A&A 561, A130 (2014)
DOI: 10.1051/0004-6361/201321692

Astronomy & Astrophysics

Resolving galaxies in time and space

II. Uncertainties in the spectral synthesis of datacubes

R. Cid Fernandes[1,2], R. M. González Delgado[2], R. García Benito[2], E. Pérez[2], A. L. de Amorim[1,2], S. F. Sánchez[2,3], B. Husemann[4], J. Falcón Barroso[5,6], R. López-Fernández[2], P. Sánchez-Blázquez[7], N. Vale Asari[1], A. Vazdekis[5,6], C. J. Walcher[4], and D. Mast[2,3]

[1] Departamento de Física, Universidade Federal de Santa Catarina, PO Box 476, 88040-900 Florianópolis, SC, Brazil
e-mail: cid@astro.ufsc.br
[2] Instituto de Astrofísica de Andalucía (CSIC), PO Box 3004, 18080 Granada, Spain
[3] Centro Astronómico Hispano Alemán, Calar Alto, (CSIC-MPG), C/Jesús Durbán Remón 2-2, 04004 Almería, Spain
[4] Leibniz-Institut für Astrophysik Potsdam, innoFSPEC Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany
[5] Instituto de Astrofísica de Canarias, vía Lactea s/n, 38200 La Laguna, Tenerife, Spain
[6] Departamento de Astrofísica, Universidad de La Laguna, 38205 Tenerife, Spain
[7] Departamento de Física Teórica, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Received 13 April 2013 / Accepted 28 June 2013

ABSTRACT

Aims. In a companion paper we have presented many products derived from the application of the spectral synthesis code STARLIGHT to datacubes from the CALIFA survey, including 2D maps of stellar population properties (such as mean ages, mass, and extinction) and 1D averages in the temporal and spatial dimensions. Our goal here is to assess the uncertainties in these products.
Methods. Uncertainties associated to noise and spectral shape calibration errors in the data and to the synthesis method were investigated by means of a suite of simulations, perturbing spectra and processing them through our analysis pipelines. The simulations used 1638 CALIFA spectra for NGC 2916, with perturbation amplitudes gauged in terms of the expected errors. A separate study was conducted to assess uncertainties related to the choice of evolutionary synthesis models, the key ingredient in the translation of spectroscopic information into stellar population properties. We compare the results obtained with three different sets of models: the traditional Bruzual & Charlot models, a preliminary update of them, and a combination of spectra derived from the Granada and MILES models. About 10^5 spectra from over 100 CALIFA galaxies were used in this comparison.
Results. Noise and shape-related errors at the level expected for CALIFA propagate to uncertainties of 0.10−0.15 dex in stellar masses, mean ages, and metallicities. Uncertainties in A_V increase from 0.06 mag for random noise to 0.16 mag for spectral shape errors. Higher-order products such as star formation histories are more uncertain than global properties, but still relatively stable. Owing to the large number statistics of datacubes, spatial averaging reduces uncertainties while preserving information on the history and structure of stellar populations. Radial profiles of global properties, and star formation histories averaged over different regions are much more stable than those obtained for individual spaxels. Uncertainties related to the choice of base models are larger than those associated with data and method. Except for metallicities, which come out very different when fits are performed with the Bruzual & Charlot models, differences in mean age, mass, and metallicity are of the order of 0.15 to 0.25 dex, and 0.1 mag for A_V. Spectral residuals are of the order of 1% on average, but with systematic features of up to 4% amplitude. We discuss the origin of these features, most of which are present in both in CALIFA and SDSS spectra.

Key words. techniques: imaging spectroscopy – galaxies: evolution – galaxies: stellar content

1. Introduction

Large optical spectroscopic surveys have dominated the panorama of extra-galactic astrophysics over the last decade. Stellar population analysis of these data has produced information on stellar masses, mean stellar ages, abundances, dust content, and diagnostics of their star formation histories (SFHs), whose analysis per se and in relation to other properties (for instance, morphology or environment) has advanced the understanding of the physics of galaxies and their evolution through cosmic time. Yet, all this information is derived from spatially unresolved spectra, and thus reflects global averages over properties that are known to vary accross the face of galaxies. Integral field spectroscopy (IFS) is a natural next step to move from global properties to the internal physics of galaxies. Dissecting the different structural components will probably lead to a much more complete picture of galaxies, and to a clearer understanding of the results previously obtained in non-spatially resolved surveys. IFS has been around for over two decades, but has only recently been promoted to the scale of modern galaxy surveys. CALIFA, the Calar Alto Integral Field Area survey (Sánchez et al. 2012; Husemann et al. 2013), is a pioneer in this blooming field.

To harvest the full potential of IFS data one needs to adapt techniques developed for spatially unresolved galaxy spectra, folding in the information on the spatial structure. In a companion article (Cid Fernandes et al. 2013, hereafter Paper I) we have presented our method for extracting stellar population

properties out of CALIFA datacubes by means of the STARLIGHT spectral-fitting software (Cid Fernandes et al. 2005). The packages developed to pre-process and partition the reduced datacubes (QBICK) and to post-process the results of the spectral synthesis (PyCASSO, the Python CALIFA Starlight Synthesis Organizer) were described in detail. The products of this work flow were examplified using data for the nearby spiral NGC 2916 (CALIFA 277). Diagnostics such as 2D maps of physical properties (mean ages and metallicities, mass surface densities, star formation rates, etc.), 1D averages in the temporal and spatial dimensions, projections of the stellar light and mass growth onto radius-age diagrams, and cuts through the $(x, y,$ age) cubes were presented to illustrate the potential of the combination of IFS data with spectral synthesis as a tool for studying galaxy evolution in time and space simultaneously.

Uncertainties in this whole process stem from *(i)* noise and calibration of the data; *(ii)* limitations of our spectral synthesis method; and *(iii)* evolutionary synthesis models, the key ingredient in the mapping of observed spectra to astrophysical information. These different sources of uncertainty, each one comprising a field of work by itself, intertwine and propagate in complex ways through our pipelines. Uncertainties in spectral synthesis products have been previously explored in the literature, such as the noise-effect simulations in Cid Fernandes et al. (2005), the model-impact study of Panter el al. (2007), and the suite of tests presented by Tojeiro et al. (2007), but never in the context of IFS.

This article complements Paper I by addressing the effects of these uncertainties. Our central goal is to provide quantitative and qualitative guidelines to allow an assessment of the robustness of results produced with the methods and tools explained in Paper I, which are being used in a series of publications by our collaboration. Pérez et al. (2013), for instance, analyzed the spatially resolved mass assembly history of 105 CALIFA galaxies, finding clear evidence for an inside-out growth for massive galaxies, and signs of outside-in growth for the less massive ones (see González Delgado et al. 2012 and 2013 for other results). Despite the focus on STARLIGHT and CALIFA, the results reported here are also useful in the broader context of stellar population analysis of IFS data; for instance, experimental datacubes recently obtained by MaNGA run through our QBICK and PyCASSO pipelines without any modification.

The first part of this paper (Sects. 2 and 3) presents extensive simulations that explore the effects of uncertainties related to the data and spectral synthesis method. Random noise and shape-changing perturbations are added to both observed and synthetic spectra of all zones of the same galaxy analyzed in Paper I. These simulations are then used to evaluate uncertainties in global properties (mean ages, masses, etc.) and higher-order products such as SFHs. This is done both at the level of individual spectra and for spatial regions cut out from the datacube.

The second part (Sect. 4) is dedicated to experiments carried out to evaluate how our results are affected by different choices of evolutionary synthesis models for the spectra of simple stellar populations (SSP). This study uses ~100 thousand spectra from 107 galaxies observed by CALIFA, all of which were processed with three sets of SSP models used in the current literature. An inspection of the quality of the spectral fits and a comparison with results obtained for SDSS galaxies is also presented.

Section 5 summarizes our main results. Throughout this paper the notation and definitions established in Paper I are followed.

2. Uncertainties associated to the data and the method: description of the simulations

A drawback of STARLIGHT is that it provides no error estimate on its output. The most straightforward (albeit computationally expensive) way to do so is to run the code several times perturbing the input data according to some realistic prescription of the errors involved. This section describes a set of extensive simulations taylored to match the characteristics of the CALIFA data. The simulations are designed to address the effects of

1. random noise,
2. shape-related calibration uncertainties, and
3. multiplicity of solutions,

upon the physical properties derived with STARLIGHT. Throughout these experiments we keep the set of SSPs fixed at the same set as used in Paper I, which is denoted "base GM" in Sect. 4, where the effects of this choice are investigated. This base comprises $N_\star = 156$ elements with 39 ages between $t = 1$ Myr and 14 Gyr, and 4 metallicities from $Z = 0.2$ to 1.5 solar. The simulations are based on the data and STARLIGHT fits of 1638 zone spectra[1] of CALIFA 277 studied in Paper I. Given the variety of properties found accross this galaxy and the fact that all other datacubes in CALIFA were processed in the same way, the uncertainty estimates presented here are representative of CALIFA data in general.

Each spectrum was perturbed ten times for each of the R, C, and E error prescriptions discussed below and refitted. This data set is then used to evaluate uncertainties in the derived properties, both at the level of individual zones and for spatially averaged properties like those shown in Figs. 10 and 11 of Paper I. Observed as well as synthetic spectra were perturbed in the simulations. The first represent real-life conditions, while the latter are essentially a theoretical exercise which explores the sensitivity of the STARLIGHT results to different types of perturbations.

2.1. Types of errors and simulations

The first type of error considered is that related to the variations on the derived properties induced by random fluctuations of the observed fluxes within the noise level. To study this, each zone spectrum was perturbed ten times with Gaussian noise with a λ-dependent amplitude given by the error spectrum ϵ_λ (hereafter OR1 runs, with O standing for observed, R for random, and 1 for the amplitude of the perturbation). Experiments were also conducted with pertubations of 2 and $3 \times \epsilon_\lambda$ (OR2 and OR3 runs, respectively) to emulate what would be obtained under conditions considerably worse than those in CALIFA[2].

A second set of simulations was carried to address the problem of the overall continuum shape calibration. STARLIGHT relies on flux-calibrated input spectra, so it is relevant to access how much its solutions vary for spectra whose shapes vary within expected uncertainty ranges. Husseman et al. (2013) found that CALIFA synthetic $g - r$ colors typically differ by 0.05 mag from

[1] 1527 of the spectra come from single spaxels, while the remaining 111 correspond to Voronoi zones that typically comprise ten spaxels.

[2] Note that these simulations add noise to observed spectra, which already contains noise. Hence, OR1 spectra, which are nominaly meant to represent variations at the level of $1\epsilon_\lambda$, effectively have errors of $\sqrt{2}\epsilon_\lambda$, i.e., 41% higher than intended. We nevertheless pretend these runs correspond to 1σ perturbations, which gives a safety margin in our analysis. Similarly, the OR2 and OR3 runs have errors 2.23 and 3.16 times larger than the original ones instead of factors of 2 and 3, respectively.

those obtained by SDSS photometry. We used this value to gauge the effects of continumm calibration uncertainty in each individual specrum of CALIFA 277. For simplicity, but with no loss of generality, these simulations keep the flux at $\lambda_r = 6231$ Å stable, such that the 0.05 mag variation in $g-r$ translates into a 0.02 dex uncertainty in the flux at $\lambda_g = 4770$ Å. To emulate the effects of such variations we produced ten versions of each spectrum by adding $A_g \times (\lambda_r - \lambda)/(\lambda_r - \lambda_g)$ to the original $\log F_\lambda$, with $A_g = \Delta \log F_{\lambda_g}$ sampled from a Gaussian distribution of zero mean and dispersion $\sigma(\log F_{\lambda 4470}) = 0.02$ (OC002 runs, where C stands for color). This produces both redder and bluer versions of the original spectrum. Experiments were also carried out for $\sigma(\log F_{\lambda 4470}) = 0.01$, and 0.04 dex (OC001 and OC004, respectively). We take OC002 runs as default for CALIFA, even though the spaxel-to-spaxel variations in the continuum calibration for a single galaxy are probably smaller than the $g - r$ dispersion quoted by Husemann et al. (2013), which was derived from the 100 galaxies in the CALIFA DR1 sample.

Finally, fits were also conducted without any perturbation at all (OE0 runs), the motivation for which is to quantify a qualitatively different source of uncertainty embedded in all other simulations. STARLIGHT provides a single best-fit set of parameters out of typically many millions tried out during its likelihood-guided sampling of the parameter space. This single solution, however, is not meant to be mathematically unique. In fact, given the pseudo-random nature of its Markov chains, STARLIGHT solutions depend on the input seed for the random-number generator. For large samples, and each CALIFA datacube is a large sample in itself, these variations have no relevant effect on the overall results of the analysis. For individual spectra, however, the multiplicity of solutions leads to uncertainties in the derived properties. The OE0 runs quantify these uncertainties, which ultimately reflect an intrinsic limitation of our spectral synthesis method.

2.2. Simulations based on synthetic spectra

The OE0, OR*, and OC* runs work upon the original observed spectra (hence the prefix O). A paralell set of simulations was carried out applying these same perturbation recipes to the synthetic spectra obtained in our STARLIGHT fits of CALIFA 277. We label these runs as SE0, SR1, SR2, SR3, SC001, SC002, and SC004 (S for synthetic).

The S runs use exactly the same sampling, masks, flags, and ϵ_λ errors as in the O runs, thus following the pattern of CALIFA data in a realistic fashion. At the same time, these are idealistic simulations in the sense that the perturbations act upon perfect input spectra instead of on observed ones. This theoretical exercise therefore overlooks inadequacies in the SSP models (including problems such as potential flux calibration problems in the libraries, limited coverage of the stellar parameter space, missing or incorrectly modeled evolutionary phases) and other assumptions involved in the modeling (namely, extinction and kinematics).

The purpose of these runs is to evaluate to which extent STARLIGHT recovers known input properties in the presence of perturbations. This kind of input versus output comparison is more commonly carried out by feeding the spectral fitting code with data generated with ad hoc parametric descriptions of the SFH (e.g., the so-called tau models), while our theroretical SFHs are inspired in those derived from actual observations. A valid criticism of these simulations is that the same base is used to generate and fit spectra. The alternative would be to generate models with some other set of $t's$ and Z', but since our base is so large (39 ages and all four available metallicities), this would make very little difference in practice.

Table 1. Notation for the simulations.

Run	Original spectra	Type of perturbation	Perturbation amplitude	Comments
(1)	(2)	(3)	(4)	(5)
OR1	observed	noise	$1 \times \epsilon_\lambda$	Gaussian
OR2	observed	noise	$2 \times \epsilon_\lambda$	Gaussian
OR3	observed	noise	$3 \times \epsilon_\lambda$	Gaussian
OC001	observed	color	0.01 dex	at λ_g
OC002	observed	color	0.02 dex	at λ_g
OC004	observed	color	0.04 dex	at λ_g
OE0	observed	none	0	
SR1	synthetic	noise	$1 \times \epsilon_\lambda$	Gaussian
SR2	synthetic	noise	$2 \times \epsilon_\lambda$	Gaussian
SR3	synthetic	noise	$3 \times \epsilon_\lambda$	Gaussian
SC001	synthetic	color	0.01 dex	at λ_g
SC002	synthetic	color	0.02 dex	at λ_g
SC004	synthetic	color	0.04 dex	at λ_g
SE0	synthetic	none	0	

Notes. (1) Nomenclature adopted to denote the different simulations. (2) The nature (observed or synthetic) of the spectra that are perturbed. (3) Type of perturbation. (4) Amplitude of the perturbations. For C runs, we list the amplitude of the variations at λ_g. Each of the 1638 spectra from CALIFA 277 were perturbed ten times according to each of these 14 recipes.

2.3. Notation

For clarity and quick reference, we summarize the notation and intended purpose of the simulations described above.

R runs (OR1, OR2, OR3, SR1, SR2, and SR3) explore the effect of noise by adding Gaussian perturbations scaled to have amplitudes 1, 2, or 3 times the nominal error ϵ_λ. C runs perturb the input spectra in their global shape by multiplying the original F_λ by a power law in λ, scaled such that the flux at λ_r remains fixed while that at λ_g differs by (on average over all realizations) 0.01 (OC001 and SC001), 0.02 (OC002 and SC002) and 0.04 (OC004 and SC004) dex with respect to the input flux. (In a $\log F_\lambda \times \lambda$ plot, these shape-pertubations correspond to straight lines anchored at λ_r and different slopes.) These runs are meant to emulate the effect of a flux calibration uncertainty in the data. E0 runs do nothing to the input spectra. They only differ in the random number seed feed to STARLIGHT. The variance among these fits reflects the multiplicity of SSP combinations which lead to a similar total spectrum.

In all cases, a prefix O indicates that the perturbations act upon the observed spectra, while an S prefix means that synthetic spectra are perturbed. Table 1 summarizes the simulation-related nomenclature used thoughout the paper.

3. Results of the simulations

The simulations described in Sect. 2 provide the material to evaluate uncertainties associated with the data and spectral synthesis method. We split the presentation of this material into an assessement of the errors in global properties (Sect. 3.1), radial profiles (Sect. 3.2), and SFHs (Sect. 3.3). We conclude this part of our investigation with a summary of the results of the simulations (Sect. 3.4).

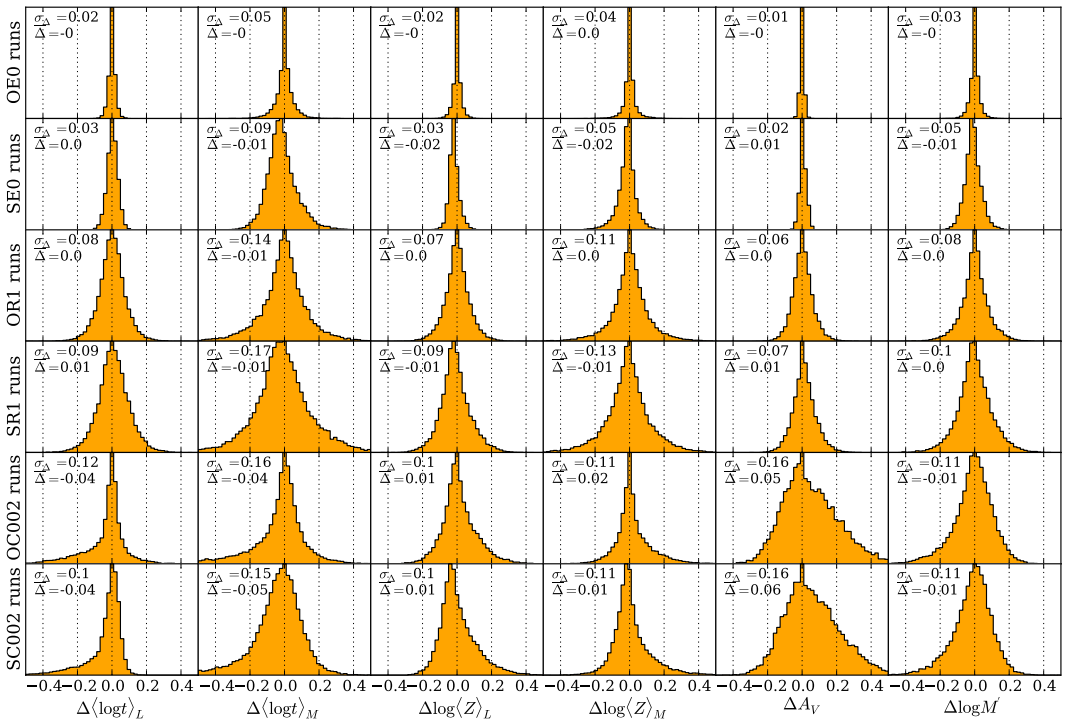

Fig. 1. Histograms of Δ (≡simulated minus original value) for the main scalar properties derived from the spectral synthesis: $\langle \log t \rangle_L$, $\langle \log t \rangle_M$, $\log\langle Z \rangle_L$, $\log\langle Z \rangle_M$, A_V and the initial stellar mass formed ($\log M'$). Each row corresponds to one set of simulations, alternating S and O runs (as labeled in the y-axis). Mean Δs and their σs are labeled in each panel. Except for the A_V column, each histogram is based on 16 380 values obtained from the 1638 zones in CALIFA 277 and ten Monte Carlo realizations of the perturbation. For clarity, the A_V histograms were drawn excluding the many $\Delta A_V = 0$ cases, which occur due to the $A_V \geq 0$ physical limit imposed on the spectral fits. Noise-free simulations (OE0 and SE0), which trace intrinsic degeneracies in spectral fits of composite stellar populations, are shown in the top two rows. Random-noise simulations (SR1 and OR1, central rows) map the uncertainties caused by statistical fluctuations of the observed F_λ fluxes. C002-simulations, where the spectral-shape is changed to emulate calibration uncertainties, are shown in *the bottom panels*.

3.1. Uncertainties in global properties

The main scalar properties derived from the spectral synthesis are mean ages and metallicities (weighted by light or mass), extinction and stellar masses[3]. These first order products of the analysis convey enough information to draw a quite detailed sketch of the 2D distribution of stellar population properties in a galaxy. Before incorporating the spatial arrangement of the spectra, however, it is useful to assess uncertainties in the individual spectra as if they came from different galaxies.

Figure 1 shows histograms of the Δ ≡ simulation − original value (i.e., that obtained from the unperturbed spectrum) of each of these quantities, and for six different sets of simulations: E0, R1, and C002 runs for the O and S versions. Statistics for all runs are listed in Table 2.

3.1.1. E runs: noise-free simulations

The OE0 and SE0 panels show that variations due to the random number seed fed to STARLIGHT have a tiny effect on the derived properties, with no bias ($\overline{\Delta} \sim 0$) and one σ uncertainty all in the second decimal place.

Intriguingly, SE0 runs have somewhat higher dispersions than OE0 runs. The mathematical reason for this behavior is subtle, but simple to grasp with the concepts developed by Pelat (1997, 1998). By construction, SE0 spectra lie within the synthetic domain (SD), i.e., the space generated by convex combinations of the base spectra[4]. A point inside this space can be exactly fitted by at least one set of parameters. Even if the solution is mathematically unique, a whole range of other solutions produce nearly identical spectra. Because of the limitations of evolutionary synthesis models, observed spectra cannot possibly be fitted exactly even in the ideal limit of perfect calibration and $S/N = \infty$ signal-to-noise ratio, so OE0 spectra are inevitably *outside* the SD. Fits of OE0 spectra therefore have access to a smaller subspace of acceptable solutions than fits of SE0 spectra, which results in the lower dispersions observed in Fig. 1. This same interpretation applies to other runs. Even when the synthetic spectra are perturbed (for instance, in the SR1 simulations), they remain closer to the SD than the corresponding O spectra, thus reaching a larger region of the solutions space.

3.1.2. R runs: random-noise simulations

Random noise at the nominal level (OR1 and SR1 runs, middle rows in Fig. 1) lead to Δ-distributions with σ around 0.1 dex in ages, metallicities and masses, and 0.05 mag in A_V. As expected (and as also seen in the E0 runs), mass-weighted ages and metallicities have broader Δ distributions than their luminosity-weighted counterparts. In any case, these dispersions are all

[3] Throughout this section we work with the initial stellar mass (M'), uncorrected for the mass expelled from stars back to the interstellar medium. Mass weighted ages and metallicities, however, do account for the returned mass.

[4] In the case of STARLIGHT and other full spectral fitting codes this space is additionally enlarged by the extinction and kinematical dimensions, but we leave this aside to simplify the argument.

Table 2. Statistics of the simulations: $\overline{\Delta} \pm \sigma_\Delta$.

Property	OE0	OR1	OR2	OR3	OC001	OC002	OC004
$\langle \log t \rangle_L$	–0.00 ± **0.02**	0.00 ± **0.08**	0.01 ± **0.13**	0.01 ± **0.16**	–0.00 ± **0.05**	–0.04 ± **0.12**	–0.09 ± **0.20**
$\langle \log t \rangle_M$	–0.00 ± **0.05**	–0.01 ± **0.14**	–0.02 ± **0.21**	–0.03 ± **0.26**	–0.00 ± **0.08**	–0.04 ± **0.16**	–0.10 ± **0.28**
$\log\langle Z \rangle_L$	–0.00 ± **0.02**	0.00 ± **0.07**	–0.00 ± **0.12**	–0.00 ± **0.16**	0.00 ± **0.05**	0.01 ± **0.10**	0.03 ± **0.16**
$\log\langle Z \rangle_M$	0.00 ± **0.04**	0.00 ± **0.11**	–0.00 ± **0.18**	0.00 ± **0.22**	0.00 ± **0.05**	0.02 ± **0.11**	0.05 ± **0.18**
A_V	–0.00 ± **0.01**	0.00 ± **0.06**	0.01 ± **0.10**	0.02 ± **0.14**	0.01 ± **0.06**	0.05 ± **0.16**	0.14 ± **0.29**
$\log M'$	–0.00 ± **0.03**	0.00 ± **0.08**	–0.00 ± **0.13**	0.00 ± **0.15**	–0.00 ± **0.05**	–0.01 ± **0.11**	–0.02 ± **0.21**
Property	SE0	SR1	SR2	SR3	SC001	SC002	SC004
$\langle \log t \rangle_L$	0.00 ± **0.03**	0.01 ± **0.09**	0.01 ± **0.14**	0.01 ± **0.18**	–0.01 ± **0.04**	–0.04 ± **0.10**	–0.10 ± **0.19**
$\langle \log t \rangle_M$	–0.01 ± **0.09**	–0.01 ± **0.17**	–0.02 ± **0.23**	–0.03 ± **0.27**	–0.02 ± **0.10**	–0.05 ± **0.15**	–0.11 ± **0.26**
$\log\langle Z \rangle_L$	–0.02 ± **0.03**	–0.01 ± **0.09**	–0.01 ± **0.14**	–0.00 ± **0.18**	–0.01 ± **0.05**	0.01 ± **0.10**	0.02 ± **0.17**
$\log\langle Z \rangle_M$	–0.02 ± **0.05**	–0.01 ± **0.13**	–0.00 ± **0.20**	0.01 ± **0.24**	–0.01 ± **0.06**	0.01 ± **0.11**	0.04 ± **0.19**
A_V	0.01 ± **0.02**	0.01 ± **0.07**	0.02 ± **0.11**	0.03 ± **0.15**	0.02 ± **0.06**	0.06 ± **0.16**	0.15 ± **0.30**
$\log M'$	–0.01 ± **0.05**	0.00 ± **0.10**	0.00 ± **0.14**	0.00 ± **0.16**	–0.01 ± **0.06**	–0.01 ± **0.11**	–0.03 ± **0.20**

Notes. For each global property the table lists the mean simulated minus original difference ($\overline{\Delta}$) and its standard deviation (σ_Δ). This last, printed in bold, indicates the level of uncertainties under different circumstances. Each column denotes one set of simulations. The upper half of the table shows results of simulations based on variations of the original observed spectrum (O runs), whereas the bottom half corresponds to runs where the synthetic spectra were perturbed (S runs) and used as input.

relatively small. Indeed, they are smaller than those induced by the use of different SSP models (Sect. 4, Fig. 11).

Table 2 reports results for simulations with noise levels 2 and 3 times worse than ϵ_λ. For the OR2 runs, for instance, the uncertainties in the properties ploted in Fig. 1 increase by roughly a factor of 2 with respect to those found in the OR1 simulations. The Δ distributions (not plotted) remain symmetric, without significant biases. We recall, however, that there is no reason to believe that the ϵ_λ values produced by the CALIFA reduction pipeline are underestimated. On the contrary, Husemann et al. (2013) showed that, if anything, they are slightly overestimated.

3.1.3. C runs: shape perturbations

The bottom panels in Fig. 1 show the results for simulations where the continuum shape was perturbed to produce (on average) 0.02 dex variations in the flux at λ_g for a fixed flux at λ_r. The one σ variations in ages, metallicities, and masses for OC002 runs are larger than those obtained with the OR1 runs, but not as large as in OR2 (Table 2). Similar conclusions apply to the corresponding S runs. Predictably, A_V is more severely affected by these global shape perturbations, with $\sigma(\Delta A_V) \sim 2.5$ times higher than those in the OR1 runs.

Unlike in other simulations, the Δ distributions are skewed, particularly those for A_V and $\langle \log t \rangle_L$. This is a by-product of the $A_V \geq 0$ limit imposed upon the fits. 582 of the 1638 zones of CALIFA 277 were fitted with $A_V = 0$ (see Fig. 4 in Paper I). About half of the shape-perturbed versions of these spectra are bluer than the original ones. Since STARLIGHT was not allowed to decrease A_V to fit them, it gets stuck at $A_V = 0$, compensating the extra blueness with younger populations, whicg in turn produces the tail toward negative $\Delta\langle \log t \rangle_L$ seen in the bottom rows of Fig. 1. This is also the reason why there are more instances of $\Delta A_V > 0$ than < 0, producing the positively skewed ΔA_V distributions.

Increasing the shape errors by a factor of 2 (C004 runs) leads to uncertainties roughly twice as large, biases of the order of -0.1 dex in ages, and $+0.15$ mag in A_V (Table 2), as well as larger skewness of the Δ distributions. Conversely, C001 runs result in uncertainties ~2 times smaller, with no bias nor much skewness.

3.1.4. Covariances

As usual in multivariate problems, variations in one property correlate with variations in others. In the context of spectral synthesis, such covariances are known as age-metallicity-extinction degeneracies: A_V affects colors while t and Z affect both absorption lines and colors. Previous work (e.g., Cid Fernandes et al. 2005; Sánchez-Blázquez et al. 2011) showed that λ-by-λ fits help in reducing these degeneracies, but do not eliminate them.

Figure 2 shows how the main scalar properies co-vary in our simulations. S runs were used in this plot-O runs produce similar plots, with slightly more compact contours. The first two columns (SR1 and SR2 runs) reveal the classical degeneracies of population synthesis, with mean ages increasing as mean metallicities decrease, and extinction variations also anti-correlated with variations in mean age. The bottom row shows how stellar masses increase as ages become older, and vice-versa.

In the R-runs, perturbations do not change the overall shape of the input spectrum. Hence, if a STARLIGHT solution turns out to be somewhat older than the original one, its extinction and/or metallicity will decrease to compensate for the redder colors and stronger absorption lines of the older stars. C-runs, on the other hand, do change colors-that is, in fact, exactly what they are designed to do. This slightly changes the nature of "degeneracies".

If the shape perturbation followed the same functional form of the extinction law, it would suffice to change A_V, leaving the population mixture essentially unchanged, varying ts and Zs at the level they would naturally vary without color perturbations (i.e, due to noise and the multiplicity of solutions). To first order, this is indeed what happens in the C-runs. Compare, for instance, the (SR1, SR2) and (SC001, SC002) pairs of $\Delta A_V \times \Delta\langle \log t \rangle_L$ panels in Fig. 2. In both pairs of plots the corresponding type of error changes by a factor of 2 in amplitude, yet one sees that the ΔA_V contours grow more from SC001 to SC002 than from SR1 to SR2. Furthermore, the ΔA_V contours are nearly vertically oriented in the C-runs, while in R-runs the anti-correlation with $\Delta\langle \log t \rangle_L$ is clear (the age-extinction degeneracy). We thus confirm the expectation that, to first order, continuum shape uncertainties primarily affect the A_V estimates.

There are still two second-order features to explain about the $\Delta A_V \times \Delta\langle \log t \rangle_L$ C-run maps in Fig. 2. First, the inner contours have a slightly positive slope, opposite to that expected from

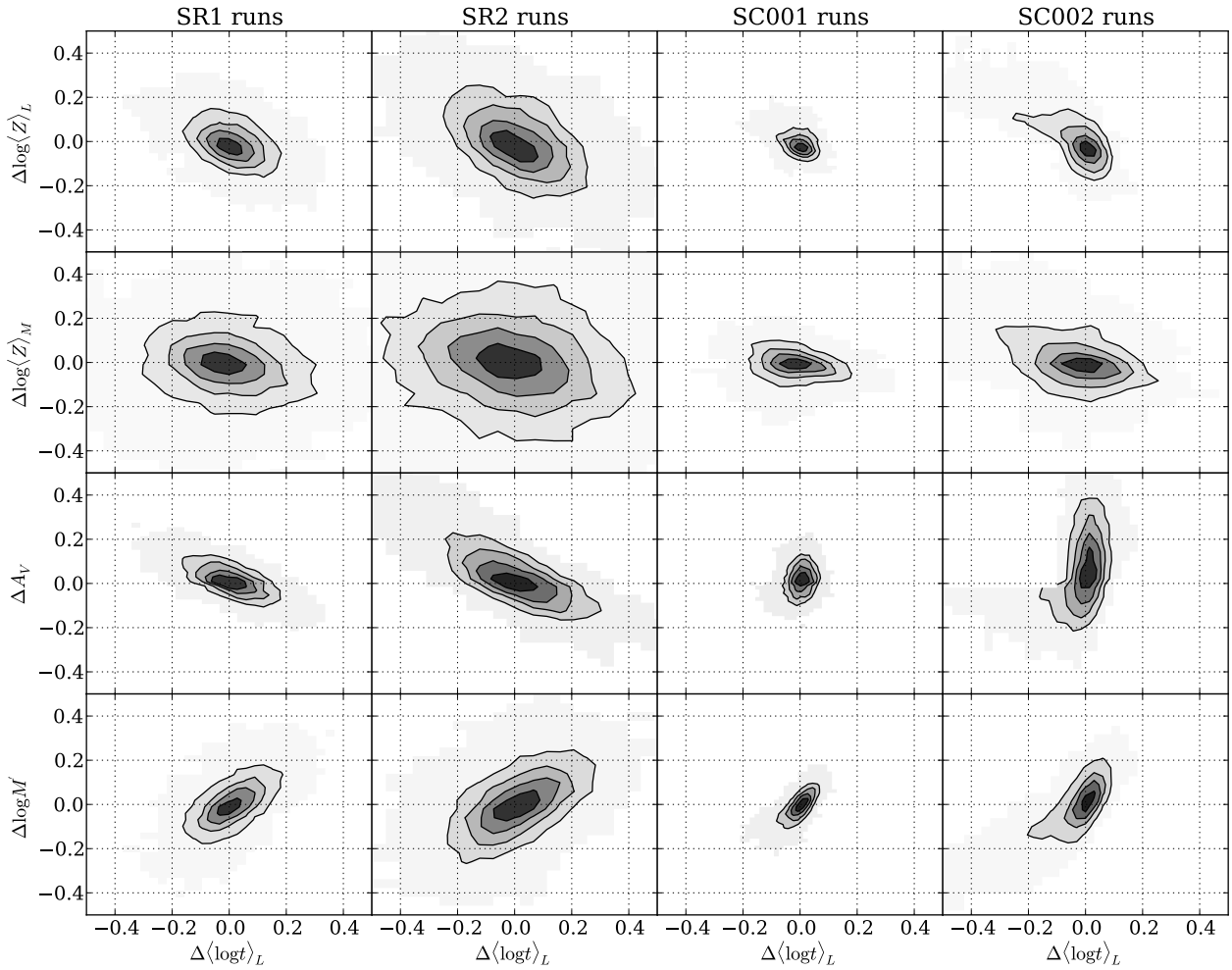

Fig. 2. Illustration of the correlated Δ variations of ages, metallicites, extincion, and masses. Each column corresponds to one set of simulations. Iso-density contours mark 20, 40, 60, and 80% of enclosed points.

the age-extincion degeneracy (clearly seen in the SR1 and SR2 panels). This behaviour arises because our shape-perturbations (where $\Delta \log F_\lambda$ changes linearly with λ) cannot be fully absorbed by the reddening curve used in the fits, because of the different functional forms. Since changing A_V alone is not quite enough to produce optimal spectral fits, STARLIGHT searches for other ways of making the synthetic spectrum redder or bluer (depending on the sign of the perturbation), and skewing the age mixture is the most effective one. In any case, the change in mean age is small: $\overline{\Delta}\langle\log t\rangle_L = -0.04$ for SC002 runs (Table 2). Secondly, for SC002 runs, the outer contours, which contain about 20% of the points, are limited to $\Delta A_V \leq 0$ whenever simulated mean ages are more than 0.1 dex younger than the original ones ($\Delta\langle\log t\rangle_L < -0.1$). Spectra along this odd-looking ridge were originally fitted with $A_V = 0$, and then perturbed to bluer colors. Since our fits do not allow $A_V < 0$, STARLIGHT fitted them with the same A_V (hence $\Delta A_V = 0$), but younger ages, leading to the pattern observed in this panel.

3.1.5. Checking the effects of imposing $A_V \geq 0$

As seen just above, the effects of a hypothetical systematic redness associated to spectroscopic calibration problems are almost entirely absorbed by biasing A_V upward, while an excessive blueness of the same nature stumbles upon the $A_V \geq 0$ constraint, leading to biases in the age mixture. An artificial way of fixing this assymetry is to allow for $A_V < 0$ in the spectral fits. A trivial statistical justification to allow for this unphysical possibility is that when A_V is truly = 0, unbiased estimates should oscillate around 0, including both negative and positive values. We also note that completely independent studies (different methods, different data) also allow for or deduce negative extinctions (e.g. Kauffmann et al. 2003, their Fig. 11).

An extra set of OE0-like simulations, but imposing $A_V > -1$, was run to investigate this issue. In practice what happens is that only zones that were originally fitted with $A_V \sim 0$ are systematicaly refitted with negative A_V. The 591 zones that result in $A_V < 0$ in these new fits include all the 582 for which we had originally found $A_V = 0$, and the remaining 9 had A_V very close to 0. The mean new A_V of these 591 zones was −0.12 mag, with a standard deviation of 0.08 mag. The lowest value reached was $A_V = -0.4$ mag, but in only 16% of the cases A_V went below −0.2 mag. The variations in other global properties (for these same 591 zones) were $\overline{\Delta}\langle\log t\rangle_L = +0.12$, $\overline{\Delta}\langle\log t\rangle_M = +0.12$, $\overline{\Delta}\log\langle Z\rangle_L = -0.07$, $\overline{\Delta}\log\langle Z\rangle_M = -0.04$, and $\overline{\Delta}\log M' = +0.03$ dex. Mean ages are the most affected property, but to an acceptably small degree, with biases of the same order as the one-σ variations produced by random and shape-related uncertainties at the nominal level (R1 and C002 simulations).

Zones originally fitted with $A_V > 0$ did not show systematic variations in their global properties in these extra runs.

Fig. 3. $\Delta\langle \log t \rangle_L$ variations for all 1638 zones × 10 realizations plotted against the S/N of the corresponding original (observed) spectra. The different panels correspond to different sets of simulations, as labeled. The lines are the running mean of $\Delta\langle \log t \rangle_L$ and the corresponding $\pm 1\sigma$ intervals.

3.1.6. The effect of S/N

The Δ histograms in Fig. 1 and the statistics in Table 2 represent the uncertainties of the CALIFA 277 data set as a whole, including zones of different S/N. The general expectation is that a lower S/N leads to larger uncertainties, such that the tails of the Δ distributions in Fig. 1 are dominated by the noisier spectra, while the cores (small Δ values) correspond to a higher S/N.

To verify this, Fig. 3 shows the $\Delta\langle \log t \rangle_L$ variations against the S/N of the original spectra. The lines represent the mean $\pm 1\sigma$ range. Uncertainties do decrease with increasing S/N in the SE0, SR1, and SR2 runs, but not in the C-runs. This is in fact expected because the shape perturbations have the same amplitude for all spectra, independent of their S/N. The right-most panel illustrates the bias toward younger ages that results from the $A_V \geq 0$ barrier.

We note in passing that, as hinted in Fig. 3, the distribution of S/N values in CALIFA 277 is skewed toward the target minimum S/N of 20 stipulated in the spatial-binning scheme explained in Paper I. This same threshold was applied to all other datacubes, so the uncertainty estimates presented here should be representative of the CALIFA data in general, even though they are based on a single galaxy.

3.2. Exploring the statistics in IFS data: uncertainties in radial profiles

So far our error analysis has treated the datacube as a collection of unrelated spectra. Spectral synthesis does not care where a spectrum comes from, so the positional information provided by IFS has no effect on the uncertainties for individual zones. However, averaging results obtained for different zones brings obvious statistical benefits, reducing uncertainties in properties derived from individual spectra.

Figure 10 in Paper I shows azimuthally averaged profiles for a series of global properties. Despite the scatter at any given radius, gradients in age and metallicity were clearly detected in CALIFA 277. Figure 4 in this paper repeats that plot, but now showing the radial profiles derived from the simulated spectra, so that we can evaluate the uncertainties in azymuthal averages. The original profiles are plotted in solid black, while colored shades represent the ranges spanned by OR1 (blue) and C002 (red) runs. Each panel also shows an (arbitrarily shifted) simulation minus the original difference profile at its bottom.

As expected, random noise does not change the radial profiles significantly. A small bias of about 0.02 mag is seen in $\overline{A_V}(R)$ for radii beyond ∼1 half light radius (HLR). This is again a side effect of the $A_V \geq 0$ constraint: when a spectrum originally fitted with $A_V \sim 0$ is perturbed, some will be refitted with $A_V > 0$, but none will be refitted with $A_V < 0$. The resulting bias propagates to minor offsets in other properties, like slightly younger ages (panels e and f). Overall, however, these are all tiny effects.

Color perturbations (red-shaded regions) produce larger effects. The main effect is on the $\overline{A_V}(R)$ profile (panel c), where the typical offset matches the expected $\overline{\Delta}(A_V) = 0.05$ mag reported in Table 2. Leaving aside this bias, the $\overline{A_V}(R)$ variations are much smaller than those in individuals zones. For instance, considering all radial bins and all 10×1638 versions of the OC002 spectra, the rms in $\Delta(\overline{A_V})$ is only 0.05 mag, while individual zones are subjected to uncertainties of $\sigma(A_V) = 0.16$ mag (Fig. 1).

Larger than typical variations occur in the nucleus for the trivial reason that it comprises a single spaxel, so that no actual statistics is performed. More robust results for the nucleus require extending its definition to include more spaxels. For instance, Pérez et al. (2013) define the nucleus as the inner 0.1 HLR, which in the case of CALIFA 277 comprises the central-most nine spaxels. Uncertainties also increase in the outer regions ($R > 2$ HLR), induced by the lower S/N of the spectra and by the decrease in the number of zones compared with intermediate radii.

Figure 4 confirms the expected gain in dealing with averaged results. The simulations show that the uncertainties in the radial profiles are just too small to question the reality of the age and metallicity gradients in this galaxy. This conclusion holds even for our most pessimistic simulations (C004 and R3). This gain comes at the expense of a loss of spatial information–compressing the x, y plane to a single dimension (R) in the example presented here. Other forms of spatial averaging can be envisaged, such as isolating structural components (bulge, disk, bar, arms, inter-arm regions). From the results above, it is easy to see that as long as enough zones are involved in the statistics, any averaging scheme should lead to small uncertainties.

3.3. Uncertainties in star formation histories

STARLIGHT outputs SFHs as population vectors: $\boldsymbol{x}$ and $\boldsymbol{\mu}$. The first gives the fractional contribution of each base element to the total flux at the chosen reference wavelength (5635 Å in our case), while the latter gives the corresponding mass fractions. The 39 ages in our base[5] sample the 1 Myr to 14 Gyr interval with a mean $\log t$ separation of just 0.1 dex. Individually, the strength of each component is not reliable (see Fig. 1 in Cid Fernandes et al. 2004 for an example), given the many ways that spectrally

[5] The base ages were chosen such that SSP spectra for ages in between t_j and t_{j+1} are adequately interpolated with the j and $j+1$ SSPs (see Mateus et al. 2006 for more details).

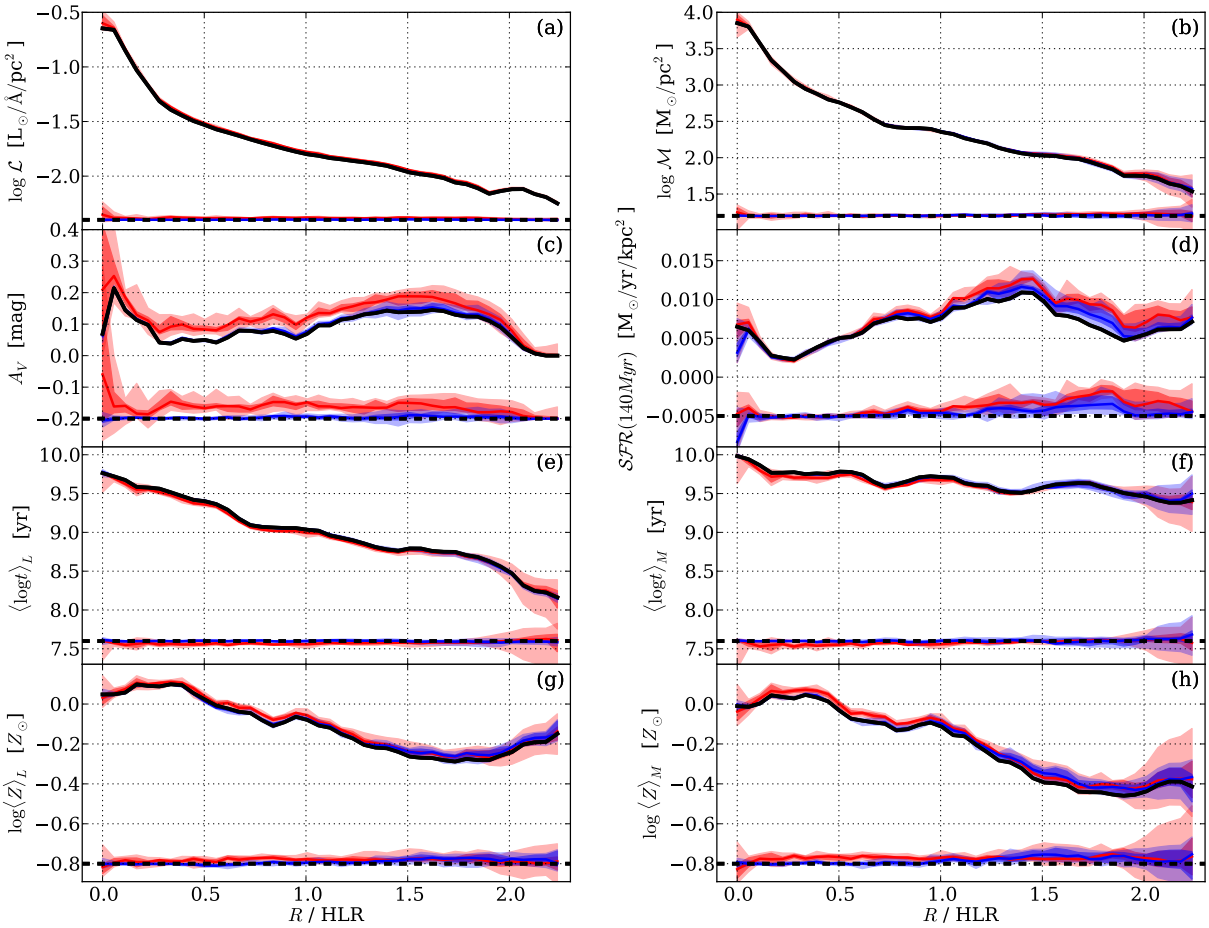

Fig. 4. Azimuthally averaged radial profiles of some global properties. **a)** Dereddened luminosity surface density at 5635 Å; **b)** stellar mass surface density; **c)** extinction; **d)** SFR surface density over the last 140 Myr; **e)** luminosity and **f)** mass-weighted mean log ages; **g)** luminosity and **h)** mass weighted mean metallicities. The original solution is shown as the solid black line, while those obtained from simulations are shown in colored shades. Light-blue shades show the full range of values found among the ten realizations of random noise (OR1 runs). The darker blue bands mark the range spanned by the six values closer to the median (marked with the blue solid line). OC002 simulations of color-change effects are plotted in red. The differences between the simulations and original profiles are shown at the bottom of each panel, with the zero level indicated by the dashed black line. The dispersion and offsets along these difference profiles give a measure of the uncertainties and biases involved in the radial profiles.

similar components can rearrange their x_j values to produce a similar total spectrum. Yet, their combination in indices such as $\langle \log t \rangle_L$ produces robust quantities, as shown by the simulations.

None of the properties studied in the previous sections relies heavily on the details of the spectral decomposition performed by STARLIGHT. Mean ages, for instance, are defined as the *first moment* of the age distributions encoded in $\boldsymbol{x}$ and $\boldsymbol{\mu}$. By construction, however, SFHs do rely on higher-order descriptions of the age distribution. This section aims to evaluate the uncertainties in our SFHs. Before doing so, we digress on the actual meaning of the "H" in SFH.

There is no standard semantics for the meaning of "history" in spectral synthesis studies. On the contrary, the same term is used in very different ways in the literature, always qualitatively referring to some measure of the how star formation unfolded over time, but quantitatively varying from estimates of mean ages to measures of the "recent" star formation (where recent can mean anything between ~10 Myr and a few Gyr, depending on the tracers), and full t-by-t descriptions with varying age resolution. In addition to different timescales, different metrics can be used to quantify the SFH, such as luminosity or mass associated to stars in a given age range, in differential or cumulative forms, etc. This diversity is particularly natural in the case of non-parametric codes such as STARLIGHT. Because no pre-established functional forms for $SFR(t)$ are imposed, one has plenty of freedom to design different indices and functions for characterizing the SFH. Paper I is itself an example of this diversity. We have presented there a series of SFH diagnostics obtained by manipulating the $\boldsymbol{x}$ and $\boldsymbol{\mu}$ vectors in different ways, always with the underlying philosophy of compressing the overly-detailed output of STARLIGHT into coarser but more robust descriptions of the SFH.

In short, there are numerous alternative ways to describe an SFH. In what follows we evaluate the uncertainties in a paper I-inspired selection of SFH-descriptions. Following the sequence of our analysis of global properties, we first present results for uncertainties in individual zones (Sect. 3.3.1) and then for spatially averaged SFHs (Sect. 3.3.2).

3.3.1. SFHs for individual zones

One way to characterize SFHs is to compress the population vectors into age ranges, an approach that dates back to the early days of spectral synthesis (Bica 1988). In the context of STARLIGHT, Cid Fernandes et al. (2004 and 2005) showed that a robust summary of the SFH can be obtained by binning $\boldsymbol{x}$

Table 3. Uncertainties in light fraction in age groups: $\overline{\Delta x} \pm \sigma_{\Delta x}$ [%].

log t range	OE0	OR1	OR2	OR3	OC001	OC002	OC004
6.0→8.2	0.0 ± **1.9**	0.1 ± **3.3**	–0.0 ± **4.9**	–0.0 ± **6.0**	0.1 ± **2.2**	0.6 ± **2.7**	1.6 ± **4.2**
8.2→9.2	0.0 ± **4.5**	–0.1 ± **9.3**	0.7 ± **14.9**	1.2 ± **18.9**	0.7 ± **6.8**	4.3 ± **14.7**	9.4 ± **23.2**
9.2→10.2	–0.1 ± **4.0**	–0.2 ± **8.7**	–1.0 ± **14.0**	–1.7 ± **17.9**	–0.8 ± **6.4**	–5.0 ± **14.8**	–11.6 ± **23.9**
6.0→7.5	–0.0 ± **1.4**	–0.2 ± **3.0**	–0.4 ± **4.7**	–0.6 ± **5.8**	0.0 ± **1.6**	0.3 ± **2.4**	0.8 ± **4.0**
7.5→8.5	–0.4 ± **3.1**	–0.7 ± **5.9**	–1.3 ± **8.4**	–1.7 ± **10.1**	–0.4 ± **3.5**	–0.4 ± **4.8**	–0.1 ± **7.2**
8.5→9.3	0.0 ± **4.4**	0.0 ± **9.4**	0.6 ± **15.1**	1.2 ± **19.2**	0.7 ± **6.7**	4.5 ± **14.7**	9.8 ± **23.3**
9.3→9.7	–0.1 ± **5.5**	–0.7 ± **12.8**	–2.3 ± **19.6**	–4.0 ± **23.8**	–0.6 ± **7.7**	–3.2 ± **13.2**	–7.4 ± **18.2**
9.7→10.2	–0.1 ± **3.4**	0.1 ± **10.4**	0.4 ± **16.4**	1.0 ± **20.1**	–0.4 ± **5.6**	–2.3 ± **11.0**	–5.0 ± **15.4**

Notes. As Table 2, but with the statistics of the light fractions in age groups for our different simulations. For each age group the table lists the mean simulated minus original difference ($\overline{\Delta}$) and its standard deviation (σ_Δ, in bold), both in percentages. The top part of the table corresponds to the young, intermediate and old groups description discussed in the text and Fig. 5, while the bottom one is for the five groups used in Fig. 6. The R1 and C002 columns are those applicable to CALIFA data.

into young (Y), intermediate (I) and old (O) groups. This $\boldsymbol{x} \rightarrow (x_Y, x_I, x_O)$ compression scheme was employed to produce the maps shown in Figs. 8 and 9 of Paper I, where these age groups were separated by frontiers placed at $t = 0.14$ and 1.4 Gyr.

We have used our simulations to assess the uncertainties in these fractions. The top three rows in Table 3 list the O-run statistics of Δx_Y, Δx_I and Δx_O, where, as before, $\Delta \equiv$ simulated – original value. The table shows that random noise at the one-σ level (R1 runs) induces uncertainties in x_Y, x_I and x_O of $\sim$ 3, 9 and 9% respectively[6]. Larger uncertainties ($\sim$15%) in x_I and x_O apply for shape-changing perturbations at the level expected for CALIFA (OC002 runs). Part of this increase is due to the aforementioned $A_V \geq 0$ effect, which forces low A_V zones that are perturbed to bluer spectra to become younger to compensate for the inability to reduce A_V below 0. This is also reflected in the biases, with $\overline{\Delta}(x_I) \sim -\overline{\Delta}(x_O) \sim +5\%$. At any rate, comparing these uncertainties with $(\overline{x_Y}, \overline{x_I}, \overline{x_O}) = (15, 25, 60)\%$, the average values over all zones in CALIFA 277, one concludes that this compact description of the SFH is reasonably robust.

Figure 5 show how the variations in these light fractions correlate with each other. Δx_O anti-correlates very strongly with Δx_I, which is not suprising given the $x_Y + x_I + x_O = 100\%$ constraint and the fact that these are the age groups that carry most of the light in CALIFA 277. The Y group, on the other hand, is roughly independent of the others.

Figure 6 shows OR1-based results for a finer-graded description using five age groups defined by log t [yr] = 6–7.5, 7.5–8.5, 8.5–9.3, 9.3–9.8, and 9.8–10.2. The anti-correlations among adjacent age groups are visible along the diagonal. The $\sigma_{\Delta x}$ values are 3, 6, 9, 13, and 10% (Table 3) for mean fractions of $\overline{x_t}$ = 13, 3, 24, 34 and 26%, respectively. This more detailed description shortens the difference between $\sigma_{\Delta x}$ values and the corresponding $\overline{x}$. The uncertainties are therefore larger, as expected, but these higher-resolution SFHs are still useful, especially if used in a statistical way.

As the name implies, STARLIGHT fits light, so $\boldsymbol{x}$-based SFH descriptions are the natural way to analyze its results. Yet, it is the stellar mass which carries the astrophysical information on the galaxy history. One can obviously group $\boldsymbol{\mu}$ elements just as we did for $\boldsymbol{x}$, but while light fractions in different age groups can be directly intercompared, mass fractions cannot. With a factor of $\sim$1000 decrease in light-to-mass ratio from the youngest to the oldest elements in the base, wildly different mass fractions, say, $\mu_Y \sim 1\%$ and $\mu_O \sim 99\%$, can be equally relevant for the fits.

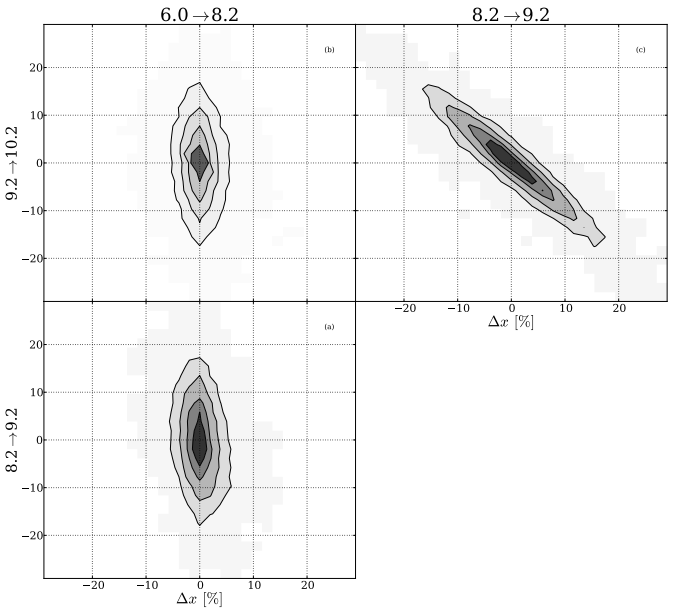

Fig. 5. Variations in the light fractions in young, intermediate, and old age groups (spanning log t = 6–8.2, 8.2–9.2 and 9.2–10.2, respectively) for OR1 simulations. Contours are drawn at 20, 40, 60, and 80% of enclosed points.

To evaluate uncertainties in the masses associated to different age groups it is better to use the masses themselves instead of mass fractions. Table 4 reports the uncertainties in log M', where the prime denotes the initial stellar mass, uncorrected for the mass returned to the interstellar medium by stellar evolution. The table reports the results of the simulations for both the (Y, I, O) and the five age groups descriptions of the SFH discussed above. Again, the fact that we are now exploring higher moments of the age distribution implies larger uncertainties than for global properties. A column-by-column comparison with the $\Delta \log M'$ statistics in Table 2 confirms this. For instance, while the total masses have uncertainties of just 0.08 dex in OR1 runs, the masses in age-groups are typically three times more uncertain, an unavoidable but still reasonable price to pay for a time-resolved description of results of the spectral synthesis.

Additional subdivisions lead to a quasi-continuous regime that is better handled in ways other than age-binning, such as

[6] Note that % here does *not* mean a relative deviation, but the very units of the x_Y, x_I, and x_O fractions.

Table 4. Uncertainties in log mass in age groups.

log t range	OE0	OR1	OR2	OR3	OC001	OC002	OC004
6.0→8.2	0.00 ± **0.14**	–0.00 ± **0.21**	–0.02 ± **0.30**	–0.03 ± **0.37**	0.00 ± **0.16**	0.02 ± **0.20**	0.06 ± **0.29**
8.2→9.2	0.00 ± **0.12**	–0.01 ± **0.25**	–0.00 ± **0.37**	–0.00 ± **0.44**	0.01 ± **0.18**	0.08 ± **0.28**	0.15 ± **0.36**
9.2→10.2	–0.00 ± **0.05**	–0.00 ± **0.12**	–0.01 ± **0.19**	–0.02 ± **0.24**	–0.01 ± **0.08**	–0.04 ± **0.20**	–0.05 ± **0.33**
6.0→7.5	–0.00 ± **0.11**	–0.02 ± **0.18**	–0.05 ± **0.27**	–0.07 ± **0.32**	–0.00 ± **0.12**	0.01 ± **0.14**	0.02 ± **0.18**
7.5→8.5	–0.00 ± **0.22**	0.01 ± **0.35**	0.06 ± **0.43**	0.09 ± **0.49**	0.00 ± **0.24**	0.03 ± **0.29**	0.12 ± **0.39**
8.5→9.3	0.00 ± **0.12**	–0.00 ± **0.24**	0.01 ± **0.35**	0.03 ± **0.41**	0.01 ± **0.18**	0.08 ± **0.28**	0.16 ± **0.36**
9.3→9.7	0.00 ± **0.14**	–0.01 ± **0.25**	–0.02 ± **0.34**	–0.03 ± **0.39**	–0.01 ± **0.17**	–0.03 ± **0.24**	–0.04 ± **0.31**
9.7→10.2	–0.00 ± **0.11**	–0.01 ± **0.24**	–0.00 ± **0.34**	0.02 ± **0.38**	–0.01 ± **0.15**	–0.03 ± **0.27**	–0.00 ± **0.36**

Notes. As Table 3, but for the (log) initial masses associated to stellar populations in different age groups. Each entry lists $\overline{\Delta \log M'} \pm \sigma_{\Delta \log M'}$ (in dex).

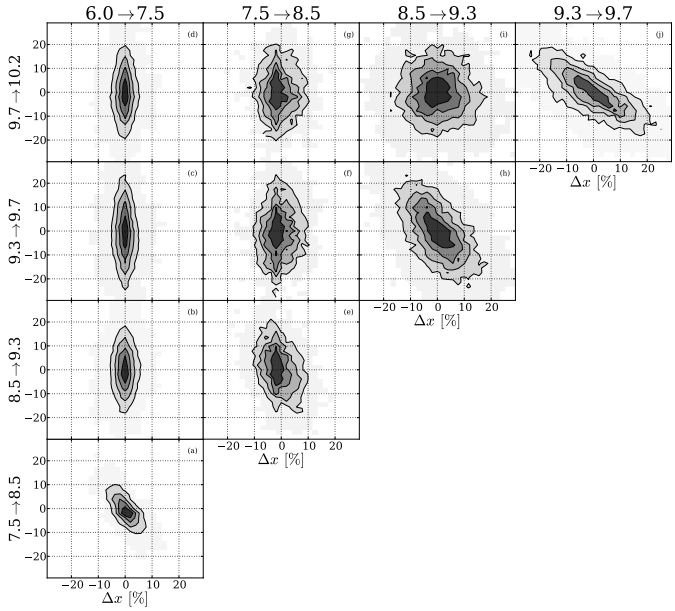

Fig. 6. As Fig. 5, but for a finer-graded description of the SFH in terms of five age groups.

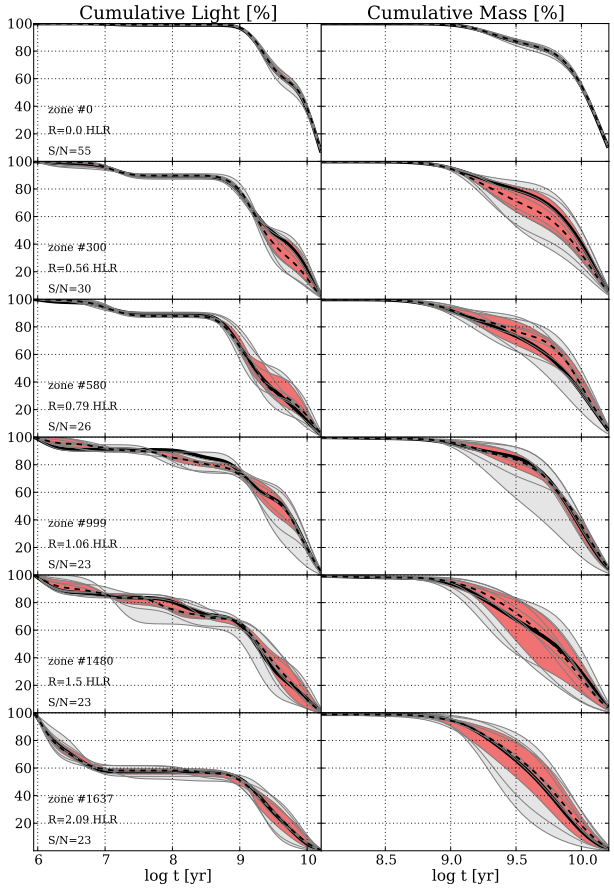

Fig. 7. SFHs of six different spatial zones, as represented by their cumulative light (*left*) and mass (*right*) distributions. Each panel shows 11 SFHs. The original one is plotted in thick black lines. Gray lines show the SFHs for the ten realizations of OR1 simulations, with a gray-shaded band tracing the full range of values. The median SFH is drawn as a dashed line, while the red band marks the range spanned by the six runs closer to the median. The zone number, distance to the nucleus, and the S/N are labeled in *the left panels*.

cumulating $\boldsymbol{x}$ or $\boldsymbol{\mu}$. Figure 7 illustrates this. The left panels show the cumulative light fraction, shown after a cosmetic smoothing by a $FWHM = 0.2$ dex Gaussian in $\log t$. Six individual zones, including the nucleus (top) and the last one (bottom), were chosen to exemplify this alternative description of SFHs and its uncertainties. The thick solid line traces the original STARLIGHT solution, while thin gray lines trace the solutions obtained for each of the ten OR1 perturbations, whose median is drawn as a dashed line. The red-shaded bands mark the range spanned by the six SFHs closer to the median at each t. The right panels show the corresponding mass assembly histories, all scaled to 100% at $t = 0$. The six zones in Fig. 7 cover different distances from the nucleus, S/N, and SFHs, thus providing representative examples of STARLIGHT results for CALIFA data.

Except for the nucleus, where the perturbed spectra and the associated SFHs are nearly indistinguishable from the original ones because of the very high S/N, there is a clear spread in cumulative light and mass distributions. The dispersion is higher for ages older than a few Gyrs, which contain a substantial fraction of the light and most of the mass. The mild spectral evolution in this age range (Conroy 2013) explains the larger uncertainties, because components can be interchanged with relatively little variation in the total spectrum. Gauging the uncertainties in the cumulative SFHs by the width of the red band, we find it to be of the order of 10–20%, comparable to those reported in Tables 3 and 4 for age-grouped quantities.

Despite uncertainties, the example SFHs in Fig. 7 reveal a gradual shift toward younger ages for larger radii. This trend is the higher-order description of the clear mean age radial gradients previously shown in panels e and f of Fig. 4, where results for individual zones at same R were averaged.

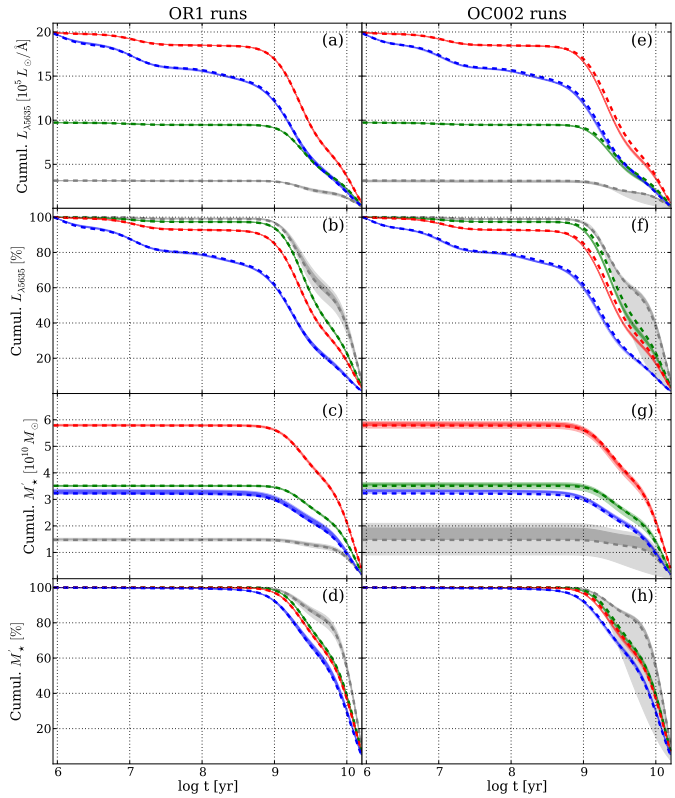

Fig. 8. Spatially averaged cumulative light and mass histories and their uncertainties for four regions: red, blue, green and grey colors correspond to regions $R > 1$, $R < 1$, $R < 0.5$ HLR and the nucleus, respectively. *Left and right panels* show OR1 and OC002 simulations, respectively. The *top four panels* show cumulative light distributions, in both absolute (**a**), **e**)) and relative (**b**), **f**)) units. Similarly, the bottom four panels show the correponding mass distributions. In each panel and for each radial region the original solution is drawn as a dashed line, while results of the simulations are shown with shaded bands, hardly visible except for the nuclear curves (grey). Light shaded bands show the ranges spanned by all 10 perturbations, and darker shades indicate the range spanned by the 6 versions closer to the median. The nuclear curves in panels **a**), **c**), **e**) and **g**) contain much less light and mass than the others, and so were multiplied by 20 for clarity.

3.3.2. Spatially averaged SFHs

In analogy with Sect. 3.2, uncertainties in the SFHs of individual zone spectra should be greatly reduced when averaging over spatial regions comprising many zones. This is confirmed in Fig. 8, which shows cumulative light and mass distributions (in both absolute and relative units) for four spatial regions: the nucleus (gray), $R < 0.5$ (green), $R < 1$ (red), and $R > 1$ HLR (blue). The smoothing here is by $FWHM = 0.5$ dex in $\log t$ for compatibility with Fig. 11 of Paper I. OR1 runs are shown in the left panels, OC002 in the right panels.

The statistical damping of the uncertainties is so strong that one can hardly notice that, as in Fig. 7, shaded bands are used to show the vertical ranges spanned by the simulations. The bands are so narrow that they look like thick lines. The exception is the nucleus (gray), where the uncertainties are visible for the simple reason that unlike for other regions, it does not involve any actual spatial averaging because it comprises a single spaxel.

Close inspection of the right panels reveals that the shape-changing simulations tend to have SFHs slightly skewed toward younger ages than SFHs inferred from spectral fits of the original data (marked by dashed lines). This is again the effect of some perturbations making low A_V zones too blue to be fitted with $A_V \geq 0$, so STARLIGHT compensates it with lower ages.

In conclusion, spatial averaging effectively eliminates the dispersions seen in the SFHs of individual zones. This remains valid even for simulations where the level of perturbations is higher than that expected for CALIFA spectra.

3.4. Summary of the simulations

The overall message from these experiments is that degeneracies inherent to stellar populations and uncertainties in the data have a limited impact upon our analysis. For the level of uncertainties expected in CALIFA, here represented by the R1 and C002 runs, the simulations yield acceptably small uncertainties in global properties. More precisely, mean ages and metallicities have one sigma variations of ∼0.1 dex when weighted in luminosity and ∼0.15 weighting by mass, while stellar masses are good to ∼0.1 dex (round numbers), with negligible biases. Uncertainties in A_V are ∼0.07 mag if only random noise is considered, increasing to 0.16 mag when shape-related uncertainties are accounted for. In the latter case, a bias of $\overline{\Delta} A_V \sim 0.06$ mag appears as a side-effect of forbidding negative A_V. Because they are a higher-order product of the spectral synthesis, SFHs are naturally less strongly constrained than global properties, but the age-grouped and cumulative descriptions explored above were found to be relatively stable against random and spectral-shape perturbations.

It is a consensus in the field of spectral synthesis that the results are best used in a statistical sense (e.g., Panter et al. 2006), and the present study endorses this view. First, averaging reduces uncertainties, as confirmed by our experiments with radial profiles (Fig. 4) and spatially averaged SFHs (Fig. 8), which showed that the results are stable against the types of uncertainties studied here. Second, and more importantly, large samples allow for *comparative* studies, which can reveal astrophysically valuable trends while mitigating worries about data or method-related systematics affecting the absolute values. This point has been abundantly explored in spectral synthesis studies of SDSS galaxies, where unequivocal relations between, for instance, age, mass, and metallicity, have been identified in spite of the uncertainties in the properties derived for individual galaxies. The analogy with IFS data is clear. In IFS each galaxy is a large sample by itself, and the very goal of IFS is to compare properties in one place with those somewhere else.

This concludes our study of data- and method-related uncertainties. This whole study was based on a single set of SSP models, a fundamental prior in the analysis. We now turn to the uncertainties resulting from this choice.

4. Uncertainties associated to evolutionary synthesis models: an empirical evaluation

SSP spectra are the key astrophysical ingredient in any spectral synthesis analysis, the dictionary that translates stellar properties into spectroscopic information, which can then be compared with data through different methods to extract physical properties and SFHs. The computation of model SSP spectra combines an assumed initial mass function (IMF) with stellar evolution theory (isochrones) and a library to provide the spectra of stars with temperature, gravity, and chemical composition matching the values dictated by the evolutionary tracks. In practice, this algebric exercise stumbles upon a series of problems that directly affect the predicted SSP spectra.

Table 5. Sets of SSP models used for spectral fitting.

Model-set	Code	Isochrones	IMF	Library	Metallicities	$N_Z \times N_t$	Ages [Gyr]
GM	SED@ + Vazdekis	Padova 2000	Salpeter	GRANADA + MILES	0.004, 0.008, 0.019, 0.033	4×39	0.001–14
CB	Galaxev	Padova 1994	Chabrier	GRANADA + MILES	0.004, 0.008, 0.02, 0.05	4×40	0.001–14
BC	Galaxev	Padova 1994	Chabrier	STELIB	0.004, 0.008, 0.02, 0.05	4×40	0.001–14

First, despite its (well-deserved) status as a founding stone of modern astrophysics, stellar evolution is still an open business. Rotation in massive stars (Levesque et al. 2012), rapid phases during the post-main-sequence evolution of intermediate-mass stars (Maraston 2005), and binarity (Li et al. 2013) are examples of phenomena that affect the radiative output of stellar populations, but are still at the forefront of research, and thus prone to uncertainties. A practical example of how incomplete treatment of evolutionary phases affects spectral synthesis analysis is discussed by Koleva et al. (2008), Cid Fernandes & González Delgado (2010), and Ocvirk (2010). Using different codes, these studies found that spectral synthesis of old Milky Way and LMC globular clusters suggests the presence of very young stellar populations accounting for some 10–20% of the optical light, an astrophysically absurd result that also happens in some old, "red and dead" galaxies (Cid Fernandes et al. 2011). These "fake young bursts" (Ocvirk 2010) are in fact by-products of inadequate modeling of the horizontal branch (and/or blue straglers, Cenarro et al. 2008). Seeking for a best match, the spectral synthesis uses young hot stars to replace the old hot ones missing in the models.

Second, spectral libraries, whether empirical or theoretical, have their own limitations (Martins & Coelho 2007). The widely used Bruzual & Charlot (2003; BC03) models, for instance, are based on the STELIB library (Le Borgne et al. 2003), a major advance in its days, but limited to 249 stars, some of which have severe spectral gaps. Its relatively poor coverage of the stellar parameter space propagates to the predicted SSP spectra, with collateral effects on the results of a spectral synthesis analysis. Koleva et al. (2008), for instance, noticed that because of the lack of truly very metal-rich stars in the library, BC03 spectra for 2.5 $Z_\odot$ SSPs behave like $\sim Z_\odot$ models of an older age. In short, this too is an evolving field, undergoing constant revisions.

In this unstable scenario, it is relevant to evaluate to which extent the results of our spectral synthesis analysis depend on the choice of SSP models. To this end, we performed spectral fits using three bases extracted from different sets of evolutionary synthesis models, listed in Table 5. We stress that this is by no means an exaustive exploration of this matter, but it suits our goal of estimating uncertainties related to this choice. Section 4.1 describes the sets of models to be compared. The CALIFA data used in this study are described in Sect. 4.2. Comparisons of the properties obtained with different sets of SSPs are presented in Sect. 4.3. Fit-quality indicators and spectral residuals are discussed in Sect. 4.4, while Sect. 4.5 compares residuals in CALIFA and SDSS spectra.

4.1. SSP spectral models

Base GM is a combination of publicly available SSP spectra from Vazdekis et al. (2010)[7], which start at an age of 63 Myr, with the González Delgado et al. (2005)[8] models for younger ages. The former are based on stars from the MILES library (Sánchez-Blázquez et al. 2006), while the latter relies on the synthetic stellar spectra from the GRANADA library (Martins et al. 2005). Only minor adjustments were needed to match these two sets of models. First, we smoothed the GRANADA models to the spectral resolution of MILES (2.5 Å *FWHM*, cf. Beifiori et al. 2011 and Falcón-Barroso et al. 2011), and then multiplied them by a factor of 1.05, estimated from a direct comparison of the predicted spectra in the 63 Myr to 1 Gyr range, where these models agree very well in both continuum shape and absorption features. These independently derived SSP spectra complement each other, providing a base suitable for spectral fitting analysis of galaxies of all types.

The evolutionary tracks in base GM are those of Girardi et al. (2000), except for the youngest ages (1 and 3 Myr), which are based on Geneva tracks (Schaller et al. 1992; Schaerer et al. 1993a,b; Charbonnel et al. 1993). The specific subset of SSPs used in the spectral fits comprises 39 ages between 1 Myr and 14 Gyr. Four metallicities are included: $Z = 0.2$, 0.4, 1, and 1.5 solar. The Salpeter IMF is adopted. Like the other bases discussed next, GM does not adequately cover the blue horizontal branch, both because it does not reach very low Z and because, as thoroughly discussed by Ocvirk (2010), these models perform poorly in this phase anyway.

Base CB is an updated version of the BC03 models (Charlot & Bruzual 2007, priv. comm.), replacing STELIB by spectra from the MILES and GRANADA libraries. The IMF is that of Chabrier (2003), and the evolutionary tracks are those by Alongi et al. (1993), Bressan et al. (1993), Fagotto et al. (1994a,b), and Girardi et al. (1996), collectively labeled "Padova 1994" by BC03. $N_\star = 160$ SSPs comprising 40 ages in the 1 Myr to 14 Gyr range and four metallicities (0.2, 0.4, 1, and 2.5 $Z_\odot$) were chosen for the spectral fits.

Finally, base BC is built from the "standard" BC03 models, whose ingredients are identical to those of base CB except for the stellar library, which is STELIB. Base BC allows us to check whether differences between results obtained with bases GM and CB are associated to isochrones and/or IMF instead of stellar libraries.

These three bases cover the same age range, and all have the same number of metallicities. All bases start at $Z = 0.004$ (0.2 $Z_\odot$)[9], but GM stops at 1.5 $Z_\odot$, while both CB and BC extend to 2.5 $Z_\odot$. Because they are both based on stars not far from the Sun, neither MILES nor STELIB actually contain many stars as metal-rich as 2.5 $Z_\odot$, hence the latter models should be interpreted with care. On the other hand, the $Z \leq 1.5\ Z_\odot$ limit of base GM may be too low to accomodate the metal-rich inner regions of galaxies, possibly leading to saturation effects. Base GM also differs from the others in IMF and isochrones. The differences in opacities and equation of state between Padova 2000 (GM) and 1994 (CB and BC) tracks produce somewhat warmer (hence bluer) stars in the red giant branch in the former than in the latter.

[7] miles.iac.es

[8] www.iaa.csic.es/~mcs/sed@

[9] We excluded lower metallicity models to avoid SSPs that are not so well covered by the stellar libraries. Base GM base could include SSPs at $Z = 0.001$, but we chose not to do this because this metallicity is not available in the CB and BC models.

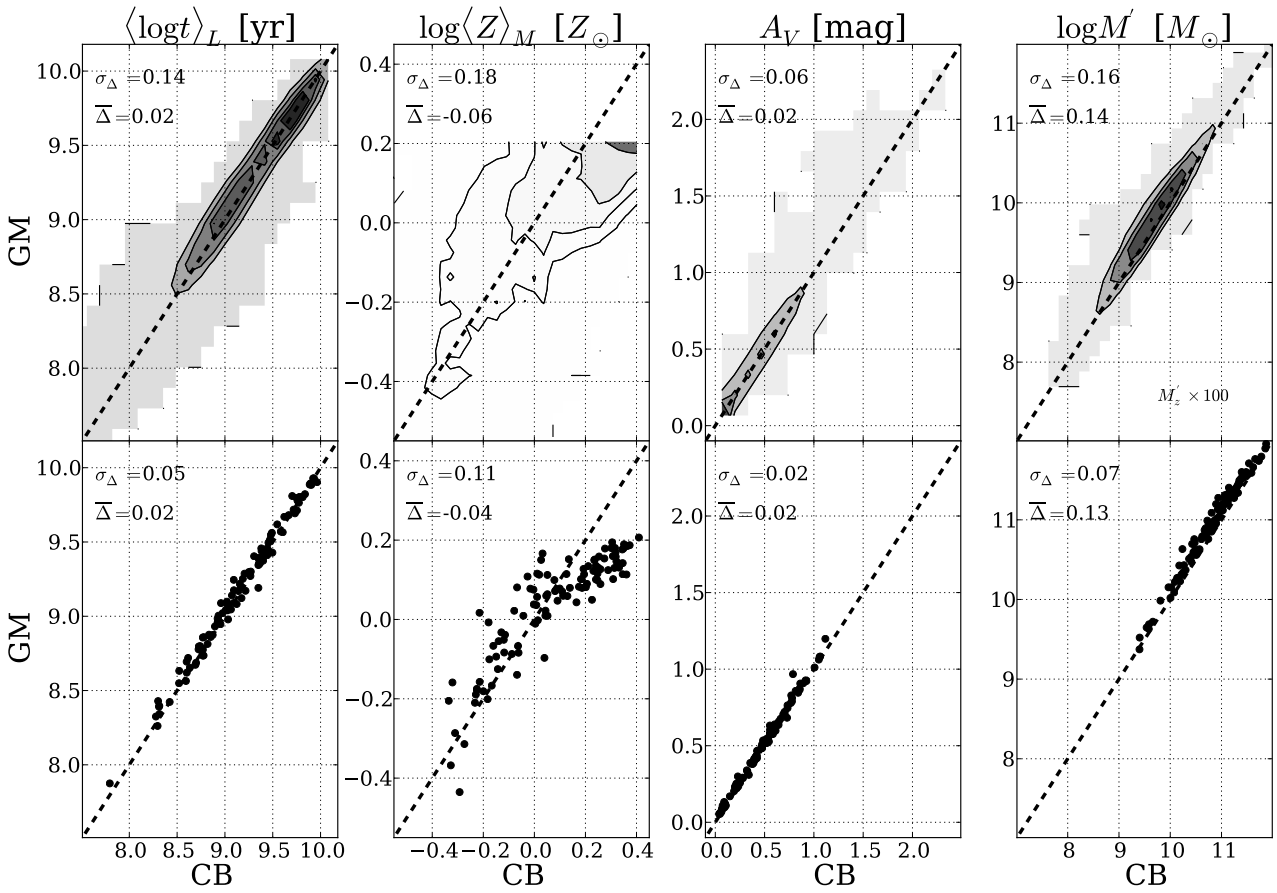

Fig. 9. Comparison of the global properties obtained with different bases GM (vertical axis) and CB (horizontal). The *top panels* compare luminosity-weighted ages ($\langle \log t \rangle_L$), mass-weighted metallicities ($\log\langle Z \rangle_M$), extinction (A_V), and initial stellar masses ($\log M'$) obtained for $\sim 10^5$ zones of 107 CALIFA galaxies, with contours drawn at every 20% of enclosed points. The *bottom panels* repeat the comparisons, but now for galaxy-wide values. A one-to-one line is drawn in all panels. Note that zone masses in the *top-right panel* are multiplied by 100 to put them on the same scale as the *bottom panel*.

In principle, one may thus expect to obtain older ages with GM than with CB and BC, but the diffences in color are very small in the spectral range of our data, only becoming noticeable at longer wavelengths (see Fig. 2 of BC03).

4.2. Data: 98291 spectra from 107 CALIFA datacubes

The experiments described below were based on data with the same characteristics and pre-processed in the exact same way as the data for CALIFA 277. The difference with respect to the simulations presented in the first part of this paper is that we now use 107 galaxies, out of which a total of 98291 spectra corresponding to individual spaxels (or Voronoi zones) were extracted following the methodology outlined in Paper I.

STARLIGHT fits of these spectra were carried out using the three bases described above. Our interest here is to identify and quantify differences in the results obtained with different model SSPs, so we overlook the spatial information and treat each spectrum as an individual. We now use this massive dataset to compare the results obtained with bases GM, CB, and BC.

4.3. Comparison of global properties

Figure 9 compares results obtained with bases CB (horizontal axis) and GM (vertical). The top panels compare values of $\langle \log t \rangle_L$, $\log\langle Z \rangle_M$, A_V and $\log M'$ for all 98921 individual zones, grayscale-coding by the density of points. Note the wide ranges of values in all panels, illustrating the diversity of properties spanned by the data. Each panel quotes the mean Δ and its standard deviation, where Δ = property(GM) − property(CB).

GM-based initial stellar masses are higher than CB ones by 0.14 dex on-average, reflecting their different IMFs[10]. Apart from this offset, the two masses agree to within 0.16 dex. Mean ages and extinctions also agree well, with dispersions of 0.14 dex in $\langle \log t \rangle_L$ and 0.06 mag in A_V, and insignificant offsets. Metallicities are obviously related, but with a substantial scatter due to a combination of the difference in evolutionary tracks and IMF, coarse Z grids, intrinsic difficulties in estimating metallicities, and the aforementioned difference in maximum Z.

To estimate the effect of this in-built difference we carried out experiments replacing the 2.5 $Z_\odot$ SSPs in CB by models interpolated to 1.5 $Z_\odot$, thus matching the Zs in GM. This modified CB base produces $\log\langle Z \rangle_M$ values that differ by $\sigma_\Delta \sim$ 0.15 dex from the GM-based ones, slightly better than the 0.18 dex obtained with the original CB. Leaving aside caveats associated to the interpolation of SSP spectra, this result indicates that the difference in maximum Z between GM and CB has a relatively modest impact on the difference between $\log\langle Z \rangle_M$ values obtained with these two bases, so that differences in evolutionary tracks and IMF are also relevant.

The bottom panels repeat the comparison of the same four properties, but now for galaxy-wide values, so there are only

[10] We recall that M' refers to the initial stellar mass. The difference between GM and CB masses increases to 0.27 dex when stellar mass loss is accounted for.

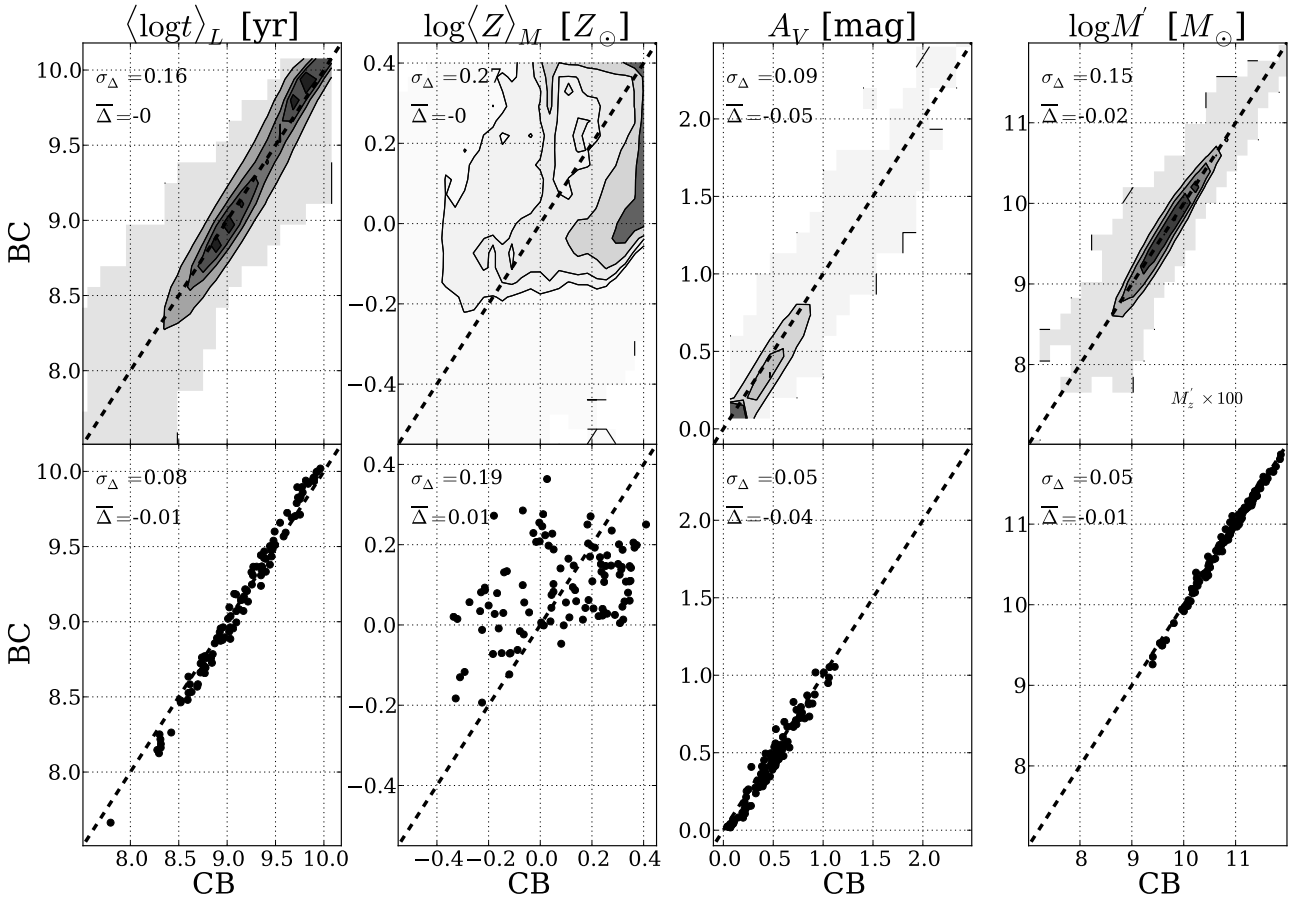

Fig. 10. As Fig. 9, but comparing results obtained with base CB (horizontal axis) and BC (vertical). Ages, extinction, and masses agree well, but BC-based mean metallicities differ significantly from both CB and GM-based estimates.

107 points[11]. These cleaner plots confirm the results from the zone-by-zone comparison. Naturally, galaxy-wide averages are much more robust than values for individual zones, so the σ_Δ dispersions are significantly lower than those quoted in the top panels. $\langle Z\rangle_M$ estimates distribute around the identity line up to $Z_\odot$, where the relation bends down due to the differences in the maximum Z between the bases. In these cases all one can be sure of is that metalicities are over-solar.

Figure 10 compares results for bases BC (vertical axis) with CB (horizontal). Masses, ages, and extinctions agree to within 0.15 dex, 0.16 dex, and 0.09 mag, respectively, with a tendency of A_V to be 0.04 mag smaller in BC fits. This relatively good level of agreement does not apply to mass weighted metallicities, which differ substantially. Given that these two sets of models coincide in isochrones, IMF and evolutionary synthesis code, this disagrement must be related to the stellar libraries. Because of the several improvements in MILES over STELIB (Sánchez-Blázquez et al. 2006), metallicities from the CB and GM bases should be considered more reliable than those obtained with BC (see also Koleva et al. 2008). This is in line with previous empirical tests by González Delgado & Cid Fernandes (2010), who showed that MILES based SSP spectra perform better than STELIB-based ones in matching the metallicities of star clusters derived from spectroscopy of individual stars or CMD analysis of the same systems.

Figure 11 shows histograms of the pair-wise Δs: GM − CB (top), BC − CB (middle), GM − BC (bottom). The width of these histograms (the σ_Δ values listed in each panel) gives a measure of the uncertainties related to the choice of evolutionary synthesis models for the spectral fits. As usual, mass-weighted properties present larger uncertainties than luminosity-weighted ones. Metallicities, for instance, have dispersions in $\Delta\log\langle Z\rangle_L$ of 0.14 and 0.19 dex (GM versus CB, and BC versus CB, respectively), lower than the 0.18 and 0.27 dex found for $\Delta\log\langle Z\rangle_M$.

The overal conclusion of these comparisons is that from a statistical perspective, physical properties obtained from a spectral synthesis analysis do not depend strongly on the choice of SSP models. Mean ages and extinctions retrieved with the three models investigated here all agree to within relatively small margins, and the same is true for stellar masses once in-built differences in IMFs are accounted for. Metallicities are generally harder to estimate, and our experiments confirm this. Unlike other properties, mean Zs are sensitive to the choice of models. Base BC gives the most discrepant results, highlighting the influence of stellar libraries.

Rounding up numbers, we find that uncertainties related to the SSP models adopted in the analysis are about 0.15 dex in $\langle\log t\rangle_L$, 0.23 dex in $\langle\log t\rangle_M$, 0.15 dex in $\log M'$, and 0.1 mag in A_V. Excluding BC models, $\log\langle Z\rangle_L$ and $\log\langle Z\rangle_M$ have uncertainties of 0.14 and 0.18 dex, respectively.

[11] M' in this case is the sum of the M'_z values for each zone. The galaxy-wide $\langle\log t\rangle_L$ is obtained from $\sum_j \log t_j \sum_z L_{jz} / \sum_j \sum_z L_{jz}$, where j is the base age index and z is a zone index. $\langle Z\rangle_M$ is computed analogously. Lastly, the galaxy-wide A_V is defined as the mean $A_{V,z}$ of all its zones.

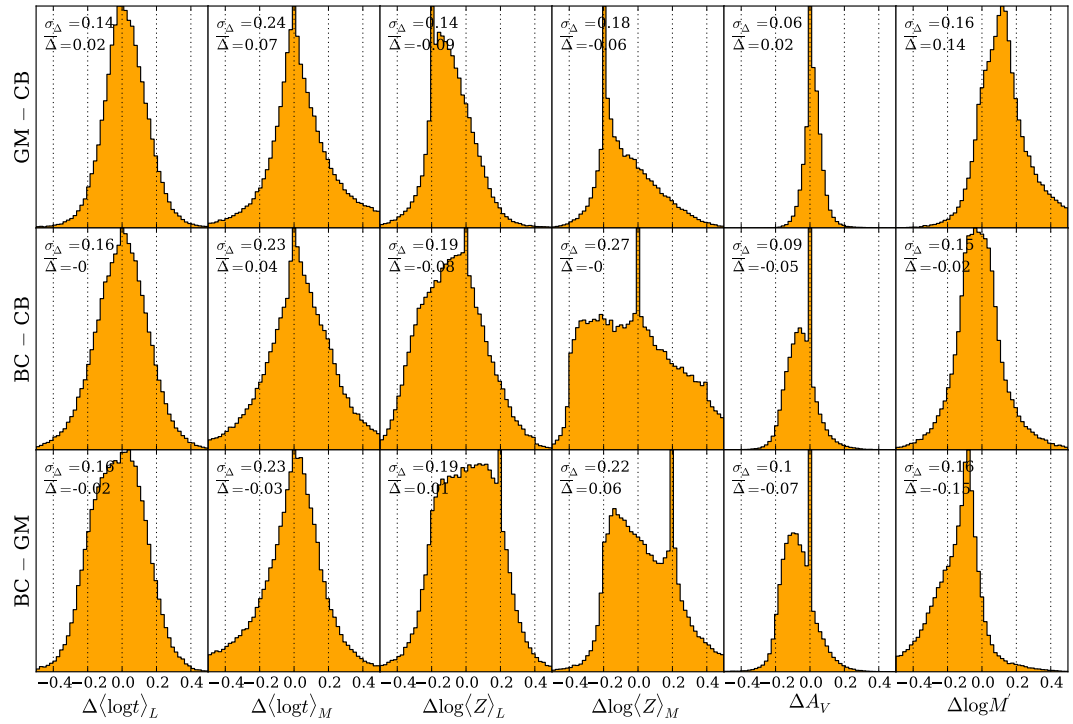

Fig. 11. Histograms of differences in global properties obtained with bases GM, CB and BC. Δs are defined as GM − CB in the *top panels*, BC − CB in the *middle*, and BC − GM in the *bottom* ones. The horizontal scale is the same as in Fig. 1 to facilitate the comparison of data- and method-related uncertainties with the model-related ones.

4.3.1. Models versus data- and method- related uncertainties

How do these uncertainties compare with those related to the data and synthesis method derived in the first part of this article?

To answer this, we focus on the OR1 and OC002 simulations described in Sect. 2.1, because they represent random noise and shape-calibration errors in CALIFA. Comparing the σ_Δ values in Fig. 11 with those for the OR1 simulations in Fig. 1, we see that the former are roughly twice as high as the latter. In fact, the σ_Δ values in the top panels of Fig. 11 are closer to those obtained in the simulations where the noise was artificially amplified by a factor or 2 (OR2 runs in Table 2). Shape-related uncertainties also lead to uncertainties somewhat smaller than those found in this section, with the exception of A_V, whose uncertainty is 0.16 mag in the OC002 runs but 0.1 mag or less in ΔA_V histograms in Fig. 11. A visual comparison of Figs. 1 and 11 corroborates the conclusion that uncertainties related to differences among SSP models are larger than those related to data and method.

These comparisons, however, are not entirely fair, since model-related uncertainties were evaluated from $\sim 10^5$ zones of 107 galaxies, while data- and method- related uncertainty estimates were based on 1638 zones of a single galaxy, perturbed ten times in the simulations. Repeating Fig. 11 for CALIFA 277 alone, we still find σ_Δ values higher than those found in Sect. 3, but by a smaller margin. The important message, therefore, is not so much that variance among current SSP spectral models leads to larger uncertainties than those related to data and method, but that both types of uncertainties are of the same order.

4.4. Quality of spectral fits

In principle, one could use the quality of the spectral fits as a criterion to favor one base over another. In practice, however, all bases studied here provide ∼ equally good fits to the data. This is illustrated in Fig. 12, which shows the distributions of two figures of merit. Green, red, and blue correspond to bases GM, CB, and BC, respectively. The histograms are nearly identical. Considering all 98291 fits, the mean (median) values of the relative spectral deviation (Eq. (6) of Paper I) are $\overline{\Delta}$ = 4.33 (3.92), 4.34 (3.93) and 4.37 (3.94) % for bases GM, CB, and BC respectively, while the corresponding mean (median) χ^2 per fitted flux are 1.14 (0.96), 1.15 (0.97), and 1.18 (0.99). Formally, this puts base GM in first place, with CB in second and BC in third, but these statistics reveal that it is clearly not possible to favour one or another model in terms of fit quality. Because of their more complete spectral libraries, we favor bases GM and CB over BC.

Figure 13 provides a more direct visualization of the quality of the spectral fits. The top panel shows (in black) the average of all zone spectra (for examples of fits to individual zones see González Delgado et al., in prep.). Each spectrum is first divided by the median flux in the 5590–5680 Å normalization window, such that all zones weight equally in the average. The average synthetic spectra are shown in green, red, and blue for fits with bases GM, CB, and BC, respectively, vertically offset for clarity. The corresponding residual spectra are shown in the bottom of the same panel, but can hardly be distinguished on this scale. This similarity persists even after zooming-in by a full order of magnitude in the flux scale, as shown in Fig. 13b. Gaps in this zoomed plot correspond to our generic emission line mask, i.e., λ-windows that were ignored in the STARLIGHT fits. Note that, because of bad-pixel flags, augmented by windows around telluric features (Paper I), the number of specra involved in these averaged residuals varies for each λ, as shown in the bottom panel.

4.5. Spectral residuals and comparison with SDSS

Clearly, the amplitude of the average spectral residuals is very small. Whereas the typical relative deviation for fits of individual

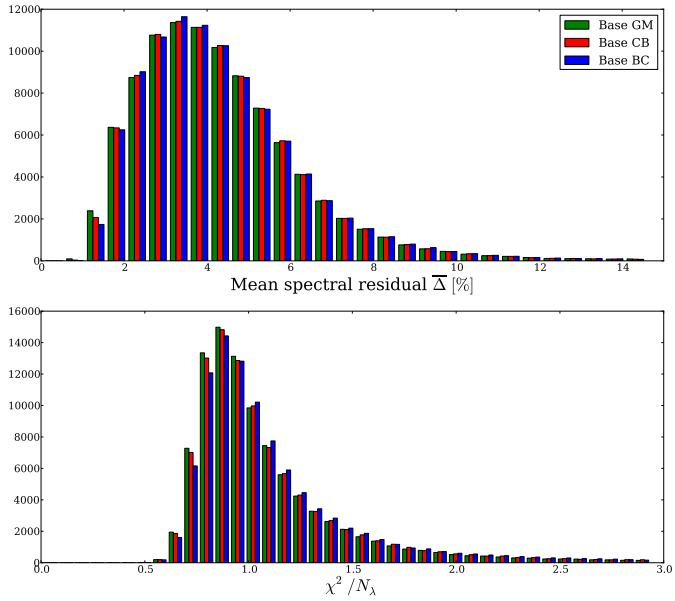

Fig. 12. *Top:* distribution of the mean percentual spectral deviation for STARLIGHT fits with the GM (bars in green), CB (red) and BC (blue) bases. *Bottom:* as above, but for the distribution of χ^2 per fitted flux. In both cases the distributions are essentially identical, reflecting the fact that the three bases provide equally good spectral fits.

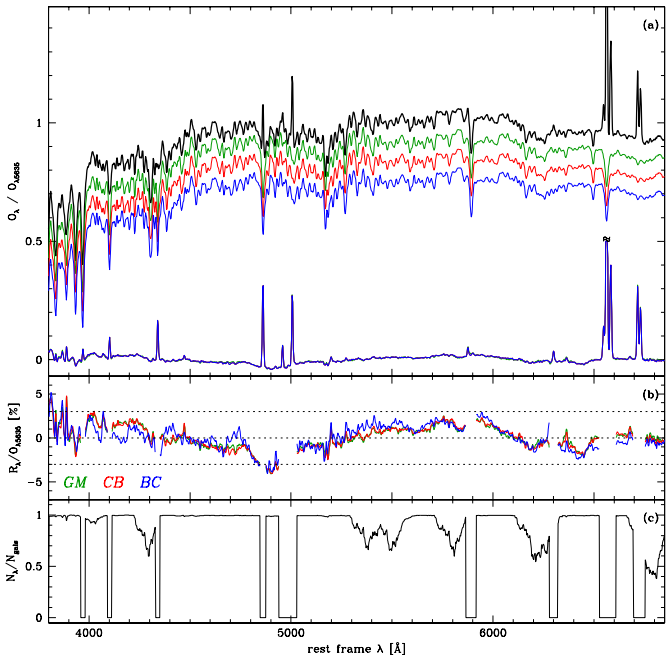

Fig. 13. **a)** Black line: mean (stacked) normalized spectrum of 98291 zones from 107 CALIFA galaxies, obtained by averaging $O_\lambda/O_{\lambda 5635}$. Green, red, and blue lines (offset for clarity): mean synthetic spectra for bases GM, CB, and BC, respectively. The corresponding $(O_\lambda - M_\lambda)/O_{\lambda 5635}$ residual spectra are shown with the same colors, but they are hardly distinguishable. **b)** Zoom of the spectral residuals in the top panel, with emission lines removed for clarity. Dashed lines mark residuals of −3, 0, and +3%. **c)** Fraction of the 98291 zones contributing to the spectral statistics in the *midde panel*.

zones is $\overline{\Delta} \sim 4\%$ (Fig. 12), for the average spectra in Fig. 13 it is a tiny $\overline{\Delta} = 1\%$. This four-fold reduction indicates that random noise is responsible for a large share of the residuals in spectral fits of individual zones. Yet, if the residuals were truly random, this reduction should be by a factor of the order of $\sqrt{98\,291} \sim 300$, not just 4. Residuals may be small, but they are definitely real.

The systematics of the spectral residuals are visible in Fig. 13b. These deviations must be related to problems in the method, data, and/or models.

The narrow positive peaks in the average residual spectrum are due to our fitting method. They are in fact emission lines not included in our generic emission line mask. [NeIII]λ3869, H8λ3889, [NI]λ5199, and [OI]λ6364 are the clearest ones in Fig. 13b, but weaker lines such as H9λ3835 and HeIλ6678 are identified upon closer inspection. The conservative clipping parameters adopted in the fits (see Paper I) prevent STARLIGHT from authomatically clipping these lines completely. Other STARLIGHT-based studies circumvented this problem by building emission line masks taylored for each spectrum individually (Mateus et al. 2006; Asari et al. 2007), a refinement that is yet to be implemented in our CALIFA pipelines.

The most noticeable features in Fig. 13b, however, are not the emission lines, but the broad $O_\lambda < M_\lambda$ trough around Hβ and the $O_\lambda > M_\lambda$ bumps in the 5800–6000 Å region. To understand whether these features are particular to CALIFA data, we compiled SDSS spectra for 50 galaxies in common with the current sample and analyzed them using the same masks, wavelength range, sampling, and STARLIGHT configuration as used for CALIFA. These spectra were then averaged as in Fig. 13, and compared with the spectral statistics obtained from the nuclear CALIFA spectra of the same 50 galaxies. Figure 14 shows the results of this comparison. The figure follows the same layout as Fig. 13, with spectral statistics for the 50 CALIFA nuclei and the corresponding SDSS spectra plotted in the left and right panels, respectively. The spikier appearance of the SDSS residuals is due to its higher spectral resolution (3 versus 6 Å FWHM), but this is irrelevant for the broad residual features discussed above.

The humps around the ∼5800 Å region of CALIFA residuals are clearly absent in the SDSS. These features are seen as shallow slanted bands in Fig. 16 of Husseman et al. (2013), where residuals for 100 galaxies were sorted by redshift. We have tracked the origin of this feature to an imperfect removal of telluric NaI bands around 5900 Å in the observed frame. In analogy with what happened with the SDSS data releases, as CALIFA data are used imperfections like this will be identified and, when possible, corrected for in future updates of the data base.

The trough around Hβ, however, is present in both SDSS and CALIFA spectra. This problem has been noticed before (Cid Fernandes 2006; Panter et al. 2007; Walcher et al. 2009), but not satisfactorily explained to date. The fact that this feature appears in both data sets suggests that the models overpredict the continuum in a ∼200 Å wide region centered around Hβ. On the other hand, the fact that the problem is present for both MILES- and STELIB-based models suggests that its origin is more likely related to data-calibration issues. This latter view is supported by studies such as the one by Ferré-Mateu et al. (2012), where STARLIGHT fits of *William Herschel* Telescope long-slit spectra of massive compact galaxies with the Vazdekis et al. (2010) models (included in our GM base) did not reveal the Hβ trough seen in Fig. 13b. While a full investigation of this matter is beyond the scope of this paper, we note that the trough tends to be more pronounced in younger systems, as previously reported by Cid Fernandes (2006, see his Fig. 5).

The negative residual seen in the region of Mg lines near 5175 Å is almost as strong as the Hβ trough. This deficit is due

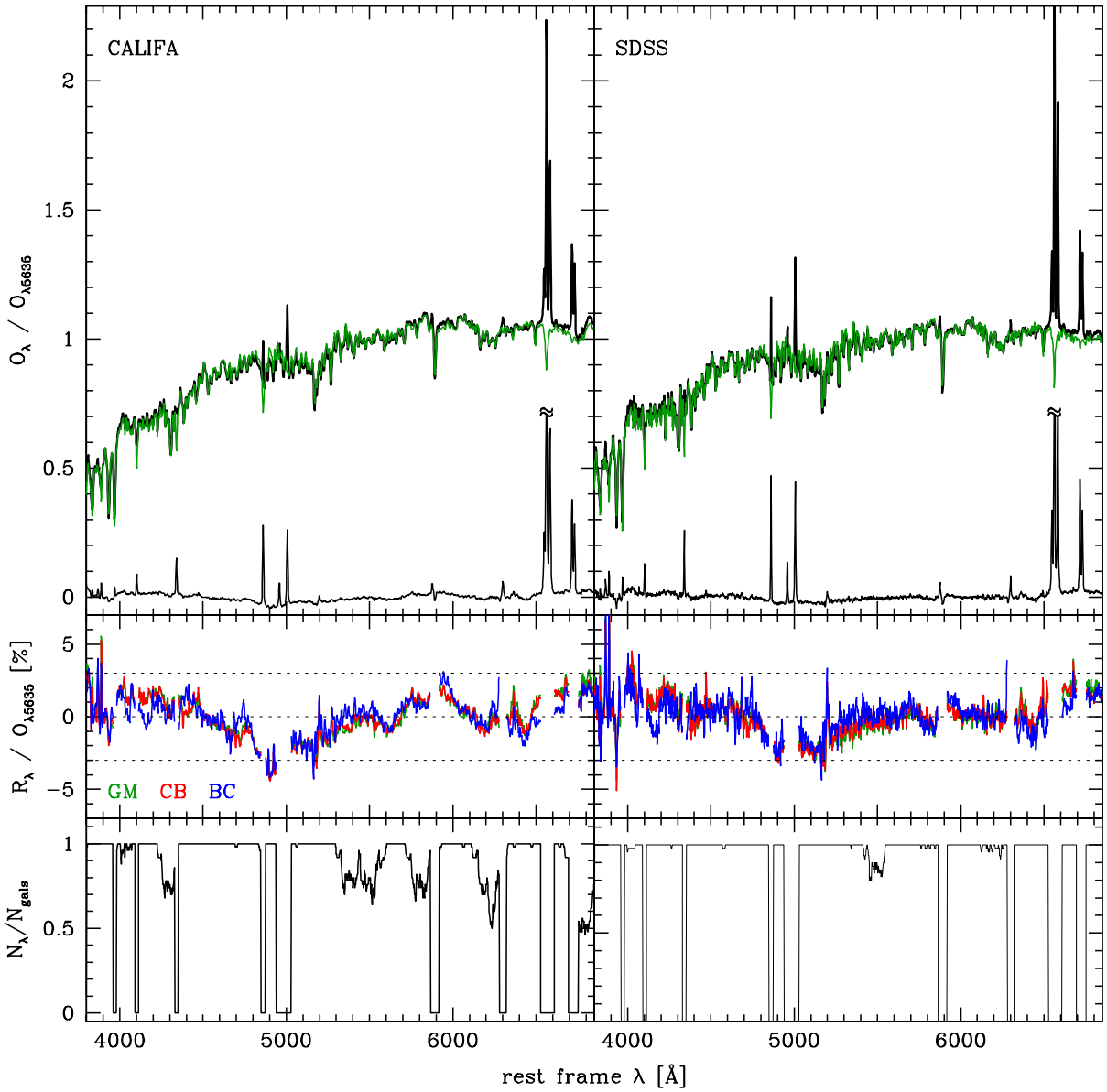

Fig. 14. As Fig. 13, but for a subset of 50 CALIFA galaxies for which SDSS spectra are available. Unlike in Fig. 13, however, *the top panels* show only results for GM-based spectral fits. *Left panels* show the spectral statistics for the nuclear extractions of CALIFA datacubes, whereas those on *the right* show results obtained from STARLIGHT fits to the SDSS spectra.

to differences in the abundances of α elements between galaxies (particular the central regions of massive ones) and stars in the stellar libraries. That this residual is stronger in Fig. 14 (nuclear spectra) than in Fig. 13 (all zones) is most likely due to α/Fe gradients. Walcher et al. (2009) reported steps toward incorporating α/Fe in spectral synthesis analysis.

4.6. Public distribution of base GM spectra

Given the good performance of the GM base in the tests discussed in this section, we make these spectra available to the community[12]. We stress that these models were produced by González Delgado et al. (2005) and Vazdekis et al. (2010), to whom all credit should be given. Our distribution just puts these two sets of models together.

[12] www.iaa.es/~rosa/AYA2010/AYA2010

5. Summary

We have performed an extensive study of the uncertainties in the stellar population properties derived from STARLIGHT fits of optical integral field spectra from the CALIFA survey. In the first part of this investigation, Monte Carlo simulations were used to explore the effects of random noise, continuum-shape calibration uncertainties and degeneracies in the spectral synthesis. The simulations are based on the data and STARLIGHT fits of 1638 zone spectra of the Sb galaxy CALIFA 277 (NGC 2916). Uncertainties in mean age, metallicity, extincion, stellar mass, and SFHs were evaluated both at the level individual zones (treating the datacube as a collection of unrelated spectra) and exploring the statistical power of IFS by means of radial profiles and other spatial averages. Second, we explored the variations in physical properties caused by using different sets of models for SSP spectra. We compared physical properties and spectral

residuals obtained from fits with three different SSP bases: (a) a combination of the Gonzalez Delgado et al. (2005) and Vazdekis et al. (2010) models (GM); (b) preliminary Charlot & Bruzual models using the GRANADA and MILES libraries (CB); (c) the standard Bruzual & Charlot (2003) models with STELIB (BC). This comparison employed $\sim 10^5$ spectra from 107 CALIFA galaxies.

Our main results can be summarized as follows:

1. The level of noise in CALIFA spectra leads to one-σ uncertainties in mean ages and metallicities of ∼0.08 dex when weighting in luminosity and ∼0.15 dex weighting by mass. Stellar masses are uncertain at the ∼0.08 dex level, while $\sigma(A_V)$ is about 0.06 mag.
2. Shape-related calibration uncertainties of 0.05 mag in $g - r$ produce uncertainties on the same order as those induced by random noise, except for A_V, which becomes uncertain at the level of $\sigma \sim 0.16$ mag and biased by +0.05 mag.
3. Time-resolved SFHs are higher-order products of the spectral synthesis, inevitably subjected to larger uncertainties than global properties. For instance, we found that, for individual zones, stellar masses formed in broad age ranges have uncertainties about three times larger than those of the mass formed over all times. While SFHs of individual zones do carry useful information, a statistical approach is more advisable.
4. Spectral synthesis results are best used in comparative studies involving large samples, and IFS data fit this requirement perfectly. Spatial averaging reduces uncertainties while preserving enough information on the history and structure of stellar populations. Radial profiles of mean ages, metallicities, mass density, and extincion are found to be very robust due to the large number statistics of datacubes. For instance, uncertainties in the data and limitations of the synthesis have a negligible impact upon the negative age and metallicity gradients detected in CALIFA 277. Similarly, SFHs averaged over different spatial regions are much more stable than those obtained for individual zones.
5. Mean ages, extinction, and stellar masses obtained with GM, CB, and BC models are reasonably consistent with one another, with one-σ differences in $\langle \log t \rangle_L$, $\langle \log t \rangle_M$, $\log M'$, and A_V of ∼ 0.15, 0.24, 0.15 dex, and 0.08 mag, respectively. GM and CB produce luminosity- (mass-) weighted metallicities consistent to within 0.14 (0.18) dex rms, while BC-based metallicities hardly correlate with the others.
   The variations induced by the use of different SSP models are larger than uncertainties due to data and method by roughly a factor of 2, highlighting the important role of this ingredient in the whole analysis.
6. The quality of the spectral fits is very nearly the same whichever set of SSP models is used. Spectral residuals are on the order of 4% for individual zones and 1% when averaged over all zones and galaxies. The strongest feature in the residual spectra is a broad but shallow (∼3%) trough around Hβ, also present in SDSS spectra, and its origin remains unidentified. Other systematic features in the residual spectra were tracked to incomplete masking of weak emission lines, imperfect removal of telluric features, and the lack of α enhanced models in the base.

These results, along with the tables and figures in this paper, fulfill our goal of assessing the uncertainties on the several products of STARLIGHT applied to CALIFA datacubes presented in Paper I. The combination of these spatial and temporal diagnostics of stellar populations with the diversity of galaxy types in the CALIFA survey offers new ways of studying galaxy evolution, that we aim to explore in other publications by our collaboration, and this paper provides quantitative and qualitative guidelines on how to interpret them appropriately.

Acknowledgements. CALIFA is the first legacy survey being performed at Calar Alto. The CALIFA collaboration would like to thank the IAA-CSIC and MPIA-MPG as major partners of the observatory, and CAHA itself, for the unique access to telescope time and support in manpower and infrastructures. We also thank the CAHA staff for the dedication to this project. R.C.F. thanks the hospitality of the IAA and the support of CAPES and CNPq. A.L.A. acknowledges support from INCT-A, Brazil. BH gratefully acknowledges the support by the DFG via grant Wi 1369/29-1. Support from the Spanish Ministerio de Economia y Competitividad, through projects AYA2010-15081 (PI RGD), AYA2010-22111-C03-03 and AYA2010-10904E (SFS), AYA2010-21322-C03-02 (PSB) and the Ramón y Cajal Program (S.F.S., P.S.B. and J.F.B.), is warmly acknowledged. We also thank the Viabilidad, Diseño, Acceso y Mejora funding program, ICTS-2009-10, for funding the data acquisition of this project. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

References

Alongi, M., Bertelli, G., Bressan, A., et al. 1993, A&AS, 97, 851
Asari, N. V., Cid Fernandes, R., Stasińska, et al. 2007, MNRAS, 381
Beifiori, A., Maraston, C., Thomas, D., & Johansson, J. 2011, A&A, 531, A109
Bica, E. 1988, A&A, 195, 76
Bressan, A., Fagotto, F., Bertelli, G., & Chiosi, C. 1993, A&AS, 100, 647
Cenarro, A. J., Cervantes, J. L., Beasley, M. A., Marín-Franch, A., & Vazdekis, A. 2008, ApJ, 689, L29
Charbonnel, C., Meynet, G., Maeder, A., Schaeller, G., & Schaerer, D. 1993, A&AS, 101, 415
Chabrier, G. 2003, PASP, 115, 763
Cid Fernandes, R. 2006, BAAA, 49, 228
Cid Fernandes, R., & González Delgado, R. M. 2010, MNRAS, 403, 780
Cid Fernandes, R., Gu, Q., Melnick, J., et al. 2004, MNRAS, 355, 273
Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska G., & Gomes J. M. 2005, MNRAS, 358, 363
Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, MNRAS, 413, 1687
Conroy, C. 2013 [arXiv:1301.7095]
Fagotto, F., Bressan, A., Bertelli, G., & Chiosi, C. 1994a, A&AS, 104, 365
Fagotto, F., Bressan, A., Bertelli, G., & Chiosi, C. 1994b, A&AS, 105, 29
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Ferré-Mateu, A., Vazdekis, A., Trujillo, I., et al. 2012, MNRAS, 423, 632
Girardi, L., Bressan, A., Chiosi, C., Bertelli, G., & Nasi, E. 1996, A&AS, 117, 113
Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371
González Delgado, R. M., Cerviño, M., Martins, L. P., Leitherer, C., & Hauschildt, P. H. 2005, MNRAS, 357, 945
González Delgado, R. M., & Cid Fernandes, R. 2010, MNRAS, 403, 797
González Delgado, et al. 2012, Highlights of Astronomy, Vol. 15, XXVIIIth IAU General Assembly
González Delgado, R., Pérez, E., Cid Fernandes, R., et al. 2013, The intriguing life of massive galaxies, Proc. IAU Symp., 295
Husemann, B., Jahnke, K., Sánchez, S. F., et al. 2013, A&A, 549, A87
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33

Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Le Borgne, J.-F., Bruzual, G., Pelló, R. et al. 2003, A&A, 402, 433
Levesque, E. M., Leitherer, C., Ekstrom, S., Meynet, G., & Schaerer, D. 2012, ApJ, 751, 67
Li, Z., Mao, C., Chen, L., Zhang, Q., & Li, M. 2013, ApJ, submitted [arXiv:1302.0927]
Maraston, C. 2005, MNRAS, 362, 799
Martins, L. P., & Coelho, P. 2007, MNRAS, 381, 1329
Martins, L. P., González Delgado, R. M., Leitherer, C., Cerviño, M., & Hauschildt, P. 2005, MNRAS, 358, 49
Mateus, A., Sodré, L., Cid Fernandes, R., et al. 2006, MNRAS, 370, 721
Ocvirk, P. 2010, ApJ, 709, 88
Panter, B., Heavens, A. F., & Jimenez, R. 2003, MNRAS, 343, 1145
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2007, MNRAS, 378, 1550
Pérez, E., Cid Fernandes, R., González Delgado, R. M., et al. 2013, ApJ, 764, L1
Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A. et al. 2012, A&A, 538, A8
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sánchez-Blázquez, P., Ocvirk, P., Gibson, B. K., Pérez, I., & Peletier, R. F. 2011, MNRAS, 415, 709
Schaerer, D., Charbonnel, C., Meynet, G., Maeder, A., & Schaller, G. 1993a, A&A, 102, 339
Schaerer, D., Meynet, G., Maeder, A., & Schaller, G. 1993b, A&AS, 98, 523
Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, A&AS, 96, 269
Tojeiro, R., Heavens, A. F., Jimenez, R., & Panter, B. 2007, MNRAS, 381, 1252
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Walcher, C. J., Coelho, P., Gallazzi, A., & Charlot, S. 2009, MNRAS, 398, L44