted IMF is negligible.

Currently, CC data have been exploited in the redshift range $0.15 < z < 2$, and we therefore want to span a grid of ages sensible for this entire redshift range. We thus consider, for the old component, a solar metallicity, exploring the ages $=$ [2, 4, 6, 8, 10, 12] Gyr, and for the young component, several values ranging from 0.1 to 1 Gyr with stellar metallicity in the range $0.4 < Z/Z_\odot < 2.5$. The choice of ages for the young component is motivated by the fact that in previous CC analysis, the effect of an extended SFH was already taken into account (e.g., a delayed exponential SFH, compatible with the colors and spectra of the data), while here we want to explore the possible bias due to a much more recent burst of star formation.

The fraction of the young population flux contribution with respect to the old population is parameterized by

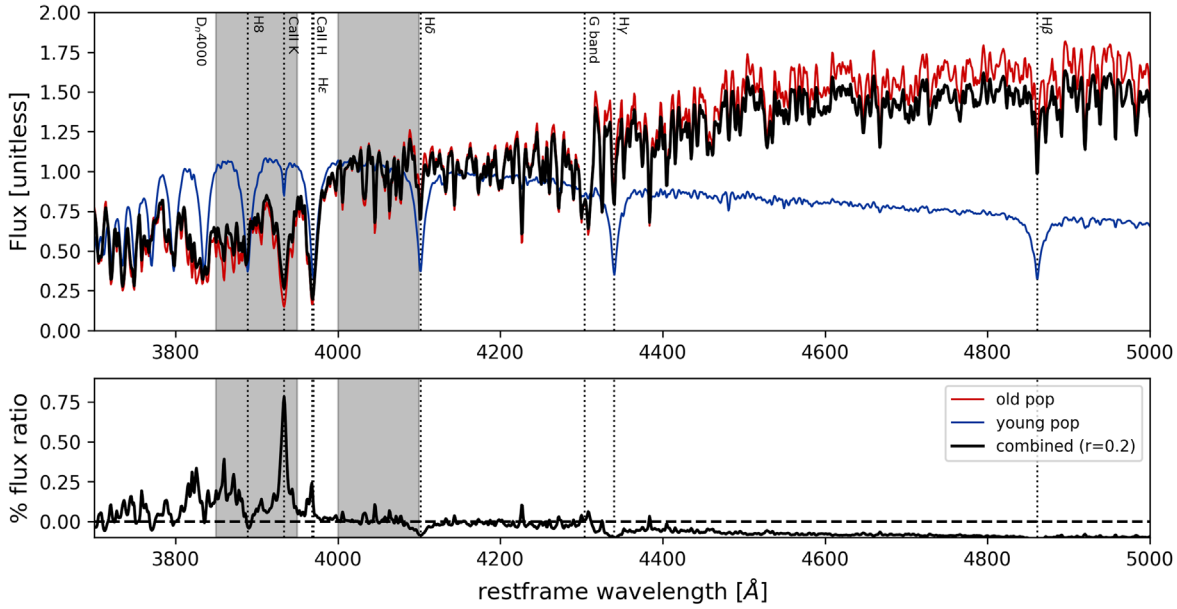

Figure 1. Synthetic spectra. For illustrative purposes, the upper panel shows the synthetic spectra of an old component (red line; 8 Gyr), a young component (blue line; 0.1 Gyr), and their combination, with a fraction $r = 0.2$ (black line). All of the spectra have been generated with BC16 models assuming a MILES library, Chabrier IMF, and solar metallicity. The lower panel shows the percentage difference between the old and combined spectra (namely, $r(\lambda) - 1$); as can be seen, the larger impact is in correspondence to the Ca II K line, due to the onset of the Ca II H/K inversion.

$r(\lambda) = F_{\rm young}/F_{\rm old}$. This definition of r depends on the bandpass used to determine $F_{\rm young}$ with respect to $F_{\rm old}$. Here we adopt two different wavelength ranges (representative of the two D_n4000 bandpasses), namely [4000–4100] Å (for r) and [3850–3950] Å (for r'), to explore the impact of the range of normalization on the results.

In Figure 1, we show, for illustrative purposes, the spectra obtained at 8 Gyr for the old population, 0.1 Gyr for the young component, and their combination, with a fraction $r = 0.2$ normalized at [4000–4100] Å. From the plots, all the absorption lines typical of the two components are evident. In particular, the older spectrum is characterized by a strong $D4000$, significant Ca II H and K lines, and the G-band feature at 4304 Å, while the younger spectrum is dominated by strong Balmer lines, namely Hβ, Hγ, Hδ, and Hϵ. The lower panel presents an example of the percentage difference between the spectrum of the old population and the combined spectrum (i.e., $r(\lambda) - 1$), showing how this ratio changes as a function of the wavelength.

Noticeably, the young component has an impact not only on the shape of the continuum, affecting the $D4000$, but also on many lines that characterize the spectrum of the passive galaxy, as we will discuss in detail in Section 4.1 and clearly shown in Figure 2. Medium-to-high-resolution spectroscopy will then provide crucial information to discriminate this effect.

Finally, to complete our analysis and study the impact of different assumed values of α-enhancement on our results, we also studied the newest (Vazdekis et al. 2016) models, considering BaSTI isochrones, different metallicities, and values of [α/Fe].

We stress here that the aim of this paper is to address the impact of a possible contamination due to a young component on $H(z)$ measurements, while we defer to Paper II a more complete discussion of how the assumed SPS models affect these measurements. However, we also note that this dependence has been extensively studied in previous works. In particular, in Moresco et al. (2012b; see Sections 3.3 and 4 and Figure 6) and Moresco et al. (2016b; see Sections 4.2 and 5.1 and Figures 4 and 5), $H(z)$ measurements have been provided assuming two completely different SPS models to calibrate the method, obtaining measurements compatible at the 1σ level, while in Moresco et al. (2012a; see Section 4.3) and Moresco et al. (2016a; see their Appendix A), the impact of this effect on the estimated cosmological parameters was studied, confirming that also in this case, all the differences are well below the 1σ level.

4. Multiple Stellar Populations: Where Is the Signal?

We begin by highlighting how different stellar populations of different ages contribute to the integrated stellar spectrum of a galaxy. To do this, we resort to the data compression algorithm MOPED (Reichardt et al. 2001), which can massively reduce the number p of data points from p to m, where m is the number of parameters in the model, without losing information. In this process and when applied to a galaxy spectrum, MOPED computes the weights to be applied to the original data points for compression. By choosing as parameters of the model the amplitude of each of the single bursts of star formation that comprise the building blocks of an SFH, we can visualize how each age contributes to the spectrum as a function of wavelength. This tells us where the signal for each age is located as a function of wavelength. These weights constitute eigenvectors: by construction, they are orthonormal. Indicating by x the input (data) spectrum (of length p) and by $\boldsymbol{b}_i$ the eigenvector for the parameter (age bin) i, the weighted (compressed) data $y_i = \boldsymbol{b}_i^T \boldsymbol{x}$ are uncorrelated. This is a very useful property, as the information about each age "bin" is uncorrelated (and uncorrelated with metallicity).

First, for the typical SDSS and VIMOS spectra (because of resolution, S/N, and wavelength coverage) considered by Moresco et al. (2012a, 2016b) and Moresco (2015), it is only possible to reconstruct eight age bins in the SFH of a galaxy (Tojeiro et al. 2007). We chose age bins such that the error in

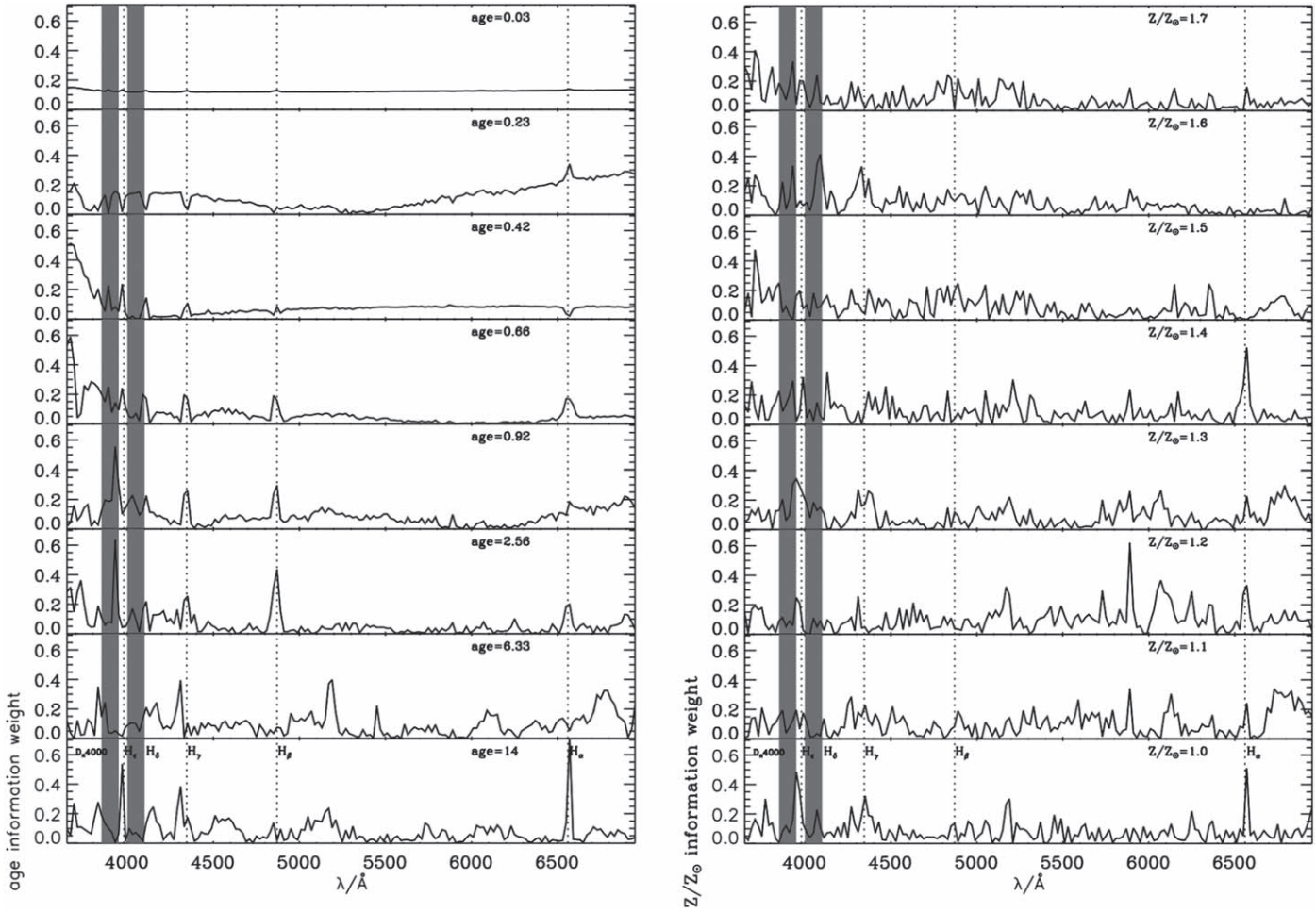

Figure 2. MOPED weights that show the information content as a function of wavelength for eight bursts at ages 0.03, 0.23, 0.42, 0.66, 0.92, 2.56, 6.33, and 14 Gyr (top to bottom). The value of the weight (*y*-axis in arbitrary units) gives a measure of how sensitive the integrated spectrum is to the stellar component of a given age (left panel) and metallicity (right panel); these 16 variables are fitted simultaneously without any restriction.

the recovered SFH is minimized. The optimal bins are centered at 0.03, 0.23, 0.42, 0.66, 0.92, 2.56, 6.33, and 14 Gyr, which are very close to being equally spaced in log space. To illustrate where the information for age and metallicity is in the spectrum, we construct a composite spectrum with constant SFH (so as to make all age components significant) and increasing metallicity (so as to make this a free variable). All 16 variables are then recovered independently. The precise choice of star formation and metallicity history will not significantly change the location of the information weights. The spectra are constructed without noise, since noise will be observation-dependent. The absolute values of the MOPED eigenvectors (i.e., the "information weight," where a value above zero shows that there is information and zero indicates a lack of information) are shown in Figure 2, where ages are youngest to oldest from top to bottom.

There are several interesting features. First, for all ages, the age information is spread over the rest-frame visible wavelength range. Second, as expected, for the youngest stellar populations, the continuum carries significant information, but the absorption lines also show significant weight. In particular, the most prominent features are the D_n4000 lines and $H_{\beta,\delta,\gamma,\epsilon}$. Note that for older stellar populations, there is (localized) information all across the wavelength range.

Finally, we can see that the youngest population is the most difficult to discriminate because most of the signal is in the continuum, but that information in the D_n4000, in particular in Ca II lines, remains. With appropriate S/N and moderate-resolution spectra like those of SDSS and VIMOS (spectral resolution $R \gtrsim 600$), it is possible to discriminate the young stellar component in passively evolving galaxies.

For the metallicity, the information is spread all over the absorption features and, very importantly, not all coincident with the age weights (by construction, the eigenvectors are orthogonal); this shows that, with sufficient S/N, the age–metallicity degeneracy can be broken.

This result is confirmed by many independent analyses. In Connolly et al. (1995), galaxy spectra were studied by orthogonal basis functions, providing the significant spectral components characterizing each particular galaxy type. Chen et al. (2012; and later Marchetti et al. 2013) used principal component analysis to both classify and measure physical properties of galaxies in the range $0.1 < z < 1$. Finally, Yip et al. (2014), from the analysis of SDSS simulated spectra, identified the region around the 4000 Å break and Hδ as primarily related to the age of the stellar population, with significant information also contained in the anticorrelation of the Ca II H and K lines.

4.1. Ca II H/K Diagnostic

Two of the main prominent features of the spectrum of a galaxy are the Ca II K line at 3934 Å and Ca II H line at 3969 Å; these also partly define the $D4000$.

In particular, in a galaxy dominated by a passive population, it is always found that the K line is deeper than the H line, as can be seen in Figure 1. On the contrary, this relation is inverted when a contribution of younger populations is present, due in particular to the presence of Hϵ λ3970 that overlaps and gets combined with the Ca II H line at 3969 Å (see Figure 1). Therefore, the ratio between the relative intensities of Ca II H and K lines can be considered as an indicator of the relative contribution of a younger population to an older one.

Here we define the ratio H/K as the ratio of the dip of Ca II H to the dip of Ca II K, and we use this quantity to analyze the impact of contamination of a younger population. This indicator was first proposed by Rose (1984) to constrain the ages of starbursts in post-starburst galaxies and later also adopted in different works (see also Rose 1985; Leonardi & Rose 1996; Wild et al. 2007; Sánchez Almeida et al. 2012).

The metallicity of massive and passive galaxies is well constrained to be almost solar, or slightly oversolar, both in the local universe (see, e.g., Gallazzi et al. 2005; Citro et al. 2016) and up to $z \sim 0.7$ (see, e.g., Moresco et al. 2016b), with some indications that this also holds in general at higher redshifts ($z \sim 1.6$; see Onodera et al. 2015). Therefore, for the old component, we assume a solar metallicity. However, this is only done for convenience and to simplify the presentation of the results, as the metallicity of CCs can be estimated from the spectrum, as was done in Moresco et al. (2016b; see also Figure 2). For the younger components, we explore three different ages, 0.01, 0.1, and 0.2 Gyr,[9] considering in each case three different metallicities: $Z/Z_{\odot} = 0.4$, 1, and 2.5 for BC16 and $Z/Z_{\odot} = 0.5$, 1, and 2 for M11. We consider a grid of values of r between zero (when the population is totally dominated by the old component) and 1 (where the two populations have the same weight).

The results are shown in Figures 3 and 4. We find that, almost independently of all the parameters (the age of the oldest population, the age of the younger contribution, its metallicity, the SPS model assumed), there is already at $r \gtrsim 0.05$ an inversion of the H/K ratio, with the Ca II H line becoming deeper than Ca II K. The figure also shows the Ca II H/K values obtained in the spectra analyzed in Moresco et al. (2012a, 2016b), having a mean value $\langle H/K \rangle = 1.17 \pm 0.05$ (at 1σ; ± 0.1 at 2σ). As can be seen in the figure, almost independent of the age of the young component, these measurements are compatible with extremely low values of r, namely $r \lesssim 0.011$ (at 1σ; $r \lesssim 0.027$ at 2σ). However, we emphasize that, in future analysis, the selection can be made even more stringent, since the optical spectrum provides enough information to reduce r below the 0.01 level. Moreover, as we demonstrate in the next sections, additional information can be gathered from other absorption and emission features in the optical spectra.

Repeating the analysis procedure by also considering the Vazdekis et al. (2016) models with different [α/Fe] ratios, we find results in complete agreement with the ones in Figures 3 and 4, with a Ca II H/K inversion found at $r = 0.037 \pm 0.003$ with varying ages and [α/Fe] ratios and even more stringent constraints on the contamination of current measurements, with $r < 0.011$ at 2σ. This demonstrates the robustness of this indicator against possible different levels of α-enhancement.

[9] For young components beyond 1 Gyr, the H/K ratio is not inverted, but these components can be discriminated with other indicators; see Section 4.4

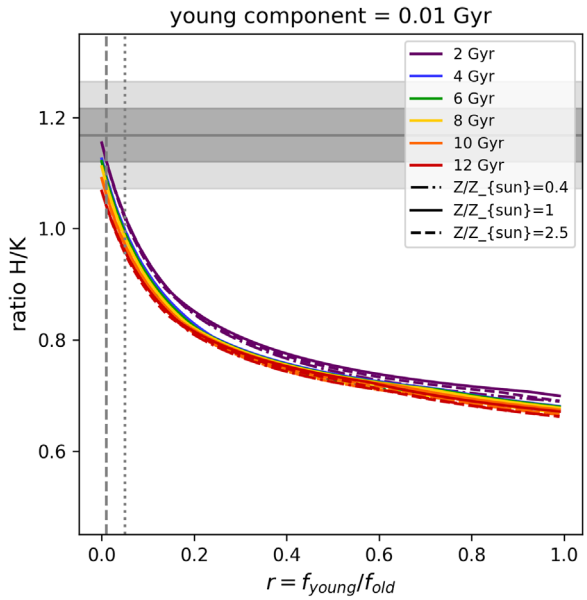

Figure 3. Ratio of the Ca II H and K lines as a function of the fraction between young (0.01 Gyr) and old component r. The dotted vertical line shows the value of r at which the two lines are equally deep ($r \sim 0.05$), while the dashed vertical line shows the value of r constrained by the data ($r < 0.01$). The gray shaded area represents the range of H/K values in the data (darker at 1σ, lighter at 2σ) from Moresco et al. (2012a, 2016b).

4.2. Presence (or Absence) of [O II] and Hα Emission Lines

The presence of an underlying young population also has an impact on emission lines in the optical wavelengths. Magris et al. (2003) provided the theoretical expectations for equivalent widths (EWs) of many emission lines as a function of the age of the population and SFH. If we focus in particular on the youngest ages (<100 Myr), we find that, almost independently of the assumed SFH, strong EWs are expected for both Hα and [O II] λ3727, with values EW(Hα) $\gtrsim$ 100–200 Å and EW([O II]) $\gtrsim$ 40 Å.

These estimates assume quite prolonged SFHs, so that at the younger ages, the population can be considered star-forming and dominated in the UV by O and B stars. We also explore the possibility of an instantaneous burst of star formation. We create photoionization models using the CLOUDY photoionization code (version 13.03; Ferland et al. 1998, 2013), assuming our galaxies as formed by a central ionizing source surrounded by a spherical cloud. We simulate the central ionizing source using the same models discussed in Section 3, where a young component (with an age of 0.01, 0.1, or 0.2 Gyr) is superimposed on an older component (with an age of 4, 8, or 12 Gyr) with different fractions r; this allows us to extrapolate the results as a function of the various parameters considered in our analysis. We assume a starting ionization parameter $\log(U) = -3.2$, which is consistent with the observations of local H II regions ($-3.2 < \log(U) < -2.9$; see Dopita et al. 2000) and star-forming galaxies (see Moustakas et al. 2006, 2010). For the ionized nebula, we assume a fixed hydrogen density $n_{\rm H} = 100\ {\rm cm}^{-3}$, which is the typical density of observed star-forming regions

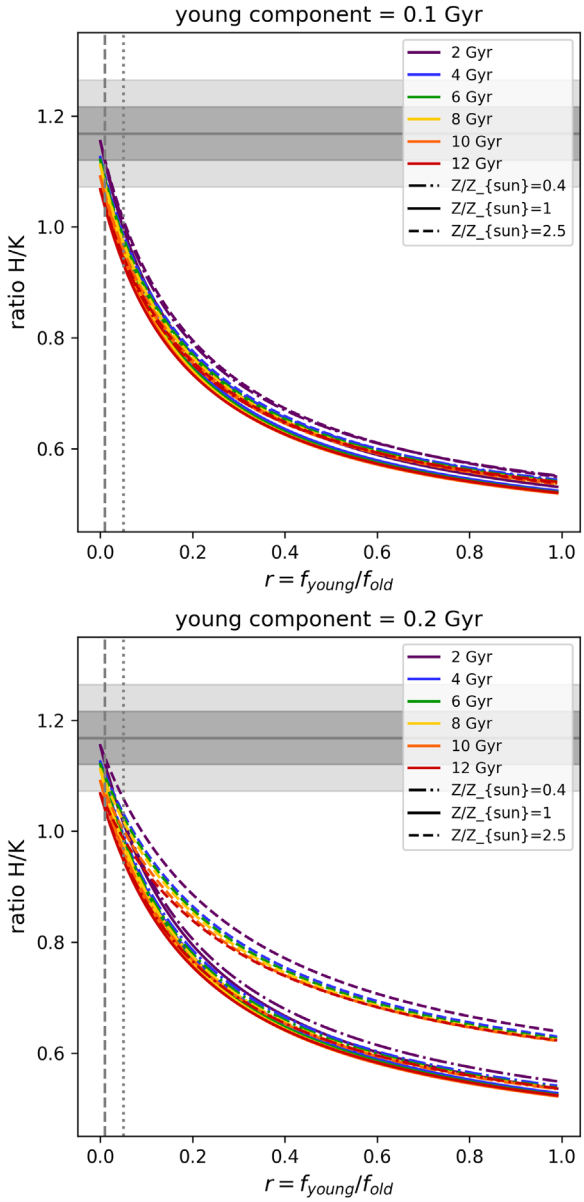

Figure 4. Same as Figure 3, but in this case, the young component is 0.1 and 0.2 Gyr old.

(Dopita et al. 2000, 2006; Kewley et al. 2001). We adopt the solar chemical composition by Asplund et al. (2005), matching the gas metallicity of the ionized nebula with the stellar metallicity of the ionizing population. Moreover, we account for the presence of dust, adopting the default CLOUDY interstellar medium (ISM) grains distribution.

We find that both [O II] and Hα emission lines are extremely sensitive to the presence of the youngest and hottest stars with ages $\lesssim$10 Myr, with significant fluxes already with a contamination $r = 0.05$, and in particular F(Hα) $\sim$ 0.2–0.9 $\times$ 10^{-15} and F([O II]) $\sim$ 0.5–3 $\times$ 10^{-15} erg s^{-1} cm^{2} (depending on the age of the old component). Similarly, at the same level of contamination, we also find relatively high values of EW, EW(Hα) $\sim$ 3.5–9.5 and EW([O II]) $\sim$ 7–30. These values increase by $\sim$70% when the contamination is $r \sim 0.1$. At older ages of the young component ($\gtrsim$100 Myr), there is a significant drop in UV emission (as can be seen also in Levesque et al. 2010; Citro et al. 2017), so that it is no longer able to produce significant emission lines; however, these other phases can be discriminated with other indicators (see Section 4.1).

To select optimal probes for the CC approach and minimize the possible contamination due to a younger population, it is crucial to carefully select galaxies not only on the basis of photometry but also from spectroscopy. In Mignoli et al. (2009), a criterion was proposed to separate passive and star-forming galaxies on the basis of a cut in [O II] (EW([O II] $<$ 5 Å). Similarly, in Wang et al. (2018), passive galaxies were separated from the other populations on the basis of the detectability of Hα emission lines, and in particular having $S/N_{H\alpha} < 3$. In Moresco et al. (2012a, 2016b), galaxies were carefully selected by rejecting objects with detectable emission lines in Hα and [O II], with EW thresholds that are compatible, given the previously discussed models, with no UV emission. In particular, in Moresco et al. (2012a), a criterion EW ([O II]) $<$ 5 Å was applied (according to Mignoli et al. 2009), and in Moresco et al. (2016b), EW([O II]) $<$ 5 Å and an S/N ([O II], Hα, Hβ, [O III]) $<$ 2. In both cases, the analysis of the stacked spectra highlighted no evidence for emission lines.

4.3. The UV Contribution

Another significant tracer of star formation is the UV flux (Kennicutt 1998). There is much observational evidence that some elliptical galaxies show the presence of UV emission. The mechanism behind this emission is still debated, with suggestions that there could be other processes than the presence of a young component producing a UV excess (e.g., see Greggio & Renzini 1990); therefore, any conclusion about a young component based on the presence of UV emission should be carefully considered as an upper limit. Typically, these processes can be discriminated on the basis of some color–color diagrams. In particular, Yi et al. (2011; but see also Han et al. 2007; Donahue et al. 2010; Boissier et al. 2018) proposed a combination of UV and optical colors to separate the UV emission coming from a young component or a UV excess. They suggested that a combination of FUV $-$ NUV and NUV $- r$ colors allows one to separate these components, with FUV $-$ NUV > 0.9 indicating a negative UV slope (hence showing evidence for UV emission) and NUV $- r < 5.4$ indicating emission due to a young component (and, respectively, NUV $- r > 5.4$ indicating emission due to a UV excess). Recently, Ilbert et al. (2013) proposed another selection criterion based on UV colors, the NUVrJ diagram, that was found to be particularly sensitive to young ages (0.1–1 Gyr; Arnouts et al. 2007; Martin et al. 2007). This selection criterion was successfully applied by many authors to safely discriminate between star-forming and quiescent populations (Ilbert et al. 2013; Moutard et al. 2016; Ownsworth et al. 2016; Davidzon et al. 2017), as opposed to the other criteria (UVJ and NUVrK, respectively; Williams et al. 2009; Arnouts et al. 2013), and can therefore be used in combination with the previously discussed ones to provide a pure, passively evolving sample of CCs. The advantage of this criterion, given its large wavelength coverage, is that it is also useful for disentangling effects due to age and extinction and in particular for discriminating possible UV emission hidden by the presence of dust. To explore how much a young component could contribute to UV fluxes (in this case, neglecting the possible contribution of obscuration due to dust), in Figure 5, we compare the combined spectra of an old (8 Gyr) and a younger (0.1 Gyr) component built with different ratios r. We find that a small value of r produces a flattening in

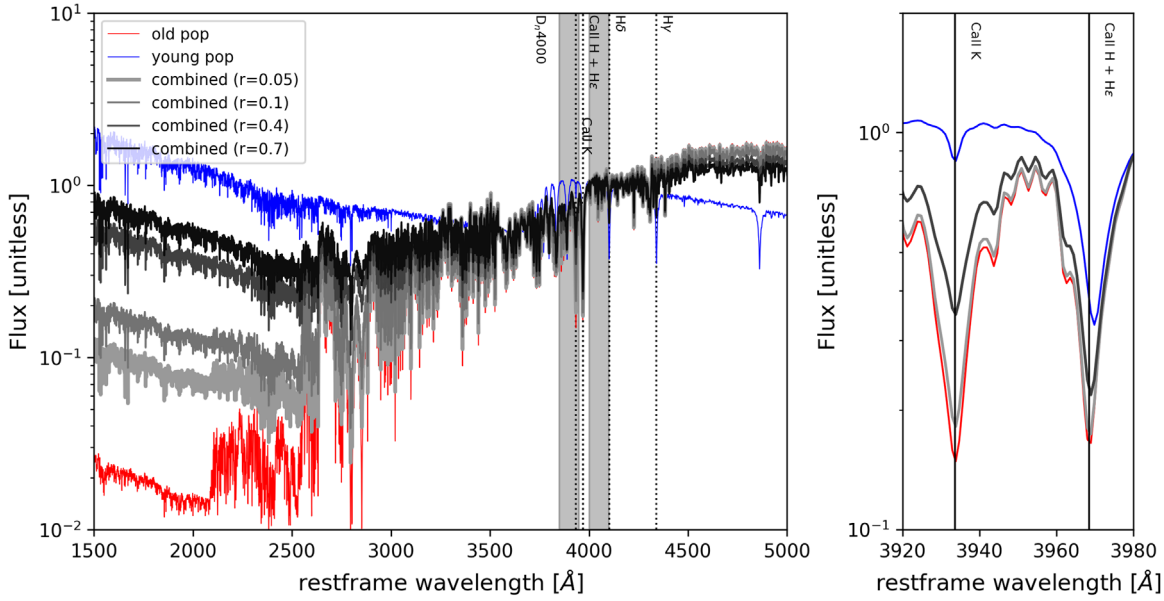

Figure 5. Contribution at UV wavelengths of different percentages of a young stellar component. The red and blue curves represent an old (8 Gyr) and a young (0.1 Gyr) component combined with different ratios to produce the black and gray curves ($r = 0.05$, 0.1, 0.4, 0.7). The impact in the UV can be seen by the total flux produced between 1500 and 2500 Å. The right panel shows a zoom-in around the Ca II H and K lines, showing its inversion as a function of r.

the spectrum at UV wavelengths but with fluxes ∼1 order of magnitude smaller than in the optical, while at relatively high values of $r \gtrsim 0.4$ we obtain a more significant excess UV flux. As demonstrated in Section 4.1, these values of r can be discriminated on the basis of the H/K diagnostic, as also shown in Figure 5. Therefore, the H/K and UV flux diagnostics are highly complementary and can be used in combination to improve the purity of the CC sample selection.

4.4. Higher-order Balmer Lines

Higher-order Balmer lines, such as Hδ and Hγ, are also powerful tools to discriminate recent episodes of star formation. This is easy to understand, since these lines are due to the abundance of H, which changes very little between different values of the metallicity content of galaxies (about 1%); therefore, it effectively measures the gravity of the stars, and thus their age. In particular, strong Hδ absorption lines (EW(H δ) $>$ 4–5 Å) have been historically used to identify post-starburst (or Hδ strong) galaxies, objects that have ceased to form stars in the past (EW([O II]) > -2.5 Å) but experienced recent episodes of star formation (e.g., see Le Borgne et al. 2006; Wilkinson et al. 2017). These indicators have been used to trace this population up to $z \sim 1.5$ (Bezanson et al. 2013). Different from Hϵ, these lines do not overlap with other absorption lines, producing a well-recognizable pattern to the one described for the Ca II H/K inversion in Section 4.1; hence, they can be used to trace star formation on slightly longer timescales than the H/K diagnostic (typically, from hundreds of Myr up to ∼1–2 Gyr, as discussed in Le Borgne et al. 2006; Bezanson et al. 2013; Wilkinson et al. 2017). Given the age of the young population, they are associated with lower values of UV emission. Therefore, by combining this indicator with the H/K ratio, it is possible to cover a wide range of recent episodes of star formation in the galaxies.

5. The Impact of a Young Component on the D_n4000 and Hubble Parameter

The D_n4000 break is a discontinuity in the spectrum of galaxies at the 4000 Å rest frame due to the blending of several absorption features and is defined as the ratio of the mean flux $\langle F_\nu \rangle$ between a red and a blue bandpass:

$$D_n4000 = \frac{(\lambda_2^{\rm blue} - \lambda_1^{\rm blue}) \int_{\lambda_1^{\rm red}}^{\lambda_2^{\rm red}} F_\nu d\lambda}{(\lambda_2^{\rm blue} - \lambda_1^{\rm red}) \int_{\lambda_1^{\rm red}}^{\lambda_2^{\rm blue}} F_\nu d\lambda}. \quad (5)$$

There have been different definitions of these bands, in particular, a broader definition with [3750–3950], [4050–4250] Å (Bruzual 1983; Hamilton 1985) and a narrower definition with [3850–3950], [4000–4100] Å (Balogh et al. 1999), which is less sensitive to reddening effects; here we adopt the narrow definition. It is important to bear in mind that D_n4000 was originally defined in flux F_ν, not in F_λ.

This index is known to depend on the age and metallicity of the stellar population (e.g., see Poggianti & Barbaro 1997 and the results obtained in Section 4). For this reason, first in Moresco et al. (2012a) and later in Moresco (2015) and Moresco et al. (2016b), the corresponding break was used to explicitly relate Equation (1) with Equation (2) and then was used to obtain a measurement of $H(z)$ from the spectral evolution of CCs. It was argued that this choice of (non-lossless) data compression is useful to decouple systematics and statistical effects in the estimate of the age of the stellar population. It has also been verified that α-enhancement has a negligible impact on this spectral feature (Moresco et al. 2012a; see Appendix A.3), representing another advantage for moving into $D4000$–z space instead of age–z space.

Here we test whether and quantify how much this index may be affected by the presence of an underlying younger stellar population.

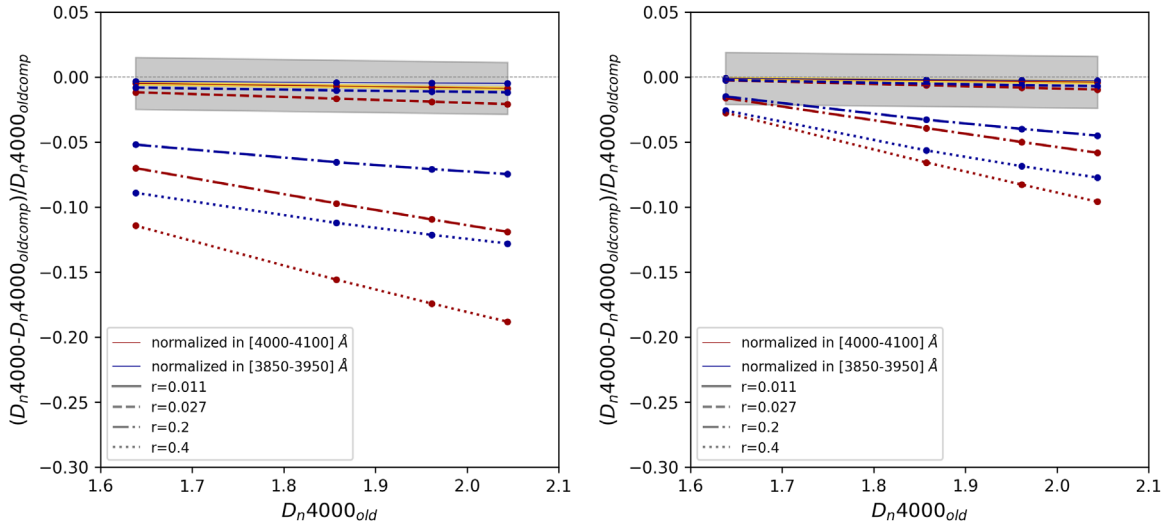

Figure 6. Percentage variation of D_n4000 due to the presence of a young component for different values of r. The left panel shows the case of a 0.1 Gyr young component, and the right panel shows a 1 Gyr young component. We explored two different combinations, with young and old component fluxes normalized in the ranges [4000–4100] Å (r; red lines) and [3850–3950] Å (r'; blue lines). Solid, dashed, dot-dashed, and dotted lines correspond, respectively, to values of $r = 0.011$, 0.027, 0.2, and 0.4. The shaded areas represent a 2% (gray) and 0.1% (yellow) statistical error on $\langle D_n4000 \rangle$, which is the typical error found in Moresco et al. (2012a, 2016b) that can be achieved with a larger ($N_{gal} \gtrsim 10{,}000$) and smaller ($N_{gal} \lesssim 1000$) statistical sample, respectively.

The impact of a young component on the D_n4000 is quantified in Figure 6, where we show the percentage variation of D_n4000 when including a young component for different values of r and r'. To explore the dependence of the results on the age of the young component, we consider two different values, 0.1 and 1 Gyr, that, as discussed in Sections 4.1 and 4.4, have a different impact on the composite spectrum.

As can be seen, the deviation from the unbiased measurement, which would also have an impact on cosmological inference, is more and more significant at increasing values of r. In the case of the youngest component (0.1 Gyr), the effect on D_n4000 ranges from $\lesssim$5% when $r < 0.05$ to $\gtrsim$20% when $r > 0.4$ (we note that, as discussed in Section 4.1, at $r = 0.05$, we are able to distinguish a young component from the H/K inversion). In the case of the oldest component (1 Gyr) instead, the impact is always $\lesssim$10%. At the same time, we see that in both cases, the impact can be kept at a negligible level, at least for current data, if the contribution is limited to small values ($r \lesssim 0.01$).

Recall that since the CC approach is based on a differential measurement of ages (or D_n4000), what would actually impact the cosmological constraints is not a systematic offset (in that case, only the absolute age or D_n4000 measurement would be biased) but rather the relative difference between different D_n4000 measurements, i.e., the slope of the relations shown in Figure 6. The statistical error associated with the measurement of the mean $\langle D_n4000 \rangle$ can be estimated from current data (Moresco et al. 2012a, 2016b); we find that, on average, the statistical percentage error ranges between 2% (in regimes with a smaller statistic with $\lesssim$1000 objects, corresponding to the $z > 0.5$ analysis; Moresco et al. 2012a) and 0.1% (in SDSS-like regimes with $\gtrsim$10,000 objects; Moresco et al. 2012a, 2016b). In Figure 6, only the low-statistics case can be seen, as the other one is almost coincident with the line width.

We find that with a lower number of galaxies, the capability of distinguishing a young component is reduced due to the larger error, and therefore its impact on cosmological measurements should be, by definition, subdominant in the total error budget; with a larger statistics, it is instead fundamental to assess the percentage of contamination r.

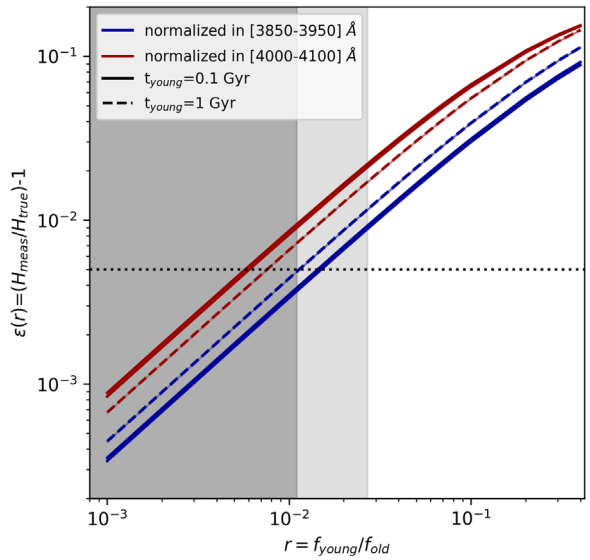

Figure 7. Impact on the measure of $H(z)$ of a young component with fraction r with respect to the old component. The different colors show the effect when the spectra are normalized in different wavelength ranges ([4000–4100] Å for the red curves, [3850–3950] Å for the blue curves), while the different line styles represent different ages of the young component (0.1 and 1 Gyr). The gray shaded areas show the ranges of r allowed by the data in Moresco et al. (2012a, 2016b) averaged across all redshifts (darker at 1σ, lighter at 2σ), and the black dotted line shows an error $\epsilon(r) = 0.5\%$. The shaded area (barely visible) between the curves represents the uncertainty due to different SPS models.

The impact of the presence of a young component with a fraction r on $H(z)$ is estimated in Figure 7. From Equation (2), we derive an analytic expression of the percentage deviation of $H(z)$ with respect to the true one,

$$\frac{H(z)_{\rm meas}}{H(z)_{\rm true}} = \frac{dD_n4000_{\rm true}}{dD_n4000_{\rm meas}} = 1 + \epsilon(r), \qquad (6)$$

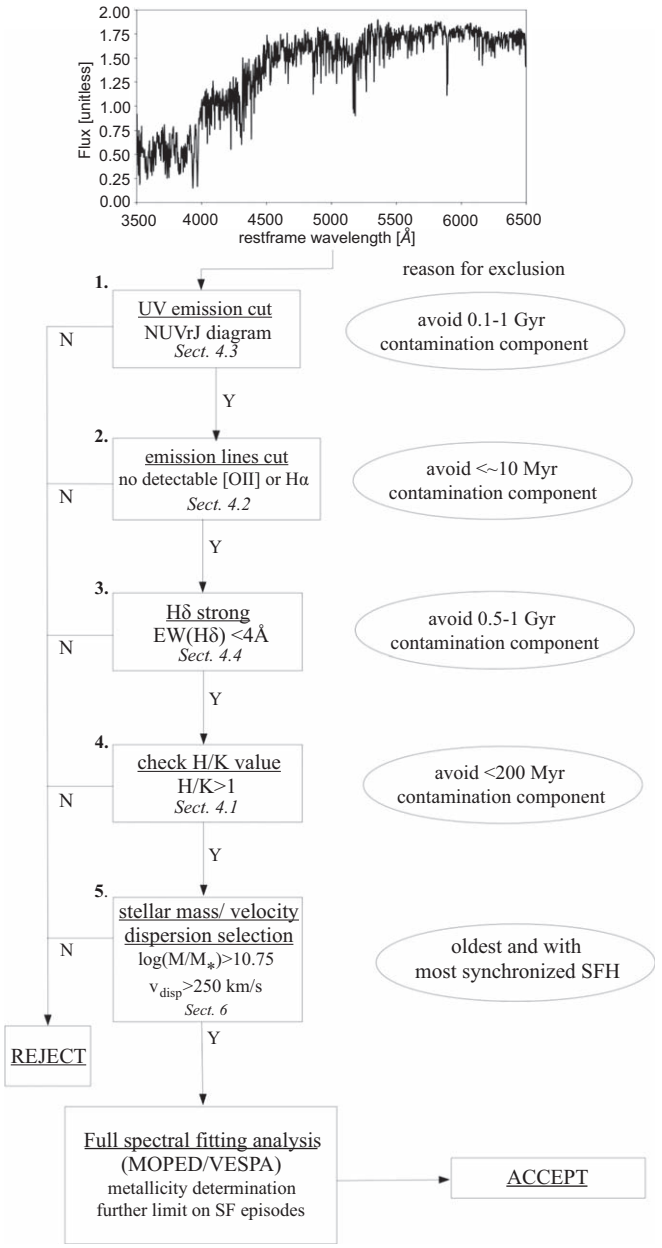

Figure 8. Workflow to optimally select CCs. This is the procedure that we have applied in our previous works. Note that all boxes need to be checked to select a CC. Selection criteria can be made more stringent for future data with improved statistics.

where $\epsilon(r)$ is the percentage variation of dD_n4000 as a function of the young component fraction r, namely $(1-\alpha)^{-1}$, where α is the slope of the relations in Figure 6. We note that formally, $r = r(z)$, since the contamination due to a young component could be different depending on the couple of points used to estimate $H(z)$; the only assumption we make here is that r is constant between the two redshifts considered to provide one $H(z)$ measurement (that should, therefore, be really close in cosmic time to avoid evolution effects, as done in Moresco et al. 2012a, 2016b). The validity of this assumption can be directly tested on the data. We note here, however, that in the CC approach, one would more likely work comparing not individual galaxies (where this effect is actually maximized) but rather mean trends of populations very close in cosmic time and selected homogeneously, helping to keep this effect under control.

In the figure, we show $\epsilon(r)$ as a function of the age of the young component (0.1 and 1 Gyr) and the chosen range of normalization. As can be seen, there is a large variation depending on the considered parameters, but we find that a young component effectively acts on $H(z)$ as an additive systematic bias that can influence the measurement up to 20% if not properly taken into account, i.e., for a contribution $r \sim 0.4$. As shown, the impact due to the different assumed SPS model is, instead, negligible. It is therefore crucial to limit this effect, as done in Moresco et al. (2012a, 2016b), by selecting an uncontaminated (or minimally contaminated) sample.

The H/K values found in these works (shown in Figures 3 and 4) allow for a maximum value of $r = 0.009$–0.013 (at 1σ, depending on the age of the youngest population; $r = 0.023$–0.035 at 2σ), which is, however, a strong upper limit on the contamination level of our sample, since the same value of H/K can be reproduced by a purely passive population at a different age. This effect translates to a maximum variation $\Delta D_n4000 < 0.6\%$–1%, which, as a consequence, affects the Hubble parameter by at most 0.4%–1% (at 1σ, considering the full variation spanned by the various models shown in Figure 7; 0.8%–2.3% at 2σ). This is below current uncertainties on the derived values of $H(z)$ with this method. We also estimate that for this systematic effect to be <0.5%, the contamination by a young component should be $r \lesssim 0.005$–0.015 (depending on the model).

In general, for unrelated errors, the associated covariance matrix (Cov) for the Hubble parameter $H(z)$ is simply the sum of the covariance due to the various errors,

$$\mathrm{Cov}_{ij}^{\mathrm{tot}} = \mathrm{Cov}_{ij}^{\mathrm{stat}} + \mathrm{Cov}_{ij}^{\mathrm{young}} + \mathrm{Cov}_{ij}^{\mathrm{model}} + \mathrm{Cov}_{ij}^{\mathrm{met}}, \quad (7)$$

where "stat," "young," "model," and "met" denote the contributions to the covariance due to statistical errors, young component contamination, dependence on the chosen stellar population model, and metallicity, respectively. These error contributions will be studied in detail elsewhere. Here we just provide the equation for $\mathrm{Cov}_{\mathrm{ij}}^{\mathrm{young}}$. Following Equation (6), at redshift z_i, the systematic error on the measurement of the Hubble parameter due to a young component residual is $H_{\mathrm{true}}\,\epsilon\,(r_i)$; therefore, its associated covariance matrix will be

$$\mathrm{Cov}_{ij}^{\mathrm{young}} = H_{\mathrm{true}}(z_1) H_{\mathrm{true}}(z_2)\,\epsilon\,(r(z_1))\,\epsilon\,(r(z_2)). \quad (8)$$

The error on H arising from a possible young component contamination is highly correlated across redshifts. However, even in current analyses, it accounts for ~0.5%–1% correlated error, while the statistical error for each data point is 5% or larger; hence, it is a subdominant contribution.

6. Toward an Optimal Selection of CCs

Finally, based on all the results discussed so far, we provide a recipe to optimally select CCs for a D_n4000 analysis to minimize the possible effects due to an underlying young component and, therefore, maximize the robustness of the cosmological results. As discussed in Moresco et al. (2012a, 2016b), the selection

criterion is crucial in the D_n4000 approach to CCs to select the purest possible sample of massive and passively evolving galaxies. Many criteria have been combined based on both photometry and spectroscopy to obtain a pure sample of massive and passively evolving galaxies.[10] Here we summarize the main ones, also discussing the possible effect of not considering (possibly due to unavailability of data) one of those. The selection workflow is outlined in Figure 8.

The fundamental steps to select the best CC candidates are as follows.

1. *Selection based on UV emission.* As discussed in Section 4.3, having UV coverage gives a first photometric indication of whether there could be a young component, which cannot be visible and distinguishable in optical photometric bands. A selection based on UV colors, such as the NUVrJ color–color diagram (Ilbert et al. 2013), can detect the presence (even at low levels) of a young population from a few hundred Myr up to ∼1 Gyr; therefore, galaxies not matching this criterion should be rejected.
2. *Check for the presence of emission lines.* As discussed in Section 4.2, the presence of Hα and [O II] emission lines is a spectroscopic indicator of ongoing star formation, due to the contribution of a young stellar population with ages ∼10–100 Myr. While different criteria can be adopted to minimize this contamination, based on either EW values (EW([O II]) < 5 Å; see Mignoli et al. 2009) or the S/N of these emission lines ($S/N_{H\alpha} < 3$; see Wang et al. 2018), only spectra with no detectable emission lines should be considered. It is fundamental to reject galaxies showing any evidence of emission lines.
3. *Check for the presence of strong Hδ absorption lines.* Complementary to item (2), a cut based on the presence of Hδ absorption lines (EW(Hδ) $>$ 4–5 Å; Le Borgne et al. 2006; Bezanson et al. 2013; Wilkinson et al. 2017) is fundamental to excluding galaxies that have experienced a recent episode of star formation, yielding a young stellar component of ∼0.5–1 Gyr (post-starburst galaxies). Galaxies with values above that threshold should therefore be excluded by the selection.
4. *Check for the Ca II H/K value.* In Section 4.1, we demonstrated how the Ca II H/K ratio is a powerful tool to diagnose the presence of an underlying young component. A check of the Ca II H/K value found in the data should allow the assessment of the corresponding level of contamination in the data and therefore the corresponding contribution to the total error induced on $H(z)$ following the recipes provided. In particular, we show that galaxies that present an inversion of this indicator (H/K < 1) are characterized by the presence of a young component (with ages $\lesssim$200 Myr) with a percentage $r \gtrsim 0.05$, and that values of H/K > 1.1 indicates a contamination $r \lesssim 0.01$.
5. *Stellar mass/velocity dispersion selection.* Finally, it is fundamental to select, among the previously selected passive galaxies, the most massive ones. In particular, in Thomas et al. (2010), it was shown not only that with increasing mass, the formation of galaxies dates progressively back in the past (therefore ensuring selection of the oldest objects at each redshift), but also that their SFH becomes more and more synchronized, making them the ideal CCs. Moreover, in Moresco et al. (2013), it was also shown how the mass cut increases the purity of the sample, since contamination is more important at smaller masses. Both a cut in stellar mass and velocity dispersion can be applied to select the most massive objects, where the second one, being a direct observable, is less dependent on model assumptions. A typical adopted cut in stellar mass is $\log(M/M_{\odot}) > 10.75$–11, roughly corresponding to a cut in velocity dispersion $v_{\rm disp} > 250$–300 km s^{-1}.

The combination of the H/K diagnostic with as many independent observables as possible allows one to avoid statistical outliers in the selection and minimize the contamination. At last, the selected spectra should be analyzed with a full spectral fitting approach like MOPED/VESPA or similar approaches (e.g., Chevallard & Charlot 2016; Citro et al. 2016). This step represents a final cross-check that will be applied on a sample that passes the selection criteria discussed above to detect possible further signs of star formation that have not been detected by other observables. In this way, the selection will rely as little as possible on modeling and fitting and as much as possible on observables, providing as a consequence a Hubble parameter measurement that is less model-dependent. As discussed in Section 4, with enough S/N in the data, this kind of analysis is able not only to provide constraints on the metallicity of the galaxy needed to apply the CC method but also to disentangle the galaxy SFH and identify the presence of further star formation episodes that were not excluded with the previous criteria (similar results have also been obtained with independent approaches; see, e.g., Citro et al. 2016). This approach maximizes the purity of the CC candidates, providing a suitable sample to constrain the expansion history of the universe.

As discussed before, some information may be unavailable; however, there is some overlap between the criteria, in particular between criteria (1), (2), and (4). On one hand, the absence of Ca II H/K inversion limits the possible level of contribution of a young component also in the UV; on the other hand, the absence of significant emission lines is directly linked to the underlying absence of young stars. A combination of criteria 1, 3, 4, and 5 or 2, 3, 4, and 5 could be, therefore, sufficient to select CCs. However, redundancy offers a more reliable selection. Selection criteria can (and must) be made most stringent for future data with increased statistics.

The selection limits discussed above are clearly dependent on the resolving power and S/N of the spectrum. The most critical features to be measured are Ca II H and K lines, and we find that, in order to detect them, medium-to-low resolving power is needed, and they have currently been detected down to $R \gtrsim 230$ (with the VIMOS low-resolution instrument) as, e.g., can be inferred from the analysis of VIPERS passive galaxies shown by Garilli et al. (2014). As for the S/N, Ca II H and K typically present an amplitude $F_{\rm max}/F_{\rm min} \sim 3$, so in order to detect them, an S/N ∼ 9 is required, which is also the typical S/N needed for a correct measurement of absorption features (see Cappellari et al. 2009). This value represents a detecting threshold that allows the Ca II H and K lines to be measurable; then the associated error (which, e.g., could be

[10] In Moresco et al. (2013), it is shown that this combination of criteria, while reducing the sample size by ∼70% with respect to a simple color cut, also significantly reduces the level of contamination.

estimated with a Monte Carlo approach) will depend on the actual S/N of the spectrum. Once it has been measured, the contamination (and its corresponding uncertainty) of a sample can be obtained from Figure 7, and it could be propagated in the error analysis with Equation (8).

The workflow discussed above is meant to be applied to individual galaxies, and as a result, the selection of galaxies without evidence of star formation (see, e.g., criterion (2)) will depend on the properties and depth of the survey. However, the procedure can be further improved by considering stacked spectra (as done in, e.g., Moresco et al. 2012a, 2016b) that will allow one to significantly increase the S/N of the spectrum and the detectability of spectral features. As an example, the stacking approach applied on BOSS data in Moresco et al. (2016b) allowed the increase of the S/N of the spectrum by a factor between 10 and 60, demonstrating the robustness of the selection by showing no evidence of emission lines even in the high-S/N stacked spectra. A more complete analysis of the detectability of features as a function of the quality of the spectrum will be addressed in a later work.

7. Conclusions

Motivated by the upcoming massive spectroscopic surveys that will significantly improve the statistics of massive and passive galaxies (especially at $z > 1$), we have discussed how to select CCs from such surveys. Because the spectra will be mostly in the optical rest frame, we have focused on this region of the spectrum.

The main systematics that could potentially affect the CC method are the dependence on the stellar metallicity estimate, the reliance on SPS models, the progenitor bias, and the presence of an underlying young component. In this analysis, we quantify the impact of a possible contamination due to a young component in two steps. Most importantly, we provide the contribution to the covariance matrix due to this effect to be taken into account in the total error budget as a function of the contamination level. This is one of the novel contributions of this paper. Moreover, we provide new spectral indicators and illustrate a clear selection workflow with which it is possible to minimize this contamination. This is also novel. While some of the individual steps were already discussed in the literature, their combination and integration were not and are presented here for the first time. Our main results are summarized as follows.

1. We identified the Ca II H/K indicator as a powerful diagnostic to identify the presence of a young underlying component. This ratio, which in galaxies dominated by passive evolution is >1, already shows an inversion for a contribution of a young component (with ages <200 Myr) of $\gtrsim$5%. Moreover, the analysis of previous data (Moresco et al. 2012a, 2016b) provided constraints of a contamination of $\lesssim$1%.
2. We analyzed the expected emission and absorption lines due to the presence of a young component, finding that it should present significant Hα and [O II] emission at young ages ($\lesssim$10 Myr) or significant Hδ absorption for a less recent star formation episode (hundreds of Myr up to $\sim$1 Gyr). The presence of a young component could also contribute to the rise of the UV flux, but in this case, we estimated it to be significant (with fluxes comparable to the ones in the optical ranges) for ratios $\gtrsim$40%; lower UV fluxes can be instead due either to a smaller young contribution or to effects due to dust extinction. In this case, however, the NUVrJ diagram selection helps in disentangling possible age–extinction degeneracies.
3. We estimated the impact of a young component with different ratios r with respect to the old one, on the D_n4000, finding a percentage variation of $\lesssim$5% when $r < 0.05$ (distinguishable from the H/K inversion) up to $\gtrsim$40% when $r \gtrsim 0.4$.
4. Finally, we propagated the previous effect, assessing its impact on the measurement of $H(z)$, providing a relation linking the systematic bias in $H(z)$ due to the presence of a young component to its percentage contribution r.

In particular, we demonstrated that the combination of these various indicators can be used to accurately discriminate the contribution of an underlying young population spanning a wide range of ages (0.01–1 Gyr). We also provide an explicit workflow to reliably select the best candidates for CCs. After this initial selection, excluding very young or localized star formation episodes, the resulting sample should undergo a further selection by mass and via a full spectral fitting to determine the (smooth) SFH and the metallicity before proceeding with a D_n4000 analysis.

To assess the impact of the above effect on our previous analysis, we have reanalyzed the data presented in Moresco et al. (2012a, 2016b) and show that indeed, the young population contamination is minimal and consistent with zero given current uncertainties. We have calculated that at most, it would bias the $H(z)$ determination by 0.4%–1% (at 1σ; 0.8%–2.3% at 2σ), well below the current errors. The fact that the diagnostic features are in the rest-frame optical spectrum opens up the possibility of using future high-resolution spectroscopic surveys to yield an optimal "golden" sample of CCs. Such a sample can, in principle, provide constraints on $H(z)$ an order of magnitude better than current ones, provided that systematic errors can be kept under control. Here we have presented a checklist and procedure to help minimize the effects due to a young stellar population contribution.

A detailed study of the other sources of systematics and how to minimize them has been presented in Moresco et al. (2012a, 2016b); let us just recall here that with massive spectroscopic surveys, the dominating effect is the one due to the stellar metallicity estimate, and an analysis of how to reduce it will follow. This opens up the exciting possibility of constructing a sample of CCs where $H(z)$ can be measured at the percent level over the past 10 Gyr of cosmic evolution ($z \sim 2$), thus testing the current ΛCDM paradigm in a cosmology model–independent way and searching for new physics.

We thank Adam Riess for useful comments and discussion. M.M., A. Cimatti, L.P., and A. Citro acknowledge the grants ASI n.I/023/12/0 "Attività relative alla fase B2/C per la missione Euclid" and PRIN MIUR 2015 "Cosmology and Fundamental Physics: illuminating the Dark Universe with Euclid". We thank the anonymous referee for the comments that helped to improve the presentation and clarify the discussion. Funding for this work was partially provided by the Spanish MINECO under projects AYA2014-58747-P and MDM-2014-0369 of ICCUB (Unidad de Excelencia María de Maeztu'). L.V. acknowledges support of European Unions Horizon 2020 research and innovation program me ERC (BePreSySe, grant agreement 725327).

Software: BC16 (Bruzual & Charlot 2003), M11 (Maraston & Strömbäck 2011), Vazdekis et al. (Vazdekis et al. 2016), CLOUDY (Ferland et al. 1998, 2013), MOPED/VESPA (Heavens et al. 2004; Tojeiro et al. 2007).

ORCID iDs

Michele Moresco https://orcid.org/0000-0002-7616-7136
Raul Jimenez https://orcid.org/0000-0002-3370-3103
Licia Verde https://orcid.org/0000-0003-2601-8770
Lucia Pozzetti https://orcid.org/0000-0001-7085-0412
Andrea Cimatti https://orcid.org/0000-0002-4409-5633
Annalisa Citro https://orcid.org/0000-0003-3948-6688

References

Arnouts, S., Le Floc'h, E., Chevallard, J., et al. 2013, A&A, 558, A67
Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. 2007, A&A, 476, 137
Asplund, M., Grevesse, N., & Sauval, A. J. 2005, in ASP Conf. Ser. 336, Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis, ed. T. G. Barnes, III & F. N. Bash (San Francisco, CA: ASP), 25
Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., & Ellingson, E. 1999, ApJ, 527, 54
Bernal, J. L., Verde, L., & Riess, A. G. 2016, JCAP, 10, 019
Bezanson, R., van Dokkum, P., van de Sande, J., Franx, M., & Kriek, M. 2013, ApJL, 764, L8
Boissier, S., Cucciati, O., Boselli, A., Mei, S., & Ferrarese, L. 2018, A&A, 611, A42
Bruzual, A. G. 1983, ApJ, 273, 105
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Butcher, H., Wells, D. C., & Oemler, A., Jr. 1983, ApJS, 52, 183
Cappellari, M., di Serego Alighieri, S., Cimatti, A., et al. 2009, ApJL, 704, L34
Cappellari, M., & Emsellem, E. 2004, PASP, 116, 138
Chabrier, G. 2003, PASP, 115, 763
Chen, Y.-M., Kauffmann, G., Tremonti, C. A., et al. 2012, MNRAS, 421, 314
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95
Citro, A., Pozzetti, L., Moresco, M., & Cimatti, A. 2016, A&A, 592, A19
Citro, A., Pozzetti, L., Quai, S., et al. 2017, MNRAS, 469, 3108
Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., & Calzetti, D. 1995, AJ, 110, 1071
Cowie, L. L., Songaila, A., & Barger, A. J. 1999, AJ, 118, 603
Crawford, S. M., Ratsimbazafy, A. L., Cress, C. M., et al. 2010, MNRAS, 406, 2569
Daly, R. A., Djorgovski, S. G., Freeman, K. A., et al. 2008, ApJ, 677, 1
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
Diaferio, A., Ostorero, L., & Cardone, V. 2011, JCAP, 10, 008
Donahue, M., Bruch, S., Wang, E., et al. 2010, ApJ, 715, 881
Dopita, M. A., Fischera, J., Sutherland, R. S., et al. 2006, ApJS, 167, 177
Dopita, M. A., Kewley, L. J., Heisler, C. A., & Sutherland, R. S. 2000, ApJ, 542, 224
Dunlop, J., Peacock, J., Spinrad, H., et al. 1996, Natur, 381, 581
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Garilli, B., Guzzo, L., Scodeggio, M., et al. 2014, A&A, 562, A23
Greggio, L., & Renzini, A. 1990, ApJ, 364, 35
Hamilton, D. 1985, ApJ, 297, 371
Han, Z., Podsiadlowski, P., & Lynas-Gray, A. E. 2007, MNRAS, 380, 1098
Heavens, A., Panter, B., Jimenez, R., & Dunlop, J. 2004, Natur, 428, 625
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Jimenez, R., Jorgensen, U. G., Thejll, P., & MacDonald, J. 1995, MNRAS, 275, 1245
Jimenez, R., & Loeb, A. 2002, ApJ, 573, 37
Jimenez, R., Verde, L., Treu, T., & Stern, D. 2003, ApJ, 593, 622
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Lavaux, G., & Wandelt, B. D. 2012, ApJ, 754, 109
Le Borgne, D., Abraham, R., Daniel, K., et al. 2006, ApJ, 642, 48
Leonardi, A. J., & Rose, J. A. 1996, AJ, 111, 182
Levesque, E. M., Kewley, L. J., & Larson, K. L. 2010, AJ, 139, 712
Lilly, S. J., & Gunn, J. E. 1985, MNRAS, 217, 551
Liu, G. C., Lu, Y. J., Xie, L. Z., Chen, X. L., & Zhao, Y. H. 2016, A&A, 585, A52
Magris, C. G., Binette, L., & Bruzual, A. G. 2003, ApJS, 149, 313
Maraston, C., Daddi, E., Renzini, A., et al. 2006, ApJ, 652, 85
Maraston, C., & Strömbäck, G. 2011, MNRAS, 418, 2785
Marchetti, A., Granett, B. R., Guzzo, L., et al. 2013, MNRAS, 428, 1424
Marín-Franch, A., Aparicio, A., Piotto, G., et al. 2009, ApJ, 694, 1498
Martin, D. C., Wyder, T. K., Schiminovich, D., et al. 2007, ApJS, 173, 342
Mignoli, M., Zamorani, G., Scodeggio, M., et al. 2009, A&A, 493, 39
Moresco, M. 2015, MNRAS, 450, L16
Moresco, M., Cimatti, A., Jimenez, R., et al. 2012a, JCAP, 8, 006
Moresco, M., Jimenez, R., Verde, L., et al. 2016a, JCAP, 12, 039
Moresco, M., Pozzetti, L., Cimatti, A., et al. 2013, A&A, 558, A61
Moresco, M., Pozzetti, L., Cimatti, A., et al. 2016b, JCAP, 5, 014
Moresco, M., Verde, L., Pozzetti, L., Jimenez, R., & Cimatti, A. 2012b, JCAP, 7, 053
Moustakas, J., Kennicutt, R. C., Jr., & Tremonti, C. A. 2006, ApJ, 642, 775
Moustakas, J., Kennicutt, R. C., Jr., Tremonti, C. A., et al. 2010, ApJS, 190, 233
Moutard, T., Arnouts, S., Ilbert, O., et al. 2016, A&A, 590, A103
O'Connell, R. W. 1988, in Towards Understanding Galaxies at Large Redshift, Vol. 141 ed. R. G. Kron & A. Renzini (Dordrecht: Kluwer), 177
Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, ApJ, 808, 161
Ownsworth, J. R., Conselice, C. J., Mundy, C. J., et al. 2016, MNRAS, 461, 1112
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, MNRAS, 378, 1550
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Poggianti, B. M., & Barbaro, G. 1997, A&A, 325, 1025
Reichardt, C., Jimenez, R., & Heavens, A. F. 2001, MNRAS, 327, 849
Riess, A. G., Casertano, S., Yuan, W., et al. 2018, ApJ, 855, 136
Rose, J. A. 1984, AJ, 89, 1238
Rose, J. A. 1985, AJ, 90, 1927
Sánchez Almeida, J., Terlevich, R., Terlevich, E., Cid Fernandes, R., & Morales-Luis, A. B. 2012, ApJ, 756, 163
Simon, J., Verde, L., & Jimenez, R. 2005, PhRvD, 71, 123001
Spinrad, H., Dey, A., Stern, D., et al. 1997, ApJ, 484, 581
Stern, D., Jimenez, R., Verde, L., Kamionkowski, M., & Stanford, S. A. 2010, JCAP, 2, 008
Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, ApJ, 621, 673
Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, MNRAS, 404, 1775
Tinsley, B. M. 1968, ApJ, 151, 547
Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, MNRAS, 381, 1252
Treu, T., Ellis, R. S., Liao, T. X., et al. 2005, ApJ, 633, 174
Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, MNRAS, 463, 3409
Verde, L., Protopapas, P., & Jimenez, R. 2013, PDU, 2, 166
Walcher, C. J., Coelho, P. R. T., Gallazzi, A., et al. 2015, A&A, 582, A46
Wang, L.-L., Luo, A.-L., Shen, S.-Y., et al. 2018, MNRAS, 474, 1873
Wild, V., Kauffmann, G., Heckman, T., et al. 2007, MNRAS, 381, 543
Wilkinson, C. L., Pimbblet, K. A., & Stott, J. P. 2017, MNRAS, 472, 1447
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Yi, S. K., Lee, J., Sheen, Y.-K., et al. 2011, ApJS, 195, 22
Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004, AJ, 128, 2603
Yip, C.-W., Mahoney, M. W., Szalay, A. S., et al. 2014, AJ, 147, 110
Zhang, C., Zhang, H., Yuan, S., et al. 2014, RAA, 14, 1221