Highlights of Spanish Astrophysics XII, Proceedings of the XVI Scientific Meeting of the Spanish Astronomical Society held on July 15 - 19, 2024, in Granada, Spain. M. Manteiga, F. González Galindo, A. Labiano Ortega, M. Martínez González, N. Rea, M. Romero Gómez, A. Ulla Miguel, G. Yepes, C. Rodríguez López, A. Gómez García and C. Dafonte (eds.), 2025

Decoding galaxies: harnessing neural networks for stellar populations in the J-PAS survey

Domínguez Sánchez, H.[1,2], Coelho, P.[3], Bruzual, G.[4] Hernán Caballero A.[1] and the J-PAS collaboration

[1] Centro de Estudios de Física del Cosmos de Aragón, Plaza de San Juan, 1, E-44001 Teruel, Spain
[2] Instituto de Física de Cantabria, Av. de los Castros, 39005 Santander, Cantabria, Spain
[3] Universidade de São Paulo, Instituto de Astronomia, Geofïsica e Ciências Atmosféricas, Rua do Matão, 1226, 05508-090, São Paulo, SP, Brazil
[4] IRyA, Campus Morelia, Michoacán, C.P. 58089, México

Abstract

J-PAS (Javalambre Physics of the Accelerating Universe Astrophysical Survey) is a groundbreaking photometric survey covering 8500 deg2 of the visible sky from Javalambre, capturing data in 56 distinct bands. This survey promises to revolutionize galaxy evolution studies by observing millions of galaxies with pseudo-spectral resolution. A crucial aspect of this analysis involves recovering stellar population (SP) parameters from the observed galaxy photometry. In this study, we demonstrate the effectiveness of a neural network, trained on synthetic J-PAS photometry from Single Stellar Population (SSP) models with realistic noise, in predicting SP parameters such as age, metallicity, and dust attenuation. To enhance the robustness of our predictions against variations in SP models, we combine training samples from different SSP libraries (eMILES, Chabrier & Bruzual, XSL). Our results indicate that the NN can accurately recover SP parameters with almost no bias and small scatter, especially for high S/N galaxies.

1 Introduction

The field of galaxy evolution aims at understanding when and how galaxies formed and evolved to become the population of observed galaxies in the local Universe. An indispensable ingredient for such studies are the SP of galaxies, and, in particular, their main age, metallicity and dust attenuation. SP can be recovered from spectroscopic or photometric observations, being the former more precise due to their higher spectral resolution (allowing for the detection of important features like absorption/emission lines or spectral indices).

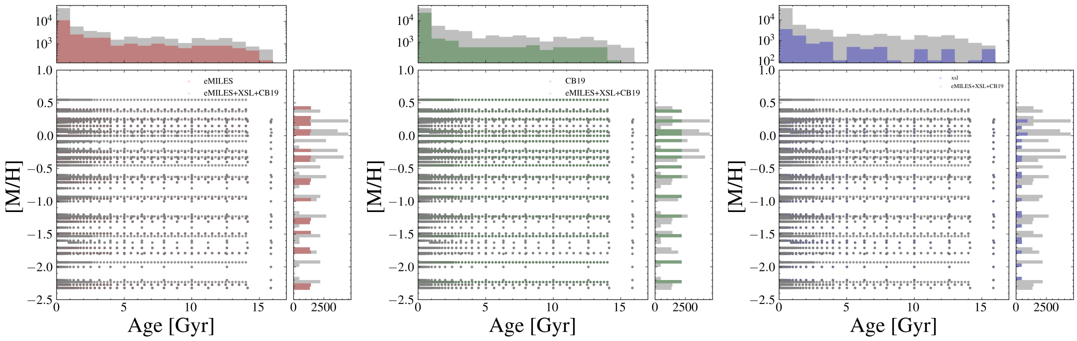

Figure 1: Age-metallicity coverage of the three SSP synthesis models used in this work to train the NN models. The gray symbols (dots and histograms) show the combination of the three SSP libraries, while in each panel we show the individual libraries: E-MILES in red (left panel), CB19 in green (middle panel) and XSL in blue (right panel). Note that the CB19 models include a very large fraction of SSP models at young ages, while the XSL models extend to very old ages.

However, spectroscopic observations are often affected by selection biases because they target specific (generally bright) objects. The J-PAS survey [2, 3] is a photometric survey specially designed to combine the best of both worlds, by observing a large fraction of the sky with a set of 56 narrow filters, providing pseudo-spectroscopic observations ($R = \delta\lambda/\lambda \sim 40$–$80$) up to magnitude $\mathrm{mag}_i \sim 22$.

SED-fitting codes (e.g. [1], [4], [6]) have been widely utilized to retrieve SP parameters from photometry, by finding the SP parameters that minimise the difference between observations and theoretical models. One of their main disadvantages is that they often require a predefined set of templates for comparison. A limited parameter space may fail to adequately represent the galaxy population. On the other hand, they can be computationally expensive if a wide range of parameter space needs to be covered. In recent years, there has been a surge in the application of machine learning algorithms, particularly neural networks (NNs), for various tasks in astronomy, with a great level of success in image analysis. However, their potential for recovering stellar populations has remained barely unexplored. In this work, we present preliminary results on NN predictions of SP parameters from J-PAS synthetic photometry trained with the combination of three SSP libraries, to have a wider coverage of the parameter space.

2 Data

2.1 Single stellar population models

We use supervised learning to train a NN to predict SP parameters (age, metallicity, dust attenuation) from J-PAS photometry. Our training sample are SSP models for which both the photometry and the SP parameters are well known. We choose this approach instead of

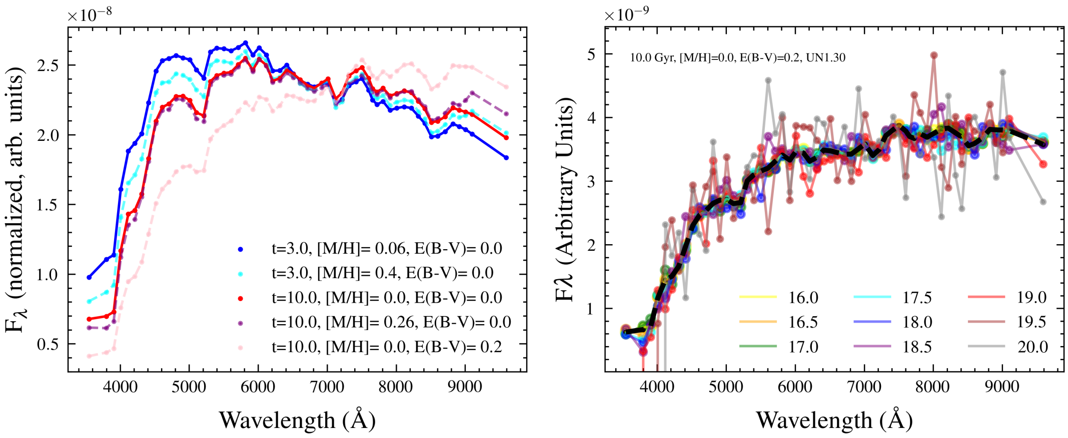

Figure 2: Left panel: Example of synthetic J-PAS photometry from eMILES SSP for a young (3 Gyr, blue colors) and an old (10 Gyr, reddish colors) SSP with different metallicities and dust attenuation, as stated in the legend.The fluxes are normalized to the mean value of the flux for each SSP. Right panel: synthetic SED for an old SP with different levels of noise, as expected for different observed magnitudes values (corresponding to the colors in the legend).

using photometry from observed galaxies, because their 'ground truth' have to be derived by other means (e.g., SED or spectra fitting) with their corresponding uncertainties and biases. We combine three different SSP libraries: E-MILES [9], CB19 [8] and the X-shooter Spectral Library [10, XSL herafter]. We refer the reader to the corresponding references for a more detailed description of the SSP libraries. By mixing different libraries in our training sample we aim to make the model more robust against spectral resolution, wavelength coverage or any bias that may arise from the parameter space sampling of the SSP libraries. Figure 1 shows the age-metallicity parameter space covered by the 26622 E-MILES models used in this work (red), the 32780 CB19 models (green) and the 9936 XSL models (blue) compared to the parameter spaced covered by the combined SP libraries combined (gray). For each metallicity-age pair, the models cover 10 different dust attenuation values (E(B-V)=[0-1]).

2.2 J-PAS photometry

J-PAS is a narrow band, wide field Cosmological Survey from the Javalambre Observatory (Teruel, Spain) with a set of 56 unique photometric filters which will allow exquisite photometric redshift determinations and precise measurement of the SP of galaxies (e.g., [5, 7]). The synthetic J-PAS photometry is generated by convolving the SSP models with the J-PAS filters. Figure 2 shows examples of the synthetic photometry from E-MILES library for a young (3Gyr) and an old (10Gyr) SSP with different metallicities and dust attenuation values. Note that the SSP models are at rest-frame wavelength (z=0) and have no noise. We introduce artificial noise in the form of random Gaussian variations to create mock photometric samples consistent with the level of noise expected at different magnitudes. The amplitude of the gaussian variations is representative of the typical uncertainties of galaxies from mini-JPAS, and is derived as the average of the photometric error in the AUTO aperture for each band, as a function of AUTO magnitude in 0.5 magnitude bins. We generate synthetic spectra of mock observed magnitudes' from 16.0 to 22.0 mag in 0.5 magnitude bins, thus, multiplying by 14 the number of SSP models (901394).

2.3 Neural Network

We employ a vanilla NN architecture to train our machine learning model. After some tests, the best configuration was obtained using as input the normalised flux data vector (56 dimensions) in each J-PAS filter (in F_ν), normalised by the median value of the flux of each SSP model). The NN is composed of a hidden layer with 128 neurons, three hidden layers with 256 neurons each, and a dense layer returning the output. `ReLu` activation is used in the intermediate layers and linear activation in the final layer. We found better results when training three independent models for each parameter (log(age), [M/H] and E(B-V)) rather than a single model with three dimensions in the output. The 'labels' (SP parameters) are also normalised by their maximum value. We combine the SSP models from the three libraries (eMILES, CB19 and XSL) and divide the sample into train/validation and test with a 64%, 16% 20% split. We train our model for 100 epochs with `adam` optimizer and mean absolute error (MAE) as loss function.

3 Results

Figure 3 shows the results for the 'age' parameter, by comparing the output versus the input for the test sample. We show the predictions for all 'mock observed magnitude' bins (i.e., synthetic photometry with different levels of noise). The results are excellent, with almost no bias and small standard deviation (μ=-0.01, σ=0.55). There is a flattening of the predicted ages for log(Age) $>$ 9.5 yr, as expected due to the little evolution of galaxy SEDs after 10 Gyr (which make them almost indistinguishable).

There is a strong dependence of the NN predictions with the level of noise added to the SSP models, as can be seen in Figure 4, where we show the bias, scatter and fraction of outliers for age, metallicity and E(B-V) estimations at different 'observed magnitude' bins. For galaxies with $\mathrm{mag}_i < 16$, the bias and scatter in age are (μ=-0.01, σ=0.15), while they increase up to (μ=-0.10, σ=0.55) for the faintest galaxies. Comparing the different parameters, E(B-V) seems to be very well recovered, even for the faintest objects, while the metallicity values show the largest bias, scatter and fraction of outliers in all magnitude bins. In general, we find excellent results up to $\mathrm{mag}_i < 19$, comparable (or even surpassing) those obtained with SED-fitting (e.g., [7])

4 Future work

The final aim of this project is to apply the NN models to real J-PAS observations to produce a Value Added Catalogue of SP. While the results presented here are very promising, we first

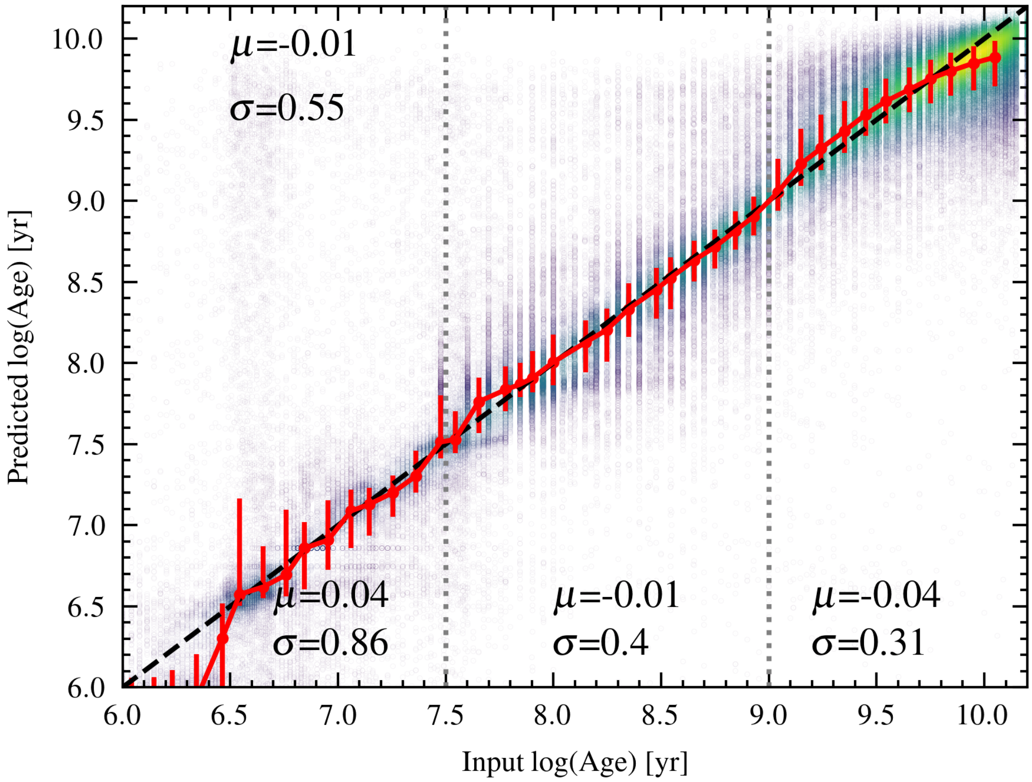

Figure 3: Comparison of the input and predicted age values for the test sample. Light blue symbols represent individual galaxies in low populated regions, while density contours are used for the crowded ones. The red lines are the median values at different age bins, while the error bars show the scatter. μ and σ show the bias and scatter for all the sample (upper text), and for 3 age bins (defined by the vertical lines).

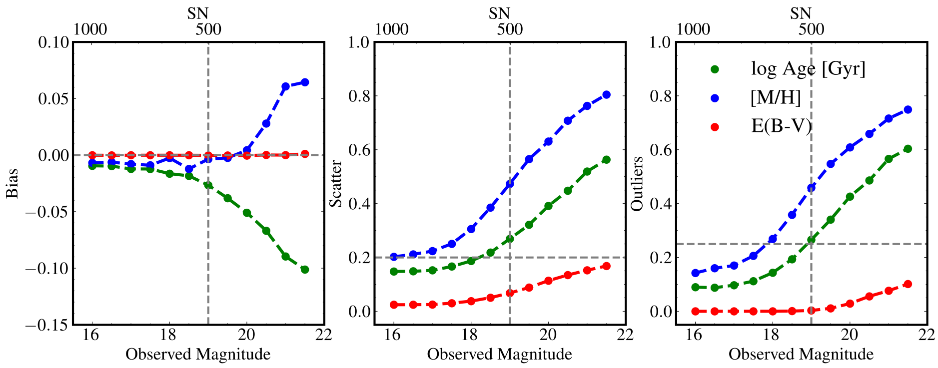

Figure 4: Bias, scatter and fraction of outliers at different 'mock observed magnitudes' – converted to SN = 1.08/magnitude_error, where magnitude_error is the average error in the i-band for each magnitude bin – for the age (green), metallicity (blue) and E(B-V) predictions.

need to test the NN-model in simulated galaxies at different redshifts and with more realistic star formation histories and/or emission lines.

Acknowledgments

HDS acknowledges the financial support from the Spanish Ministry of Science and Innovation and the European Union - NextGenerationEU through the Recovery and Resilience Facility project ICTS-MRR-2021-03-CEFCA and financial support provided by the Governments of Spain and Aragón through their general budgets and the Fondo de Inversiones de Teruel and The Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033 y FEDER, Una manera de hacer Europa) with grant PID2022-138896NA-C54, as well as financial support by RyC2022-030469-I grant, funded by MCIN/AEI/10.13039/501100011033 and FSE+. PC acknowledges support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) process number 2021/08813-7. The NN was trained using the computer resources at Artemisa, funded by the European Union ERDF and Comunitat Valenciana as well as the technical support provided by the Instituto de Física Corpuscular, IFIC (CSIC-UV).

References

[1] Arnouts, S., Ilbert, O. 2011, astrophysics Source Code Library, record ascl:1108.009

[2] Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints, arXiv:1403.5237

[3] Bonoli, S., Marín-Franch, A., Varela, J., et al. 2021, A&A, 653, A31

[4] Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103

[5] González Delgado, R. M., García-Benito, R., Pérez, E., et al. 2015, A&A, 581, A103

[6] Magris, C. G., Mateu, P. J., Mateu, C., et al. 2015, PASP, 127, 16

[7] Mejía-Narváez, A., Bruzual, G., Magris, C. G., et al. 2017, MNRAS, 471, 4722

[8] Plat, A., Charlot, S., Bruzual, G., et al. 2019, MNRAS, 490, 978

[9] Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, MNRAS, 463, 3409

[10] Verro, K., Trager, S. C., Peletier, R. F., et al. 2022, A&A, 661, A50