*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*

# CHEmical Survey analysis System (CHESS) - Exploring the Milky Way metallicity gradient using open clusters

**Martínez Fernández, J. E.[1], Özdemir, S.[1], and Smiljanic, R.[1]**

[1] Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, ul. Bartycka 18, 00-716, Warsaw, Poland

## Abstract

In recent years, several surveys have supplied vital data for studies on star clusters. These surveys, completed or in progress, offer photometric, astrometric, and spectroscopic details for numerous stars in the Milky Way and nearby galaxies. The Gaia mission is of special importance because it aids in determining cluster membership, discovering new clusters, and providing a large set of stellar atmospheric parameters. As open clusters can be found in a wide range of Galactocentric distances, they are valuable tracers of the chemical enrichment history of the disk of the Milky Way. This work presents preliminary findings from our ongoing investigation into the chemical abundances within the Galactic disk. This analysis is part of the development of a new spectral analysis pipeline called CHESS (CHEmical Survey analysis System). CHESS aims to automate the analysis of large stellar spectral datasets and provide high-quality chemical abundances for as many elements as possible, which is key to the study of Galactic archaeology. One of the first steps of CHESS is to use unsupervised machine learning algorithms directly on the spectra to perform what we call a similarity analysis. This analysis aims to identify stars with similar stellar parameters without the need to conduct a preliminary radiative transfer analysis. At the same time, this similarity analysis can be used as a consistency check for the quality of atmospheric parameters in large catalogues. In this work, we discuss the similarity analysis applied to open cluster stars to investigate and validate the radial metallicity gradient of the Galaxy given by external catalogues, in comparison with previous results available in the literature.

## 1 Introduction

The study of open clusters (OCs) is essential to understand the formation and evolution of stars in the Galactic disk [6, 13]. Spread throughout the Galactic disk, the stars within each OC were formed from single molecular clouds [12], making OCs useful tracers of the chemical composition of their parent molecular cloud at the time of their formation [23]. Open clusters also cover a wide range of ages, and their ages can be calculated with greater precision compared to field stars. However, most OCs are relatively young as they tend to be disrupted over time and mix with the field.

One of the areas where precise abundances for OC stars are valuable is the study of the Galactic radial metallicity gradient and its evolution over time. The shape and evolution in the radial distribution of chemical abundances are useful to test models of Galactic formation and evolution [4]. Various objects can be used to study the radial gradient, but OCs are particularly important due to their precisely determined ages, broad age range [3], and high precision with which their abundances can be measured.

In this context, our team is developing a pipeline called CHEmical Survey Analysis System (CHESS) to analyse Ultraviolet and Visual Echelle Spectrograph (UVES) [5] archival spectra using the differential analysis technique [14]. The goal is to determine both stellar parameters and elemental abundances of a target with respect to a well-studied reference star. This approach minimises systematic errors, while achieving high precision ($< 0.05$ dex) and accuracy at the same time. However, this technique is only applicable to groups of stars with similar parameters (effective temperature, surface gravity, and metallicity). To identify the groups in which we can apply this technique, we developed a method that clusters stars based solely on the similarity of their spectra. Additionally, in the course of testing the method, we applied this initial stage of the pipeline to investigate potential biases in large catalogues of parameters with respect to the stellar parameters of a set of reference stars. This correction helps refine the original metallicity gradient obtained from such catalogues.

## 2 Data sample

The spectroscopic data used in this work are publicly accessible through the Science Archive Facility of the European Southern Observatory (ESO) [20]. The spectra we used were obtained with the UVES instrument at the Unit Telescope 2 of the Very Large Telescope (VLT), Paranal Observatory, Chile. UVES is a high-resolution spectrograph that covers the ultraviolet and visual range from 300 to 1100 nm. The analysis we discuss in this work focuses on the spectra of F,G, and K-type stars in the field of OCs, but the ultimate goal is to re-analyse the entire UVES archive.

The complete sample we used in this work consists of two parts: the reference stars and the OC stars. For the OC stars, we conducted a cone search with a 20 arcmin diameter centred on the coordinates of each OC [11]. For our initial purposes, cluster membership was not required. The only additional criteria applied were to select spectra with resolution $R \geq 30\,000$ and to exclude spectra with a signal-to-noise ratio (S/N) lower than 10. This selection provided 8323 spectra from stars in the field of 371 open clusters. The UVES spectra of the reference stars were also downloaded from the ESO Archive. In total, following the same criteria as before, we obtained 1018 spectra of 267 reference stars with different signal-to-noise ratios and wavelength coverages.

The reference star lists used come from the following catalogs: **Titans I**. Metal-poor dwarfs and subgiants [9]; **Titans II**. Metal-poor giant stars [10]; Gaia "**Golden sample**" [7]; **Gaia FGK** benchmark Stars [22]; **Gaia-ESO** survey K2 stars [25]; **Gaia-ESO** survey hot benchmarks (O to B-type stars) [17].

## 3 Methodology

The spectra downloaded from the ESO Archive vary in properties as a result of different observational setups of the UVES instrument. To be able to apply the method to find similar stars based on the spectra, these spectra had to be corrected and standardised before this part of the analysis. This involves homogenising them to have the same spectral resolution and spectral binning, with the resolution set to the minimum in our sample, which is 30 000. We determined and applied the radial velocity correction using a grid of template spectra that covers our parameter space. Additionally, for the continuum normalisation, we used a software called "SUPPNet"[21] to perform this step automatically, due to the impossibility of manually performing this task in this amount of data.

The homogenised data are then prepared for the first analysis of the CHESS pipeline. This stage, which we call similarity analysis, involves identifying groups of similar spectra that match our selected sample of reference stars. This grouping enables us to find similar stars before conducting the full spectroscopic analysis. The method employs the unsupervised machine learning algorithm known as t-SNE (t-distributed Stochastic Neighbor Embedding [24]) for dimensional reduction. Here, the high-dimensional data consist of the homogenised spectra that we aim to analyse. We apply t-SNE directly to these spectra, treating each wavelength value as a coordinate in a high-dimensional space. The algorithm then generates a lower-dimensional 2D "projection map", where each point represents an individual spectrum. Spectra that appear in clusters on this map are those with similar characteristics.

However, the wavelength coverage of the spectra varies drastically in our sample, depending on the configuration used for the observation. For the initial development of CHESS, we decided to limit our analysis to wavelengths between 400 and 700 nm. Within this range, we defined specific subregions in which we applied the similarity analysis algorithm. These subregions are carefully chosen to maximise the number of spectra they cover while minimising their number and ensuring that they contain features sensitive to various stellar parameters. In this work, we show results based on what we refer to as region number 2: **R2**: 484.81 - 490.44 nm, with 970 pixel dimensions.

The primary feature of the t-SNE projection map is its ability to cluster spectra with similar characteristics. To validate these projection maps, we colour-coded them based on stellar parameter values from various external catalogues. Across all selected regions, regardless of the catalogue used, we observe a clear separation between giant and dwarf stars, as shown in Fig. 1. We developed a method for defining regions of similar stars in the projection map (explained in Martínez Fernández, J. E. in preparation). This method identifies stars near each reference object and was calibrated using both parameters from external catalogues and the differences in the spectra. Independently of the external catalogue we used, the spectra identified as similar to the reference object remain consistently similar. The catalogues we use to validate our results are the following:

- Parameters for 175 million stars from Gaia DR3 using the XGBoost algorithm [2];
- StarHorse 2 catalogue with stellar parameters, distances, and extinction for 362 million stars [1];
- The *Gaia*-ESO spectroscopic survey sample of almost 7 000 stars observed with UVES [8, 19];
- Parameters for 220 million sources from a data-driven analysis of the Gaia XP spectra [26].

## 4 Results and discussion

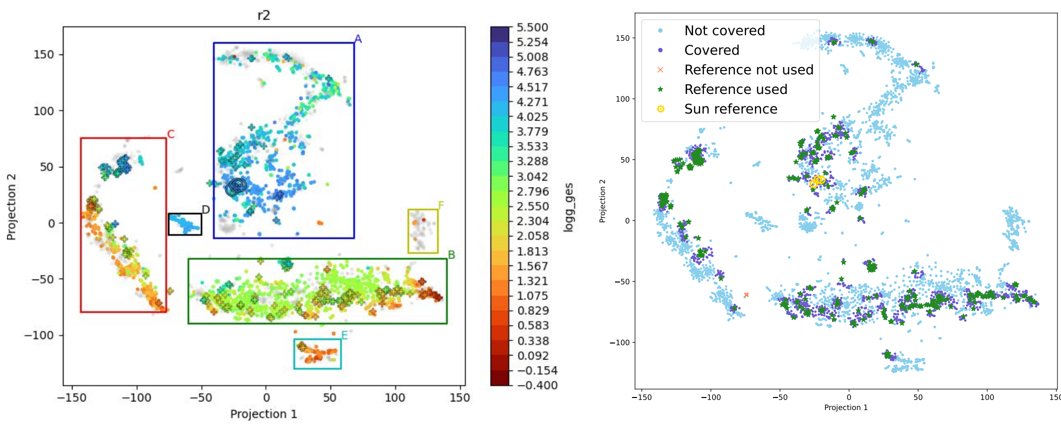


Figure 1: *Left*: t-SNE projection map of the region R2 colour-coded with the surface gravity values from the Gaia-ESO catalogue. *Right*: Restricted group selection for the bias correction.

In the projection map, two primary groups are distinctly separated, which correspond to dwarf and giant stars, along with a few other smaller but well-defined groups. In Fig. 1, left panel, they are labelled as: **Section A**: Dwarfs are found in this group and they display a temperature gradient increasing from the bottom to the top of the section, where the hottest O- and B-type stars are found; **Section B**: Giants of higher metallicity with a very flat gradient in the stellar parameters; **Section C**: Giants of low metallicity are grouped in this region; **Section D**: These are solar-type stars; however, they are separated due to reduction problems creating artefacts in the spectra; **Section E**: This region has not yet been fully understood, but seems to contain mostly very bright giants; **Section F**: This region groups variable stars, in particular Cepheids.

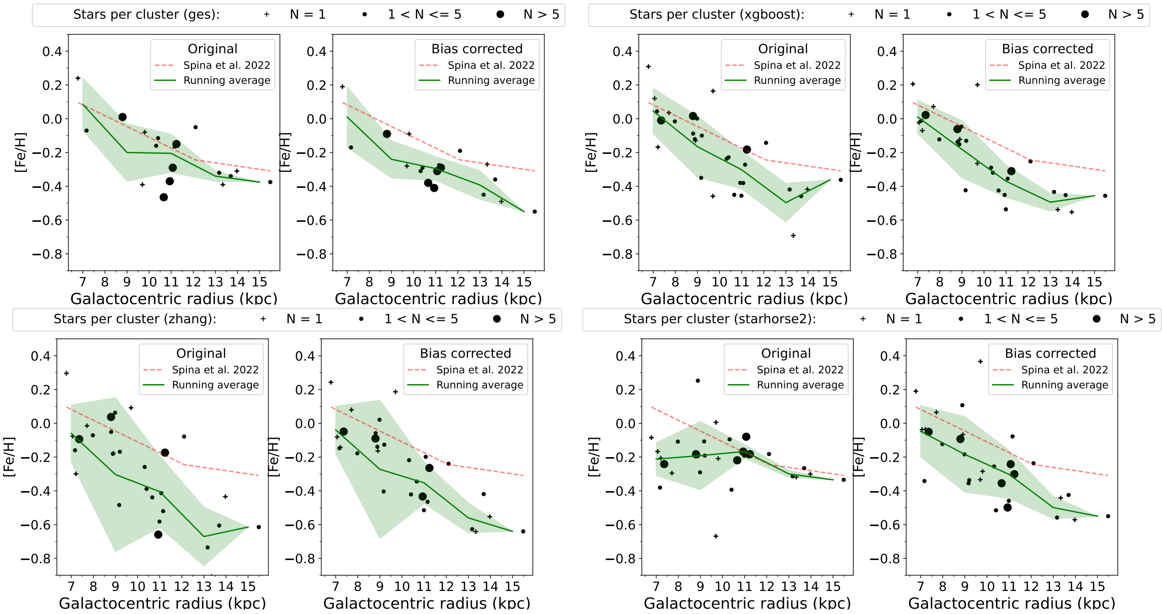

Figure 2: Bias corrections in the different catalogues. *Top-Left*: GES catalogue [8, 19]. *Top-Right*: XGBoost catalogue [2]. *Bottom-Left*: Zhang catalogue [26]. *Bottom-Right*: StarHorse 2 catalogue [1].

Using the projection map, we can select groups of stars that exhibit spectra very similar to the ones of the reference stars. This is seen in the right plot of Fig. 1, where the selected similar stars assigned to each reference object are highlighted. We can then compare the median values of the stellar parameters in each group, using the values from the catalogues mentioned in Section 3, with the values of their reference object. Assuming the values of the reference object to be accurate, with this comparison, we can estimate the typical biases affecting the parameters of the surrounding stars. We can then use the bias-corrected metallicities, to estimate the metallicities of the open clusters, based on a scale defined by our set of reference stars. For this step, we crossmatched our sample with the catalogue of [11], that provides membership probabilities and distances for OCs.

As a result, we can generate Fig. 2, which illustrates the main changes in the metallicity gradient across all catalogues of stellar parameters after the bias correction. The dispersion decreases with, in some cases, a significant improvement observed in the metallicities, resulting in a clearer gradient emerging after correction. Furthermore, we find that the metallicity gradients before and after correction appear to be slightly more metal-poor compared to the gradient obtained in another study [23]. This difference is currently being investigated.

## Acknowledgments

J. E. Martínez Fernández, S. Özdemir, and R. Smiljanic acknowledge support from the National Science Centre, Poland, project 2019/34/E/ST9/00133. Based on data obtained from the ESO Science

Archive Facility with DOI: `https://doi.eso.org/10.18727/archive/50` [18]. This research has made use of: NASA's Astrophysics Data System Bibliographic Services; the SIMBAD database, operated at CDS, Strasbourg, France; the VizieR catalogue access tool, CDS, Strasbourg, France [15]. The original description of the VizieR service was published in [16].

## References

[1] Anders, F., Khalatyan, A., Queiroz, A. B. A., et al. 2022, A&A, 658, A91

[2] Andrae, R., Rix, H.-W., & Chandra, V. 2023, ApJS, 267, 8

[3] Bonatto et al. 2006, A&A, 446, 121

[4] Chiappini et al. 2001, ApJ, 554, 1044

[5] Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4008, Optical and IR Telescope Instrumentation and Detectors, ed. M. Iye & A. F. Moorwood, 534–545

[6] Friel et al. 2013, in Planets, Stars and Stellar Systems Vol. 5, Ed.: Oswalt, T. D.; Gilmore, G.

[7] Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023c, A&A, 674, A1

[8] Gilmore, G., Randich, S., Worley, C. C., et al. 2022, A&A, 666, A120

[9] Giribaldi, R. E., da Silva, A. R., Smiljanic, R., & Cornejo Espinoza 2021, A&A, 650, A194

[10] Giribaldi, R. E., Van Eck, S., Merle, T. 2023, A&A, 679, A110

[11] Hunt, E. L. & Reffert, S. 2024, A&A, 686, A42

[12] Krause et al. 2020, Space Sci. Rev., 216, 64

[13] Krumholz et al. 2019, ARA&A, 57, 227

[14] Nissen and Gustafsson 2018, A&ARv, 26, 6

[15] Ochsenbein, F. 1996, The VizieR database of astronomical catalogues

[16] Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23

[17] Pancino, E., Lardo, C., Altavilla, G., et al. 2017, A&A, 598, A5

[18] Percheron, I. & Hanuschik, R. 2020, UVES reduced data obtained by standard 980 ESO pipeline processing

[19] Randich, S., Gilmore, G., Magrini, L., et al. 2022, A&A, 666, A121

[20] Romaniello, M., Arnaboldi, M., Barbieri, M., et al. 2023, The Messenger, 191, 29

[21] Różański, T., Niemczura, E., Lemiesz, J., Posiłek, N., & Różański, P. 2022, A&A, 659, A199

[22] Soubiran, C., Creevey, O. L., Lagarde, N., et al. 2024, A&A, 682, A145

[23] Spina et al. 2022a, Univ., 8, 87

[24] van der Maaten, L. & Hinton, G. 2008, Journal of Machine Learning Research, 9, 2579

[25] Worley, C. C., Jofré, P., Rendle, B., et al. 2020, A&A, 643, A83

[26] Zhang, X., Green, G. M., & Rix, H.-W. 2023, MNRAS, 524, 1855