SF2A 2023
M. N'Diaye, A. Siebert, N. Lagarde, O. Venot, K. Baillié, M. Béthermin, E. Lagadec, J. Malzac, J. Richard (eds)

THE SPECTROSCOPIC DATA OF THE GAIA DR3 CATALOG: PERFORMANCE AND USE

P. A. Palicio[1]

Abstract. Galactic Archaeology has experienced a great advance in the last decade thanks to the development of large surveys like APOGEE, LAMOST and GALAH. The *Gaia* mission (ESA) has contributed to this revolution by providing precise astrometric and photometric data for more than one billion stars, radial velocities for the bright subset of them and, since *Gaia* DR3, the stellar parameters and chemical abundances of 5.6 million sources observed by the on-board RVS spectrograph. The extraction of the atmospheric parameters and composition for such an unprecedented large volume of stars has been possible thanks to the implementation of the GSP-Spec module in the *Gaia* pipeline. In this work, we describe the performance of GSP-Spec and present its output to the community. We demonstrate the quality of the GSP-Spec data and provide a guide of best practices for its correct usage, as well as illustrate some scientific applications.

Keywords: stars: fundamental parameters – stars: abundances – methods: data analysis – Galaxy: stellar content

1 Introduction

Gaia is the most advanced astrometric mission to date (Perryman et al. 2001; Gaia Collaboration et al. 2016). Since its launch in December 2013, it has continued the legacy of Hipparcos (Perryman & Hassan 1989) by providing positions, parallaxes and proper motions with errors of few microns. In addition to the astrometry, *Gaia* provides the line-of-sight velocities of a subset of bright stars whose spectra have been collected by the on-board Radial Velocity Spectrograph (Katz et al. 2004, RVS), which operates in the wavelength domain of the near infra-red calcium triplet (846-870 nm) with a resolution of 11,500 (Cropper et al. 2018). This full kinematic information for *Gaia* sources was included, for first time, as part of the second data release (Gaia Collaboration et al. 2018, *Gaia* DR2), in which the astrometry and line-of-sight velocities were extended and updated in the subsequent *Gaia* eDR3 (Gaia Collaboration et al. 2021) and *Gaia* DR3 (Gaia Collaboration et al. 2023b) data releases, respectively. On the contrary, the stellar parameters, chemical abundances and other valuable information extracted from the RVS spectra were not publicly available until the latest *Gaia* data release. Thus, *Gaia* DR3 constitutes the first data-set containing the results of the GSP-Spec module (Recio-Blanco et al. 2023) –one of the eighth Data Processing and Analysis Consortium (DPAC) collaboration unit (CU8) modules – devoted to the analysis of the RVS combined spectra and currently incorporated to the *Gaia* pipeline. Since *Gaia* has been continuously observing the sky for years in more stable conditions compared to ground-base telescopes, it allows a more homogeneous treatment of the spectra and modelling of the systematics. This results in the largest stellar spectroscopic characterisation catalog ever published, totalling almost 5.6 million *Gaia* sources. From the scientific point of view, this characterisation is key for understanding the structure of the Milky Way. In particular, the chemical abundances shed light on Galaxy formation and evolution through the composition of the individual stars, since they are synthesised in the stellar interior and pollute the inter-stellar medium (ISM) through different channels. The relative contribution of these channels leads to a characteristic footprint that preserves the information of the birthplace of the stars. For example, it is well known the thick disc stars show higher [α/Fe] abundances compared to its thinner counterpart (Lee et al. 2011), while stars accreted from satellites tend to be α-poor compared to these born in the Milky Way

[1] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France

© Société Française d'Astronomie et d'Astrophysique (SF2A) 2023

with the same metallicity (Nissen & Schuster 2010). In this work, we describe the performance of the GSP-Spec module and summarise its main products, with special emphasis on the treatment of the output data for the proper usage on scientific applications. Finally, we illustrate some novel results already published made with GSP-Spec data.

2 The GSP-Spec module

2.1 The GSP-Spec architecture

As mentioned in the previous section, GSP-Spec is one of the modules included in the *Gaia* pipeline for the analysis of the RVS spectra. In contrast to other modules, it uses no additional input information (i.e., parallaxes, photometry, extinction, etc.) for performing the analysis. GSP-Spec encompasses two distinct workflows with different architecture, MatisseGauguin and the supervised machine learning algorithm ANN. The former incorporates various sub-modules to tackle multiple tasks, while the latter comprises three layers to analyse the spectra: an input layer with 800 neurons, an intermediate layer containing a variable number of neurons ranging from 50 to 100, and an output layer of four neurons.

While both workflows yield the four basic atmospheric parameters (effective temperature $T_{\rm eff}$, surface gravity logg, global metallicity [M/H] and global alpha abundance [α/Fe]) and their respective quality flags as output, only MatisseGauguin provides additional information concerning the chemical composition and the inter-stellar absorption through the estimates of thirteen chemical abundances, the cyanogen equivalent width and two diffuse inter-stellar bands (DIBs).

The structure of the MatisseGauguin module is summarised in Fig. 4 of Recio-Blanco et al. (2023). Given an input RVS spectra, a decision tree called DEGAS gets an initial estimation of the atmospheric parameters. Then, the Matisse algorithm updates this seed by performing a multi-regression method until it converges. The output parameters of Matisse are refined by the Gauguin module, which implements a Gauss-Newton algorithm that computes the dimension in the parameter space that minimises the difference between the observed and the synthetic spectra. If needed, this process is repeated to get a satisfactory subtraction of the continuum, especially for cool stars, by performing a sigma-clipping procedure based on the comparison of the observed and the interpolated synthetic spectra. The quality of this fitting is quantified by the $\log \chi^2$ parameter, included in the final output. Previously to its ingestion in the pipeline, the Matisse algorithm had been trained with the synthetic grid to optimise the b-functions, required for determining the weight of each pixel during the application stage.

Once the stellar parameters are determined, the Gauguin algorithm is re-implemented for the extraction of the chemical abundances from individual lines, adding an extra normalisation routine focused on the spectral region near the considered line (Santos-Peral et al. 2020). The algorithm stops when the iteration step is smaller than one hundredth the step of the abundance grid. Finally the DIB and the cyanogen parameters are determined by an iterative Gaussian profile fitting algorithm.

In order to account for the flux error of the input spectra, the whole procedure described above is repeated fifty times using different random realisations of the input spectra, whose distortion depends on their signal-to-noise ratio. We consider the median of the outputs of all of these realisations as estimators, while the 16 and 84th percentiles are associated with the uncertainty. For those elements with multiple lines, we combine the individual abundances from each line by computing their weighted mean, in which the weight is determined by the inverse of their uncertainties. A more detailed explanation involving the rejection of bad estimates can be found in Section 6.8 of Recio-Blanco et al. (2023).

2.2 The synthetic spectral grids

As mentioned in the previous section, MatisseGauguin uses grids of synthetic spectra as reference for determining the stellar parameters and the chemical abundances. These grids were generated with the version 19.2 of the TURBOSPECTRUM code (Plez 2012) assuming the MARCS atmospherics models (Gustafsson et al. 2008) and the list of atomic and molecular lines of Contursi et al. (2021). As Fig. 1 illustrates, the four-dimensional grid used in the determination of the atmospheric parameters contains more than fifty thousand spectra spanning the range 2750 K $\leq T_{\rm eff} \leq$ 8000 K, -0.5 dex $\leq$ logg $\leq$ 5.5 dex, -0.5 dex $\leq$ [M/H] $\leq$ 5.5 dex, -0.4 dex $\leq [\alpha/{\rm Fe}] \leq$ 0.8 dex with steps of $\Delta T_{\rm eff}$ = 250K, Δlogg = 0.5 dex, Δ[M/H] = 0.25 dex and $\Delta[\alpha/{\rm Fe}]$ = 0.1 dex. Once a synthetic spectrum is generated, it is degraded to the RVS resolution (11,500) and re-sampled using a wavelength step of 0.03 nm within 876.0 and 870. nm to mimic the observed spectra.

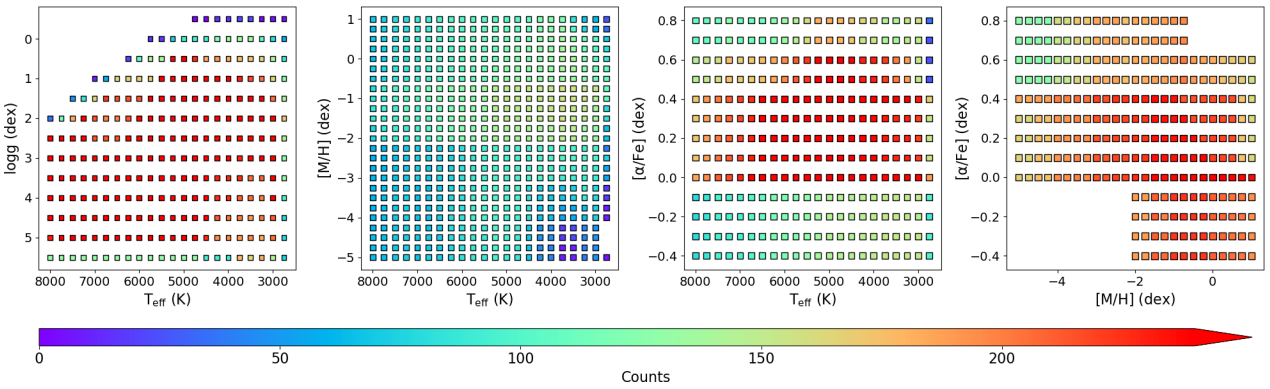

Fig. 1. Distribution of the nodes of the four-dimensional synthetic grid in different projections. **Left:** Number of synthetic spectra simulated in the logg vs. $T_{\rm eff}$ plane. **Middle-left:** Number of synthetic spectra per node in the [M/H] vs. $T_{\rm eff}$ plane. **Middle-right:** Distribution of synthetic spectra in the [α/Fe] vs. $T_{\rm eff}$ plane. **Right:** [α/Fe] vs. [M/H] diagram with the number of synthetic spectra in this projection.

Regarding the abundance determination, a five-dimensional grid is computed for each chemical element by considering 9-10 possible values of the individual abundance, generally ranging from -2.0 dex to $+2.0$ dex with the exception of the α elements, whose range of values depends on the metallicity to mimic that of the Milky Way. Compared to the four-dimensional grid described above, these five-dimensional grids are more restricted in the atmospheric parameter space, excluding the cool stars and the very metal poor (¡-3 dex) regime among other variations. However, the synthetic grids associated with the Fe, Ti and Ca abundances contain the nodes at such low metallicites. In total, the number of synthetic spectra for each individual element ranges from $\sim 470,000$ to almost $591,000$ for the mentioned Fe, Ti and Ca cases.

2.3 The line list

GSP-Spec makes use of a compilation of atomic and molecular lines summarised in Contursi et al. (2021). After an exhaustive evaluation based on the dependence of these lines with the abundance, the existence of blends with other chemical species, and the consistency with lines of the same element, we reduce the initial list to 49 lines of 12 different elements: N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce and Nd.

For each line in the final list, we consider a wavelength domain where the flux variation correlates with the abundance in the simulated spectra of Arcturus, Procyon, and solar-like stars. Similarly, we define a broader wavelength domain for the extra normalisation (referred as *second normalisation* in Recio-Blanco et al. 2023) associated with each line (see Section 2.1). For consecutive lines of the same elements, in which common abundance and/or second normalisation windows are applicable, we tested whether the inclusion in a shared wavelength domain improves the results or not. Furthermore, in the specific case of the calcium triplet, abundances are determined by exploring both sides of the line wings while excluding the core to avoid NLTE effects. This results in an 'effective" number of lines of 33. After testing multiple combinations of abundance and second normalisation windows, the final limits for these wavelength domains are summarised in Table B.1 of Recio-Blanco et al. (2023).

3 GSP-Spec data availability

The running of the GSP-Spec module was performed at the Data Processing Centre CNES (DPCC) in Toulouse (France), where it analysed almost seven million sources with fifty repetitions each due to the error propagation (see Section 2.1). It required more than one hundred thousand computing hours spread on 2,100 cores, resulting in a total execution time of one hundred and fifty hours. Among the total volume analysed, approximately 1.3 million sources were discarded during the validation of the processing stage, leading to the released sample of 5.6 million stars.

Similarly to other *Gaia* products, the tables with the GSP-Spec output have been ingested into the *Gaia*

archive*. More precisely, the results of MatisseGauguin are located in the GAIADR3.ASTROPHYSICAL_PARAMETERS table while these of ANN can be found in the supplementary GAIADR3.ASTROPHYSICAL_PARAMETERS_SUPP table, where the parameters associated with each module can be identified by the subfix of the field names. As usual, these data is accessible by performing an ADQL query on the *Gaia* archive web-site (see Appendix B of Gaia Collaboration et al. (2023a) for examples of queries with scientific purposes).

4 Calibration

The comparison of the GSP-Spec results with these from large ground-based surveys like APOGEE-DR17 (Abdurro'uf et al. 2022), GALAH-DR3 (Buder et al. 2021) and RAVE-DR6 (Steinmetz et al. 2020) reveals some trends with the GSP-Spec parameters that must be corrected. The giants analysed by GSP-Spec show a significant bias in logg compared to that reported in literature, while their global metallicity [M/H] seems to be under-estimated. On the contrary, dwarf stars show good agreement in surface density but an excess of global metallicity. No significant bias has been observed in effective temperature. In order to account for the mentioned discrepancies, we propose a polynomial correction based on the measured value of logg of the form

$$parameter_{calibrated} = parameter_{uncalibrated} + \sum_{n=0}^{N} p_n (logg_{uncalibrated})^n \tag{4.1}$$

where N is the degree of the polynomia and the coefficients p_n can be found in Table 3 of Recio-Blanco et al. (2023), where an additional calibration for the metallicies of stars in open clusters has been provided. Similarly, analogous corrections have been suggested for the the global $[\alpha/Fe]$ and the individual chemical abundances, with the exception of [Zr/Fe], [Nd/Fe] due to low number statistics, and of [Ce/Fe] because of an offset in the solar value. These calibrations are, however, restricted to specific ranges of logg. For the particular case of the [α/Fe], [Ca/Fe] and [S/Fe] estimates, other corrections based on the effective temperature $T_{\rm eff}$ are provided.

Since the corrections proposed above were obtained for a selection of stars that spans a wide range in logg or $T_{\rm eff}$, other calibrations for more specific samples in the parameter space might improve the agreement with the literature (for example, see the case of the [Mg/Fe] abundance of Spitoni et al. 2023). In any case, we recall the calibration of GSP-Spec parameters is mandatory for their correct usage. Additional resources for the calibration of GSP-Spec parameters can be found in the *Gaia* DR3 web-page†.

5 Flagging system

Both the MatisseGauguin and the ANN workflows return a string of integer numbers, called *flags*, encoding the quality of their respective outputs. In total, a system of 41 flags is adopted for the GSP-Spec output, in which only the first twelve are common to both workflows (the remaining 29 are exclusive to MatisseGauguin). These 41 flags can be group into three families concerning the spectral features (first 13 flags), the chemical abundance determination (the next 26 flags) and the cyanogen/DIB fittings (last 2 flags). In general, flag values range from "0" for the high quality parameters to "9" for the bad estimates (sometimes removed from the table). Since their definition is quite technical, we provide here a brief description of the first family and refer to Section 8 in Recio-Blanco et al. (2023) for a more detailed explanation of the whole set.

The first three flags *vbroadT*, *vbroadG* and *vbroadM* indicate the bias induced by the rotational velocity of the star on the atmospheric parameters $T_{\rm eff}$, logg and [M/H], respectively. We evaluated these biases by simulating the broadening effects on a set of synthetic spectra, which is used as input of GSP-Spec. As a result, this artificial rotation leads to values of $T_{\rm eff}$, logg and [M/H] different from the nominal ones (see Figure 2 for an example). Depending on the level of such distortion, a flag from "0" to "2" is assigned to the stellar parameter. In the same way, the biases induced by the errors in the line-of-sight velocity correction were explored following a similar procedure but using shifted synthetic spectra instead of broadened ones. The quantification of these biases are summarised in the next three flags *vradT*, *vradG* and *voadM* for the $T_{\rm eff}$, logg and [M/H], respectively.

The *fluxNoise* flag is associated with the bias induced by the flux error of the input spectra, which was explored by observing the dispersion of the four atmospheric parameters $T_{\rm eff}$, logg, [M/H] and [α/Fe] along the fifty random realisations mentioned in Section 2.1. The *extrapol* flags determines the quality of the solution

*https://gea.esac.esa.int/archive/

†https://www.cosmos.esa.int/web/gaia/dr3-gspspec-metallicity-logg-calibration

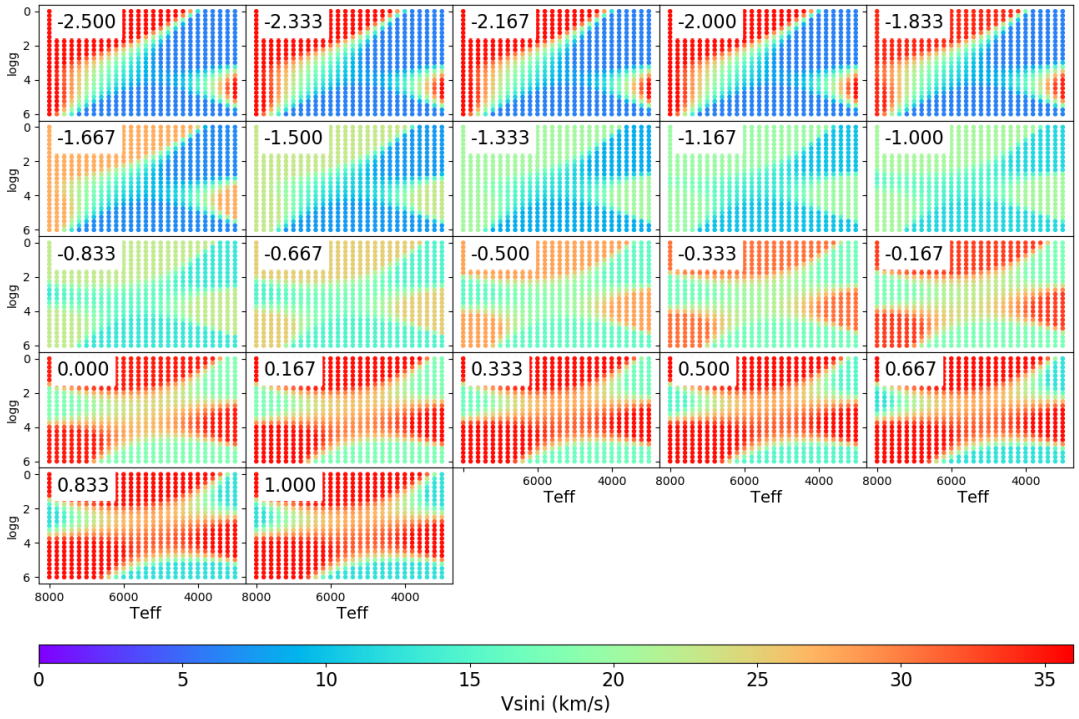

Fig. 2. Limiting rotational velocity values $Vsini$ (colour-code) in the Kiel diagram leading to a *vbroadT* flag equal to zero (i.e., a bias of $\Delta T_{\rm eff} < 250$ K). Each subpanel corresponds to the metallicity $[M/H]$ indicated in the upper left corners.

found by MatisseGauguin or ANN depending on the distance to the borders of the training grids, and it was used to exclude some sources with bad estimations from the *Gaia* archive during the post-processing phase.

There are four flags related to undesired features in the RVS spectra: *negativeFlux* and *nanFlux* informs about the presence of pixels with negative or NaN fluxes, respectively. Similarly, *nullFluxErr* is equal to nine if no flux error is provided by the input spectra, leading to a zero-variance atmospheric parameter estimations during the random realisation stage; while the *emission* flag is nine if the CU6 has detected an emission line in the spectra.

In the cool regimes, where there are many molecular absorption lines, the estimation of the atmospheric parameters is specially difficult, resulting in an unphysical combination of $T_{\rm eff}$ and logg values. For these stars, we assign a non-zero *KMgiantFlag* value and artificially set their effective temperatures and surface gravities to 4250 ± 500 K, 1.5 ± 1 dex, respectively. As for the calibrations (cf. Section 4), the usage of the flags previously introduced is mandatory for the correct application of the GSP-Spec data and avoid misleading results. As an initial recommendation, we suggest selecting the stars with the first thirteen flags equal to zero and adjusting the initial guess as needed.

6 Scientific application of GSP-Spec

Since the publication of the GSP-Spec products with *Gaia* DR3, many groups have been exploiting this information for scientific purposes. For sake of exemplification, we have selected few of them to illustrate the quality of the GSP-Spec parametrization and the multiple possibilities of its application. Gaia Collaboration et al. (2023a) constitutes one of the major compilation of examples of usages of GSP-Spec data. Their analysis of accreted structures identified in the energy vs. angular momentum reveal a chemical pattern consistent with their *ex-situ* origin, showing a deficiency of α-elements compared to the Milky Way stars at the same metallicity (Nissen & Schuster 2010). Similarly, they explored the chemistry of the kinematic and dynamic structures reported in previous works, such as the ridges in the velocity profile (Ramos et al. 2018), the spiral in the Z-V_Z plane (Antoja et al. 2018), and the over-densities in the action space J_R vs. L_Z (Trick et al. 2019; Trick 2022), to find clear correlations between the location of the stars in these structures and their chemical composition. By selecting massive stars in the Kiel diagram, Gaia Collaboration et al. (2023a) observed spatial inhomogeneities

in the Galactic Plane revealing the presence of spiral arms, being the innermost ones more metal rich than the external arms. This connects with the work of Poggio et al. (2022), who analysed the spatial distribution of three populations classified as bright giant stars by GSP-Spec to find an excess of metallicity in these regions where the spiral arms are observed. They also report a varying metallicity gradient with azimuth for their sample of old stars. Analysing the α-element abundances, Spitoni et al. (2023) found an evidence of a third infall of gas that triggered the star formation in the solar vicinity 2.7 Gyr ago, likely related to the pericenter passage of the Sagittarius dwarf galaxy. This model can also explain the trend of the cerium abundances with metallicity studied in Contursi et al. (2023), who found a good agreement between the GSP-Spec [Ce/Fe] estimates and these reported in literature, even though the latter might show some systematics. They also reported radial and vertical gradients for the [Ce/Fe] and [Ce/H] abundances.

7 Conclusions

Gaia DR3 constitutes the largest all-sky chemical survey ever published, with atmospheric parameters and chemical abundances for nearly 5.6 million sources. This was possible thanks to the implementation of the GSP-Spec module in the *Gaia* pipeline, which automatised the analysis of the combined, velocity-corrected spectra given as input from the CU6. As a result, it provided very competitive data compared to large ground-base spectroscopic surveys, although calibration and filtering are required for its correct application. A wide variety of scientific results has already been published using the GSP-Spec catalog, including the evidence of a recent gas infall, the metallicity excess in the spiral arm regions, as well as the gradients in the distribution of heavy elements among others .

I acknowledge the invitation from the organisation of the "Archéologie Galactique avec *Gaia*/DR3: un an après" session, as part of the French Astronomy and Astrophysics Society (SF2A) meeting held in Strasbourg (France). I thank the financial support from the Centre national d'études spatiales (CNES). This work presents results from the European Space Agency (ESA) space mission Gaia. Gaia data are being processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the Gaia MultiLateral Agreement (MLA).

References

Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, ApJS, 259, 35
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Nature, 561, 360
Buder, S., Sharma, S., Kos, J., et al. 2021, MNRAS, 506, 150
Contursi, G., de Laverny, P., Recio-Blanco, A., & Palicio, P. A. 2021, A&A, 654, A130
Contursi, G., de Laverny, P., Recio-Blanco, A., et al. 2023, A&A, 670, A106
Cropper, M., Katz, D., Sartoretti, P., et al. 2018, A&A, 616, A5
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Recio-Blanco, A., Kordopatis, G., et al. 2023a, A&A, 674, A38
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023b, A&A, 674, A1
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Katz, D., Munari, U., Cropper, M., et al. 2004, MNRAS, 354, 1223
Lee, Y. S., Beers, T. C., An, D., et al. 2011, ApJ, 738, 187
Nissen, P. E. & Schuster, W. J. 2010, A&A, 511, L10
Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
Perryman, M. A. C. & Hassan, H. 1989, ESA Bulletin, 58, 77
Plez, B. 2012, Turbospectrum: Code for spectral synthesis, Astrophysics Source Code Library
Poggio, E., Recio-Blanco, A., Palicio, P. A., et al. 2022, A&A, 666, L4
Ramos, P., Antoja, T., & Figueras, F. 2018, A&A, 619, A72
Recio-Blanco, A., de Laverny, P., Palicio, P. A., et al. 2023, A&A, 674, A29
Santos-Peral, P., Recio-Blanco, A., de Laverny, P., Fernández-Alvar, E., & Ordenovic, C. 2020, A&A, 639, A140
Spitoni, E., Recio-Blanco, A., de Laverny, P., et al. 2023, A&A, 670, A109
Steinmetz, M., Guiglion, G., McMillan, P. J., et al. 2020, AJ, 160, 83
Trick, W. H. 2022, MNRAS, 509, 844
Trick, W. H., Coronado, J., & Rix, H.-W. 2019, MNRAS, 484, 3291