Hierarchical Bayesian approach for estimating physical properties in nearby galaxies: Age Maps (Paper II)

M. Carmen Sánchez-Gil [1,2]★ Emilio J. Alfaro,[2]★ Miguel Cerviño,[2,3,4]★ Enrique Pérez,[2]★ Joss Bland-Hawthorn [5] and D. Heath Jones[6]

[1]*Facultad de Ciencias, Universidad de Cádiz, 11510 Puerto Real, Spain*
[2]*Instituto de Astrofísica de Andalucía (CSIC), 18008 Granada, Spain*
[3]*Centro de Astrobiología (CSIC/INTA), 28850 Torrejón de Ardoz, Madrid, Spain*
[4]*Instituto de Astrofísica de Canarias, c/vía Láctea s/n, 38205 La Laguna, Tenerife, Spain*
[5]*School of Physics, Sydney Institute for Astronomy, University of Sydney, Darlington, NSW 2006, Australia*
[6]*English Language and Foundation Studies Centre, University of Newcastle, Callaghan NSW 2308, Australia*

ABSTRACT

One of the fundamental goals of modern astrophysics is to estimate the physical parameters of galaxies. We present a hierarchical Bayesian model to compute age maps from images in the H α line (taken with Taurus tunable filter, TTF), ultraviolet band (GALEX far UV, FUV), and infrared bands (Spitzer 24, 70, and 160 μm). We present the burst ages for young stellar populations in a sample of nearby and nearly face-on galaxies. The H α to FUV flux ratio is a good relative indicator of the very recent star formation history (SFH). As a nascent star-forming region evolves, the H α line emission declines earlier than the UV continuum, leading to a decrease in the H α/FUV ratio. Using star-forming galaxy models, sampled with a probabilistic formalism, and allowing for a variable fraction of ionizing photons in the clusters, we obtain the corresponding theoretical ratio H α/FUV to compare with our observed flux ratios, and thus to estimate the ages of the observed regions. We take into account the mean uncertainties and the interrelationships between parameters when computing H α/FUV. We propose a Bayesian hierarchical model where a joint probability distribution is defined to determine the parameters (age, metallicity, IMF) from the observed data (the observed flux ratios H α/FUV). The joint distribution of the parameters is described through independent and identically distributed (i.i.d.) random variables generated through MCMC (Markov Chain Monte Carlo) techniques.

Key words: methods: observational – methods: statistical – galaxies: spiral – galaxies: starburst – galaxies: star formation – galaxies: structure.

1 INTRODUCTION

This work carries on the study of a sample of nearby galaxies, where star-forming regions are spatially resolved, in order to place the relationship between star formation, ultraviolet, and H α emission on a stronger empirical foundation (hereinafter Paper I Sánchez-Gil et al. 2011). This paper focuses on the tools and mathematical methodology applied, based on a Bayesian model that yields more accurate results. A first approach to this Bayesian methodology can be found in Sánchez Gil et al. (2015). We refer to these two papers for more details about the motivation and interest in the study of the star formation history (SFH) and star formation rate (SFR) in

galaxies, in particular by means of the comparison between H α and ultraviolet emission as a tracer of recent SFH.

In this paper, we continue with the study of the age maps started in Paper I, for three new galaxies, using the new age-dating methodology from Bayesian inference approach applied pixelwise. Appendix D contains the resulting age maps for the rest of the galaxies sample, corresponding to Paper I.

Section 2 describes the data and its reduction, the pixel-based mapping, as well as the extinction correction via the total Infrared (TIR) to FUV ratio.

The stellar population model adopted and its uncertainties when applied pixelwise are described in Section 3. In our previous work (Sánchez Gil et al. 2015), we used the original STARBURST99 (Leitherer et al. 1999; Vázquez & Leitherer 2005) code results. Here, STARBURST99 has been modified to bring it closer with the theoretical model distribution assumed for the H α to FUV ratio (which also

★ E-mail: mcarmen.sanchez@uca.es (MCSG); emilio@iaa.es (EJA); mcs@iaa.es (MC); eperez@iaa.es (EP)

Table 1. Galaxy parameters (Sourced from *NASA Extragalactic Database*).

Galaxy	NGC 1068 (M 77)	NGC 5236 (M83)	NGC 5457 (M101)
RA (J2000)	$02^h 42^m 40.7^s$	$13h37m00.9s$	$14^h 03^m 12.5^s$
DEC (J2000)	$-00°00'48.0''$	$-29°51'56.0''$	$+54°20'56.0''$
Type	(R)SA(rs)b	SAB(s)c	SAB(rs)cd
Redshift	0.003793	0.001711	0.000804
Distance (Mpc)	14.4	4.5 ± 0.3	6.7
pc arcsec^{-1}	69.8	21.82	32.5
Inclination (deg)	$40° \pm 3$	$24°$	$18°$
Dim. (arcmin)	7.1×6.0	2.9×11.5	28.8×26.9
M_B	9.61	8.20	8.31

Note: Distance reference: Bland-Hawthorn et al. (1997) for NGC 1068; M83 Thim et al. (2003); M101 Tully et al. (2008) – Scale in pc per arcsec of the final images, and the age maps plots, which pixel scale is 1.5″/px. – Inclination angle reference: Bland-Hawthorn et al. (1997) for NGC 1068; M83 Foyle et al. (2012); M101 Bosma, Goss & Allen (1981)

differs with the one assumed in the previous work). Specifically, to obtain the number of ionizing photons, $Q(H)$, or the theoretical correlation ρ between H α and FUV luminosities.

Section 4 presents the new Bayesian inference methodologies proposed for age estimation generalized to estimate any parameter. However, as we explain below, only the age variable is sensitive to the ratio H α/FUV. In Section 4.1, the age maps assume pixel independence, keeping the spatial resolution at pixel value, and providing a sample of the age posterior probability function given the observed flux ratios for each pixel of the image. In Section 4.2, we study how image segmentation may affect several issues, such as the possible dependence between adjacent pixels of sub-sampling effects of the IMF from stellar population models.

Finally, Section 5 presents an analysis of the age maps and the resulting age patterns in view of possible mechanisms of galaxy structure and evolution, and the generation of galactic spiral arms. Section 6 contains a brief discussion and the conclusions.

2 OBSERVATIONS

We study the case of three new galaxies in addition to the original sample in Paper I, which are also included in Appendix C for comparison between the present Bayesian methodology and our earlier empirical model. This sample contains face-on nearby galaxies across a range of star-forming types. Low inclination angles mitigate the effects of extinction and scattering as well as wavelength shift in H α due to galactic rotation. Their proximity allows sufficient spatial resolution to resolve individual star-forming structures within spiral arms. A summary of the main physical properties of the sample is given in Table 1.

The H α images were taken with the Taurus Tunable Filter (TTF, Bland-Hawthorn & Jones 1998; Jones & Bland-Hawthorn 2001; Jones, Shopbell & Bland-Hawthorn 2002) on the William Herschel Telescope (WHT) during 1999 March 4–6. Conditions were photometric with stable seeing of 1.0 arcsec. TTF was tuned to a bandpass of width $\Delta\lambda = 20$ Å centred at $\lambda_c = 6570$ Å. The intermediate-width R0 blocking filter ($\lambda_c/\Delta\lambda = 6680/210$ Å) was used to remove transmission from all but a single interference order. The pixel scale was 0.56 arcsec.

The Far UV images were obtained from the Nearby Galaxies Survey of the Galaxy Evolution Explorer mission (NGS survey, GALEX, Martin et al. 2005). This survey contains well-resolved imaging (1.5 arcsec pix) of 296 and 433 nearby galaxies for

GR2/GR3 and GR4 releases, respectively, in two passbands: a narrower far-ultraviolet band (FUV; $\lambda_{\rm eff}/\Delta\lambda = 1516/268$ Å), and a broader near-ultraviolet band (NUV; $\lambda_{\rm eff} = 2267/732$ Å).

Ancillary 24, 70, and 160μ data from the Multiband Imaging Photometer for Spitzer (MIPS) were used to provide estimates of extinction, in the same way as in Paper I. For M83 and M101, we obtained IR data from the Spitzer Local Volume Legacy Survey[1] (LVL, Lee et al. 2011). These images were resampled to a common 1.5 arcsec/px scale (same as the 24μ MIPS and FUV images), and combined into an image of total far infrared (TIR) flux, according to $F_{\rm TIR} = \zeta_1 \nu F_\nu(24\mu) + \zeta_2 \nu F_\nu(70\mu) + \zeta_3 \nu F_\nu(160\mu)$, with $[\zeta_1, \zeta_2, \zeta_2] = [1.559, 0.7686, 1.347]$ (Dale & Helou 2002). For NGC 1068, only 70 and 160μ data were available from the Very Nearby Galaxy Survey[2] (VNGS, Bendo et al. 2012). In this case, the TIR flux was estimated from the latter as $S_{\rm TIR} = c_{70}S_{70} + c_{160}S_{160}$, according to equation (4) of Galametz et al. (2013) (where $c_{70} = 0.999 \pm 0.023$ and $c_{160} = 1.226 \pm 0.017$, from their Table 3), which is a quite reliable fit, with a coefficient $R^2 = 0.97$.

Internal reddening is corrected using a straight relation between the extinction $A_{\rm FUV}$ and $F_{\rm TIR}/F_{\rm FUV}$ from equation (2) of (Buat et al. 2005). The $A(H\alpha)$ extinction factor was derived through the relation $A_{\rm FUV} = 1.4A(H\alpha)$ (Boissier et al. 2005).

With all images on a common scale of 1.5 arcsec/px, our pixel-by-pixel technique becomes straightforward to implement. An example of the processed frames in H α, FUV, and TIR can be found in Fig. 1. The top panels display the different images for M83. Artefacts of the data reduction in the centre of the Hα and TIR images can be seen. Those pixels were masked in the final age map. NGC 1068, the bottom panels, show the range of morphologies in a single galaxy at different wavelengths. The strong effect of the central AGN is also quite evident.

The estimated calibration uncertainties for the MIPS images are 2 per cent, 5 per cent, and 9 per cent for the 24, 70, and 160μ data, respectively (Dale et al. 2007). Average relative errors of the respective images are $E_{H\alpha} \leq 5$ per cent, $E_{\rm FUV} \simeq 15 - 25$ per cent, and $E_{\rm TIR} \leq 10$ per cent, resulting in an overall uncertainty in $F_{H\alpha}/F_{\rm FUV}$ (reddening corrected) flux ratio of $\sim$23–28 per cent. More details on the data, data reduction, and extinction correction can be found in Paper I.
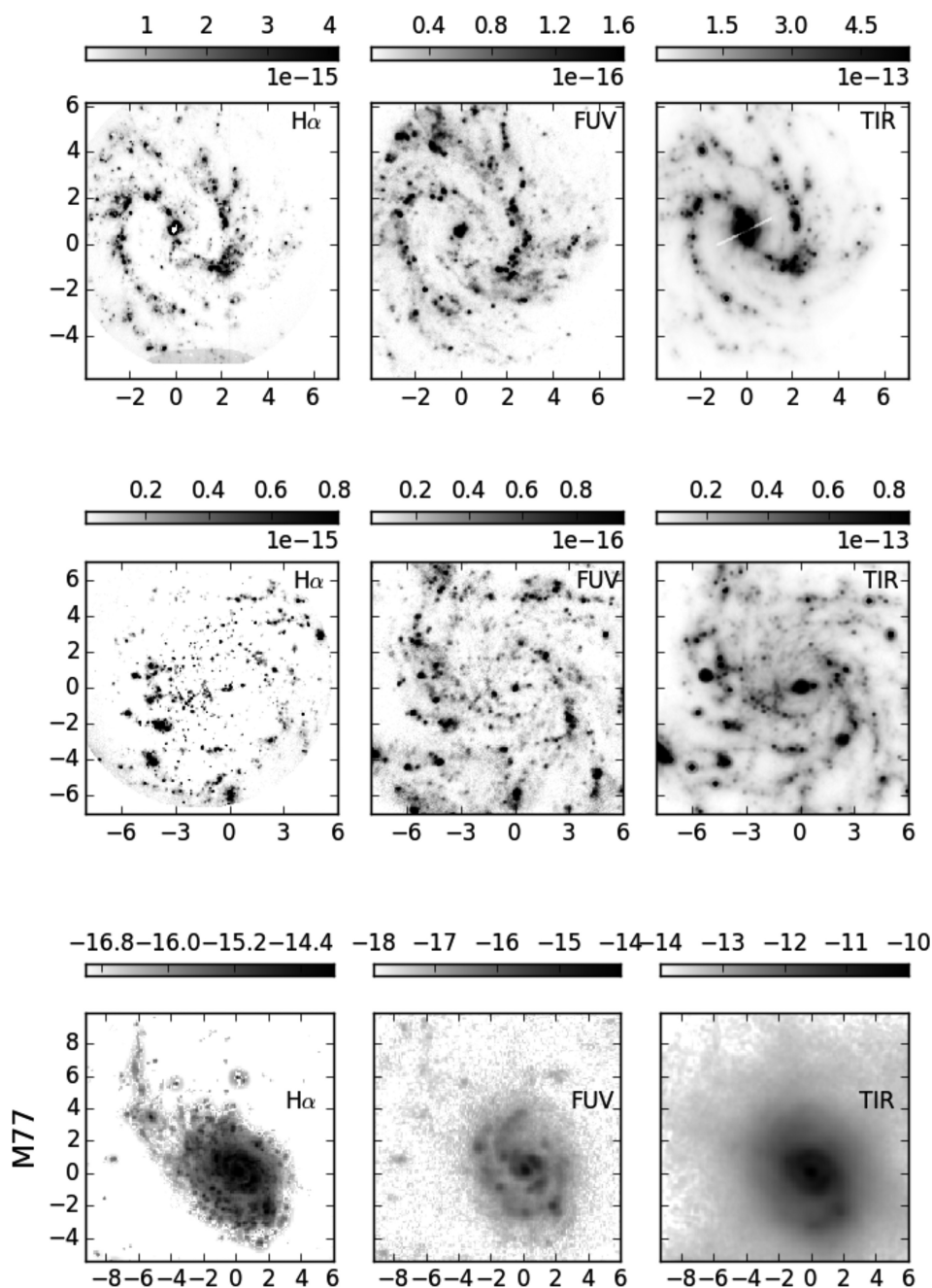
3 STELLAR POPULATION MODELS

To obtain the number of ionizing photons $Q(H)$ and the stellar and nebular contribution to the FUV/GALEX band, we implement the probabilistic formalism by Cerviño & Luridiana (2006) into STARBURST99 synthesis models (version 7.0.1 August 2014 Leitherer et al. 1999; Vázquez & Leitherer 2005). The original STARBURST99 code has been modified to obtain $Q(H)$ and the FUV/GALEX band emission for each of the stars along each isochrone (SB99 obtains such values for the ensemble after computing the total spectra of the cluster). We have included the FUV/GALEX filter response (as provided by the SVO filter service[3]) to customize SB99 to obtain this quantity for each star. Finally, we have obtained the nebular contribution to FUV/GALEX using the implementation of the nebular continuum used by SB99. $Q(H)$ values are converted to H α fluxes using the conversion factors provided by Leitherer & Heckman (1995), but multiplied by a factor

[1] http://irsa.ipac.caltech.edu/data/SPITZER/LVL
[2] http://irsa.ipac.caltech.edu/data/Herschel/VNGS/overview.html
[3] http://svo2.cab.inta-csic.es/theory/fps3/

Figure 1. The processed frames for the three galaxies. By row: M83 (top), M101 (centre), and M77 (bottom). By columns: H α images taken with the TTF at the WHT, in $\mathrm{erg\,s^{-1}cm^{-2}}$ (left), FUV images from GALEX survey, in $\mathrm{erg\,s^{-1}cm^{-2}\,\mathring{A}^{-1}}$ (centre), and MIPS/SPITZER infra red images, in units of $\mathrm{erg\,s^{-1}cm^{-2}}$ (right). Images have been resampled to have identical size, orientation, and pixel scale (1.5 arcsec pixel^{-1}). The coordinates and axis scales are given in kpc, with respect to the galaxy centre (North up, East left).

$f_{Q(H)} = 1 - f_{\mathrm{scp}}$ to account for the escape of ionizing photons, f_{scp} (e.g. Mas-Hesse & Kunth 1991), hence our conversion formula is $L(\mathrm{H}\alpha) = 1.36 \times 10^{-12}\, Q(H)\,(1 - f_{\mathrm{scp}})$. Simialry, we multiply the nebular component in FUV/GALEX by $(1 - f_{\mathrm{scp}})$ before adding it to the stellar component.

As a result, we obtain the isochrone table used by SB99 including, for each star, the H α flux, the nebular contribution to the FUV/GALEX luminosity, the stellar FUV/GALEX luminosity, and the contribution to the total luminosity (i.e., its IMF-weighted value). Isochrones are computed in the age interval from 0.1 to 20 Myr with a linear time-step of 0.1 Myr for the Geneva evolutionary

track set. We use standard mass loss rates for metallicity values of 0.040, 0.020, 0.008, 0.004, and 0.001, corresponding to the evolutionary tracks presented in Schaller et al. (1992), Charbonnel et al. (1993) and Schaerer et al. (1993a,b). Stellar libraries are from BASEL (Lejeune, Cuisinier & Buser 1997) for intermediate and low-mass stars, and from Smith, Norris & Crowther (2002) for massive and Wolf–Rayet (WR) stars when present in the cluster. The code was run using a Salpeter (1955) IMF slope within the mass range of $1 - 120\,\mathrm{M_\odot}$ with a total mass of $3.14\,\mathrm{M_\odot}$, which is the mean stellar mass in such IMF interval. In this way, our results are naturally normalized to the total number of stars instead of the total

mass, a requirement to compute properly the theoretical covariance matrix.

The resulting components in each isochrone table were added following prescriptions by Cerviño & Luridiana (2006) to obtain the mean values, variances (expressed as effective number of stars, $\mathcal{N}_{\rm eff}$, where $1/\mathcal{N}_{\rm eff} = (\sigma/\mu)^2$, see Cerviño et al. 2002), skewness (γ_1), and kurtosis (γ_2) of the stellar luminosity functions of $L({\rm H}\alpha)$ and $L({\rm FUV})$, and the covariance between both luminosities.[4] We also obtained the Lowest Luminosity Limit (LLL) (Cerviño et al. 2003; Cerviño & Luridiana 2004) for H α and FUV, which gives the luminosity of the brightest individual star in the model for the given band; i.e. an observed cluster less luminous than the LLL could not be modelled by the mean value of the ensembles obtained by SSPs models, since there is a confusion between the emission of the ensemble and the emission of a single star, and the overall probability distribution function (PDF) of the theoretical integrated luminosity, and not just its mean value, must be taken into account (Cerviño & Luridiana 2004).

The resulting evolution of the H α/FUV ratio mean values is shown in Fig. 2, where we have tested that the mean fluxes and the ratio for the case of $f_{\rm scp} = 0$ ($f_{Q(H)} = 1$) are coincident with SB99 results without modifications, which are also coincident with the parameters used in Sánchez-Gil et al. (2011) with slight modifications due to the variation of stellar atmosphere models.

The right-hand panel of Fig. 2 shows the resulting theoretical correlation coefficient between Hα and FUV. The correlation coefficient is almost one up to 3 Myr, reflecting the fact that the same type of stars dominate both quantities. The correlation decreases abruptly during the Wolf–Rayet phase (WR) of the cluster, since Hα is dominated by the WR stars, but the FUV is mainly produced by hot stars in the Main sequence. After the WR phase in the cluster, the correlation coefficient increases to values around 0.6. It increases with metallicity, which depends on the balance between main sequence stars which still produce ionizing fluxes (O–B stars mainly) and those that only produce FUV flux (B–A spectral types). The metallicity dependence is explained by the variation of the main sequence with metallicity: at low metallicity, the main sequence is hotter, hence FUV fluxes are produced by both ionizing and non-ionizing stars and the correlation coefficient is lower with respect to larger metallicities where FUV and ionizing stars are coincident. Finally, there is a small variation of the correlation coefficient with $f_{\rm scp}$ since the escape of photons only affects the correlation coefficient by the nebular contribution to the FUV flux.

Fig. 3 shows the variance (expressed in terms of effective number $\mathcal{N}_{\rm eff}$ of stars), skewness (γ_1), and kurtosis (γ_2) obtained from theoretical models normalized to $N_* = 1$. We recall that the effective number of stars scales with the number of stars; the skewness scales with the inverse of the square root of the number of stars, and the kurtosis scales with the inverse of the number of stars (Cerviño & Luridiana 2006). As reference, a relative dispersion of 4 per cent requires that the used resolution element contains 6.25×10^4 stars; in that case, $\gamma_1 = 0.08$ and $\gamma_2 = 0.008$ (assuming values of $\mathcal{N}_{\rm eff} = 0.01$, $\gamma_1 = 20$, and $\gamma_2 = 500$ when normalized to the number of stars). Such values imply that the underling dis-

tribution can be well represented by a Gaussian. In the case of a relative dispersion of 20 per cent, only 2500 stars are required, but the values of γ_1 and γ_2 increase to 0.4 and 0.2, respectively, and the underling distribution deviates from Gaussian (see Appendix A for details).

The use of a pixelwise age dating technique allows age mapping of the youngest stellar population without prior assumptions about the spatial distribution of the star-forming regions. It also provides a spatial characterization of the age distribution for HII regions in galaxies within the local volume through their spatially resolved spiral arms or other galactic structures. However, this pixel-based technique is subject to some systematic effects, including (i) the pixel-sized luminosities, and (ii) potential interaction between surrounding regions through adjacent pixels.

The model validity was checked using a single stellar population model in pixel-sized regions; it was checked by comparing our pixel-sized FUV luminosity with the possible LLL values in our age range. We note that a similar test for $L({\rm H}\alpha)$ depends on the factor $f_{\rm scp}$, so the LLL test is not decisive in this case. The result is shown in Fig. 4, where the left-hand panel shows the value of the LLL for $L({\rm FUV})$ at the five metallicities; it also includes, as reference, the mean FUV integrated luminosity for $N_* = 10^4$. The middle panel shows the histogram of pixel $L({\rm FUV})$ for the three galaxies, using the distances quoted in Table 1. Except for most pixels of M77, all the pixel-based values have luminosities below the LLL in FUV, for ages between 0 and 10 Myr (17 Myr for the lower metallicity). The right-hand panel shows the distribution of the pixel-based regions in the H α/FUV versus $L({\rm FUV})$ plane. The colour lines show the position of the mean values obtained with the synthesis models for the case of 10^4 stars. The black line shows the region covered by individual normal stars and the grey line the location of individual WR stars. We note that any pixel-sized region with $2 \leq \log({\rm H}\alpha/{\rm FUV}) \leq 2.4$ requires the presence of WR stars, and that pixel-sized regions with $\log({\rm H}\alpha/{\rm FUV}) > 2.4$ cannot be explained with synthesis models and require some extra ionization source. The extreme case that would be explained by the models are regions with $\log({\rm H}\,\alpha/{\rm FUV}) \sim 2.5$, which would correspond to regions completely dominated by nebular emission without stellar content, hence age determinations would not be performed. In M 83, those regions with an excess in this flux ratio actually correspond with those holding WR stars detected by Kim et al. (2012).

Although we are clearly aware of the IMF sampling issues discussed in Cerviño et al. (2003), Cerviño & Luridiana (2004, 2006), and Cerviño (2013), we apply a quasi 'standard' SSP analysis to the age estimation. We calculate model H α and FUV fluxes, ages, masses, etc. assuming that the IMFs, even though well populated, contain an intrinsic scatter and that the integrated H α and FUV luminosities can be modelled by a 2D-Gaussian distribution, with the correlation coefficient described previously and a fixed standard deviation of 4 per cent for each theoretical luminosity. We have also computed the results for a standard deviation of 20 per cent and this does not alter significantly the main results. Although not perfect, given the failure of the Gaussian approximation at larger standard deviations, it is a first step in the inclusion of sampling effects in the modelling of stellar clusters. A detailed discussion of the relevance of sampling for this work is given in Appendix A.

We also examined the effect of adjacent pixels, to verify whether the amount of H α and FUV fluxes in a given pixel reflect the number of ionizing O and B stars in that pixel. In Paper I, we explored a range of different spatial binning scales. These results showed that age structures and gradients remain the same irrespective of the binning scale used (cf. Fig. 6 of Paper I), confirming the

---

[4]Note in papers previous to Cerviño & Luridiana (2006) the correlation coefficient was obtained under the incorrect assumption of Poissonian statistics in each isochrone component, instead of multinomial statistics. For the correct formula to compute variances and covariances between H α and FUV/GALEX luminosities for each possible case, see Cerviño & Luridiana (2006) and Cerviño (2013).
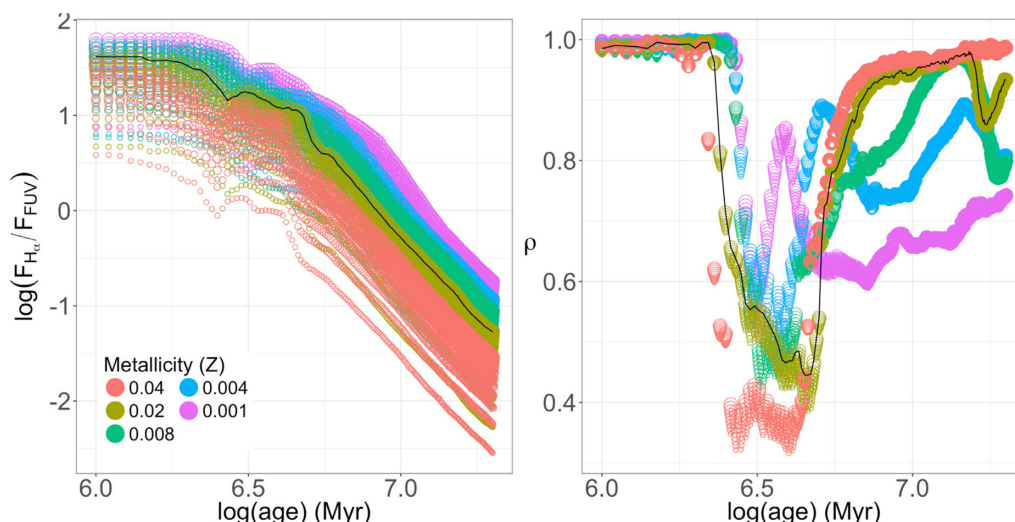
Figure 2. Models from SB99. On the left, the model H α/FUV ratio versus age, ranging from 1 to 20 Myr, in log units. On the right, the theoretical correlation between H α and FUV luminosities versus age. Both relationships are given for different combinations of the parameters: The metallicity Z is represented with different colours, from $Z = 0.001$ (purple points) to $Z = 0.04$ (in red). The fraction of ionizing photons, $f_{Q(H)}$, is coded with different point sizes, ranging 10–100 per cent from the smallest to the largest. The black line in both plots represents these model relationships for solar metallicity, $Z = 0.02$, and $f_{Q(H)} = 100$ per cent.
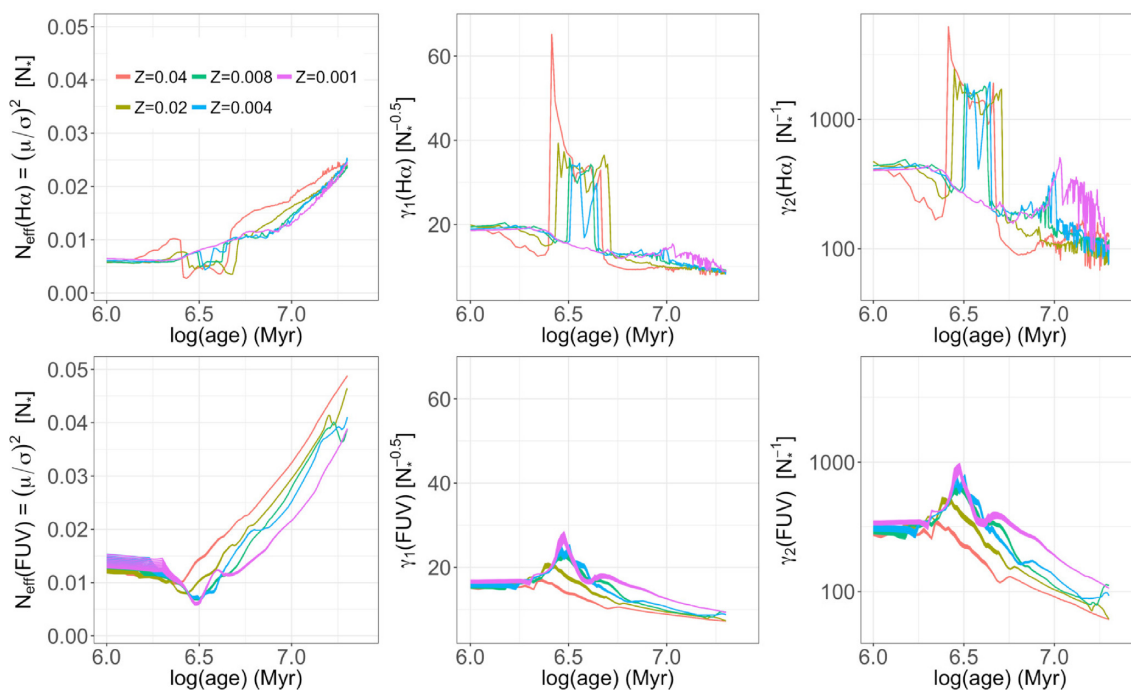

Figure 3. Variance expressed in terms of effective number of stars (left), skewness (middle), and kurtosis (right) obtained from theoretical models. Top panels show the values for H α and bottom panels show the values for FUV as a function of the age. Metallicity follows the same colour code as Fig. 2.

robustness of measurements to the effects of binning. In Section 4.2, we present a different approach to deal with the possible influence of adjacent pixels, as well as the low-mass pixels below the LLL, under a Bayesian inference framework. We apply an image segmentation technique, to account for the possible effects of the spatial dependence in terms of adjacent regions (or pixels), grouping together neighbouring pixels that carry, on average, the same true value of the observed measured quantity (H α/FUV flux ratio). In this manner, we can address both systematic effects. Not only the dependence between adjacent regions but also the pixelwise appli-

cability of SSP models, since many of the grouped larger regions are now above the LLL threshold.

A similar approach can be found in Casado et al. (2017). These authors implement a Bayesian Technique for Multi-image Analysis (BATMAN), focused on Integral Field Spectroscopy (IFS) data cubes. Unlike our fully Bayesian approach, BATMAN's algorithm is rather a Bayesian approximation. The parameter estimation, carried out as usual, requires the selection of the most probable tesselation of the image performed by means of an iterative procedure to select the best model. Another significant difference is that the presence
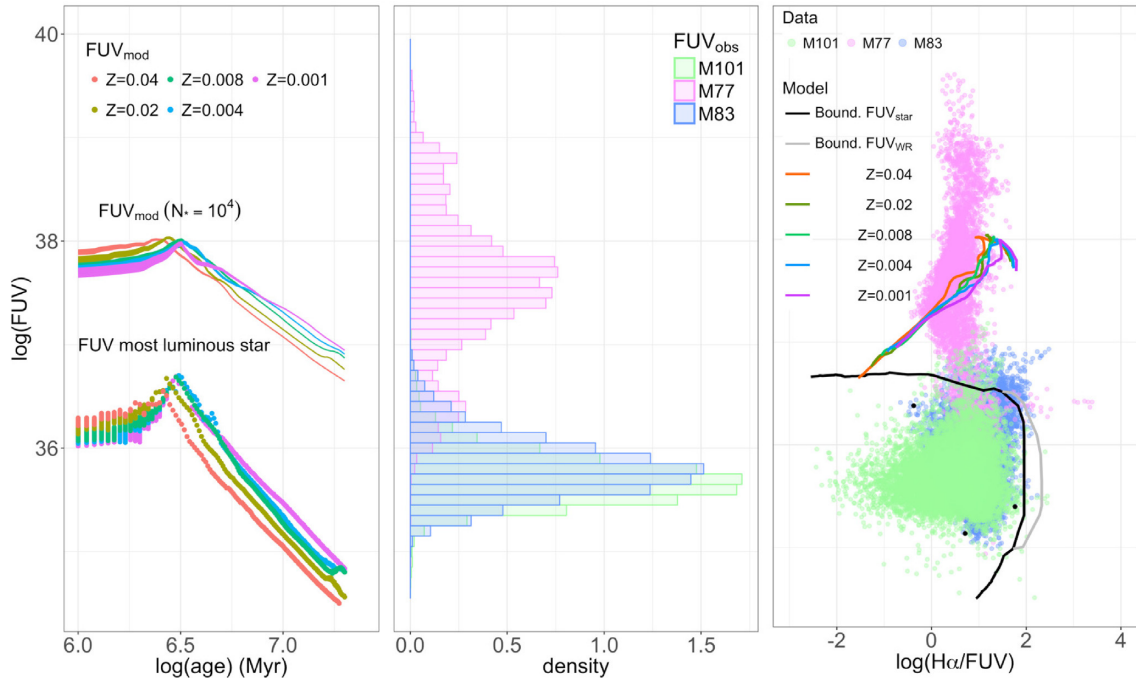
Figure 4. Left: Time evolution of the FUV LLL for different metallicities; colours as in Fig. 2. The mean FUV integrated luminosity for 10^4 stars is also shown for reference. Centre: Histograms of the FUV luminosities of the pixel-sized data of the three galaxies. Right: distribution of the pixel-sized data in the H α/FUV versus L(FUV) plane. Solid black and grey lines show the boundaries of individual normal and WR stars, respectively in this plane. Data points in between both lines correspond to pixel-sized values that can be only explained by the presence of WR stars; data points with ratio larger than the boundary of the WR line require some extra ionization source since they are not compatible with the current models. The plot also shows the position of H α/FUV and the mean FUV integrated luminosity and for 10^4 stars for the case of $f_{\rm scp} = 0$ as reference. Finally, the three black point correspond to the positions of the pixels shown in Fig. 6.

of gradients poses a challenge to BATMAN's algorithm, since its segmentation model does not consider the presence of gradients inside the regions. In our case, the presence of age gradients, which are indeed expected, do not seem to affect the image segmentation.

A limitation of image segmentation techniques is the loss of spatial resolution. More complex and elaborate Bayesian models, such as latent Gaussian models for spatial modelling can be found (Rue et al. 2017). Not only the spatial correlation structure is estimated, with no loss of image resolution, but also some fixed effects and Gaussian random effects can be included in the linear predictor of the model fitting. However, these models are computationally challenging and time-consuming. An approximation for a faster inference is implemented in the R-INLA[5] package of R; however, it has a steep learning curve, requiring a deeper understanding of both the Bayesian and the latent Gaussian models for a proper implementation. An example of its successful application to IFU spectra can be found in González-Gaitán et al. (2018). This type of data encodes some autocorrelated structures, and it is very important to evaluate the spatial information. Besides, the loss of spatial resolution in these cases is quite significant (see their Fig. 4 for a qualitative comparison with other image segmentation and spatial algorithms methods). This is not an issue with our images, that have high-enough spatial resolution to resolve individual star-forming structures within spiral arms, even after image segmentation. See bottom left panels of Figs. 9, 11, and 13. In fact, the morphology of the segmented images traces the same global and local spatial patterns as the pixel-based age maps.

Finally, we want to stress that there is an additional fundamental difference between our approach and those using IFU data (Casado et al. 2017; González-Gaitán et al. 2018 which also include a Bayesian analysis, or de Amorim et al. 2017, which use a Voronoi technique). In the case of IFU data, a major issue is the requirement of increasing the low *S/N* of some spaxels and avoid data sparsity; in our case, we do not have such a problem, since photometric data have a larger *S/N*, and those pixels below a certain H α flux level are considered to belong to the background, so they are masked in the resulting maps.

In brief, Section 4.2 proposes a simple and fast fully Bayesian approach to take into account the spatial dependence by means of image segmentation. We will see that the loss of spatial resolution is not important, in part due to the spatial characteristics of the data. We will show how this method provides good results within the aim of this work, the global determination of the age patterns along and across the spiral arms.

4 A BAYESIAN FRAMEWORK FOR MODELLING THE RATIO IMAGES

4.1 Hierarchical Bayesian Model

In this section, we address the problem of deriving the galaxy age map from observed flux ratio images, namely $\hat{r} = \hat{F}_{\rm H\alpha}/\hat{F}_{\rm FUV}$, by establishing a probabilistic framework relating the random variables involved in the problem. These relationships will be formulated in terms of a joint probability distribution, given the observations and their uncertainties through a hierarchical Bayesian model (HBM; Gelman et al. 2003). Specifically, we want to describe the probabil-

M.C. Sánchez-Gil et al.
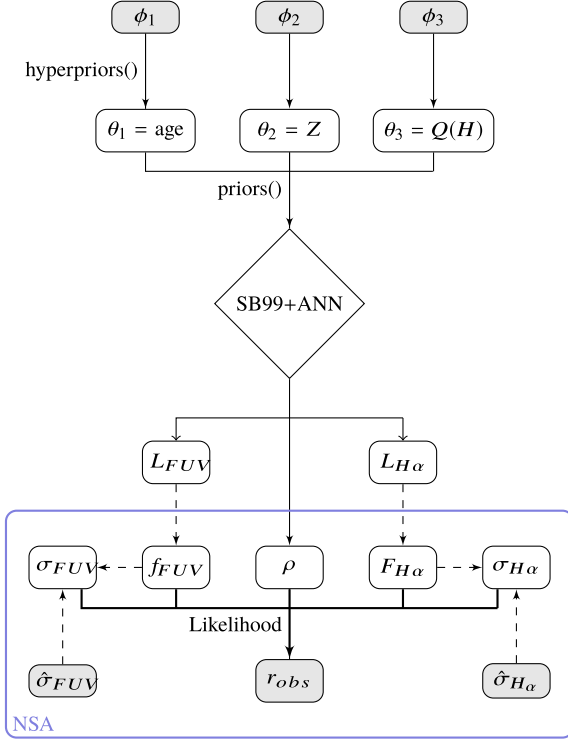
Figure 5. The hierarchical Bayesian model in plate notation. Blank shaded nodes represent random fixed (or observed) values, respectively. Arrows represent the kind of relationship between variables: probabilistic (solid arrows) and deterministic (dashed arrows). In the central part of the graphical model, we show deterministic calculations done by SB99 scripts and the artificial neural network (ANN). The blue square at the bottom, encircles the variables involved in the Likelihood (7), as well as the observed ratio which are the input for the NSA.

ity distribution of age given $\hat{r}$, which will be derived by marginalization as we will see below. This procedure or algorithm is applied pixel by pixel throughout the image, keeping the spatial resolution of the flux ratio images, see top right panels of Figs. 9, 11, and 13.

Fig. 5 shows our HBM in plate notation, namely a graphical model representing the previous relationships. Nodes in the graphical model are circles and squares representing random variables and fixed values, respectively. Inside the circles and squares there are numbers representing the dimension of values in each case, and a shaded circle (square) means that the variable has been observed. Arrows represent the relationships (solid arrows probabilistic, dashed arrows deterministic) between variables.

Let $\mathcal{H}$ represent the set of all population synthesis parameters held fixed, such as the SFH (instantaneous), IMF (see Salpeter 1955), evolutionary tracks, as well as the extinction correction applied to the data. And let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ be the parameters to be estimated: θ_1 the age of the region under study (the image pixel), θ_2 the metallicity, and θ_3 the fraction of ionizing photons. Every parameter θ_i is connected with the hyperparameter ϕ_i in the graphical model of Fig. 5. That is, there is a probability distribution once the ϕ_i is fixed. Specifically we set uniform prior distributions for all the parameters, $\theta_i \sim \mathcal{U}(\phi_i)$, where $\phi_1 = (1, 20)$, corresponding to the age (θ_1) ranging from 1 to 20 Myr; $\phi_2 = (0.001, 0.02)$, for the metallicity (θ_2). SB99 only covers five metallicity values, i.e. $^1/_{20}Z_\odot = 0.001$, $^2/_5 Z_\odot = 0.008$, $^1/_5 Z_\odot = 0.004$, $Z = Z_\odot = 0.02$

(solar) and $2Z_\odot = 0.04$; and $\phi_3 = c(0.1, 1)$, the fraction of ionizing photons $Q(H)$ (θ_3) ranging from 10 per cent to 100 per cent. This is, priors for $\boldsymbol{\theta}$ were set as

$$\theta_1 \sim \mathcal{U}(0.1, 20)$$
$$\theta_2 \sim \mathcal{U}(0.001, 0.04)$$
$$\theta_3 \sim \mathcal{U}(0.1, 1) \tag{1}$$

Let us notice that we deal with a finite set of values for θ_i, and therefore a finite number of model flux ratios derived from them. See left-hand panel in Fig. 2.

As described below, our main objective is to obtain a posterior distribution of the parameters θ_i, using an iterative algorithm in our HBM. The main consequence of using SB99 scripts and the iterative algorithm is the increase of CPU time. In order to obtain a continuous range for the model flux ratios or parameters, rather than such discrete set of θ_i values, we set up an Artificial Neural Network (ANN) to interpolate the grid of parameters. A thorough introduction to ANN can be found in Haykin (1999).

Specifically, our ANN is a *multilayer perceptron* with four layers: two hidden layers with 30 and 50 nodes, respectively. The input layer for the parameters $\boldsymbol{\theta}$ and the output layer for the Hα and FUV luminosities (see Fig. B1). This topology was selected by doing cross-validation through repeated random sub-sampling validation (90 per cent of the dataset for training and 10 per cent for validating the ANN) and assessing the fit using the mean squared error (MSE).

Fig. 5 shows the use of the ANN by dashed arrows linking the parameters θ_i with H–α and FUV luminosities (i.e. a deterministic relationship characterized by the ANN). These dashed arrows also include the calculations to obtain the modelled H–α and FUV fluxes, and therefore the model flux ratio $r = F_{\text{H}\alpha}/F_{\text{FUV}}$ for a specific pixel, according to

$$F_{\text{H}\alpha} = 4\pi D^2 L_{\text{H}\alpha}, \ (\text{erg s}^{-1}\,\text{cm}^{-2}) \tag{2}$$

$$f_{\text{FUV}} = 4\pi D^2 L_{\lambda\text{FUV}}, \ (\text{erg s}^{-1}\,\text{cm}^{-2}\,\text{Å}^{-1}) \tag{3}$$

with the same units as the observed data.

In order to explain completely the HBM, let us focus on the right-hand side of Fig. 5, showing the relationships between unknown r and observed $\hat{r}$ flux ratios. We apply this technique pixel by pixel (as described in Paper I), assuming a bivariate normal distribution for the observed fluxes as a result of the convolution between the observational/model uncertainties and the unknown fluxes. Specifically,

$$(\hat{F}_{H\alpha}, \hat{f}_{\text{FUV}}) \sim \mathcal{N}((F_{H\alpha}, f_{\text{FUV}}), \Sigma), \tag{4}$$

where Σ is defined according to

$$\Sigma = \begin{pmatrix} \sigma_{H\alpha}^2 & \rho\sigma_{\text{mod},H\alpha}\sigma_{\text{mod,FUV}} \\ \rho\sigma_{\text{mod},H\alpha}\sigma_{\text{mod,FUV}} & \sigma_{\text{FUV}}^2 \end{pmatrix}. \tag{5}$$

Here, ρ is the theoretical correlation between the Hα and FUV fluxes; and $\sigma^2 = \hat{\sigma}^2 + \sigma_{\text{mod}}^2$ accounts for both the observational and the model uncertainties.

Therefore, $\hat{r} = \hat{F}_{H\alpha}/\hat{F}_{\text{FUV}}$ is the ratio of two correlated normal random variables, whose exact distribution is given by Hinkley (1969) as

$$\psi(r) = \frac{b(r)d(r)}{\sqrt{2\pi}\sigma_{H\alpha}\sigma_{\text{FUV}}a^3(r)} \left(2\Phi\left(\frac{b(r)}{\sqrt{1-\rho^2}a(r)}\right) - 1\right) +$$
$$+ \frac{\sqrt{1-\rho^2}}{\pi\sigma_{H\alpha}\sigma_{\text{FUV}}a^2(r)} \exp\left\{-\frac{c}{2(1-\rho^2)}\right\}, \tag{6}$$

where r is the model flux ratio, and parameters $a(r)$, $b(r)$, c, and $d(r)$ are defined as

$$a(r) = \left(\frac{r^2}{\sigma_{H\alpha}^2} - \frac{2\rho r}{\sigma_{H\alpha}\sigma_{FUV}} + \frac{1}{\sigma_{FUV}^2} \right)^{1/2}$$

$$b(r) = \frac{r F_{H\alpha}^2}{\sigma_{H\alpha}^2} - \frac{\rho(F_{H\alpha} + r f_{FUV})}{\sigma_{H\alpha}\sigma_{FUV}} + \frac{f_{|r,,FUV}}{\sigma_{FUV}^2}$$

$$c = \frac{F_{H\alpha}^2}{\sigma_{H\alpha}^2} - \frac{2\rho F_{H\alpha} f_{FUV}}{\sigma_{H\alpha}\sigma_{FUV}} + \frac{f_{FUV}^2}{\sigma_{FUV}^2}$$

$$d(r) = \exp\left\{ \frac{b^2(r) - ca^2(r)}{2(1 - \rho^2)a^2(r)} \right\}.$$

Here, ρ is the correlation between $F_{H\alpha}$ and f_{FUV} and Φ is the cumulative density distribution of the standard normal. We recall that the values assumed for $\hat{\sigma}$ are 5 per cent for $F_{H\alpha}$ and 25 per cent for $\hat{f}_{FUV}$ (as shown in Section 2), and that σ_{mod} has been taken as 4 per cent in both cases (see Section 3 and Appendix A). Then, we can define the probabilistic relationship between unknown and observed flux ratios according to the likelihood

$$p(\hat{r}|\boldsymbol{\theta}, \mathcal{H}) = \psi(\hat{r}|F_{H\alpha}, f_{FUV}, \sigma_{H\alpha}, \sigma_{FUV}, \rho) \qquad (7)$$

Fig. 5 explains how we obtain observed ratios from parameters $\boldsymbol{\theta}$ through an HBM by using the graphical model. However our main objective is to move in the reverse order, i.e. to obtain suitable parameters $\boldsymbol{\theta}$ from the observed ratios $\hat{r}$.

The joint posterior probability distribution $p(\boldsymbol{\theta}|\hat{r})$ can be rewritten by using Bayes' theorem according to

$$p(\boldsymbol{\theta}|\hat{r}) = \frac{p(\hat{r}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\hat{r})} \propto p(\hat{r}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \qquad (8)$$

where $p(\hat{r})$ is a normalization constant, which for our purposes can be ignored. We will assume independence between parameters θ_i, or equivalently, the prior distribution $p(\boldsymbol{\theta}) = p(\theta_1)p(\theta_2)p(\theta_3)$, where $p(\theta_i)$ is uniform distribution as defined above. The likelihood, $p(\hat{r}|\boldsymbol{\theta})$, has been defined previously in equation (7).

In the Bayesian framework, inference proceeds by estimating the posterior distribution $p(\boldsymbol{\theta}|\hat{r})$ and then marginalizing to obtain the age distribution of the region under study according to

$$p(age|\hat{r}) = p(\theta_1|\hat{r}) = \iint p(\boldsymbol{\theta}|\hat{r}) \, d\theta_2 d\theta_3, \qquad (9)$$

where the posterior distribution $p(\boldsymbol{\theta}|\hat{r})$ was characterized by an independent and identical distributed sample obtained by using the Nested Sampling Algorithm (NSA; Skilling 2006). Although we refer to these authors, a brief summary is given in Appendix B.

Finally, we obtain the posterior distribution of the parameter of interest (age) by marginalizing nuisance parameters in the posterior distribution as set in equation (9). Marginalized distributions for the other two parameters can be obtained in the same way.

Fig. 6 shows an example of the resulting posterior distributions. For the case of M83, we have chosen three pixels of different estimated ages (1.2, 5.5 and 10 Myr) and plotted (left to right, respectively) the histograms of the samples from the marginalized distributions.

On the top, we have the marginalized posterior probability distributions of the age parameter. From an initial, non-informative, uniform prior $\mathcal{U}(0.1, 20)$ Myr, we can observe how well defined is the estimated posterior distribution. In general, the median or the (uni)modal values are close from each other, close to symmetrical distributions, and well differentiated. As mentioned in Section 1, the ratio Hα/FUV is very sensitive to age variations for young SF regions.

The middle and bottom panels of Fig. 6 show the marginalized posterior distributions of the metallicity and the fraction of ionizing photons, respectively. The assumed priors for these parameters were also uniform. However, it is remarkable that their posterior distributions are too spread and flat, remaining nearly uniform in many cases. This result should indicate that this method is not able to determine these parameters. In fact, the metallicity is a parameter quite difficult to estimate.

In order to check how much the selection of priors in these other properties may influence the age estimate, we compute the posterior probability density function $p(\boldsymbol{\theta}|\hat{r})$ for different priors. The results, for a given flux ratio of $\hat{r} = 1.13$ (a young region), are shown in Fig. 7. The first column corresponds to the uniform priors assumed along this work. In the second column only the prior on metallicity is not uniform, $\theta_2 = Z \sim \mathcal{N}_{(0,\infty)}(0.02, 0.01)$ truncated Normal, negative values are not allowed. Similarly, in the third column for the fraction of ionizing photons, $\theta_3 = f_{Q(H)} \sim B(5, 2)$ Beta distribution, assuming higher values of $f_{Q(H)}$ to be more likely. The fourth column shows the results when both priors are not uniform distributed. For other flux ratio values or different prior selections, e.g. $\theta_2 \sim B*(2, 2)$, or $B*(5, 2)$ (Beta distribution transformed from the unit interval [0,1] to [0.001,0.04]), $\theta_3 \sim B(2, 2)$, or $U(0.5, 1)$, etc, the results were similar.

Fig. 7 shows that, except for the age, the posterior PDF of the other parameters is dominated by the prior. Besides, the influence of the selection of the prior distribution in these other parameters is almost negligible on the age estimate. The mode (and median) of the different age posterior PDFs are pretty close. In fact, it seems that adding some prior information results in narrower and taller posterior density functions around the mode.

The three last rows display how correlated are the posterior marginals for the different parameters. Note, how the age correlates with the fraction of ionizing photons and the metallicity, whereas these two parameters are not only nearly uncorrelated but nearly independent, as well as their priors. These parameters are not determined by the Hα/FUV ratio. Therefore, their posteriors remain dominated by the priors distributions assumed, as it is clearly seen in the figure. Any posterior sample occupies nearly the whole space of parameters, except for the age. We confirm that this methodology, based on the Hα/FUV flux ratio, is robust and efficient for dating H II regions but not for determining other parameters such as the metallicity.

4.2 Image Segmentation

In the following, we describe an image segmentation technique to assess the possible effects of spatial dependence in terms of adjacent regions (or pixels), grouping together as a single region those neighbouring pixels that carry the same average true value of the measured quantity.

In Paper I, this was achieved by calculating the age maps after re-scaling the Hα/FUV flux ratio images at 3×3 or 6×6 pixel binning. We found that not only the main structures and global age patterns remain unchanged, but also some local age gradients. This level of pixel bining is in agreement with the results from several variograms computed in M83, where we found that the level of spatial dependence is roughly under bins of 5×5 pixels.

Both segmentation techniques imply a loss of resolution, but it is worst in the rebinning case. The proposed image segmentation technique, based also on a hierarchical Bayesian approach, is an improvement compared to re-binning the images as in Paper I.
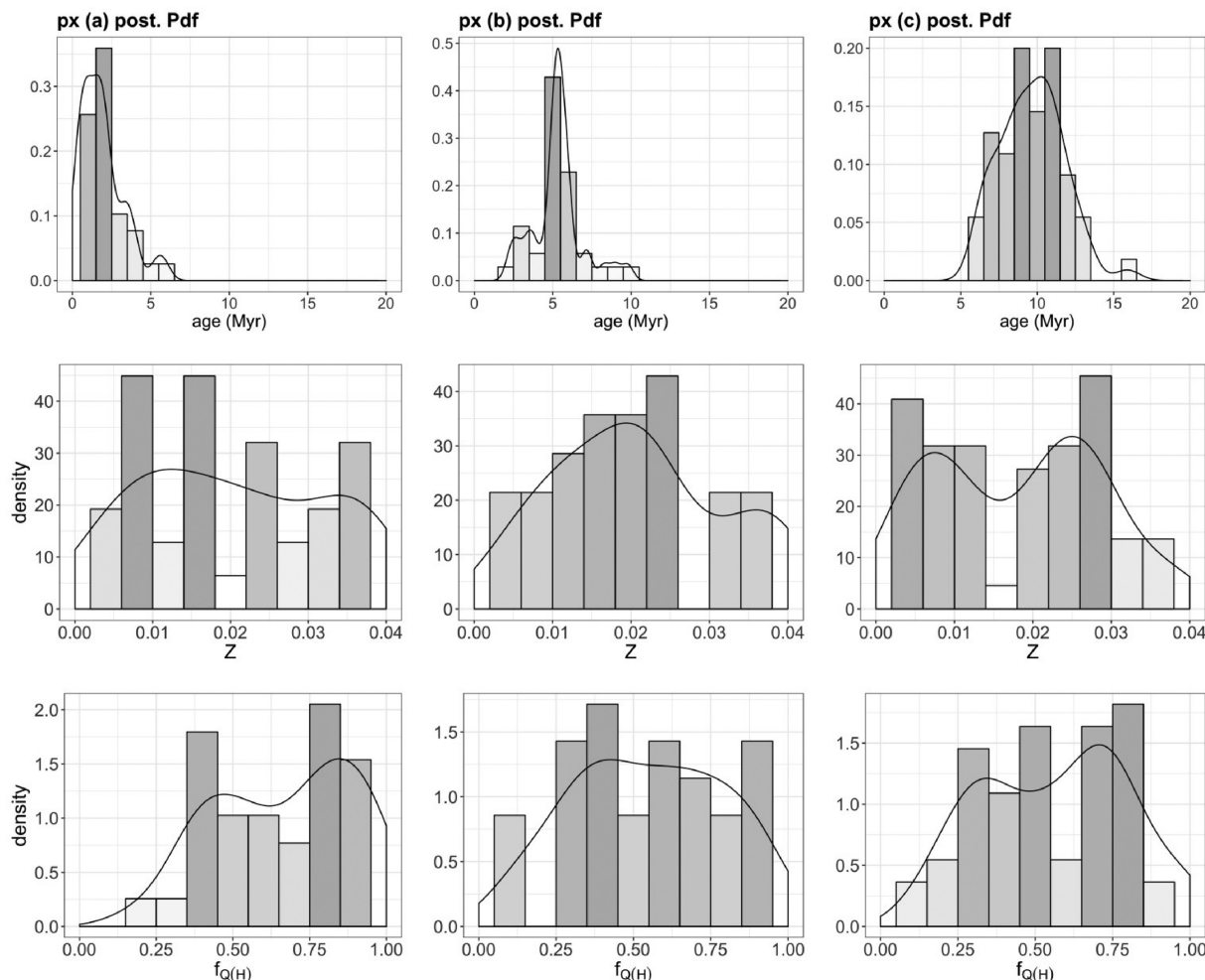
Figure 6. Example marginal posterior probability distributions of different parameters for three image pixels of different age. The black lines denote kernel density estimates. The scale fill gradient correspond to the density (from white lower to grey higher). These plots highlight the sensitivity of the proposed methodology for the age estimation, unlike other parameters.

We segment the Hα/FUV ratio image, in terms of homogeneous values, to model the effects of the spatial dependence (in terms of adjacent pixels). We work with regions resulting from clustering several pixels. This approach is driven by the assumption that pixels with similar Hα/FUV ratios will share similar inferred properties. Segmentation maps serve to identify structures sharing common properties relevant to the interpretation of the age map.

The purpose of image segmentation is to cluster pixels into homogeneous classes, without prior definition of those classes, based only on spatial coherence. We present a fully Bayesian approach, based on the Potts model for the image reconstruction (Marin & Robert 2014, Chapter 8).

Consider the 'true' image as a random bidimensional array $x = \{x_i, i \in \mathcal{I}\}$ whose elements are indexed by the lattice $\mathcal{I}$, the location of the pixels, and related through a neighborhood relation. The four nearest neighbors of the i−th pixel (m, n) are $(m, n − 1)$, $(m, n + 1)$, $(m − 1, n)$, and $(m + 1, n)$ respectively, denoted as $j \sim i$.

This neighborhood relation is translated into a probabilistic dependence by means of Markov Random Field (MRF), where each x_i takes a finite set of values. The conditional distribution of any pixel $i \in \mathcal{I}$, given the rest of the pixels of the image, depends only on the values of their neighbors, denoted by $n(i)$, $p(x_i|x_{-i}) = p(x_i|x_{n(i)})$.

Denote the observed flux by y, considered 'noisy' in the sense that the measured flux of a pixel is not observed exactly but with some perturbation (instrument noise, reduction process, etc). Both objects x and y are arrays, with each entry of x taking a finite number of values, for numerical convenience. Each entry of y takes real values.

The aim is to draw inference on the 'true' image x, given an observed noisy image y. We are thus interested in the posterior distribution of x given y, provided by Bayes' theorem

$$p(x|y) \propto f(y|x)p(x). \tag{10}$$

We assume a Potts model for the prior on x, a specific family of distributions inspired from particle physics in order to structure images and other spatial structures in terms of local homogeneity (Wu 1982).

$$p(x|\beta) = \frac{1}{Z(\beta)} exp \left\{ \beta \sum_{j \sim i} \mathbb{1}_{x_j = x_i} \right\} \tag{11}$$

where $Z(\beta) = \sum_{i \in \mathcal{I}} exp \left\{ \beta \sum_{j \sim i} \mathbb{1}_{x_j = x_i} \right\}$ is the normalizing constant of the Potts model with G categories.
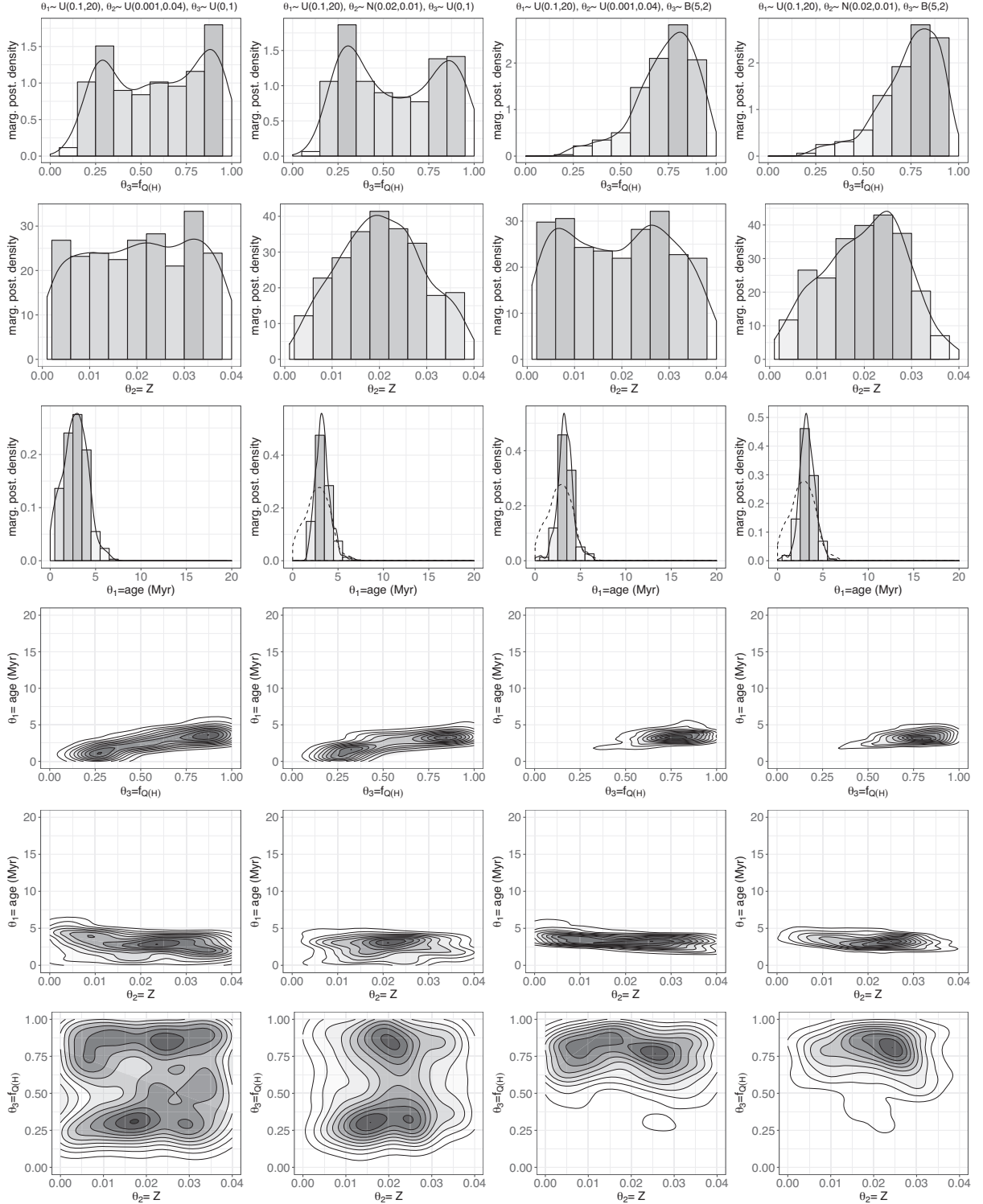
Figure 7. Posterior marginals and pairwise correlations for the three parameters, given different prior selections and a flux ratio $\hat{r} = 1.13$. The first column corresponds with uniform priors for the three parameters, $\theta_1 \sim \mathcal{U}(0.1, 20)$, $\theta_2 \sim \mathcal{U}(0.001, 0.04)$, and $\theta_3 \sim \mathcal{U}(0.1, 1)$ At the second column only the prior on metallicity is not uniformed distributed, $\theta_2 = Z \sim \mathcal{N}_{(0,\infty)}(0.02, 0.01)$, i.e. a truncated Normal to non negative values. While at the third column $\theta_3 = f_{Q(H)} \sim B(5, 2)$, a Beta distribution assuming highest values of $f_{Q(H)}$ to be more likely. Last column shows the results when both priors are not uniformly distributed. The first three rows are the same as in Fig. 6: the black solid lines denote the corresponding kernel density estimates. The dashed lines denote the density when all priors are uniformly distributed, for comparison purposes. The three last panels rows display how correlated are the posterior marginals for the different parameters.

The likelihood $f(\boldsymbol{y}|\boldsymbol{x})$ describes the link between the observed image and the underlying classification of homogeneous flux ratios. That is, it gives the distribution of the noise. We will make the assumption that the observations in $\boldsymbol{y}$ are conditionally independent of $\boldsymbol{x}$ and Gaussian,

$$f(\boldsymbol{y}|\boldsymbol{x}, \sigma, \mu_1, \ldots, \mu_G) = \prod_{i \in \mathcal{I}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu_{x_i})^2\right\}.$$

(12)

For a fully Bayesian approach, we have to give the distribution of the hyperparameters β, σ, and $\mu_1, \ldots, \mu_G$, respectively. Since there is no additional information about any of these nuisance parameters we assume uniform and independent hyper priors.

The Potts model parameter β, which represents the strength between neighbouring pixels, is uniform distributed over $\beta \sim \mathcal{U}(0, 2)$, (e.g. Marin & Robert 2014; Stoehr 2017, and references within). Above this critical value $\beta_c = 2.269$ (for the case of four neighbour relation) the Markov chain is no longer irreducible, converging to one of two different stationary distributions, depending on the starting point. The distribution becomes multimodal, known as phase transition in particle physics.

The mean value of each homogeneous class $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_G) \sim \mathcal{U}(\boldsymbol{\mu}; y_{min} \le \mu_1 \le \ldots \le \mu_G \le y_{max})$. This is a generalization of the treatment for an image where y represents its noisy version of the true colour or grey level (and not necessarily an integer).

The recorded values of $\boldsymbol{y}$ represent the ratio of the Hα and FUV fluxes (e.g. in the range $(-2, 3)$ for M83 example). We classify this ratio image into $G = 5$ homogeneous regions or classes, with mean μ_g. The number of classes is inspired in the previous work in Paper I.

For the noise variance it is assumed an uniform prior on $\log(\sigma)$, or equivalently $p(\sigma^2) \propto \sigma^{-2}$.

Finally, the posterior distribution for $\boldsymbol{x}$ is

$$P(x_i = g|\boldsymbol{y}, \beta, \sigma^2, \boldsymbol{\mu}) \propto exp\left\{\beta \sum_{j \sim i} \mathbb{1}_{x_j = x_i} - \frac{1}{2\sigma^2}(y_i - \mu_g)^2\right\}$$

(13)

Appendix C describes in more detail the implementation of a hybrid Gibbs algorithm for sampling. The underlying algorithm addresses the reconstruction of an image distributed from a Potts model based on a noisy version of this image. The purpose of image segmentation is to cluster pixels into homogeneous classes without preference and based only on the spatial coherence of the structure.

Once the image has been classified into G different homogeneous classes, with μ_g the mean value of the ratio Hα/FUV for each class, we can apply both methodologies: the one used in Paper I and the one presented here. In Section 4.1, we compare these ratio values with the SB99 model to assign an age range to that homogeneous region. In particular, for M83 these mean values are: $\mu_1 = 0.3045$, corresponding to ages >6 Myr; $\mu_2 = 0.6139$ and $\mu_3 = 0.8807$ for the age range $5 - 6$ Myr; $\mu_4 = 1.151$ for $3 - 4$ Myr; and $\mu_5 = 1.5195$ for ages $1 - 2$ Myr.

The resulting age map is also discrete (bottom-left plot in Fig. 9), but the structures of the age patterns are more consistent with the Hα/FUV image, as seen comparing the flux ratio image (top-left) with the age map obtained in Paper I (bottom-right). When compared, the age patterns and structures of these two discrete age maps are quite similar, the main difference being just the age range assignment. It attests the robustness of the results in Paper I, and their consistency with the present analysis. The loss of resolution in the

resulting segmented age map is rather negligible under the point of view of the study of age patterns. The main aim of these studies is to get the age structures and patterns along and across the spiral arms, rather than giving an absolute age.

We can conclude that the effect of spatial dependence with adjacent regions does not affect the study of the age patterns when this dating technique is applied pixel by pixel. The Hα/FUV ratio proves to be a robust estimator of the age, local gradients, and global age patterns remain unchanged despite some loss of spatial resolution.

5 RESULTS: BAYESIAN AGE MAPS

Figs. 9, 11, and 13, show the resulting age maps for the galaxies of this study, applying the three proposed age dating methodologies. In them we compare the two methods presented at the current work (discussed in Sections 4.1) and 4.2, as well as the previous method (Paper I).

The top left-hand panels show the Hα/FUV ratio images for each galaxy, after a 3σ noise filter masking the background or fainter Hα emission pixels. The remaining pixels with strong Hα emission, understood as whole or partial H II regions, define the spiral arms and star forming regions clearly. Relative age patterns are already within this ratio image, where the higher Hα to FUV values correspond with the younger regions, whereas the darker pixels of lower ratio values represent older regions. This distinction between 'younger' and 'older' is under a very young age scale framework, since Hα emission is mainly available up to $\sim$15–20 Myr. That is, we are studying the spatial distribution of the very recent star forming regions.

The bottom left panels, (ii), display the age maps following the Bayesian age dating methodology of Section 4.2, assuming dependence between adjacent pixels. The Hα to FUV flux ratio image has been segmented into five or four heterogeneous average $<F_{H\alpha}/F_{FUV}>$ regions. In this manner, we deal with issues affecting age dating by means of a SSP model, such as IMF subsampling, luminosity or mass thresholds, or spatial influence from adjacent regions. In so doing, we lose spatial resolution into discretized age maps.

Top right panels,(i), are the age maps obtained with the Bayesian approach presented in Section 4.1, assuming spatial independence of the fluxes from each pixel. The higher advantage of this last technique is not only the great resolution of the age maps, but also obtaining a posterior age distribution function in each pixel of the image. We are able to give probabilities or select any statistical moment since a sample of the posterior probability function for the age is given. To match this with the same criterion of Paper I, and for plotting purposes, each pixel of the image shows the Mode of the age posterior probability distribution.

Bottom right panels, (iii), are the age maps applying the previous methodology of Paper I. In this case we also get a 'discrete' or categorized age map, into four age ranges: <4, $4 - 6$, $6 - 9$, and >9 Myr. The colour scale has been matched to the other Bayesian age maps for comparison. We take the median age for each age range but the >9 Myr range by 9 Myr.

Figs. 9, 11, and 13 show the results from the previous Paper I methodology (iii) to be robust. The look broadly the same, although the new age maps fits better with the observed structures in the Hα/FUV flux ratio image. Moreover, the present methodologies let us to get the posterior PDF of the age, and so a continuous age map, as well as statistical estimator of probabilities. That is. it also improves not only qualitatively the resolution of the age patterns, resulting in richer age patterns structures, but the quality of the
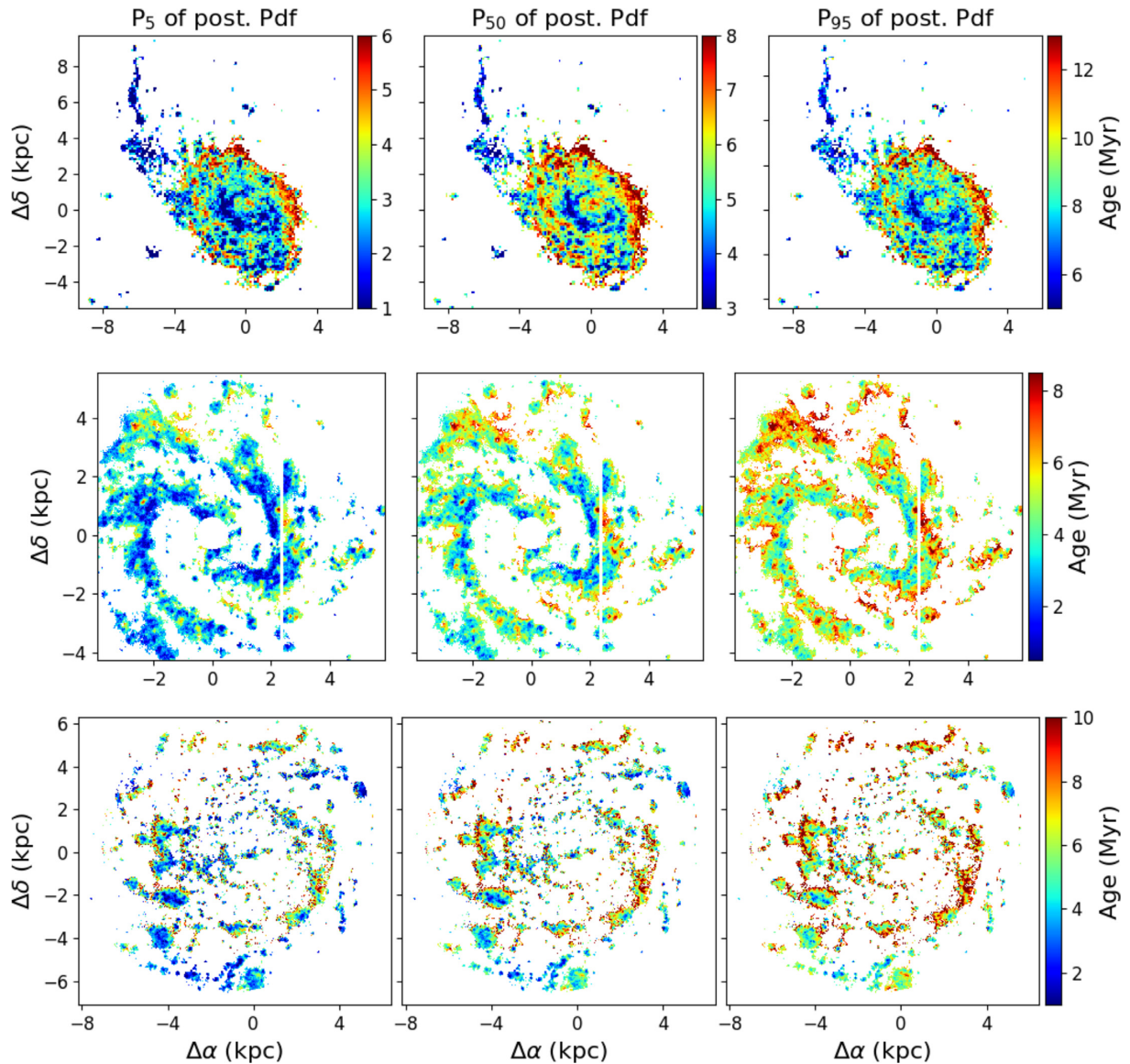
Figure 8. Comparison between age maps obtained in Section 4.1, showing the central 90 per cent credible interval of the age posterior distribution. I.e. the true value of the age will be in between these values with a probability of 90 per cent. They provide an idea of the age estimation uncertainty. Age maps for NGC 1068 are on the top, each one with their own age scale, highligthing the age pattern structure remains nearly the same. M83 is in the middle panels, and the bottom panels correspond with M101. The latter share the same age scale to exhibit as the general age pattern structure remains, except an average 'zero' point.

information and the results. Once the data are given, we get the conditional probability of the model parameters of the observed data, and not only a single value estimator.

Fig. 8 gives the uncertainties for the age estimation in the (i) age maps, when the posterior probability distribution is available. It shows the central 90 per cent credible interval of the age posterior distribution; i.e. the true value of the age will be in between these values with a probability of 90 per cent. On the left it shows the 5th percentile of the age distribution. The central age map corresponds to the median age, P50, and the 95th percentile on the left. After such comparison we check out the robustness of the methodology for age dating.

The corresponding age maps for NGC 1068 are on the top, each one with their own age range scale. Despite the different age scales, the general age patterns are much the same. The only difference

is a zero point offset in age. The middle and bottom panels show the corresponding percentiles for M83 and M101, respectively. For these maps we leave the same age scale to highlight the offset point between them. Once again the general age patterns remain the same. On average, this zero point value is less than ± 2 Myr for the three galaxies, even $\lesssim \pm 1$ Myr in the case of M83.

Finally, in Appendix D we show the age maps (i) and (iii) for the galaxy sample studied in Paper I, for the different methodologies to be compared. Several percentiles of the distribution are shown to give the corresponding age uncertainties. We can infer the same physical analysis and conclusions from the new age maps as was done in Paper I. The general age patterns results are the same, with the exception of the resulting continuous age maps what improves the resolution of the age patterns and structures, as well as the potential of the Bayesian approach.
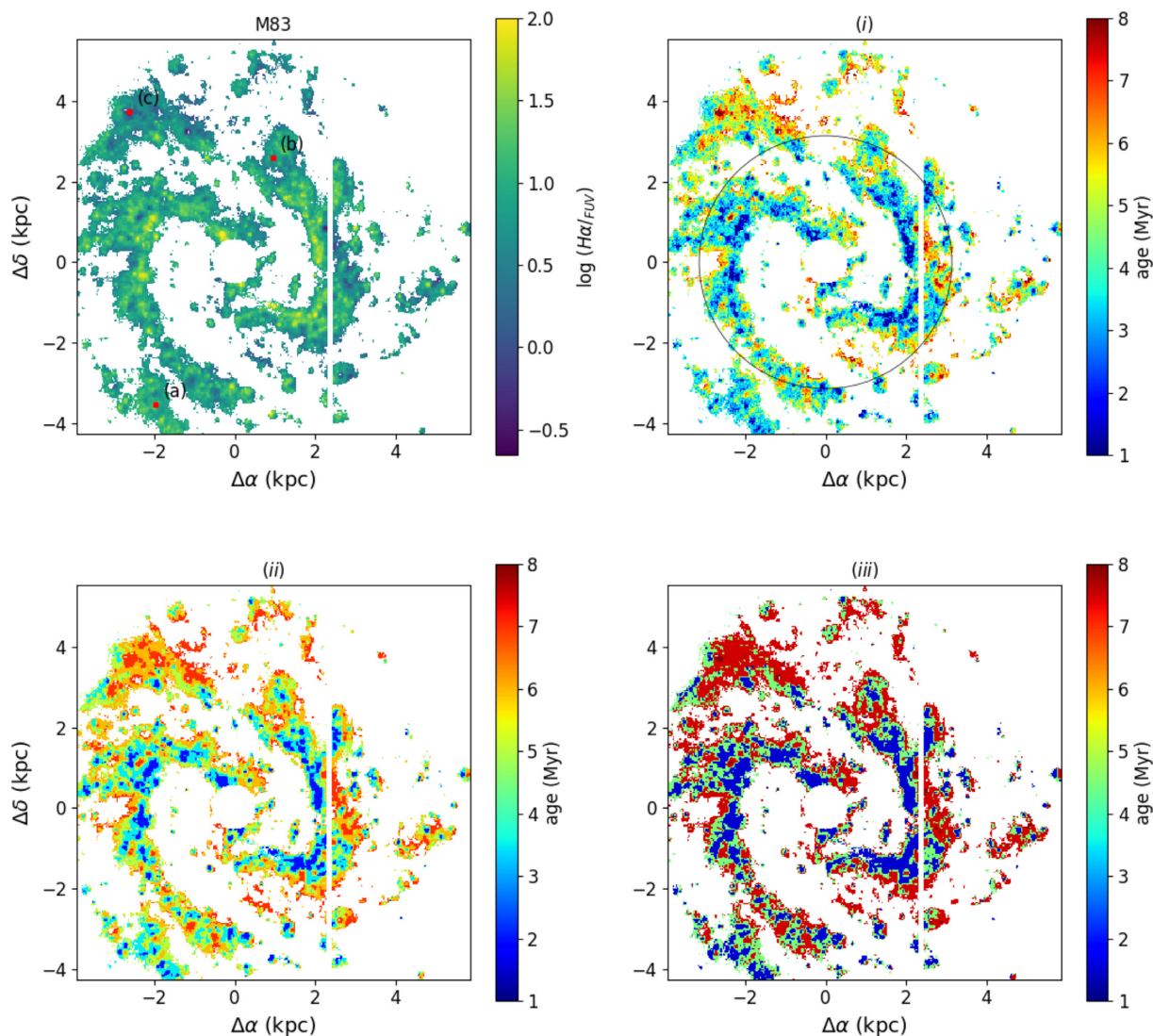
Figure 9. *Top left:* Image of the H α to FUV ratio for M83 (in log units). The higher values of $F_{H\alpha}/F_{FUV}$ flux ratio (in yellow) denote the youngest SF regions. While the lower ratios indicate 'older' ages; *Top right (i):* Age map for M83 obtained with the presented Bayesian approach in Section 4.1, assuming independence between pixels, and taking the mode of the posterior probability distribution of the age parameter at each pixel of the image. The black circle indicates the corotation radius, at $\sim$3 kpc (see text); *Bottom left (ii):* Age map obtained with the Bayesian approach in Section 4.2, assuming dependence between adjacent pixels and after an image segmentation of the $F_{H\alpha}/F_{FUV}$ flux ratio, into five heterogeneous regions. The colour scale has been matched to the previous age map, at the top left; *Bottom right (iii):* Age map for M83, applying the previous methodology of Paper I. In this case we get a 'discrete' or categorized age map into four classes or age ranges: $<4, 4-6, 6-9$ and >9 Myr. The colour scale has been matched to the Bayesian age maps for comparison purposes (taking the median age for each age range, except for the last one it is taken the minimum). The vertical white stripes of the images correspond with some artefacts from the Hα image, so they were masked out.

5.1 M83

M83 (NGC 5236), or the 'Southern Pinwheel' galaxy, is a nearby barred-spiral galaxy. The nearest galaxy in the sample (4.5 Mpc; Karachentsev et al. (2002)), its proximity and low inclination angle gives spatial resolution, not only to resolve its spiral arms but also to observe a wealth of detail in individual star-forming structures within the arms.

M83 is a metal-rich spiral galaxy, with a radial metallicity gradient which flattens at large radii (Bresolin et al. 2016). The observable used in our study is not sensitive to metallicity, we could not see nay gradient or pattern in the metallicity maps. The middle panels of Fig. 6 show the marginal posterior probability distribution of the metallicity for three pixels (a), (b), and (c) with different flux ratio

values and location within the disk (see top left plot at Fig. 9). The Hα/FUV flux ratio is quite sensitive to age variations, but metallicity behaves as a free parameter in a our model, free to fit the data.

The resulting age maps (i), (ii), and (iii), see Fig. 9, are similar to that obtained for M 51 in Paper I, dominated by a young population of stars with less than 6–7 Myr. The age structure exhibits gradients across the spiral arms, with the younger stars toward the inner edge while the older stars are located approaching the outer edges within the corotation radius, as expected from the density wave theory (Roberts 1969; Martínez-García, González-Lópezlira & Bruzual-A 2009). The corotation radius of $\sim$3 kpc ($\sim$2.3$'$ − 2.4$'$ (Hirota et al. 2014) at a distance of 4.5 Mpc) is plotted as a black circle in the top right panel of Fig. 9. We can also observe the inverse age pattern
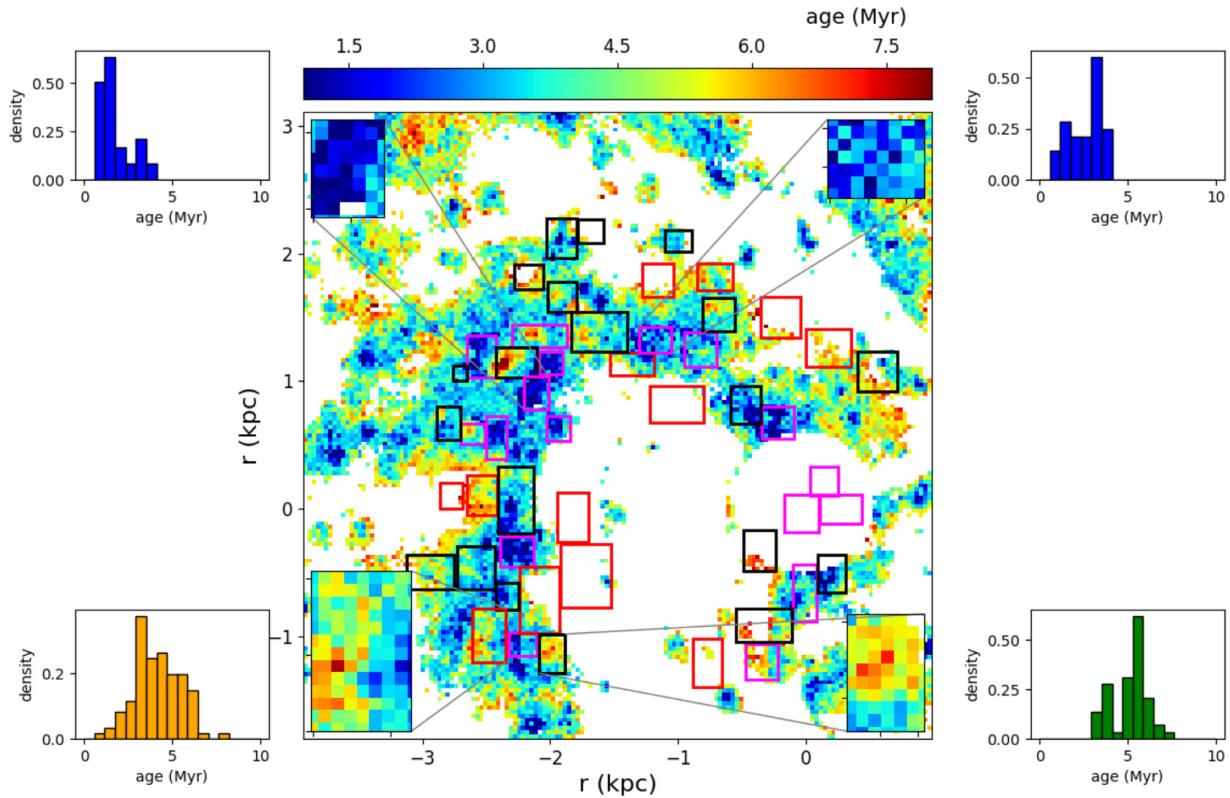
Figure 10. Zoom of the nuclear region and part of the eastern spiral arm and inter-arm region of M83, showing 50 regions from (Kim et al. 2012) for comparison. These regions were age-dated with data from HST/WFC3, with the aim of studying the relationship of the spatial variations of stellar ages with the evolution of the galaxy and the star formation triggering mechanisms. The regions with ages $1 - 10$ Myr are plotted with magenta boxes, ages of $10-20$ Myr in black, and for ages greater than 20 Myr in red.

outside corotation; i.e. the younger population are preferentially located radially in the outer side of the arms. The eastern arm exhibits pretty well how the youngest population, bluest pixels, change from the inner to outer arm side when crossing the corotation radius.

Fig. 10 displays 50 regions of an average size 260 pc × 280 pc, which cover the nuclear region and part of the eastern spiral arm and inter-arm region (Kim et al. 2012, see their fig.1). These authors selected these regions with the aim of studying the spatial variations of stellar ages in M83, looking for evidence of the evolution of the galaxy and star formation triggers. Their age scale is similar to ours, with regions classified into 3 age ranges: encompassing $1 - 10$ Myr, which covers the majority of our age range. These are plotted with magenta boxes at Fig. 10. It can be seen that many of these regions also correspond with our younger ages. Two examples are magnified at the top left and top right sides. The blue histograms represent the age distribution of two of these regions, with ages less than 5 Myr.

Regions with ages of $10-20$ Myr are shown with black boxes. They correspond to our oldest ages, since this is the limit for Hα emission. We observe in Fig. 10 how many of these regions are indeed dominated by our older ages (orange and red pixels). But there are also some of these regions more likely to be intermediate ages, such as the case in the bottom left of the plot.

Outlined red boxes show the third group, with ages greater than 20 Myr. Most of these do not show Hα in our maps, as they correspond to the post-nebular phase. In general we find quite a good agreement between the two results.

For the youngest star forming regions, our age maps find a similar scenario as that described by Kim et al. (2012), and in agreement with density wave theory. These authors also found that younger (10 Myr) stars are found mainly in concentrated aggregates along the active star forming regions in the spiral arm. Intermediate age stars are located downstream, on the opposite side from the dust lane, as expected based on density wave models, and the older stars more dispersed (see their fig. 12). Kim et al. (2012) argue stars form primarily in star clusters and then disperse on short timescales to form the field population. Moreover, Wolf-Rayet stars, which are taken into account in our SSP models, correlate with the position of many of the youngest regions.

Dobbs & Pringle (2010) study the mechanisms triggering star formation in galaxies and their evolution. They discuss the locations of age-dated stellar clusters as a possible discriminant of the origin/source of the excitation mechanism for the spiral structure. Under the assumption that stellar clusters form predominantly within the spiral arms (higher gas density regions), the distribution of the age-dated clusters through out a spiral could give some clues to the mechanism for spiral arm formation. These authors found diverse spatial distributions for clusters of different ages, depending on the underlying dynamics of the galaxy and the spiral excitation mechanism.

Despite the different age scales (the age in Dobbs & Pringle (2010) varies from ∼2–130 Myr, see their fig. 2), we note that their models for a global age pattern still agree with our age maps for the youngest stellar populations (up to around 15 or 20 Myr).
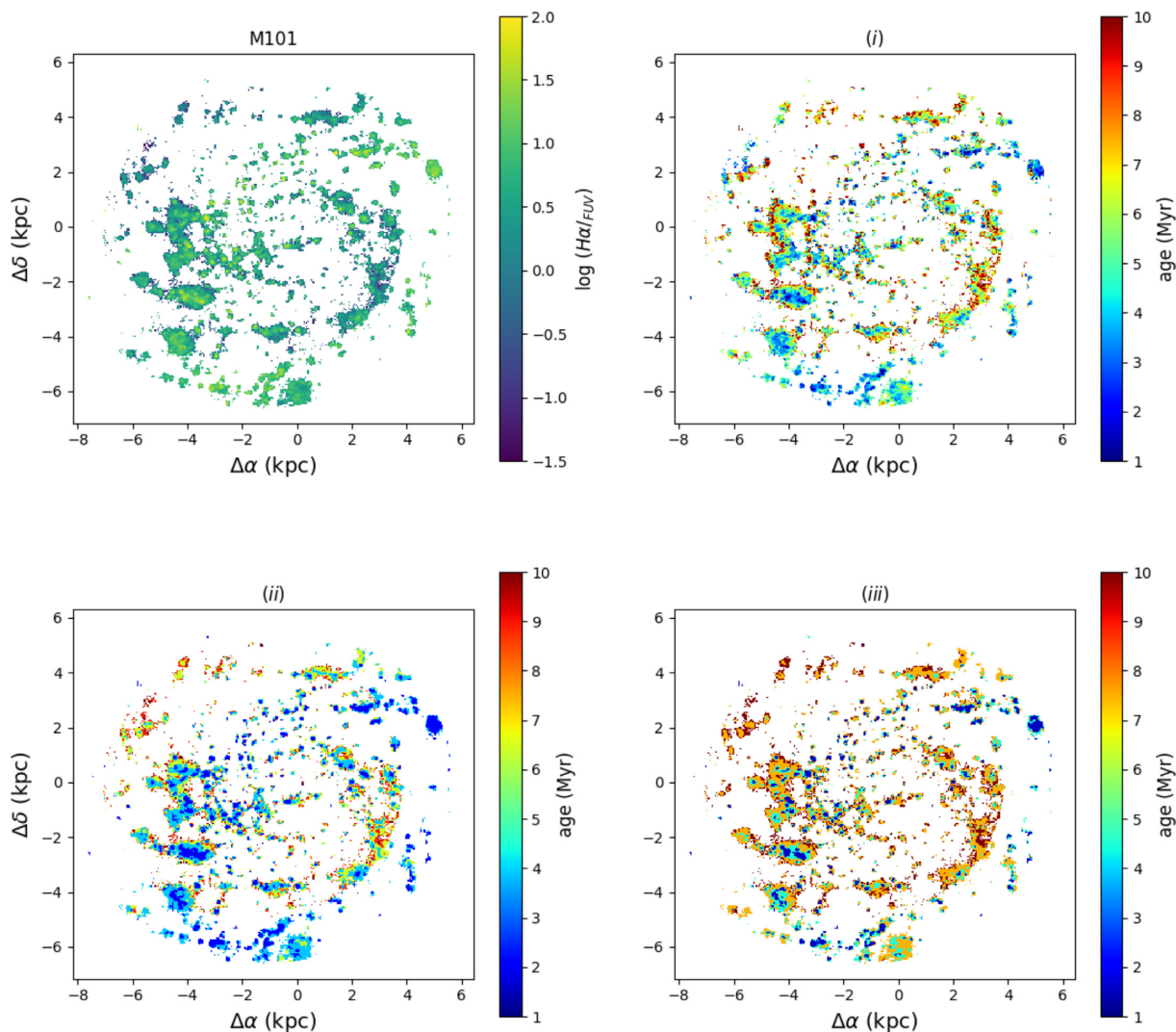
Figure 11. As Fig. 9, but for M101.

For both M83 and M101 (Figs. 9 and 11), we can identify the age distribution for a galaxy model of a constant pattern speed with a bar, and a floculent spiral, respectively (cf. their figure 2). Dobbs & Pringle (2010) describe the younger stars in the spiral arms or bar with older stars downstream in the interarm regions for the former model. The distribution of stellar clusters is more complicated in floculent spirals, because each segment of a spiral arm tends to contain clusters of a similar age. This can be observed in the figure at 11 for M101, mostly at the extreme southern segments.

5.2 M101

M101 is a nearby, face-on, giant spiral galaxy. At a distance of 6.7 kpc (Bosma et al. 1981), it provides an enough spatial resolution to study the underlying stellar populations. It is also an excellent laboratory for studying stars, as OB-type stars, blue giants, yellow supergiants, and red supergiants, as well as stellar clusters, H II regions, and supernova remnants (Grammer & Humphreys 2013, 2014, and references within).

Grammer & Humphreys (2014) explore the M101 SFH as a function of the radial distance and its effect on the emission properties and stellar populations. They find that the mass fraction for

stars younger than 16 Myr (roughly comparable to our age range) is 15 per cent − 35 per cent in the inner regions, compared to less than 5 per cent in the outer regions. This percentage is greater than 50 per cent in the inner regions for stars younger than 35 Myr. That is, the inner regions are dominated by a young stellar populations. Comparing the results in their figure 8 within the reach of our data, 8 kpc, we see that 85 per cent and 95 per cent of stellar mass fraction have ages younger that 15 or 20 Myr under 10 kpc in radius, in good agreement with our results.

As in Fig. 9 for M83, Fig. 11 shows the age maps for M101; (i) and (ii), derived with the methods introduced in this paper, and (iii) as in Paper I. A common characteristic feature of these age maps is their morphologies, with a structure more floculent than spiral. Furthermore, we detect some an age gradient tending toward younger stars in the 'outer' disk. It is not actually the outer disc but the middle region, since we only observe around 8 kpc, due to the circular Taurus aperture.

According to Lin et al. (2013), the radial age profile of M101 presents a younger bulge, comprising an older inner region of the disk with steeper age gradient, and an outer disk region with a flatter gradient. These gradients and ages are not comparable with our age maps, since they are in Gyr units. However, at youngest age
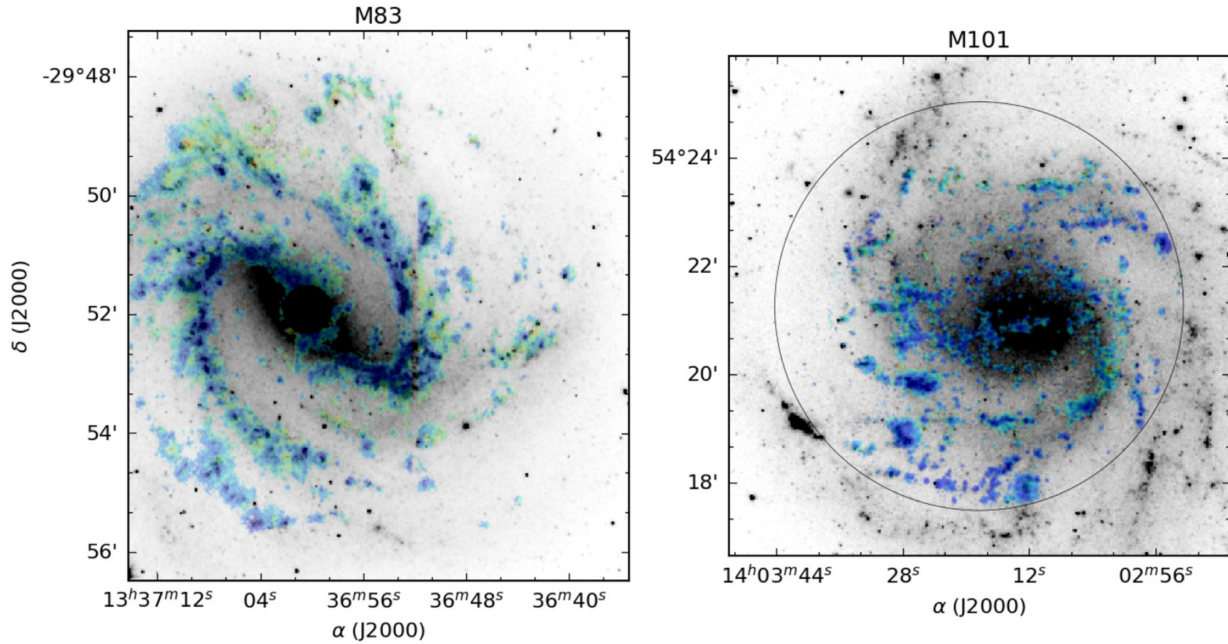
Figure 12. Comparison between the resulting morphology of star forming regions for M83 and M101. Despite both being spiral galaxies, the former has a well defined spiral estructure, whereas the latter has flocculent spiral arms segments. The background image is a 3.6 μm IRAC image from the *Spitzer* survey, where it is effectively observed the characteristic spiral morphology in both cases. Black circle defines the TTF field of view for the Hα image, leaving outside part of the disk and spiral arms of M101.

scales there are similar trends. The interaction with another galaxy of the M101 group and consequent gas accretion could trigger this star formation in the outer regions. The low gas metallicity gradient of M101 with respect to other spiral galaxies also supports an interacting or recent merger scenario (Lin et al. 2013, and references within).

Lin et al. (2013) also describe spurious arms full of H II regions much younger than the interarm in the inner disk (albeit within a different framework, of wider age scales in Gyr and providing the whole SFH of M101).

Fig. 12 shows a comparison between the morphologies of the star forming regions in M83 and M101. The background IRAC 3.6 μm images show clearly their spiral structure. However, while the age map of M83 presents a well-defined structure, whereas the corresponding M101 image shows flocculent spiral arms segments. The black circle in the M101 image defines the TTF field of view for the Hα image. Despite the regions it leaves outside the field of view, this does not affect the resulting global morphology for this galaxy. The inner disk has a clear flocculent structure.

The difference between the morphologies of the star forming regions of both M83 and M101 is interesting. It is likely related to the processes triggering the star formation, as well as the subsequent evolution.
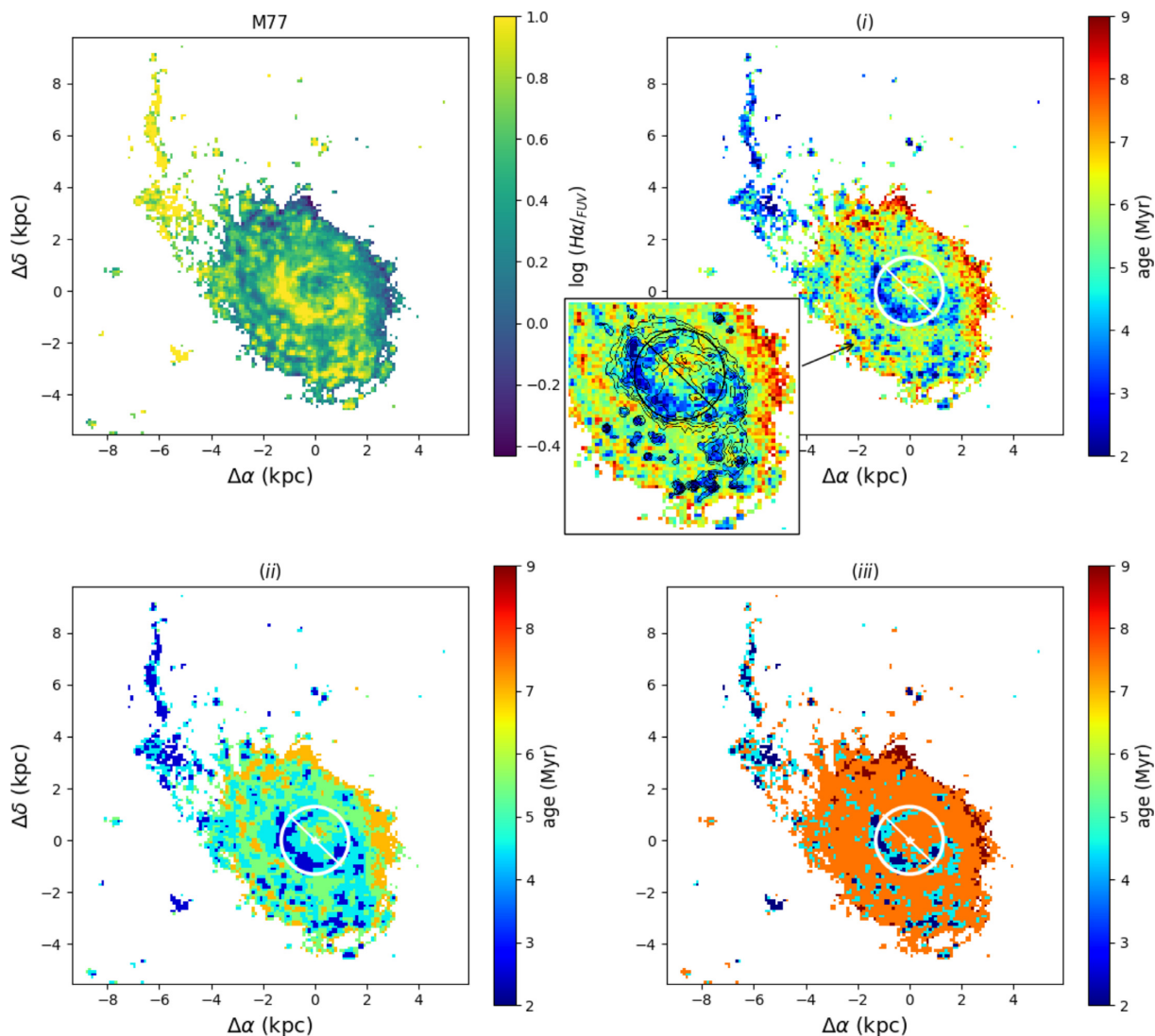
5.3 NGC 1068

NGC 1068 is an early-type Sb, barred spiral galaxy, and the closest (14.4 Mpc, $1'' = 72$ pc; Bland-Hawthorn et al. 1997) luminous Seyfert 2 galaxy. It is considered the prototype Seyfert 2 galaxy (Khachikian & Weedman 1974). Its brightness, nuclear activity, proximity, and orientation make it an excellent laboratory (in a single physical framework) to study the Seyfert nucleus, inner disk structure, as well as the unifying, dusty torus

model for Seyfert galaxies (e.g. Orr & Browne 1982; Davies, Sugai & Ward 1998; Bruhweiler et al. 2001; López-Gonzaga et al. 2014).

Likely the best studied active galaxy in the local universe, it has been subject of numerous studies, at many different wavelengths. Signs of the AGN are evident at nearly all wavelengths, such as the bright cones of photoionized gas detected both in optical and X-rays wavelenghts (Veilleux et al. 2003, and references within). It has a large-scale oval and a nuclear bar with a pseudo-bulge which is very massive with respect to its central black hole (e.g. Kormendy & Ho 2013). The ionization source of the gas in the circumnuclear region is due mainly to the AGN, so age determination with the proposed methodologies based on H α/FUV is not reliable for this innermost region. The location of the AGN is marked with a star in the age maps of Fig. 12. Diagnostic line ratio plots have shown that photoionization is the preferred mechanism for high excitation gas in NGC 1068 (Nishimura et al. 1984; Evan & Dopita 1986; Bergeron et al. 1989).

Once the background has been subtracted, the age maps (see Fig. 13) do not show well-defined spiral arms as M83, indicating a lower intensity of SF activity across the disk. The circumnuclear ring or pseudo-ring (the star-forming or starburst ring) has been clearly resolved into tightly-wound spiral arms at the ends of the 3 kpc bar (Bland-Hawthorn et al. 1997), outlined by the white line segment at Fig. 13. This starburst ring is clearly defined in our age maps by the bluest, and therefore youngest, structures inside and around the corotation radius, outlined with a white/black circumference at Fig. 13, at $r \sim 1.3$ kpc (García-Burillo et al. 2014). The H II regions in the circumnuclear ring are as bright as a string of M82-type galaxies, and compete in terms of bolometric luminosity with the nucleus (Davies et al. 1998). Telesco et al. (1984) found that approximately the half of the bolometric luminosity of NGC 1068, $\sim 3 \times 10^{11} L_\odot$, corresponds to the Seyfert nucleus whereas the other half to the starburst region, where most of the CO(3-2) flux

Figure 13. As Fig. 9, but for NGC 1068. The AGN location is marked with a star in all the age maps, as well as the 3 kpc bar (Bland-Hawthorn et al. 1997), outlined by a white line, and the ≈1.3 kpc corotation radius (García-Burillo et al. 2014) with a white (black in the zoomed central region) circle. The central part of the disk is zoomed in the (i) age map, plotting the highest levels, greater than percentiles P_{75}, P_{80}, P_{85}, P_{90}, and P_{95}, of the age contours. These contours define the starburst ring, and some of the south-west spiral arm and concentrated knots of SF.

in the disk is detected (García-Burillo et al. 2014). The existence of the circumnuclear ring is attributed to gas settling between the inner Lindbland resonances (ILRs) as a result of the action of a barred gravitational potential (Telesco & Decher 1988). Therefore vigorous SF is expected to occur as a direct result of the increased cloud density in the ILR.

The three age maps in Fig. 13, coincide with the location of the youngest regions, the bluest ones, mainly within the brightest regions of the ring, around the corotation radius, with some knots of high SF regions throughout the disk, and at north-east plume out to a radius of ∼7–8 kpc from the nucleus. However, multi-line imaging and long-slit spectroscopy of the gas in this north-east complex or filament, roughly aligned with the ionization cone on smaller scale, revealed contributions from the AGN hard radiation to the high ionization of the gas in this region (Veilleux et al. 2003). If photoionization is not the main source of the gas ionization, the ages determined by H α/FUV are not reliable, as

it also happens within the circumnuclear region, due to the AGN effects.

In the rest of the disk, the younger population is concentrated in the inner ring, as mentioned above. It seems to extend along the south-west spiral arm at larger scale, as well as it concentrates in some knots of SF regions across the disk. The intermediate ages are within the disk, and the older ones are found at the edges.

The relevance or strength of the ring luminosity, being host of the youngest population in the disk, is checked by plotting contour levels from the H α image. The highest levels appear only around this ring. To see this feature in more detail, the central part of the disk is zoomed in the (i) age map of Fig. 13, combining the ages and highest levels of age contours. Only those levels greater than the P_{75} percentile are shown, which coincide with the ring and some of the south-west spiral arm, but not in the plume.

6 DISCUSSION

In this paper, we present two Bayesian inference algorithms that allow us to evaluate the age of the stellar population of a galaxy at pixel resolution. The methodology of Section 4.1 provides the posterior probability density function for different parameters of the stellar population synthesis models, including age, metallicity and the fraction of ionizing photons. Only the former results in sensitive to the Hα/FUV flux ratio. We focused on the age variable, with a unimodal distribution and well defined central value in most cases, Fig. 6. The accuracy of age estimation in the cases studied is better than ±2 Myr. However total uncertainty could be strongly influenced when applying synthetic stellar models to single pixels, more than by the own model uncertainties. This imprecision is affected by the mass of the underlying stellar population, whose parameters are estimated. The measurement of this uncertainty is not trivial when the total brightness of the pixel is lower than the Lower Luminosity Limit (LLL) of the model (Cerviño & Luridiana 2006), and is beyond the scope of this work.

The other Bayesian approach developed in Section 4.2 try to mitigate the effect of these luminosity/mass thresholds or IMF sub-sampling issues, when SSP models are directly applied to pixel-wise regions. On the other hand, we get a discretization of the age map, and so higher resolution is lost. Nevertheless the main structures remain recognizable, or otherwise unchanged.

In general, the age patterns we show appear to be quite robust to sampling and only sensitive to the Hα/FUV flux ratio. So the importance of these two new techniques is, on the one hand, the huge potential provided by the Bayesian inference and holding the age posterior probability function (a sample of the posterior). And in the other case, to deal with the uncertainties derived when applying SSP models to pixel-wise size regions. The present work focuses mainly on the formulation of a new methodology to estimate the age from UV, optical, and infrared images by Bayesian inference.

The number of arms shown, the sharpness and width of these arms, as well as the separation and location of large star-formation complexes along the arms are some of the features that can give us some first hints about the nature of the physical mechanisms within a spiral galaxy. However, the morphology by itself can sometimes be misleading. Thus we also need physical observables, drawing from their galactic spatial distribution to reconcile model predictions from different spiral pattern generators, Dobbs & Pringle (2010). Age map analysis is one of the best ways to test these aims.

These proposed methodologies only refer to the youngest stellar populations (i.e. the distribution or location of the youngest stellar regions, less than 20 Myr), from the very recent to less recent episodes of SF along the galactic disk. These are mainly in the arms and interarm regions of the grand-design galaxies. The use of age maps to elucidate the origin of the spiral arms is outside the scope of this paper. However we are able to highlight how age maps show very distinctive morphologies characteristics for three spiral galaxies.

ACKNOWLEDGEMENTS

REFERENCES

Barker S., de Grijs R., Cerviño M., 2008, A&A, 484, 711
Bendo G. J., Boselli A., Dariush A., 2012, MNRAS, 419, 1833
Bergeron J., Petitjean P., Durret F., 1989, A&A, 213, 61
Bland-Hawthorn J., Gallimore J. F., Tacconi L. J. et al., 1997, Ap&SS, 248, 9
Bland-Hawthorn J., Jones D. H., 1998, PASA, 15, 44
Bland-Hawthorn J., Lumsden S. L., Voit G. M., Cecil G. N., Weisheit J. C., 1997, Ap&SS, 248, 177
Boissier S., Gil de Paz A., Madore B. F. et al., 2005, ApJ, 619, L83
Bosma A., Goss W. M., Allen R. J., 1981, A&A, 93, 106
Bresolin F., Kudritzki R.-P., Urbaneja M. A. et al., 2016, ApJ, 830, 64
Bruhweiler F. C., Miskey C. L., Smith A. M., Landsman W., Malumuth E., 2001, ApJ, 546, 866
Buat V., Iglesias-Páramo J., Seibert M. et al., 2005, ApJ, 619, L51
Casado J., Ascasibar Y., García-Benito R. et al., 2017, MNRAS, 466, 3989
Cerviño, M., Luridiana, V., 2006, A&A, 451, 475
Cerviño M., 2013, New Astron. Rev., 57, 123
Cerviño M., Luridiana V., 2004, A&A, 413, 145
Cerviño M., Luridiana V., Pérez E., Vílchez J. M., Valls-Gabaud D., 2003, A&A, 407, 177
Cerviño M., Valls-Gabaud D., Luridiana V., Mas-Hesse J. M., 2002, A&A, 381, 51
Charbonnel C., Meynet G., Maeder A., Schaller G., Schaerer D., 1993, A&AS, 101, 415
Dale D. A., Gil de Paz A., Gordon K. D. et al., 2007, ApJ, 655, 863
Dale D. A., Helou G., 2002, ApJ, 576, 159
Davies R. I., Sugai H., Ward M. J., 1998, MNRAS, 300, 388
de Amorim A. L., García-Benito R., Cid Fernandes R. et al., 2017, MNRAS, 471, 3727
Dobbs C. L., Pringle J. E., 2010, MNRAS, 409, 396
Evans I. N., Dopita M. A., 1986, ApJl, 310, L15
Foyle K., Wilson C. D., Mentuch E. et al., 2012, MNRAS, 421, 2917
Galametz M., Kennicutt R. C., Calzetti D. et al., 2013, MNRAS, 431, 1956
García-Burillo S., Combes F., Usero A. et al., 2014, A&A, 567, A125
Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2003, Bayesian Data Analysis. Chapman and Hall, New India Book Agency
González-Gaitán S., de Souza R. S., Krone-Martins A., Cameron E., Coelho P., Galbany L., Ishida EEO, MNRAS, 2018, 482, 3880
Grammer S., Humphreys R. M., 2013, AJ, 146, 114
Grammer S., Humphreys R. M., 2014, AJ, 148, 58
Haykin S., 1999, Neural Networks, A Comprehensive Foundatio. Prentice-Hall, Pearson Education, Inc. Upper Saddle River, New Jersey 07458
Hinkley D. V., 1969, On the Ratio of Two Correlated Normal Random Variables, Biometrika Vol. 56. Oxford Univ. Press, UK, 635. Available at: https://www.jstor.org/stable/i315463
Hirota A., Kuno N., Baba J. et al., 2014, PASJ, 66, 46
Jones D. H., Bland-Hawthorn J., 2001, ApJ, 550, 593
Jones D. H., Shopbell P. L., Bland-Hawthorn J., 2002, MNRAS, 329, 759
Karachentsev I. D., Sharina M. E., Dolphin A. E. et al., 2002, A&A, 385, 21
Khachikian E. Y., Weedman D. W., 1974, ApJ, 192, 581
Kim H., Whitmore B. C., Chandar R. et al., 2012, ApJ, 753, 26
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Lee J. C., Gil de Paz A., Kennicutt R. C., Jr. , et al. 2011, ApJS, 192, 6
Leitherer C., Heckman T. M., 1995, ApJS, 96, 9
Leitherer C., Schaerer D., Goldader J. D. et al., 1999, ApJS, 123, 3
Lejeune T., Cuisinier F., Buser R., 1997, A&AS, 125, 229
Lin L., Zou H., Kong X. et al., 2013, ApJ, 769, 127
López-Gonzaga N., Jaffe W., Burtscher L., Tristram K. R. W., Meisenheimer K., 2014, A&A, 565, A71
Marin J.-M., Robert C. P., 2014, Bayesian Essentials with R. Springer, New York
Martin D. C., Fanson J., Schiminovich D. et al., 2005, ApJ, 619, L1
Martínez-García E. E., González-Lopezlira R. A., Bruzual-A G., 2009, ApJ, 694, 512
Mas-Hesse J. M., Kunth D., 1991, A&AS, 88, 399

Nishimura M., Kaneko N. Toyama K., 1984, AAP, 130, 46
Orr M. J. L., Browne I. W. A., 1982, MNRAS, 200, 1067
Roberts W. W., 1969, ApJ, 158, 123
Rue H., Riebler A., Sø rbye S. H., Illian J. B., Simpson D. P., Lindgren F. K., 2017, Annu. Rev. Stat. Appl., 4, 395
Salpeter E. E., 1955, ApJ, 121, 161
Schaerer D., Charbonnel C., Meynet G., Maeder A., Schaller G., 1993a, A&AS, 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
Skilling J., 2006, Bayesian Anal., 1, 833
Smith L. J., Norris R. P. F., Crowther P. A., 2002, MNRAS, 337, 1309
Stoehr J., 2017, A review on statistical inference methods for discrete Markov random fields, Cornell University Library.preprint (arXiv: 1704.03331)

Sánchez-Gil M. C., Jones D. H., Pérez E. et al., 2011, MNRAS, 415, 753
Sánchez Gil M. C., Berihuete A., Alfaro E. J., Pérez E., Sarro L. M., 2015, J. Phys. Conf. Ser., 633, 012140
Telesco C. M., Becklin E. E., Wynn-Williams C. G., Harper D. A., 1984, ApJ, 282, 427
Telesco C. M., Decher R., 1988, ApJ, 334, 573
Thim F., Tammann G. A., Saha A. et al., 2003, ApJ, 590, 256
Tully R. B., Shaya E. J., Karachentsev I. D. et al., 2008, ApJ, 676, 184–205
Veilleux S., Shopbell P. L., Rupke D. S., Bland-Hawthorn J., Cecil G., 2003, AJ, 126, 2185
Vázquez G. A., Leitherer C., 2005, ApJ, 621, 695
Wu F. Y., 1982, Rev. Mod. Phys., 54, 235

APPENDIX A: SAMPLING EFFECTS AND THE RATIO OF TWO CORRELATED NORMAL RANDOM VARIABLES

Along this work, we assume that the underling distributions of H α and FUV luminosities can be described as a Gaussian distribution, therefore their ratio distribution is the ratio of two correlated Gaussian distributions, that is we take into account the correlation between the H α and the FUV flux from the same source.

Let us first show the mathematical steps to obtain the ratio distribution analytically, to compare it with the results of synthesis models. More details can be found in Hinkley (1969). Given two normal random variables $X \sim \mathcal{N}(\mu_x, \sigma_x)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y)$, which are correlated $\rho \neq 0$, and assuming the joint density of (x, y) is $g(x, y)$ a bivariate normal distribution.

The probability distribution function of the ratio $Z = X/Y$ is obtained by marginalizing the joint distribution function $g(x, y)$, and replacing x by zy into the equation (A1).

$$p_z(z) = \int_{-\infty}^{\infty} |y| g(zy, y) dy =$$

$$= \int_{-\infty}^{\infty} |y| \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} exp \left\{ \frac{-1}{2(1-\rho^2)} \left(\frac{(zy - \mu_x)^2}{\sigma_x^2} - \frac{2\rho(zy - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right) \right\} dy =$$

$$= \int_{-\infty}^{\infty} \frac{|y|}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} exp \left\{ \frac{-1}{2(1-\rho^2)} \left[y^2 \left(\frac{z^2}{\sigma_x^2} - \frac{2\rho z}{\sigma_x \sigma_y} + \frac{1}{\sigma_y^2} \right) - 2y \left(\frac{z\mu_x}{\sigma_x^2} - \frac{\rho(z\mu_y + \mu_x)}{\sigma_x \sigma_y} + \frac{\mu_y}{\sigma_y^2} \right) + \left(\frac{\mu_x^2}{\sigma_x^2} - \frac{2\rho\mu_x\mu_y}{\sigma_x \sigma_y} + \frac{\mu_y^2}{\sigma_y^2} \right) \right] \right\} dy$$

$$\text{(A1)}$$

Defining the following parameters (A2), as a function of the ratio z and the parameters of the bivariate normal distributions,

$$a(z) = \sqrt{\frac{z^2}{\sigma_x^2} - \frac{2\rho z}{\sigma_x \sigma_y} + \frac{1}{\sigma_y^2}} \qquad b(z) = \frac{z\mu_x}{\sigma_x^2} - \frac{\rho(z\mu_y + \mu_x)}{\sigma_x \sigma_y} + \frac{\mu_y}{\sigma_y^2}$$

$$c = \frac{\mu_x^2}{\sigma_x^2} - \frac{2\rho\mu_x\mu_y}{\sigma_x \sigma_y} + \frac{\mu_y^2}{\sigma_y^2} \qquad d(z) = exp \left\{ \frac{b^2(z) - ca^2(z)}{2(1-\rho^2)a^2(z)} \right\}$$

$$\text{(A2)}$$

Equation (A1) is simplified into equation (A3)

$$p_z(z) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} |y| exp \left\{ -\frac{1}{2} \left(\frac{y^2 a^2(z)}{(1-\rho^2)} - 2\frac{yb(z)}{(1-\rho^2)} + \frac{c}{(1-\rho^2)} \right) \right\} dy =$$

$$= \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} |y| exp \left\{ -\frac{1}{2} \left(\left(\frac{ya(z)}{\sqrt{1-\rho^2}} - \frac{b(z)}{a(z)\sqrt{1-\rho^2}} \right)^2 + \left(\frac{c}{(1-\rho^2)} - \frac{b^2(z)}{(1-\rho^2)a^2(z)} \right) \right) \right\} dy =$$

$$= \frac{d(z)}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} |y| exp \left\{ -\frac{1}{2} \left(\frac{ya(z)}{\sqrt{1-\rho^2}} - \frac{b(z)}{a(z)\sqrt{1-\rho^2}} \right)^2 \right\} dy$$

$$\text{(A3)}$$

We separate the integral into the positive and negative ranges in equation (A3), and apply the change of variable $u = (ya^2(z) - b(z))/a(z)\sqrt{1 - \rho^2}$

$$p_z(z) = \frac{d(z)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \left(\int_{-\infty}^0 -\mathbf{y}\, exp\left\{ -\frac{1}{2}\left(\overbrace{\frac{ya^2(z) - b(z)}{a(z)\sqrt{1-\rho^2}}}^{u} \right)^2 \right\} \mathrm{d}y + \int_0^\infty \mathbf{y}\, exp\left\{ -\frac{1}{2}\left(\overbrace{\frac{ya^2(z) - b(z)}{a(z)\sqrt{1-\rho^2}}}^{u} \right)^2 \right\} \mathrm{d}y \right)$$

$$= \frac{d(z)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \left(\int_{-\infty}^{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}} -\left(\mathbf{u} + \frac{\mathbf{b(z)}}{\mathbf{a(z)}\sqrt{1-\rho^2}}\right) \frac{\sqrt{1-\rho^2}}{\mathbf{a(z)}} exp\left\{ -\frac{1}{2}u^2 \right\} \frac{\sqrt{1-\rho^2}}{a(z)} \mathrm{d}u \right. +$$

$$\left. + \int_{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}}^\infty \left(\mathbf{u} + \frac{\mathbf{b(z)}}{\mathbf{a(z)}\sqrt{1-\rho^2}}\right) \frac{\sqrt{1-\rho^2}}{\mathbf{a(z)}} exp\left\{ -\frac{1}{2}u^2 \right\} \frac{\sqrt{1-\rho^2}}{a(z)} \mathrm{d}u \right) =$$

$$= \frac{d(z)\sqrt{1-\rho^2}}{2\pi\sigma_x\sigma_y a^2(z)} \left(\int_{-\infty}^{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}} -u\, e^{-\frac{1}{2}u^2} - \frac{b(z)}{a(z)\sqrt{1-\rho^2}} e^{-\frac{1}{2}u^2} \mathrm{d}u + \int_{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}}^\infty u\, e^{-\frac{1}{2}u^2} + \frac{b(z)}{a(z)\sqrt{1-\rho^2}} e^{-\frac{1}{2}u^2} \mathrm{d}u \right) =$$

$$= \frac{d(z)\sqrt{1-\rho^2}}{2\pi\sigma_x\sigma_y a^2(z)} \frac{b(z)}{a(z)\sqrt{1-\rho^2}} \left(-\int_{-\infty}^{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}} e^{-\frac{1}{2}u^2} \mathrm{d}u + \int_{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}}^\infty e^{-\frac{1}{2}u^2} \mathrm{d}u \right) + \tag{A4}$$

$$+ \frac{d(z)\sqrt{1-\rho^2}}{2\pi\sigma_x\sigma_y a^2(z)} \left(\int_{-\infty}^{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}} -u\, e^{-\frac{1}{2}u^2} \mathrm{d}u + \int_{\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}}^\infty u\, e^{-\frac{1}{2}u^2} \mathrm{d}u \right) = \tag{A5}$$

The integrals between the parenthesis in equation (A4) are the cumulative distribution functions (CDFs) of the standard normal distribution, $\Phi(x)$, by a factor of $\sqrt{2\pi}$. And the two integrals in equation (A5) are finite and immediate integrals. They are simplified into equation (A6), by means of the relation $-\Phi(-a) + \Phi(a) = \Phi(a) - 1 + \Phi(a) = 2\Phi(a) - 1$.

$$p_z(z) = \frac{b(z)d(z)}{\sqrt{2\pi}\sigma_x\sigma_y a^3(z)} \left(-\Phi\left(\frac{-b(z)}{a(z)\sqrt{1-\rho^2}}\right) + \Phi\left(\frac{b(z)}{a(z)\sqrt{1-\rho^2}}\right) \right) + \frac{d(z)\sqrt{1-\rho^2}}{2\pi\sigma_x\sigma_y a^2(z)} 2exp\left\{ -\frac{1}{2}\frac{b^2(z)}{(1-\rho^2)a^2(z)} \right\} =$$

$$= \frac{b(z)d(z)}{\sqrt{2\pi}\sigma_x\sigma_y a^3(z)} \left(2\Phi\left(\frac{b(z)}{a(z)\sqrt{1-\rho^2}}\right) - 1 \right) + \frac{\sqrt{1-\rho^2}}{\pi\sigma_x\sigma_y a^2(z)} exp\left\{ \frac{-c}{2(1-\rho^2)} \right\} \tag{A6}$$

Once the analytical equation (A6) has been found, it is compared with the results of our synthesis models. To do so, we have computed simulations which assume the same IMF as used in Section 3 (Salpeter 1955) with mass limits of $1 - 120\,M_\odot$, $Z = 0.020$, and $f_{esc} = 0$. We have computed two sets of 10^4 Monte Carlo simulations for the IMF. The first set contains 5000 stars per simulation (see Fig. A1) and the second one 500 stars (see Fig. A2). Once the IMFs are obtained we follow the evolution of each cluster. Figs. A1 and A2 show the joint distribution of Hα and FUV luminosity (top-left), the histograms of the marginal distributions of FUV (top-right) and Hα(bottom-left), and the Hα/FUV ratio distribution obtained from the simulations for an age of 2.8 Myr (bottom-right). We note that this age corresponds to the worse values of skewness γ_1 and kurtosis γ_2, where more deviation of the Gaussian case is expected (c.f. Fig. 3). The associated Gaussian distributions, with mean and variance obtained in Section 3 for the corresponding number of stars in each set, are overplotted in the histograms as well. We also overplot a Gaussian approximation for the H α/FUV ratio with mean equals to the ratio of the mean values of H α and FUV, and its dispersion is calculated by standard error propagation analysis (see, e.g. appendix in Cerviño et al. 2002).

Fig. A1, with 5×10^3 stars per simulation (similar to $1.5 \times 10^4\,M_\odot$ in the $1 - 120\,M_\odot$ range), shows that the Gaussian approximation for the integrated luminosities and the Hα/FUV ratio analytical distribution are in good agreement for Monte Carlo simulations. This result is consistent with the associated values of $\gamma_1(H\alpha) = 0.55$ and $\gamma_2(H\alpha) = 0.48$ for Hα, and $\gamma_1(FUV) = 0.26$ and $\gamma_2(FUV) = 0.09$ for FUV, associated to this number of stars. FUV is approximated by a Gaussian due to its lower γ_1 and γ_2 values. We also note that for such number of stars, metallicity and age, the relative standard deviation is around a 20 per cent, which is about the maximum test value we have assumed in our analysis. We check that the ratio distribution is also similar to a Gaussian distribution with a relative standard deviation of a 16 per cent.

Fig. A2 shows the latter analysis but for 5×10^2 stars per simulation (similar to $1.5 \times 10^3\,M_\odot$ in the $1 - 120\,M_\odot$ range). In this case, the mean value of $\log L(FUV) = 36.7$ is close to the LLL for this age. The luminosity distributions show an asymmetry with a high luminosity tail, and skewness and kurtosis values, $\gamma_1(H\alpha) = 1.74$ and $\gamma_2(H\alpha) = 4.80$, that make more appreciable the detach from gaussianity. The ratio analytical distribution also departs from the Monte Carlo simulations, and differences with the Gaussian case are also more relevant. However the global behavior of Monte Carlo distribution is still captured by the Hα/FUV ratio distribution. A peculiar feature of the ratio distribution obtained from Monte Carlo simulations is the presence of three local maximums. This is an artifact due to the assignation of close atmosphere model in the computation of the isochrones and it is implicit in SB99 models, what produces an artificial discretization of colours (see a discussion in Cerviño & Luridiana 2006). Such effect is also visible in the joint distribution of Hα and FUV which shows a discrete and well-aligned structure. In the case of lower number of stars (equivalent to lower masses or lower luminosities for a given age) the detach from a Gaussian distribution is larger, and a proper analysis would require to interpolate atmosphere models to avoid numerical artefacts in the Monte Carlo distributions.
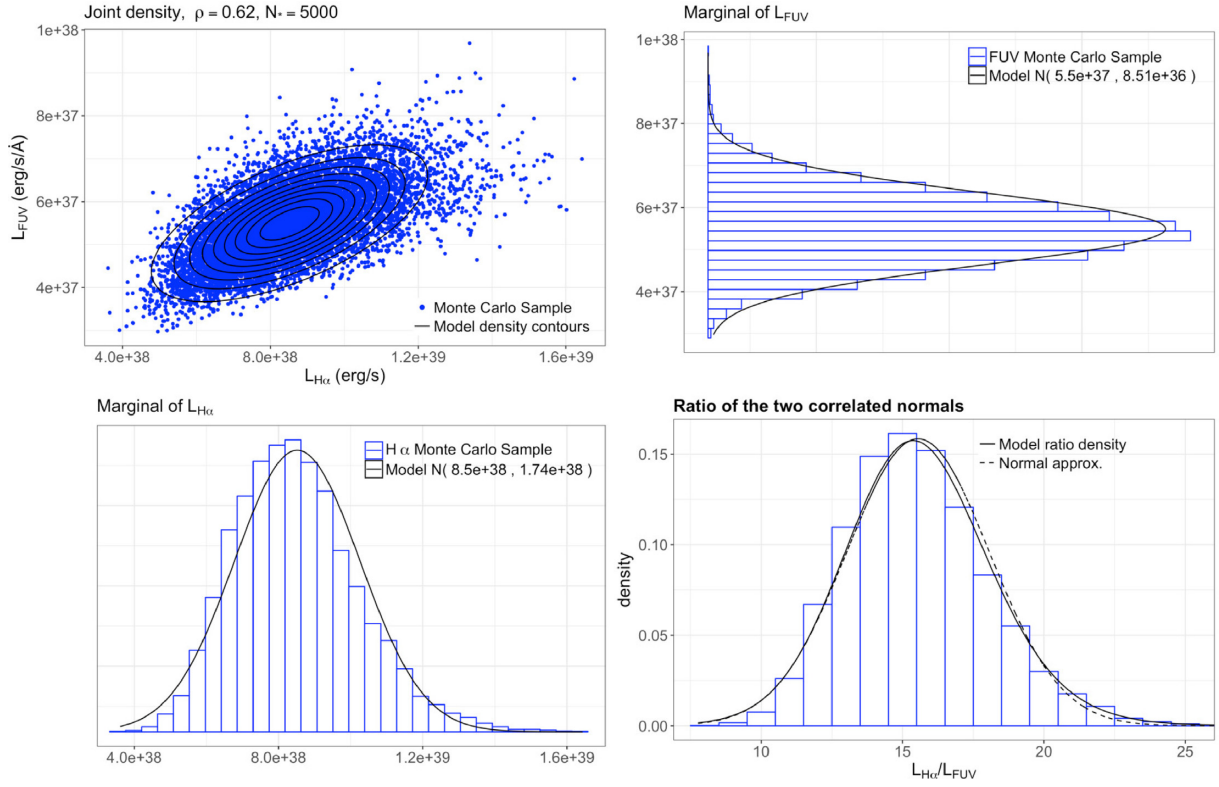
Figure A1. Joint distribution of Hα and FUV (top-left), histograms of the distribution of FUV (top-right) and Hα (bottom-left), and the ratio distribution obtained from the simulations for an age of 2.8 Myr obtained from 10^4 Monte Carlo simulations with 5000 stars each. Overplotted is a Gaussian distribution with the mean and variance obtained in Section 3. We also plot the analytic ratio distribution.
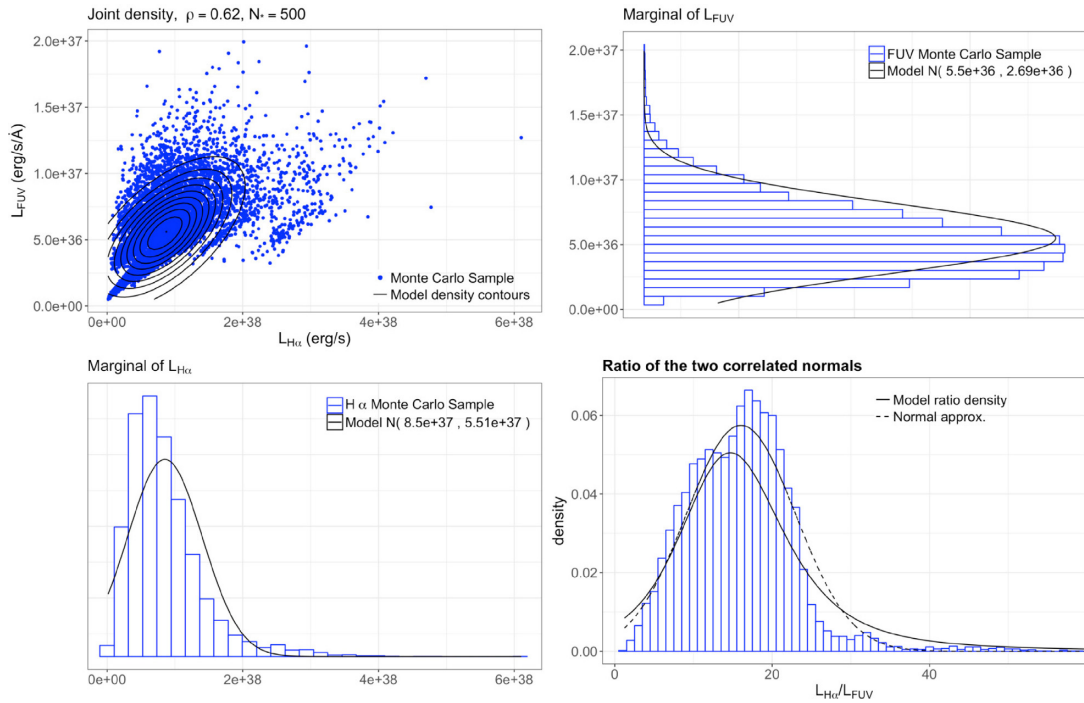
Figure A2. As Fig. A1, but for pixels with 500 stars.
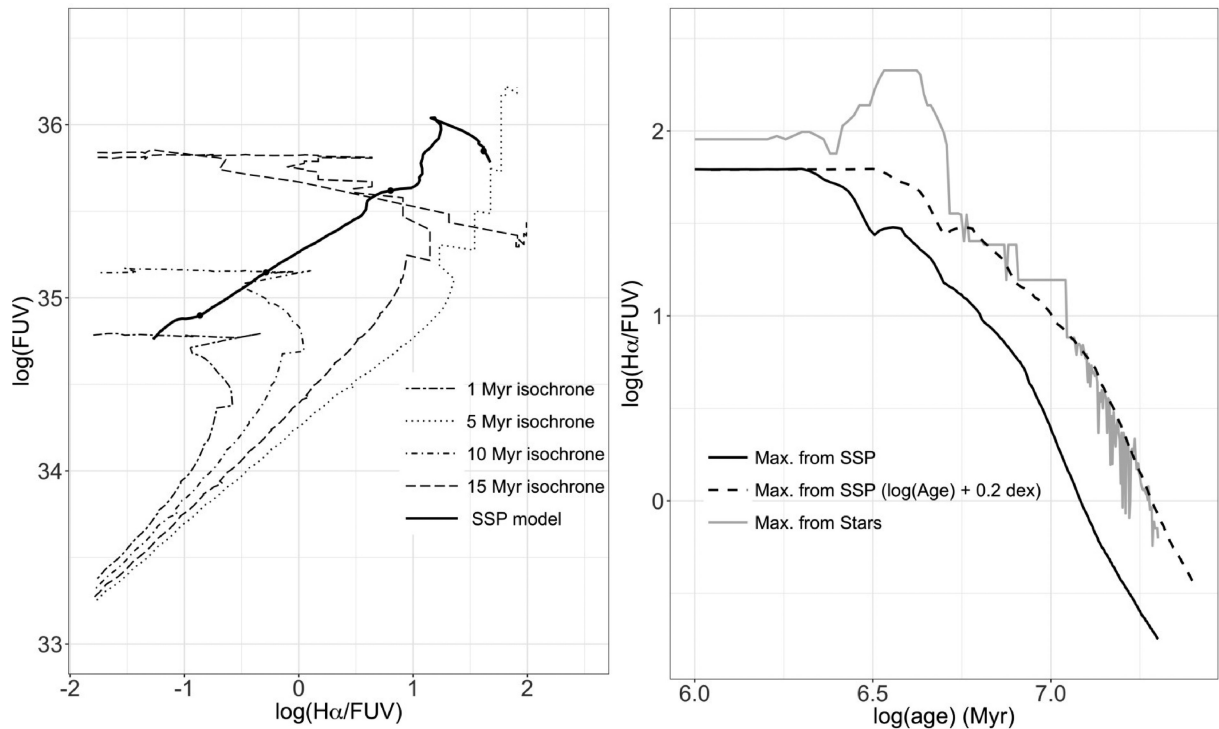
Figure A3. Left: Comparison of the position of SB99 isochrones at 1, 5, 10, and 15 Myr with Z=0.020 and synthesis models results in the Hα/FUV vs. FUV plane. Black symbols shows the synthesis models results at the quoted ages for the isochrones. The stairs-like behavior is due to the use of close atmosphere model approach in SB99 computations. Right: Maximum values of the Hα/FUV ratio from the full isochrone and SSP models set.

Given that in our pixel-by-pixel analysis we find some pixel values below the LLL, what is the error in our age determinations due to sampling effects? Notwithstanding, we can still evaluate qualitatively this effect by comparing the synthesis models results with the used isochrones, which is a general technique for cases of extreme sampling (see, e.g. Barker, de Grijs & Cerviño 2008).

The general situation is shown in the left-hand panel of Fig. A3, which shows isochrones for $Z = 0.020$ and ages of 1, 5, 10, and 15 Myrs. Synthesis models computations are overplotted, the same ages are shown as black points, and the FUV luminosity has been multiplied by 100 for clarity. The isochrones exhibit a step-like behaviour due to the close-atmosphere assignation approach, as described above. It can be observed that Hα/FUV ratio is close to the main sequence turn-off defined by the isocrones at the corresponding ages. SSP results are at the left of the main sequence turn-off except during the WR phase (5 Myrs), where they are at the right of the turn-off. Such a situation produces an asymmetry in ages inferences in the most extreme case of sampling effects (each pixel containing a single star).

In this situation, the ages inferred from the use of mean values of synthesis models and the Hα/FUV ratio would be severely overestimated. This is because the Hα/FUV ratio of stars at a given age would span towards low values. On the other hand, the same ages would only be moderately underestimated since the possible maximum Hα/FUV ratio for a given age is similar to the maximum stellar value. I.e., the actual age would be much younger (lower) than the inferred one, but for sure it cannot be much older (larger) than our inference. Right panel in A3 shows the comparison of the maximum Hα/FUV ratio that a single star would reach as a function of the age. The maximum mean value of Hα/FUV is obtained from synthesis models results considering all possible cases for metallicity and f_{scp} considered in this work. Except for ages lower than the WR phase age, the use of the mean value would be obtained by standard SSP codes produced, in the worst case, an underestimate of $\log(Age)$ of 0.2 dex.

For a final evaluation of sampling effects with our methodology, we have computed sets of 10^4 Monte Carlo simulations for each metallicity and the considered age range. Assuming the number of stars is in the range from 50 to 5×10^4, following a power law with exponent -1 in order to increases the relevance of sampling effects and assuming too a flat distribution of f_{scp}. The implicit distributions are similar to the those presented in Section 4.1, except by the inclusion of the distribution in the number of stars and the use of discrete values in the Z distribution. Despite there is still the problem of using close atmosphere model approach, this set of simulations allows us to obtain a first order comparison of the method performance of Section 4.1.

The method followed here is a naïve direct inversion of the problems from Monte Carlo simulations. This is, the resulting distributions are obtained from Monte Carlo simulations in the rectangular region defined by $H\alpha_{obs} \pm \sigma_{obs}(H\alpha)$ and $FUV_{obs} \pm \sigma_{obs}(FUV)$. We note that it is only a first-order approximation (a correct analysis would require to weight the region by a 2D uncorrelated Gaussian distribution, and to interpolate atmosphere models, which is outside the scope of this paper).

Fig. A4 displays the distribution of ages for the same points shown in Fig. 6. We note that the age distribution is similar for large (young) and small (old) Hα/FUV values, although Monte Carlo simulations show a fat tail that extends towards lower ages, and a narrow tail for larger ages. As explained above, this feature dues to the asymmetry between isochrones and synthesis models computations, as well as our assumption for the distribution of the number of stars which favors extreme sampling effects. The situation is more extreme for the
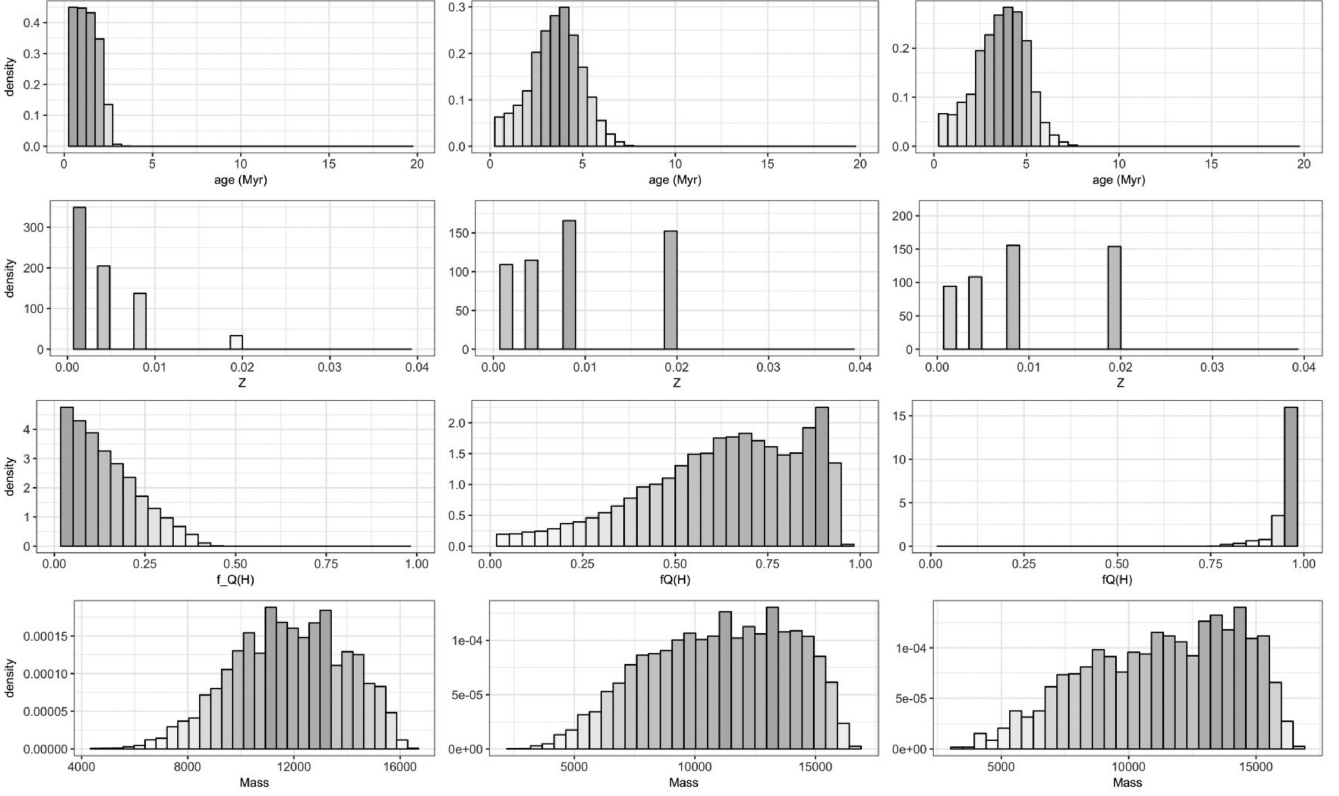
Figure A4. Samples of the probability distributions of the different parameters obtained forma a direct inversion of Monte Carlo simulations, for some pixels of the image with different modal ages. These plots can be compared with the ones in Fig. 6.

intermediate Hα/FUV values. For well-sampled clusters there is an equilibrium between normal and WR stars, whereas for subsampled pixels normal stars (hence lower ages) are preferred.

APPENDIX B: NESTED SAMPLING ALGORITHM FOR SAMPLING THE MARGINAL POSTERIOR DISTRIBUTION FOR THE AGE PARAMETER

In the Bayesian framework, inference proceeds by estimating the posterior distribution $p(\boldsymbol{\theta}|\hat{r})$ and then marginalizing to obtain the age distribution of the region under study according to equation (9)

$$p(\text{age}|\hat{r}) = p(\theta_1|\hat{r}) = \iint p(\boldsymbol{\theta}|\hat{r}) \, d\theta_2 d\theta_3$$

where the posterior distribution $p(\boldsymbol{\theta}|\hat{r})$ was characterized by an independent and identical distributed sample obtained by using the Nested Sampling Algorithm (NSA; Skilling 2006).

The main objective of NSA is the estimation of the *evidence*, $p(\hat{r})$, but as a by-product we can obtain an independent sample of the posterior distribution. The NSA is based on the relationship between the likelihood $L(\boldsymbol{\theta}) = p(\hat{r}|\boldsymbol{\theta})$ and the *prior volume* $X(\lambda)$ defined as the bulk of prior distribution contained within an iso-contour of the likelihood:

$$X(\lambda) = \int_{p(\hat{r}|\boldsymbol{\theta})>\lambda} p(\boldsymbol{\theta}) \, d\boldsymbol{\theta}. \tag{B1}$$

As λ increases $X(\lambda)$ decreases from $X(0) = 1$ to $X(\infty) = 0$. See Skilling (2006), their Figs. 4 and 5 as a perfect example of the nested sampling strategy and how it works. The likelihood and the evidence are related by

$$p(\hat{r}) = \int p(\hat{r}|\boldsymbol{\theta}) dX \tag{B2}$$

The key of the NSA is to evaluate numerically integral (B2) using directly the prior volume X, instead of rastering over $\boldsymbol{\theta}$, which becomes impractical for high dimensions (we only deal with three parameters). For any size of $\boldsymbol{\theta}$, the evaluation of the integral in (B2) becomes a one-dimensional integral over the unit range

$$p(\hat{r}) = \int_0^1 L(X) \, dX \tag{B3}$$
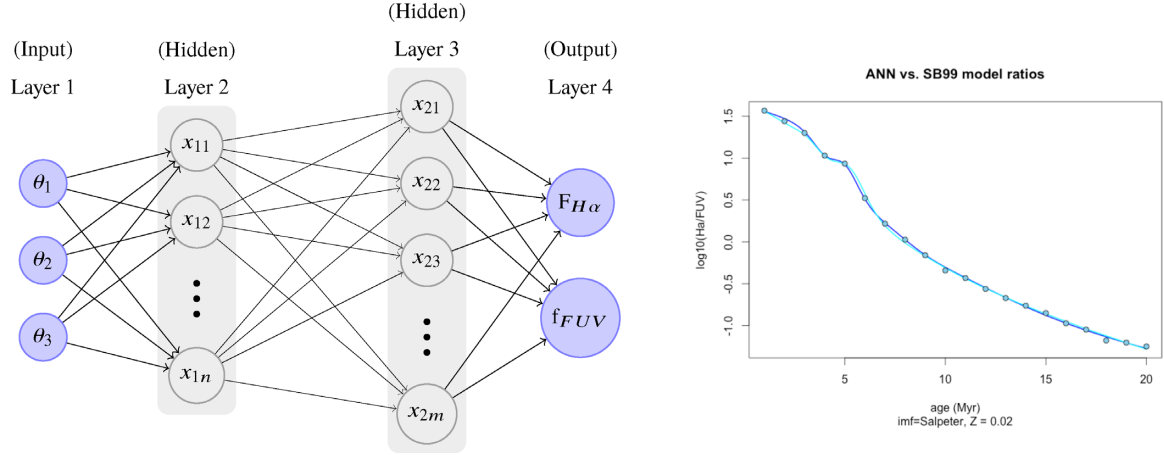
Figure B1. *On the left:* Multilayer Perceptron for the estimation of the Hα and FUV luminosities from the SED given by SB99 (output layer), depending on the input parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (age, IMF, Z)$. The ANN was developed with two hidden layers with n=20 and m=50 nodes respectively, and it was trained with 70 per cent and validate with 30 per cent of the grid points. The uncertainties due this interpolation were consider negligible. *On the right:* Comparison of the ANN (interpolation for the Hα to FUV flux ratio given different parameters than in SB99. Blue line corresponds to the ratios from the ANN given straightforward the model ratios, whereas the cyan line when the Hα and FUV fluxes are calculated separately. They are in good agreement.
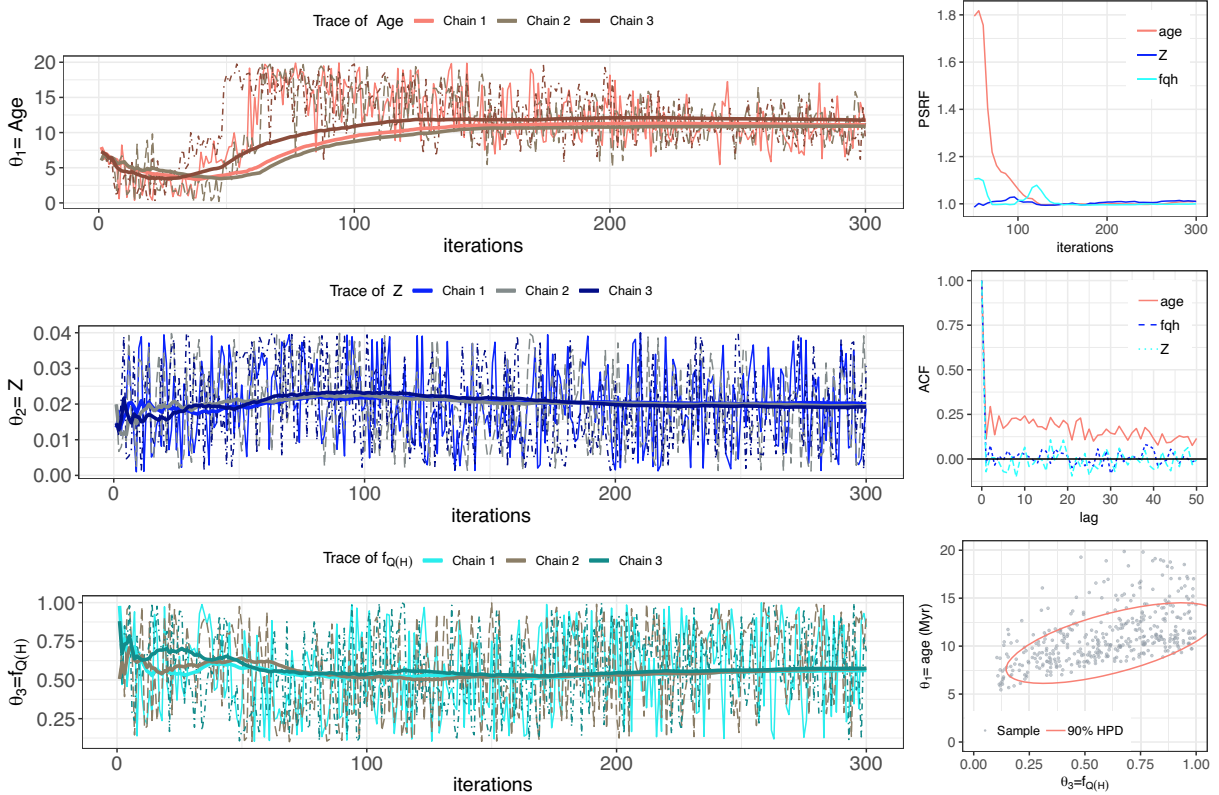
Figure B2. Assessing mixing and convergence. *On the left:* Trends of the estimated parameters, for three different chains obtained from the NSA algorithm application. After the first 100 iterations, all the chains are mixed and stable. *On the right:* On the top, the Potential Scale Reduction Factor (PSRF) for the three parameters, always lower than 1.1, indicates a fast convergence of the algorithm. At the middle a plot with the level of autocorrelation in each property, up to a maximum number of 50 lags. At the bottom, an example of the convergence towards the target distribution (red line indicates the 90 per cent Highest Posterior Density interval). After a burn-in period of 100 iterations, the chain rapidly converges towards a specific region of the space of parameters in regard of the age, according to the given Hα to FUV flux ratio. Whereas it remains spread on the $f_{Q(H)}$, as it is expected and already shown in the bidimensional posteriors distributions at Fig. 7.

being $L(X)$ the inverse function of $X(\lambda)$, and $dX = p(\boldsymbol{\theta})d\boldsymbol{\theta}$. The numerical evaluation of (B3) implies dividing the prior X into tiny elements and sorting them by likelihood volume

$$0 < X_m < \ldots < X_2 < X_1 < 1$$

Thus the integral (B3) can be estimated as a weighted sum of the corresponding likelihoods:

$$p(\hat{r}) \simeq \sum_{i=1}^{m} L(X_i)w_i \tag{B4}$$

with $w_i = (X_{i-1} - X_{i+1})/2$.

The resulting sequence of parameters $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^N$, for a run with N objects, already allows posterior samples to be extracted from the evidence calculation as

$$p(\boldsymbol{\theta}^i|\hat{r}) = \frac{L_i e^{-i/N}}{p(\hat{r})} \tag{B5}$$

Thorough explanation and details of the NSA can be found at Skilling ([2006](#)). Algorithm 1 sketches the NSA applied to each pixel of the observed flux ratio image $\hat{r} = \hat{F}_{H\alpha}/\hat{F}_{FUV}$. The full codes implemented in R can be found in https://github.com/carmensg/Age-maps. Fig. B2 shows several tests for assessing mixing and convergence of the chains.

Algorithm 1: Nested Sampling Algorithm

Data: the observed flux ratio, $\hat{r} = \hat{F}_{H\alpha}/\hat{F}_{FUV}$,
of a certain pixel of the ratio image.
Result: evidence $p(\hat{r})$ calculation, and the nested sequence of params. $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$
initialization;
$x_0 \leftarrow 1$
$z_0 \leftarrow 0$ ▷ z is the evidence, i.e. $p(\hat{r})$
$j \leftarrow 0$
repeat
    Draw N objects $\boldsymbol{\theta}_1^{(j)}, \ldots, \boldsymbol{\theta}_N^{(j)}$ from prior (1)
    for $i = 1, \ldots, N$ **do**
        Use $\boldsymbol{\theta}_i^{(j)}$ to generate the model params.

        $\{F_{H\alpha}, f_{FUV}, \sigma_{H\alpha}, \sigma_{FUV}, \rho\}_i$

        with SB99+ANN, equations (2) to (5)
        Compute the likelihood $L_i^{(j)} \leftarrow$ equation (6)
    end
    $j \leftarrow j + 1$
    Record $L_j = \min_{1 \leq i \leq N}\{L_i^{(j-1)}\}$
    $x_j \leftarrow e^{-j/N}$ ▷ *crude estimation*
    $w_j \leftarrow \frac{1}{2}\left(e^{-(j-1)/N} - e^{-(j+1)/N}\right)$
    $z_j \leftarrow z_{j-1} + L_j w_j$ ▷ *approx. by equation* (B4)
until $L_{max}x_j < fz_j$ & $j < M$;
    ▷ terminate the main loop when the largest current likelihood would not increase the current evidence by more than small fraction f

---

APPENDIX C: IMAGE SEGMENTATION

We describe the fully Bayesian approach for the model presented at Section 4.2. Where the aim is to cluster pixels into homogeneous classes, without prior definition of those classes, based only on spatial coherence. Suppose that the 'true' image $\boldsymbol{x}$ can only take a discrete set of values $\mu_1, \ldots, \mu_G$, and the assumed prior on $\boldsymbol{x}$ is a Potts model with G categories, equation (11). The variables are considered as random bidimensional arrays $\boldsymbol{x} = \{x_i, i \in \mathcal{I}\}$, whose elements are indexed by the lattice $\mathcal{I}$ (the location of the pixels), and related through a neighborhood relation.

The observed image $\boldsymbol{y}$ is a degraded version of $\boldsymbol{x}$ by additive Gaussian noise, where y_i are conditional independent given $\boldsymbol{x}$. So the distribution of $\boldsymbol{y}$ is given by equation (12).

Using Bayes theorem we know how to construct the posterior distribution of $\boldsymbol{x}|\boldsymbol{y}$. The aim is to draw inferences about $\boldsymbol{x}$ based on the posterior $p(\boldsymbol{x}|\boldsymbol{y})$.

Algorithm 2: Hybrid Gibbs Algorithm for the Image segmentation

---

Data: y observed flux ratio image, G categories

Result: sampling the posterior distr. of x over the full conditional distr. of the model parameters $\{\mu, \sigma, \beta\}$

initialization;

$x, \mu^{(0)}, \sigma^{2(0)} \leftarrow 0, \beta^{(0)} \leftarrow 1$

$\{xcum_{ig}\}_{|\mathcal{I}| \times G} \leftarrow 0$ for estimation of the $\hat{x}_i^{MPM}$ (C5)

for $t = 1, \ldots, niter$ **do**

    $S(x) \leftarrow 0$

    for $i = 1, \ldots, |\mathcal{I}|$ **do**

        for $g = 1, \ldots, G$ **do**

            $n_{i,g} = \sum_{j \sim i} \mathbb{1}_{x_i = g}$ store num. of neighbours of the class g

        end

        Draw $x_i | y, \mu^{(t-1)}, \sigma^{2(t-1)}, \beta^{(t-1)}$ from post. d. (13)

        $xcum_{i,x_i} \leftarrow xcum_{i,x_i} + 1$

        $S(x) \leftarrow S(x) + n_{i,g}$

    end

    for $g = 1, \ldots, G$ **do**

        Draw $\mu_g^{(t)}$ from truncated normal (C2)$\mathcal{N}\left(\frac{\sum_{i \in \mathcal{I}} \mathbb{1}_{x_i = g} y_i}{\sum_{i \in \mathcal{I}} \mathbb{1}_{x_i = g}}, \frac{\sigma^{2(t-1)}}{\sum_{i \in \mathcal{I}} \mathbb{1}_{x_i = g}} \right) \mathbb{1}_{\left[\mu_{g-1}^{(t)}, \mu_{g+1}^{(t-1)} \right]}$

    end

    Draw $\sigma^{2(t)}$ from $Inv - \Gamma\left(|\mathcal{I}|^2 / 2, \sum_{i \in \mathcal{I}} (y_i - \mu_{x_i})^2 / 2 \right)$ (C3)

    Draw $\beta^{(t)}$ from post. distr. (C4) with MCMC sampler:

        Draw $\tilde{\beta}$ from the uniform proposal:

          $\mathcal{U}\left(\beta^{(t-1)} - h, \beta^{(t-1)} + h \right)$

        Compute the acceptance ratio:

    $\rho = \left\{ \frac{Z(\beta^{(t-1)})}{Z(\tilde{\beta})} exp\left((\tilde{\beta} - \beta^{(t-1)}) S(x) \right) \right\} \wedge 1$

    Set $\beta^{(t)} = \tilde{\beta}$ with probability ρ

end

We treat the parameters β, σ, μ as variables, this is as hyper parameters. As they are nuisance parameters, and there is no additional prior information, the chosen priors are the following uniforms

$$\beta \sim \mathcal{U}(0, 2)$$

$$\mu_g \sim \mathcal{U}(y_{min}, y_{max})$$

$$p(\sigma^2) \propto \sigma^{-2}$$

as described in Section 4.2.

Assuming independent priors for these parameters, the joint posterior distribution is therefore

$$p(x, \beta, \sigma^2, \mu | y) \propto p(x, \beta, \sigma^2, \mu)\, p(y | x, \beta, \sigma^2, \mu)$$

$$\propto p(\beta)\, p(\sigma^2)\, p(\mu)\, p(x | \beta)\, p(y | x, \beta, \sigma^2, \mu)$$

$$\propto \frac{(\sigma^2)^{-(|\mathcal{I}|/2 + 1)}}{Z(\beta)} \times exp\left\{ \beta \sum_{j \sim i} \mathbb{1}_{x_j = x_i} - \frac{1}{2\sigma^2} \sum_{i \in \mathcal{I}} (y_i - \mu_{x_i})^2 \right\} \tag{C1}$$

For sampling purposes (implemented in the hybrid Gibbs Algorithm 2), it is necessary the full conditionals distributions of the parameters (see Marin & Robert 2014, for more information). The conditional distribution for x is given in equation (13), $P(x_i = g | y, \beta, \sigma^2, \mu) \propto exp\left\{ \beta \sum_{j \sim i} \mathbb{1}_{x_j = g} - \frac{1}{2\sigma^2} (y_i - \mu_g)^2 \right\}$.

Given x the pixels associated to each category g can be separated, so the parameters μ_g can be simulated indepentdently as

$$P(\mu_g | y, x, \sigma^2) \propto exp\left\{ -\frac{1}{2\sigma^2} \sum_{i : x_i = g} (y_i - \mu_g)^2 \right\}$$

$$\propto exp\left\{ -\frac{n_g}{2\sigma^2} \left(\mu_g - \frac{s_g}{n_g} \right)^2 \right\} \tag{C2}$$

a truncated normal on $[y_{min}, y_{max}]$, or $[\mu_{g-1}, \mu_{g+1}]$ if we establish the ordering $y_{min} \leq \mu_1 \leq \ldots \leq \mu_G \leq y_{max}$ (setting $\mu_0 = y_{min}$ and $\mu_{G+1} = y_{max}$). Where $n_g = \sum_{i \in \mathcal{I}} \mathbb{1}_{x_i = g}$ and $s_g = \sum_{i \in \mathcal{I}} \mathbb{1}_{x_i = g} y_i$.

The conditional distribution of σ^2 is

$$P(\sigma^2|\boldsymbol{y}, \boldsymbol{x}) \propto (\sigma^2)^{-(|\mathcal{I}|/2+1)} exp \left\{ \frac{-1}{2\sigma^2} \sum_{i \in \mathcal{I}} (y_i - \mu_{x_i})^2 \right\} \tag{C3}$$

an inverse gamma distribution with parameters $|\mathcal{I}|^2/_2$ and $\sum_{i \in \mathcal{I}} (y_i - \mu_{x_i})^2/_2$.

And finally, the conditional distribution of β is

$$p(\beta|\boldsymbol{x}) \propto \frac{1}{Z(\beta)} exp \left(\beta \sum_{j \sim i} \mathbb{I}_{x_j=x_i} \right) \mathbb{I}_{[0,\beta_{crit}]} \tag{C4}$$

since β only depends on $\boldsymbol{x}$. Together the difficulty of the calculation of the normalising coonstant $Z(\beta)$, we have this is no for longer a Potts model, or some known distribution function. We follow the *path sampling* methodology to estimate $Z(\beta)$, described in Marin & Robert (2014), in order to implement the MCMC sampler at the end of the Algorithm 2.

To give an estimation of the corresponding image $\boldsymbol{x}$ of the flux ratio, given the observed one $\boldsymbol{y}$, classified by G homogeneous regions with a common flux ratio mean μ_g we use the associated marginal posterior mode (MPM) estimator.

$$\widehat{x}_i^{MPM} = argmax_{1 \le g \le G} \mathbb{P}^\pi (x_i = g|y), i \in \mathcal{I}$$

$$\simeq \max_{1 \le g \le G} \sum_{n=1}^{N} \mathbb{I}_{x_i^{(n)}=g} \tag{C5}$$

The later equation gives an approximation of the MPM based on a simulation $\{\boldsymbol{x}^{(n)}\}_{n=1,...,N}$ from the posterior distribution of $\boldsymbol{x}$, Marin & Robert (2014).

All the codes implemented in R can be found in https://github.com/carmensg/Age-maps.

APPENDIX D: AGE MAPS FOR THE GALAXIES FROM PAPER I

Next, the age maps for the galaxies sample in Paper I are also included. In Figs. D1–D6, top plots show the comparison between the age maps obtained in Paper I with those applying the Bayesian approach, from Section 4.1.

At bottom plots it is checked the uncertainties in age estimation when it is used another estimator than the Mode. For example the percentiles of the resulting sample from the age posterior probability function. The extreme cases determine a 90 per cent posterior credible interval for the age estimation.
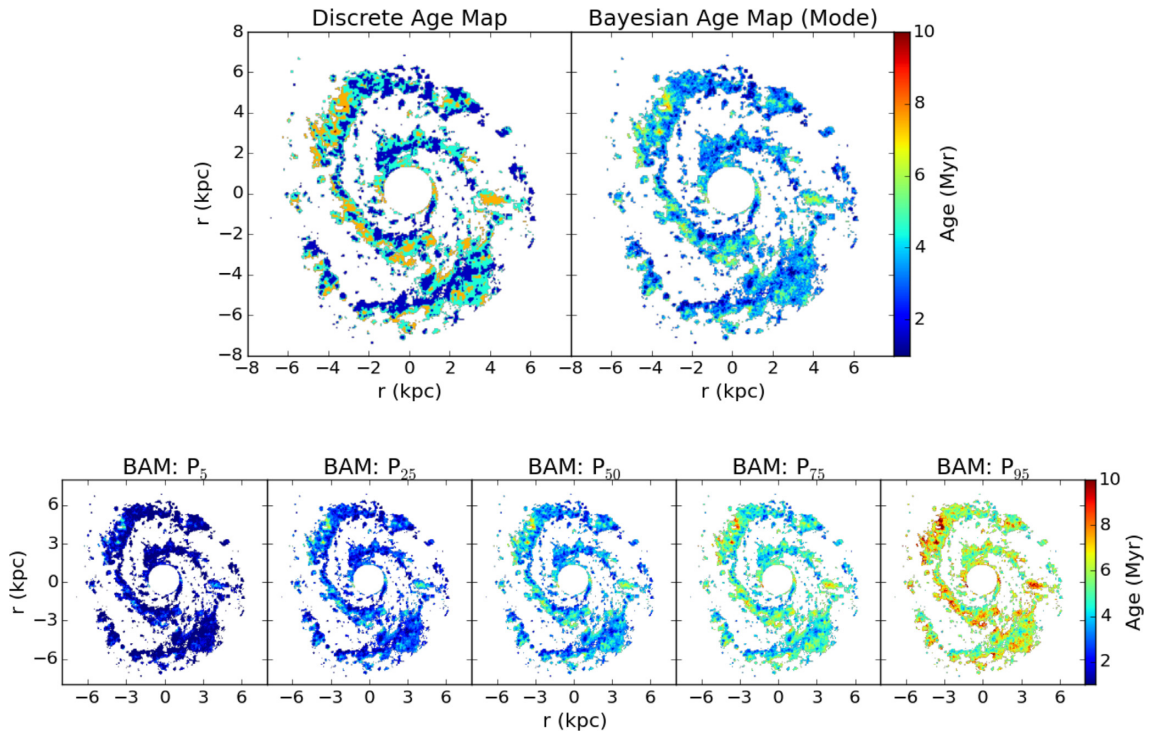
Figure D1. M51 age maps. *Top left*: The discrete age map obtained in Paper I. *Top right*: Bayesian age map, calculated by method of Sec. 4.1. *Bottom*: Different Bayesian age maps (BAM), calculated in Sec. 4.1, but taken this time 5 per cent, 25 per cent, 50 per cent, 75 per cent, and 95 per cent percentiles from the age posterior probability distribution. Axes are the distance to the galactic centres (kpc units).
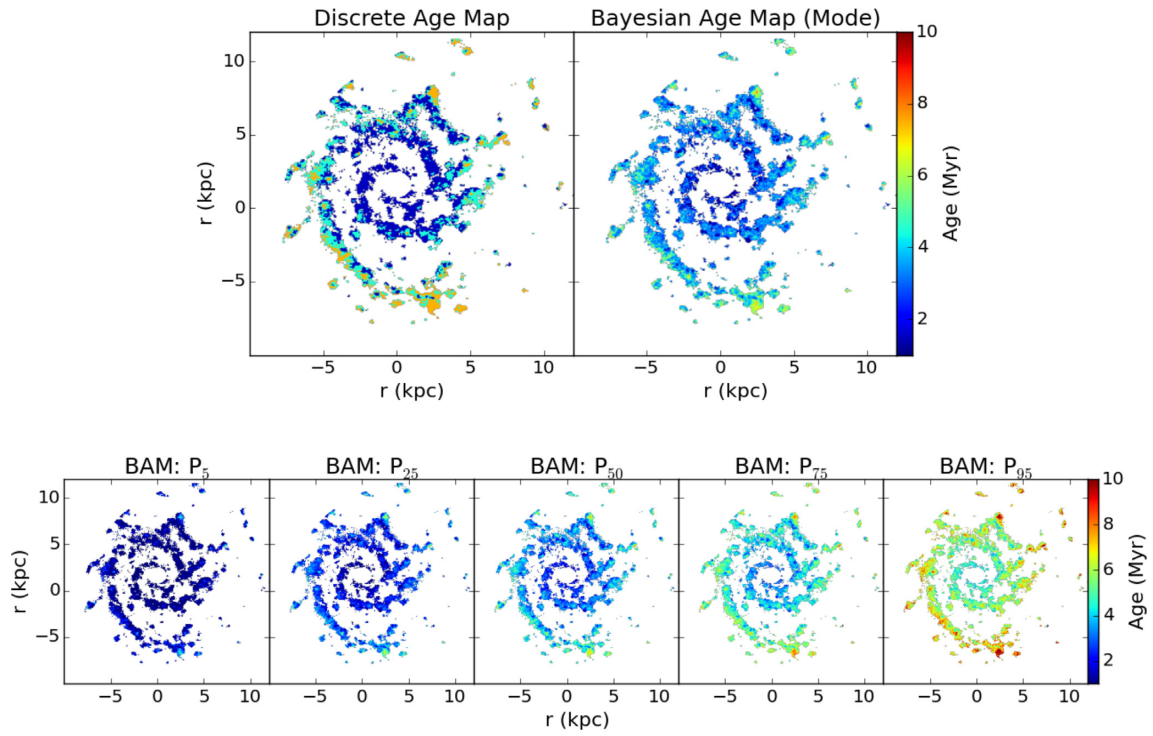
Figure D2. Same as Fig. D1, but for M74.
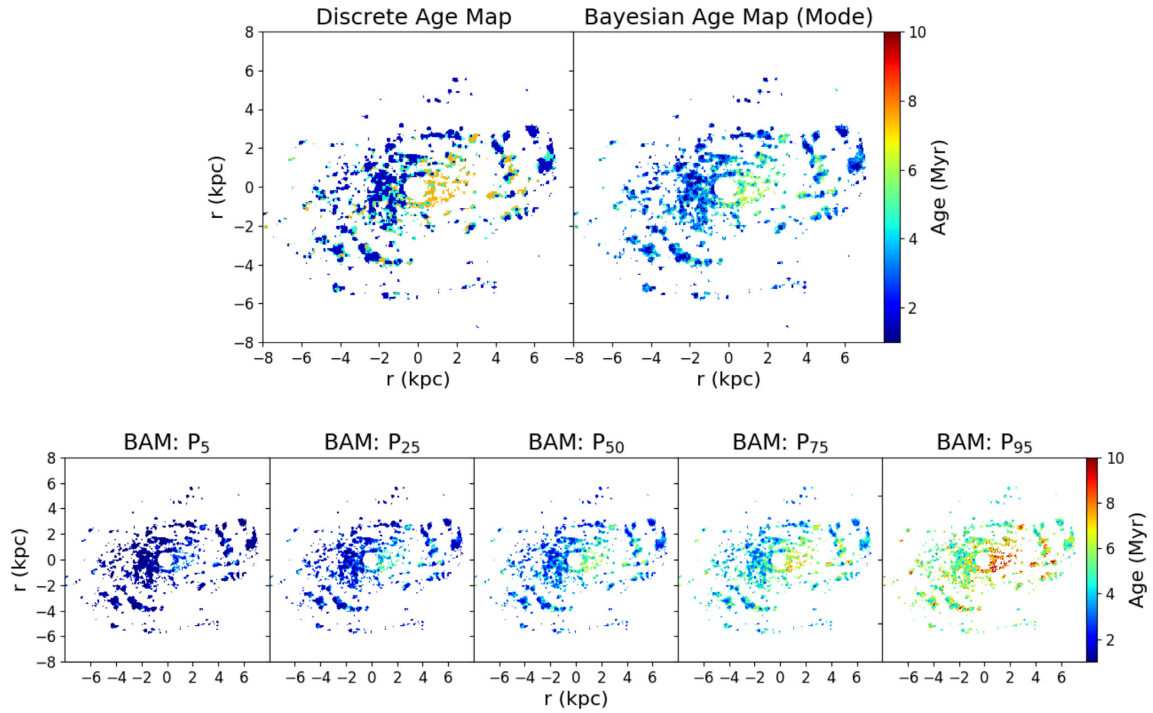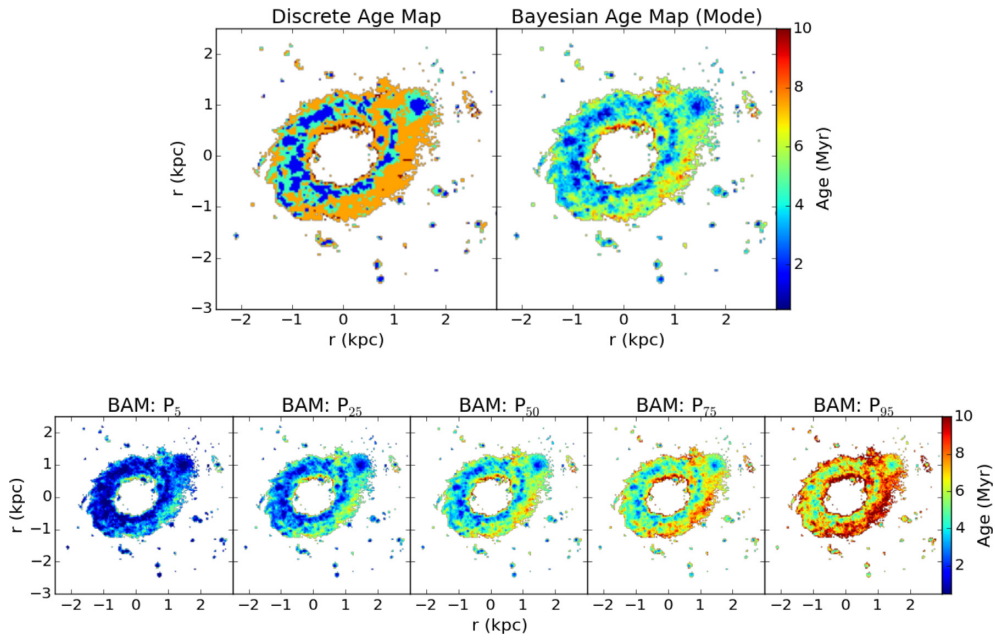
Figure D3. Same as Fig. D1, but for M63.

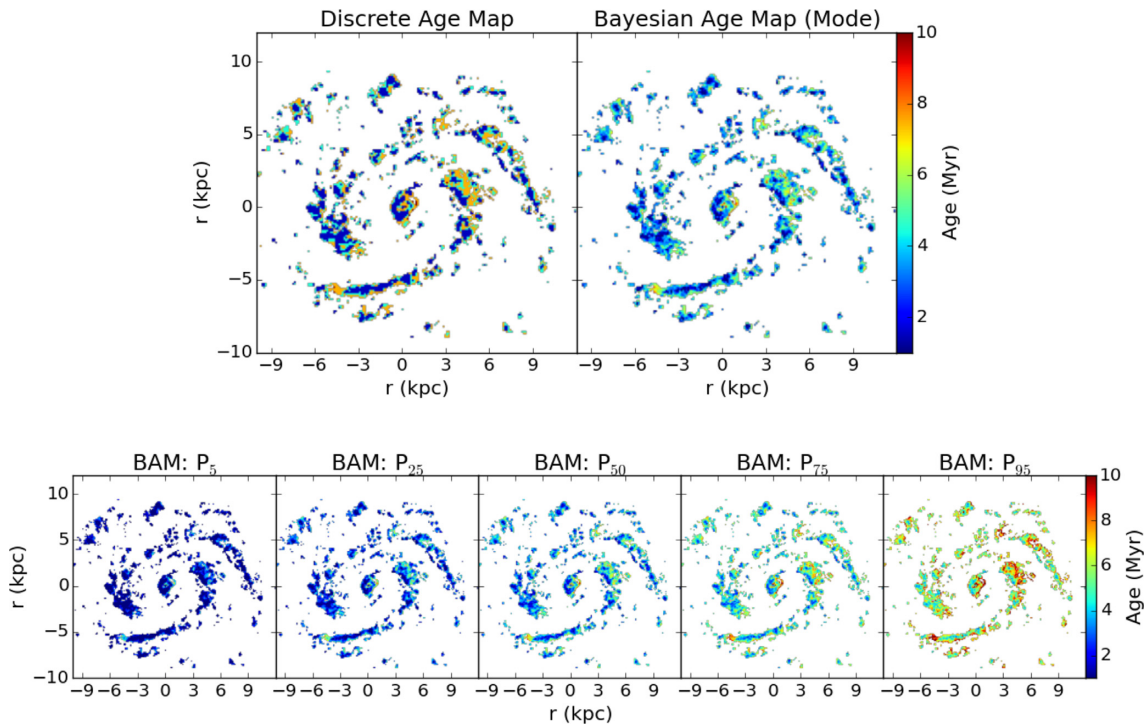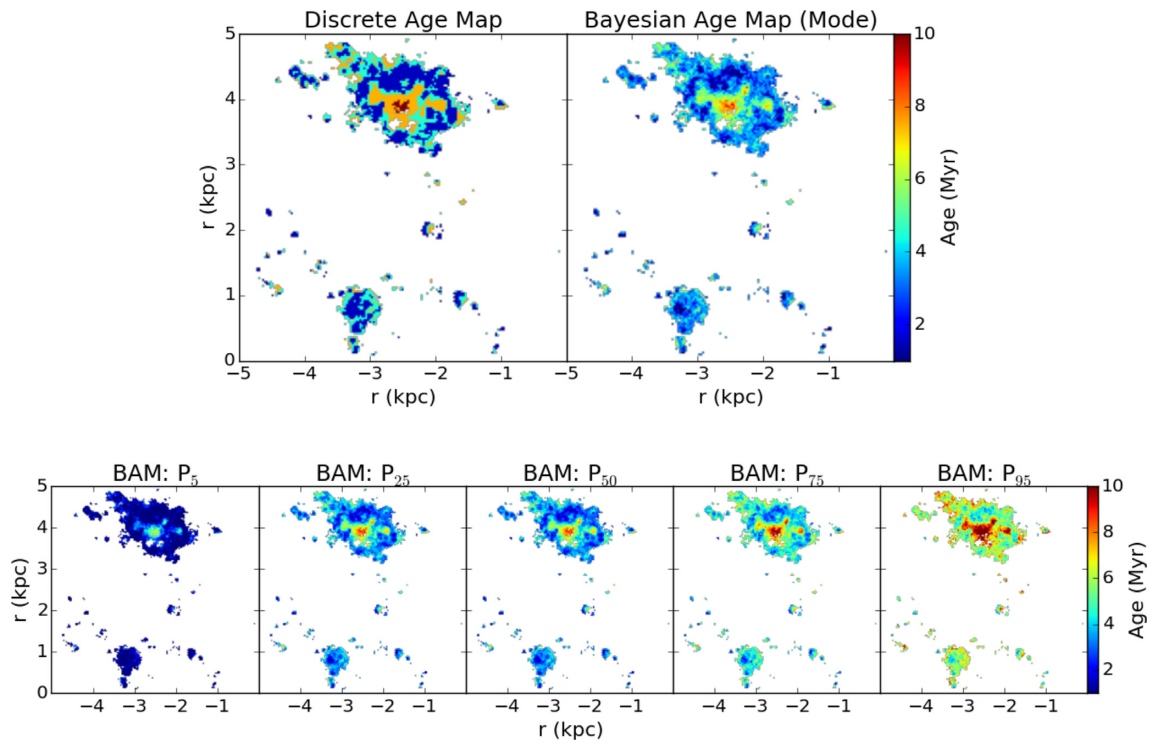**Figure D4.** Same as Fig. D1, but for M94.

Figure D5. Same as Fig. D1, but for M100.

Figure D6. Same as Fig. D1, but for IC2574.

This paper has been typeset from a TEX/LATEX file prepared by the author.