



Publication Year	2023
Acceptance in OA	2025-02-25T14:52:23Z
Title	Gaia Data Release 3. Analysis of RVS spectra using the General Stellar Parametrised from spectroscopy
Authors	Recio-Blanco, A., de Laverny, P., Palicio, P. A., Kordopatis, G., Álvarez, M. A., Schultheis, M., Contursi, G., Zhao, H., Torralba Elipe, G., Ordenovic, C., Manteiga, M., Dafonte, C., Oreshina-Slezak, I., Bijaoui, A., Frémat, Y., Seabroke, G., Pailler, F., Spitoni, E., POGGIO, Eloisa, Creevey, O. L., Abreu Aramburu, A., Accart, S., Andrae, R., Bailer-Jones, C. A. L., Bellas-Velidis, I., Brouillet, N., BRUGALETTA, Elisa, Burlacu, A., Carballo, R., Casamiquela, L., Chiavassa, A., Cooper, W. J., Dapergolas, A., Delchambre, L., Dharmawardena, T. E., DRIMMEL, Ronald, Edvardsson, B., Fouesneau, M., Garabato, D., García-Lario, P., García-Torres, M., Gavel, A., Gomez, A., González-Santamaría, I., Hatzidimitriou, D., Heiter, U., Jean-Antoine Piccolo, A., Kontizas, M., Korn, A. J., LANZAFAME, Alessandro, Lebreton, Y., Le Fustec, Y., Licata, E. L., Lindstrøm, H. E. P., Livanou, E., Lobel, A., Lorca, A., Magdaleno Romeo, A., Marocco, F., Marshall, D. J., Mary, N., Nicolas, C., Pallas-Quintela, L., Panem, C., Pichon, B., Riclet, F., Robin, C., Rybizki, J., Santoveña, R., Silvelo, A., SMART, Richard Laurence, Sarro, L. M., SORDO, Rosanna, Soubiran, C., Süveges, M., Ulla, A., VALLENARI, Antonella, Zorec, J., Utrilla, E., Bakker, J.
Publisher's version (DOI)	10.1051/0004-6361/202243750
Handle	http://hdl.handle.net/20.500.12386/36218
Journal	ASTRONOMY & ASTROPHYSICS
Volume	674

Gaia Data Release 3

Analysis of RVS spectra using the General Stellar Parametrising from spectroscopy

A. Recio-Blanco^{1,*}, P. de Laverny¹, P. A. Palicio¹, G. Kordopatis¹, M. A. Álvarez², M. Schultheis¹, G. Contursi¹, H. Zhao¹, G. Torralba Elipe², C. Ordenovic¹, M. Manteiga³, C. Dafonte², I. Oreshina-Slezak¹, A. Bijaoui¹, Y. Frémat⁴, G. Seabroke⁵, F. Pailler⁶, E. Spitoni¹, E. Poggio¹, O. L. Creevey¹, A. Abreu Aramburu⁷, S. Accart⁶, R. Andrae⁸, C. A. L. Bailer-Jones⁸, I. Bellas-Velidis⁹, N. Brouillet¹⁰, E. Brugaletta¹¹, A. Burlacu¹², R. Carballo¹³, L. Casamiquela^{10,14}, A. Chiavassa¹, W. J. Cooper^{15,16}, A. Dapergolas⁹, L. Delchambre¹⁷, T. E. Dharmawardena⁸, R. Drimmel¹⁶, B. Edvardsson¹⁸, M. Fouesneau⁸, D. Garabato², P. García-Lario¹⁹, M. García-Torres²⁰, A. Gavel¹⁸, A. Gomez², I. González-Santamaría², D. Hatzidimitriou^{21,9}, U. Heiter¹⁸, A. Jean-Antoine Piccolo⁶, M. Kontizas²¹, A. J. Korn¹⁸, A. C. Lanzafame^{11,22}, Y. Lebreton^{23,24}, Y. Le Fustec¹², E. L. Licata¹⁶, H. E. P. Lindstrøm^{16,25,26}, E. Livanou²¹, A. Lobel⁴, A. Lorca²⁷, A. Magdaleno Romeo¹², F. Marocco²⁸, D. J. Marshall²⁹, N. Mary³⁰, C. Nicolas⁶, L. Pallas-Quintela², C. Panem⁶, B. Pichon¹, F. Riclet⁶, C. Robin³⁰, J. Rybizki⁸, R. Santoveña², A. Silvelo², R. L. Smart¹⁶, L. M. Sarro³¹, R. Sordo³², C. Soubiran¹⁰, M. Süveges³³, A. Ulla³⁴, A. Vallenari³², J. Zorec³⁵, E. Utrilla²⁷, and J. Bakker⁷

(Affiliations can be found after the references)

Received 10 April 2022 / Accepted 17 May 2022

ABSTRACT

Context. The chemo-physical parametrisation of stellar spectra is essential for understanding the nature and evolution of stars and of Galactic stellar populations. A worldwide observational effort from the ground has provided, in one century, an extremely heterogeneous collection of chemical abundances for about two million stars in total, with fragmentary sky coverage.

Aims. This situation is revolutionised by the *Gaia* third data release (DR3), which contains the parametrisation of Radial Velocity Spectrometer (RVS) data performed by the General Stellar Parametrising-spectroscopy, GSP-Spec, module. Here we describe the parametrisation of the first 34 months of *Gaia* RVS observations.

Methods. GSP-Spec estimates the chemo-physical parameters from combined RVS spectra of single stars, without additional inputs from astrometric, photometric, or spectro-photometric BP/RP data. The main analysis workflow described here, MatisseGauguin, is based on projection and optimisation methods and provides the stellar atmospheric parameters; the individual chemical abundances of N, Mg, Si, S, Ca, Ti, Cr, Fe I, Fe II, Ni, Zr, Ce and Nd; the differential equivalent width of a cyanogen line; and the parameters of a diffuse interstellar band (DIB) feature. Another workflow, based on an artificial neural network (ANN) and referred to with the same acronym, provides a second set of atmospheric parameters that are useful for classification control. For both workflows, we implement a detailed quality flag chain considering different error sources.

Results. With about 5.6 million stars, the *Gaia* DR3 GSP-Spec all-sky catalogue is the largest compilation of stellar chemo-physical parameters ever published and the first one from space data. Internal and external biases have been studied taking into account the implemented flags. In some cases, simple calibrations with low degree polynomials are suggested. The homogeneity and quality of the estimated parameters enables chemo-dynamical studies of Galactic stellar populations, interstellar extinction studies from individual spectra, and clear constraints on stellar evolution models. We highly recommend that users adopt the provided quality flags for scientific exploitation.

Conclusions. The *Gaia* DR3 GSP-Spec catalogue is a major step in the scientific exploration of Milky Way stellar populations. It will be followed by increasingly large and higher quality catalogues in future data releases, confirming the *Gaia* promise of a new Galactic vision.

Key words. stars: fundamental parameters – stars: abundances – methods: data analysis – Galaxy: stellar content

1. Introduction

The chemo-physical characterisation of stars is at the core of stellar physics and Galactic studies, but also, through the analysis of unresolved stellar populations, of extragalactic physics. Stellar spectra encode a wealth of information that we have now learned to decrypt. The light emitted by a star is absorbed by the atoms and molecules present in its own atmosphere. This creates spectral absorption lines whose profiles depend on the physical properties of the star and the abundances of the different absorbing chemical species. Our understanding of stellar spectra, used to decode

the enclosed information on the nature of stars, relies on a complex and extensive theoretical framework, including (among others) nuclear, atomic, and molecular physics, stellar atmosphere physics, element nucleosynthesis, and radiative transfer theory.

Before development of the necessary background theoretical knowledge, stellar spectra motivated the definition of stellar types and luminosity classes. These were the fruit of a classification effort categorising stars based on the identification and strength of their spectral features. Therefore, the chemo-physical parametrisation of stellar spectra has its roots in the large observational campaigns of the beginning of the 20th century (cf. Cannon & Pickering 1918) and the seminal works leading to the Morgan-Keenan classification (cf. Morgan et al. 1943).

* Corresponding author: A. Recio-Blanco,
e-mail: Alejandra.Recio-Blanco@oca.eu

The development of CCD detectors and, more recently, of multiobject facilities has resulted in the ability of even small telescopes to acquire large numbers of stellar spectra. Pioneering projects such as the Geneva Copenhagen Survey (Nordström et al. 2004), RAVE (Steinmetz et al. 2006), and SEGUE (Yanny et al. 2009), followed by archival parametrisation projects like AMBRE (de Laverny et al. 2012) and a worldwide observational effort illustrated by the *Gaia*-ESO Survey (Gilmore et al. 2012), LAMOST (Zhao et al. 2012), APOGEE (Majewski et al. 2017), and GALAH (Martell et al. 2017) characterise era of Galactic spectroscopic surveys. In parallel to the above-mentioned ground-based efforts, the design and preparation of the *Gaia* space mission, including the Radial Velocity Spectrometer (RVS; for a historical overview see Katz et al. 2004; Cropper et al. 2018, and references therein), opened new horizons in the observation of Milky Way stellar populations, and delivered on the promise of an unprecedentedly extensive spectroscopic survey (Wilkinson et al. 2005).

This rapid evolution of observational capabilities brought to the fore the need for automated parametrisation tools, enabling fast and homogeneous processing of extensive data sets. Once again, pioneering efforts followed the trail of the first spectroscopic surveys (e.g. Allende Prieto et al. 2006, among others) and the *Gaia* space project (Recio-Blanco et al. 2006). A variety of mathematical approaches have been developed and applied since then. These include different optimisation, projection, and classification methods used as part of model-driven or data-driven approaches (see e.g. Recio-Blanco 2014; Allende Prieto 2016; Jofré et al. 2019, and references therein).

Gaia observations started on 25 July 2014. The wavelength range covered by the RVS is [846–870] nm, and its medium resolving power is $R = \lambda/\Delta\lambda \sim 11\,500$ (Cropper et al. 2018). The present work describes the parametrisation of the first 34 months of *Gaia* RVS observations by the General Stellar Parametriser from spectroscopy (GSP-Spec) module of the Astrophysical parameters inference system (Apsis, Creevey et al. 2023). Apsis is the heritage of the previously described scientific pathway and the outcome of a long-term effort: from the development of innovative methodologies (Recio-Blanco et al. 2006; Bijaoui et al. 2010, 2012; Kordopatis et al. 2011) to their integration into the *Gaia* Data Processing and Analysis Consortium (DPAC) framework (Bailer-Jones et al. 2013), their tailoring to *Gaia*/RVS prelaunch characteristics (Recio-Blanco et al. 2016; Dafonte et al. 2016) and their first publication as part of the *Gaia* Data Release 3 (DR3; *Gaia* Collaboration 2023c). This effort results in the largest catalogue of stellar chemo-physical parameters ever published, which is simultaneously the first of its kind from a space spectroscopic survey and with all-sky coverage.

Section 2 presents the GSP-Spec goals and output parameters. This is followed by a description of the input *Gaia* RVS data (Sect. 3) and reference synthetic spectra grids (Sect. 4). The spectral line selection used for individual abundance analysis is explained in Sect. 5. The two GSP-Spec analysis workflows, MatisseGauguin and artificial neural network (ANN), are described in detail in Sects. 6 and 7, respectively. Section 8 presents the performed validation of GSP-Spec outputs as part of *Gaia* DR3 operations, and defines the implemented quality flags. Section 9 is devoted to the comparison of GSP-Spec results to literature data and suggested calibrations. Finally, in Sects. 10 and 11 we present the GSP-Spec catalogue and our conclusions.

2. Goals and outputs of GSP-Spec

The GSP-Spec module implements a purely spectroscopic treatment. It estimates the stellar chemo-physical parameters from

combined RVS spectra of single stars. No additional information is considered from astrometric, photometric, or spectrophotometric BP/RP data¹.

In particular, GSP-Spec estimates (i) the stellar effective temperature T_{eff} , reported as `teff_gspspec`; (ii) the stellar surface gravity expressed in logarithm $\log(g)$ ², reported as `logg_gspspec`; (iii) the stellar mean metallicity $[M/H]$ ³ defined as the solar-scaled abundances of all elements heavier than He and reported as `mh_gspspec`; (iv) the enrichment of α -elements⁴ with respect to iron ($[\alpha/Fe]$), reported as `alphafe_gspspec`; (v) the individual abundances of 13 chemical species ($[N/Fe]$, $[Mg/Fe]$, $[Si/Fe]$, $[S/Fe]$, $[Ca/Fe]$, $[Ti/Fe]$, $[Cr/Fe]$, $[Fe\ I/M]$, $[Fe\ II/M]$ ⁵, $[Ni/Fe]$, $[Zr/Fe]$, $[Ce/Fe]$, $[Nd/Fe]$, (reported as `Xfe_gspspec` or `Xm_gspspec`, with X being the chemical species), including the number of used spectral lines (`Xfe_gspspec_nlines`) and the line-to-line scatter (`Xfe_gspspec_linescatter`); (vi) a CN differential abundance proxy reported as `cnew_gspspec`; (vii) the equivalent width (EW) and fitting parameters of the diffuse interstellar band (DIB) at 862 nm, reported as `dibew_gspspec`, `dibp1_gspspec` and `dibp2_gspspec`; (viii) a goodness-of-fit (*gof*) over the entire spectral range reported as `logchisq_gspspec`; and (ix) a quality flag chain (the use of which is highly recommended) considering different error sources affecting the output parameters and reported as `flags_gspspec`.

Two different procedures, MatisseGauguin and ANN, described in Recio-Blanco et al. (2016), are applied to estimate the stellar atmospheric parameters (T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$). Individual chemical abundances and DIB parameters are estimated by the GAUGUIN algorithm (Recio-Blanco et al. 2016) and Gaussian fitting methods (Zhao et al. 2021), respectively, and are only produced by the MatisseGauguin analysis workflow⁶. The goodness-of-fit and the quality flag chain are provided for both the MatisseGauguin and ANN parametrisation. It is worth noting that, for each star, parameter uncertainties are estimated from 50 Monte-Carlo realisations⁷ of its RVS spectrum, considering flux uncertainties. For each realisation, a new complete analysis is implemented, including atmospheric parameters, individual chemical abundances, and CN and DIB parameters. From this analysis, upper and lower confidence values are respectively provided from the 84th and 16th quantiles of the resulting parameter and abundance distributions and reported with the suffix `_upper` and `_lower`, respectively (cf. Sect. 6.7).

The DR3 GSP-Spec analysis is available through two archive tables: the MatisseGauguin workflow provides 101 fields for 5 594 205 stars in the *AstrophysicalParameters* table, and the

¹ A separate Apsis module, the GSP from photometry is in charge of the stellar parametrisation from BP/RP data, using constraints from astrometric and stellar isochrones (Andrae et al. 2023).

² g being in cm s^{-2} .

³ In the following, we adopt the standard abundance notation for a given element X : $[X/H] = \log(X/H)_* - \log(X/H)_\odot$, where (X/H) is the abundance by number, and $\log \epsilon(X) \equiv \log(X/H) + 12$.

⁴ O, Ne, Mg, Si, S, Ar, Ca, and Ti are considered as α -elements and vary in lockstep.

⁵ Fe I and Fe II abundance enhancements with respect to the mean metallicity are estimated and respectively called *fem_gspspec* and *feII_m_gspspec* in the *AstrophysicalParameters* table.

⁶ It is worth mentioning that MatisseGauguin algorithms have been conceived assuming a white Gaussian noise framework.

⁷ This number of realisations has been optimised through simulations to ensure a good sampling of the associated parameter distributions, and taking into account the computation time allocated to GSP-Spec. We note that this Monte-Carlo procedure does not take into account uncertainties in the radial velocity correction, which have been considered through analysis flags (cf. Sect. 8.2).

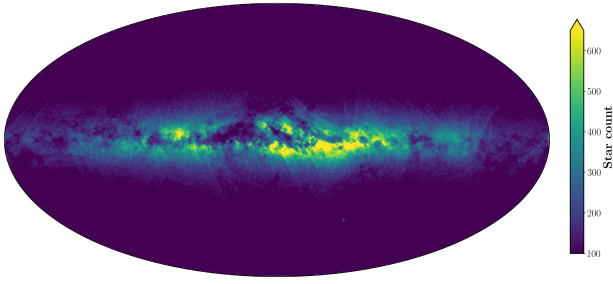


Fig. 1. Global all-sky spatial density distribution of all the GSP-Spec parametrised stars. This HEALPix map (Górski et al. 2005) in Galactic coordinates has a spatial resolution of 0.46° and at least 100 stars are contained in each resolution element.

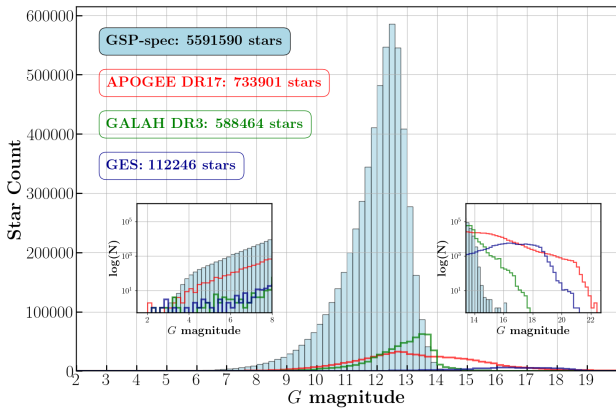


Fig. 2. *Gaia*-magnitude distribution of all the GSP-Spec parametrised stars. The APOGEE, GALAH, and GES magnitude distributions are shown for comparison in red, green, and blue, respectively.

ANN workflow provides 13 fields for 5 524 387 stars in the *AstrophysicalParametersSupp* table with an added `_ann` suffix in the parameter names. Figure 1 shows the spatial distribution in (l, b) Galactic coordinates of all the GSP-Spec parametrised stars. One can see that most stars are located close to the Galactic plane, as expected, although larger latitudes are still very well sampled with at least 100 stars per resolution element. The small-scale structures close to the Galactic plane are caused by interstellar absorption. Figure 2 illustrates the magnitude distribution of all the GSP-Spec parametrised stars in the G -band. The parametrised stars can be seen to cover a large range of magnitudes, starting from the brightest objects (about 4000 of them have $G < 6$, i.e. about two-thirds of the sky visible to the naked eye) to the faintest ones up to $G \sim 16$ (more than half a million and $\sim 100\,000$ have $G > 13$ and $G > 13.5$, respectively). This very high number statistics can also be appreciated for the magnitude bins with the highest number of stars. For instance, the bin $12.4 \leq G\text{-mag} < 12.6$ contains as many stars as published by the large ground-based spectroscopic survey GALAH. For comparison, Fig. 2 also shows the magnitude distributions of the largest ground-based spectroscopic surveys whose spectral resolution is larger than the RVS one: APOGEE-DR17 (Abdurro'uf et al. 2022), GALAH-DR3 (Buder et al. 2021), and the *Gaia*-ESO Survey (GES) (Gilmore et al. 2022; Randich et al. 2022). The highest number statistics of the *Gaia* GSP-Spec catalogue is achieved for $G < 13.6$ mag. For magnitudes fainter than about $G \sim 14.0$, APOGEE dominates with about 100 000 stars. GES also complements *Gaia* DR3 data at such fainter magnitudes with several tens of thousands, while GALAH has only a few

thousand stars fainter than this data release of GSP-Spec. We note that the number of stars parametrised by GSP-Spec will strongly increase with the next *Gaia* data releases, being about a factor ten larger in DR4 as a result of the spectra signal-to-noise ratio (S/N) increase with repeated observations (and hence with observing time).

The GSP-Spec analysis module is coded in Java following DPAC requirements, and is executed at the Data Processing Centre C hosted by the Centre National d'Etudes Spatiales (CNES) in Toulouse, France. During DR3 operations, about 6.9 million spectra were processed by the module in $\sim 110\,000$ h, spread over ~ 2100 cores (execution time of around 130 h, all the cores not being fully dedicated to GSP-Spec). The necessary RAM to run GSP-Spec is 25–30 GB. Therefore, the total execution time to derive the two sets (MatisseGauguin and ANN) of four atmospheric parameters, the 13 individual chemical abundances, the CN differential abundance proxy, the DIB fitting parameters, and all the associated uncertainties and goodness of fit is about one second per spectrum for one Monte-Carlo realisation of the noise.

An illustration of the GSP-Spec parameterisation was published as a *Gaia Image of the Week*⁸. GSP-Spec parameters are also used in the *Gaia* DR3 chemical cartography analysis (Gaia Collaboration 2023a,b). GSP-Spec is the main spectroscopic parameteriser module of the *Gaia* Apsis pipeline, independent of other modules, and feeds some of them executed afterwards in the module chain. The GSP-Spec methodology was largely tested on ground-based spectroscopic observations resulting from different projects, such as RAVE (Steinmetz et al. 2006), GES (Recio-Blanco et al. 2014), and AMBRE (de Laverny et al. 2013), among others.

3. Input *Gaia* RVS data

As input, GSP-Spec uses combined RVS spectra (averaged over multiple transits) and their flux uncertainties per wavelength pixel (wlp) over the 846–870 nm spectral domain. Prior to the GSP-Spec module operations, the stellar radial velocity (V_{Rad} , Katz et al. 2023) is used to Doppler shift RVS CCD spectra to the rest frame before combining them into a mean RVS spectrum (Seabroke et al., in prep.). The actual RVS wavelength range extends into the filter wings (845–872 nm, see Cropper et al. 2018, Fig. 16), and the cut to 846–870 nm minimises border effects. In addition, the spectra are normalised at the local pseudo-continuum and are resampled to a wavelength bin width of 0.01 nm (2400 wavelength points, wlp hereafter) by the DPAC/Coordination Unit6 (CU6) pipelines.

It is important to note that GSP-Spec reassesses the continuum placement during the parameterisation procedure (see Sect. 6.3). Moreover, the spectra are rebinned from 2400 to 800 wlp , sampled every 0.03 nm (without reducing the spectral resolution thanks to the RVS oversampling), which increases their S/N. The RVS spectra analysed by GSP-Spec during DR3 operations were selected to have $S/N > 20$ before resampling. The considered S/N corresponds to the `rv_expected_sig_to_noise` value provided by the CU6 analysis (Seabroke et al., in prep.).

It is worth mentioning that, although the mean RVS spectra serve as an input to GSP-Spec, subsequent filtering of V_{Rad} was not propagated to GSP-Spec outputs for DR3. This means that there are a very small number of stars with GSP-Spec parameters, but not V_{Rad} (Appendix A). A subset of RVS mean spectra (999 995, all having V_{Rad}) are published for the first time in

⁸ https://www.cosmos.esa.int/web/gaia/iow_20210709

DR3 (Seabroke et al., in prep.). These articles detail the overlap of the published mean RVS spectra with GSP-Spec parameters (and other *Gaia* parameters).

4. Input and training synthetic spectra grids

GSP-Spec performs a model-driven parametrization for which stellar flux dependencies on atmospheric parameters and surface chemical abundances are interpreted through the comparison of the observed spectra with theoretical (*synthetic*) ones. For this purpose, we have computed large grids of synthetic RVS spectra with different combinations of stellar atmospheric parameters (T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$) and individual chemical abundances ($[X/Fe]$, with X being the considered element, with the exception of Fe I and Fe II for which $[X/M]$ is used). They span the entire parameter space of Galactic stellar populations with a detailed coverage that allows to reach the required parametrization precision. The use of these grids is three-fold: (i) training the GSP-Spec MATISSE (cf. Sect. 6.1) and ANN (cf. Sect. 7) algorithms before their application; (ii) acting as reference models for the algorithm performing on-the-fly regressions (GAUGUIN), and (iii) anchoring the normalization and DIB analysis procedures to reference flux values.

As a consequence, a 4-dimensional grid of spectra in T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$ (cf. Sect. 4.2) and 5-dimensional grids for twelve chemical elements with the fifth dimension being $[X/Fe]$ (cf. Sect. 4.3) are provided as input for GSP-Spec together with the learning functions of the parametrization algorithms. These synthetic spectra are calculated through a procedure previously implemented for the AMBRE Project (de Laverny et al. 2012). We refer to a detailed description of the AMBRE grid to de Laverny et al. (2013). In the following, we particularly focus on several improvements considered for the GSP-Spec module.

4.1. Set of MARCS atmosphere models

The reference spectra are computed using MARCS atmosphere models (Gustafsson et al. 2008). We first selected 13 848 models that covered the following parameter space: 2600 to 8000 K for T_{eff} in steps of 200 or 250 K (below or above 4000 K, respectively), -0.5 to 5.5 for $\log(g)$ (step of 0.5 dex), and -5.0 to 1.0 dex for the mean metallicity (step of 0.25 dex for $[M/H] > -2.0$ dex and 0.5 dex for lower $[M/H]$ values). For each metallicity, all the available $[\alpha/Fe]$ -enrichments were considered. In practice, this corresponds to models with $[\alpha/Fe]$ -values varying between at most -0.4 dex and $+0.8$ dex, around the classical relation observed for Galactic populations: $[\alpha/Fe] = 0.0$ dex for $[M/H] \geq 0.0$ dex, $[\alpha/Fe] = +0.4$ dex for $[M/H] \leq -1.0$ dex and $[\alpha/Fe] = -0.4 \times [M/H]$ for $-1.0 \leq [M/H] \leq 0.0$ dex. We point out, however, that not all values of $[\alpha/Fe]$ were always available for a given set of T_{eff} , $\log(g)$, and $[M/H]$. Moreover, we only selected models for dwarfs (defined as $\log(g) > 3.5$) with plane-parallel geometry and a microturbulent-velocity parameter of 1.0 km s^{-1} whereas spherical geometry with a mass of $1 M_{\odot}$ and $V_{\text{micro}} = 2 \text{ km s}^{-1}$ were considered for giants ($\log(g) \leq 3.5$). Then, in order to improve the covering of the parameter space (particularly in the $[\alpha/Fe]$ dimension for which we adopted a step of 0.1 dex), we filled this first selection of MARCS models by models interpolated linearly, using the tool developed by T. Masseron and available on the MARCS website⁹. The resulting

grid of MARCS atmosphere models adopted in the present work contains 35 803 models.

4.2. The 4-D spectra grid in T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$

For each adopted MARCS atmosphere model, a synthetic spectrum has been computed with the TURBOSPECTRUM code (version 19.1.2, Plez 2012) between 842.0 nm and 874.0 nm (i.e. a wider spectral domain than the one covered by the RVS spectra, in order not to be affected by border effects when simulating the RVS-like spectra) and adopting an initial wavelength step of 0.001 nm (i.e. corresponding to a spectral resolution larger than $\sim 300\,000$). We considered the Solar abundances of Grevesse et al. (2007), and specific atomic and molecular line lists. These line lists contain millions of lines and have been checked (and, when necessary, some atomic lines were calibrated) with observed spectra of benchmark reference stars (see Contursi et al. 2021, for more details). For dwarfs (defined as above for the MARCS models by $\log(g) > 3.5$), the spectra were computed assuming one-dimensional plane-parallel atmospheric model while for giants ($\log(g) \leq 3.5$) a spherical geometry is considered. Both cases assume hydrostatic and local thermodynamic equilibria. Similar stellar masses as in the MARCS models were adopted for the computation. Moreover, consistent $[\alpha/Fe]$ -enrichments were considered in the model atmosphere and the synthetic spectrum calculation. Finally, we used an empirical law for the microturbulence parameter. This parametrized relation is a function of T_{eff} , $\log(g)$ and $[M/H]$ and has been derived from V_{micro} literature values for the *Gaia*-ESO Survey (Bergemann et al., in prep.). The spectra were computed in the air and then converted into vacuum wavelengths thanks to the relation of Birch & Downs (1994). It is worth noting that no stellar rotation or macro-turbulence broadening were included in these spectra. The impact of this assumption in the derived stellar parameters has been estimated from simulations and accounted through quality flags (Appendix C.1). These flags are a function of the *vbroad* parameter value of each star (available in the *gaia_source* table) but also of T_{eff} , $\log(g)$ and $[M/H]$.

The high-resolution spectra were then convolved and resampled in order to mimic real observed RVS spectra. For that purpose, we adopted a broadening instrumental profile corresponding to the RVS spectral resolution, keeping only the 846–870 nm domain and adopting the sampling of 0.03 nm chosen for the parametrization within GSP-Spec (800 *wlp*, see Sect. 3). In practice, this convolution was performed thanks to tools developed for the DR3 version of the CU6 pipeline (Sartoretti et al. 2018). It assumes a Gaussian ALong-scan line spread function and adopts the median resolving power value known at the beginning of CU8's DR3 processing phase ($R = 11\,500$, Cropper et al. 2018).

Finally, for the stellar atmospheric parameters estimation (see Sects. 6 and 7), this original grid of RVS-like synthetic spectra has been filled adopting a cubic Catmull-Rom (Catmull et al. 1974), a quadratic or linear 1D interpolation, depending on the number of neighbour models available. The final 4D grid contains 51,373 spectra with a constant step of 250 K, 0.5, 0.25 dex and 0.1 dex in T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$, respectively. Figure 3 illustrates the covered parameter space.

4.3. 5D spectra grids for individual chemical abundance estimations

For the derivation of individual chemical abundances with the GAUGUIN method (Sect. 6.4), we have computed sets of 5D grids for which the first four dimensions are the ones of the 4D

⁹ <https://marcs.astro.uu.se/>

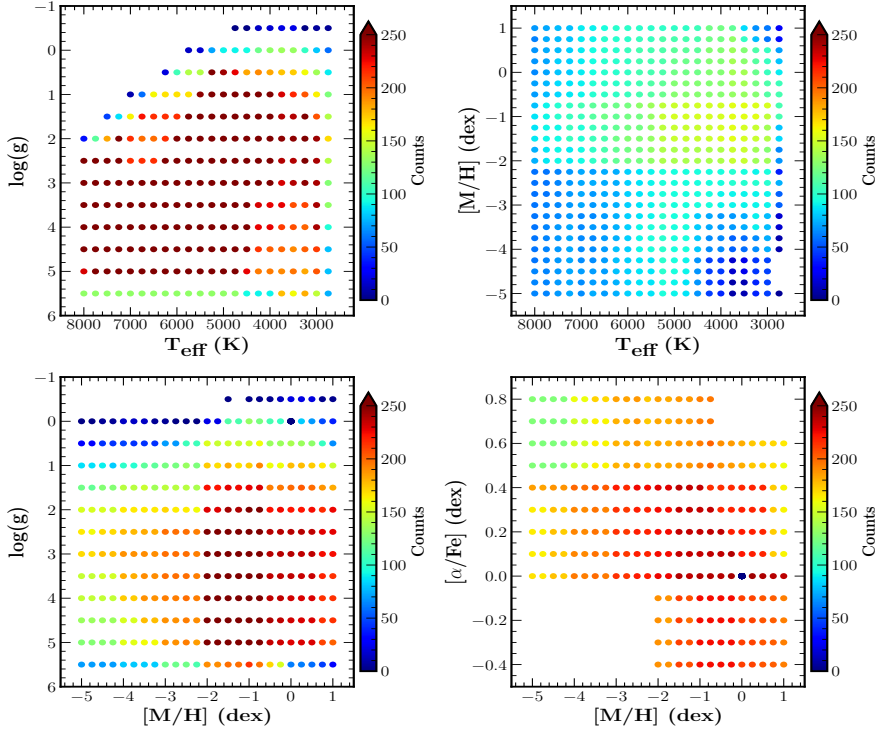


Fig. 3. Distribution in the 4D parameter space of the GSP-Spec reference grid, that contains the 51 373 synthetic spectra adopted for the stellar parametrisation. The colour-code refers to the number of available spectra in each 2D projection. For the derivation of the chemical abundance of a given chemical element X with the GAUGUIN method, 21 spectra are computed for most combinations of the four atmospheric parameters by varying the individual abundance of X (12 different species were considered: N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce, Nd).

grid described above while the fifth dimension corresponds to the abundance values of a specific chemical species $[X/Fe]$ (with the exception of Fe I and Fe II for which $[X/M]$ is used). The considered chemical elements, X , are N, Mg, Si, S, Ca, Ti, Cr, Fe, Ni, Zr, Ce, Nd. These species have been chosen due to the availability of at least one of their atomic lines in the RVS spectral domain, following a careful line quality selection (see Sect. 5).

For these 5D grids, we considered a subsample of the MARCS models selected in Sect. 4.1: $T_{\text{eff}} > 3500$ K, $[M/H] > -3.0$ dex and any values of $\log(g)$ and $[\alpha/Fe]$, except for Ca, Fe and Ti. Some atomic lines of these three atoms can indeed be detected at the very metal-poor regime and we therefore computed their 5D grids for any $[M/H]$ values, i.e. down to $[M/H] = -5.0$ dex. The adopted variations in the chemical element dimension are from -2.0 to $+2.0$ dex around $\epsilon(X) = \epsilon(X)_{\odot} + [M/H] + K_{\alpha}$, with a step of 0.2 dex (i.e. 21 different abundance values). K_{α} is assumed to be equal to zero for all elements except the α -species for which it follows a similar variation with the metallicity as $[\alpha/Fe]$: $K_{\alpha} = 0.0$ for $[M/H] \geq 0.0$ dex, $K_{\alpha} = +0.4$ for $[M/H] \leq -1.0$ dex and $K_{\alpha} = -0.4 \times [M/H]$ for $-1.0 \leq [M/H] \leq 0.0$ dex.

In total, we have computed twelve 5D grids of $\sim 478\,400$ spectra each, except for Ca, Fe and Ti whose grids contain $\sim 590\,750$ spectra since they cover the entire metallicity regime of the atmosphere model grids.

5. Line and wavelength interval selection for individual abundance analysis

As mentioned above, the reference synthetic spectra grids contain all the atomic and molecular lines collected by Contursi et al. (2021). Most of these lines are too weak and/or blended and can therefore not easily be used to derive reliable chemical diagnostics. To choose the adequate spectral intervals for individual abundance estimation, we implemented a careful line selection procedure and a thorough definition of the wave-

length intervals for abundance estimation and local normalisation described below.

5.1. Selection of unblended lines

First, we looked for unblended lines through visual inspection of synthetic spectra at high-resolution ($R \sim 100\,000$) and at the resolution of RVS ($R \sim 11\,500$). The atmospheric parameters of four well-known reference stars were adopted: two cool giants (Arcturus and μ Leo), one cool dwarf (the Sun), and one hot dwarf (Procyon)¹⁰. In particular, we looked at (i) the flux contribution of each chemical species (including the 12 atomic elements and the most abundant molecules) by computing specific spectra with highly enhanced abundances, and (ii) the existing blends assuming super-solar metallicities and high enhancements in α -elements. This led to an initial selection of about 130 isolated atomic lines belonging to a dozen different atoms and five CN lines¹¹ that could be useful for chemical diagnostics. In particular, we identified interesting lines of some heavy elements (Zr, Ce, and Nd) and one line of singly ionised iron at $\lambda = 858.794$ nm, as suggested by Contursi et al. (2021), to complement iron abundance based on Fe I lines (see Sect. 8.7.2). The correct simulation of these lines was verified through the comparison of synthetic spectra to high-resolution observed spectra for the four mentioned benchmarks.

Second, the previous selection was confirmed by examining the observed RVS spectra of a few stars with atmospheric parameters close to those of the reference ones. By visual inspection, we kept only the lines showing the highest sensitivity to abundance variations in at least one of the inspected spectra,

¹⁰ The adopted parameters for these stars can be found in Contursi et al. (2021).

¹¹ In our tests, CN was the sole identified molecule with rather unblended lines but this work has to be extended towards cooler stars ($T_{\text{eff}} < 4000$ K) for future *Gaia* releases.

excluding those for which blends were still suspected from lines of different chemical elements within ~ 0.3 nm.

5.2. Selection of abundance and local normalization windows

To further optimise the line selection and the chemical analysis procedure, we carefully defined two wavelength windows around the selected lines used in the abundance estimation and the local normalisation, respectively. These were defined after visual inspection of the Arcturus, solar, and Procyon spectra at the RVS resolution, maximising the wavelength domains (and therefore the information on the abundance and the continuum placement) and avoiding nearby lines. To ensure the reliability of the finally selected windows, chemical abundances were derived using GAUGUIN for a set of about 10 000 RVS spectra and slight variations of the window interval. This allowed us to exclude window definitions producing very discrepant results (≥ 0.5 dex) with respect to other lines of the same element and their average value.

For line doublets and triplets, the merger of each line within a single abundance determination and normalisation window was adopted whenever possible. These are referred to as *merged multiplets* hereafter. In the particular case of the Ca II IR triplet, to minimise NLTE effects, two abundance windows were defined at the line wings, avoiding the cores (i.e. up to six independent abundance estimates can be provided for the three Ca II lines).

5.3. Line selection based on line-to-line abundance scatter

Finally, from the above set of unblended lines, we performed an additional selection to optimise the line-to-line scatter. For some species (N, Si, S, Ca, Ti, Cr, Fe I), the final element abundance was computed by combining the results of the different single line (or merged multiplets) abundances of the same element (cf. Sect. 6.8)¹². For that purpose, we computed 50 Monte-Carlo flux realisations of the above-mentioned set of RVS spectra, considering their corresponding flux covariances. For each spectral line, the median and the inter-quartile range (IQR) were estimated from the derived abundance distribution. We then explored all the possible line combinations to evaluate the contribution of each line to the mean abundance, as well as its effect on the total number of estimates. A mean abundance is derived for each line combination, weighted by the inverse of the individual line abundance IQR. These weights were set to zero if they had $\text{IQR} > 0.5$ dex to avoid low-quality estimations. For each chemical element, the combination of lines minimising the line-to-line scatter and maximising the correlation¹³ of the mean abundance with $[\text{M}/\text{H}]$ (for iron-peak elements) or $[\alpha/\text{Fe}]$ (for α -elements) was selected.

The final list of 33 lines (some being merged multiplets) selected for the individual abundance analysis is provided in Appendix B and Table B.1, together with their associated windows for chemical analysis and normalisation. We refer also to the two figures of Sect. 6 for some examples of observed and model spectra that help to identify most of these lines. We note that Zr, Ce, and Nd lines passed all the above tests and are therefore adopted for abundance estimation. Similarly, the Fe II line

¹² Other derived abundances (Mg, Ni, Fe II, Zr, Ce, Nd) rely on only a single line or a single abundance determination in the case of merged multiplets.

¹³ We first performed a linear fit and then obtained the slope, the intercept, the median, and the MAD of the distance $|\text{data-fit}|$.

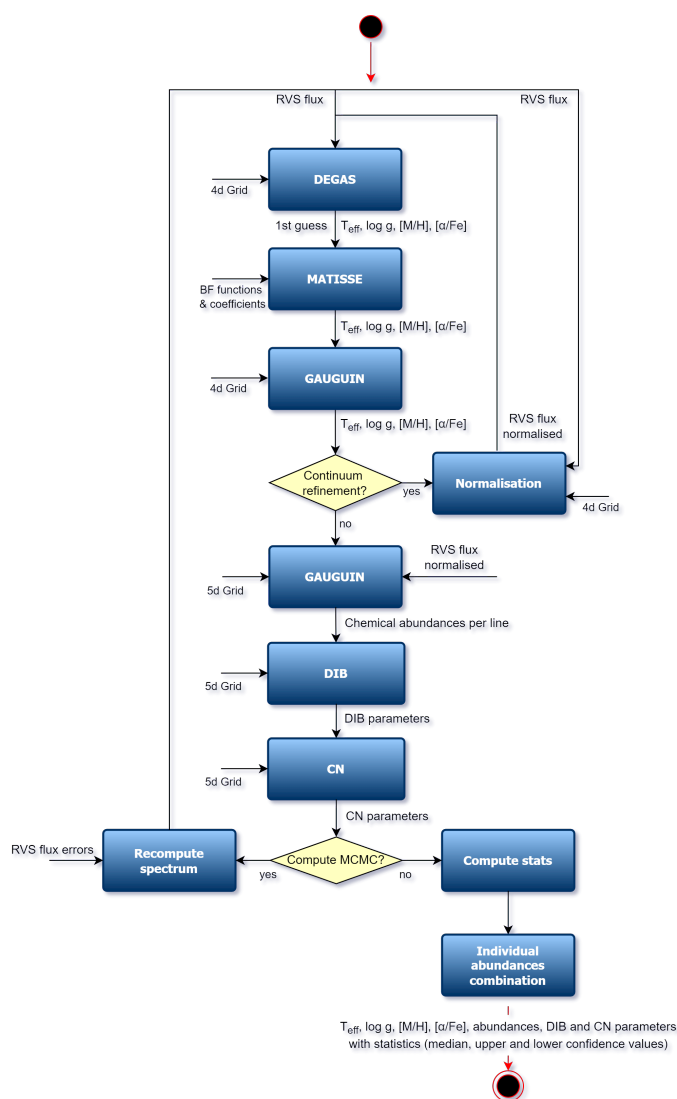


Fig. 4. Complete MatisseGauguin workflow that estimated stellar atmospheric parameters (T_{eff} , $\log(g)$, $[\text{M}/\text{H}]$, and $[\alpha/\text{Fe}]$), individual chemical abundances of 12 species, CN, and DIB parameters (see Sect. 6 for detailed description).

is also conserved. As a consequence, the GSP-Spec module is able to estimate abundances of neutral and singly ionised iron (see Sect. 8.7.2).

6. GSP-Spec MatisseGauguin analysis workflow

This section describes the MatisseGauguin analysis workflow in sequential order. We reiterate that MatisseGauguin produces the GSP-Spec fields published in the *AstrophysicalParameters* table, including stellar atmospheric parameters, individual chemical abundances, and DIB parameters.

The complete workflow of MatisseGauguin is summarised in Fig. 4. In addition, to illustrate the MatisseGauguin parametrisation performance, the challenging automated fit of two observed high-S/N spectra¹⁴ is presented in Figs. 5 and 6. The presented synthetic spectra are computed using the atmospheric parameters and chemical abundances estimated by GSP-Spec for those

¹⁴ These two spectra are not part of the set of RVS spectra published in the *Gaia* DR3.

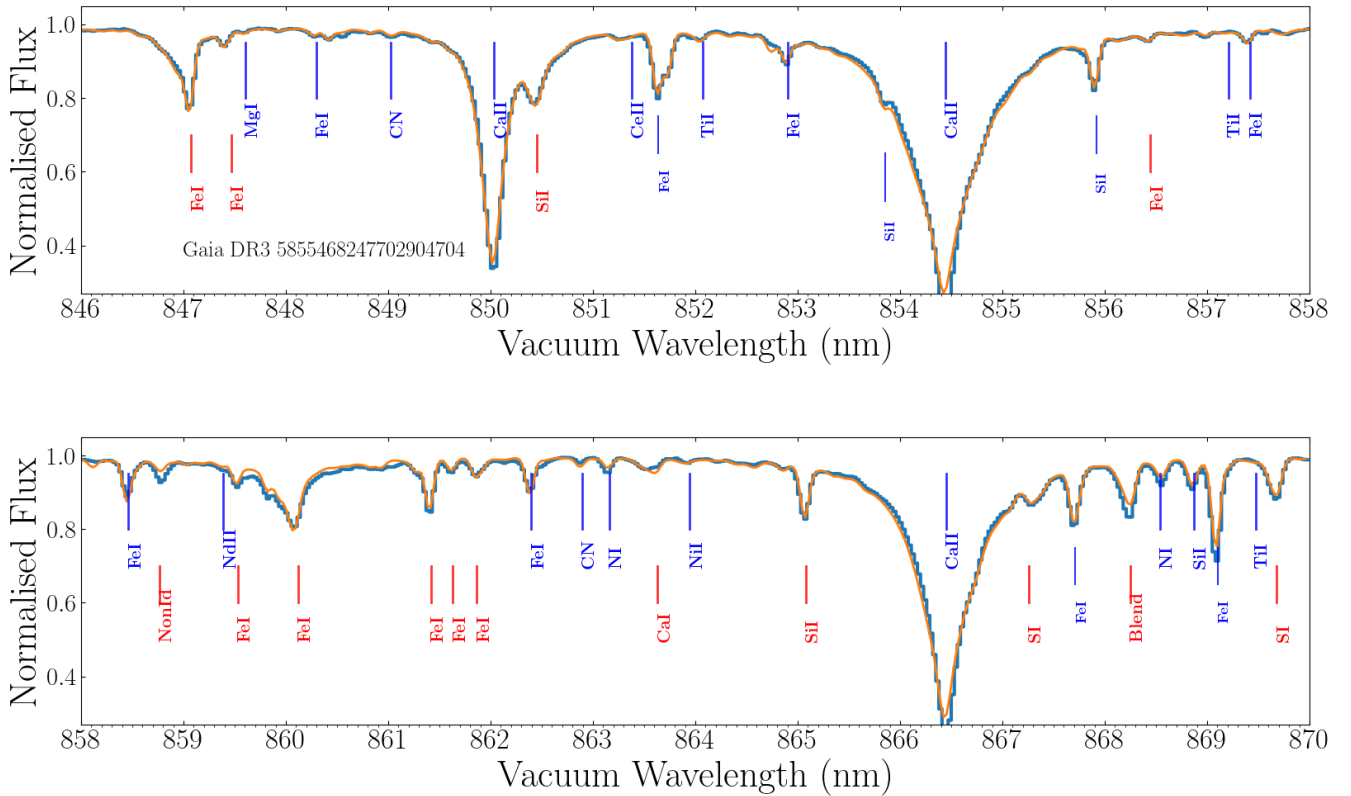


Fig. 5. Observed (blue histogram) and synthetic (orange line) spectra of the Cepheid variable star *Gaia* DR3 5855468247702904704. The observed spectrum has a very high S/N (equal to 884) and its histogram bin size corresponds to the wavelength sampling adopted for the analysis (0.03 nm, 800 *wlp*). The synthetic spectrum was computed from the GSP-Spec MatisseGauguin atmospheric parameters ($T_{\text{eff}} = 5477$ K, $\log(g) = 1.44$, $[M/H] = 0.07$ dex, $[\alpha/Fe] = 0.11$ dex) and individual chemical abundances, was then convolved by a rotational profile to reproduce the CU6 estimated broadening velocity (15.6 km s⁻¹) and, finally, was degraded to the RVS spectral resolution and sampling. The atomic lines identified in blue belong to the chemical species whose abundances were derived by the GAUGUIN method (the local normalisation performed for the chemical analysis of these selected lines was not considered in the figure for clarity reasons). The lines in red were not analysed in the shown spectrum because of suspected blends in the present case. The feature around 868.3 nm is a blend of Si+FeI+SiII plus probably other potential unidentified lines. The NonId feature at ~858.8 nm is a blend of the Fe II line described in Sect. 8.7.2 (seen in orange) and of unidentified lines that cannot be reproduced with the present line list.

two stars. The identification of several spectral lines is included in the figures. It is worth noting that the combination of the automated MatisseGauguin parametrisation with our reference spectra is able to find an excellent match with the observations that confirms the quality and the high precision of the observed RVS spectra and the input reference spectra grids.

6.1. MATISSE stellar atmospheric parameters

To initialise the whole MatisseGauguin procedure, a first guess of T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$ is derived using the DEGAS decision-tree method (Bijaoui et al. 2010), which considers the entire parameter space of the 4D grid (see also Kordopatis et al. 2011, 2013, for first applications to observed spectra).

Subsequently, the MATISSE algorithm (Recio-Blanco et al. 2006) is applied following an iterative procedure in the parameter estimation. This allows the user to overcome problems caused by a non-linear variation of the spectral flux with the stellar parameters. MATISSE is a local multi-linear regression method, resulting from the projection of the full input spectrum onto a set of vectors (called BF functions in Fig. 4). These vectors (and the associated coefficients) account for the sensitivity, at each wavelength, of the stellar flux to variations of a given parameter (ΔT_{eff} , $\Delta \log(g)$, $\Delta [M/H]$ or $\Delta [\alpha/Fe]$); they are derived dur-

ing a training phase based on the noise-free 4D reference grids, and correspond to regions of the entire parameter space, spanning ± 500 K in T_{eff} , ± 0.5 dex in $\log(g)$, ± 0.25 dex in $[M/H]$, and ± 0.20 dex in $[\alpha/Fe]$. The noise optimisation is taken into account by employing a Landweber algorithm during the covariance matrix inversion and which is adapted to each scientific application (see Recio-Blanco et al. 2006, for more details). The MATISSE projection is first applied at the DEGAS solution in a local environment of ± 500 K in T_{eff} , ± 0.5 dex in $\log(g)$, ± 0.25 dex in $[M/H]$, and ± 0.20 dex in $[\alpha/Fe]$ (corresponding to the parameter space region of each training function). This produces a second solution around which MATISSE is applied again. This iterative procedure is repeated until convergence (i.e. the solution stays within the local environment), within a maximum of ten iterations.

6.2. GAUGUIN refinement of the atmospheric parameters

The GAUGUIN algorithm is then applied around the final MATISSE solution of the previous step, considering a local environment of ± 250 K in T_{eff} , ± 0.5 dex in $\log(g)$, ± 0.25 dex in $[M/H]$, and ± 0.20 dex in $[\alpha/Fe]$. GAUGUIN (Bijaoui et al. 2012; Recio-Blanco et al. 2016) is a classical, local optimisation method implementing a Gauss-Newton algorithm. It is based on

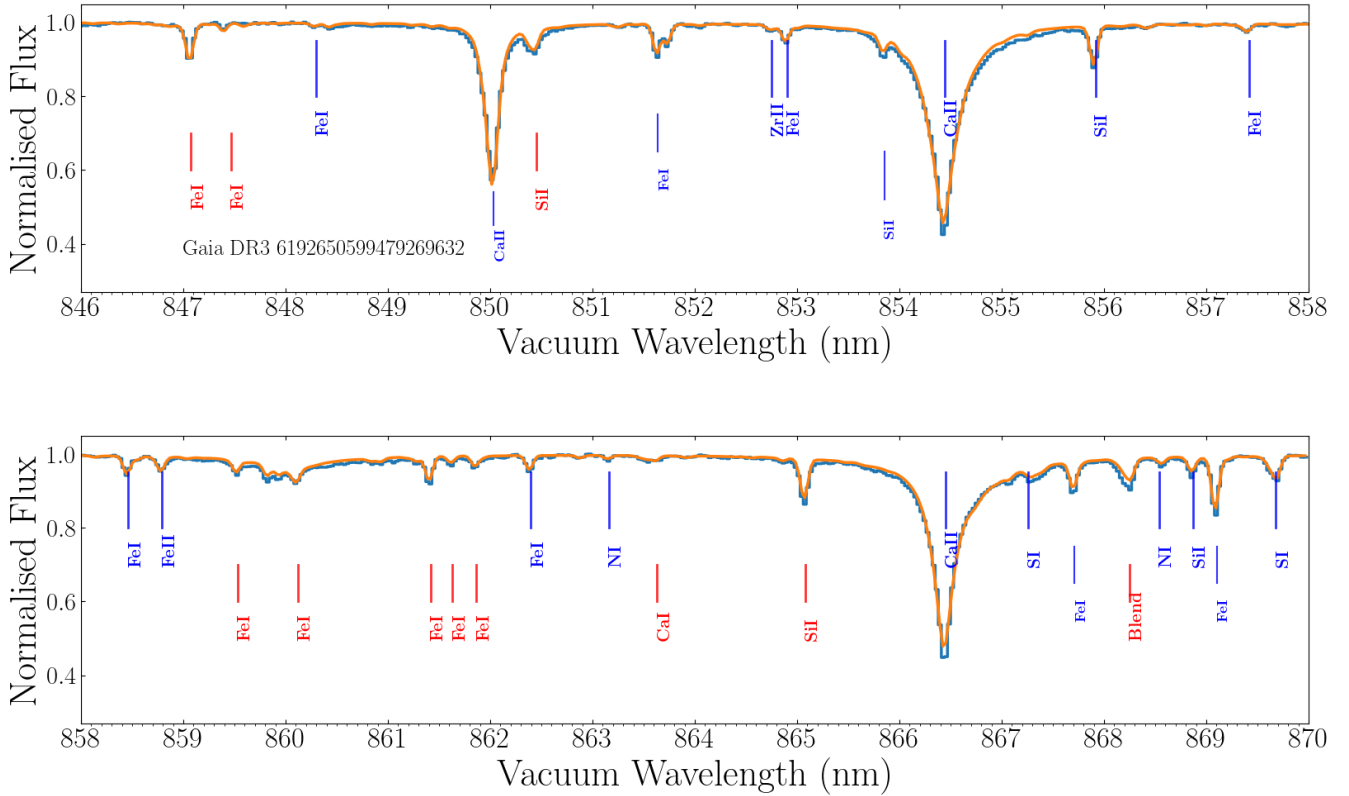


Fig. 6. Same as Fig. 5 but for the hot dwarf *Gaia* DR3 6192650599479269632 whose MatisseGauguin atmospheric parameters are $T_{\text{eff}} = 6754$ K, $\log(g) = 4.38$, $[M/H] = -0.03$ dex, and $[\alpha/Fe] = 0.15$ dex ($S/N = 408$). No rotational profile was applied as no broadening velocity was estimated (suspected low-rotating star).

a local linearisation around a given set of parameters that are associated with a reference synthetic spectrum (via linear interpolation of the derivatives). It is designed to find the direction in the parameter space that has the highest negative gradient as a function of distance (defined as the flux difference between the observed and synthetic spectra). Once this direction is found, the method proceeds in an iterative way, by modifying the initial guess of the studied parameter and re-calculating the gradient again, until convergence of the parameter solution. A few iterations are carried out through linearisation around the new solution until the algorithm converges towards the minimum distance. In practice, and to avoid trapping into secondary minima, we recall that GAUGUIN is initialised by parameters independently determined by MATISSE. At the end of this process, the final MatisseGauguin solution in T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$ is provided as input to the spectrum normalisation procedure.

6.3. Spectra re-normalisation and iterations on atmospheric parameters

The parameter solution of the previous step is used to re-estimate the continuum placement. This step is particularly important in the case of cool stars, which have pseudo-continuum flux values that can be much lower than one. The continuum placement and normalisation procedure is described in detail in Santos-Peral et al. (2020). In this step, the spectrum flux is normalised over the entire RVS wavelength domain. For this purpose, the observed spectrum (O) is compared to an interpolated synthetic one from the 4D reference grid (S) with the same atmospheric parameters. First, the most appropriate wavelength points of the residuals ($\text{Res} = S/O$) are selected using an iter-

ative procedure implementing a linear fit to Res followed by a σ -clipping. The residual trend is then fitted with a third-degree polynomial. Finally, the refined normalised spectrum is obtained after dividing the observed spectrum by a linear function resulting from the fit of the residuals.

This renormalised spectrum is then fed back to the first step described in Sect. 6.1 to re-estimate the atmospheric parameters using the new spectra normalisation. This loop is performed five times (a sufficient number to reach convergence), iterating on the parameters and the continuum placement.

The parameters of the converged solution in T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$ is then saved, as well as the final normalised spectrum. A goodness-of-fit (gof) between the observed and a synthetic spectrum interpolated to the atmospheric parameters is computed. The logarithm of this gof is reported in the *AstrophysicalParameters* table under the *logchisq_gspspec* field. The provided gof value reports the goodness of fit with respect to the observed spectrum, not including Monte-Carlo variations of the flux (see Sect. 6.7).

6.4. GAUGUIN chemical abundances per spectral line

Considering the final atmospheric parameters solution and normalised spectrum, each of the 33 selected atomic lines (see Table B.1) is then analysed with GAUGUIN to estimate the chemical abundance of the related chemical element causing the line absorption.

First, for each line l associated with the chemical element X , a specific 1D grid in the $[X/Fe]$ abundance space is generated. To this purpose, the corresponding 5D grid presented in Sect. 4.3 is interpolated at the stellar T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$ values

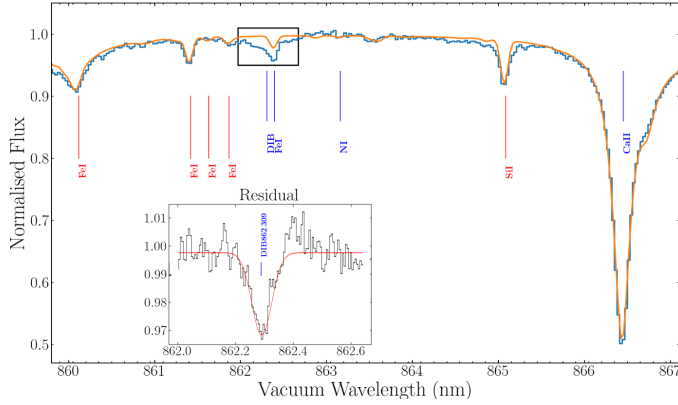


Fig. 7. Similar to Fig. 5 but for the metal-poor hot subgiant *Gaia* DR3 4378933739135936000 around its DIB feature. The insert is a zoom onto the flux residual between observed and model spectra around the DIB. It has been renormalised and the DIB characteristics are measured thanks to the Gaussian fit shown in red ($EW = 0.0244$ nm and central wavelength $p_1 = 862.309$ nm). The MatisseGauguin atmospheric parameters of this star are $T_{\text{eff}} = 6414$ K, $\log(g) = 3.75$, $[M/H] = -0.61$ dex, and $[\alpha/Fe] = +0.42$ dex ($S/N = 293$ and CU6 broadening velocity equal to 17.1 km s $^{-1}$).

of the adopted MatisseGauguin solution (cf. Sect. 6.3). This 1D reference spectra grid covers the entire normalisation wavelength range. It includes a large range of abundance variations in $\epsilon(X)$. Second, a local normalisation around the line is performed (Santos-Peral et al. 2020). A minimum quadratic distance is then calculated between the reference grid and the observed spectrum, providing a first guess of the abundance estimate $[X/Fe]_0^l$. This initial guess is then optimised using the GAUGUIN algorithm, which iterates through linearisation around the successive new solutions. The algorithm stops when the relative difference between two consecutive iterations is less than a given value (one-hundredth of the grid abundance step) and provides the final abundance estimation of each line ($[X/Fe]^l$).

6.5. Diffuse interstellar band parameters

Once the atmospheric parameters and the individual abundances have been derived, the next step of the MatisseGauguin workflow is to evaluate the presence of any DIB signature around ~ 862 nm. For each RVS spectrum, we first perform a local renormalisation on the spectrum around the DIB feature (over 35 \AA around 862 nm). We then pass a preliminary detection of the DIB profile by fitting a Gaussian profile to produce initial guesses for the fitting and eliminate cases where noise is at the same level as or exceeds the depth of the possible detection of the DIB. Only detections above the 3σ -level are considered as true detections. In order to perform the main fitting process of the DIB, we then separate our sample into cool ($3500 < T_{\text{eff}} \leq 7000$ K) and hot star samples ($T_{\text{eff}} \geq 7000$ K). For cool stars, the observed spectrum is divided by a synthetic spectrum whose atmospheric parameters are provided by MatisseGauguin. The residual, assumed to correspond to the DIB profile, is then renormalised and fitted by a Gaussian function (see Fig. 7). For hot stars for which no lines are found close to the DIB feature, a Gaussian process similar to Kos (2017) is applied where the DIB profile is fitted by a Gaussian process regression (Gershman & Blei 2012).

For each detected DIB feature, we determine its equivalent width (EW), the central wavelength of the fitted Gaussian (p_1), its depth (p_0), the width of the Gaussian profile (p_2), and their

uncertainties which are estimated based on Monte-Carlo Markov chain realisations (see Sect. 6.7). We remind the reader that, for a Gaussian, $p_2 = FWHM/(2 \cdot \sqrt{\ln 2})$ with $FWHM$ being the full width at half maximum. The EW is computed assuming a Gaussian profile: $EW = \sqrt{2\pi} \times |p_0| \times p_2/C$ where C is the continuum level.

Finally, two quality flags (DIB_q and QF) for the DIB parameters were implemented in order to allow a selection of the best determinations, depending on the science application (see e.g. Gaia Collaboration 2023b). The first quality flag DIB_q is included in the GSP-Spec quality flag string chain and its value varies from zero (highest quality) to five (lowest quality) and is equal to nine when no DIBs are measured; we refer to Sect. 8.9 for its definition. The second flag QF is defined during the preliminary detection of the DIB profile and provides the reason why DIB_q has been fixed to nine for a given spectrum.

If the depth of the fitted profile is smaller than 3σ the noise level, we do not consider this case to be a true detection and assign it $QF = -1$. Finally, stars with effective temperatures cooler than 3500 K are automatically disregarded because their spectrum is crowded by molecular lines, leading to undetectable DIB ($QF = -2$).

6.6. Cyanogen differential abundance proxy

In the spectra of cool stars, a couple of cyanogen lines can be seen (their wavelength identification can be found in Fig. 5, although the lines are weaker in the illustrated spectrum with respect to cooler stars). Five interesting CN lines were initially identified when building the line list. The tests performed in the line-selection process presented in Sect. 5 selected one of these CN lines as a reliable CN over- or underabundance proxy in the spectra of cool stars.

This CN line is centred at 862.884 nm and a window of 0.15 nm has been selected around it for its analysis. As for the DIB feature of cool stars, the observed spectrum is divided by the corresponding synthetic spectrum, interpolated to the atmospheric parameters of the star derived by MatisseGauguin. This synthetic spectrum assumes the solar-scaled values of carbon and nitrogen abundances $[C/Fe] = [N/Fe] = 0.0$ dex. We then estimated the EW of the residual by adopting the same Gaussian fitting procedure as for the DIB parameters of cool stars. This CN proxy is therefore an indicator of the strength of the line with respect to the standard value, and reveals a CN underabundance or overabundance (positive or negative EW, respectively). In addition, the central wavelength and the width of the residual feature are also derived from the above-mentioned Gaussian fit, as already implemented for the DIB.

6.7. Propagation of flux uncertainties

The estimation of a star's atmospheric parameters, chemical abundances from individual lines, DIB, and CN-index parameters described above is performed from the input RVS spectrum, without considering the associated flux uncertainties per wlp . To estimate parameter uncertainties induced by the spectral noise, the complete MatisseGauguin workflow is rerun 50 times to analyse the same number of different Monte-Carlo realisations of the stellar spectrum. Upon each realisation, the input stellar flux per wlp F_i is modified according to the corresponding flux uncertainty of that wlp , σ_{F_i} . In particular, each realisation is computed by adding or subtracting a ΔF_i at each wlp i , randomly

sampling a Gaussian distribution of standard deviation equal to σ_{Fi} and centred at zero.

The complete Monte-Carlo implementation produces a total set of 50 values for each estimated parameter (T_{eff} , $\log(g)$, $[M/H]$, $[\alpha/Fe]$, individual line abundances $[X/Fe]^l$, DIB, and CN indexes). For each of the corresponding parameter distributions, we compute the median and the lower and upper confidence values, from the 50th, 16th, and 84th quantiles, respectively. The median value of each parameter is saved as the adopted parameter estimation in the GSP-Spec catalogue. Both the lower and upper confidence levels are also published. In summary, this procedure allows parameter uncertainties to be properly estimated for each star, and for them to be tailored to the quality of the associated spectrum, but also to its stellar type and chemical abundance pattern. It is important to note that, in this way, the uncertainties on individual line abundances $[X/Fe]^l$ propagate the atmospheric parameters ones, as new $[X/Fe]^l$ values are computed upon each realisation for the new T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$ estimations. In addition, asymmetric uncertainties around the finally considered median value are provided thanks to the lower and upper confidence levels. This Monte Carlo treatment is made possible thanks to the extremely fast application of the GSP-Spec analysis (cf. Sect. 2).

6.8. Individual element chemical abundances

As explained in Sect. 6.4, GAUGUIN provides chemical abundances for each of the 33 atomic lines of Table B.1, called $[X/Fe]^l$. The final chemical abundances per element $[X/Fe]$ are derived by combining the independent abundance estimates $[X/Fe]^l$ of all the available lines of the same species. To this purpose, a mean abundance per element is calculated, weighted by the inverse of the $[X/Fe]^l$ uncertainty of each line (defined as the upper minus lower confidence values of the $[X/Fe]^l$ abundance distribution provided by the 50 Monte-Carlo realisations). The published abundances are $[N/Fe]$, $[Mg/Fe]$, $[Si/Fe]$, $[S/Fe]$, $[Ca/Fe]$, $[Ti/Fe]$, $[Cr/Fe]$, $[Fe\text{ I}/M]$, $[Fe\text{ II}/M]$ ¹⁵, $[Ni/Fe]$, $[Zr/Fe]$, $[Ce/Fe]$, and $[Nd/Fe]$. Their associated lower and upper confidence values are also published and were calculated as the weighted mean of the $[X/Fe]^l$ ones.

7. The GSP-Spec ANN workflow

The ANN algorithm is based on supervised learning and provides a different parameterisation of the RVS spectra, independent from the MatisseGauguin workflow. ANN projects the RVS spectra onto the label space of the astrophysical parameters. We trained the network on the same grid of reference synthetic spectra as MatisseGauguin (see Sect. 4), in this case adding noise according to the different S/N scales in the observed spectra (Manteiga et al. 2010).

The ANN architecture is feed-forward with three fully connected neuron layers. The input layer has as many neurons as wlp in the spectrum (800) whereas the output layer has four neurons corresponding to the number of estimated parameters. The number of neurons in the hidden layer was empirically determined between 50 and 100 for nets trained with low- to high-S/N spectra, respectively. In the same way, we determined the learning rate in the range [0.001, 0.2]. The activation function

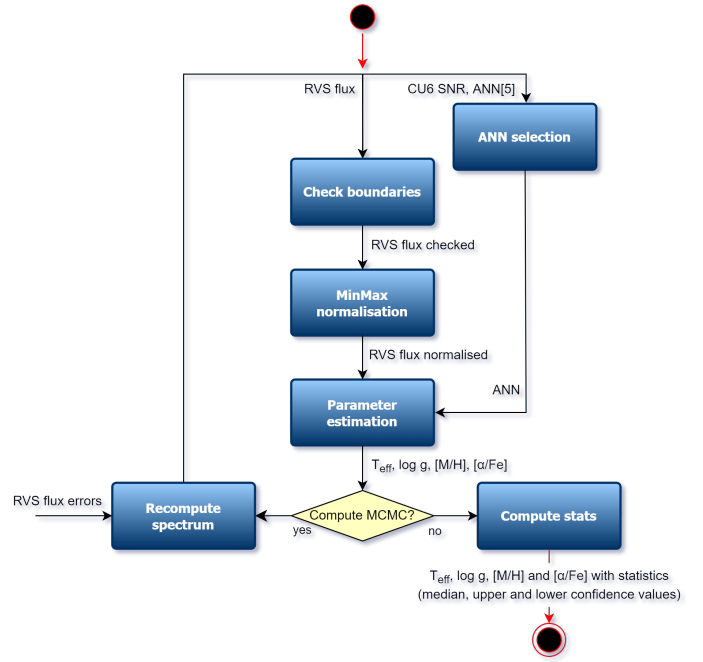


Fig. 8. ANN workflow that provides the second set of the main stellar atmospheric parameters (T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$).

Table 1. Equivalent S/Ns between ANN networks and RVS spectra.

ANN	25	30	35	40	50
RVS	[20, 24]	(24, 40]	(40, 68]	(68, 108]	(108, ∞)

selected for input and output layers is linear, whereas the logistic function was selected for the hidden layer.

The training procedure is performed with the backpropagation function, which can be interpreted as a problem of minimisation of the error existing between the obtained and desired outputs. In order to avoid overtraining, and to select the ANN that leads to the best generalisation, the early stopping procedure was used, finalising the training process when the performance starts to degrade, and obtaining the net that minimises the error.

The effectiveness of the ANN depends on the input ordering. For that reason, we perform ten trainings with different ordering, selecting the one with minimum error. For each train, weights initialise in the range $[-0.2, 0.2]$, and we established a limit of 1000 iterations because we observed that beyond that number, the training process does not improve but the computational cost increases.

The ANN parameterisation procedure that estimates the second set of GSP-Spec atmospheric parameters (T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$) is published in the *AstrophysicalParametersSupp* table and is summarised in Fig. 8. Specifically, the present ANN version included in GSP-Spec proceeds as follows:

ANN selection: ANN behaves well in the presence of noise (Manteiga et al. 2010), confirming that it is a robust method when estimating astrophysical parameters for relatively low-S/N spectra. As there is no noise model for the *Gaia* RVS spectra, we empirically determined the relation between the noise given by CU6 and the Gaussian noise that we need to use to train the nets. The corresponding values are shown in Table 1. For each RVS input spectrum, we then used its S/N value,

¹⁵ We provide iron abundances with respect to the mean metallicity following the implementation of the reference grids. The classical $[Fe/H]$ can be easily obtained by adding $[M/H]$ to $[Fe\text{ I}/M]$ or $[Fe\text{ II}/M]$.

provided by CU6, to select which net performs the parameter estimation.

Check boundaries: Some RVS spectra have zero flux values at the beginning or at the end of their spectral range. These are often caused by radial velocity corrections and could lead to large flux variations in the borders and cause ANN malfunctions. To avoid this behaviour, we truncated these zero flux values and adopted the mean of the flux spectrum for these *wlp*.

Normalisation: A minimum–maximum scaling procedure is applied to the RVS spectra, equalising it to avoid geometric biases during the training stage in order to guarantee that all the inputs are in a comparable range.

Parameter estimation: Once the net has been selected, it is fed with the normalised spectrum to estimate T_{eff} , $\log(g)$, $[M/H]$ and $[\alpha/Fe]$. The net returns these estimations normalised, and so a denormalisation procedure is applied to return the values in the expected range.

Monte-Carlo iterations using flux uncertainties: The same procedure as for MatisseGauguin (see Sect. 6.7) is also applied for ANN to estimate the parameter uncertainties caused by flux errors. We therefore obtain the median and the lower and upper confidence values of each AP again.

8. Validation and *flags_gspspec* quality flag chain

The GSP-Spec output after operations has been carefully checked and validated, considering different potential error sources. Following this validation procedure, a quality flag chain (*flags_gspspec*) is implemented (cf. Table 2)¹⁶. In this chain, a value of 0 is the best, and 9 is the worst, generally implying the parameter masking. This allows the user to publish all kinds of quality results, satisfying the more or less restrictive needs of different science applications. Nevertheless, this implies that considering these quality flags is mandatory for correct use of the GSP-Spec parameters and abundances. If not applied, results of low quality for a given application could be unconsciously included in the analysis, severely affecting its conclusions.

The following subsections review the different reasons for failure, potential bias, and the uncertainty sources considered in the GSP-Spec validation, and following the characters ordering in the quality flag chain. Several associated figures and tables can be found in Appendix C.

8.1. Parameterisation biases induced by rotational and macroturbulence line broadening (*vbroad* flags)

GSP-Spec is trained with reference spectra assuming no rotation (see Sect. 4). At the RVS spectral resolution, the parameterisation tolerance to broadened spectra through rotational ($V \sin i$) and/or macroturbulence broadening has to be flagged according to tests with synthetic data.

Potential biases in T_{eff} , $\log(g)$, and $[M/H]$ induced by stellar rotation were therefore modelled using a dedicated set of synthetic RVS spectra, which were broadened with different $V \sin i$ values from 0 to 70 km s⁻¹. For simplicity, we assumed in what follows that the line broadening factor produced by CU6 (*vbroad*) is well reproduced by only mimicking a stellar rotation. The estimated biases (ΔT_{eff} , $\Delta \log(g)$, $\Delta [M/H]$) induced by rotational broadening are a function of T_{eff} and $\log(g)$. Metallicity dependencies are also observed, with metal-poor objects

being more affected than metal-rich ones (a consequence of their smaller number of lines that can be used for the parametrisation).

First, using this data set, we identified the limiting $V \sin i$ values inducing a bias larger than $\Delta T_{\text{eff}} = 2000$ K. We then modelled the parameter dependence of the $V \sin i$ values leading to that maximum admitted bias by fitting a third-order polynomial with variable T_{eff} , $\log(g)$, and $[M/H]$, as shown in Fig. C.1. To avoid extrapolation issues, upper and lower limits were imposed on the third-order interpolation polynomial during the post-processing. This function was finally adopted during post-processing to mask the corresponding GSP-Spec results (Flag *vbroad* $T = 9$ in Table 2). Similarly, we applied this procedure to define three other values for this quality flag, depending on the amplitude of the predicted induced bias in T_{eff} : Flag *vbroad* $T = 0, 1, \text{ or } 2$ for stars with a possible bias $\Delta T_{\text{eff}} \leq 250$ K, $250 < \Delta T_{\text{eff}} \leq 500$ K, and $500 < \Delta T_{\text{eff}} < 2000$ K, respectively.

Exactly the same procedure was adopted for defining the flags associated with a bias in $\log(g)$ and $[M/H]$ induced by the rotational and macroturbulence line broadening. Their detailed definition is given in Table C.1.

8.2. Parameterisation biases induced by radial velocity uncertainty (*vrad* flags)

In a very similar way, we investigated the possible bias induced by radial velocity uncertainties, because the GSP-Spec parametrisation is performed whilst assuming that the observed spectra are perfectly at rest-frame. The examination of GSP-Spec unfiltered results reveals that large V_{Rad} errors (provided by CU6) are preferentially found in specific regions of the output atmospheric parameter space (combinations of T_{eff} , $\log(g)$, and $[M/H]$ where no stars are expected, or at extremely high or low $[\alpha/Fe]$). This is an important illustration of the expected parametrisation sensitivity to V_{Rad} uncertainties.

We therefore investigated the amplitude of possible biases in T_{eff} , $\log(g)$, and $[M/H]$ caused by V_{Rad} errors varying between 0 and 10 km s⁻¹ using specific synthetic spectra. Again, metal-poor stars (with a lower number of lines available for the parametrisation) were found to be more affected than metal-rich ones. As described above for the *vbroad* flags, specific third-order polynomials with variable T_{eff} , $\log(g)$, and $[M/H]$ were then fitted to define the values associated with three *vrad* flags. Their precise definition is given in Table C.2.

8.3. Parameter uncertainties due to flux noise (*fluxNoise* flag)

The parametrisation is affected by uncertainties in the observed fluxes, that is, the noise at each wavelength leading to a mean S/N over the entire wavelength domain. To quantify this effect and as already explained in Sect. 6.7, flux uncertainties are taken into account by GSP-Spec through 50 Monte-Carlo realisations of the spectral flux for each star. The GSP-Spec parametrisation is then performed for those 50 spectrum realisations and parameter uncertainties (noted σ , hereafter) are defined from the 16th and 84th quantiles of the obtained distributions. To enable a rapid selection of results in the GSP-Spec catalogue from the estimated parameter uncertainties, we defined a specific quality flag (*fluxNoise*). This flag simultaneously considers uncertainties in T_{eff} , $\log(g)$, $[M/H]$, and $[\alpha/Fe]$, labelling results of progressively higher precision from *fluxNoise* = 5 to *fluxNoise* = 0. The exact conditions imposed during the post-processing for the noise uncertainty quality flags are indicated in Tables C.3 and C.4 for MatisseGauguin and ANN, respectively. It is worth

¹⁶ We note that the flags associated with the ANN results correspond to the first 12 flags of this table.

Table 2. Definition of each character in the GSP-Spec quality flag string chain (*flags_gspspec*), including the possible values (Col. 3) and the related subsection and tables providing further information (Col. 4).

Chain character number - name	Considered quality aspect	Possible adopted values	Related subsection and table
1 vbroadT	<i>vbroad</i> induced bias in T_{eff}	0,1,2,9	8.1 & C.1
2 vbroadG	<i>vbroad</i> induced bias in $\log(g)$	0,1,2,9	8.1 & C.1
3 vbroadM	<i>vbroad</i> induced bias in [M/H]	0,1,2,9	8.1 & C.1
4 vrادT	V_{Rad} induced bias in T_{eff}	0,1,2,9	8.2 & C.2
5 vrادG	V_{Rad} induced bias in $\log(g)$	0,1,2,9	8.2 & C.2
6 vrادM	V_{Rad} induced bias in [M/H]	0,1,2,9	8.2 & C.2
7 fluxNoise	Flux noise induced uncertainties	0,1,2,3,4,5,9	8.3 & C.3, C.4
8 extrapol	Extrapolation level of the parametrisation	0,1,2,3,4,9	8.4 & C.5, C.6
9 negFlux	Negative flux <i>wlp</i>	0,1,9	8.5 & C.7
10 nanFlux	NaN flux <i>wlp</i>	0,9	8.5 & C.7
11 emission	Emission line detected by CU6	0,9	8.5 & C.7
12 nullFluxErr	Null uncertainties <i>wlp</i>	0,9	8.5 & C.7
13 KMgiantPar	KM-type giant stars	0,1,2	8.6 & C.8
14 NUpLim	Nitrogen abundance upper limit	0,1,2,9	8.7 & C.9
15 NUncer	Nitrogen abundance uncertainty quality	0,1,2,9	8.7 & C.10
16 MgUpLim	Magnesium abundance upper limit	0,1,2,9	8.7 & C.9
17 MgUncer	Magnesium abundance uncertainty quality	0,1,2,9	8.7 & C.10
18 SiUpLim	Silicon abundance upper limit	0,1,2,9	8.7 & C.9
19 SiUncer	Silicon abundance uncertainty quality	0,1,2,9	8.7 & C.10
20 SUpLim	Sulphur abundance upper limit	0,1,2,9	8.7 & C.9
21 SUncer	Sulphur abundance uncertainty quality	0,1,2,9	8.7 & C.10
22 CaUpLim	Calcium abundance upper limit	0,1,2,9	8.7 & C.9
23 CaUncer	Calcium abundance uncertainty quality	0,1,2,9	8.7 & C.10
24 TiUpLim	Titanium abundance upper limit	0,1,2,9	8.7 & C.9
25 TiUncer	Titanium abundance uncertainty quality	0,1,2,9	8.7 & C.10
26 CrUpLim	Chromium abundance upper limit	0,1,2,9	8.7 & C.9
27 CrUncer	Chromium abundance uncertainty quality	0,1,2,9	8.7 & C.10
28 FeUpLim	Neutral iron abundance upper limit	0,1,2,9	8.7 & C.9
29 FeUncer	Neutral iron abundance uncertainty quality	0,1,2,9	8.7 & C.10
30 FelUpLim	Ionised iron abundance upper limit	0,1,2,9	8.7 & C.9
31 FelUncer	Ionised iron abundance uncertainty quality	0,1,2,9	8.7 & C.10
32 NiUpLim	Nickel abundance upper limit	0,1,2,9	8.7 & C.9
33 NiUncer	Nickel abundance uncertainty quality	0,1,2,9	8.7 & C.10
34 ZrUpLim	Zirconium abundance upper limit	0,1,2,9	8.7 & C.9
35 ZrUncer	Zirconium abundance uncertainty quality	0,1,2,9	8.7 & C.10
36 CeUpLim	Cerium abundance upper limit	0,1,2,9	8.7 & C.9
37 CeUncer	Cerium abundance uncertainty quality	0,1,2,9	8.7 & C.10
38 NdUpLim	Neodymium abundance upper limit	0,1,2,9	8.7 & C.9
39 NdUncer	Neodymium abundance uncertainty quality	0,1,2,9	8.7 & C.10
40 DeltaCNq	Cyanogen differential equivalent width quality	0,9	8.8 & C.12
41 DIBq	DIB quality flag	0,1,2,3,4,5,9	8.9 & C.13

Notes. Flag names are split into three categories: Parameter flags (green), individual abundance flags (blue), and EW flags (orange). All flags concern the MatisseGauguin parameters, while only the parameter flags except KMgiantPar are applied to ANN results.

noting that stars with extremely poor quality parameters, such as, for instance, those without any distinction between giants and dwarfs ($\sigma \log(g) > 2$ dex) or between F, G, and K stellar types ($\sigma T_{\text{eff}} > 2000$ K) are filtered out during the post-processing (*fluxNoise* = 9) and do not appear in the finally published catalogue.

8.4. Extrapolation level (*extrapol* flag)

Due to extrapolation, the GSP-Spec parameter solution could be located outside the parameter space of the training grid (cf.

Fig. 3) for either one or several parameters. In addition, censored training occurs near the grid borders. In order to flag those extrapolated results for which the parametrisation is less reliable, we have implemented a specific flag (*extrapol*) that is indicative of the extrapolation level.

The definition of this flag is reported in Tables C.5 and C.6 for MatisseGauguin and ANN, respectively, depending on the availability (or not) of a *gof* and the distance between the parameter solutions and the grid borders. The flag value depends on the level of extrapolation: from results near the grid limits (*extrapol* = 4) to no extrapolation at all (and therefore a more

reliable solution, $extrapol=0$). Again, sources without a gof and with T_{eff} values outside the 2500 to 9000 K interval or $\log(g)$ values outside the -1 to 6 dex range were filtered out ($extrapol=9$) during the post-processing and do not appear in the final catalogue.

8.5. RVS flux issues or emission line flags

MATISSE and GAUGUIN being model-driven methods that essentially aim to maximise the goodness of fit between an observation and a set of templates, any significant and/or systematic difference between the RVS spectra and the reference grid can introduce biases in the results. These differences can be associated with the RVS spectra processing, or be inherent to the stellar physics assumptions adopted when computing the reference grid (stellar activity being one example). When such issues randomly affect a wlp , then it can be very difficult (if not impossible) to properly take them into account during the analysis. We have implemented four specific flags to identify such cases. Their definition is presented below and is summarised in Table C.7.

RVS spectral anomalies can manifest as wlp that have a negative flux (flag $negFlux$), or a flux (or associated variance) that is not a number ($nanFlux$ and $nullFluxErr$ flags, respectively). Whereas such caveats do not necessarily alter the RV determination, they can hamper parameterisation estimates relying specifically on the affected wlp . For instance, some tens of stars have a couple of wlp with negative flux. They are predominantly found in the cores of the strongest Ca II lines and result from an oversubtraction of the straylight during the spectrum production. This leads to a modified line profile and could indeed affect the parameterisation. Similarly, NaN flux values can appear in the spectra. As explained in Seabroke et al. (in prep.), wlp are masked in the CCD sample. When these are averaged, a chance alignment of these masks when there are few CCD spectra pixels contributing to a particular wavelength bin in the combined spectrum could lead to a NaN flux value, which happens more often near the edges. The GSP-Spec treatment partly overcomes this problem thanks to the rebinning (from 2400 to 800 wlp) of the oversampled input spectra. For this rebinning, a median flux is computed every three wlp , excluding NaN values. As a consequence, NaN flux values in the rebinned spectra only remain if the three averaged wlp are equal to NaN. To filter out those rare cases, we have implemented the specific $nanFlux$ flag. Finally, if no flux variance is associated with a wlp , then the derived parameter uncertainty is unreliable or impossible to estimate. This is reported by the $nullFluxErr$ flag.

As presented in Table C.5, while the $nanFlux$ and the $nullFluxErr$ flags lead to a systematic exclusion of the source from the final catalogue (only values equal to 9 have been implemented), the $negFlux$ flag can also be equal to 1 (one or two wlp with negative flux values) or 0 (no negative wlp at all). However, for the reasons described above, we recommend preferentially selecting stars with $negFlux=0$.

On the other hand, emission lines due to stellar activity are inherent to the stellar properties and carry important information about the observed star. However, the physical conditions that lead to the emission lines are not considered in our grid of synthetic spectra. Therefore, if a star shows signs of activity, its GSP-Spec parameters should also be discarded and considered unreliable. We used the $CU6_is_emission$ flag provided by the CU6 to detect such stars, and forced them to have a GSP-Spec flag $emission=9$ to reject them.

8.6. Parametrisation quality of K and M type giants

The parameterisation of cool stars with effective temperatures below 4000 K is known to be complex due to their crowded spectra, which results from the increasing presence of atomic and, especially, molecular lines. This aggravates normalisation issues and parameter degeneracies, in particular for metal-rich stars. During the GSP-Spec validation process, a correlation was found between the minimum flux value (F_{min}) of the spectra of giant stars with $T_{\text{eff}} \lesssim 4000$ K and their estimated $\log(g)$. In particular, in this cool temperature regime, objects with higher $\log(g)$ values present larger F_{min} values than expected when compared to those of slightly hotter giants with similar $\log(g)$ and S/N values. This reveals a parameterisation problem, as the pseudo-continuum should present lower values for cooler stars for which the line-crowding increases, and not vice versa. We have therefore implemented a specific flag ($KM\text{-typestars}$) that takes this issue into account. This flag depends on the F_{min} value and the gof in order to take account of the influence of the S/N on F_{min} . As reported in Table C.8, stars with $KM\text{-typestars}$ equal to 1 and 2 have corrected T_{eff} and $\log(g)$ with uncertainties reflecting the GSP-Spec parameterisation problems encountered for these stars: $T_{\text{eff}} = 4250 \pm 500$ K and $\log(g) = 1.5 \pm 1$.

8.7. Quality of individual chemical abundances

We checked the reliability of all the abundance estimates, including their uncertainties across the Kiel diagram ($\log(g)$ vs. T_{eff} plot) and taking into account the S/N. As a result of this process, we defined two flags for each individual abundance. Their definitions are given in Table C.9 and C.10 (with associated coefficients in Table C.11).

On one hand, as expected, the estimation quality depends on the strength of the spectral lines of the studied element, which varies with T_{eff} , $\log(g)$, $[M/H]$, and the abundance of the element itself. To help the user to deal with this effect, we implemented the individual abundance upper limit flag ($XUpLim$), which is an indicator of the line depth with respect to the noise level. This flag is based on an estimate of the detectability limit (*upper-limit*) that depends on the line atomic data, the stellar parameters, the line broadening, and the S/N. We note that, for the definition of this $XUpLim$ flag, we adopted a GSP-Spec internal estimate of the S/N that could slightly differ from the published $rv_expected_sig_to_noise$. The closer is the derived abundance to this upper limit, the higher the flag value and the abundances should therefore be used more cautiously.

On the other hand, for low-S/N spectra (with a limiting S/N depending on the analysed lines), the reliability of the associated abundance uncertainties can be underestimated. This is due to the fact that the maximum allowed abundance value in the reference grids is $[X/Fe]=2.0$ dex, preventing higher values in the abundance distribution associated with the flux noise Monte-Carlo realisations. This effect depends on the line detectability and the S/N. As a consequence, we defined a second individual abundance flag ($XUncer$) labelling the reliability of the associated abundance uncertainty taking into account its dependence on the stellar type (T_{eff} , $\log(g)$, and $[M/H]$), the S/N estimate, and/or the gof . Moreover, it is worth noting that the distance between the $[X/Fe]$ upper confidence level and the grid upper border is also a good indicator of the estimate reliability.

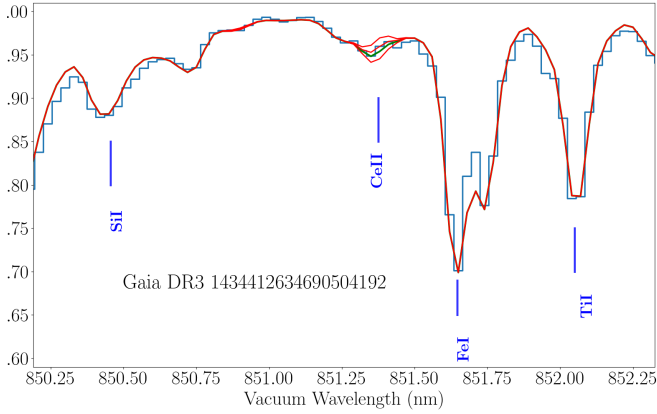


Fig. 9. Fit of the RVS spectrum (blue histogram) of the RGB star *Gaia* DR3 1434412634690504192 around its cerium line. The model in green corresponds to the GAUGUIN solution $[\text{Ce}/\text{Fe}] = 0.26$ dex (in excellent agreement with the literature value) whereas those in orange have $[\text{Ce}/\text{Fe}] = -2.0$ dex (almost no cerium) and ± 0.2 dex around the GAUGUIN abundance, respectively. The S/N is 907 and the broadening velocity is equal to 10.4 km s^{-1} . See text for more details.

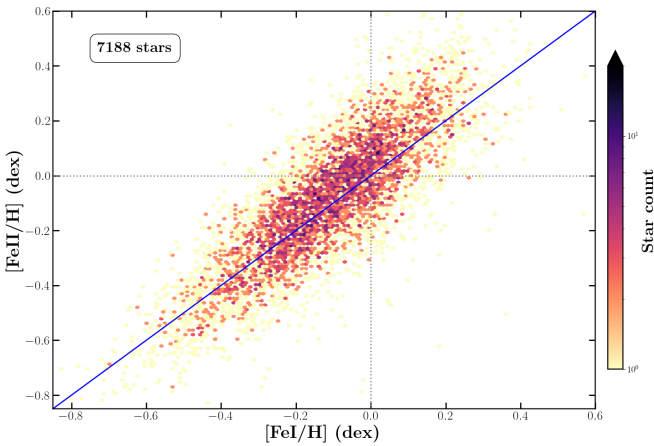


Fig. 10. Comparison between iron abundances measured from the proposed Fe II line at 858.79 nm and from all the other Fe I lines. The Spearman correlation coefficient is equal to 0.82. See text for more details.

8.7.1. Validation of heavy element abundances

An important illustration of the quality of the GSP-Spec abundance analysis with GAUGUIN is provided by the derivation of heavy element abundances, the estimation of which seemed too challenging for the RVS resolution. As an example, Fig. 9 shows the RVS spectrum of a red giant branch (RGB) star around its cerium line. The Ce abundance $[\text{Ce}/\text{Fe}] = 0.26$ dex with lower and upper confidence levels being 0.18 and 0.38 dex, respectively) was derived from the MatisseGauguin parameters: $T_{\text{eff}} = 4157 \text{ K}$, $\log(g) = 1.09$, $[\text{M}/\text{H}] = -0.4$ dex, and $[\alpha/\text{Fe}] = 0.12$ dex. The GSP-Spec abundance flags are $\text{CeUpLim} = \text{CeUncer} = 0$. This star was previously analysed by Forsberg et al. (2019) who derived a very consistent $[\text{Ce}/\text{Fe}] = 0.22$ dex, adopting very similar atmospheric parameters. It is important to note that GSP-Spec cerium abundances are on the same scale as those found by Forsberg et al. (2019) with a null median difference for the overlapping sample. This confirms the high quality of the GSP-Spec chemical analysis and of the *Gaia*/RVS spectra.

8.7.2. Validation of the singly ionised iron abundance

The specific case of Fe II abundances merits discussion. When building the GSP-Spec line list, Contursi et al. (2021) identified an unknown line at 858.79 nm (in the vacuum) and proposed that it is actually an Fe II feature. Because of its unblended nature in the RVS spectra of hot stars (see Fig. 6), it has been included in the line list used by GAUGUIN for the individual abundance analysis (Table B.1).

In Fig. 10, Fe II abundances are compared to Fe I ones in the atmospheric parameters regime where both estimates are possible at the same time. Both iron abundances were calibrated as suggested in Sect. 9. We selected stars with all 13 atmospheric parameter flags equal to zero together with a rather strict quality selection using the two abundance flags: $XU_{pLim} \leq 1$ and $XU_{uncer} = 0$. We also selected only stars in which the Fe II line is easily detected ($6000 < T_{\text{eff}} < 7200 \text{ K}$). The agreement between both iron abundances is excellent. The Spearman correlation coefficient is equal to 0.82 and increases up to 0.89 when selecting the ~ 2500 stars with $S/N > 300$.

We can therefore safely conclude that this 858.79 nm line is indeed a very good metallicity proxy and probably corresponds to an absorption produced by an iron-peak element, Fe II being the best candidate as suggested by Contursi et al. (2021).

8.8. Quality of cyanogen differential equivalent width ($\Delta\text{CN}q$)

To validate CN parameters, literature data were used to identify cool RGB and AGB stars for which CN lines are expected to be present. The flag associated with the EW of this CN abundance proxy ($\Delta\text{CN}q$) is defined in Table C.12; it depends on the three line-broadening flags ($vbroad$), the S/N, the gof , and the measured line position (p_1).

8.9. DIB quality flag ($DIBq$)

To quantify the quality of the DIB analysis, we defined a specific flag, ranging from $DIBq = 0$ (highest quality) to 5 (lowest quality). When no DIB is measured ($DIBq = 9$), another flag QF , not included in the flag chain, details the reasons as to why no measurements were performed (see Sect. 6.5). Its definition depends on the p_0 and p_2 parameters but also on the global noise level (Ra) defined by the standard deviation of the $(data - model)$ residual between 860.5 and 864 nm as well as on the local noise level Rb , that is, the $(data - model)$ residual within the DIB profile. Table C.13 explains the definition of the $DIBq$ flag and Fig. C.2 shows its flow chart. As discussed in Gaia Collaboration (2023b), we recommend the adoption of the most reliable DIB parameters ($DIBq = 0, 1, 2$) as well as (i) a good central wavelength measurement $862.0 < p_1 < 862.6 \text{ nm}$, (ii) a rather small uncertainty on the EW measurement ($\text{err}(EW)/EW < 0.35$), and (iii) a good stellar parametrisation (first 13 GSP-Spec flag being smaller than 2).

9. Known parameter and abundance biases

After the previous evaluation of the parameter quality through a flagging system, internal and external biases were studied, taking into account the implemented flags. The result of this analysis is presented in this section for MatisseGauguin atmospheric parameters and abundances (Sect. 9.1 includes a summary of the proposed solutions at the end) and for ANN atmospheric parameters (Sect. 9.2). Several figures and tables associated with this

section can be found in Appendix E. In some cases, simple calibrations with low-degree polynomials are suggested. It is worth noting that published DR3 GSP-Spec data are deliberately uncalibrated, and so users are able to (i) use the raw data that come from the GSP-Spec processing, (ii) apply, whenever suggested, the calibrations presented in this paper, and (iii) perform a new calibration tailored to their scientific analysis.

On one hand, although specific work has been done on the optimisation of the reference synthetic spectra grids, the observed biases can be partially due to mismatches between observations and reference synthetic spectra if some physical aspects not considered in the modelling (e.g. stellar rotation, macroturbulence, departures from local thermodynamic and hydrostatic equilibria) become non-negligible for some parameters of certain types of stars. This has been partially taken into account with parameter flags (e.g. *vbroadT*, *vbroadG*, *vbroadM* flags). We recall that the parametrisation of cool stars is often challenging (see e.g. Sect. 8.6 and Soubiran et al. 2022), and even higher resolution surveys in the literature can exhibit biases.

On the other hand, it is worth noting that the observed biases with respect to the literature can also have their origin in methodological and theoretical assumption differences with respect to those adopted in this work, such as different atmosphere models, atomic data, or reference solar abundances. In addition, several ground-based spectroscopic surveys have applied *ad hoc* offset corrections as a result of their calibration procedures. Finally, the presented global biases with respect to the literature depend on the relative proportion of stars in the various reference catalogues as a function of the S/N and the analysed parameter space.

Finally, it is important to mention that reference catalogues have their own biases. Although literature references are generally calibrated (while *Gaia* archive data are not), this does not remove all the existent trends, as shown by some recent works (e.g. Soubiran et al. 2022). As a consequence, it cannot be excluded that the observed trends in the comparison with external catalogues are partly due to biases that are still present in the literature data.

As a consequence of all the above mentioned points, the results of the bias analysis presented in the following have to be cautiously and thoroughly considered. We recommend that the user adapt any bias correction to the targeted scientific goal and selected sample.

9.1. GSP-Spec MatisseGauguin biases

The MatisseGauguin workflow produces both atmospheric parameters and individual chemical abundances. Estimation biases have been evaluated for each case and are presented in the two following subsections. We have chosen to present the $[\alpha/\text{Fe}]$ biases together with those of individual abundances, as the underlying spectral indicators are dominated by the Ca II IR triplet lines and, as a consequence, the behaviour of $[\alpha/\text{Fe}]$ is very similar to that of the $[\text{Ca}/\text{Fe}]$ abundance. Section 9.1.3 summarises the observed biases and the proposed solutions.

9.1.1. Analysis of T_{eff} , $\log(g)$, and $[\text{M}/\text{H}]$

In this section, we compare the GSP-Spec MatisseGauguin T_{eff} , $\log(g)$, and $[\text{M}/\text{H}]$ with the latest data releases of three major ground-based spectroscopic surveys, namely APOGEE-DR17 (Abdurro'uf et al. 2022), GALAH-DR3 (Buder et al. 2021), and RAVE-DR6 (Steinmetz et al. 2020). We filtered the literature samples based on both the associated uncertainties of the published parameters (≤ 500 K, 0.5, 0.3 dex, for T_{eff} , $\log(g)$ and

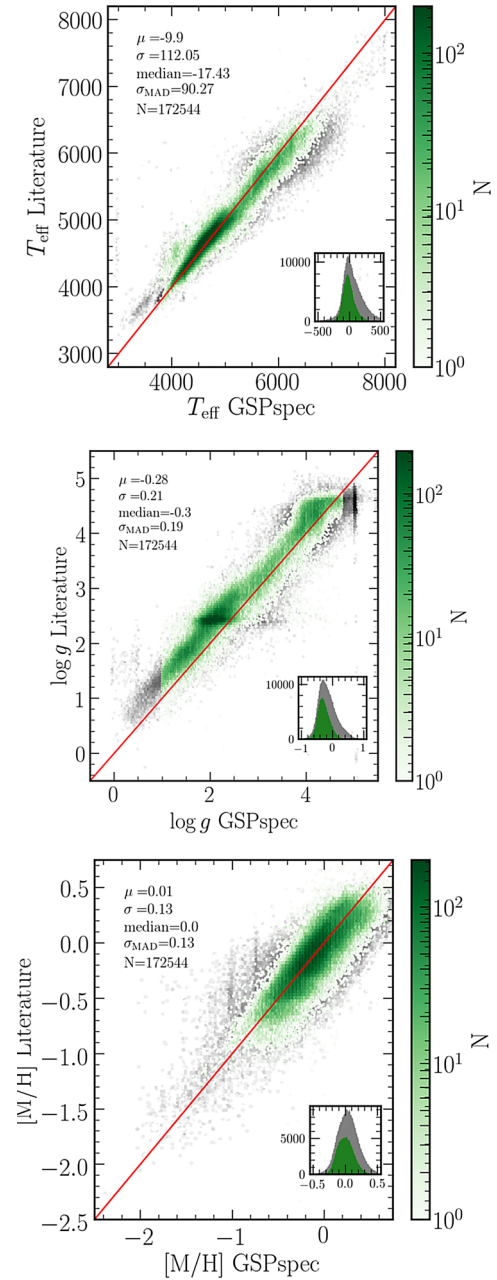


Fig. 11. Density plots comparing GSP-Spec MatisseGauguin parameters with literature data (APOGEE-DR17, GALAH-DR3, RAVE-DR6). Green and grey show the best- and medium-quality subsamples, respectively (see text for details about these samples). The histograms inside each plot show the difference between the literature and the GSP-Spec parameters. Mean (μ), standard deviation (σ), median, robust standard deviation (derived from the MAD), and the number of stars (N) of the offsets for the best-quality subset are annotated inside each box.

metallicity/iron abundance, respectively) and the reliability flags (following the suggestions of each of the respective surveys). In total, a sample of $\sim 8 \times 10^5$ stars (among which $\sim 7.5 \times 10^5$ unique targets) were selected in such a way. The three panels in Fig. 11 show how the main atmospheric parameters compare when all of the first 13 GSP-Spec flags are equal to zero (best quality sample, $\sim 1.7 \times 10^5$ stars plotted in green) and when we allow them to be smaller than or equal to one, except for the *KMgiantPar*, which we insist must be equal to zero and the *fluxnoise* flag