A&A 585, A93 (2016)
DOI: 10.1051/0004-6361/201425030
© ESO 2015

Astronomy & Astrophysics

Stellar parametrization from *Gaia* RVS spectra

A. Recio-Blanco[1], P. de Laverny[1], C. Allende Prieto[2,3], D. Fustes[4], M. Manteiga[5], B. Arcay[4], A. Bijaoui[1], C. Dafonte[4], C. Ordenovic[1], and D. Ordoñez Blanco[6]

[1] Laboratoire Lagrange (UMR 7293), Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, CS 34229, 06304 Nice, France
e-mail: arecio@oca.eu
[2] Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
[3] Universidad de La Laguna, Dept. de Astrofísica, 38206 La Laguna, Tenerife, Spain
[4] Departamento de Tecnologías de la Información y de las Comunicaciones, Universidade da Coruña, 15071 A Coruña, Spain
[5] Departamento de Ciencias de la Navegación y de la Tierra, Universidade da Coruña, 15011 A Coruña, Spain
[6] Geneva Observatory, University of Geneva, 51 ch. des Maillettes, 1290 Versoix, Switzerland

Received 19 September 2014 / Accepted 30 September 2015

ABSTRACT

Context. Among the myriad of data collected by the ESA *Gaia* satellite, about 150 million spectra will be delivered by the Radial Velocity Spectrometer (RVS) for stars as faint as $G_{\rm RVS} \sim 16$. A specific stellar parametrization will be performed on most of these RVS spectra, i.e. those with enough high signal-to-noise ratio (S/N), which should correspond to single stars that have a magnitude in the RVS band brighter than ∼14.5. Some individual chemical abundances will also be estimated for the brightest targets.
Aims. We describe the different parametrization codes that have been specifically developed or adapted for RVS spectra within the GSP-Spec working group of the analysis consortium. The tested codes are based on optimisation (FERRE and GAUGUIN), projection (MATISSE), or pattern-recognition methods (Artificial Neural Networks). We present and discuss each of their expected performances in the recovered stellar atmospheric parameters (effective temperature, surface gravity, overall metallicity) for B- to K-type stars. The performances for determining of [α/Fe] ratios are also presented for cool stars.
Methods. Each code has been homogeneously tested with a large grid of RVS simulated synthetic spectra of BAFGK-spectral types (dwarfs and giants), with metallicities varying from $10^{-2.5}$ to $10^{+0.5}$ the solar metallicity, and taking variations of ±0.4 dex in the composition of the α-elements into consideration. The tests were performed for S/N ranging from ten to 350.
Results. For all the stellar types we considered, stars brighter than $G_{\rm RVS} \sim 12.5$ are very efficiently parametrized by the GSP-Spec pipeline, including reliable estimations of [α/Fe]. Typical internal errors for FGK metal-rich and metal-intermediate stars are around 40 K in $T_{\rm eff}$, 0.10 dex in log(g), 0.04 dex in [M/H], and 0.03 dex in [α/Fe] at $G_{\rm RVS} = 10.3$. They degrade to 155 K in $T_{\rm eff}$, 0.15 dex in log(g), 0.10 dex in [M/H], and 0.1 dex in [α/Fe] at $G_{\rm RVS} \sim 12$. Similar accuracies in $T_{\rm eff}$ and [M/H] are found for A-type stars, while the log(g) derivation is more accurate (errors of 0.07 and 0.12 dex at $G_{\rm RVS} = 12.6$ and 13.4, respectively). For the faintest stars, with $G_{\rm RVS} \gtrsim 13-14$, a $T_{\rm eff}$ input from the spectrophotometric-derived parameters will allow the final GSP-Spec parametrization to be improved.
Conclusions. The reported results, while neglecting possible mismatches between synthetic and real spectra, show that the contribution of the RVS-based stellar parameters will be unique in the brighter part of the *Gaia* survey, which allows for crucial age estimations and accurate chemical abundances. This will constitute a unique and precious sample, providing many pieces of the Milky Way history puzzle with unprecedented precision and statistical relevance.

Key words. stars: fundamental parameters – Galaxy: stellar content

1. Introduction

The European Space Agency *Gaia* mission was successfully launched on December 19 2013 from the French Guiana Space Centre in Kourou. *Gaia* is an ambitious astrometric, photometric, and spectroscopic survey of a large part of the Milky Way: about 1% of the Galactic stellar content down to $V \sim 20$th magnitude will be observed several tens of times. *Gaia* will thus revolutionise our view of the Galaxy, together with our understanding of its formation and evolution history. Apart from the astrometric instrument, two low-resolution spectrophotometers (Blue and Red Photometers, BP/RP) and the Radial Velocity Spectrometer (RVS) are in operation onboard *Gaia*. Recent summaries of characteristics of these instruments, the main goals, and expected scientific impact can be found in de Bruijne (2012), Bailer-Jones et al. (2013)[1].

The analysis of the *Gaia* data is being performed by the Data Processing and Analysis Consortium (DPAC), composed of nine Coordination Units (CU). One of them, CU8 Astrophysical Parameters, is in charge of the classification and the parametrization of the observed targets (see Bailer-Jones et al. 2013). The CU8/Working Group that is responsible for the main parametrization of the RVS spectra is the Global Stellar Parametrizer – Spectroscopy (GSP-Spec). To complement this GSP-Spec parametrization, other CU8 modules (Extended Stellar Parametrizer, ESP) will also estimate atmospheric parameters from RVS spectra of more specific types of stars as emission-line stars (ESP-ELS), hot stars (ESP-HS), cool stars (ESP-CS), and ultra-cool stars (ESP-UCD) (see Bailer-Jones et al. 2013). GSP-Spec is composed of three research groups that have different and complementary expertises in automated stellar classification from spectral data. In this article, we present the parametrization performed within this working group for RVS spectra.

[1] See also http://www.cosmos.esa.int/web/gaia/science-performance

The RVS is a high-resolution integral-field spectrograph that will survey the whole sky at a rate of about 100 spectra per second, producing about 15 billion spectra during the mission. From the time of its preliminary real performances, the RVS will collect spectra with large enough signal-to-noise ratio (S/N) to derive the radial velocity of stars that are brighter than $G_{\rm RVS} \lesssim 16$ (i.e. about 150 million stars, $G_{\rm RVS}$ being the *Gaia* magnitude of the targets through the RVS filter). This limiting magnitude corresponds to $V \lesssim 17.3$ for a solar-type star (for the corresponding magnitudes in other photometric *Gaia* bands, see Table 2 and Fig. 4 presented in Sect. 3.2). Several tens of millions of stars will be observed by the RVS down to a magnitude of $G_{\rm RVS} \lesssim 13$, and about five million stars down to $G_{\rm RVS} \lesssim 12$.

The RVS will provide spectra in the CaII IR triplet region (from 847 to 871 nm[2]) at a spectral resolution of ∼11 200[3]. In addition to the CaII strong lines, weak lines of Si I, Ti I, Fe I, etc. are encountered in the RVS spectral range and for late-type star spectra. In hotter (A-F types) star spectra, weak lines of N I, S I, and Si I appear around the Ca II lines and the strong Paschen hydrogen lines (see Figs. 1 and 2). Even hotter ($T_{\rm eff} \gtrsim 15\,000$ K) stellar spectra contain lines of N I, Al II, Ne I, and Fe II, whereas the Ca II lines start to decrease, and some He I lines appear.

However, the *Gaia* commissioning phase has revealed that the RVS suffers from (i) a level of scattered light that is higher than expected and variable with time, and CCD position (mainly sunlight scattered around sunshields) and thus an increased noise for part of the spectra, together with (ii) a time-variable throughput loss due to mirror contamination that reduces the collected signal by a few tenths of a magnitude. This last issue is regularly corrected thanks to decontamination campaigns that reduce the loss to acceptable levels. Moreover, the *Gaia* DPAC has put in place a new version of the onboard software, resulting in a data collection scheme that is more robust against the stray light. This has mainly been achieved by lowering the autocollimation width of read-out windows in the RVS (and possibly the astrometric field) so that less noise from stray light is accumulated. The new video-processing unit software that makes this possible has already been uploaded to the satellite (Fleitas et al. 2015). In addition, following the actual RVS performances that were revealed by the commissioning phase, it has also been decided that every RVS spectra will be provided in nominal high-resolution mode to minimise the background contamination that is a function of the window width. Initially, a binning by a factor of three for stars fainter than $G_{\rm RVS} < 10$ was planned (e.g. Bailer-Jones et al. 2013), decreasing the effective resolution to around 7500. This possibility has now been definitely abandoned.

In this paper, the above mentioned post-launch RVS characteristics are taken into account and the new S/N-magnitude relation recently published by the European Space Agency has been adopted[4]. In addition, the expected final GSP-Spec performances are given for the actual RVS resolution of ∼11 200. Nevertheless, the influence of the effective resolution change, at a constant S/N value, is studied in Sect. 3.4.

The main goal of the RVS is to measure the radial velocity of the stars to get their full 3D space motions when combined with the *Gaia* proper motions. However, the RVS data will be also very useful to parametrize brighter stars observed by *Gaia* in complement to the parametrization that as been performed independently from the two more sensitive low-resolution spectrophotometers (see Liu et al. 2012 for an estimate of the expected parametrization performances with BP/RP data, together with the performance improvements presented in Bailer-Jones et al. 2013; and Andrae et al., in prep.). We point out that, in the present work, we do not consider any (a priori) input from the BP/RP parametrization for GSP-Spec, although this is one of the alternatives implemented in the *Gaia* global data-processing pipeline developed by CU8 and called Astrophysical parameters inference system (Apsis, see Bailer-Jones et al. 2013).

Since *Gaia* scans the whole sky, each target will be observed several tens of times depending on their location on the sky (with an average of 40 epochs per star for the RVS at the end of the mission, assuming the nominal five-year mission). As a consequence, the S/N of the combined spectra will increase with time during the mission and we will have to reparametrize these better quality spectra that are delivered by the successive releases. In the following, we consider RVS end-of-mission spectra that are a combination of the successive individual observations. It is expected that any star brighter than $G_{\rm RVS} \lesssim 14$ (i.e. several tens of millions) will be parametrized by GSP-Spec. The estimated stellar parameters will be the effective temperature ($T_{\rm eff}$), the surface gravity ($\log(g)$), the global metallicity ([M/H]), and the abundance of α-elements versus iron ([α/Fe]). In a second step, and whenever possible (depending on spectral type, metallicity, S/N, radial velocity shifts...), the individual chemical abundances of several species as Fe, Ca, Mg, Ti, and Si will be also measured. This should be performed for about five million sources with an expected accuracy of about 0.1 dex for $G_{\rm RVS} \lesssim 12$ because of a specific optimization method that is under development (Guiglion et al. 2014).

The amount of spectra to parametrize requires the use of automated methods that have to be fast enough and able to manage different types of stars. Therefore, GSP-Spec is currently composed of different independent codes, which are based on different algorithms. The advantage of having different codes is to get independent parameter estimates and to get the best parameters across the whole studied space (a given algorithm could provide excellent results for some parameter combinations and/or S/N, but rather poor results for others). Therefore, every RVS spectra will be parametrized by the various GSP-Spec codes, and most stars will be assigned to several parameter estimates with quality flags.

In this paper, we describe the GSP-Spec codes and their expected performance for different types of stellar populations as they have been implemented at the *Gaia* launch epoch. Increasingly, optimized versions of the GSP-Spec module of Apsis (see Bailer-Jones et al. 2013) are delivered at each operations cycle. GSP-Spec is expected to be running in operations cycle 4 in 2017, with a possible contribution from the third *Gaia* data release (planned around 2018).

In the following, we first present the codes specifically developed or adapted to the RVS stellar spectra by GSP-Specin Sect. 2 to estimate their stellar atmospheric parameters, together with their enrichment in α-elements with respect to iron. In Sect. 3, we then detail the methodology adopted for testing and comparing these different codes, and their GSP-Spec performances on simulated spectra at different RVS magnitudes are investigated and discussed in Sect. 4. The end-of-mission GSP-Spec expected parametrization is described in Sect. 5 for different types of stars and S/N. We then provide a comparison of the expected performances for the parametrization from BP/RP

[2] The RVS red edge has been shifted from 874 to 871 nm with respect to the original configuration, following a change in the RVS filter effective transmission. In the present work, we have thus adopted this 871 nm cut-off for the simulated spectra.

[3] The design specification of $R \sim 10\,500$ being exceeded.

[4] See http://www.cosmos.esa.int/web/gaia/sn-rvs

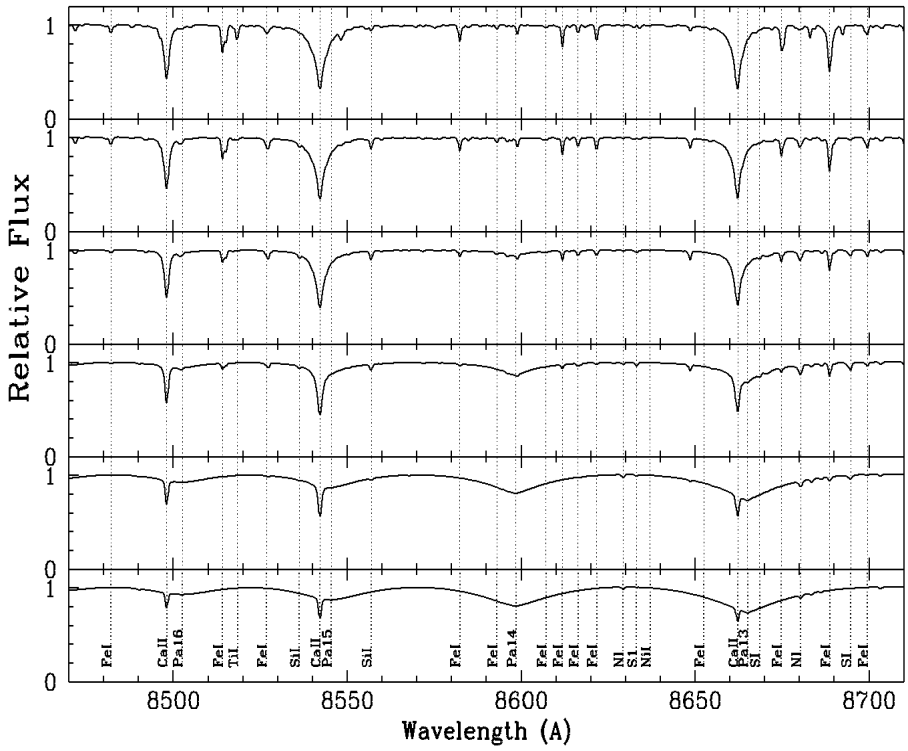

Fig. 1. Noise-free simulated RVS spectra for late B- to K-spectral-type metal-rich dwarf stars, included in the *reference* grids. The adopted effective temperatures are 11 000 K, 9000 K, 7500 K, 6500 K, 5500 K, and 4500 K *from bottom to top*, respectively. The other stellar parameters are kept constant at $\log(g)$ = 4.5 cm/s^2, [M/H] = +0.0 dex, and [α/Fe] = +0.0 dex. The identified lines refer to most of the strongest atomic transitions that are present in the different spectra.

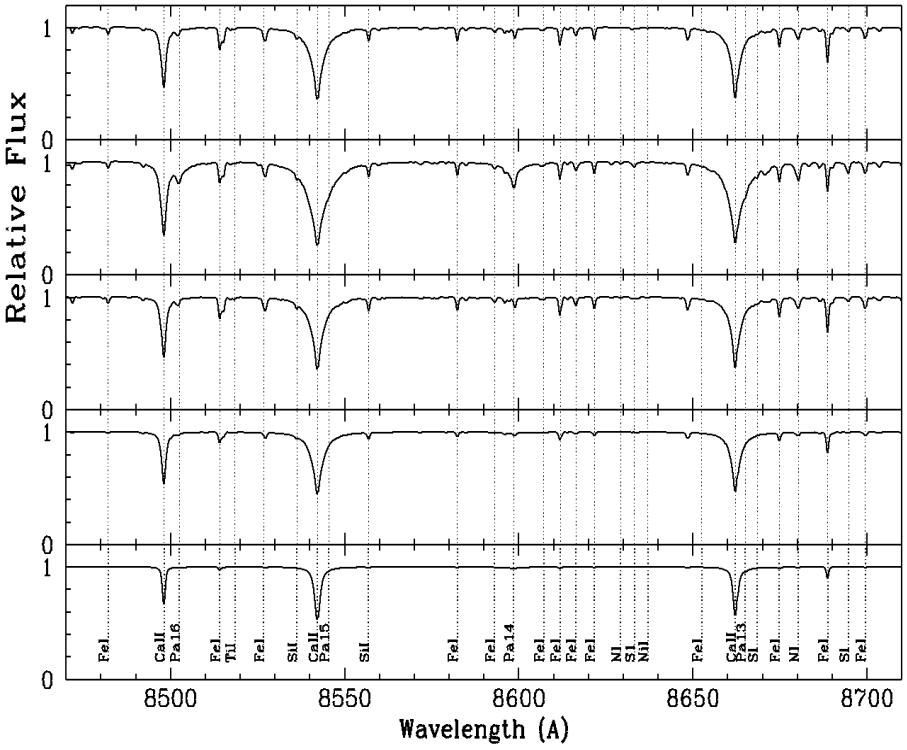

Fig. 2. Same as Fig. 1 but for cool stars only. Taking as reference a Solar spectrum defined by 5750 K, $\log(g)$ = 4.5 cm/s^2, [M/H] = +0.0 dex, [α/Fe] = +0.0 dex and $R \sim 11\,000$ (*top panel*), the following panels show the effect of (*from top to bottom*): a change in surface gravity to $\log(g)$ = 2.0 cm/s^2 (*second panel*); and changes in the chemical composition (*third, fourth, and bottom panels*, showing respectively, [M/H] = +0.5 dex and [α/Fe] = +0.2 dex; [M/H] = −1.0 dex and [α/Fe] = +0.4 dex; [M/H] = −2.0 dex and [α/Fe] = +0.4 dex).

photometry (Sect. 6) and from some ground-based spectroscopic surveys (Sect. 7). We finally conclude this work in Sect. 8.

2. The GSP-Spec automated parametrization codes

In this work, four different codes for estimating the atmospheric stellar parameters and the overall [α/Fe] chemical index (the latter, only for cool FGK-spectral type stars) have been tested for GSP-Spec: FERRE, GAUGUIN, MATISSE, and Artificial Neural Networks (ANN). These codes have already been extensively described in previous papers. We hereafter refer to the indicated references for their detailed description. Below, we provide only a brief summary of these codes, and we mainly focus on the specificities that have been developed for GSP-Spec. Furthermore, we point out that most of these codes have already been used to separately parametrize real stellar spectra, which have been collected at a similar resolution and over the same wavelength region as the ones of the RVS (see references below).

However, the present work is the first to illustrate and compare their performances in the context of *Gaia*.

Within the different data-mining approaches that have been developed so far, GSP-Spec parametrization algorithms belong to the class that uses the mapping with reference models in a continuously variable space. In addition, they belong to the three main families of parametrization: optimization methods, projection methods, and pattern recognition.

- The optimization codes (FERRE and GAUGUIN) perform a distance minimization between the full input spectrum and the reference spectra grid.
  FERRE[5] (Allende Prieto et al. 2006) is a FORTRAN90 code that compares the χ^2 between models and observations to identify the optimal set of atmospheric parameters (and/or abundances) of a star. The search is performed using the Nelder-Mead (1965) algorithm. The model evaluation is based on linear interpolation in a pre-computed grid, which is held on the computer's memory for speed. The observed and (interpolated) model spectra are forced to have the same mean value. Multiple searches are performed for each spectrum, initialized at 100 random positions in the grid. The code is parallelized over multiple cores using OpenMP and is made available upon request.
  GAUGUIN (Bijaoui et al. 2012) is a classical, local optimization method implementing a Gauss-Newton algorithm. It is based on a local linearization around a given set of parameters that are associated with a reference synthetic spectrum (via linear interpolation of the derivatives). A few iterations are carried out through linearization around the new solutions, until the algorithm converges towards the minimum distance. In practice, and to avoid trapping into secondary minima, GAUGUIN is initialized by parameters independently determined by other algorithms such as MATISSE (this is noted hereafter as MATISSE$_G$, see below). GAUGUIN is part of the analysis pipeline of the *Gaia*- ESO Survey (GES, Gilmore et al. 2012) for the GIRAFFE spectra (Recio-Blanco et al., in prep.).
- The MATISSE algorithm (Recio-Blanco et al. 2006) is a projection method in the sense that the full input spectra are projected into a set of vectors derived during a learning phase, based on the noise-free reference grids (see Sect. 3.1). These vectors are a linear combination of reference spectra and could be viewed roughly as the derivatives of these spectra with respect to the different stellar parameters. MATISSE is, thus, a local multi-linear regression method. Furthermore, a two-step procedure in the parameter estimation has been implemented within MATISSE to tackle non-linearity problems. Other recent applications of MATISSE to very large amounts of real observed spectra are the *AMBRE* project (de Laverny et al. 2013; Worley et al. 2012), the RAVE fourth Data Release (Kordopatis et al. 2013), and the *Gaia* – ESO Survey (see Recio-Blanco et al., in prep.).
- Finally, the ANN code uses a pattern recognition approach to parametrize the spectra.
  The implementation of the ANN method for GSP-Spec has already been presented in Manteiga et al. (2010). To summarise, the architecture is a feed-forward network with three layers, trained with the online error-back propagation algorithm. We generate one network for each stellar parameter where the number of neurons in the input layer coincides with the number of points of the format that was selected as signal representation. The output layer consists of the parameter to be predicted. The activation function for the hidden and output neurons of the network is a sigmoidal function. We also set the number of hidden neurons as the minimum between half the dimensionality of the input signal and 200 hidden units at maximum to reduce the computational burden of the training process. To deal with the initialization dependence and possible local minima, a series of training procedures are performed, until a near-optimal value is reached. On the other hand, early stopping is also used by means of a validation dataset to avoid the problem of over fitting, so that the network state which best generalizes is kept. Furthermore, to generalize the application to random spectra, we use 100 noised reference grid spectra during the training phase (see below). Finally, we made use of the wavelet decomposition to obtain new signal representations (low-pass filter) which are used as the network inputs, both in the training and testing phases. In practice, this means that the results of this code are obtained by adopting the approximations of the first, second, or third level in the wavelet pyramid (depending on the S/N value) as input for the ANN instead of the full spectra.

Table 1. Summary of technical aspects of the tested codes or algorithms.

	Training phase	Programming language	Already implemented in Apsis
FERRE	No	Fortran	No
GAUGUIN	No	Fortran & Java	Yes
MATISSE	Yes	Fortran & Java	Yes
ANN	Yes	Java	No

As outputs, these four codes provide the three atmospheric parameters ($T_{\rm eff}$, $\log(g)$, and [M/H]), and the [α/Fe] chemical index (for FGK stars only) estimates, together with their associated uncertainties (except for the ANN). For some of them (FERRE, GAUGUIN, MATISSE), a quality control parameter, which is based on the goodness of the fit between the input spectrum and an interpolated one at the estimated parameters, is also provided. In Table 1, we report more details on some technical aspects of these codes.

3. Adopted methodology for quantifying the code performances

We present here the implemented homogeneous tests, including the used data and the analysis methodology, that estimate and compare the expected performances of the different GSP-Spec codes. The ultimate goal is to determine the optimal application fields of the different codes in terms of stellar types and S/N. This will lead to the definition of the quality flags that will be assigned to the different parameter estimates within the GSP-Spec pipeline.

In this section, the data used for the tests and the considered methodology for the analysis of the results are described. The different subsections present the steps of the procedure that was followed. In particular:

- The codes have been trained (when necessary, depending on the algorithm strategy, see Sect. 2) using large grids

[5] The code is now public and available at: http://hebe.as.utexas.edu/ferre

of noise-free, RVS simulated synthetic spectra that are described in Sect. 3.1[6].
- Then, to perform a homogeneous comparison of the methods, parametrization tests have been performed using noised random grids. The six adopted S/N values are 350, 150, 125, 40, 20, and 10. Those test grids contain interpolated spectra with arbitrary parameter values (see Sect. 3.2) that span the whole grid parameters range.
- A subsample of the previously noised random spectra, excluding non-physical parameter combinations and correctly populating the Hertzsprung-Russel diagram, has been selected (cf. Sect. 3.3).
- Finally, we have tackled the problem of the recently abandoned RVS spectra rebin for stars with $G_{\rm RVS} > 10$ (cf. Sect. 3.4). The influence of this change in the GSP-Spec performances is quantified thanks to specific tests with one of the tested algorithms.

In this work, we have favoured the use of synthetic spectra instead of real ground based observed ones (RVS data not being available) to explore any possible combination of the four parameters over a wide range of possible values and to keep a good homogeneity between the tests. We are fully aware that synthetic spectra may not be perfectly realistic, when compared to observed stars for some parameter combinations, but this will not affect the comparison of codes. Of course, the application to real observed spectra will lead to larger errors (external ones) mostly because of the possible mismatches that could exist between synthetic and observed spectra. These effects will be quantified (and possibly corrected) during the mission as a result of calibration with *reference stars* that will be (or already has been for some of them) accurately parametrized (see Bailer-Jones et al. 2013).

Finally, in the following we neglect any effects that could be induced by wrong normalization or radial velocity corrections of the input spectra. Within the *Gaia* analysis pipeline, such problems will be examined, in collaboration with GSP-Spec, by the CU6 in charge of providing RVS spectra to CU8. The synthetic spectra described below are all, therefore, at the rest frame and normalized. However, as the estimation of the atmospheric parameters and individual abundances can be quite sensitive to the pseudo-continuum normalization, an automated renormalization of the input spectra has already been implemented within GSP-Spec. Indeed, the RVS spectra can be renormalized through an iterative procedure coupled with the atmospheric parameters estimates. Such an iterative procedure has already be shown to be very successful when automatically applied to real spectra (Kordopatis et al. 2011, 2013; Worley et al. 2012, and the GES pipeline, Recio-Blanco et al., in prep.). More specifically, we also point out that Kordopatis et al. (2011) have already shown that parametrization errors that are induced by normalization uncertainties as large as 3% of the continuum level for RVS-like spectra can be neglected if such an iterative renormalisation is implemented. This is however not considered in the following tests since we focus on already perfectly normalized input synthetic spectra.

3.1. Grids of reference spectra

The reference grids are a collection of noise-free, high-resolution normalized synthetic spectra that have been computed from Kurucz model atmospheres. We favoured these Kurucz models, contrary to previous works based on MARCS model atmospheres (see, for instance, de Laverny et al. 2012) because they allow us to consider consistent spectra for cool FGK and hot BA-spectral-type stars. The original computed spectra preserved the continuum slope, but then we continuum-normalized each spectrum by (i) fitting iteratively a line and σ-clipping ten times the points that fell between 0.1σ and 3σ below and above the fit, respectively; and (ii) dividing the original spectrum by this final fit. We refer in the following to such *normalized* spectra, in which the slope in the spectrum, within the RVS wavelength domain, is not conserved. In practice, thanks to the estimates of stellar distance, and after accurate flux calibration, it could also be possible to analyse *Gaia*/RVS absolute flux spectra. This possibility is however not considered in this work.

With regard to the grid calculations, these were computed using Castelli & Kurucz (2003) model atmospheres and a line list from Kurucz[7] that was enhanced with damping constants from Barklem et al. (2000) for atomic transitions. The calculations were done using the ASSET synthesis code (Koesterke et al. 2008; Koesterke 2009), sampling the spectra with steps corresponding to 1 $\mathrm{km\,s^{-1}}$. The values of the solar reference abundances in the synthesis are from Asplund et al. (2005), while those of the corresponding atmospheric structures are from Grevesse & Sauval (1998). The abundances of the α-elements were changed for the synthesis, but not in the model atmospheres.

Four parameters were considered for building the FGK-spectral type grid. These parameters are the effective temperature $T_{\rm eff}$ that varies from 4000 to 8000 K (with a step of 250 K), the surface gravity $\log(g)$ varying between 2.0 to 5.0 $\mathrm{cm/s^2}$ (step of 0.5 dex), the mean metallicities varying from $10^{-2.5}$ to $10^{+0.5}$ the solar metallicity (with a step of 0.5 dex), and variations of $\pm$0.4 dex in the enrichment in the α-chemical species with respect to iron (step of 0.2 dex). This cool-star reference grid contains 5 831 spectra.

For hotter stars (from late B- to F-spectral types), only the first three parameters were considered since the spectra become almost metal line-free with increasing $T_{\rm eff}$. The effective temperature for this hot-star grid varies from 7000 to 11 500 K (step of 500 K). The surface gravity and mean stellar metallicity ranges together with their variation steps are identical to those of the cool-star grid (without variations in [α/Fe]). We end up with a reference grid of 490 hot stellar spectra. We point out that any possible combination of the above-mentioned stellar parameters has been considered to build these "square" cool and hot grids (i.e. without gaps in any of these three or four parameters).

Furthermore, these very high-resolution grids have then been degraded to mimic the RVS instrumental effects by convolution with a Gaussian profile for the spectral resolution ($R \sim 11\,200$) and adopting a sampling of ~0.024 nm/pixel (1 125 pixels in total). Additionally, a sampling of ~0.072 nm/pix (375 pixels) was considered to produce two low resolution grids ($R \sim 7500$). They were used to analyse the influence of the post-launch effective resolution change (cf. Sect. 3.4).

In the following, we will refer to these two RVS synthetic spectra grids as the *reference* grids. For illustration, some spectra corresponding to the late B- to K-spectral types are shown in Fig. 1, together with the identification of some of their strongest lines. Moreover, Fig. 2 shows some cool stars' spectra that are representative of different Galactic populations. Finally, the

[6] We point out, however, that the non-trained codes also use the spectra grid for their parametrization, based on a fitting process.

[7] From his website `kurucz.harvard.edu`

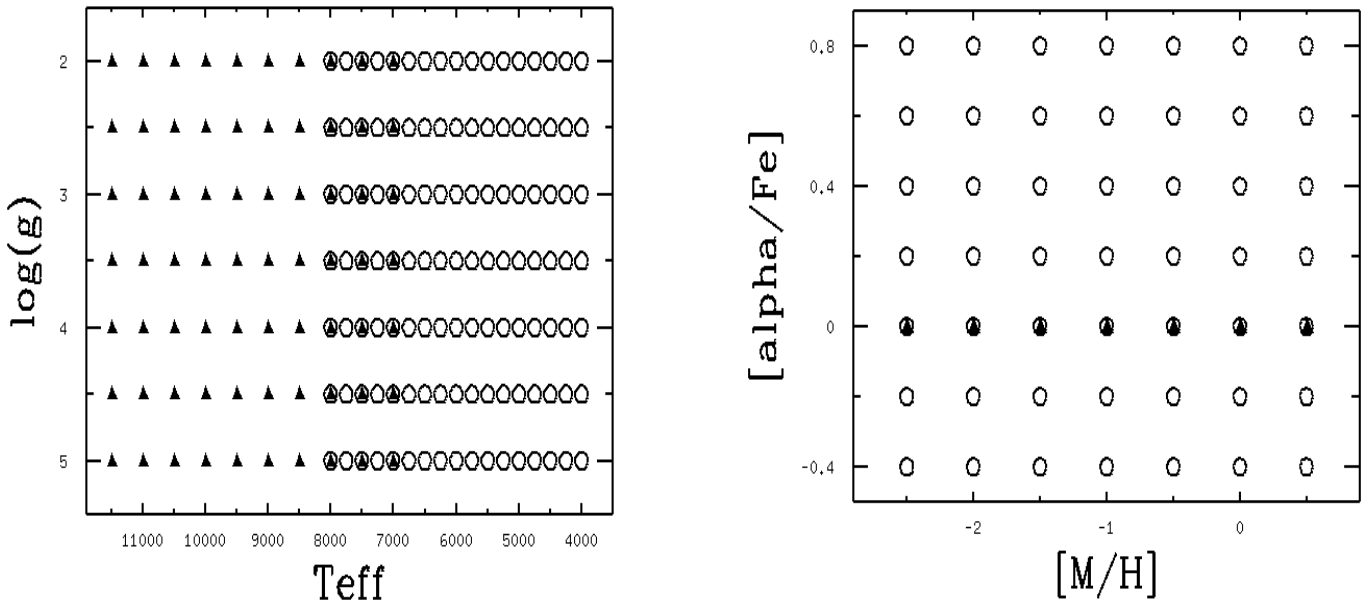

Fig. 3. Distribution of the *reference* synthetic spectra in the atmospheric parameter space ($T_{\rm eff}$, log(g), [M/H] and, [α/Fe] are in K, cm/s^2, dex, and dex, respectively). The cool-star and hot-star grids are shown with open circles and filled triangles, respectively. Any combination of the shown stellar parameters has been considered when building these grids, except for the hot-star grid in which no variations in [α/Fe] have been considered (see filled triangles in the right panel).

distribution of the atmospheric parameters and the [α/Fe] chemical index of the reference grids are shown in Fig. 3.

3.2. Grid of noised random spectra

From the *reference* high-resolution ($R \sim 11\,200$) and low-resolution ($R \sim 7500$) noise-free grids described above, interpolated noised synthetic spectra have been generated to test the codes.

The interpolations were performed at random combinations of the four parameters. The S/N was then simulated by adding noise (see below) to these interpolated spectra. The six adopted S/N values are 350, 150, 125, 40, 20, and 10. They correspond to the RVS filter magnitudes $G_{\rm RVS}$ values equal to 8.4, 10, 10.3, 11.8, 12.6, and 13.4 according to the most recent RVS performances, including post-launch studied instrumental effects, like stray light contamination, actual line spread function profiles, effects of window decentring, and light loss.

We again point out that this added noise will allow us to estimate the parameter uncertainties caused by internal and instrumental errors and not the external errors, which are mainly dominated by possible synthetic spectra mismatches. Moreover, we report the magnitudes in the G and V-bands corresponding to $G_{\rm RVS} = 13.5$ for different stellar types in Table 2 and the ($G - G_{\rm RVS}$) versus ($V - G$) relation for the range of $T_{\rm eff}$ studied in the present work in Fig. 4. These estimated V, G, and $G_{\rm RVS}$-magnitudes and associated colours are based on the colour transformations provided by Jordi et al. (2010) together with the photometric colours – $T_{\rm eff}$ relations of Ramírez & Meléndez (2005) and Boyajian et al. (2012) for cool and hot stars, respectively.

Practically speaking, we have simulated end-of-mission RVS spectra, based on the instrument performance information available to us. The CCD windows for the spectra, and therefore their noise properties, depend mainly on the source brightness. On the

Table 2. V and G magnitudes corresponding to $G_{\rm RVS} = 13.5$ for some stellar types as defined in Sect. 5.

Stellar type	V (mag)	G (mag)
B dwarf	13.69	13.65
A dwarf	13.83	13.77
F metal-poor dwarf	14.25	14.12
F metal-rich dwarf	14.32	14.18
G metal-poor dwarf	14.54	14.33
G metal-rich dwarf	14.76	14.48
K metal-poor dwarf	15.06	14.66
K metal-rich dwarf	15.37	14.82
G metal-poor giant	14.54	14.33
G metal-rich giant	14.76	14.48
K metal-poor giant	15.14	14.70
K metal-rich giant	15.52	14.89

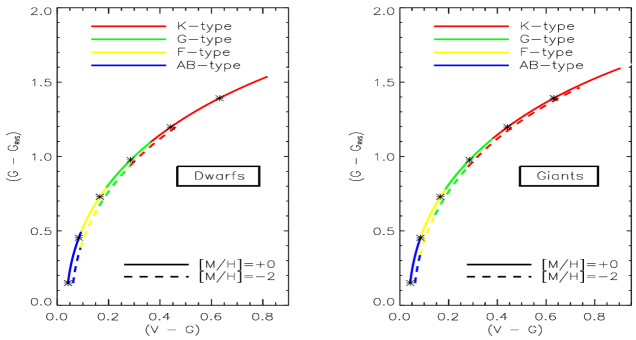

Fig. 4. ($G - G_{\rm RVS}$) versus ($V - G$) relation for the stellar types (colour-coded) defined in Sect. 5. For clarity reasons, the dotted lines (metal-poor stars) have been slightly (+0.02) horizontally shifted and the dwarf and giant regimes separated. The ($G - G_{\rm RVS}$) and ($V - G$) indexes only vary with the ($B - V$) colour index but the ranges covered in both axis are dependent on the metallicity. The crosses along the curves refer to ($B - V$) varying from 0.0 to 1.5 (step of 0.25) from bottom to top.

other hand, we take into account that the binning in the spectral direction corresponds to one RVS pixel[8].

The average signal per binned pixel in one RVS spectrum I (single visit and single CCD) is approximately defined by the brightness in that band as $G_{\rm RVS} = -2.5{*}\log(I \text{ x } q) + 22.5866$, where q, the number of binned pixels, is equal to 1260 for the high resolution and $q = 420$ for the low one. We account for Poisson shot noise in the data, and the CCD read-out noise, assumed to be 4 e-. Since the final RVS spectra will be accumulated from a variable number of visits, depending mainly on the location of a source in the sky, and since objects typically cross three CCDs per visit, we simulate individual observations (spectra acquired per CCD per visit), and then combine them to produce an end-of-mission spectrum for each source.

The final high- and low-resolution noised grids contain 20 000 random spectra at each selected $G_{\rm RVS}$ magnitude. They cover a wide range of atmospheric parameter values, including some non-physical combinations (unreal stars). Actually, 10 000 random spectra have been interpolated in each high/low resolution and cool/hot *reference* grids, having stellar parameters independent of the $G_{\rm RVS}$ values. Moreover, another 100 random noised spectra were selected to train the ANN code independently (see Sect. 2). These grids are made available in FITS format upon request to the authors.

3.3. Final definition of noised random spectra samples

To determine the code performances, we selected a subsample of the previously described 20 000 random spectra, based on the following criteria:

- We first retrieved PARSEC v1.1 isochrones from Bressan et al. (2012) to correctly populate the Hertzsprung-Russell diagram (in terms of stellar parameters combinations, and not of stellar lifetimes) as expected for a galaxy like the Milky Way, with different stellar populations (assuming a standard initial mass function over a long period of star formation). The selected isochrones correspond to ages of 1 and 13 Gyr and metal content $Z = 10^{-4}$ and 0.06 (i.e. [M/H] = −2.2 and +0.7 dex).
- We then selected spectra that have a ($T_{\rm eff}$, $\log(g)$) combination located between these two isochrones from among the 20 000 interpolated spectra described in Sect. 3.2.
- Finally, for any selected ($T_{\rm eff}$, $\log(g)$) combination, we chose all the available interpolated spectra (whatever their [M/H] and [α/Fe] values are). These criteria led to the selection of 9067 random spectra.
- In addition, and to consider the few hot giant stars that could be present in the RVS surveyed volume, we also retrieved an isochrone that is representative of younger stars (100 Myr) with solar metallicity ($Z = 0.19$). We then selected all the giants having (i) $T_{\rm eff}$ that varyies between 9000 and 11 000 K and; (ii) a surface gravity comprised of between 3.5 cm/s^2 and the isochrone gravity values. This procedure added 898 stars to the total sample selected.

The final high and low resolution *random* grids (as called hereafter) are composed of 9965 noised random spectra at each of the six selected $G_{\rm RVS}$ or S/N-values. Their location in the Hertzsprung-Russell diagram is shown in Fig. 5. Their distribution in metallicity and α-enrichment is perfectly flat over the ranges [−2.5, +0.5] and [−0.4, +0.8], respectively, because of their random nature.

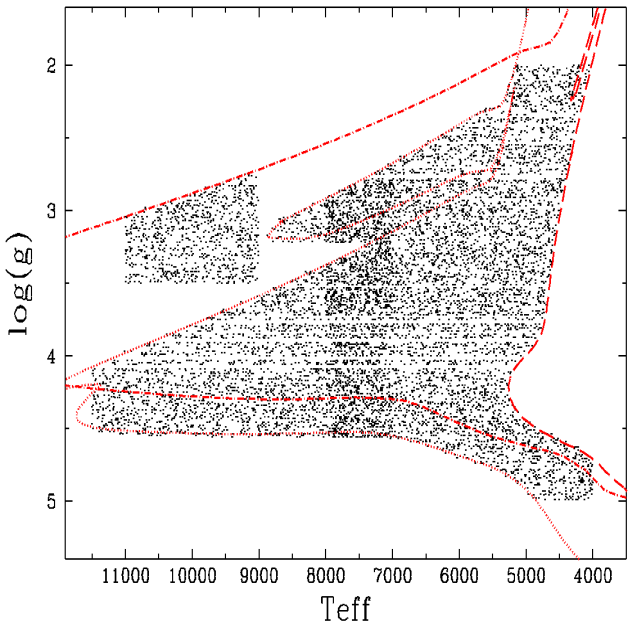

Fig. 5. Hertzsprung-Russell diagram of the random spectra (black dots) produced for the code performance tests ($T_{\rm eff}$ and $\log(g)$ are in K and cm/s^2, respectively). The isochrones in long-dashed, dotted, and dash-dotted lines correspond, respectively, to $Z = 0.06$ and 13 Gyr, $Z = 10^{-4}$ and 1 Gyr, and $Z = 0.019$ and 100 Myr. The higher density of *random* spectra between 7000 and 8000 K is a consequence of the temperature overlap between the two different *reference* grids, for hot and cool stars (see Sect. 3.1), from which the random spectra have been interpolated.

3.4. Influence of the post-launch abandoned RVS spectra rebin on GSP-Spec performances

As explained in the Introduction, following the actual RVS performances revealed by the commissioning phase, it has been decided that every RVS spectra will be provided in the nominal high-resolution mode to minimize the background contamination. Before launch, a binning by a factor of three for stars fainter than $G_{\rm RVS} < 10$ was planned (e.g. Bailer-Jones et al. 2013), decreasing the effective resolution of their spectra to around 7500. The parametrization algorithms tested for GSP-Spec have therefore faced the problem of correctly evaluating this recent effective resolution change in their performances.

In this work, we have decided to analyse and to report the influence of this post-launch revision, while providing up-to-date performance expectations that agree with the actual nominal RVS configuration. The resolution change issue has been tackled through the following steps:

- One of the codes, MATISSE$_G$ (already integrated in the Apsis *Gaia* DPAC pipeline) was run both on all the noised random spectra with the nominal resolution of $R \sim 11\,200$, and on their corresponding rebinned spectra with effective resolution $R \sim 7500$. The four left panels of Fig. 6 show the MATISSE$_G$ results for cool random spectra of $S/N \sim 20$ at the nominal high resolution (HR) compared to those for the rebinned lower resolution (LR) spectra of the same stars and same S/N value. No particular trends are found between

[8] Note that the RVS CCDs, as well as the other *Gaia* instruments, are operated in TDI mode, and hence talking about pixels (or *sample* as often used in *Gaia* literature) is not accurate, since the signal in any sample of a spectrum has been accumulated over many pixels as the star crosses the focal plane, but we will use this term by way of analogy.

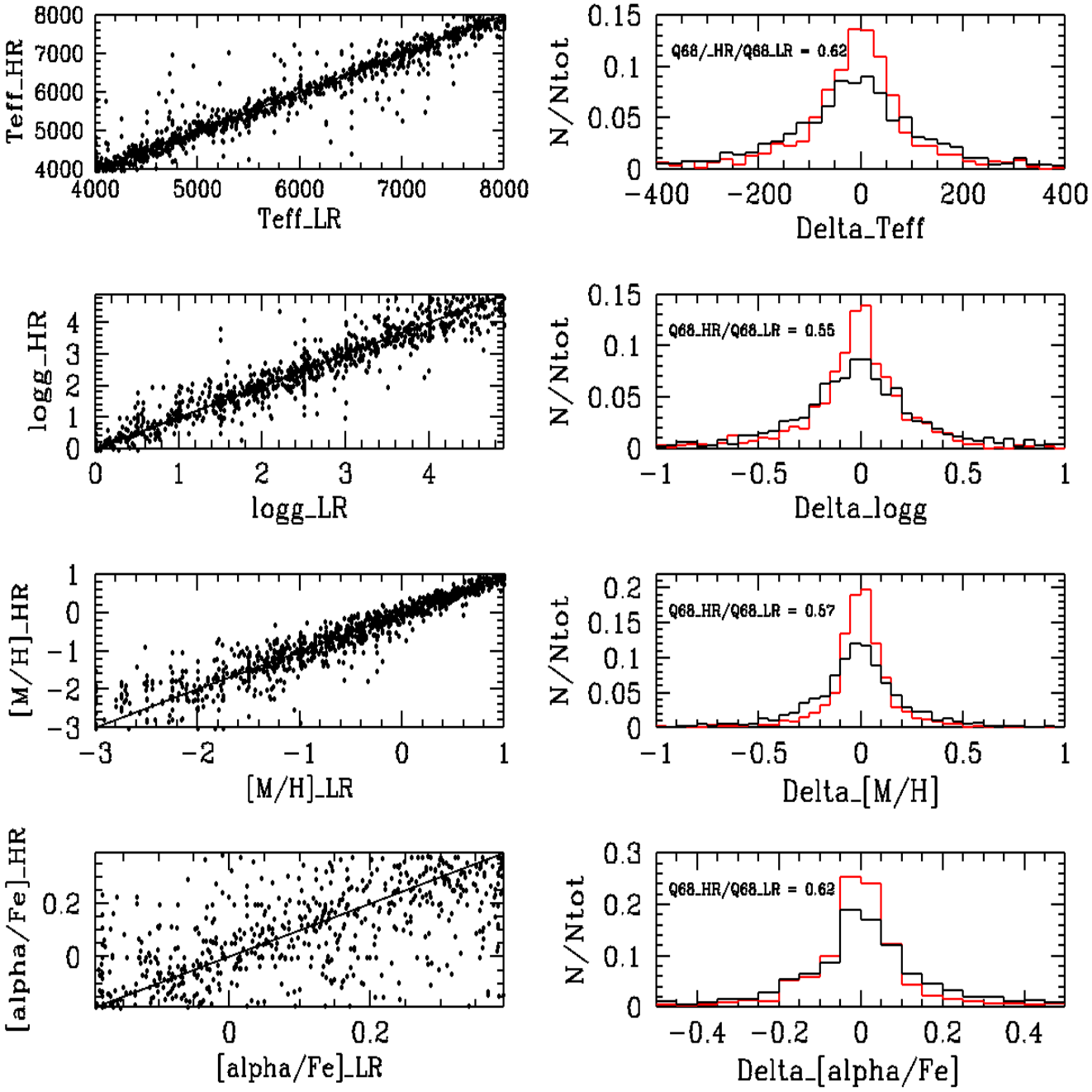

Fig. 6. *Left panels*: MATISSE$_G$ results for cool random spectra of $S/N \sim 20$ at the nominal high resolution (HR) compared to those for the lower resolution (LR) spectra of the same stars (and same S/N value). *Right panels*: normalized distribution of residuals for the high-resolution spectra (red curve) and for the low-resolution ones (black curve). The ratio of the 68% quantile values of both distributions is given for each stellar parameter.

both solutions, justifying the possibility of correcting the parametrization performances for this rebinning change in the input data.

– The influence of the resolution in the parametrization precision was quantified by estimating the variation of the 68% quantile of the error (residuals) distributions (see also Sect. 4.2) resulting from the resolution change only. The four right panels of Fig. 6 present normalized distribution of residuals for the $S/N \sim 20$ high-resolution spectra results and for the low-resolution ones of cool random stars. The ratio of the 68% quantile values of both distributions is given for each stellar parameter. Similar ratios have been estimated for all the S/N values considered in this paper (cf. Table 3). In general, we can see that the gain in the precision of the parameters at constant S/N, when passing from the low resolution rebinned spectra to the nominal high resolution ones, is of about one third.

– The results of the FERRE and ANN models which had been trained on rebinned low-resolution spectra for S/N 125, 40, 20, and 10 (fainter stars, as expected before launch), could now be rescaled to what we would expect if they had been trained on high-resolution spectra. The performances of the

Table 3. Ratio of the 68% quantile values of the error distributions for nominal high-resolution data and low effective resolution data (in the sense $Q68_{HR}$ divided by $Q68_{LR}$), as estimated with MATISSE$_G$.

S/N	125	40	20	10
G_{RVS}	10.3	11.8	12.6	13.4
T_{eff}	0.62	0.66	0.62	0.62
log(g)	0.44	0.48	0.55	0.51
[M/H]	0.62	0.62	0.57	0.68
[α/Fe]	0.72	0.73	0.68	0.83

Notes. The results for different parameters and S/N values are presented.

codes for the fainter stars, quantified through the 68% quantile of the residuals distribution, have been corrected by multiplying by the high-to-low resolution ratios derived in the previous step for each stellar parameter and each S/N value.

In summary, thanks to the above described synthetic spectra samples that consider the most up-to-date S/N-magnitude relation, and dealing properly with the actual RVS configuration and resolution, the parametrization tests presented here can be used to confidently estimate the future GSP-Spec performances.

4. Performance comparison of the different parametrization codes

We present in this section the performances of the tested methods for the sample of noised random FGK-type and BA-type stars defined in Sect. 3.3. In the following, the reported and discussed performances are those obtained by the FERRE and ANN codes, together with those of MATISSE, which has been locally improved by GAUGUIN (referred to as MATISSE$_G$, hereafter). In particular, the erosion of the performance of these codes, as the information contained in the spectra decreases (for example, for increasing noise and lack of spectral signatures) is analysed here, to understand the behaviour of each method and its best applicability domain.

4.1. Distribution of residuals

To evaluate each code's performance, we first computed the differences between the recovered (i.e. *estimated*) and real (*input*) atmospheric parameters, $\Delta\theta = \theta_{rec} - \theta_{real}$ with θ referring to T_{eff}, log(g), [M/H] and [α/Fe]. The $\Delta\theta$ will be referred as the *residuals*, hereafter.

Figures 7 and 8 show the distribution of these $\Delta\theta$ residuals obtained with the three methods that were tested for the random spectra of cool and hot stars, respectively. Both figures illustrate the results obtained at different G_{RVS} magnitudes (G_{RVS} = 8.4, 10.3, and 12.6, which correspond to S/N values of 350, 125, and 20).

It can be seen that the residual distributions are always very peaked and depart from a perfect Gaussian-like distribution only at the faintest magnitudes (the fact that both figures are in logarithmic scale has to be taken into account). Almost no outliers (spectra that are parametrized with an error that is well outside the main distribution) are seen. Moreover, these distributions are not biased at all, except for the faintest hot-star spectra, where small biases appear only for some methods.

Moreover, we notice that, at a given magnitude, all the methods recover the four atmospheric parameters with a rather similar quality. The performances are particularly excellent for the best-quality spectra (G_{RVS} = 8.4, which is also true as long as $G_{RVS} < 10$, see Fig. 9 for instance). The residual distributions get wider as the noise increases, although the large majority of the spectra (see, for instance, the discussion on the $Q68_\theta$ below) are always recovered with an acceptable accuracy, even at G_{RVS} = 12.6. As expected, the parametrization in T_{eff} and [M/H] of the hottest, faintest stars is of poorer quality, since these spectra lack sensitive spectroscopic signatures (see Fig. 1 where few atomic lines are seen, except the Ca II triplet, when $T_{eff} \gtrsim 8000$ K). However, the appearance of the broad Paschen lines still allows a very good estimate of the stellar surface gravity, even for very low quality spectra. By contrast, the surface gravity of the late-type stars is always the most difficult parameter to recover as already shown in several previous studies.

4.2. Quantification of method performance

To quantify each method performance, the 68% quantile of the $\Delta\theta$ distributions ($Q68_\theta$, hereafter) was adopted. This quantile can be viewed as the 1σ error of the parameter recovery in the case of a perfect Gaussian distribution of $\Delta\theta$. This is, however, not always the case, in particular, for low S/N. Another statistical estimate of the performances is the systematic error (or bias) that corresponds to the mean of the differences between the recovered and real parameters ($\langle\Delta\theta\rangle$). In our case, the biases are almost always very small for every method, compared to the $Q68_\theta$ quantiles (see previous subsection). They can, thus, be neglected for our purpose and, as a consequence, will not be discussed hereafter.

We first point out that we favoured these Q68 quantiles over other possible statistical indicators (such as the root mean square, rms, or the mean absolute residual, MAR), since we believe that the Q68 are more representative of the real performance (in terms of the bulk of the tested data) of the methods, particularly at low S/N where the parametrization is more difficult to perform. As an illustration of this, we compared the MAR_θ and rms_θ with the $Q68_\theta$ (see Table 4) for all the random FGK-spectral-type stars (about 6000 spectra) and three G_{RVS} magnitudes considered in the present article. In this table, the reported numbers are obtained with the FERRE code but our conclusions are independent of the adopted method. Although the MAR_θ and $Q68_\theta$ are very similar for the best quality spectra ($G_{RVS} < 10.3$), the MAR_θ are always smaller than the $Q68_\theta$ for every atmospheric parameter and that this departure increases for decreasing S/N. For instance, the MAR_θ gets close to the 60% quantiles when G_{RVS} = 13.4 (S/N = 10). This results from the fact that, for fainter spectra, the $\Delta\theta$ distributions start to depart from a pure Gaussian distribution (see Figs. 7 and 8). By contrast, the rms_θ tend to correspond to higher quantiles than the $Q68_\theta$. As it is well known, this is caused by the high sensitivity of the rms to outliers. However, for the faintest magnitudes, these two statistical indicators become closer to each other and the reported errors agree with each other.

4.2.1. General cool and hot random samples: noise and effective temperature effects

As a first step in the comparison, it is important to understand how the parametrization codes react to i) S/N degradation and ii) the general *palette* of spectral types (and therefore stellar effective temperatures) that the GSP-Spec module will have to deal

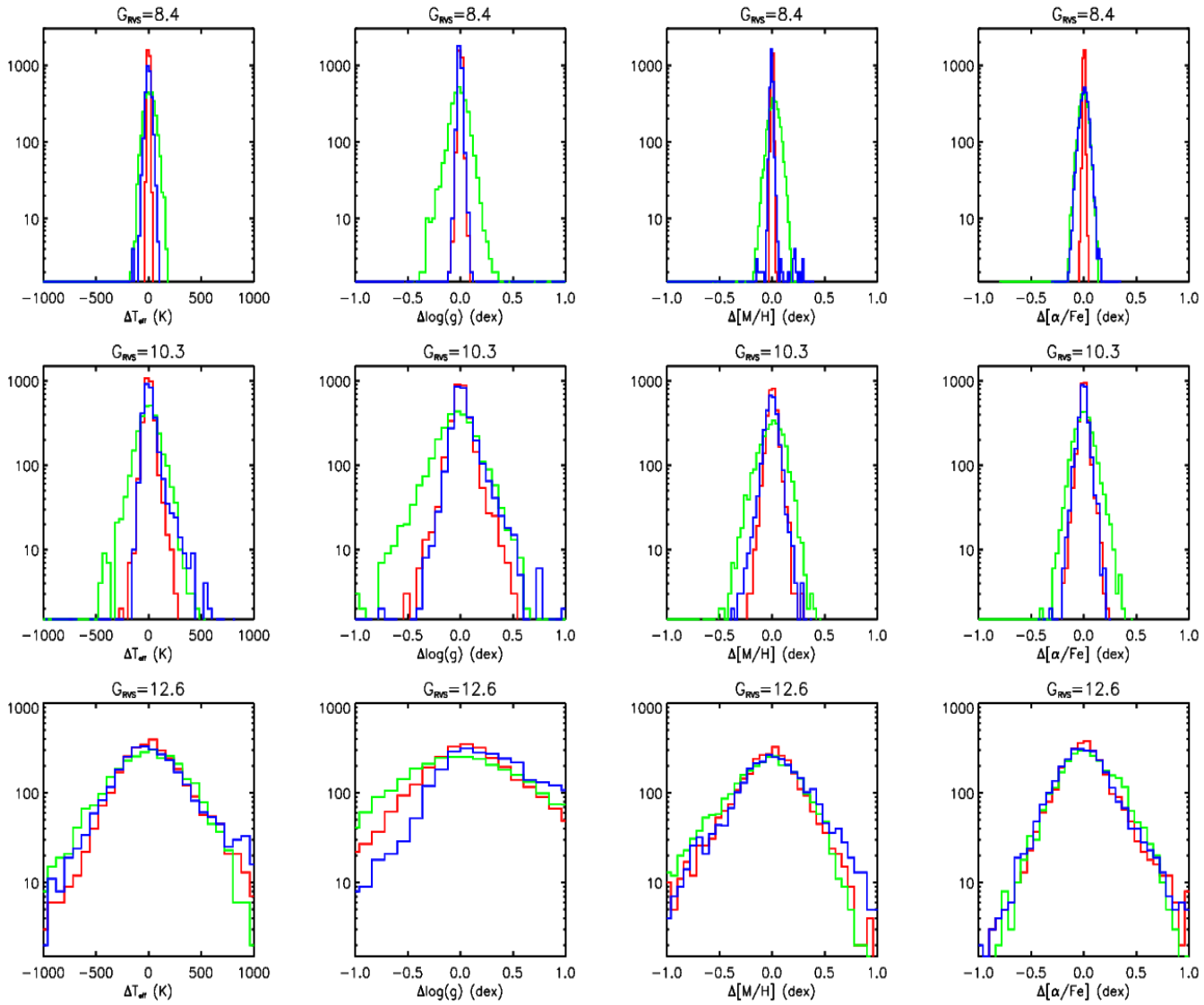

Fig. 7. Distributions of the residuals in the recovered atmospheric parameters ($\Delta\theta = \theta_{\rm rec} - \theta_{\rm real}$) for a subsample of cool random spectra with $G_{\rm RVS}$ = 8.4, 10.3, and 12.6 (i.e. S/N values of 350, 125, and 20, *from top to bottom*, respectively) and defined by 4000 < $T_{\rm eff}$< and 8000 K. The different colours represent the different tested methods: FERRE, ANN, and MATISSE$_G$ in red, green and blue, respectively.

Table 4. Comparison between the statistical performance indicators Q68$_\theta$, MAR$_\theta$, and rms$_\theta$ for the FERRE code.

	Q68$_\theta$			MAR$_\theta$			rms$_\theta$		
S/N	125	40	10	125	40	10	125	40	10
$G_{\rm RVS}$	10.3	11.8	13.4	10.3	11.8	13.4	10.3	11.8	13.4
$T_{\rm eff}$ (K)	32.1	103.3	381.8	33.4 (Q69)	101.4 (Q68)	327.0 (Q62)	56.1 (Q84)	153.7 (Q80)	436.0 (Q73)
log(g) (dex)	0.05	0.15	0.49	0.05 (Q68)	0.15 (Q67)	0.40 (Q61)	0.08 (Q82)	0.22 (Q79)	0.57 (Q71)
[M/H] (dex)	0.05	0.14	0.36	0.05 (Q68)	0.13 (Q65)	0.29 (Q60)	0.08 (Q82)	0.18 (Q76)	0.38 (Q71)
[α/Fe] (dex)	0.04	0.14	0.35	0.04 (Q68)	0.12 (Q66)	0.27 (Q58)	0.09 (Q83)	0.18 (Q78)	0.35 (Q68)

Notes. We report the quantiles corresponding to the MAR$_\theta$ and the rms$_\theta$ in parenthesis in the last six columns.

with. To this purpose, Figs. 9 and 10 illustrate the degradation of each code's performance with increasing noise for the late-type and early-type stars samples, respectively.

In addition, Figs. 11 and 12 show the scatter plots of the method-to-method comparison results at $SNR \sim 125$ ($G_{\rm RVS} \sim 10.3$) and $SNR \sim 20$ ($G_{\rm RVS} \sim 12.6$) respectively. No important trends are found between the results of the different methods, confirming the consistency, with a higher or lower agreement, between the three types of results.

From the above-mentioned plots, several conclusions can be derived:

– Most of the stars are well parametrized in $T_{\rm eff}$, log(g) and [M/H] (together with [α/Fe] for the cool stars) by the three methods, working completely independently. This reinforces the idea that our estimates are robust.

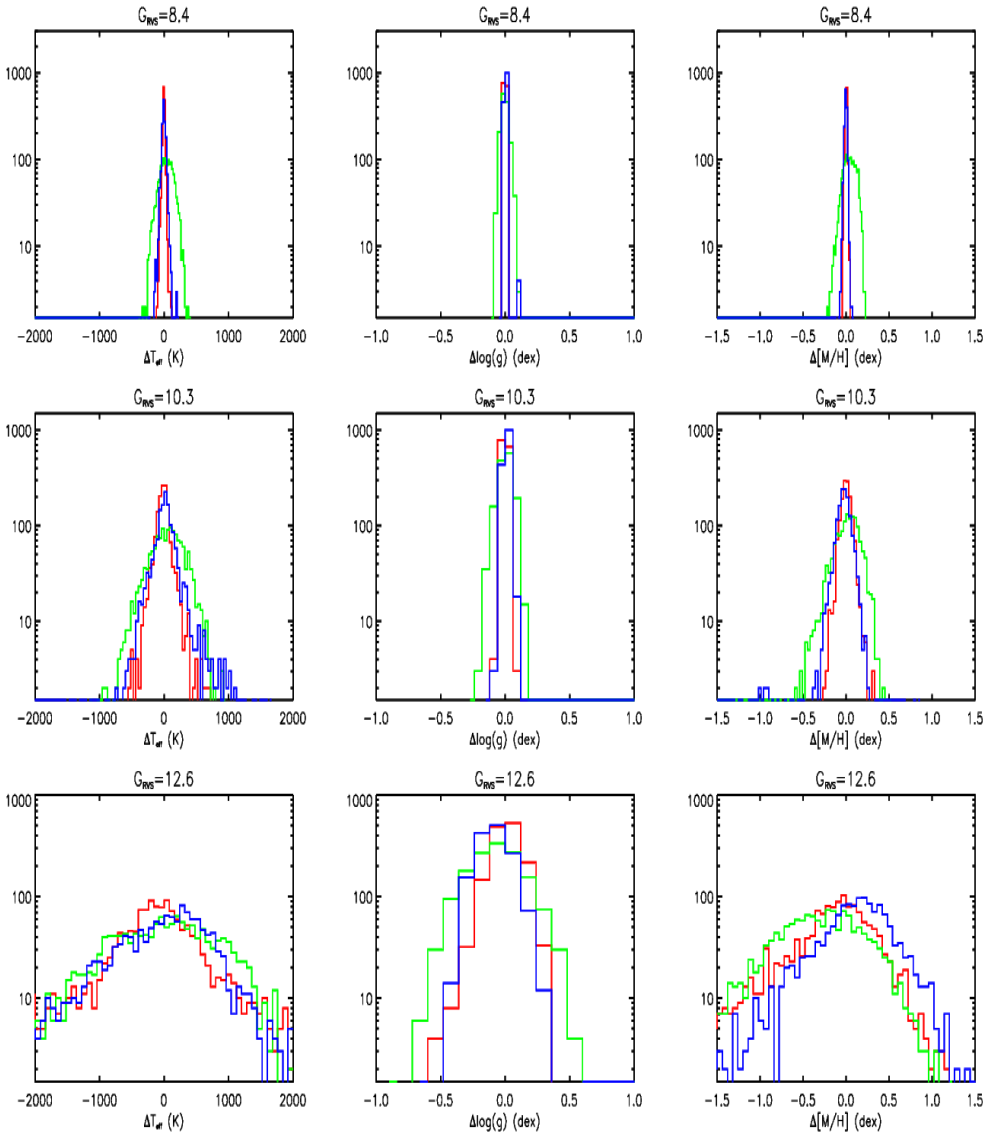

Fig. 8. Same as Fig. 7, but for a subsample of hot random spectra defined by $T_{\rm eff} > 8\,000$ K.

- In the good-to-intermediate quality regime (for $G_{\rm RVS} \lesssim 12.5$) and for cool and hot stars, two completely independent methods (FERRE and MATISSE$_G$) produce very compatible results with no significative differences between both codes.
- In the low-quality regime (for $G_{\rm RVS} \gtrsim 12.5$) and cool stars, the three methods (FERRE, ANN and MATISSE$_G$) give similar results (see also Fig. 7 bottom panel), although the ANN method seems to perform slightly better for very low S/N spectra ($G_{\rm RVS} \sim 13.5$).
- For $G_{\rm RVS} \gtrsim 12.5$ and hot stars, MATISSE$_G$ solutions seem slightly more robust.

4.2.2. Gravity and metallicity effects

To complete the robustness evaluation of the tested codes, and to understand their optimal applicability domains, we need to analyse their performances as a function of two additional parameters: stellar metallicity and surface gravity. To this purpose, we have chosen to illustrate two particular cases, concentrating on G- and K-spectral type stars: rather metal-rich giants (Fig. 13) and dwarf stars with intermediate-to-low metallicities (Fig. 14). These two stellar types correspond to extreme cases of the GSP-Spec performances since (i) late-type giants are more easily parametrized than corresponding dwarfs and (ii) metal-rich star spectra exhibit many more spectral lines that facilitate their parametrization (see Sect. 5.1). The results show the following tendencies:

- In the good-to-intermediate quality regime (for $G_{\rm RVS} \lesssim 12.5$), the FERRE and MATISSE$_G$ codes are confirmed as the two methods that provide the best results, independently of the metallicity and the gravity.
- In the low-quality regime (for $G_{\rm RVS} \gtrsim 12.5$) and for metal-rich stars, the three methods show similar performances, with ANN and MATISSE$_G$ codes being slightly better for $\log(g)$.
- In the low-quality regime (for $G_{\rm RVS} \gtrsim 12.5$) and for metal poor stars, the ANN method again provides the best results.

4.2.3. Summary of each code's performance

In summary, from the above performance comparison, we can infer the following characteristics for each method application to the RVS data:

- The FERRE parametrization is always very satisfactory with good results for any parameters for any type of star.

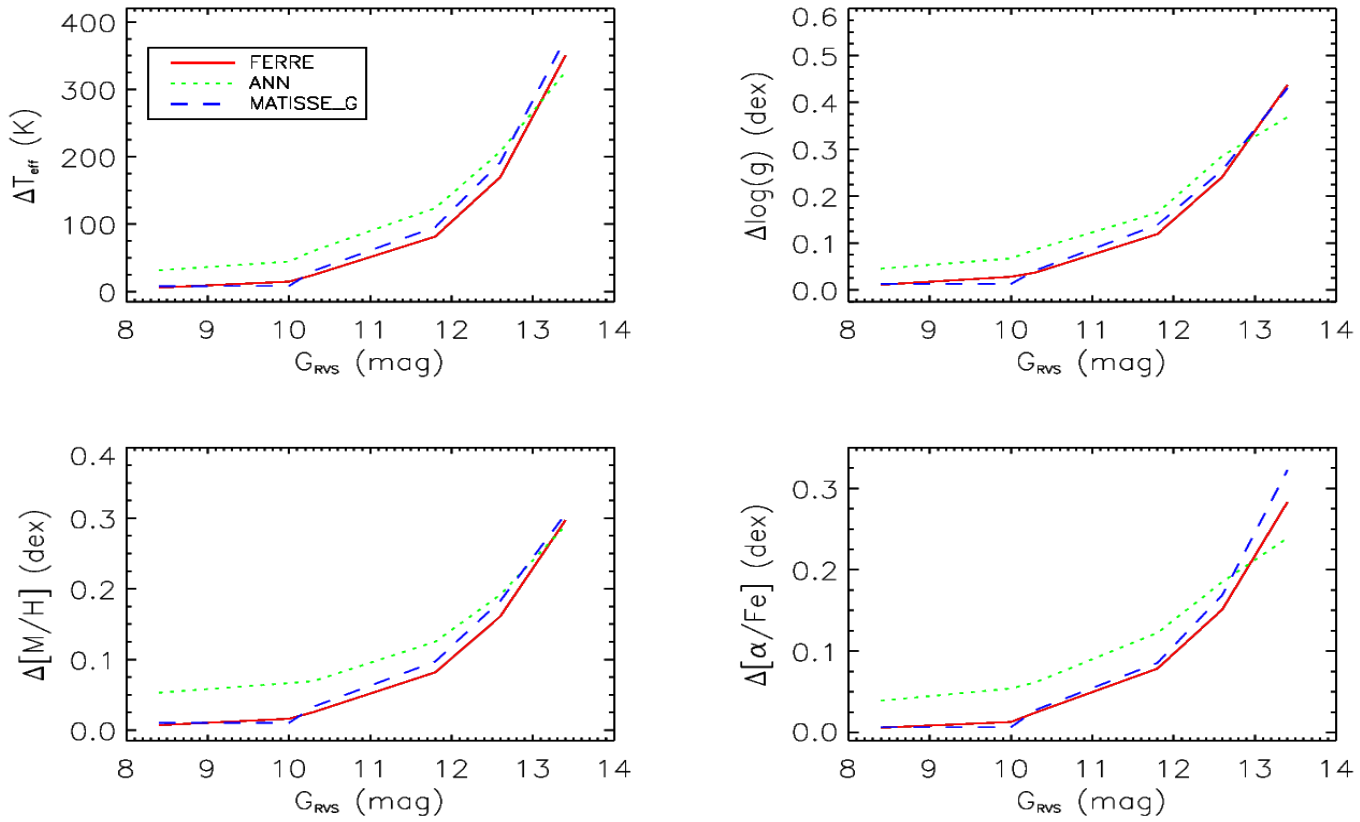

Fig. 9. Variation of the code performances (quantified by the 68% quantile) as a function of the magnitude, for the subsample of random cool stars defined by 4000 < $T_{\rm eff}$ < and 8000 K and [M/H] ≥ −1.0 dex (2951 spectra in total).

However, for the faintest spectra, the performances degrade noticeably, leading to rather poorly classified cool metal-poor dwarfs.

- The MATISSE$_G$ method performs similarly to FERRE with satisfactory parametrization for every situation. MATISSE$_G$ actually produces slightly better results when $G_{\rm RVS} \lesssim 10.5$ but slightly worse for lower S/N. In addition, MATISSE$_G$ can sometimes produce the best estimates when information on the physical parameters is poor, although still not completely degraded at $G_{\rm RVS} = 13.5$ (as for metal-rich cool giants and early-type stars).
- The ANN code always provides the best results for late-type stars when $G_{\rm RVS} \simeq 13.5$ (see Fig. 15). We stress that such stars represent the largest sample collected by *Gaia* RVS. However, for early-type stars, MATISSE$_G$ sometimes performs better.
- It seems clear that the present version of FERRE and MATISSE$_G$ are very sensitive to the Gaussian-like dominated noise simulated in this work. This will be improved in the near future with the development of an optimized version of these codes when real RVS spectra will be available, together with a precise knowledge of their noise properties. The adopted filtering for the ANN method will also need to be adapted as a consequence.

5. Expected parametrization performances for *Gaia* RVS end-of-mission data

From the previous examination of the different parametrization codes, we derived the final (end-of-mission) GSP-Spec expected results by choosing the optimal method for each applicability domain. We first point out that the same code solution was adopted for all four (or three for the early-type stars) atmospheric parameters, to avoid a mix of physically inconsistent parameters. This selection was performed through the following main rules, based on the conclusions of Sect. 4:

- For $G_{\rm RVS} < 12.6$ ($S/N > 20$), the average of the FERRE and MATISSE$_G$ solutions has been adopted as the final GSP-Spec performances since no significant differences appear between both methods.
- For $G_{\rm RVS} \geq 12.6$ ($S/N \geq = 20$) and FGK stars, the average of the FERRE, MATISSE$_G$, and ANN is chosen for metal-rich and intermediate-metallicity stars, while the ANN method is favoured for late-type metal-poor spectra.
- $G_{\rm RVS} \geq 12.6$ ($S/N \geq 20$) and hot stars, MATISSE$_G$ solutions have been selected.

In any case, the GSP-Spec pipeline will also provide the individual results for the different codes in order to avoid any possible discontinuities between the different parameter and/or S/N regimes. Such discontinuities could be accidentally produced by these adopted rules. We point out, however, that this should be avoided because of an accurate validation phase that is based on the analysis of benchmark stars (see an example of this procedure within GES in Recio-Blanco et al., in prep.).

First of all, to describe the code performances for specific types of stars (in terms of spectral type, luminosity, and metallicity), we defined different stellar classes as characterised by the following atmospheric parameters ranges:

1. Metallicity ranges:
   - metal-poor stars: −2.25 ≤ [M/H] < −1.25 dex, roughly corresponding to Halo stars;

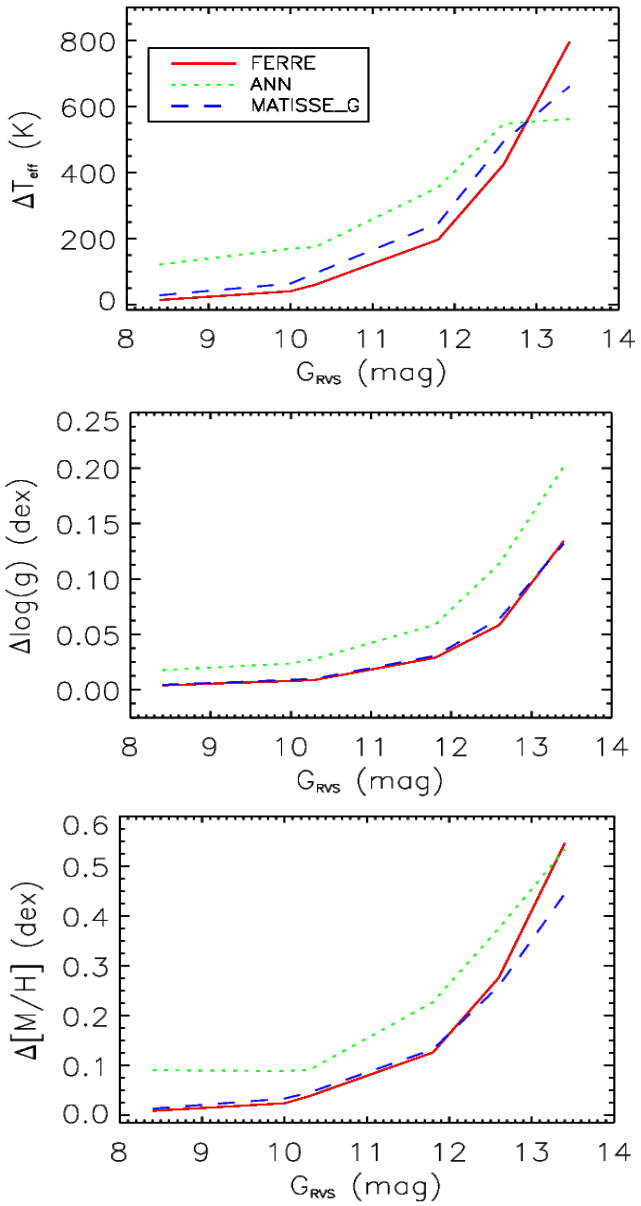

Fig. 10. Same as Fig. 9 but for the subsample of hot stars, defined by $T_{\rm eff} > 8000$ K and [M/H] ≥ -1.0 dex (1457 spectra in total).

  - intermediate-metallicity: $-1.25 \leq$ [M/H] < -0.5 dex, typical of the Galactic thick disk;
  - metal-rich stars: $-0.5 \leq$ [M/H] ≤ 0.25 dex, roughly corresponding to the Galactic thin disc.
2. Gravity ranges:
  - giant stars: $2.5 \leq \log(g) < 3.5$ cm/s^2;
  - dwarf stars: $3.5 \leq \log(g) \leq 4.5$ cm/s^2.
3. Effective temperature ranges:
  - B-type stars: $10\,000 \leq T_{\rm eff} \leq 11\,500$ K;
  - A-type stars: $7500 \leq T_{\rm eff} \leq 9500$ K;
  - F-type stars: $6000 \leq T_{\rm eff} \leq 7000$ K;
  - G-type stars: $5000 \leq T_{\rm eff} < 6000$ K;
  - K-type stars: $4000 \leq T_{\rm eff} < 5000$ K.

This led to the 30 stellar classes (15 for dwarfs and 15 for cool giants)[9]. The end-of-mission GSP-Spec parametrization performances for these different classes and $G_{\rm RVS}$ magnitudes are presented in Table 5. The following subsections analyse and discuss the obtained results.

[9] Some of these stellar classes correspond to rather infrequent real stars, particularly for the hot ones.

5.1. Performances for late-type stars

FGK-spectral-type stars will represent the majority of the *Gaia* RVS targets and therefore, special attention has to be given to the GSP-Spec parametrization capabilities of their spectra. Figures 16 and 17 illustrate the expected errors (defined as the 68% quantile of the $\Delta\theta$ distributions, cf. Sect. 4) for G-type dwarfs and K-type giants, respectively. The different curves on each panel correspond to the three metallicity intervals defined above and reported in Table 5.

For stars with $G_{\rm RVS} \lesssim 12.5$, the FGK stars parametrization is accurate enough to precisely characterize the stellar properties (typical errors are smaller than 0.1 dex in [M/H] and [α/Fe]) and, therefore, to conduct Galactic population studies as already performed from ground-based Galactic archaeology surveys (see Recio-Blanco et al. 2014, for instance). This accuracy in the stellar atmospheric parameters will allow, as a second step, the quite accurate determination of individual chemical abundances (see, for instance, Guiglion et al. 2014). This is especially true for metal-rich and intermediate metallicity stars, that will be the most abundant ones in the magnitude volume sounded by the RVS. For the faintest stars at $G_{\rm RVS} \geq 12.5$–13.0, where the noise amplitude becomes too strong with respect to the available stellar spectroscopic signatures, the accuracy in estimation of the parameters degrades (with metallicity errors in the range 0.2 to 0.5 dex, for instance).

On the other hand, the dependency on the stellar metallicity for the accuracy of the parameters is illustrated by the clear separation of the continuous (metal-rich), dashed (metal-intermediate) and dotted (metal-poor) curves. As expected, metal-rich stars are more easily parametrized than the cool metal-poor ones, whatever the S/N are. This result is evidently caused by the number of spectroscopic signatures (lines sensitive to the atmospheric parameters and abundances) that are available to perform the spectral analysis and that dramatically decreases below [M/H] $\lesssim -0.5$ dex. This is also illustrated in Figs. 18 and 19, which show the dependences of the errors in $T_{\rm eff}$, $\log(g)$, and [M/H] on the three atmospheric parameters. First of all, the three right panels of Fig. 18, show how the errors in $T_{\rm eff}$ (right upper panel), $\log(g)$ (right middle panel), and [M/H] (right bottom panel) depend on the stellar metallicity. In each case, two different curves are shown: the evolution of the 68% quantile for $G_{\rm RVS} = 10.3$ (red continuous line, $S/N = 125$) and for $G_{\rm RVS} = 11.8$ (blue dashed line, $S/N = 40$). As expected, the tendencies are more clearly appreciated at $G_{\rm RVS} = 11.8$ as the parametrization is more sensitive to loss (or gain) of information at lower S/N.

First, it can be appreciated that the errors in the three parameters increase as the metallicity decreases, with a higher sensitivity for the metallicity error itself. In addition, the errors in $T_{\rm eff}$ increase as the metallicity decreases down to about [M/H] $= -1.5$ dex. For lower stellar metallicities, the error in $T_{\rm eff}$ remains almost constant and practically independent of [M/H]. For the metal-poor stars, this is because the only useful temperature indicator that remains is the CaII triplet, which is present even at very low metallicities (see Fig. 2). A similar behaviour can be appreciated for the $\log(g)$ errors as a function of [M/H] (right middle panel). However, to distinguish possible differences between FGK-type and early-type stars, Fig. 19 shows the evolution of the errors in $\log(g)$ and [M/H] on the same two

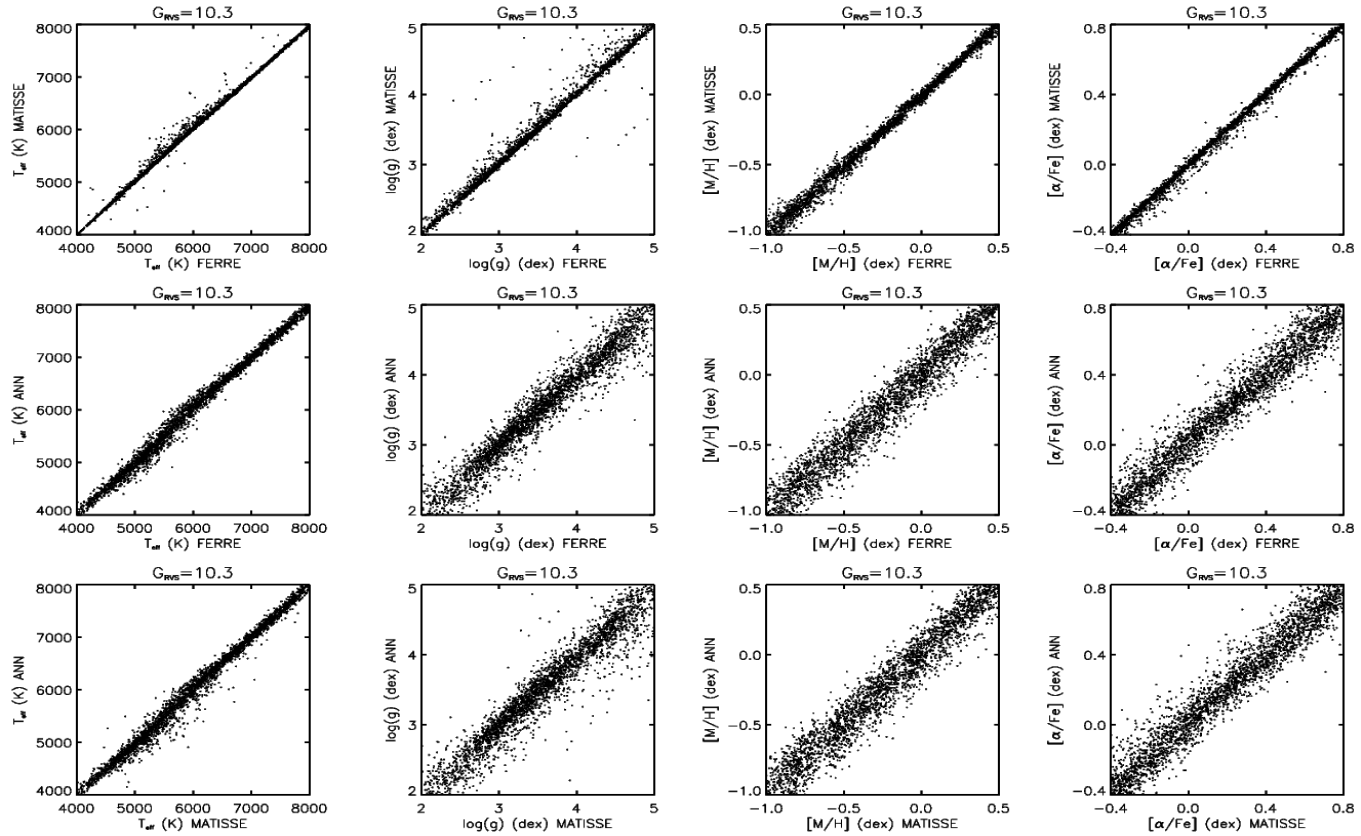

Fig. 11. Comparison of results method-to-method, at $SNR \sim 125$ ($G_{\rm RVS} \sim 10.3$), for the subsample of random cool stars defined in Fig. 7.

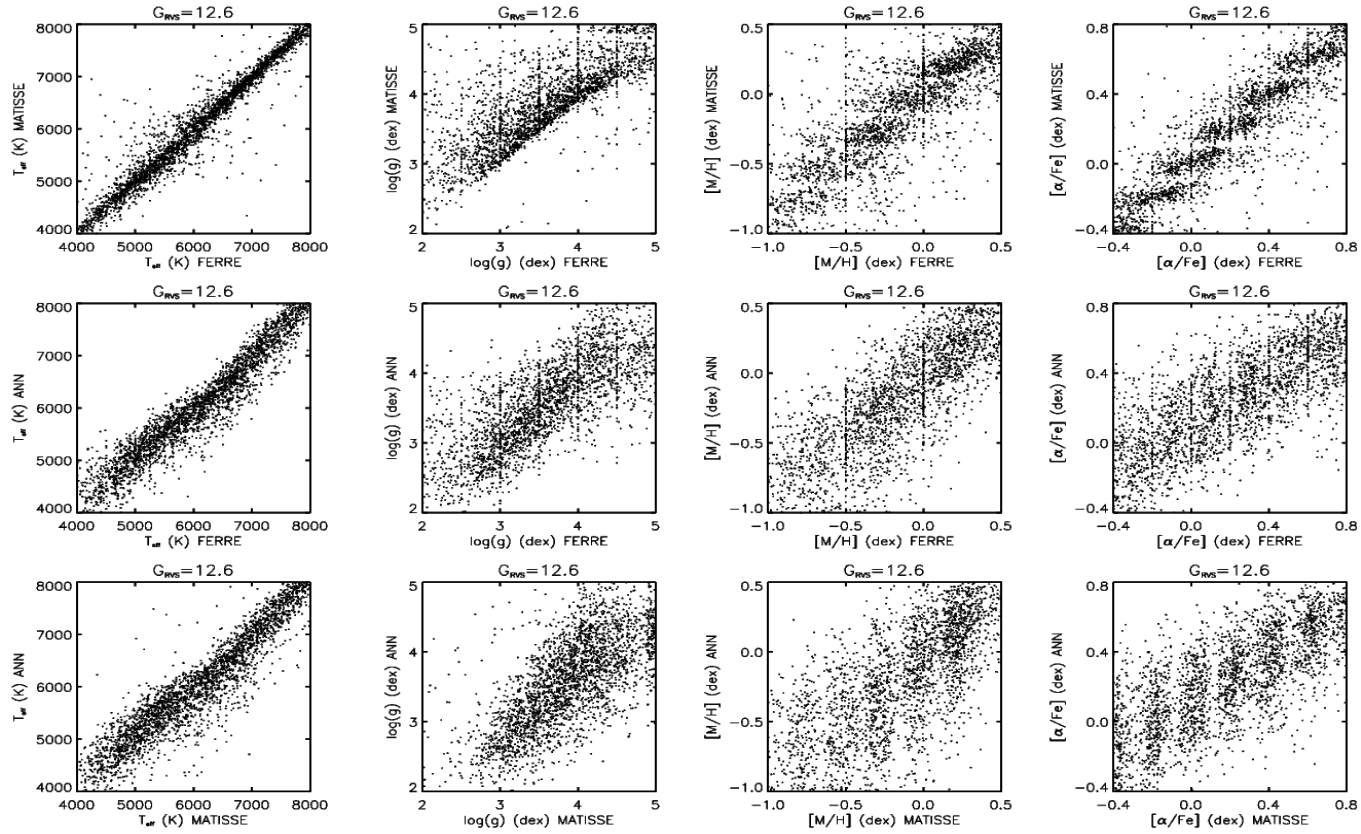

Fig. 12. Same as Fig. 11 but at $SNR \sim 20$ ($G_{\rm RVS} \sim 12.6$).

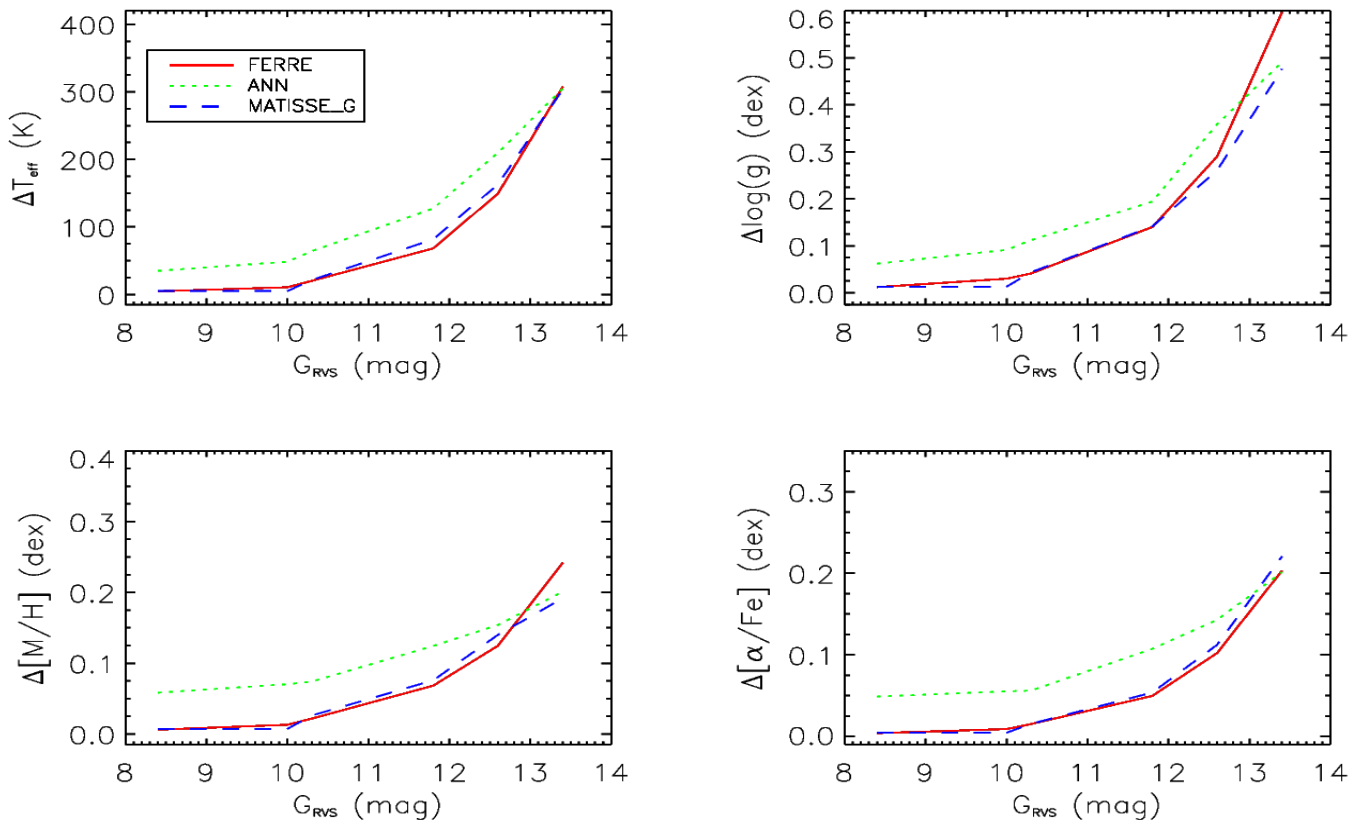

Fig. 13. Same as Fig. 9 but for a subsample of random cool giant stars defined by $T_{\rm eff} < 6000$ K, $\log(g) < 3.5$ cm/s^2, and [M/H] ≥ -0.5 dex (557 stars in total).

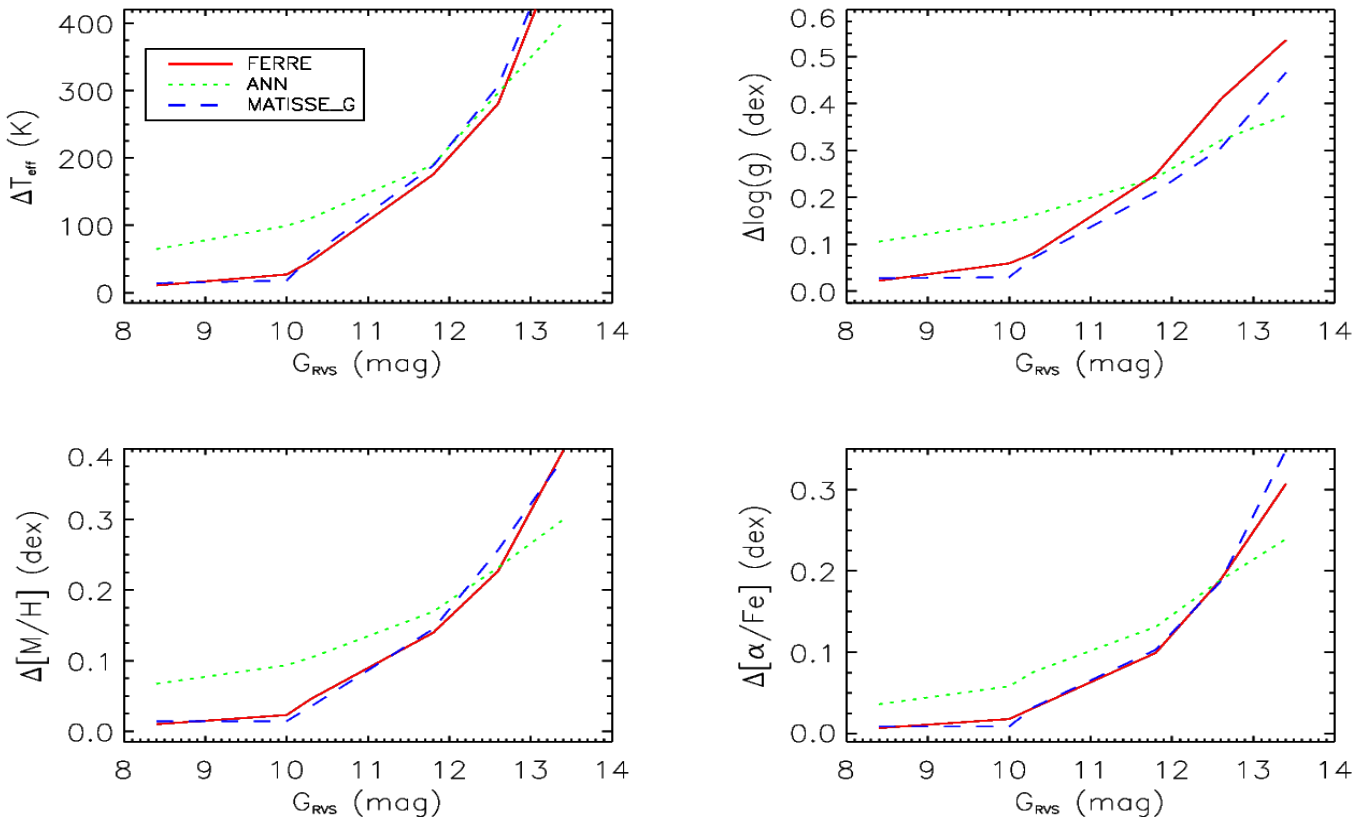

Fig. 14. Same as Fig. 9 but for a subsample of random cool dwarf stars defined by $T_{\rm eff} < 6000$ K, $\log(g) \geq 3.5$ cm/s^2, and $-1.25 \leq$ [M/H] < -0.5 dex (376 stars in total).

Table 5. Expected end-of-mission performances for GSP-Spec (quantified by the 68% quantile) for the different stellar classes defined in Sect. 5.

	$T_{\rm eff}$ (K)						log(g) (dex)						[M/H] (dex)						[α/Fe] (dex)					
S/N	350	150	125	40	20	10	350	150	125	40	20	10	350	150	125	40	20	10	350	150	125	40	20	10
$G_{\rm RVS}$(mag)	8.4	10.0	10.3	11.8	12.6	13.4	8.4	10.0	10.3	11.8	12.6	13.4	8.4	10.0	10.3	11.8	12.6	13.4	8.4	10.0	10.3	11.8	12.6	13.4
Dwarfs																								
B metal-rich	35	89	138	382	478	744	0.01	0.01	0.01	0.02	0.04	0.11	0.01	0.03	0.05	0.13	0.32	0.51						
B interm. met.	36	105	144	420	490	808	0.01	0.01	0.01	0.02	0.05	0.12	0.02	0.05	0.07	0.19	0.38	0.58						
B metal-poor	42	108	149	429	499	784	0.01	0.01	0.02	0.03	0.05	0.12	0.02	0.06	0.10	0.27	0.45	0.65						
A metal-rich	8	21	31	101	210	323	0.01	0.01	0.01	0.03	0.07	0.12	0.01	0.01	0.02	0.07	0.14	0.18						
A interm. met.	8	21	35	104	226	353	0.01	0.01	0.01	0.03	0.07	0.12	0.01	0.02	0.03	0.11	0.26	0.34						
A metal-poor	9	24	35	104	232	393	0.01	0.01	0.01	0.04	0.07	0.14	0.01	0.03	0.05	0.17	0.33	0.45						
F metal-rich	7	19	34	98	179	336	0.01	0.03	0.04	0.10	0.15	0.26	0.01	0.02	0.02	0.08	0.12	0.20	0.01	0.01	0.02	0.06	0.14	0.26
F interm. met.	10	23	40	134	198	394	0.01	0.03	0.05	0.12	0.16	0.27	0.01	0.03	0.06	0.14	0.20	0.39	0.01	0.02	0.05	0.14	0.24	0.27
F metal-poor	13	30	51	134	198	425	0.02	0.05	0.06	0.14	0.17	0.29	0.02	0.05	0.08	0.21	0.23	0.43	0.02	0.05	0.09	0.22	0.25	0.29
G metal-rich	7	19	26	99	166	385	0.02	0.05	0.06	0.17	0.21	0.27	0.01	0.01	0.02	0.08	0.13	0.20	0.01	0.01	0.02	0.07	0.15	0.22
G interm. met.	14	33	62	178	295	460	0.03	0.08	0.12	0.18	0.25	0.28	0.01	0.03	0.06	0.16	0.25	0.32	0.01	0.02	0.04	0.10	0.20	0.27
G metal-poor	23	69	105	200	326	487	0.05	0.13	0.15	0.22	0.28	0.32	0.02	0.06	0.11	0.22	0.31	0.38	0.01	0.04	0.06	0.14	0.18	0.25
K metal-rich	4	7	22	44	143	255	0.01	0.04	0.06	0.12	0.22	0.28	0.01	0.01	0.02	0.06	0.11	0.20	0.01	0.01	0.02	0.04	0.15	0.19
K interm. met.	8	16	27	74	219	305	0.02	0.05	0.06	0.13	0.25	0.30	0.01	0.02	0.02	0.06	0.14	0.28	0.01	0.01	0.03	0.06	0.16	0.19
K metal-poor	19	47	80	183	250	422	0.04	0.15	0.15	0.20	0.28	0.33	0.02	0.05	0.08	0.17	0.27	0.36	0.01	0.03	0.04	0.15	0.19	0.22
Giants																								
B metal-rich	11	31	32	107	365	450	0.01	0.01	0.01	0.04	0.06	0.12	0.01	0.02	0.03	0.12	0.28	0.38						
B interm. met.	13	32	45	139	390	473	0.01	0.01	0.01	0.04	0.06	0.12	0.01	0.03	0.06	0.19	0.40	0.45						
B metal-poor	16	33	52	194	411	493	0.01	0.01	0.01	0.04	0.06	0.12	0.02	0.05	0.09	0.27	0.43	0.48						
A metal-rich	7	27	33	97	209	369	0.01	0.01	0.01	0.04	0.07	0.12	0.01	0.02	0.02	0.07	0.13	0.17						
A interm. met.	9	27	34	102	229	382	0.01	0.01	0.01	0.04	0.07	0.14	0.01	0.02	0.04	0.12	0.22	0.30						
A metal-poor	11	27	36	110	235	403	0.01	0.01	0.01	0.04	0.08	0.14	0.01	0.04	0.04	0.17	0.31	0.43						
F metal-rich	5	10	18	63	134	253	0.01	0.02	0.03	0.08	0.16	0.22	0.01	0.01	0.02	0.06	0.13	0.16	0.01	0.01	0.02	0.07	0.14	0.25
F interm. met.	6	16	22	69	147	265	0.01	0.03	0.03	0.11	0.17	0.22	0.01	0.03	0.04	0.11	0.17	0.22	0.01	0.02	0.04	0.12	0.22	0.27
F metal-poor	6	17	24	71	148	280	0.01	0.04	0.04	0.11	0.18	0.32	0.02	0.05	0.09	0.21	0.20	0.24	0.02	0.05	0.10	0.21	0.28	0.26
G metal-rich	5	12	22	92	177	350	0.01	0.03	0.04	0.15	0.22	0.36	0.01	0.01	0.02	0.08	0.14	0.16	0.01	0.01	0.02	0.06	0.13	0.23
G interm. met.	10	26	47	166	254	373	0.02	0.06	0.10	0.21	0.30	0.44	0.01	0.03	0.05	0.16	0.23	0.25	0.01	0.02	0.04	0.12	0.19	0.24
G metal-poor	15	43	54	170	265	383	0.04	0.11	0.12	0.24	0.34	0.44	0.02	0.06	0.07	0.18	0.27	0.31	0.01	0.04	0.06	0.17	0.21	0.26
K metal-rich	5	11	21	64	147	237	0.02	0.04	0.06	0.17	0.29	0.43	0.01	0.01	0.02	0.06	0.12	0.17	0.01	0.01	0.02	0.04	0.10	0.20
K interm. met.	6	18	29	91	211	333	0.02	0.06	0.08	0.21	0.33	0.50	0.01	0.02	0.03	0.11	0.19	0.28	0.01	0.01	0.03	0.09	0.17	0.24
K metal-poor	13	39	65	200	289	444	0.04	0.10	0.13	0.28	0.34	0.52	0.02	0.05	0.07	0.19	0.25	0.31	0.01	0.04	0.05	0.14	0.19	0.26

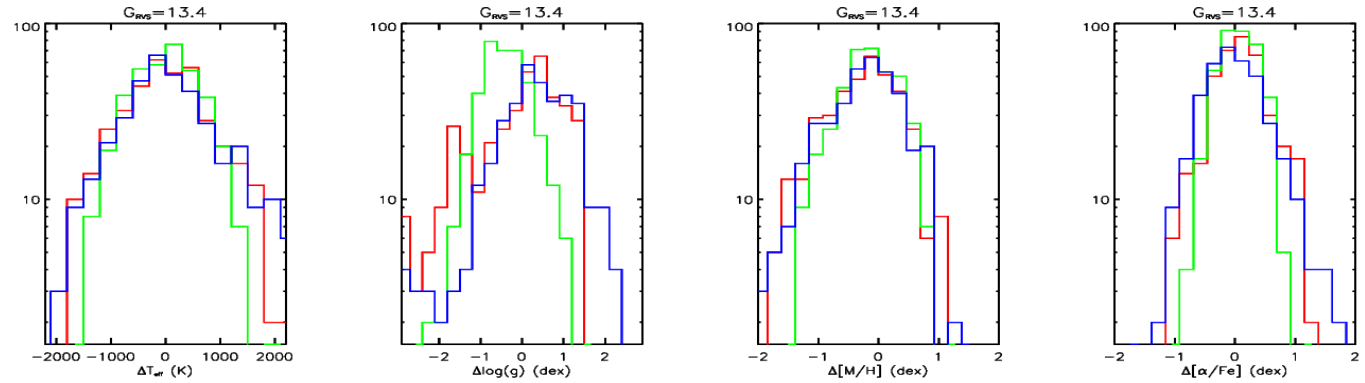

Fig. 15. Same as Fig. 7 but for the subsample of Fig. 14 and $G_{\rm RVS}$ = 13.4.

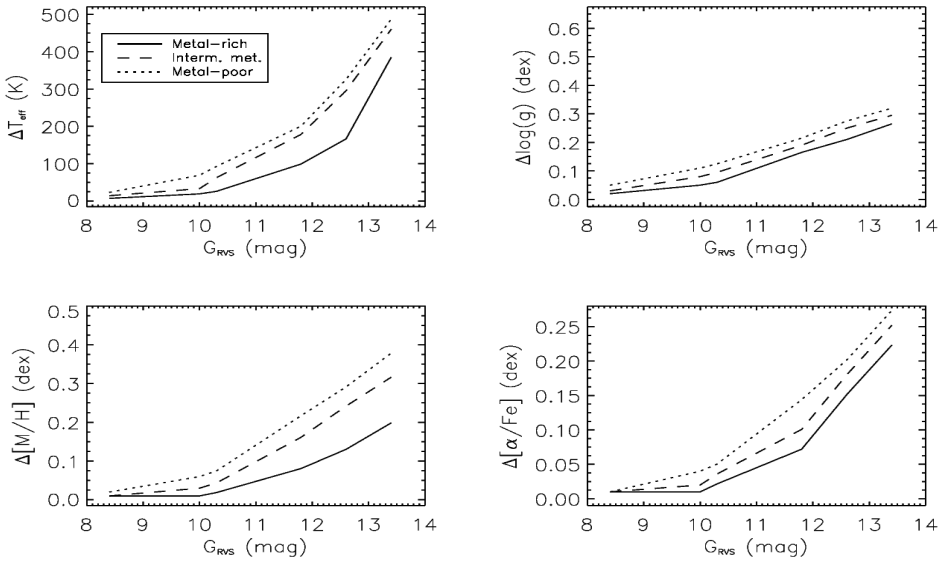

Fig. 16. Variation of the end-of-mission GSP-Spec performances (quantified by the 68% quantile of the residuals) as a function of increasing magnitudes for the *G*-dwarf stars, as defined in Sect. 5.

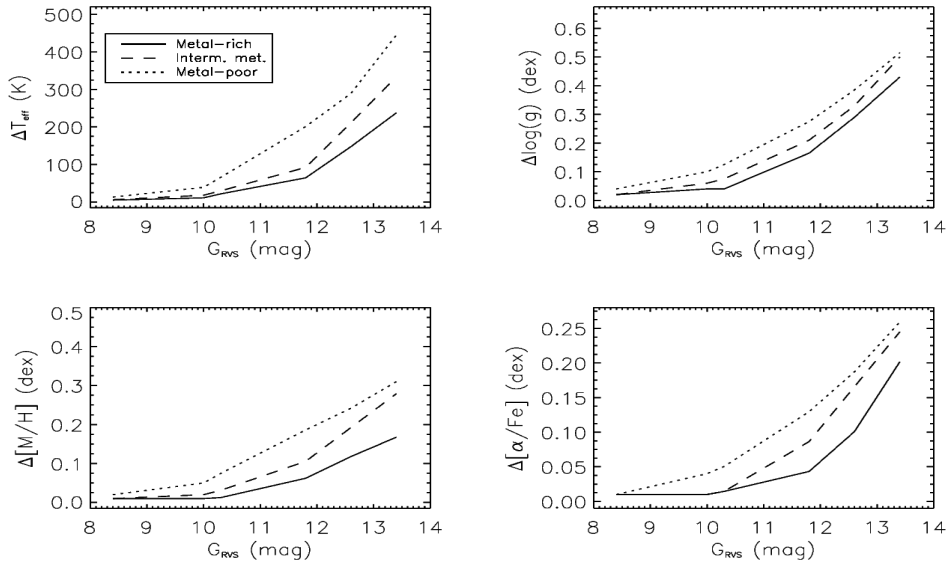

Fig. 17. Same as Fig. 16, but for the K-giant stars as defined in Sect. 5.

parameters, but separating FGK and early-type stars. In fact, the surface gravity error shows a clear metallicity dependence (right upper panel of Fig. 19) down to [M/H] = −1.5 dex, and no dependence for lower metallicity stars.

Another important physical parameter that influences the parametrization performances is the effective temperature. This is illustrated in the three left panels of Fig. 18. In the range that affects FGK-type stars ($T_{\rm eff}$ approximately between 4000 and 7000 K), the behaviour of errors in $T_{\rm eff}$, log(g), and [M/H] shows a maximum around ~5500 K. From that point, the errors decrease in both directions, i.e. for lower and higher $T_{\rm eff}$. In the first case (for lower $T_{\rm eff}$ stars), molecular signatures start to

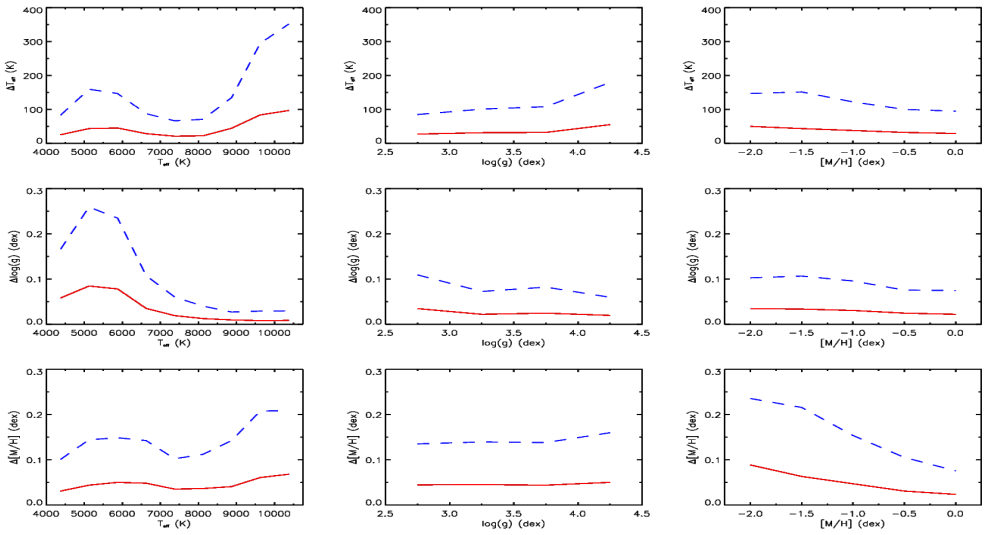

Fig. 18. Variation of the 68% quantile of the residuals for $T_{\rm eff}$, log(g), [M/H] for early- and late-type stars (end-of-mission GSP-Spec performances) as a function of the real atmospheric parameters for $G_{\rm RVS}$ = 10.3 and 11.8 (S/N = 125 and 40) in solid red and dashed blue lines, respectively. The adopted bins in $T_{\rm eff}$, log(g), [M/H] are 750 K, 0.5 dex and 0.5 dex, respectively.

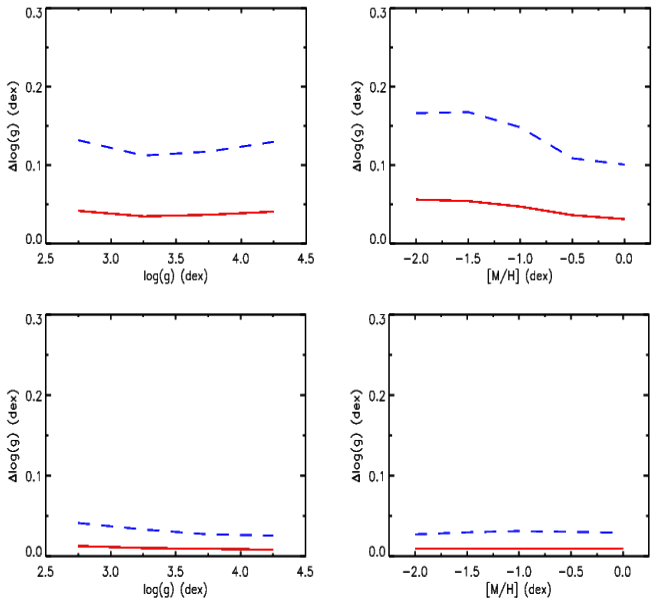

Fig. 19. Same as Fig. 18, but for the residuals in log(g)for cool-FGK and hot-BA stars separately (*upper and lower panels*, respectively).

be visible in the spectra, being more abundant as the temperature decreases. Those molecular signatures are sentitive to both $T_{\rm eff}$ and log(g), since the formation of molecules is favoured for lower temperatures and higher gas pressure (and therefore log(g)). In the second case (for stars with $T_{\rm eff}$ higher than about 6000 K), the appearance of the hydrogen Pachen lines and their rapid change with $T_{\rm eff}$ brings a precious gravity indicator that reduces the errors in log(g) and breaks the $T_{\rm eff}$-log(g) degeneracy (cf. Sects. 5.2 and 5.3). As a consequence, the derivation of $T_{\rm eff}$ and [M/H] are also improved for these hot stars.

Finally, the gravity influence on the stellar parametrization is illustrated by the middle panels of Fig. 18. The derivation of the $T_{\rm eff}$ (upper middle panel) and the [M/H] (bottom middle panel) seem more difficult for dwarf stars than for giants, while the behaviour seems different for the estimation of log(g) (central panel). In practice, the left panels of Fig. 19, which show the residuals of log(g) as a function of log(g) for FGK-type (upper left panel) and early-type stars (bottom left one), clarify the situation. For FGK stars, the gravity determination is, in fact, also more difficult for dwarfs than for giants, following the same tendency as for deriving temperature and metallicity. More generally, in both cases (FGK dwarfs and giants), the gravity is the more difficult parameter to estimate. This problem with the surface gravity is mostly caused by the lack of neutral and ionized lines of the same element in the RVS spectral domain. In any case, even for the lowest quality RVS spectra, the dichotomy between dwarf and giant stars will still be distinguishable.

In summary, the *Gaia* RVS data of FGK-type stars will allow accurate studies of the Galactic disc and halo populations. In particular, for stars with metallicity higher than around −0.5 dex – the majority of the RVS survey – the metallicity and α-enrichment estimates will be very accurate, with typical errors smaller than 0.1 dex in [α/Fe] down to $G_{\rm RVS} \sim 12.5$ (a few tens of million of stars). In addition, K-giants will allow Galactic studies to be performed up to distances of ∼5 kpc (for $G_{\rm RVS}$ = 12, and low extinction regions), or even ∼12 kpc (for $G_{\rm RVS}$ = 13.5).

5.2. Performances for early-type stars

Figure 20 shows the expected errors (defined again as the 68% quantile of the $\Delta\theta$ distributions) for A-type stars. As for FGK stars, three different curves are reported for each stellar parameter, illustrating the behaviour for metal-rich, metal-intermediate and metal-poor stars.

First of all, we can conclude that the parametrization of hot stars with $G_{\rm RVS} \lesssim 12.5$ is expected to be very good (and actually excellent for the stellar surface gravity). As an example, the typical error in [M/H] for hot metal-rich stars is smaller than 0.1 dex down to that magnitude. In fact, this results from the pressure sensitivity of the Paschen lines that is a classical luminosity indicator for early-type stars, especially for $T_{\rm eff} \gtrsim 9000$ K. This comes from the pressure-dependence of the Stark effect. As a consequence, thanks to this important gravity indicator, hot stars do not show a dependence on the gravity estimation accuracy for any atmospheric parameter (cf. Fig. 18 left middle panel, and Fig. 19 bottom panels). Only a small degradation in the S/N is detected. In contrast, both the estimation of the effective temperature and the metallicity are very sensitive to $T_{\rm eff}$. In fact, the number of metallic lines drastically decreases for the spectra of hot stars.

Finally, even for the faintest stars, the parameter accuracy is high enough to allow their classification into the main stellar classes (spectral subtypes, dwarf/subgiant/giant luminosity

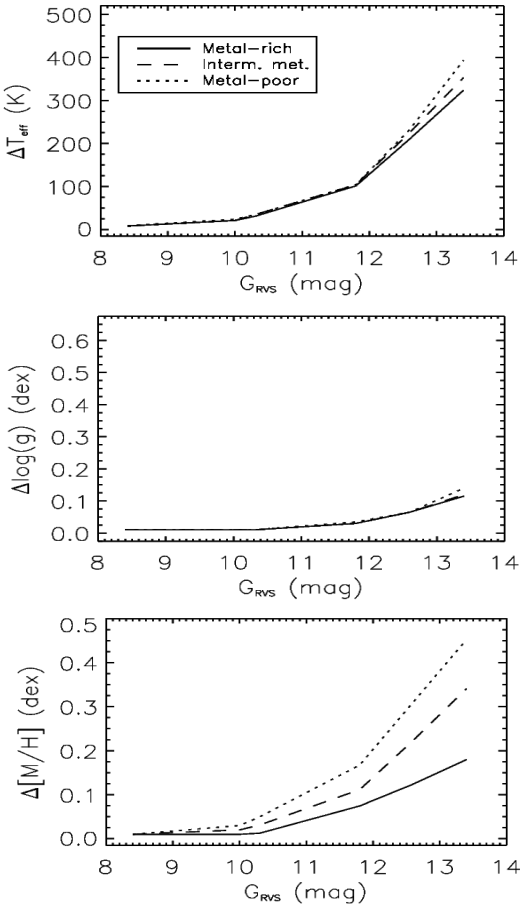

Fig. 20. Same as Fig. 16, but for the A-dwarf stars, as defined in Sect. 5.

classes, with errors in $\log(g)$ lower than approximately 0.2 dex). The stellar metallicity is expected to be recovered with an error smaller than 0.8 dex for the faintest early-type stars (being smaller than 0.3 dex for A-type metal-rich stars). We note that A-type dwarfs are bright stars that allow the RVS sounded volume to extend out to distances of 5 Kpc from the Sun (for $G_{\rm RVS} = 14$).

5.3. Error correlations and parameter degeneracies

One important aspect to consider in any parametrization exercise is the existence of error correlations. These not only provide information on the robustness of the results, but also on the possible physical sources of parameter degeneracies. Figure 21 shows the parameter error correlations at a given magnitude, $G_{\rm RVS} = 10.3$ ($S/N = 125$), which was chosen for being the high quality regime, although with a high enough error amplitude for their analysis. Different colours have been assigned to FGK dwarfs (red), K-giants (green) and BA-type stars (blue). First of all, a strong correlation is visible between the errors in surface gravity and effective temperature for the cool FGK-dwarfs. This comes from a known degeneracy between these two parameters (Kordopatis et al. 2011). On the one hand, the wings of the CaII lines, carrying much of the information in the RVS domain, grow proportionally to $\log(g)^{1/3}$ for cool main sequence stars, but they also depend greatly on the $T_{\rm eff}$. This implies that differences in the spectra with rather different parameters are very small, causing the error correlations seen in Fig. 21. This degeneracy is more significant in the low metallicity regime, where less metallic lines, carrying additional information on $\log(g)$, remain in the spectra. Moreover, as a consequence of the $T_{\rm eff}$-$\log(g)$ degeneracy, the third atmospheric parameter, [M/H], is also more difficult to constrain, also showing error correlations with the other two.

On the other hand, K-giants do not suffer as much as dwarf stars from this $T_{\rm eff}$-$\log(g)$ error correlation, as shown by the green points of Fig. 21. This is because, as already discussed in Sect. 5.1, the parameterisation is easier for giant stars than for dwarfs, with more uncorrelated parameter variations.

Finally, BA-type stars (blue points in Fig. 21), because of the presence of a strong and sensitive surface gravity indicator (the Paschen lines), are not affected by $T_{\rm eff}$-$\log(g)$ or [M/H]-$\log(g)$ degeneracies This is illustrated by the flat behaviour of the $\log(g)$ errors as a function of the $T_{\rm eff}$ ones (left panel of Fig. 21), and the absence of a relation between the [M/H] errors and the $\log(g)$ ones (right panel). Only a small correlation between the metallicity errors with the temperature ones (middle panel) seems to exists, which agrees with the discussions of Sect. 5.2.

6. Comparison with the expected performances from *Gaia* spectrophotometric data

In a study rather similar to the present one, Liu et al. (2012) reported the expected performances of stellar parameterisation from *Gaia* BP/RP spectrophotometry. Liu et al. (2012) analysed the results of various tested methods within the context of the DPAC/*Working Group: Global Stellar Parametrizer – Photometry* (*GSP-Phot*). Some of the end-of-mission *GSP-Phot* expected performances have since been recently updated in Bailer-Jones et al. 2013, see their Table 4; see also Andrae et al., in prep.) The present section discusses the comparison between the expected end-of-mission parametrization performances between RVS (GSP-Spec) and BP/RP (*GSP-Phot*) data for stars brighter than $G_{\rm RVS} \lesssim 15$. We recall that stars fainter than $G_{\rm RVS} \sim 15$ will only have BP/RP-based parameters. Moreover, it should also be pointed out that the present GSP-Spec analysis was performed for simulated RVS spectra that are not affected by any interstellar extinction, contrary to the Liu et al. (2012) results that rely on BP/RP spectra, which show a large range of a priori unknown interstellar extinction. As a consequence, we will therefore assume in the following discussion that GSP-Spec is insensitive to interstellar extinction and we will only consider the Liu et al. (2012) results related to the smallest extinctions. Finally, although the tested random samples, the adopted statistical criteria, and the detailed performances for different types of stars and magnitudes are not exactly the same in our study, or in the Liu et al. (2012) or the Bailer-Jones et al. (2013) ones, a rough quantification of the expected differences can still be performed. We point out that, in the following (as in the substance of all these papers), the reported uncertainties in the recovered stellar atmospheric parameters refer to internal errors only, i.e. relative star-to-star uncertainties.

For the purpose of this *GSP-Phot* and GSP-Spec comparison, we adopted the relationship between the G-band (*Gaia* white light) and $G_{\rm RVS}$ magnitudes already presented in Table 2 and Fig. 4. Subsequently, the $G_{\rm RVS}$ magnitude is found to be brighter than the G-band magnitude of about 0.3 mag for A-type stars, of 0.6 to 0.7 mag for F-type stars (with the magnitude range reporting the metallicity effect from metal-poor to metal-rich stars), of 0.8 to 1.0 mag for G-type stars and of 1.2 to 1.4 mag for the K spectral type. The variation between cool giants and cool dwarf stars is weak and has been neglected.

First, the study of Liu et al. (2012) reveals that the *GSP-Phot* stellar parametrization was performed with almost the

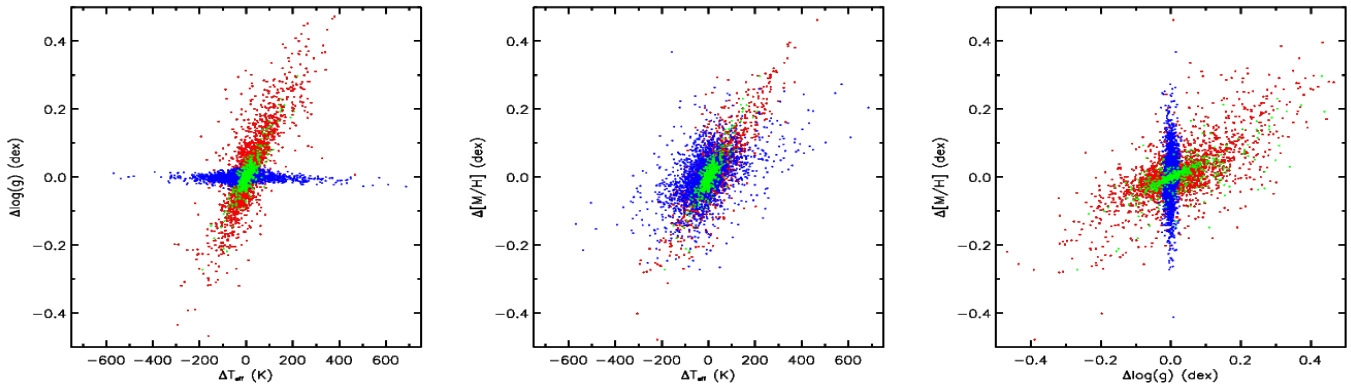

Fig. 21. Correlations between the residuals of the main atmospheric parameters for spectra at $G_{\rm RVS} = 10.3$ ($S/N = 125$). Cool FGK-dwarf, cool K-giant, and hot BA-spectral-type stars are plotted in red, green, and blue, respectively. The main shape of these correlations does not change with $G_{\rm RVS}$, only their amplitude varies.

same efficiency, providing that $G \lesssim 15$, and only starts to degrade for fainter stars. This is also confirmed by the Table 4 of Bailer-Jones et al. (2013), in which the performances at $G = 9$ and $G = 15$ are almost identical. On the contrary, the GSP-Spec parametrization degrades earlier for $G_{\rm RVS}$fainter than ~11 or 12 mag, depending on the metallicity.

As pointed out in Sect. 4.2, the $Q68_\theta$ and rms_θ statistical indicators can be assumed to be almost identical for low S/N RVS spectra. This allows us to compare our Table 5 and the Table 4 of Bailer-Jones et al. (2013) to deduce approximate performance differences. Of course, this comparison only shows the tendencies suggested by tests on simulated data, and neglects external errors, mismatches between real data and models, parametrization methods optimisation, etc., and can be summarised as follows:

- Bright dwarf and giant A-type stars ($G_{\rm RVS} \lesssim 12.5$) should probably be always better parametrized in $T_{\rm eff}$, $\log(g)$, and [M/H] from their RVS spectra. We recall that very good surface gravities and global metallicities (with an accuracy greater than 0.1 dex) will be available for this type of star from their GSP-Spec parametrization. For A-stars that are fainter than $G_{\rm RVS} \sim 12.5$, although $\log(g)$ and [M/H] are still better estimated from their RVS spectra, the effective temperature derived from BP/RP data should be more accurate.
- The GSP-Spec effective temperature of F-type stars should probably be favoured for $G_{\rm RVS} < 12.5$. Their GSP-Spec surface gravity and global metallicity should also be adopted, providing that $G_{\rm RVS} \lesssim 13$. Their accuracy should be better than 0.1 dex when $G_{\rm RVS} < 12$.
- GSP-Spec stellar parameters of GK-spectral type stars should be adopted providing that $G_{\rm RVS} \lesssim 12.5$.
- The global [α/Fe] chemical index will be estimated from the GSP-Spec pipeline only for FGK-spectral type stars. Uncertainties of approximately 0.1 dex (or even smaller) are expected when $G_{\rm RVS} \lesssim 12{-}12.5$, depending on the metallicity.

In summary, for all the considered stellar types, it can be concluded that stars brighter than $G_{\rm RVS} \sim 12.5$ ($S/N = 20$) can be very efficiently parametrized by the GSP-Spec pipeline, including good estimations of the [α/Fe] chemical index. From these stellar atmospheric parameters, individual chemical abundances (such as Fe, Ca, Ti, Si) can be derived with an expected uncertainty less than 0.1 dex for most of the RVS sample that have about $G_{\rm RVS} \lesssim 12$ ($S/N \gtrsim 35$), i.e. for a few million targets. For the faintest stars, which are better parametrised from their BP/RP photometry, a $T_{\rm eff}$ input from *GSP-Phot* as an initial condition for GSP-Spec enables an improvement of its final $\log(g)$, [M/H] and [α/Fe] estimates. Such a *GSP-Phot*/ GSP-Spec link has already been implemented in the Apsis processing system developed by the CU8, and combined performance tests are being implemented.

Finally, we also stress that the spectral parametrization for extinct stars should be easier with the GSP-Spec pipeline, since a $T_{\rm eff}$-extinction degeneracy appears in the parametrization of BP/RP low-resolution spectra, having too large line-of-sight interstellar extinction (assuming that their brightness in the RVS band is not too faint to collect high enough S/N spectra). In these cases, a feedback from GSP-Spec to *GSP-Phot*, as a second iteration of the analysis cycle, will also improve the estimations of final parameters.

7. Comparison with other spectroscopic surveys

A suite of ground-based vast stellar spectroscopic surveys that map the Milky Way is revolutionizing the observational information about Galactic stellar populations. Their synergy with the *Gaia* mission relies not only on the sounded spatial volume, but also on their spectral resolution and covered wavelength domain. These two characteristics primarily determine their corresponding performances in the stellar parameters and estimation of chemical abundances.

The Sloan Digital Sky Survey project, in its series of operations (SDSS I, II, and III) has published about 250 000 spectra ($R = 1800$) from the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009). SDSS spectra have only provided limited information on the structures revealed in the SDSS photometry, but they produced $T_{\rm eff}$ and $\log(g)$ estimates to 250 K and 0.5 cm/s^2 respectively, and [Fe/H] abundances to 0.3 dex for stars with $14 < r < 19$ mag (Schlesinger et al. 2012). SEGUE data overlap the RVS targets only in the RVS fainter magnitudes domain. Because of the larger wavelength coverage of the SEGUE data, its $T_{\rm eff}$ estimations are generally better than the expected GSP-Spec ones. However, the higher RVS resolution should allow more precise measurements of $\log(g)$, and [M/H].

The RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006) is obtaining accurate radial velocities (<5 km s^{-1}) and global metallicities for 5×10^5 stars with $J < 12$ from spectra with $R = 7500$. RAVE is also estimating the individual abundances of some elements for several thousand stars. This project

is probing essentially the Galactic discs populations, because of their particularly bright magnitude limit that corresponds to about $G_{\rm RVS} = 12$ for solar type stars. In terms of parameter estimations, the RAVE internal errors at $SNR = 10$ (about $G_{\rm RVS} = 12$) are of 350 K for $T_{\rm eff}$, 0.5 dex for $\log(g)$, and 0.3 dex for [M/H] for solar-type stars (cf. Kordopatis et al. 2013, Table 1). More generally, the GSP-Spec performances should always be better than the RAVE ones, as is expected from the RVS fainter magnitude limit and higher resolution.

More recently, the Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST; Zhao et al. 2012) project has implemented a survey dedicated to Galactic exploration (LEGUE; Deng et al. 2012). The LEGUE survey plan includes spectra for 2.5 million stars that are brighter than $r < 19$, and an additional five million stars brighter than $r < 17$. The magnitude distribution depends on the telescope throughput and the survey resolution is $R = 1800$, which is much lower than the RVS one. Xiang et al. (2015) estimate the uncertainties of the LAMOST stellar parameter pipeline to be about 150 K in $T_{\rm eff}$, 0.25 dex in $\log(g)$ and 0.15 dex in [Fe/H]. On the one hand, for red giant stars, Liu et al. (2014) report typical errors in metallicity in the range of 0.15 to 0.30 dex. On the other hand, similarly to the SEGUE survey, LEGUE data mainly overlap the RVS observations in the faint magnitudes domain, for which BP/RP data will also be available.

The results of the first low-resolution surveys revealed the key role of the chemical information of stars to disentangle the Milky Way's stellar population puzzle, motivating a new era of ground based, high-resolution spectroscopic surveys. Three of them will be active from the ground during 2015–2019: The *Gaia*-ESO Survey (GES; Gilmore et al. 2012), the SDSS Apache Point Observatory Galactic Evolution Experiment (APOGEE; Eisenstein et al. 2011), and the Galactic Archaeology with AAO HERMES (GALAH; Zucker et al. 2012) survey. All these surveys, thanks to their larger wavelength coverage and resolution, will provide more accurate parameters than *Gaia*/RVS for a subsample of stars. Nevertheless, only the GALAH survey, which targets about one million stars with $V < 14$, is expected to have an important overlap with the RVS. In any case, this overlap will only correspond to less than one tenth of the RVS targets with $G_{\rm RVS} < 13$. The GES survey is mainly targeting faint stars ($14 < V < 19$) thanks to the Very Large Telescope FLAMES/GIRAFFE facility ($R \sim 20\,000$), and it will primarily complement the *Gaia*/BPRP parameter estimations. In the RVS magnitude domain, only a small GES sample of 10^4 G-stars within 2 kpc of the Sun ($12 < V < 14.5$, which correspond to about $11 < G_{\rm RVS} < 13.5$), is being observed with the FLAMES/UVES spectrograph ($R = 40\,000$). Finally, the APOGEE survey, is preferentially targetting high extinction regions of the disc and the bulge in the range $8 < H < 13.8$. Although the magnitude coverage overlaps the RVS one, the APOGEE targeted fields are characterized by a high stellar crowding that limits the RVS observations. As a result, APOGEE will mostly complement the RVS survey near the Galactic plane, rather than overlapping it.

In conclusion, the RVS-based stellar parameters will provide precious information about Galactic populations in the bright part of the *Gaia* volume. This corresponds to a number of stars tens of times higher than what will be provided by currently ongoing and planned spectroscopic surveys from the ground. These surveys, especially those at high spectral resolution, will nevertheless be crucial for the validation of *Gaia*/RVS parameters and for complementing them with precise chemical abundances for a subsample of stars.

8. Conclusion

After having analysed the results of different independent methods, in this work we have estimated the end-of-mission expected parametrization performances of the *Gaia* DPAC pipeline (GSP-Spec) in charge of the RVS stellar spectra atmospheric parameters and chemical abundances derivation. The estimated accuracies, as a function of stellar types and magnitudes, are summarised in Table 5 and in Figs. 16 to 20.

The reported uncertainties in the recovered stellar atmospheric parameters refer to internal errors only, i.e. relative star-to-star uncertainties. In many cases, total errors will be dominated by external ones (partly caused by the possible synthetic spectra mismatches with respect to the real, observed ones) and they will be estimated from the analysis of real *Gaia* RVS spectra of benchmark reference stars during a results validation phase. Nevertheless, the internal errors reported here permit us to clearly identify, and quantify in detail, the enormous variety of science cases that will be obtained from the interpretation of pure *Gaia* data (without any need for references to external catalogues).

The GSP-Spec pipeline will be optimised in the light of the first analysed real RVS spectra next year. Increasingly improved versions of the Apsis GSP-Spec module are delivered during each operations cycle. GSP-Spec is expected to be running in operations cycle 4 in 2017, with a possible contribution from the third *Gaia* data release. The current GSP-Spec version, integrated in the general Apsis chain, already meets the tight requirements in processing speed (17 Mflops per source), which is needed to repeatedly treat tens of million of spectra.

Our tests, including first estimations of the impact caused by the onboard detected stray light contamination, show that the contribution of the RVS-based stellar parameters will be unique for stars with $G_{\rm RVS} \lesssim 12.5$ (a few tens of million of stars). On the one hand, the GSP-Spec parameters will probably be more accurate than the majority of the parameters derived from the spectrophotometry in that magnitude range. This will allow, thanks to the use of the *Gaia* parallaxes, a better estimation of the stellar evolution phase and, as a consequence, of the isochrone-based age estimations (for which the effective temperature accuracy is a dominant source of error). Accurate stellar ages will be one of the revolutions in Milky Way astrophysics that the *Gaia* mission will accomplish, and the RVS data will strongly contribute to it, sharpening our view of Galactic history in a volume of very precise measurements (up to ~8 kpc from the Sun for K-giants and ~1 kpc for *G*-dwarfs).

On the other hand, accurate metallicity and chemical abundance measurements as the [α/Fe] content are today recognized as crucial information for the understanding of the highly complex evolution of Galactic stellar populations. As an example, the classical kinematically-based definitions of the thin and the thick disc populations have blurred our comprehension of the Galactic disc substructure (cf. Bovy et al. 2012; Recio-Blanco et al. 2014). The RVS chemical abundance estimations, with an accuracy greater than 0.1 dex, will therefore be a unique and precious sample of several millions of stars from which many pieces of the Milky Way history puzzle will be available with unprecedented precision and statistical relevance.

Acknowledgements. We thank the Centre National d'Etudes Spatiales (CNES, France) and the French CNRS/INSU for continuous support for the preparation of the *Gaia* mission. This work benefited from travel support from the European Science Foundation through the GREAT Research Network Program. Some of the computations were done on the Mesocentre SIGAMM machine, hosted by

the Observatoire de la Côte d'Azur. The first two authors would like to thank Naia for her numerous (and efficient) attempts to postpone the revision of this paper. We warmly thank C.A.L. Bailer-Jones for his constructive remarks, which helped to improve this article. We are also sincerely grateful to D. Katz for his help with regard to the in-flight RVS characteristics.

References

Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804
Asplund, M., Grevesse, N., & Sauval, A. J. 2005, in Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis, eds. T. G. Barnes, III, & F. N. Bash, ASP Conf. Ser., 336, 25
Bailer-Jones, C. A. L., Andrae, R., Arcay, B., et al. 2013, A&A, 559, A74
Barklem, P. S., Piskunov, N., & O'Mara, B. J. 2000, A&AS, 142, 467
Bijaoui, A., Recio-Blanco, A., de Laverny, P., & Ordenovic, C. 2012, Statistical Methodol., 9, 55
Bovy, J., Rix, H.-W., Liu, C., et al. 2012, ApJ, 753, 148
Boyajian, T. S., McAlister, H. A., van Belle, G., et al. 2012, ApJ, 746, 101
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Castelli, F., & Kurucz, R. L. 2003, in Modelling of Stellar Atmospheres, eds. N. Piskunov, W. W. Weiss, & D. F. Gray, IAU Symp., 210, 20
de Bruijne, J. H. J. 2012, Ap&SS, 341, 31
de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, A&A, 544, A126
de Laverny, P., Recio-Blanco, A., Worley, C. C., et al. 2013, The Messenger, 153, 18
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RA&A, 12, 735
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Fleitas, J., Davidson, M., Katz, D., Els, S., & de Bruijne, J. 2015, in ESA/DPAC Note: GAIA-CO-TN-ESAC-JMF-012-01-00
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
Grevesse, N., & Sauval, A. J. 1998, Space Sci. Rev., 85, 161
Guiglion, G., Recio-Blanco, A., & de Laverny, P. 2014, in IAU Symp. 298, eds. S. Feltzing, G. Zhao, N. A. Walton, & P. Whitelock, 408
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Koesterke, L. 2009, in AIP Conf. Ser. 1171, eds. I. Hubeny, J. M. Stone, K. MacGregor, & K. Werner, 73
Koesterke, L., Allende Prieto, C., & Lambert, D. L. 2008, ApJ, 680, 764
Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011, A&A, 535, A106
Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, AJ, 146, 134
Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, MNRAS, 426, 2463
Liu, C., Deng, L.-C., Carlin, J. L., et al. 2014, ApJ, 790, 110
Manteiga, M., Ordóñez, D., Dafonte, C., & Arcay, B. 2010, PASP, 122, 608
Ramírez, I., & Meléndez, J. 2005, ApJ, 626, 465
Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, MNRAS, 370, 141
Recio-Blanco, A., de Laverny, P., Kordopatis, G., et al. 2014, A&A, 567, A5
Schlesinger, K. J., Johnson, J. A., Rockosi, C. M., et al. 2012, ApJ, 761, 160
Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645
Worley, C. C., de Laverny, P., Recio-Blanco, A., et al. 2012, A&A, 542, A48
Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, MNRAS, 448, 822
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Zhao, G., Zhao, Y., Chu, Y., Jing, Y., & Deng, L. 2012, ArXiv e-prints [arXiv:1206.3569]
Zucker, D. B., de Silva, G., Freeman, K., Bland-Hawthorn, J., & Hermes Team 2012, in Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, eds. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto, ASP Conf. Ser., 458, 421