

# Retrieval of the physical parameters of galaxies from WEAVE-StePS-like data using machine learning

J. Anghopo<sup>1,\*</sup>, B. R. Granett<sup>1</sup>, F. La Barbera<sup>2</sup>, M. Longhetti<sup>1</sup>, A. Iovino<sup>1</sup>, M. Fossati<sup>3</sup>, F. R. Ditrani<sup>1,3</sup>, L. Costantin<sup>4</sup>, S. Zibetti<sup>5</sup>, A. Gallazzi<sup>5</sup>, P. Sánchez-Blázquez<sup>6,7</sup>, C. Tortora<sup>2</sup>, C. Spiniello<sup>8,2</sup>, B. Poggianti<sup>9</sup>, A. Vazdekis<sup>10,11</sup>, M. Balcells<sup>10,11,12</sup>, S. Bardelli<sup>13</sup>, C. R. Benn<sup>12</sup>, M. Bianconi<sup>14</sup>, M. Bolzonella<sup>13</sup>, G. Busarello<sup>2</sup>, L. P. Cassarà<sup>15</sup>, E. M. Corsini<sup>16,9</sup>, O. Cucciati<sup>13</sup>, G. Dalton<sup>8,17</sup>, A. Ferré-Mateu<sup>10,11</sup>, R. García-Benito<sup>18</sup>, R. M. González Delgado<sup>18</sup>, E. Gafton<sup>12</sup>, M. Gullieuszik<sup>9</sup>, C. P. Haines<sup>19,1</sup>, E. Iodice<sup>2</sup>, A. Ikhsanova<sup>16</sup>, S. Jin<sup>8,20</sup>, J. H. Knapen<sup>10,11</sup>, S. McGee<sup>14</sup>, A. Mercurio<sup>21,2</sup>, P. Merluzzi<sup>2</sup>, L. Morelli<sup>19,1</sup>, A. Moretti<sup>9</sup>, D. N. A. Murphy<sup>22</sup>, A. Pizzella<sup>16,9</sup>, L. Pozzetti<sup>13</sup>, R. Ragusa<sup>2</sup>, S. C. Trager<sup>20</sup>, D. Vergani<sup>13</sup>, B. Vulcani<sup>9</sup>, M. Talia<sup>23,13</sup>, and E. Zucca<sup>13</sup>

(Affiliations can be found after the references)

Received 14 March 2024 / Accepted 7 June 2024

## ABSTRACT

**Context.** The *William Herschel* Telescope Enhanced Area Velocity Explorer (WEAVE) is a new, massively multiplexing spectrograph that allows us to collect about one thousand spectra over a 3 square degree field in one observation. The WEAVE Stellar Population Survey (WEAVE-StePS) in the next 5 years will exploit this new instrument to obtain high-S/N spectra for a magnitude-limited ( $I_{AB} = 20.5$ ) sample of  $\sim 25\,000$  galaxies at moderate redshifts ( $z \geq 0.3$ ), providing insights into galaxy evolution in this as yet unexplored redshift range.

**Aims.** We aim to test novel techniques for retrieving the key physical parameters of galaxies from WEAVE-StePS spectra using both photometric and spectroscopic (spectral indices) information for a range of noise levels and redshift values.

**Methods.** We simulated  $\sim 105\,000$  galaxy spectra assuming star formation histories with an exponentially declining star formation rate, covering a wide range of ages, stellar metallicities, specific star formation rates (sSFRs), and dust extinction values. We considered three redshifts (i.e.  $z = 0.3, 0.55$ , and  $0.7$ ), covering the redshift range that WEAVE-StePS will observe. We then evaluated the ability of the random forest and K-nearest neighbour algorithms to correctly predict the average age, metallicity, sSFR, dust attenuation, and time since the bulk of formation, assuming no measurement errors. We also checked how much the predictive ability deteriorates for different noise levels, with  $S/N_{I,obs} = 10, 20$ , and  $30$ , and at different redshifts. Finally, the retrieved sSFR was used to classify galaxies as part of the blue cloud, green valley, or red sequence.

**Results.** We find that both the random forest and K-nearest neighbour algorithms accurately estimate the mass-weighted ages,  $u$ -band-weighted ages, and metallicities with low bias. The dispersion varies from  $0.08$ – $0.16$  dex for age and  $0.11$ – $0.25$  dex for metallicity, depending on the redshift and noise level. For dust attenuation, we find a similarly low bias and dispersion. For the sSFR, we find a very good constraining power for star-forming galaxies,  $\log sSFR \geq -11$ , where the bias is  $\sim 0.01$  dex and the dispersion is  $\sim 0.10$  dex. However, for more quiescent galaxies, with  $\log sSFR \leq -11$ , we find a higher bias, ranging from  $0.61$  to  $0.86$  dex, and a higher dispersion,  $\sim 0.4$  dex, depending on the noise level and redshift. In general, we find that the random forest algorithm outperforms the K-nearest neighbours. Finally, we find that the classification of galaxies as members of the green valley is successful across the different redshifts and S/Ns.

**Conclusions.** We demonstrate that machine learning algorithms can accurately estimate the physical parameters of simulated galaxies for a WEAVE-StePS-like dataset, even at relatively low  $S/N_{I,obs} = 10$  per Å spectra with available ancillary photometric information. A more traditional approach, Bayesian inference, yields comparable results. The main advantage of using a machine learning algorithm is that, once trained, it requires considerably less time than other methods.

**Key words.** galaxies: evolution – galaxies: formation – galaxies: general – galaxies: stellar content

## 1. Introduction

Over the last two decades, several wide-area photometric and spectroscopic surveys have greatly improved our understanding of galaxy formation and evolution. Most notably, the combination of wide-area and pencil-beam spectroscopic surveys, including the Sloan Digital Sky Survey (SDSS; York 2000), the Galaxy And Mass Assembly (GAMA; Hopkins et al. 2013), zCOSMOS (Lilly et al. 2009), the VIMOS Public Extragalactic Redshift Survey (VIPERS; Guzzo et al. 2014), and 3D-HST (Momcheva et al. 2016), have pushed the boundaries of galaxy formation studies to a few billion years after the Big Bang. These spectroscopic efforts usually target well-known

fields, where multi-wavelength imaging campaigns provide deep complementary datasets, often covering the UV to near-infrared (NIR) parts of the electromagnetic spectrum. These surveys include the SDSS imaging survey, the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Koekemoer et al. 2011), UltraVISTA (McCracken et al. 2012), and the Hyper Suprime-Cam Subaru Strategic Prime (HSC-SSP; Aihara et al. 2018). Exploitation of these data in combination with numerical hydrodynamic simulations (Crain et al. 2015; Schaye 2015; Springel et al. 2017) and semi-analytic models (Somerville & Davé 2015, and references within) has allowed us to greatly advance our understanding of galaxy formation and evolution mechanisms.

One of the key discoveries of recent decades has been the existence of a bimodality in the star formation activity of galaxies

\* Corresponding author; james.angthopo@inaf.it

(Strateva et al. 2001; Baldry et al. 2004; Faber et al. 2007), with two distinct populations of galaxies: one star-forming population characterised by blue optical colours and a separate red-der population of quiescent galaxies. This bimodality holds locally and at high redshifts, up to  $z \approx 3\text{--}4$  (Fritz et al. 2014; de Graaff et al. 2021). Moreover, a broad separation of the galaxy population into two classes is found in different parameters, including colour–colour (Williams et al. 2009), star formation rate (or colour)–stellar mass (Salim 2014; Trayford et al. 2016; Phillipps et al. 2019; Wright & Lagos 2019; Nelson et al. 2019), and  $D_n(4000)$ –velocity dispersion (Anghopo et al. 2019) planes. Finally, an important sample of transitioning galaxies straddle the two main populations, often referred to as galaxies in the green valley (GV). The study of GV galaxies is essential to understanding the quenching mechanism of star formation. Faber et al. (2007) proposed multiple evolutionary paths across the GV depending on the quenching mechanism. This was further supported by Schawinski et al. (2014), who found that there are two distinct quenching timescales depending on the morphology of the galaxy: elliptical galaxies are thought to have a more rapid quenching than spiral galaxies.

To gain insights into the physical mechanisms related to the quenching of star formation and the origin of the aforementioned bimodality, it is essential to obtain accurate estimates of galaxy physical parameters. These parameters can be obtained by fitting observed galaxies to synthetic spectral energy distribution templates (Gavazzi et al. 2002; Ilbert et al. 2006) or by comparing spectral indices of observed galaxies to templates obtained with different star formation histories (SFHs; Gallazzi et al. 2005; Costantin et al. 2019; Anghopo et al. 2020; Ditrani et al. 2023). Furthermore, with access to a large quantity of spectroscopic and photometric data, sophisticated spectral fitting codes such as STARLIGHT (Cid Fernandes et al. 2005), pPXF (Cappellari 2017, 2023), BAGPIPES (Carnall et al. 2018), and Prospector (Johnson et al. 2021) have been developed to estimate various physical parameters and to reconstruct the SFHs of galaxies from joint fits of the available spectroscopy and photometry. A large scatter, however, remains in the estimates of these parameters due to the different physical recipes assumed by different codes. Pacifici et al. (2023) explore such differences for 15 different spectral fitting algorithms when used to estimate stellar masses, star formation rates (SFRs), and effective dust attenuation ( $A_V$ ) at  $z \sim 1$  and 3. They find systematic differences in the estimations of physical parameters by different algorithms of 0.1–0.3 dex.

Despite these systematics, modern spectral fitting codes provide accurate estimates of the physical parameters of galaxies, often thanks to Bayesian inference methods and advanced algorithms used to explore the high-dimensional likelihood space. However, existing and upcoming large surveys that will measure the spectro-photometry of tens of thousands to millions of galaxies are bringing the field into the realm of big data, making the computational time of traditional fitting codes a variable of growing interest. As an alternative, machine learning (ML) algorithms have been applied to estimate galaxy parameters. These codes have been used to estimate photometric redshifts (Ball et al. 2008; Li et al. 2022), SFRs or specific star formation rates (sSFRs; Stensbo-Smidt et al. 2017; Davidzon et al. 2019; Euclid Collaboration 2023), and metallicities (Simet et al. 2021), with a good agreement found between true and predicted values. ML algorithms can be broadly divided into two categories – supervised and unsupervised. Supervised ML algorithms require data with labels and are mostly used for the purposes of regression and classification. Unsupervised ML algorithms do not use any labels for the data and instead learn to

characterise the distribution of the dataset on their own. Unsupervised methods are predominantly used to group data points with similar properties as a tool for dimensionality reduction and data compression or to identify outliers.

Our study uses the former method – supervised ML for regression. We make use of random forest (RF) and K-nearest neighbour (KNN) algorithms to estimate physical parameters – including the average age of the stellar population, both the mass-weighted age (mwa) and the  $u$ -band-weighted age (uwa), the metallicity, the sSFR, the  $A_V$ , and the time from the bulk of the star formation – using both spectral and photometric information. Most of the analysis is performed on a simulated dataset that mimics the upcoming observations of the WHT Enhanced Area Velocity Explorer Stellar Population Survey (WEAVE-StePS; Iovino et al. 2023b) at the 4.2 m *William Herschel* Telescope (WHT) in La Palma. The WEAVE spectrograph has a large field of view,  $\sim 3$  square degrees, and huge multiplexing capabilities, with nearly 1000 spectra observed in a single pointing (Jin et al. 2024). The spectral coverage spans from  $\sim 3660\text{--}9590 \text{ \AA}$  at a resolution of  $R \sim 5000$  (Dalton et al. 2012, 2016). WEAVE-StePS aims to observe a magnitude-limited ( $I_{AB} = 20.5$ ) sample of  $\sim 25\,000$  galaxies, mostly between  $z = 0.3$  and  $0.7$  with median  $S/N_{I,obs} = 10$  per  $\text{\AA}$ . This survey is designed to bridge the redshift gap between SDSS and LEGA-C (van der Wel et al. 2016). The combination of the three surveys will allow us to directly study galaxy evolution on a long and continuous span of cosmic time, over nearly 8 billion years.

The paper is structured as follows: In Sect. 2 we describe the two ML algorithms and the models used to create the synthetic galaxy templates. We also describe the procedure to simulate realistic spectra and photometry with measurement uncertainties. Section 3 outlines our main findings, and we test how well we can retrieve the physical parameters, both in the absence of measurement errors and with realistic uncertainties on spectra and photometry. Additionally, we investigate the variations in retrieval capabilities for data with different S/N values and redshifts. In Sect. 4 we classify galaxies into three groups – blue cloud (BC), GV, and red sequence (RS) – and we discuss the completeness of the classification. Section 5 discusses our results, outlining the methodology’s caveats and potential limitations. Finally, in Sect. 6 we summarise our main results. Throughout the paper, we assume  $H_0 = 69.6 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.286$ , and  $\Omega_\Lambda = 0.714$  (Bennett et al. 2014). Magnitudes are given in the AB system (Oke 1974) unless otherwise stated.

## 2. Methodology

This section describes the two ML algorithms applied in this paper, the library of templates adopted, and how we used them to obtain observed magnitudes and spectra that simulate WEAVE-StePS observations.

### 2.1. Machine learning methods

To retrieve the physical parameters, we made use of two ML algorithms for regression, RF (Breiman 2001) and KNN (Altman 1992), as they are two well-known algorithms that are simple to implement and yield accurate results (Stensbo-Smidt et al. 2017; Bonjean et al. 2019). These algorithms are implemented using the `scikit-learn v1.4.0` Python package (Pedregosa et al. 2011). The algorithms take as input the set of galaxy spectro-photometric observables and output the predicted physical parameters. They are first trained on a galaxy dataset with known

physical parameters. Once trained, new observables can be input to determine estimates of the physical parameters. We used a simulated dataset of galaxy spectro-photometric measurements for which we have the underlying model physical parameters. We used 90% of the dataset for training the algorithms, and the remaining 10% of the sample for testing and validation. We randomised the data selection so that both the training and testing samples are fully representative of each other. Generally, ML algorithms perform better when the input values are normally distributed and centred on 0 with a standard deviation of 1. Thus, before running the algorithms for training or testing, we pre-processed the galaxy parameters by subtracting the mean values and dividing by the standard deviation, computed from the training set. For parameters that range over many orders of magnitude, such as the sSFR, we used the base ten logarithm of the value and then standardise to mean 0 and standard deviation 1.

The first algorithm we considered, the RF, uses an ensemble of decision trees. A decision tree consists of a sequence of rules applied to determine the output value. It is constructed by iteratively splitting the training sample according to the values of the observables (features). A different feature is used to make the partition at each tree level. The process can be stopped after a certain number of splits – this stopping point is known as the maximum depth. The final nodes, known as leaf nodes, have physical parameter values assigned to them, which are returned to give the parameter estimate. The final estimate from the RF is determined by averaging the estimates from each decision tree. We tested the algorithm with a number of trees ranging from 1 to 100 (see Sect. 3). We find that a larger number of trees in the forest gave more robust results, although the performance is not sensitive to the precise number (see Appendix A for details). We thus opted for 100 trees. The trees were expanded until each leaf node contained a single galaxy.

The KNN algorithm is the second algorithm that we used for the purposes of estimating the physical parameters of our templates. The algorithm stores the training sample, then when given observables to evaluate, it finds the nearest neighbours in the multi-dimensional parameter space of the training sample. A weighted average of the estimates from each neighbour determines the returned parameter estimate. We weighted the neighbours according to the inverse of the Euclidean distance. Similarly to the RF, we tested the performance with the number of neighbours ranging from 1 to 100. The optimal value depends on the dataset’s size and distribution; in our case, we opted to use 100 neighbours for our runs.

## 2.2. Template library

We tested the ML algorithms described in the previous section using a library of spectral templates. The library is based on the Bruzual & Charlot (2003) models (2016 revised version, hereafter CB16) assuming a Chabrier initial mass function (Chabrier 2003). The CB16 models cover a wide spectral range using different stellar spectral libraries. The optical wavelength is based on the MILES stellar library (Sánchez-Blázquez et al. 2006), which covers the range  $3525 < \lambda < 7500 \text{ \AA}$  at the resolution of  $2.5 \text{ \AA}$ . The UV part of the spectra,  $\lambda \leq 3525 \text{ \AA}$ , is purely theoretical, based on Martins et al. (2005) models with a resolution of  $1 \text{ \AA}$ . Finally, the NIR part of the spectra is based on the BaSeL semi-empirical library (Westera et al. 2002) at a resolution of  $3 \text{ \AA}$ .

We built our template library assuming a simple SFH represented by an exponentially decaying SFR:

$$\text{SFR}(t) \propto e^{-t/\tau}, \quad (1)$$

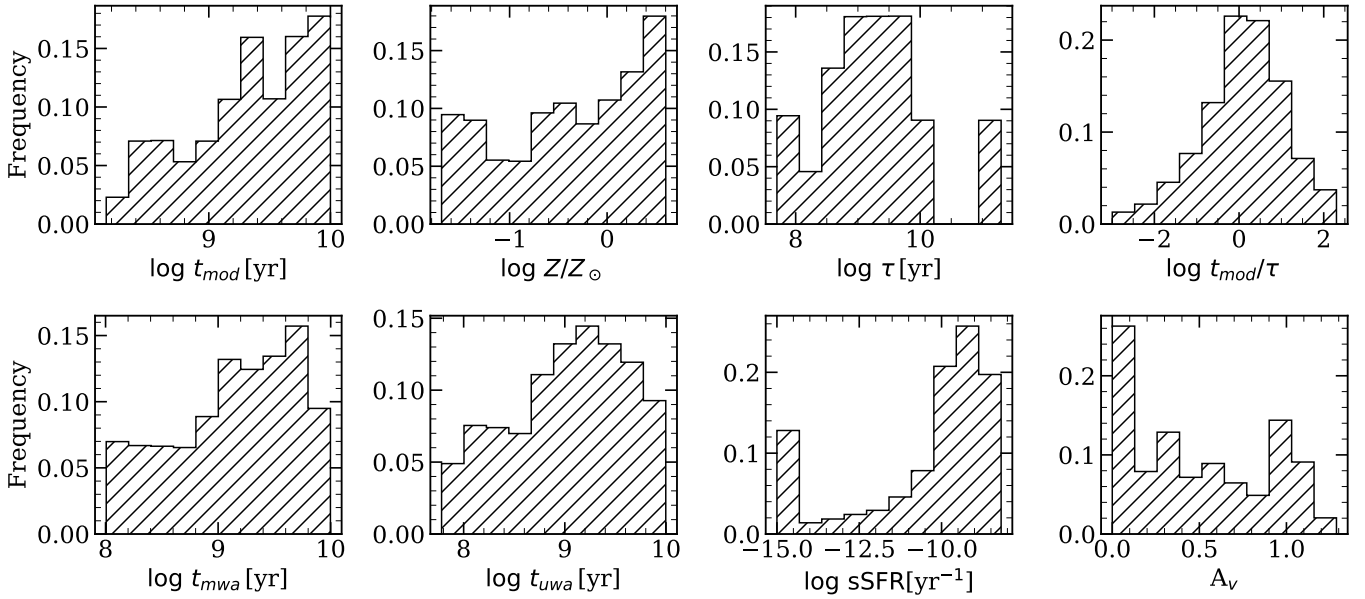
where  $t$  is the time elapsed since the onset of star formation in Gyrs, and  $\tau$  is the rate of decay of the SFR. The model ages,  $t$ , vary from 0.1 to 10.04 Gyr, where we have a total of 59 different ages<sup>1</sup>. For  $\tau$  we considered 20 different values ranging from 0.05 Gyr to 10 Gyr. Between 0.05 and 1 Gyr, the values of  $\tau$  were set with a linear step size of 0.1 Gyr to represent quenched old galaxies, while for  $\tau \geq 1$  Gyr, we used a logarithmic step of  $\sim 0.09$  dex. We also added two more sets of templates with  $\tau \sim 100$  and 200 Gyr to mimic a constant SFH. The metallicity varies from  $\log Z/Z_{\odot} = -1.69$  to  $+0.6$ . Since BC16 models are distributed at discrete values of metallicity  $\log Z/Z_{\odot} = -1.69, -0.40, +0.0, +0.40$ , we created templates covering a continuous distribution of this parameter by interpolating between the spectra on random values between those obtained on the fixed grid. We reached a maximum value of  $+0.6$  by linear extrapolation of the template spectra obtained with  $\log Z/Z_{\odot} = +0.00$  and  $+0.40$ . Finally, we adopted the Charlot & Fall (2000) dust prescription, which applies dust attenuation depending on the galaxy’s SFH. We chose the total optical depth in the V band,  $\hat{\tau}_V$ , to range from 0 to 3, where for each  $\tau$  and metallicity we have a model for  $\hat{\tau}_V = 0$  and 3. In addition to these two optical depths, we randomly generated two additional values from a uniform distribution between 0 and 3. We note that the prescription treats dust in younger stellar populations,  $\leq 10^7$  yr, differently to that in older stellar populations. Less dust affects the older population, owing to the diffuse interstellar medium, and is formulated as  $\mu \hat{\tau}_V$ . We selected  $\mu = 0.3$ . In total, we have  $\sim 105\,000$  galaxy templates, which do not contain any emission lines, as we assumed that their contribution has already been removed from the simulated spectra. We note that correcting for nebular emissions from stellar contributions is challenging and may introduce systematic errors in the estimation of certain absorption lines. As a worst-case scenario, we assumed that the systematics associated with emission line corrections are so high that the absorption features potentially affected by emission line residuals (namely Mgb, Balmer lines, and Fe5015) are no longer usable. We therefore performed a test run excluding all these indices from the spectral analysis and verified that this extreme choice induces only minimal variations in the age and metallicity estimates. The major effect is present for the age estimates, where removing all the Balmer lines (H $\beta$ , H $\gamma$ , and H $\delta$ ) in the absence of photometric information causes an increase in the uncertainty in all age estimates by roughly 0.06 dex.

Figure 1 shows the distribution of the physical parameters of our templates. We used the uwa and mwa to retrieve ages that are sensitive mostly to young stellar populations and to all stars regardless of age, respectively. We defined the sSFR as

$$\text{sSFR} = \frac{\text{SFR}}{M_{\star}^{\text{live}} + M_{\star}^{\text{rem}}}, \quad (2)$$

where SFR is the instantaneous SFR in  $M_{\odot}/\text{yr}$ , and  $M_{\star}^{\text{live}}$  and  $M_{\star}^{\text{rem}}$  are the stellar mass in  $M_{\odot}$  stored in living stars and remnants, respectively. For galaxies with  $\log(\text{sSFR}) \leq -15$ , we set it to  $-15$  as it has been found that measurements of SFRs,

<sup>1</sup> The changes in the observables, such as the spectral indices, do not scale linearly with the changes in the age of the stellar population. Therefore, the time steps,  $t_{\text{mod}}$ , are chosen ad hoc so that we have a more uniform distribution in the average ages of our templates while ensuring that we do not have a strong over-density region in the observable plane, such as the H $\beta$  vs. [MgFe]<sup>1</sup>. Note that since we tried to ensure uniformity in multiple dimensions, we do not have complete uniformity in any single given dimension.



**Fig. 1.** Histograms of the physical parameters of the templates created using the BC16 models. *Top panels:* distribution of, from left to right, model age, i.e. age since the birth of the simulated galaxy, metallicity, rate of decay of star formation ( $\tau$ ), and time since the bulk of the star formation ( $t_{\text{mod}}/\tau$ ). *Bottom panels:* histograms of mwa and uwa, sSFR, and effective dust attenuation ( $A_V$ ). See the main text for more details on the distribution of these parameters.

and thus sSFRs, are highly uncertain below a certain threshold (Brinchmann et al. 2004; Donnari et al. 2019) since there is a high degree of degeneracy between observables and sSFRs. Finally, we note that the Charlot & Fall (2000) prescription applies to dust depending on the SFH; thus, for a given  $\hat{\tau}_V$ , the  $A_V$  will differ for different  $\tau$ . This leads to the non-uniform distribution shown above.

### 2.2.1. Observables

We considered two sets of observables from each template: spectral indices and photometric magnitudes, observables that will be available for the upcoming WEAVE-StePS data. These quantities are computed at three redshifts,  $z = 0.3, 0.55$ , and  $0.7$ . Figure 2 shows a typical spectrum of an old stellar population with all of the observables available. We selected a particular set of indices available in the UV and optical, following the reasoning outlined in Costantin et al. (2019) and Table 1 of Ditrani et al. (2023). To summarise this reasoning, the optical spectral indices were selected specifically as they are sensitive to the age and metallicity of the population but not affected by other physical phenomena, such as the specific elements abundance, the chromospheric emission from the stellar atmosphere, flux calibration, or being highly sensitive to the initial mass function. In the UV part of the spectrum, all spectral indices are considered Mgwide, which is heavily affected by flux calibration. While  $D_n(4000)$  is also affected by flux calibration, it is less so compared to Mgwide as it is measured over a shorter wavelength range. However, we still accounted for uncertainties due to flux calibration (see Sect. 3.2).

The photometric data consists of the observed magnitudes covering the observed wavelength range from the UV to the NIR (see Sect. 2.2.3). The inset of Fig. 2 shows the filters that we used (GALEX FUV, GALEX NUV, CFHT  $u^*$ , HSC  $g, r, i, z, y$ , VISTA  $J, H$ , and  $K_s$ ). Observations in all these filters are available in the

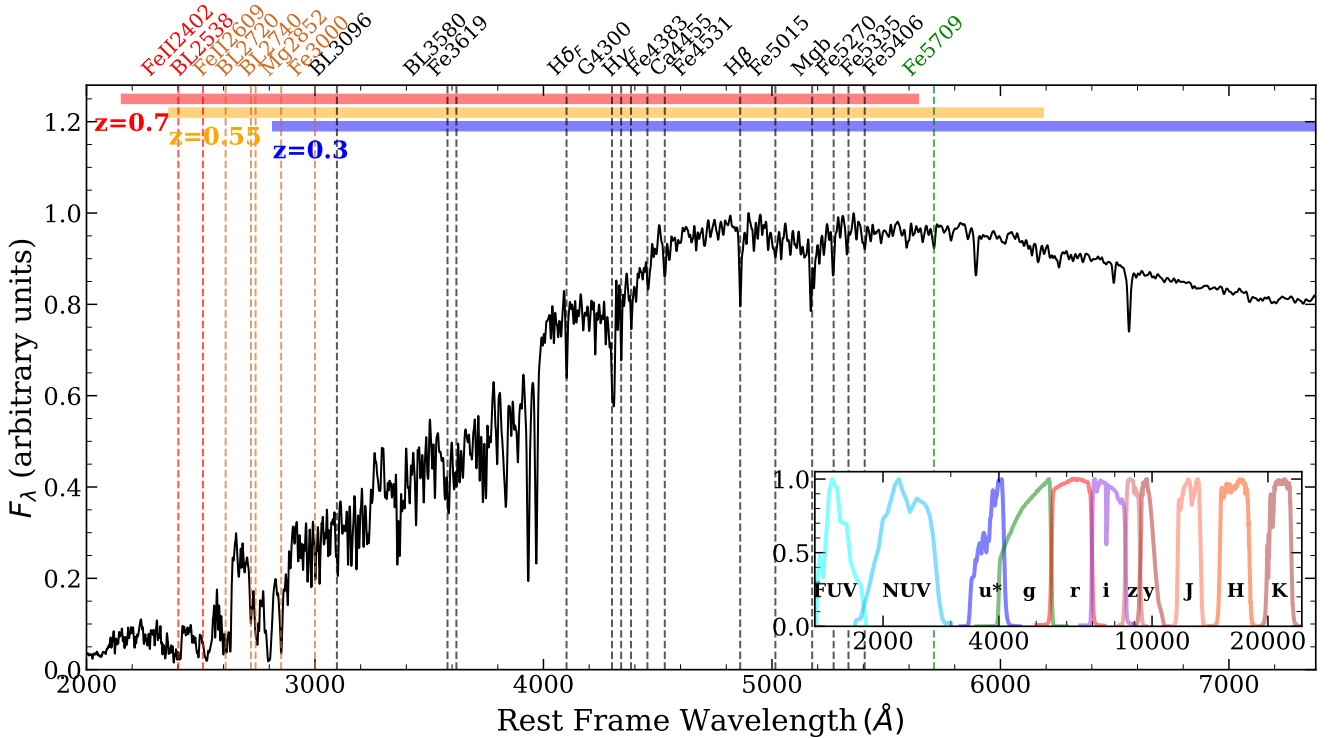
COSMOS field (Scoville et al. 2007), which will be observed by WEAVE-StePS (Iovino et al. 2023b).

### 2.2.2. Simulated WEAVE-StePS spectra

This section briefly describes how we obtained simulated WEAVE-StePS spectra at each redshift value and S/N value used in our analysis. Full details can be found in Costantin et al. (2019).

To create a set of realistic WEAVE-StePS observed spectra, we started by convolving each template spectrum (see Sect. 2.2) with a common velocity dispersion of  $\sigma_* = 200$  km/s, well suited to the massive tail of the galaxy stellar mass distribution covered by WEAVE-StePS (see below and Zahid et al. 2016). The convolved templates are then shifted to the observed wavelengths corresponding to the redshift value considered and trimmed to cover the WEAVE LR-MOS mode wavelength range, including the WEAVE CCD gaps. These ‘observed’ model spectra are converted from flux units to counts on the CCD, using the total WEAVE throughput, which includes the signal lost due to atmospheric transmission and the optics of the WHT and the WEAVE spectrograph. The expected noise due to the WEAVE CCD detectors and the Poisson noise of the sky background are added in quadrature to the Poisson noise present in the signal itself in such a way as to obtain the desired S/N in the  $I$  band. Finally, the spectra are converted back into flux units. We note that this methodology results in different S/Ns depending on the considered wavelength.

In our analysis, we considered three different signal-to-noise ratios and redshifts, chosen according to values expected within WEAVE-StePS (Iovino et al. 2023b). Most galaxies are expected to be observed at  $z = 0.3$ ,  $N_{\text{gal}} \sim 550 \pm 60 \text{ deg}^{-2}$  per  $\Delta z = 0.1$  and have  $\log M_*/M_\odot \geq 10.2$ . The distribution of galaxies observed by the WEAVE-StePS will extend to  $z \sim 0.8$ , but  $z = 0.7$  with  $N_{\text{gal}} \sim 80 \pm 20 \text{ deg}^{-2}$  is the highest redshift that still contains a statistically good sample of galaxies. As a third



**Fig. 2.** Example of a normalised spectrum for one of our templates. The vertical lines show the central wavelength in the definition of the spectral indices used here. The horizontal bar at the top shows the wavelength coverage available for the WEAVE spectrograph at  $z = 0.3$  (blue),  $0.55$  (yellow), and  $0.7$  (red). The black lines show spectral indices available for all three redshifts, while the green line indicates indices measurable only at  $z = 0.3$  and  $0.55$ . The orange and red lines show the spectral indices available at  $z = 0.55$  and  $0.7$  and only at  $z = 0.7$ , respectively. While not shown here, we also measured the  $4000 \text{ \AA}$  break. We note that while some spectral indices have their central bandpass within the WEAVE wavelength coverage at a given redshift, they still cannot be used as their blue or red bandpass falls at the very border or outside the wavelength range available to WEAVE at a given redshift (see [Ditrani et al. 2023](#), and references within for the spectral index definitions). The inset shows the filters we used to calculate the observed magnitudes.

redshift value we chose  $z = 0.55$ , a representative mid-point value of the redshift distribution, with  $N_{\text{gal}} \sim 310 \pm 40 \text{ deg}^{-2}$  per  $\Delta z = 0.1$ . At  $z = 0.55$  we will observe galaxies with stellar mass,  $\log M_{\star}/M_{\odot} \gtrsim 11.0$ , while at  $z = 0.7$ , the sample will have  $\log M_{\star}/M_{\odot} \gtrsim 11.3$ . At  $z = 0.3$ , we have  $\sim 105\,700$  simulated spectra at each S/N level. This number decreases to  $\sim 97\,000$  and  $\sim 89\,000$  at  $z = 0.55$  and  $z = 0.7$ , respectively, as we excluded all galaxies that are older than the age of the Universe.

Using the simulated WEAVE-StePS spectra, we measured the set of UV and optical spectral indices available in the spectral ranges shown in Fig. 2. To assess the impact of measurement errors, we generated 1000 Monte Carlo realisations for each spectrum according to the error model described above for a given S/N. For each realisation, we computed the spectral indices. We then computed the statistical error on the indices from the standard deviations of the 1000 realisations. We note that  $D_n(4000)$  will be sensitive to the spectro-photometric calibration; therefore, to account for such systematics, we added a 5% lower limit on its uncertainty.

### 2.2.3. Simulated photometry

From the model templates, we calculated the observed magnitudes in each filter bandpass by transforming the spectra to the corresponding redshift and convolving the model spectrum with the correct filter bandpass. For the uncertainty in these observed magnitudes, we assigned the photometric ancillary data that will be available in WEAVE-StePS to each simulated galaxy ([Iovino et al. 2023b](#)). In particular, we considered the photo-

metric catalogues covering the COSMOS field from the UV to NIR. To assign uncertainties to the simulated photometric measurements from our templates, we used the COSMOS catalogue ([Laigle et al. 2016](#)) to fit a relation between the observed magnitudes and their formal uncertainties in each band. We normalised the templates to have  $I_{\text{AB}} = 20.5$ , which is the WEAVE-StePS magnitude limit. We note that we added an error of 0.05 mag in quadrature to set a lower limit on the uncertainty to account for possible systematics.

## 3. Results

A well-known issue in retrieving the stellar population properties of galaxies is the age-metallicity degeneracy, especially for the older population ([Worthey 1994](#)), which is further complicated by the presence of dust. This is largely due to the fact that different combinations of age, metallicity, and dust produce similar spectral shapes, making it impossible to disentangle the components unless we have a sufficient S/N and a large enough wavelength range. Therefore, in this section, we first examine how well ML algorithms retrieve the physical parameters on perfect observables without measurement errors. We then analyse how well the ML algorithms can predict the same physical parameters for our simulated galaxies containing realistic noise. Regarding the spectral indices, we performed this test for S/N levels of 10, 20, and 30. In each case, the noise in the photometric data was fixed to that of the WEAVE-StePS COSMOS ancillary catalogue. For each S/N level, we also tested the retrieval capability

of the ML algorithms at three different redshifts,  $z = 0.3, 0.55,$  and  $0.7,$  respectively.

### 3.1. Simulated data with no errors

For the initial test case, we used the observables without noise. We quantified the success of the ML algorithms using either the bias and the Spearman correlation coefficient (scc) when considering noiseless data or the bias, dispersion ( $\sigma$ ), and the fraction of outliers when dealing with noisy data (see the next section)<sup>2</sup>. The bias is estimated as  $\bar{x} = \text{median}(\Delta x_i)$ , where  $\Delta x_i = x_i^{\text{pred}} - x_i^{\text{true}}$ . For the dispersion, we used the median absolute deviation (MAD) estimate and it is calculated using the formula

$$\sigma = 1.4826 \times \text{MAD}(\Delta x_i), \quad (3)$$

where  $\text{MAD}(\Delta x_i) = \text{median}(|x_i - \text{median}(x_i)|)$ . In the following, we present the dispersion computed in moving average bins, where the bin size is larger than the bin step. The bin and step sizes change depending on the physical parameter, as the range of minimum and maximum values changes. Finally, we considered a data point an outlier when  $\Delta x_i > 2\sigma$ .

Figure 3 shows the comparison between predicted values and true values for the mwa (top left panel) and uwa (top middle panel), metallicity (top right panel), sSFR (bottom left panel),  $A_V$  (bottom middle panel), and  $\log(t_{\text{mod}}/\tau)$  (bottom right panel) assuming no noise on the observables. The values quoted are the bias and Spearman's correlation coefficient (scc). We note that the quoted bias value is the median bias across all bins. The uncertainty values represent the 25<sup>th</sup> and 75<sup>th</sup> percentiles of the distribution, showing the scatter in bias in different bins. We do not present the dispersion values since they are negligible in this case of perfect data. Hence, we also excluded the outlier fractions. This low scatter is demonstrated by the high scc values calculated for the physical parameters, as discussed below. There is a good prediction for both the mwa and uwa, metallicity and  $A_V$ , as the bias is low, with both the median bias and the percentiles close to 0. In addition, the scc is high  $\sim 0.99$ , indicating a strong correlation between true and predicted values with little scatter. Looking at the difference between predicted and true (bottom sub-panels), we find similar results, but it is evident that RF consistently outperforms KNN. Our retrieval capability of sSFRs is robust at high specific star formation, as the bias values are  $\sim -0.002$  and  $-0.01$ , but at very low specific star formation,  $\log \text{sSFR} \lesssim -12$  and  $-11$ , there is a clear bias for RF and KNN, respectively, with KNN having a stronger bias. This bias may be because there is limited variance in the observed parameters of templates at low sSFRs, thus creating a high degeneracy even for perfect data. Therefore, when averaging over 100 trees/neighbours, the high level of degeneracy in observed parameters for low sSFRs will result in the selection of numerous galaxies with lower sSFRs, thus underestimating the sSFR.

Finally, we have  $\log t_{\text{mod}}/\tau$ , which can be considered a second-order physical parameter that indicates the number of e-folds since the peak of star formation. We find that the retrieval of this parameter is biased due to degeneracies that exist at low and high values. At very low values,  $\log t_{\text{mod}}/\tau < -1$ , the galaxies are actively forming stars and are either very young, with

$t_{\text{mod}} \sim 0$  Gyr, or have a long decay time,  $\tau$ . For these effectively young stellar populations, the observables are relatively insensitive to the precise value of the scale time  $\log t_{\text{mod}}/\tau$ . At the low end, the bias is  $+0.25$  for RF and  $+0.6$  for KNN. At high values,  $\log t_{\text{mod}}/\tau > 1$ , star formation has ended, and the stellar populations are evolving passively. In this regime, the observables are not strongly sensitive to the scale time, and the bias is  $-0.06$  for RF and  $-0.2$  for KNN.

### 3.2. Simulated data with noise

In the following subsections, we explore how the parameter estimation performs on realistic data. Here, we introduce various levels of noise on the spectral indices ( $S/N_{I,\text{obs}} = 10, 20,$  and  $30$ ) following the steps described in Sect. 2.2.2. We note that we do not change the errors in the photometry since it will come from archival data of the COSMOS field with known photometric errors. In addition, we also tested the performance of the algorithms to predict physical parameters at three different redshifts,  $z = 0.3, 0.55,$  and  $0.7$ . At each redshift we set the maximum age of each template to be the age of the Universe ( $t_{\text{mod}} = 10.27$  Gyr at  $z = 0.3$ ,  $t_{\text{mod}} = 8.27$  Gyr at  $z = 0.55$  and  $t_{\text{mod}} = 7.34$  Gyr at  $z = 0.7$ ). Finally, while the metallicity of the templates covers the range  $-1.69 \lesssim \log Z/Z_{\odot} \lesssim +0.6$ , we restricted it for the testing sample to  $-1.39 \lesssim \log Z/Z_{\odot} \lesssim +0.4$  to avoid significant bias at the lowest and highest metallicity values.

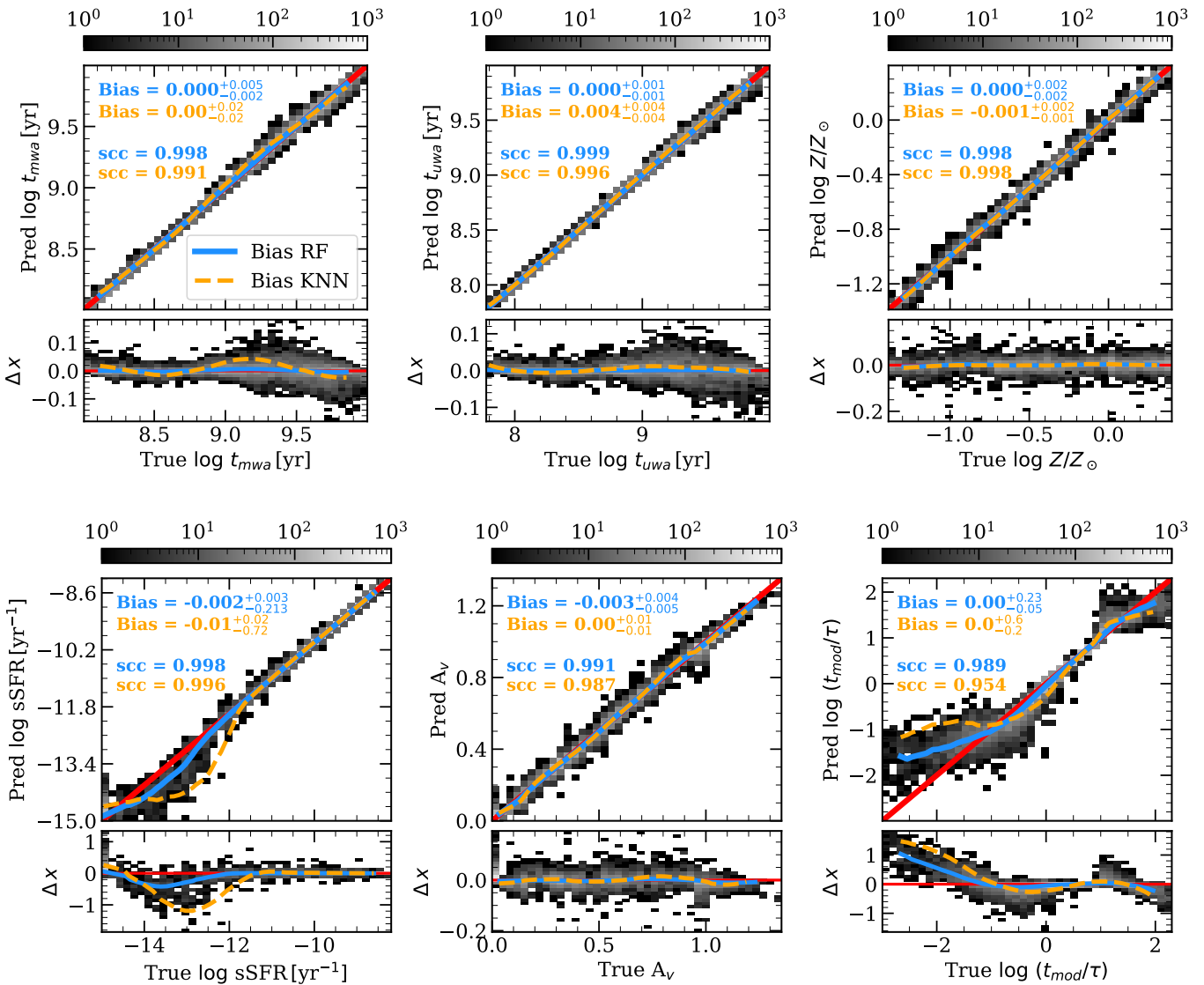
#### 3.2.1. Results for $z = 0.3$ and $S/N = 10$

This sample will define the baseline performance. The ML algorithms are trained on the simulated sample with noise. As neither KNN nor RF directly take the uncertainty as inputs, we perturbed the simulated spectrum (with noise) five times, in a Gaussian manner, and re-measured the spectral indices. Similarly, we also perturbed the photometric data five times in accordance with the noise. This procedure grows the training sample five-fold; the testing sample size was not modified, but we also perturbed the testing sample once according to the noise.

Figure 4 shows the results obtained for the mwa, uwa, metallicity, sSFR,  $A_V$ , and  $\log t_{\text{mod}}/\tau$  using the RF algorithm.

The retrieval capability worsens compared to when there were no errors on the spectro-photometric observables; however, most of the physical parameters have neither a large bias nor scatter. For the mwa, the bias reaches 0.1 dex at most. For the uwa, we find a similar trend; however, in general, both the bias and  $\sigma$  are smaller. The metallicity shows a negligible bias  $< 0.1$  dex. This is particularly true for high metallicity,  $\log Z/Z_{\odot} \gtrsim +0.0$ , where the bias and the dispersion tend to be the lowest. The  $A_V$  also has a small bias, straddling the 0 value, except for  $A_V \gtrsim 1.0$ , where there is a general underestimation for the effective dust attenuation, by  $\sim 0.1$  mag. The sSFR also has low bias and  $\sigma$  for  $\log \text{sSFR} \gtrsim -11.0$ . However, for simulated galaxies considered quiescent, at  $\log \text{sSFR} \lesssim -11.0$ , we find a strong bias and large  $\sigma$ . For  $\log t_{\text{mod}}/\tau$  we generally find high bias and dispersion. Although we observe an increase in the bias for most physical parameters, it is not significantly different from the test case where the observables have no error, as in most instances, the bias is still close to zero. Finally, we observe no clear difference in the relationship between the retrieval capability of the physical parameters and the age group of galaxies, where we classified the galaxies into three groups based on their stellar populations: young ( $t_{\text{mwa}} \leq 1$  Gyr), intermediate ( $1 \leq t_{\text{mwa}} \leq 4$  Gyr), and old ( $t_{\text{mwa}} \geq 4$  Gyr). We note that the main exception to this trend is the metallicity, where the older

<sup>2</sup> We have two sets of statistics, one for perfect simulated data and one for simulated data with noise. We do not quantify  $\sigma$  or fraction of outliers with perfect simulated data as this leads to non-physical interpretations of results.



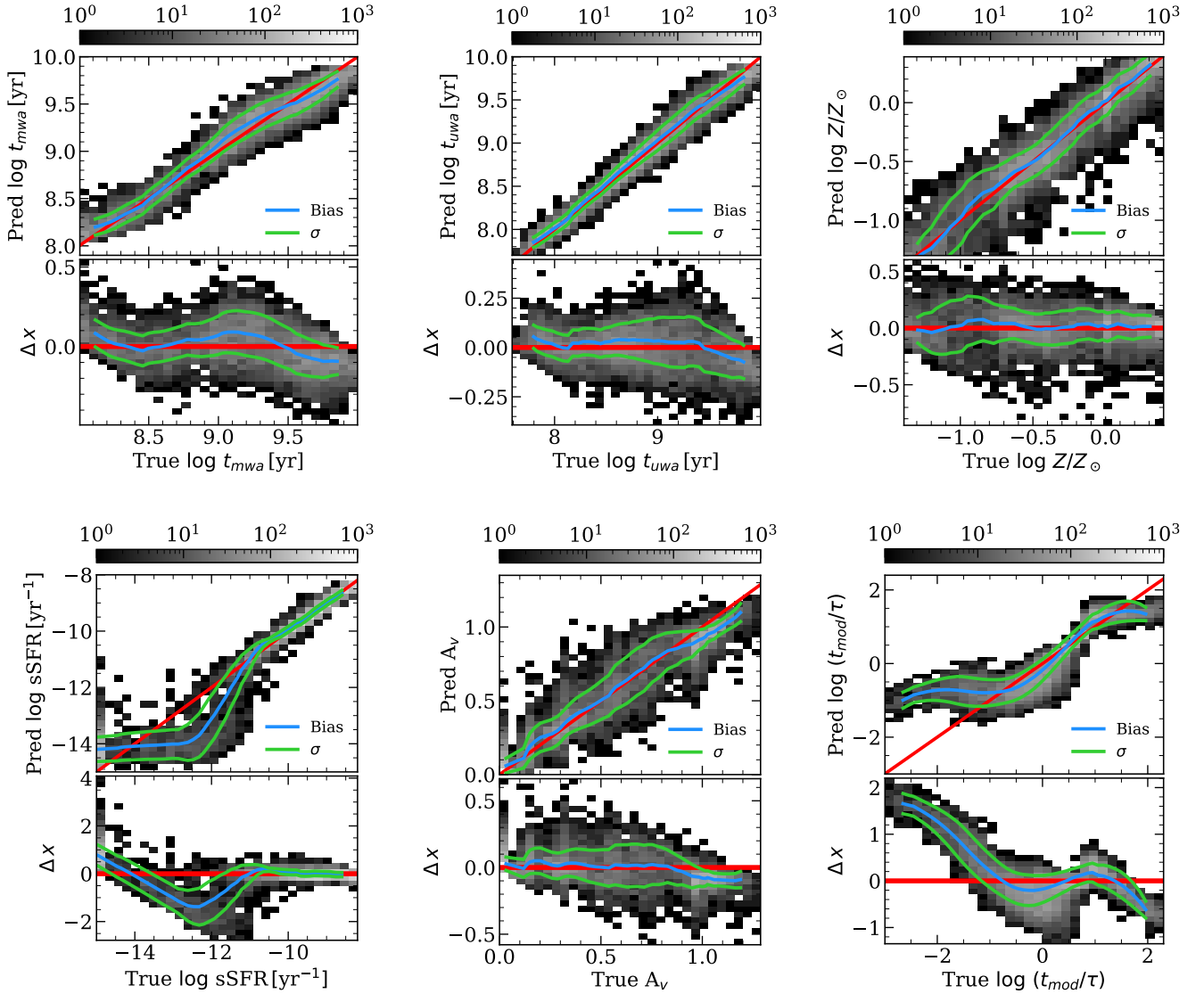
**Fig. 3.** Retrieval of the physical parameters assuming no noise on data. *Top panels:* density plots showing predicted vs. true values for the mwa (left), uwa (middle), and metallicity (right) derived using RF. The blue lines show the bias computed as a running average with fixed bin step and size. The bin size is six times larger than the bin step, and therefore the same galaxies are present in multiple bins; this ensures we have a sufficient number of galaxies in each bin. The blue labels refer to the median bias and Spearman’s correlation coefficient (label scc in the plot) for RF. For comparison purposes, we have included a line showing the bias for KNN and the values of the median bias and the scc in orange. The errors quoted with bias are the 25<sup>th</sup> and 75<sup>th</sup> percentile. The smaller, lower panels show the difference between predicted and true values. The red line always shows the one-to-one line or the 0 value for the variable considered. The colour bar at the top indicates the number of galaxies within a pixel in the density plot. *Bottom panels:* same results but for the sSFR (left), dust attenuation,  $A_v$  (middle), and  $t_{\text{mod}}/\tau$  (right).

stellar population exhibits smaller  $\sigma$  values for low metallicities in comparison to the younger and intermediate stellar populations. This trend was also observed in a previous study by Ditrani et al. (2023) using simulated galaxies and also in real data (see Fig. 7 of Gallazzi et al. 2005).

We note that the results discussed in this section use the mean of the 100 trees or of the 100 nearest neighbours – as given by the output of the scikit-learn python package. While it is possible to extract the median using additional functions for RF, for KNN, this requires modifying the source code itself, which is beyond the scope of this paper. Therefore, we continued to use the mean estimates of 100 trees/neighbours for a fairer comparison between the two algorithms. Nevertheless, we tested how much the results vary when we use the mean and median of 100 trees for RF. Using the median gives results comparable to the

mean except at extremes, the minimum and maximum bins along their true value, where the bias is slightly smaller. For example, there is a difference of  $\leq 0.03$  dex for the mwa and uwa, and  $\leq 0.09$  mag for  $A_v$  when using the median rather than mean at the boundaries. For metallicity, the difference between bias and  $\sigma$  is negligible. Finally, for sSFRs, while the degradation of the sSFR estimates begins at  $\log \text{sSFR} \leq -11.2$ , when using the median, the bias values are larger, as expected, since we imposed a lower limit of  $-15.0$ .

Table 1 presents the median bias,  $\sigma$  and the fraction of outliers for RF and KNN. The uncertainties given are the 25<sup>th</sup> and 75<sup>th</sup> percentiles, showing the variation in both bias and  $\sigma$  for different bins. While the median bias is comparable to that obtained in the no-noise case,  $\sim 0.00$ , the bias variations are generally larger as seen by the quartile ranges, reflecting what is



**Fig. 4.** Physical parameters derived using RF, assuming  $S/N = 10$  at  $z = 0.3$ . *Top:* results for the mwa (*left*), uwa (*centre*), and metallicity (*right*). *Bottom:* sSFR,  $A_V$ , and  $t_{\text{mod}}/\tau$ , from *left to right*. Each sub-panel on the bottom shows the difference between predicted and true values ( $\Delta x$ ). The blue lines show the bias estimated using moving averages. Similarly, the lime-green lines show the  $\sigma$ . The red line shows the one-to-one line (top sub-panels) or 0 value (bottom sub-panels). Finally, the colour bar shows the number of galaxies in the density plot.

seen in Fig. 4. For RF (KNN), both the mwa and uwa have a median  $\sigma \leq 0.10$  (0.16) dex, and metallicity has a median  $\sigma$  of  $\sim 0.14$  (0.22) dex. The median bias for the sSFR is relatively low  $-0.03$  ( $-0.1$ ) dex; however, we see that the scatter on the dispersion for different bins is much larger, with 25<sup>th</sup> and 75<sup>th</sup> percentile values of  $-0.80$  ( $-0.70$ ) dex and  $0.05$  (0.20) dex, respectively. The  $A_V$  has a lower median bias for RF, whereas for KNN, we see a typical bias of 0.1 mag. Similarly, KNN has a larger scatter than RF by  $\sim 0.04$  mag. Finally,  $\log t_{\text{mod}}/\tau$  is found to be the worst constrained parameter as it is the one with the highest bias,  $\sim -0.1$  ( $-0.2$ ) dex, and  $\sigma$ ,  $\sim 0.28$  (0.27) dex. Overall, RF consistently performs better than KNN, as seen in previous literature in other contexts (Thanh Noi & Kappas 2018). We note that the main exception is in the fraction of outliers, where RF always has a slightly larger outlier fraction. The fraction of outliers is influenced by two factors: how compact the distribution of  $\Delta x$  is in a given bin and how Gaussian it is. For RF, we typically find a strongly peaked distribution with asymmetric tails. This leads to a low  $\sigma$  but a high fraction of outliers. On the other hand, KNN

has a broader peak, resulting in higher  $\sigma$ , and reduced tails. In this case, the  $\Delta x$  distribution is closer to Gaussian, giving a lower fraction of outliers despite having a larger  $\sigma$ .

### 3.2.2. Higher S/N at $z = 0.3$

In this section, we analyse how the previous results improve when spectral data are affected by lower noise values. For this test, we considered  $S/N = 20$  and  $30$  and show  $S/N = 10$  for comparison. Figure 5 shows the bias and  $\sigma$ , obtained using RF, as a function of the true physical parameter values. The black, blue and red lines show the values for  $S/N = 10$ , 20, and 30, respectively. Overall, we find that improving the  $S/N$  of the spectra does not significantly improve the bias for most of the physical parameters. Regardless of the physical parameter or the true value of the parameter we are considering, the difference in bias between  $S/N$ s is  $\leq 0.05$  dex for the age, metallicity, and sSFR and  $\leq 0.05$  mag for  $A_V$ . The only exception to this is for  $\log t_{\text{mod}}/\tau$ , where for the lowest value,  $\log t_{\text{mod}}/\tau \lesssim -1.2$ , we see that the

**Table 1.** Average statistics in all bins at different combinations of redshifts and S/Ns.

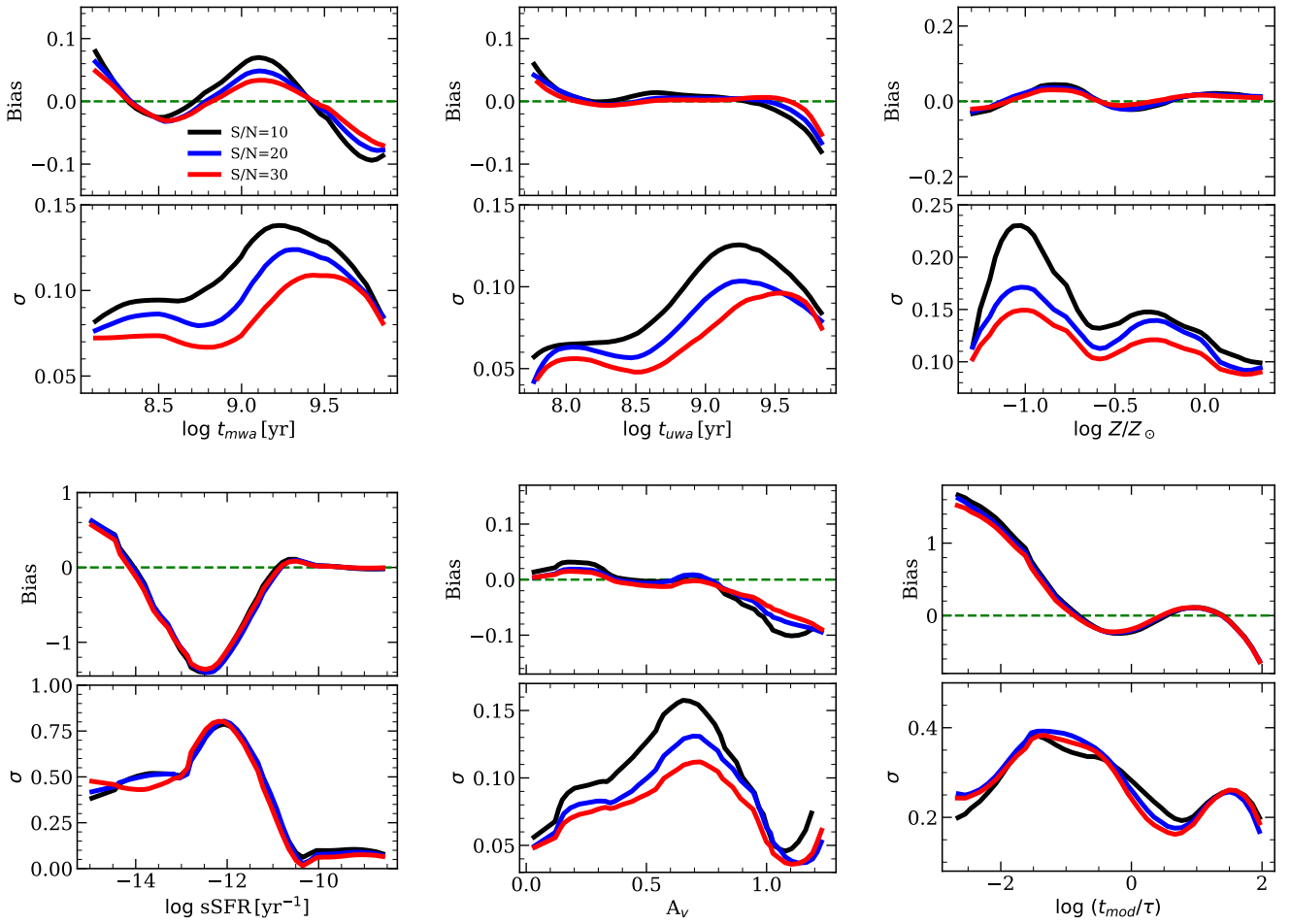
Physical Parameter	Random forest			K-Nearest neighbour		
	Bias	$\sigma$	$f_{\text{outlier}} (\%)$	Bias	$\sigma$	$f_{\text{outlier}} (\%)$
$z = 0.3, S/N = 10$						
$\log t_{\text{mwa}} [\text{yr}]$	$0.00^{+0.05}_{-0.03}$	$0.10^{+0.02}_{-0.01}$	4.94	$-0.01^{+0.08}_{-0.07}$	$0.16^{+0.03}_{-0.04}$	3.15
$\log t_{\text{uwa}} [\text{yr}]$	$0.01^{+0.01}_{-0.01}$	$0.08^{+0.02}_{-0.02}$	5.38	$-0.03^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.01}$	4.74
$\log Z/Z_{\odot}$	$0.01^{+0.01}_{-0.03}$	$0.14^{+0.02}_{-0.02}$	7.37	$0.00^{+0.05}_{-0.01}$	$0.20^{+0.01}_{-0.02}$	5.87
$\log \text{sSFR} [\text{yr}^{-1}]$	$-0.03^{+0.05}_{-0.80}$	$0.4^{+0.1}_{-0.3}$	9.28	$-0.1^{+0.2}_{-0.7}$	$0.45^{+0.09}_{-0.26}$	7.90
$A_V$	$0.00^{+0.01}_{-0.08}$	$0.09^{+0.04}_{-0.02}$	11.80	$-0.1^{+0.1}_{-0.2}$	$0.13^{+0.02}_{-0.02}$	5.06
$\log t_{\text{mod}}/\tau$	$0.1^{+0.7}_{-0.2}$	$0.27^{+0.05}_{-0.06}$	7.40	$0.1^{+0.9}_{-0.3}$	$0.23^{+0.07}_{-0.03}$	5.08
$z = 0.3, S/N = 20$						
$\log t_{\text{mwa}} [\text{yr}]$	$0.01^{+0.03}_{-0.04}$	$0.09^{+0.03}_{-0.01}$	5.63	$-0.01^{+0.07}_{-0.07}$	$0.14^{+0.01}_{-0.04}$	3.16
$\log t_{\text{uwa}} [\text{yr}]$	$0.00^{+0.01}_{-0.01}$	$0.07^{+0.03}_{-0.01}$	6.63	$-0.02^{+0.02}_{-0.01}$	$0.09^{+0.02}_{-0.01}$	4.88
$\log Z/Z_{\odot}$	$0.01^{+0.01}_{-0.02}$	$0.13^{+0.02}_{-0.02}$	7.51	$0.02^{+0.02}_{-0.02}$	$0.16^{+0.02}_{-0.02}$	6.14
$\log \text{sSFR} [\text{yr}^{-1}]$	$-0.03^{+0.04}_{-0.86}$	$0.4^{+0.1}_{-0.4}$	9.95	$-0.1^{+0.1}_{-0.9}$	$0.42^{+0.05}_{-0.28}$	9.15
$A_V$	$0.00^{+0.01}_{-0.07}$	$0.08^{+0.03}_{-0.02}$	11.63	$-0.01^{+0.03}_{-0.12}$	$0.11^{+0.02}_{-0.02}$	6.49
$\log t_{\text{mod}}/\tau$	$0.0^{+0.7}_{-0.2}$	$0.27^{+0.06}_{-0.06}$	7.79	$0.1^{+0.9}_{-0.3}$	$0.23^{+0.05}_{-0.04}$	5.50
$z = 0.3, S/N = 30$						
$\log t_{\text{mwa}} [\text{yr}]$	$0.00^{+0.03}_{-0.03}$	$0.08^{+0.02}_{-0.01}$	6.64	$-0.01^{+0.06}_{-0.06}$	$0.11^{+0.02}_{-0.02}$	3.19
$\log t_{\text{uwa}} [\text{yr}]$	$0.00^{+0.01}_{-0.01}$	$0.06^{+0.02}_{-0.01}$	6.98	$-0.01^{+0.01}_{-0.01}$	$0.08^{+0.02}_{-0.01}$	4.37
$\log Z/Z_{\odot}$	$0.01^{+0.01}_{-0.01}$	$0.11^{+0.02}_{-0.01}$	6.11	$0.01^{+0.02}_{-0.01}$	$0.12^{+0.02}_{-0.01}$	5.82
$\log \text{sSFR} [\text{yr}^{-1}]$	$-0.03^{+0.03}_{-0.83}$	$0.4^{+0.1}_{-0.3}$	10.51	$-0.1^{+0.1}_{-0.9}$	$0.4^{+0.1}_{-0.3}$	9.30
$A_V$	$-0.01^{+0.01}_{-0.05}$	$0.08^{+0.02}_{-0.02}$	11.32	$-0.02^{+0.02}_{-0.09}$	$0.10^{+0.02}_{-0.02}$	6.77
$\log t_{\text{mod}}/\tau$	$0.0^{+0.6}_{-0.2}$	$0.26^{+0.06}_{-0.06}$	8.18	$0.1^{+0.9}_{-0.3}$	$0.22^{+0.04}_{-0.04}$	5.46
$z = 0.55, S/N = 10$						
$\log t_{\text{mwa}} [\text{yr}]$	$0.00^{+0.03}_{-0.03}$	$0.10^{+0.01}_{-0.01}$	5.67	$0.01^{+0.06}_{-0.08}$	$0.15^{+0.03}_{-0.04}$	3.63
$\log t_{\text{uwa}} [\text{yr}]$	$0.00^{+0.00}_{-0.01}$	$0.08^{+0.03}_{-0.01}$	5.72	$-0.03^{+0.04}_{-0.02}$	$0.11^{+0.03}_{-0.01}$	4.97
$\log Z/Z_{\odot}$	$0.01^{+0.02}_{-0.02}$	$0.17^{+0.03}_{-0.03}$	8.20	$0.02^{+0.08}_{-0.02}$	$0.24^{+0.01}_{-0.02}$	4.73
$\log \text{sSFR} [\text{yr}^{-1}]$	$0.00^{+0.02}_{-0.79}$	$0.3^{+0.2}_{-0.2}$	9.86	$0.0^{+0.2}_{-0.9}$	$0.5^{+0.1}_{-0.3}$	8.17
$A_V$	$0.01^{+0.01}_{-0.07}$	$0.08^{+0.03}_{-0.02}$	12.18	$-0.01^{+0.09}_{-0.17}$	$0.13^{+0.01}_{-0.01}$	4.74
$\log t_{\text{mod}}/\tau$	$0.0^{+0.7}_{-0.2}$	$0.25^{+0.07}_{-0.05}$	8.73	$0.0^{+0.9}_{-0.3}$	$0.22^{+0.07}_{-0.02}$	5.40
$z = 0.7, S/N = 10$						
$\log t_{\text{mwa}} [\text{yr}]$	$0.00^{+0.03}_{-0.02}$	$0.09^{+0.01}_{-0.01}$	5.93	$0.00^{+0.06}_{-0.10}$	$0.13^{+0.02}_{-0.04}$	4.01
$\log t_{\text{uwa}} [\text{yr}]$	$0.00^{+0.01}_{-0.01}$	$0.07^{+0.02}_{-0.00}$	6.02	$-0.03^{+0.05}_{-0.02}$	$0.10^{+0.03}_{-0.02}$	4.39
$\log Z/Z_{\odot}$	$0.02^{+0.01}_{-0.02}$	$0.16^{+0.03}_{-0.02}$	8.32	$0.04^{+0.09}_{-0.07}$	$0.24^{+0.00}_{-0.01}$	5.00
$\log \text{sSFR} [\text{yr}^{-1}]$	$0.00^{+0.02}_{-0.61}$	$0.3^{+0.3}_{-0.2}$	9.39	$0.0^{+0.1}_{-0.9}$	$0.5^{+0.1}_{-0.4}$	9.25
$A_V$	$0.01^{+0.01}_{-0.07}$	$0.07^{+0.04}_{-0.02}$	13.24	$-0.01^{+0.08}_{-0.16}$	$0.13^{+0.01}_{-0.02}$	6.11
$\log t_{\text{mod}}/\tau$	$0.0^{+0.7}_{-0.2}$	$0.29^{+0.04}_{-0.08}$	9.01	$0.0^{+0.9}_{-0.3}$	$0.22^{+0.06}_{-0.04}$	5.49

**Notes.** The left and right set of columns show results for RF and KNN, respectively. The first three sets of rows show results at  $z = 0.3$  for  $S/N = 10$ ,  $S/N = 20$ , and  $S/N = 30$ . The bottom two sets of rows show results for  $S/N = 10$  at  $z = 0.55$  and  $z = 0.7$ . The leftmost columns indicate the physical parameters of interest for each set of rows. We estimate the bias, the  $\sigma$ , and the percentage of outliers for each algorithm and physical parameter. The uncertainty quoted is the 25<sup>th</sup> and 75<sup>th</sup> percentiles.

bias (slightly) decreases at higher S/Ns. Compared to the bias, increasing the S/N decreases  $\sigma$  for most of the physical parameters. For the mwa and uwa, the largest difference in  $\sigma$  is seen for the simulated galaxies in the range  $8.5 \leq \log t \leq 9.5$  yr. We see a decrease in  $\sigma$  for metallicity as we increase the S/N; this is significant at low metallicity  $\log Z/Z_{\odot} \lesssim -0.5$ . Additionally, there is a gradual decrease in the difference between  $\sigma$  of different S/Ns as we increase in metallicity.  $A_V$  shows a gradual improvement in constraint capability with respect to the S/N, as we see a decrease in  $\sigma$  with an increase in the S/N in the range  $0.15 \leq A_V \leq 0.8$ . For sSFR, improvement in the S/N results in comparable  $\sigma$ . Finally, we also see no strong trend between

the S/N and  $\sigma$  for  $\log t_{\text{mod}}/\tau$ . Overall, while there is no clear improvement in bias, we find a clear decrease in  $\sigma$  for all physical parameters except the  $t_{\text{mod}}/\tau$  and sSFR at higher S/Ns.

Table 1 also lists the median bias,  $\sigma$ , and the fraction of outliers for both RF and KNN for  $S/N = 20$  and  $30$  at  $z = 0.3$ . The uncertainties give the 25<sup>th</sup> and 75<sup>th</sup> percentiles of the bias and  $\sigma$ . Similar to the figure, we find minimal change in the median bias for all of the physical parameters,  $\sim 0.0$  dex or mag, irrespective of the S/N. We note that while the sSFR and  $\log t_{\text{mod}}/\tau$  have median bias  $\sim 0.0$  dex, they have very high scatter on the bias at different bins, as shown by the percentiles. For example, for the sSFR we find a bias of  $\sim -0.86$  dex at different S/Ns



**Fig. 5.** Change in bias (sub-top panels) and  $\sigma$  (sub-bottom panels) as a function of the true physical parameter value for RF at  $z = 0.3$ . *Top panels:* results of, from left to right, the mwa, uwa, and metallicity for S/N = 10 (black), 20 (blue), and 30 (red). Note that while the S/Ns for the spectral indices change, the uncertainties assumed for the photometric data remain unchanged for the different cases. For the bias plot, the dashed green line indicates the zero point. *Bottom panels:* results for the sSFR,  $A_V$ , and  $t_{\text{mod}}/\tau$ .

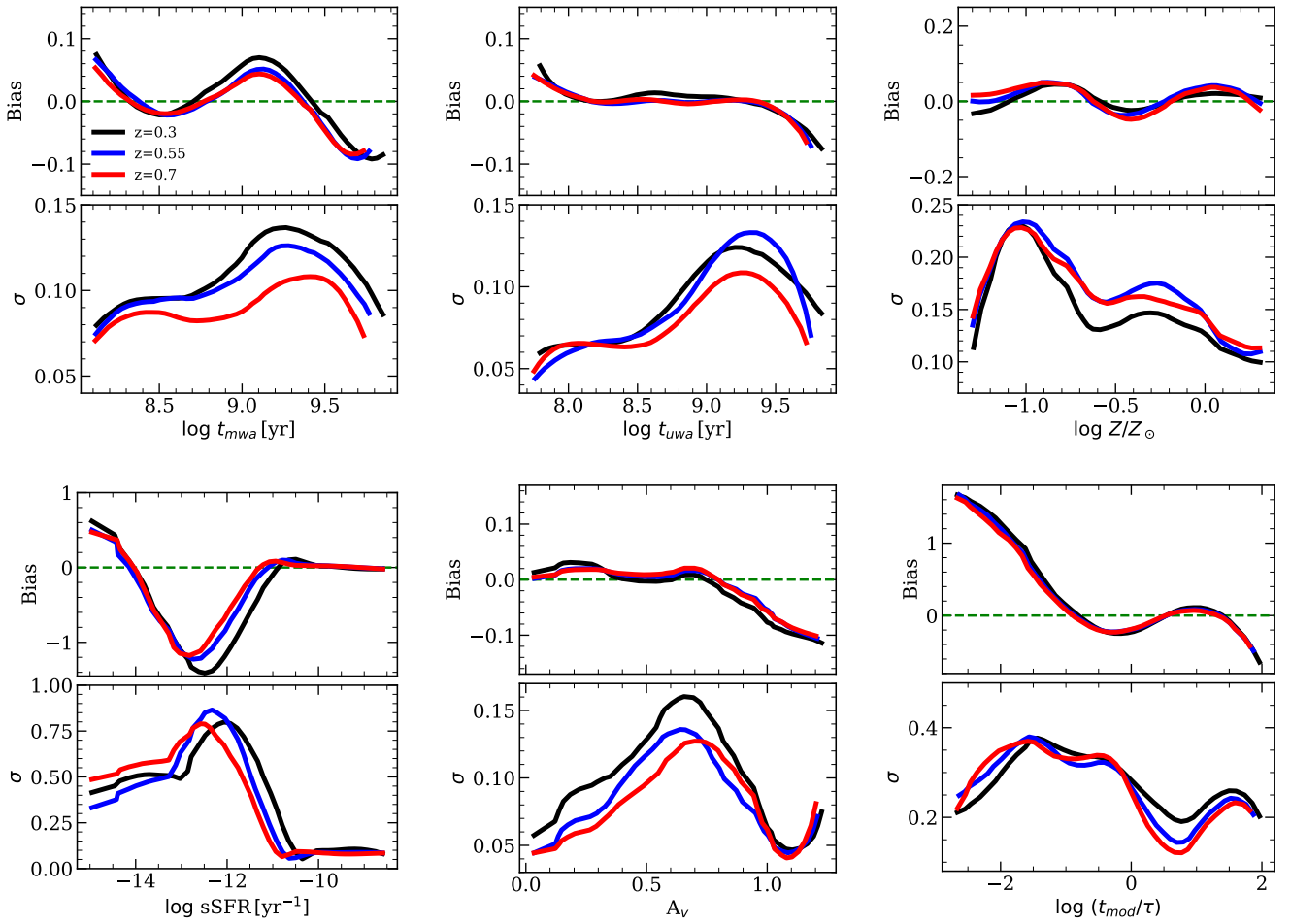
as shown by the 25<sup>th</sup> percentile value. Similarly, for  $t_{\text{mod}}/\tau$  we find the 25<sup>th</sup> and 75<sup>th</sup> percentile values for bias to be  $\sim -0.2$  and 0.7 dex, respectively. The  $\sigma$  shows a gradual decrease with increasing S/N for most parameters, by  $\sim 0.02$  dex for age and  $\sim 0.01$  mag for  $A_V$ ,  $\sim 0.03$  dex for metallicity. For the sSFR and  $\log t_{\text{mod}}/\tau$ , while we see differences in the curves, we do not see a strong trend when using average statistics. Generally, RF has (marginally) lower bias than KNN, by  $\sim 0.02$  dex. Similarly, RF has a lower dispersion than KNN for most physical parameters, varying between  $\sim 0.02$  to 0.6 dex. Again, the main exception to such a trend is seen for  $\log t_{\text{mod}}/\tau$ , where KNN has a lower dispersion by  $\sim 0.04$  dex. Finally, KNN generally has fewer outliers than RF owing to the Gaussian-like distribution of the differences between predicted and true values.

### 3.2.3. Samples at different redshifts

In the following section, we quantify differences in retrieving galaxy physical parameters at redshift higher than 0.3 for S/N = 10. Increasing the galaxy redshift affects both the spectroscopic and the photometric information. Indeed, the higher the redshift, the bluer the observed rest-frame spectral window, thus including a different set of indices. In addition to having access to more UV spectral indices, the same spectral indices

at different redshifts will possess different S/Ns because they will fall in different observed wavelength regions of the spectrograph. Furthermore, the same photometric filters will sample bluer rest-frame spectral windows. This results in the same simulated galaxy having different magnitudes and uncertainties at different redshifts.

Figure 6 shows the bias and  $\sigma$  as a function of the true physical parameter for  $z = 0.3, 0.55$  and  $0.7$  in black, blue and red lines, respectively, at S/N = 10. These results are for the RF algorithm. We find no significant differences in the bias for the mwa and uwa, metallicity, or  $A_V$ . We note that the maximum age of the simulated galaxies is lower at higher redshifts due to the younger age of the Universe. There is an improvement in the prediction of the sSFR with increasing redshift, as there is a notable reduction in the bias, where the underestimation of the sSFR starts at  $-11.5$ , at  $z = 0.7$ , rather than at  $-11.0$ . This is due to two factors: at higher redshifts we have a lower number of quiescent galaxies. Therefore, when averaging over 100 trees, we sample over a lower number of galaxies with  $\text{sSFR} = -15.0$ , thus reducing the bias in the estimate of the sSFR. Furthermore, and more importantly, at higher redshifts we have access to a larger number of UV spectral indices, and the observed magnitudes sample a bluer region of the spectra, which are more sensitive to the recent SFH and provide a better constraint



**Fig. 6.** Same as Fig. 5 but for S/N = 10 at  $z = 0.3$  (black), 0.55 (blue), and 0.7 (red). Again, these results are obtained with RF.

on the sSFR (Vazdekis et al. 2016; López Fernández et al. 2016; Salvador-Rusiñol et al. 2020). For  $t_{\text{mod}}/\tau$ , we see only minor improvements in bias as the redshift increases. In contrast to bias, there is a clear improvement in  $\sigma$  for both mwa and uwa estimations with an increase in redshift. This effect is most evident for older ages,  $\log t_{\text{mwa}} \text{ (yr)} \gtrsim 9.0$  and  $\log t_{\text{uwa}} \text{ (yr)} \gtrsim 8.5$ . The better estimation of age is also visible in Costantin et al. (2019), which uses a more traditional Bayesian approach. This trend is likely due to the larger number of UV spectral indices and the bluer rest frame magnitudes available at higher redshifts. In contrast, the  $\sigma$  for metallicity are comparable at different redshifts. This could be due to higher redshift observations probing less of the IR part of the spectrum, which, combined with optical observed magnitudes, is known to help constrain metallicity. For  $0.15 \lesssim A_V \lesssim 0.8$  mag, we find a larger  $\sigma$  at  $z = 0.3$  than at higher redshifts. We note that at  $z = 0.55$  and  $z = 0.7$ , the  $\sigma$  values are comparable. At higher redshift, the observed magnitudes sample the bluer part of the spectrum, which is known to be more affected by dust, hence the slight improvement in our  $A_V$  prediction. Finally, the  $\sigma$  of  $\log t_{\text{mod}}/\tau$  shows no strong correlation with any redshift.

Table 1 also lists the bias,  $\sigma$  and percentage of outliers for the various physical parameters at  $z = 0.55$  and 0.7. This supports the results found in Fig. 6, where the bias is comparable between redshifts for most parameters. The main exception is for sSFR, where while the median value is similar, the 25<sup>th</sup> percentile value is closer to 0 by 0.19 dex at  $z = 0.7$  than at 0.3. The median of

$\sigma$  is marginally lower for both the mwa and uwa. Again for the sSFR, we see a decrease in  $\sigma$  of 0.1 dex between  $z = 0.3$  and 0.7. In contrast, the metallicity  $\sigma$  values are comparable, a difference of only 0.03 dex as we increase the redshift. There is no clear trend in regards to  $t_{\text{mod}}/\tau$ . In comparison, KNN shows a decrease in median  $\sigma$  for a mwa (uwa) of 0.03 (0.01) dex between  $z = 0.3$  and 0.7. The median  $\sigma$  for KNN does not have any strong correlation with redshift for the sSFR,  $A_V$ , or  $\log t_{\text{mod}}/\tau$ ; however, overall they display the same behaviour as RF. In contrast, we see a negligible increase in the median  $\sigma$  of 0.04 dex for metallicity between  $z = 0.3$  and 0.7. While RF and KNN follow the same trend with an increase in  $z$ , RF again finds a better constraint than KNN.

#### 4. A test case: Red, green, and blue galaxies

As an example of how quantities retrieved from ML algorithms can be applied to real data, we consider here the use of sSFRs to classify galaxies into three categories: BC, mostly star-forming galaxies, GV, galaxies in the process of quenching, and RS, mostly quiescent galaxies (Salim 2014, and references within). Once we group simulated galaxies into BC, GV, and RS, we check for completeness of our classification for different S/Ns, at each redshift. We note that here we do not look at the purity of our classification as this is heavily dependent on both the distribution and fraction of galaxies in BC, GV, and RS. While completeness might also be affected by the actual distribution of

**Table 2.** sSFR boundary between BC and GV ( $\text{sSFR}_{\text{BC,GV}}$ ) and GV and RS ( $\text{sSFR}_{\text{GV,RS}}$ ) for the three redshifts explored in this work, i.e.  $z = 0.3, 0.55,$  and  $0.7$ .

$z$	$\log \text{sSFR}_{\text{BC,GV}}$	$\log \text{sSFR}_{\text{GV,RS}}$
0.3	-10.1	-11.6
0.55	-10.0	-11.5
0.7	-9.9	-11.4

**Table 3.** Completeness of BC, GV, and RS.

	Random forest			K-Nearest neighbour		
	SN = 10 (Completeness)					
Redshift	$z = 0.3$	$z = 0.55$	$z = 0.7$	$z = 0.3$	$z = 0.55$	$z = 0.7$
BC	0.99	0.99	0.99	0.99	0.99	0.99
GV	0.78	0.86	0.88	0.52	0.55	0.64
RS	0.99	0.98	0.98	0.97	0.96	0.96
	SN = 20 (Completeness)					
BC	0.99	0.99	0.99	0.99	0.99	0.99
GV	0.80	0.88	0.90	0.60	0.69	0.75
RS	0.99	0.98	0.97	0.98	0.98	0.97
	SN = 30 (Completeness)					
BC	0.99	0.99	0.99	0.99	0.99	0.99
GV	0.81	0.88	0.92	0.67	0.76	0.83
RS	0.99	0.98	0.98	0.98	0.98	0.98

**Notes.** The left and right set of columns show result for RF and KNN. Top, middle and bottom sets of rows shows completeness at  $S/N = 10, 20,$  and  $30$ . We tabulate results at each of three redshifts,  $z = 0.3, 0.55,$  and  $0.7$ .

galaxies, as objects close to the border are more likely to be misclassified than those at the centre, it is likely more robust against the fraction of populations in each region.

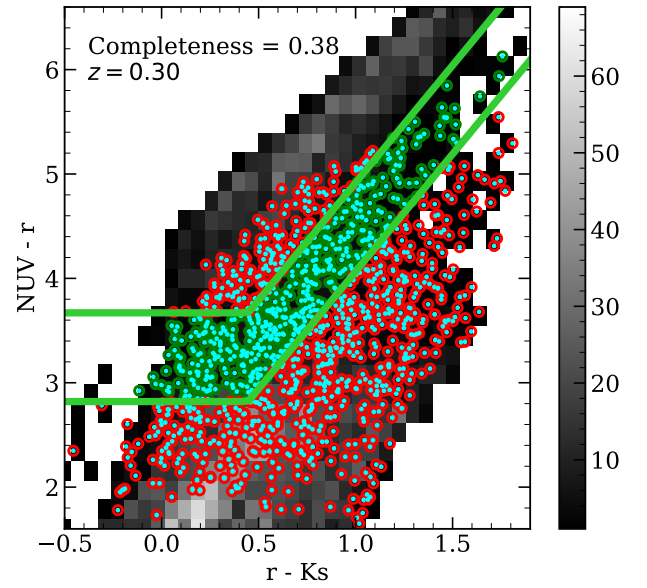
While there are several ways to define BC, GV, and RS, we utilise the definition based on sSFR here. We adopt a cut-off value for the sSFR that evolves with cosmic time, as this has been observed in both hydrodynamical simulations and observations (Fritz et al. 2014; Trayford et al. 2016; Phillipps et al. 2019; Wright & Lagos 2019; Jian et al. 2020). To define the cut-off boundary between star-forming and quiescent galaxies, we use the formula:  $\log(\text{sSFR}(\text{yr}^{-1})) = -11.0 + 0.5z$  (Furlong et al. 2015). We then consider 0.75 dex above (below) the cut-off point to define the border between BC (RS) and GV, resulting in a GV width of 1.5 dex in the sSFR. Table 2 shows the GV limits obtained by such a method, where  $\log \text{sSFR}_{\text{GV,RS}}$  is the border between GV and RS, and  $\log \text{sSFR}_{\text{BC,GV}}$  is the boundary between BC and GV.

Completeness is defined as

$$\text{Completeness} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (4)$$

where TP is the number of true positives, that is, simulated galaxies correctly identified to be in BC, GV, or RS; the FN is the number of false negatives, that is, galaxies misclassified as belonging to the ‘wrong’ region. Table 3 shows the completeness obtained with RF and KNN at each redshift for different S/Ns.

The completeness of galaxies in BC and RS is consistently very high ( $\geq 0.96$ ) irrespective of the S/N or ML algorithm we use. On the other hand, for GV galaxies, completeness is lower, although RF still performs relatively well, with completeness  $\geq 0.75$ . We note that not only does the GV completeness improve



**Fig. 7.** Distribution of galaxy templates in the colour–colour diagram. The lime-green lines define the demarcation used to select GV galaxies (Moutard et al. 2020). Cyan dots are GV galaxies, identified using the true sSFR at  $z = 0.3$ . Green (red) circles around the cyan dots show galaxies that are correctly identified (misidentified) as GV galaxies according to colour–colour selection. The colour bar on the side shows the number of galaxies.

with increasing S/N, but we also find higher values with increasing redshift, with an increase of  $\sim 10\%$  from  $z = 0.3$  to  $0.7$ . For KNN, while there is also an improvement with S/N and redshift, we find that the overall completeness is  $\leq 0.65$ , at the lowest S/N and redshift. This stems from the significant bias in predicting sSFRs with KNN (see Sect. 3.2), where we consistently underestimate the true value. The underestimation leads to misclassifying a GV galaxy into an RS one, while the opposite scenario is unlikely.

To compare our results with those obtained with other methods, we also compute the completeness of GV classification using a NUVrKs diagram (Arnouts et al. 2013). The density plot in Fig. 7 shows the distribution of simulated galaxies in colour–colour space. The completeness for this method is 0.38,  $\sim 0.40$  (0.30) lower than for RF (KNN) at the same redshift. We note that at different redshifts, we have different numbers of galaxies, and also, the definition of demarcation lines varies (see equation 1 of Moutard et al. 2020). Therefore, we repeated the analysis at  $z = 0.55$  and  $z = 0.7$ , finding that the completeness of NUVrKs-selected GV galaxies drops to 0.36 and 0.32, respectively. Methods that use the sSFR to define GV galaxies are expected to perform better than simple colour selections: the sSFR is the physical parameter that defines GV galaxies, and it spans a wide range of rest-frame NUVrKs colours. Therefore, a simple colour–colour selection of GV galaxies is prone to large incompleteness. On the contrary, the use of ML to estimate the sSFR with photometry and spectral indices has small uncertainties, leading to high completeness.

## 5. Discussion

We highlight the capabilities of two ML algorithms to retrieve the physical properties of galaxies based on photometric and spectroscopic information. We analysed how results vary for

simulated galaxy spectra at different redshifts, with  $S/N_{I,obs}$  of 10, 20, and 30, respectively. In the following, we discuss some caveats and possible limitations of our assumptions (Sect. 5.1). In addition, we compare our results with those obtained using a more traditional Bayesian approach (Sect. 5.2). Finally, since spectroscopic data can have a wide range of S/Ns, we analyse how much our results vary when the training and testing data have different S/Ns (Sect. 5.3).

### 5.1. Photometry versus spectroscopy

Both photometry and spectroscopy provide valuable constraints on the physical parameters of galaxies. Therefore, many state-of-the-art spectral fitting algorithms combine the two types of information (Carnall et al. 2018; Johnson et al. 2021; Cappellari 2023). In this section, we discuss the ability of ML algorithms to retrieve physical parameters of galaxies based on spectroscopy alone or with spectro-photometry information, discussing limitations due to observations and stellar population models.

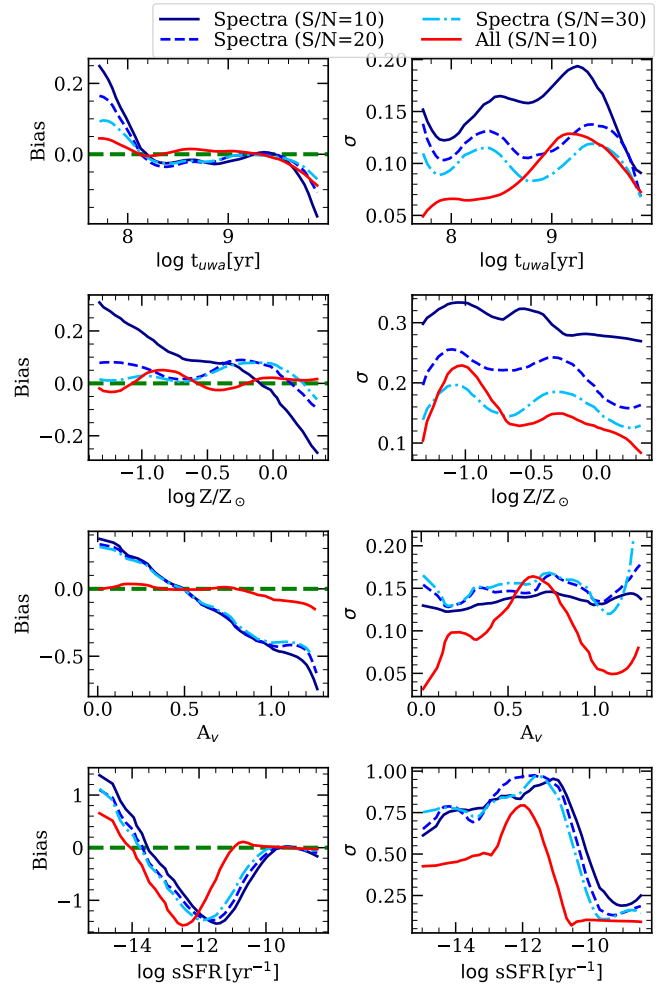
Figure 8 shows how well we can retrieve the uwa, metallicity,  $A_V$ , and sSFR using only spectroscopic information at  $z = 0.3$  (i; blue curves), an  $I$ -band S/N (per Å) of 10, 20, and 30; (ii; red curves), and both photometry (all bands) and spectroscopy, the latter with  $S/N = 10$ . Panels in the left (right) column of Fig. 8 show the bias (dispersion,  $\sigma$ ) on retrieved quantities.

For age and metallicity, at  $S/N = 10$ , the predictions generally give the highest bias values (especially at the extremes of the distribution) and  $\sigma$ . At  $S/N = 20$ , we find improvement in bias and  $\sigma$  – most noticeable for younger ages. However, the  $\sigma$  for spectra is still larger but with differences of less than 0.1 dex. At  $S/N = 30$ , the bias and  $\sigma$  are comparable both when using spectroscopic information only and photometry plus spectroscopy, although the latter method still provides lower  $\sigma$  values for ages younger than  $\log t_{uwa}[\text{yr}] \sim 8$ . We note that, while not shown, the mwa shows the same trend as the uwa.

In contrast to these trends, there is limited constraining capability on  $A_V$  solely based on spectral information, independent of the S/N, as spectral indices are measured over short wavelength ranges and are therefore insensitive to dust. We note the low constraining ability is due to  $D_n(4000)$  being sensitive to dust (MacArthur 2005). For the sSFR, we find a noticeable difference when using only spectral or all information; for the former, the increase in bias and  $\sigma$ , towards lower sSFR values, starts at  $\log \text{sSFR} \sim -10$  rather than at  $-11$ . While increasing the S/N tends to decrease the bias and  $\sigma$ , the degradation of constraining capability for  $\log \text{sSFR}$  remains at about  $-10$ .

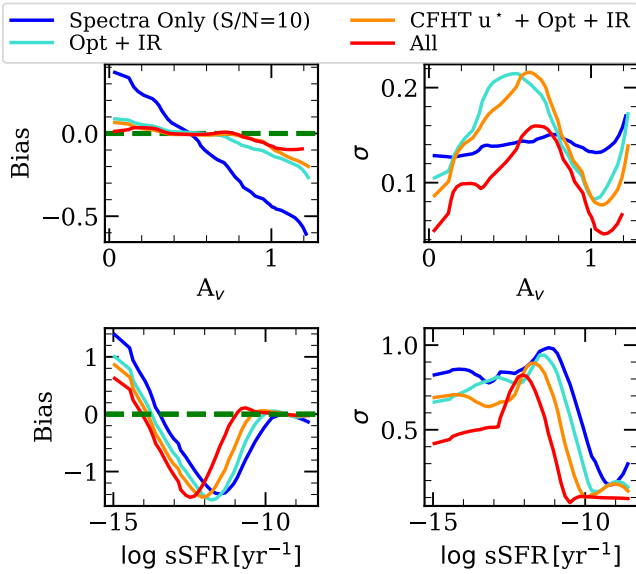
While we find that the inclusion of photometry gives us a higher constraining capability for the  $A_V$  and, more importantly, for the sSFR, we warn the reader that the uncertainties considered here do not account for all sources of systematics – as we consider only formal photometric errors from the COSMOS catalogue (Laigle et al. 2016). While we set a lower limit on photometric errors by adding 0.05 mag in quadrature to the formal errors on magnitudes, the true uncertainties, due to systematics, may be larger for several reasons. A significant source of systematics on photometry is dust, as the reddening correction depends on the adopted dust model. Different dust models may lead to large variations on measured magnitudes, with such systematics propagating to colours, thus affecting the retrieval of physical parameters (Salim & Narayanan 2020; Pacifici et al. 2023).

While NIR wavebands are less affected by dust, they are also susceptible to systematics due to differences among different stellar population models. To illustrate this point, we compare magnitudes obtained from BC16 and EMILES (Vazdekis et al.



**Fig. 8.** Bias (left) and  $\sigma$  (right) of the uwa, metallicity,  $A_V$ , and sSFR retrieved using either spectral indices only, at different S/Ns, or both spectroscopic and photometric information. The solid dark blue, dashed blue, and dash-dotted sky-blue curves show results for S/N = 10, 20, and 30, respectively. For comparison, the red curves show results obtained using both observed magnitudes and spectral indices at S/N = 10. The dashed green line in the left panels indicates zero bias. The results are for the RF algorithm at  $z = 0.3$ .

2015) simple stellar population (SSP) models. We considered only ages  $\geq 1$  Gyr, where EMILES stellar population models are ‘safe’ in the NIR spectral range ( $\lambda \geq 8950 \text{ \AA}$ , for all metallicities; see Vazdekis et al. 2015). While magnitude differences between the two sets of models are small in the optical (less than a few percent), large differences are found in the NIR. In particular, for an age of  $\sim 1$  Gyr at solar metallicity, we find differences of  $\geq 0.2$  mag in the observed  $H$  and  $K_s$  bands at  $z = 0.3$ . Such systematic differences are due to EMILES models relying on the empirical library Indo-US library (Valdes et al. 2004) whereas BC16 uses the semi-empirical library BaSeL 3.1 (Westera et al. 2002). Systematics in the models also affect the UV part of the spectra. In the UV, EMILES models rely on fully empirical stellar library, whereas BC16 is purely theoretical. The comparison of CFHT  $u^*$  magnitudes shows a maximum difference of 0.26 mag for a SSP age of  $\sim 1.4$  Gyr, at solar metallicity. These systematics may lead to higher uncertainties than what we have assumed on our magnitudes. Another possible source of systematics is that spectroscopic and photometric



**Fig. 9.** Same as Fig. 8 but also showing results when considering different observed magnitude combinations at  $z = 0.3$ . Note for all results shown, we also use all available spectral indices, at  $z = 0.3$ , with  $S/N = 10$ . The turquoise lines show results for observed optical (these consist of HSC  $g, r, i, z$ , and  $y$  filters) and IR magnitudes (VISTA  $H, J$ , and  $K_s$ ). The dark orange lines show results for observed optical, IR, and CFHT  $u^*$  magnitudes.

observations usually cover different galaxy regions. Both, slit and aperture spectroscopy usually cover the central parts of the galaxies, while magnitudes obtained from photometry are integrated in larger apertures. This may lead to discrepancies in the estimates of various physical parameters, depending on whether spectroscopic or photometric data are used. Last but not least, one should bear in mind that while optical photometry is affected by the age-metallicity degeneracy, for an unresolved stellar population, spectral indices are able to partially break it (e.g. Worthey 1994; Arimoto 1996). The combinations of UV and optical (Kaviraj et al. 2007) or optical and IR (Carter et al. 2009) colours have been shown to be successful at partially breaking this degeneracy. However, they suffer from other issues; for example, the UV and optical combination suffers from age–dust degeneracy.

In addition to systematics, one usually has limited access to far-UV (FUV) and near-UV (NUV) photometric data, because of the faint flux of galaxies at these wavelengths. For example, within the WEAVE-StePS, in the CFHTLS-W4 field, only 4% (20%) of the sources have FUV (NUV) photometry, while for the ELAIS field, only 10% of galaxies have GALEX data. In contrast, for  $I_{AB} < 20.5$  at  $z > 0.3$ , the COSMOS field has 60% and 70% of galaxies with FUV and NUV photometry, respectively. Finally, even when FUV photometry is available, its interpretation for old stellar populations is not trivial, as this spectral range includes contributions from stars that are difficult to model (e.g. very young stars and/or hot evolved stars; see Le Cras et al. 2016; Salvador-Rusiñol et al. 2020).

Despite such issues, photometry is a reasonable source of information for a sensible estimation of the  $A_V$  and sSFR. Without photometry, our estimates of the sSFR worsen significantly, affecting the completeness. For instance, when using only spectral indices at  $z = 0.3$ , the GV completeness drops to 0.43, 0.51, and 0.56 for  $S/N = 10, 20$ , and 30, respectively. Although these values are relatively low, they are still higher than that obtained

from the NUVrKs diagram (i.e. 0.38; see Sect. 4). A relevant question is what happens to the constraining capability on  $A_V$  and sSFR, when only limited photometric information is available. We address this point in Fig. 9, showing the bias and  $\sigma$  on the  $A_V$  and sSFR, for simulated galaxies at  $z = 0.3$ . The blue and red lines in Fig. 9 are the same as in Fig. 8: they show constraints from spectroscopy only, with  $S/N = 10$ , and photometry plus spectroscopy, respectively. The figure also shows results when spectroscopy is combined with (i) only optical and IR photometry (turquoise lines); and (ii) optical, IR, and  $u$ -band photometry (dark orange lines). Access to photometric information improves the accuracy of the  $A_V$  and, more importantly, the sSFR. When only optical and IR photometry are used, there is a marginal improvement in the estimate of the sSFR, leading to a relatively low completeness of 0.52 for GV galaxies, comparable to that for spectroscopy only, with  $S/N = 20$  (see above). On the other hand, adding CFHT  $u^*$  to optical and IR magnitudes (see the dark orange lines in Fig. 9) leads to a more significant improvement of sSFR estimates. The deterioration now starts at  $\log sSFR \sim -10.5$ , increasing GV completeness up to 0.62. This is particularly relevant for the WEAVE-StePS, where in the CFHTLS-W4 and COSMOS fields 99% and 70% of galaxies have  $u$ -band magnitudes at  $z > 0.3$  and  $I_{AB} < 20.5$ .

## 5.2. Comparison with Bayesian methodology

In this section, we compare results obtained from ML techniques with those from a more traditional Bayesian approach following the method outlined in Gallazzi et al. (2005) and Zibetti et al. (2017). To make a fair comparison with previous works presenting results for WEAVE-StePS simulated data (Costantin et al. 2019; Ditrani et al. 2023), we consider only UV and optical spectral indices, assuming  $S/N_{I,obs} = 10$ , excluding the photometric information. Under the assumption that errors are normally distributed, the goodness of the fit for a given statistical model is given by

$$\chi^2 = \sum_i \left[ \frac{O_i - M_i}{\sigma_i} \right]^2, \quad (5)$$

where  $O_i$  are observed values (i.e. index line strengths),  $\sigma_i$  their errors, and  $M_i$  are model values. Equation (5) can be converted into a posterior probability distribution function, defined as  $\mathcal{L} \propto e^{-\chi^2/2}$ . The predicted value of a given quantity is estimated by marginalising the probability distribution function with respect to it, taking median values (see Ditrani et al. 2023). We note that for a fair comparison, we also use median values for RF, namely we consider the median rather than the mean of 100 trees from the output of RF algorithm.

For a given physical parameter, we grouped simulated galaxies into groups of 1200 objects according to the true values. For each bin, we calculated the bias and  $\sigma$ . Table 4 shows quantities for the age and metallicity from both the RF algorithm and the Bayesian approach.

We find the largest differences in bias between the two methods for bins that contain either the oldest metal-poor populations ( $\log Z/Z_\odot = -1.17$ ) or the oldest metal-rich populations ( $\log Z/Z_\odot = 0.38$ ), particularly those at the boundaries of the parameter space. For the lowest and highest metallicity bins, the difference in bias (between the RF and the Bayesian methods) are 0.06 and 0.05 dex, respectively (see Table 4). Similarly for the  $u$ wa, we have the largest difference, of 0.18 Gyr, in the oldest age bin,  $\text{Age}_u \sim 7.94$  Gyr. In contrast to the bias, we find comparable results between both methods in their  $\sigma$ .

**Table 4.** Comparison of bias and  $\sigma$  of metallicity and age obtained using RF and Bayesian approach.

	RF		Bayesian	
	Bias	$\sigma$	Bias	$\sigma$
$\log Z/Z_{\odot}$				
-1.17	0.13	0.37	0.07	0.35
-0.67	0.10	0.29	0.07	0.27
-0.40	0.07	0.35	0.06	0.34
-0.05	0.00	0.32	0.05	0.32
0.22	-0.08	0.24	0.04	0.23
0.38	-0.19	0.23	-0.14	0.20
Age <sub>u</sub> (Gyr)				
0.09	0.02	0.03	0.02	0.03
0.30	0.00	0.11	-0.01	0.10
0.61	-0.03	0.20	-0.04	0.19
1.37	-0.07	0.52	-0.04	0.47
4.19	-0.23	1.34	-0.25	1.24
7.94	-2.55	1.35	-2.37	1.27

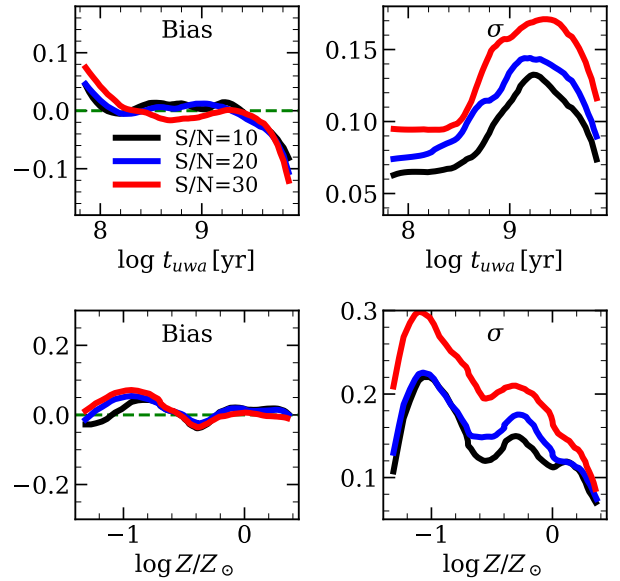
**Notes.** Top and bottom rows tabulates the bias and  $\sigma$  for metallicity and age across different bins.

While both the Bayesian and RF approaches perform similarly, the main advantage of RF is that it is orders of magnitude faster than the Bayesian approach, especially when the training has been completed. Such a large difference in time is due to the fact that ML methods have a lower computational complexity than Bayesian inference. Once the tree is constructed, the RF simply goes down the different branches to retrieve the physical parameter, while Bayesian inference requires a comparison with each template. Previous works in the literature have also noted differences in computational time between various types of ML algorithms and classical methods (Stensbo-Smidt et al. 2017; Domínguez Sánchez et al. 2018; Davidzon et al. 2019).

Finally, there have been studies that use more complex ML algorithms in astronomy for different purposes. For example, Martínez-Solaache et al. (2023) used an artificial neural network (ANN) and RF to classify objects into quasars, galaxies or stars, finding ANN to outperform RF, given proper calibration. In addition, Hunt et al. (2024) carried out a similar analysis to us, where they predicted the average ages of galaxies from the GAMA survey using ANN. They trained the ANN to predict average ages based on the optical spectral indices, finding results that are comparable to ours.

### 5.3. Varying the signal-to-noise of the training and testing samples

Throughout Sect. 3, we assumed the spectra in both the training and testing samples have the same S/N. For a given S/N, we included five random realisations of each galaxy in the training sample. The benefits of such a methodology are outlined by Shy et al. (2022), namely that it becomes possible for the ML algorithms to incorporate the uncertainties on the observables. However, when dealing with real data, one may expect to observe a large number of spectra (e.g.  $\sim 25\,000$  target galaxies from WEAVE-StePS) with a wide range of signal-to-noise ratios. Since training the ML algorithm for the S/N of each galaxy would be extremely time-consuming, we tested the effect of setting the S/N of the training sample to a different value than that of the testing sample. In the following, we discuss the impact of this test on our predictions of age and metallicity.



**Fig. 10.** Bias (left panels) and  $\sigma$  (right panels) of the uwa (top panels) and metallicity (bottom panels) as a function of true values. Black, blue, and red lines correspond to training samples with S/N = 10, 20, and 30, respectively. Note that for the testing sample, we set S/N = 10 in all three cases.

Figure 10 shows the change in bias and  $\sigma$  for both age and metallicity when the training sample has a S/N of 10, 20, and 30 (see the black, blue, and red lines, respectively), and the testing sample has a S/N of 10. There are no significant variations in the age bias with an increase in the difference in S/N between the training and testing samples. The same applies to metallicity, although for S/N = 30, we have a slightly higher bias (within 0.03 dex) at the lowest metallicity (see the red and black curves in the lower-left panel of Fig. 10). On the contrary, we find a clear trend of  $\sigma$  with the S/N of the training sample. For age and metallicity, the  $\sigma$  increases as the difference in S/N between training and testing samples increases, implying a deterioration in the estimate of physical parameters. The difference in  $\sigma$  between S/N = 10 and 30 amounts to  $\sim 0.05$  dex ( $\sim 0.10$  dex) for the uwa (metallicity), whereas the difference between S/N = 10 and 20 is considerably smaller,  $\leq 0.02$  dex, for both quantities. This suggests that for galaxies with S/Ns close to 10, assuming a fixed S/N = 10 for the training sample, should have negligible impact, for most of the applications, on the estimate of physical parameters. This is actually the case for the WEAVE-StePS, where one expects most of the source to have spectra with S/Ns in the range from 8.8 to 16.2 (see Fig. 9 of Iovino et al. 2023b). Figure 10 suggests that, in this case, assuming S/N = 10 to create the training sample will give results fully consistent with those shown in Fig. 4, where both the training and testing samples were assumed to have the same S/N. This result is encouraging and opens up new opportunities in the analysis of massive datasets using ML techniques.

## 6. Summary and conclusion

With the increased number of upcoming spectroscopic surveys, the astrophysics community will have access to a large number of high-resolution spectra with a wider spectral coverage than ever before, coupled to rich photometric information. These observations span from the local Universe to high redshifts and will cover both the UV and optical rest frames (Euclid Collaboration 2023; Iovino et al. 2023a,b). With access to such a large quantity of data,

there is an increasing need to find faster ways to extract accurate estimates of physical parameters.

In this work, for the first time, we tested the predictive capabilities of two easy-to-implement ML algorithms (KNN and RF) in estimating the average age (both the  $m_w$  and  $u_w$ ), metallicity, sSFR,  $A_v$ , and  $t_{\text{mod}}/\tau$  using photometric and spectral information. We measured UV and optical spectral indices, which are sensitive to age and metallicity, in template spectra and obtained synthetic photometry in the representative UV, optical, and NIR bands.

First, we analysed how well we can retrieve the physical parameters, assuming no uncertainty in our observed measurements. We then produced realistic simulations of WEAVE-StePS-like spectra for  $S/N = 10, 20, \text{ and } 30$  at redshifts  $z = 0.3, 0.55, \text{ and } 0.7$  following the method outlined in Costantin et al. (2019). We also assigned uncertainties to the magnitudes calculated for each galaxy template using the photometric ancillary data. We find that, in general, both RF and KNN are able to accurately predict most of the parameters, with a low bias, for all  $S/N$ s and at different redshifts. The main exception was for the sSFR and  $t_{\text{mod}}/\tau$ . For the sSFR, while the median bias was  $\sim -0.03$ , the lower percentiles ranged from  $\sim 0.61$ , at  $z = 0.7$ , to  $\sim 0.86$ , at  $z = 0.3$  (see Table 1). We note that this large scatter in the bias is due to an underestimation of the sSFR, specifically below  $\lesssim -11.0$  (see Figs. 5 and 6). The  $\sigma$  were also relatively low for RF (KNN); for the age we found a maximum  $\sigma \sim 0.10$  (0.16) dex for  $S/N = 10$  at  $z = 0.3$ , which decreased to  $\sim 0.08$  (0.11) dex for  $S/N = 30$ . At higher redshifts, we find a similar constraint. For metallicity, the  $\sigma$  on the estimates vary from  $\lesssim 0.11$ – $0.16$  (0.12–0.25) dex for different  $S/N$ s at  $z = 0.3$ . At higher redshifts for  $S/N = 10$ , we find a slightly worse constraint,  $\sigma \sim (0.17) 0.24$  dex. The  $A_v$  has a typical  $\sigma$  of 0.09 (0.10–0.13) mag. Again, the sSFR has a different constraining power depending on the value. For star-forming galaxies,  $\log \text{sSFR} \gtrsim -10$ , we find a typical  $\sigma$  of  $\lesssim 0.1$  (0.2) dex, which increases to 0.3–0.4 dex at lower values, depending on the redshift. For RF and KNN, we find no strong trend between redshifts,  $S/N$ s, and constraint ability for  $t_{\text{mod}}/\tau$ .

In addition, the retrieved sSFR was utilised to classify galaxies into three categories, BC, GV and RS, and it was checked for completeness. The completeness was satisfactory for RF within each region, with GV values of  $\gtrsim 0.75$  at  $z = 0.3$ . At a redshift of  $z = 0.7$  we find a much higher completeness,  $\sim 0.90$ . In comparison, the completeness in BC and RS was much higher,  $\gtrsim 0.96$ . While KNN performed similarly well for BC and RS, the GV completeness was lower and ranged from  $\sim 0.52$ – $0.67$  and  $\sim 0.64$ – $0.84$  at  $z = 0.3$  and  $0.7$ , respectively. Although KNN had a lower completeness, it can still be used to accurately estimate the sSFR, thus giving a higher completeness than what we found for colour–colour selection,  $\sim 0.35$  (see Sect. 4). Finally, with a more traditional approach, Bayesian statistics, we observe outcomes similar to those of the ML algorithms. Nevertheless, considering the computational time, the ML algorithm is considerably faster than the traditional method. This is especially true since most of the computational time needed for ML is for training, which is only required once. Once trained, the prediction of physical parameters for any new set of galaxies ( $\sim 9600$ ) can be achieved within seconds. Therefore, ML techniques are an excellent and efficient tool that can be used to exploit the high-quality data that will be available from WEAVE-StePS, and the huge quantity of data<sup>3</sup> from upcoming<sup>4</sup>

large spectroscopic surveys from the MULTIPLEXED SURVEY TELESCOPE (MUST; Zhang et al. 2023), the MAUNA KEA SPECTROSCOPIC EXPLORER (MSE; The MSE Science Team 2019), and the WIDE-FIELD SPECTROSCOPIC TELESCOPE (WST; Mainieri et al. 2024).

*Acknowledgements.* Funding for the WEAVE facility has been provided by UKRI STFC, the University of Oxford, NOVA, NWO, Instituto de Astrofísica de Canarias (IAC), the Isaac Newton Group partners (STFC, NWO, and Spain, led by the IAC), INAF, CNRS-INSU, the Observatoire de Paris, Région Île-de-France, CONCYT through INAOE, Konkoly Observatory (CSFK), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Lund University, the Leibniz Institute for Astrophysics Potsdam (AIP), the Swedish Research Council, the European Commission, and the University of Pennsylvania. The WEAVE Survey Consortium consists of the ING, its three partners, represented by UKRI STFC, NWO, and the IAC, NOVA, INAF, GEPI, INAOE, and individual WEAVE Participants. Please see the relevant footnotes for the WEAVE website and for the full list of granting agencies and grants supporting WEAVE. This work makes use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC), <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the *Gaia* Multilateral Agreement. This work makes use of data from the VIMOS VLT Deep Survey, the VIPERS-MLS database and the HST-COSMOS database, operated by CeSAM/Laboratoire d’Astrophysique de Marseille, France. This work makes use of observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT), which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work makes use of data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. J.A. acknowledges financial support from INAF-WEAVE funds, program 1.05.03.04.05 and INAF-OABrera funds, program 1.05.01.01. A.I., F.D., M.L. acknowledge financial support from INAF Mainstream grant 2019 WEAVE StePS 1.05.01.86.16 and INAF Large Grant 2022 WEAVE StePS 1.05.12.01.11. R.G.B. and R.G.D. acknowledge financial support from the Severo Ochoa grant CEX2021-001131-S funded by MCIN/AEI/10.13039/501100011033 and to PID2022-141755NB-I00. A.F.M. acknowledges support from RYC2021-031099-I and PID2021-123313NA-I00 of MICIN/AEI/10.13039/501100011033/FEDER,UE,NextGenerationEU/PR M.Bi. acknowledges support from STFC grant numbers ST/N021702/1 G.B., M.Bo., F.R.D., A.I., F.L.B., M.L., P.M., B.P., C.T., D.V., S.Z., and A.R.G. acknowledge financial support from INAF funds, program 1.05.01.86.16 – Mainstream 2019. L.C. acknowledges support by grant PIB2021-127718NB-100 and PID2022-139567NB-I00 from the Spanish Ministry of Science and Innovation/State Agency of Research MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe. Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. JHK acknowledges grant PID2022-136505NB-I00 funded by MCIN/AEI/10.13039/501100011033 and EU, ERDF, E.M.C. and A.P. are supported by the MUR grant PRIN INAF 2022 C53D2300085-0006 and Padua University grants DOR 2021-2023. We also like to thank Daniela Bettoni for providing suggestions to improve this paper.

## References

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, 70, S4  
 Altman, N. S. 1992, *Am. Stat.*, 46, 175  
 Angthopo, J., Ferreras, I., & Silk, J. 2019, *MNRAS*, 488, L99  
 Angthopo, J., Ferreras, I., & Silk, J. 2020, *MNRAS*, 495, 2720  
 Arimoto, N. 1996, *ASP Conf. Ser.*, 98, 287  
 Arnouts, S., Le Floch, E., Chevallard, J., et al. 2013, *A&A*, 558, A67  
 Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, *ApJ*, 600, 681  
 Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2008, *ApJ*, 683, 12  
 Bennett, C. L., Larson, D., Weiland, J. L., & Hinshaw, G. 2014, *ApJ*, 794, 135  
 Bonjean, V., Aghanim, N., Salomé, P., et al. 2019, *A&A*, 622, A137  
 Breiman, L. 2001, *Mach. Learn.*, 45, 5  
 Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, 351, 1151  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000  
 Cappellari, M. 2017, *MNRAS*, 466, 798  
 Cappellari, M. 2023, *MNRAS*, 526, 3273

<sup>3</sup> <https://ingconfluence.ing.iac.es/confluence/display/WEAV/The+WEAVE+Project>

<sup>4</sup> <https://ingconfluence.ing.iac.es/confluence/display/WEAV/WEAVE+Acknowledgements>

- Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, *MNRAS*, **480**, 4379
- Carter, D., Smith, D. J. B., Percival, S. M., et al. 2009, *MNRAS*, **397**, 695
- Chabrier, G. 2003, *PASP*, **115**, 763
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., & Gomes, J. M. 2005, *MNRAS*, **358**, 363
- Costantin, L., Iovino, A., Zibetti, S., et al. 2019, *A&A*, **632**, A9
- Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, **450**, 1937
- Dalton, G., Trager, S. C., Abrams, D. C., et al. 2012, *SPIE Conf. Ser.*, **8446**, 84460P
- Dalton, G., Trager, S., Abrams, D. C., et al. 2016, *SPIE Conf. Ser.*, **9908**, 99081G
- Davidzon, I., Laigle, C., Capak, P. L., et al. 2019, *MNRAS*, **489**, 4817
- de Graaff, A., Bezanson, R., Franx, M., et al. 2021, *ApJ*, **913**, 103
- Ditrani, F. R., Longhetti, M., La Barbera, F., et al. 2023, *A&A*, **677**, A93
- Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, *MNRAS*, **476**, 3661
- Donnari, M., Pillepich, A., Nelson, D., et al. 2019, *MNRAS*, **485**, 4817
- Euclid Collaboration (Bisigello, L., et al.) 2023, *MNRAS*, **520**, 3529
- Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, *ApJ*, **665**, 265
- Fritz, A., Scodreggio, M., Ilbert, O., et al. 2014, *A&A*, **563**, A92
- Furlong, M., Bower, R. G., Theuns, T., et al. 2015, *MNRAS*, **450**, 4486
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, **362**, 41
- Gavazzi, G., Bonfanti, C., Sanvito, G., Boselli, A., & Scodreggio, M. 2002, *ApJ*, **576**, 135
- Guzzo, L., Scodreggio, M., Garilli, B., et al. 2014, *A&A*, **566**, A108
- Hopkins, A. M., Driver, S. P., Brough, S., et al. 2013, *MNRAS*, **430**, 2047
- Hunt, L. J., Pimblet, K. A., & Benoit, D. M. 2024, *MNRAS*, **529**, 479
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841
- Iovino, A., Mercurio, A., Gallazzi, A. R., et al. 2023a, *The Messenger*, **190**, 22
- Iovino, A., Poggianti, B. M., Mercurio, A., et al. 2023b, *A&A*, **672**, A87
- Jian, H.-Y., Lin, L., Koyama, Y., et al. 2020, *ApJ*, **894**, 125
- Jim, S., Trager, S. C., Dalton, G. B., et al. 2024, *MNRAS*, **530**, 2688
- Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, *ApJS*, **254**, 22
- Kaviraj, S., Rey, S. C., Rich, R. M., Yoon, S. J., & Yi, S. K. 2007, *MNRAS*, **381**, L74
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24
- Le Cras, C., Maraston, C., Thomas, D., & York, D. G. 2016, *MNRAS*, **461**, 766
- Li, R., Napolitano, N. R., Feng, H., et al. 2022, *A&A*, **666**, A85
- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, *ApJS*, **184**, 218
- López Fernández, R., Cid Fernandes, R., González Delgado, R. M., et al. 2016, *MNRAS*, **458**, 184
- MacArthur, L. A. 2005, *ApJ*, **623**, 795
- Mainieri, V., Anderson, R. I., Brinchmann, J., et al. 2024, ArXiv e-prints [arXiv:2403.05398]
- Martínez-Solaache, G., Queiroz, C., González Delgado, R. M., et al. 2023, *A&A*, **673**, A103
- Martins, L. P., González Delgado, R. M., Leitherer, C., Cerviño, M., & Hauschildt, P. 2005, *MNRAS*, **358**, 49
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *ApJS*, **225**, 27
- Moutard, T., Malavasi, N., Sawicki, M., Arnouts, S., & Tripathi, S. 2020, *MNRAS*, **495**, 4237
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Comput. Astrophys. Cosmol.*, **6**, 2
- Oke, J. B. 1974, *ApJS*, **27**, 21
- Pacifici, C., Iyer, K. G., Mobasher, B., et al. 2023, *ApJ*, **944**, 141
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Phillipps, S., Bremer, M. N., Hopkins, A. M., et al. 2019, *MNRAS*, **485**, 5559
- Salim, S. 2014, *Serb. Astron. J.*, **189**, 1
- Salim, S., & Narayanan, D. 2020, *ARA&A*, **58**, 529
- Salvador-Rusiñol, N., Vazdekis, A., La Barbera, F., et al. 2020, *Nat. Astron.*, **4**, 252
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, *MNRAS*, **371**, 703
- Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, *MNRAS*, **440**, 889
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Shy, S., Tak, H., Feigelson, E. D., Timlin, J. D., & Babu, G. J. 2022, *AJ*, **164**, 6
- Simet, M., Chartab, N., Lu, Y., & Mobasher, B. 2021, *ApJ*, **908**, 47
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Springel, V., Pakmor, R., Pillepich, A., et al. 2017, *MNRAS*, **475**, 676
- Stensbo-Smidt, K., Gieseke, F., Igel, C., Zirm, A., & Steenstrup Pedersen, K. 2017, *MNRAS*, **464**, 2577
- Strateva, I., Ivezić, Z., Knapp, G. R., et al. 2001, *AJ*, **122**, 1861
- Thanh Noi, P., & Kappas, M. 2018, *Sensors*, **18**, 18
- The MSE Science Team (Babusiaux, C., et al.) 2019, ArXiv e-prints [arXiv:1904.04907]
- Trayford, J. W., Theuns, T., Bower, R. G., et al. 2016, *MNRAS*, **460**, 3925
- Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., & Bell, D. J. 2004, *ApJS*, **152**, 251
- van der Wel, A., Noeske, K., Bezanson, R., et al. 2016, *ApJ*, **223**, 29
- Vazdekis, A., Coelho, P., Cassisi, S., et al. 2015, *MNRAS*, **449**, 1177
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, *MNRAS*, **463**, 3409
- Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, *A&A*, **381**, 524
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, *ApJ*, **691**, 1879
- Worthey, G. 1994, *ApJS*, **95**, 107
- Wright, R. J., Lagos, C., d. P., Davies, L. J. M., et al. 2019, *MNRAS*, **487**, 3740
- York, D. G., Adelman, J., Anderson, J., et al. 2000, *AJ*, **120**, 1579
- Zahid, H. J., Geller, M. J., Fabricant, D. G., & Hwang, H. S. 2016, *ApJ*, **832**, 203
- Zhang, Y., Jiang, H., Shectman, S., et al. 2023, *PhotoniX*, **4**, 16
- Zibetti, S., Gallazzi, A. R., Ascasibar, Y., et al. 2017, *MNRAS*, **468**, 1902

- 
- 1 INAF-Osservatorio Astronomico di Brera, via Brera 28, 20121 Milano, Italy
  - 2 INAF – Osservatorio Astronomico di Capodimonte, Via Moiriello 16, 80131 Napoli, Italy
  - 3 Università degli studi di Milano-Bicocca, Piazza della scienza, 20125 Milano, Italy
  - 4 Centro de Astrobiología (CAB), CSIC-INTA, Ctra. de Ajalvir km 4, Torrejón de Ardoz, 28850 Madrid, Spain
  - 5 INAF – Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy
  - 6 Departamento de Física Teórica, Universidad Autónoma de Madrid, 28049 Madrid, Spain
  - 7 Instituto de Física de Partículas y del Cosmos (IPARCOS), Universidad Complutense de Madrid, 28040 Madrid, Spain
  - 8 Dept. Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK
  - 9 INAF–Osservatorio Astronomico di Padova, vicolo dell’Osservatorio 5, 35122 Padova, Italy
  - 10 Instituto de Astrofísica de Canarias, IAC, Vía Láctea s/n, 38205 La Laguna (S.C. Tenerife), Spain
  - 11 Departamento de Astrofísica, Universidad de La Laguna, 38206, La Laguna (S.C. Tenerife), Spain
  - 12 Isaac Newton Group of Telescopes, ING, 38700 La Palma (S.C. Tenerife), Spain
  - 13 INAF – Osservatorio di Astrofisica e Scienza dello Spazio, Via P. Gobetti 93/3, 40129 Bologna, Italy
  - 14 School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK
  - 15 INAF – IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
  - 16 Dipartimento di Fisica e Astronomia “G. Galilei”, Università di Padova, vicolo dell’Osservatorio 3, 35122 Padova, Italy
  - 17 RAL, Space, Science and Technology Facilities Council, Harwell, Didcot OX11 0QX, UK
  - 18 Instituto de Astrofísica de Andalucía (CSIC), PO Box 3004, 18080 Granada, Spain
  - 19 Instituto de Astronomía y Ciencias Planetarias de Atacama (INCT), Universidad de Atacama, Copayapu 485, Copiapó, Chile
  - 20 Kapteyn Astronomical Institute, Rijksuniversiteit Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
  - 21 Dipartimento di Fisica “E.R. Caianiello”, Università degli studi di Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy
  - 22 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
  - 23 Università di Bologna – Department of Physics and Astronomy, Via Gobetti 93/2, 40129 Bologna, Italy

## Appendix A: Optimising ML algorithms

For the optimisation of the ML algorithms, it is necessary to fine-tune the hyperparameters as these will affect the prediction of the physical parameters that we considered. We use both sets of observables (the spectral indices and observed magnitude) as input. For the output, we estimate the physical parameters —  $m_{\text{wa}}$ ,  $u_{\text{wa}}$ ,  $t_{\text{mod}}/\tau$ , metallicity,  $A_{\text{v}}$ , and sSFR. The hyperparameters are adjusted for multiple runs, and two different statistics, R-squared ( $R^2$ ) and root-mean-square-error (RMSE), are used to assess the effectiveness of the algorithms. The  $R^2$  is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i^{\text{pred}} - x_i^{\text{true}})^2}{\sum_{i=1}^n (x_i^{\text{true}} - \bar{x}_i^{\text{true}})^2}, \quad (\text{A.1})$$

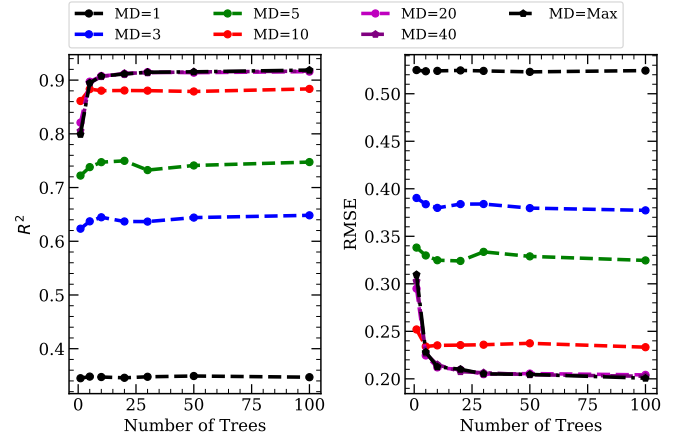
where  $\bar{x}_i^{\text{true}}$  is mean value of the physical parameter and  $n$  is the number of samples. Similarly RMSE is formulated as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i^{\text{pred}} - x_i^{\text{true}})^2}{n}}. \quad (\text{A.2})$$

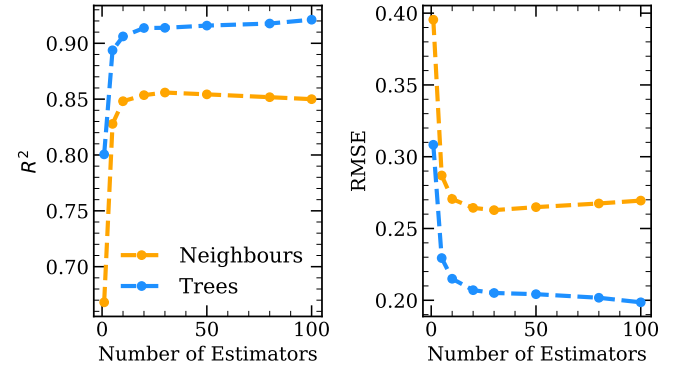
These statistics are commonly used to evaluate the goodness of fit for regression. The  $R^2$  value ranges from 0 to 1, where a value  $\sim 1$  indicates a good fit. Unlike  $R^2$ , RMSE provides an absolute measure of fit, where a lower value indicates a better fit. We note that these statistics are calculated for each physical parameter, such as the  $m_{\text{wa}}$  and  $u_{\text{wa}}$ , and that the final statistics quoted are averaged across all physical parameters considered.

For the RF, we focus on tuning two hyperparameters: the number of trees and the maximum depth (MD). Figure A.1 shows the changes in  $R^2$  and RMSE for different values of these hyperparameters. It is important to note that for our templates, MD=Max corresponds to a range of values between  $44 < \text{MD} < 58$ , depending on the particular run, due to the random nature of the RF algorithm. Both  $R^2$  and RMSE exhibit higher sensitivity to changes in MD compared to the number of trees. Both  $R^2$  and RMSE show similar values for  $\text{MD} > 20$  and number of trees  $> 25$ .

For the KNN algorithm, the hyperparameter that was optimised was the number of neighbours,  $K$ . Figure A.2 shows the  $R^2$  (left) and RMSE (right) values for different values of  $K$ . The performance of KNN is dependent on the value of  $K$ , but beyond  $K \geq 20$  there is minimal change in the  $R^2$  and/or RMSE statistics. To mitigate overfitting, a larger  $K$  value is chosen than the point at which the statistics flatten. Therefore, 100 neighbours are selected. For comparison, the statistics obtained for the RF algorithm, using MD=Max, are also plotted. It is noted that although the plateau of the statistics begins around the same number of estimators for both algorithms, the  $R^2$  and RMSE values also indicate a better fit of the model. Finally, we also checked the variance around the mean for  $R^2$  and RMSE for both algorithms.



**Fig. A.1.** Variation in  $R^2$  (left) and RMSE (right) for different numbers of trees and maximum depths (MDs). The number of trees varies (1, 5, 10, 20, 30, 50, and 100), while the MDs we consider are 1 (dashed black line with circles), 3 (dashed blue line with circles), 5 (dashed green line with circles), 10 (dashed red line with circles), 20 (dashed magenta line with circles), 40 (solid purple lines with pentagon markers), and ‘Max’ (solid black lines with pentagon markers). The statistics are averaged over all six parameters that we consider. Note that these runs assume  $S/N=10$ , and the simulated templates are at  $z = 0.3$ .



**Fig. A.2.** Variation in  $R^2$  (left) and RMSE (right) with the change in the number of estimators, neighbours for KNN (dashed orange line), and trees for RF (dashed blue line). These statistics were computed by considering the difference between true and predicted values. These results assume the spectra have  $S/N=10$  and the templates are at  $z = 0.3$ .

We set the number of trees or nearest neighbours ranging from 1 to 100, and for each configuration we carried out 30 independent runs of KNN and RF. For each run, we calculated the  $R^2$  and RMSE statistics and computed the variance in these statistics. We find low evidence of overfitting as both KNN and RF exhibit minimal variance with standard deviation  $\lesssim 10^{-3}$ , in  $R^2$  and RMSE, regardless of the number of trees/neighbours.