Scientific report. GREAT ESF Exchange Grant

Maximizing the scientific exploitation of Gaia RVS and Gaia-ESO Survey data for faint magnitude stars

Diego Fustes

January 11, 2013

1 Purpose of the visit

The Gaia Radial Velocity Spectrograph (RVS) data will be a precious complement to the BP/RP spectrophotometric data for many galactic archeology objectives. In particular, in addition to the information about the stellar radial velocities, a complementary determination of other stellar parameters will be possible. Effective temperature, surface gravity, global metallicity and, whenever possible, the abundance of alpha-elements with respect to iron will be derived from RVS data by the Generalized Stellar Parametrizer-spectroscopy (GSPspec) algorithm, integrated in the APSIS analysis chain of the Gaia DPAC CU8. The parameters derived by the GSPspec algorithm will not suffer from the degeneracies between effective temperature and interstellar absorption, in opposition to the measurements provided by the analysis of BPRP data. They will offer a precious independent estimation of the metallicity that can be particularly accurate for several millions of stars in the solar neighbourhood. This allows an improvement of the age estimations and therefore the age-metallicity relations in the solar neighbourhood, with important implications for our detailed understanding of the local galactic disc, based on an unpreceedently high number of stars. Moreover, the RVS data, centered in the infrared calcium triplet, offer the possibility of obtaining other chemical diagnostics of galactic populations as the alpha-elements content with respect to iron. This will allow the chemical tagging of the stars observed in the solar neighbourhood, increasing the scientific impact of Gaia in our knowledge of the relative influence of proceses as radial migration, stellar accretion, and Bulge-Disc interaction, in the formation of the Galactic Disc.

The above mentioned added value of RVS data is nevertheless injured by the degeneracies between effective temperature and surface gravity that exist in the infrared calcium triplet domain at the RVS resolution. Those degeneracies are more important for low signal to noise data, in particular for stars fainter than about $V = 14mag$. Particular analysis methods have to be implemented for this faint magnitude limit of the RVS sample, that will represent a huge fraction the final RVS catalog.

The visit object of this funding proposal will strength an on-going collaboration between the group of the University of A Coruña and the Nice one with a specific goal: to improve, test and apply to real data several spectrum analysis techniques that tackle the parameter degeneracy problem. The Gaia team at the University of A Coruña has been testing an algorithm based on Artificial Neural Networks (ANNs) trained with the error backpropagation algorithm. In our case, a preprocessing stage, prior to the proper process of parameterization of the stellar spectra, was included in order to refine the algorithm performance, especially for cases of low S/N. The fact that the spectral features, that are sensitive to the different astrophysical parameters, (APs, T_{eff} , $logg$, metallic lines, molecular bands, and α element lines), are broad or narrow, together with the consideration that they are located at specific wavelengths along the spectrum, suggested that a multilevel Wavelet decomposition could provide good results for a selective filtering of the information. An extensive ensemble of tests have been performed, applying synthetic spectra calculated in the domain of the Gaia RVS instrument.

Finally, the work done during this visit will allow us to efficiently compare different analysis approaches and to converge to an optimized, hybrid analysis that will permit to maximize the scientific return of the Gaia/RVS and Gaia-ESO spectroscopic surveys.

2 Work carried during the visit

In the time shared in Nice, both groups have put several methodologies in common. To start with, we have generated a complete simulation of RVS spectra for stars covering a wide range of parameters (T_{eff}, $logg$, $[M|H]$, $[\alpha|Fe]$), including a *training* dataset, a *validation* dataset and a *testing* dataset, for each combination of spectra brightness (the SNR) and the type of star (A or FGK). Futhermore, we have built a complete resimulation which includes radial velocity (RV) errors in the spectra, in order to assess the impact of them in the obtained parameterizations. With these datasets at hand, we can evaluate the performance of all developed algorithms in a unified framework.

In order to assess the performance of the ANN algorithm and the Wavelet preprocessing methods, we first used the results obtained, from the *validation* dataset, to select the best input representation in function of the spectra brightness (the SNR), the type of star (A or FGK) and the astrophysical parameter to be estimated. Then, we evaluate the parameters given with the best ANN method when working with the *test* dataset in order to discuss the scientific results of the method, looking at the systematics in the residuals and the estimations given for the different stellar populations. Finally, all this procedure has been repeated with the RV noised datasets, and future tests, including spectrum normalization issues, have been designed.

3 Main results

Figures 1 and 2 show the 70th percentile of the residuals given by each data representation for each combination of star type, magnitude and parameter, when the *validation* dataset is presented to the ANNs. We can observe that the Wavelet representation outperforms the original one in most cases. The gain obtained by the Wavelets is small when working with FGK stars, nevertheless it becomes evident in the case of A stars parameterization of T_{eff} and $logg$.

Once selected the best representation for each case, the test dataset has been used to assess the performance of the ANNs. We can observe the parameterization given by the ANNs, decomposed by population type, as is shown in figure 3

The following conclusions can be drawn from the consequent analysis of residuals (not completely shown for brevity):

- T_{eff} **estimation**: It can be observed that the parameterizations for FGK stars are fairly accurate up to $Grvs = 14$, without any perceptible systematic. However, the T_{eff} estimation for A stars performs poorly, and it seems to be a systematic under-estimation for the hotter stars. This is expected since the relevant lines are less prominent in the spectra of hot stars. The Wavelet filtering improves the results significantly in the case of hot stars, since the relevant spectral lines are broad and the method drops part of the high-frequency noise.

- $logg$ **estimation**: The estimation of this parameter performs accurately up to $Grvs = 13$ for FGK stars and $Grvs = 14$ for A stars. Therefore, the parameterization of $logg$ performs better for hotter stars. Physically, this is due to the Paschen spectra lines, which become strong from 6,000K up. Such lines are broad, which explains the significant improvement provided by the Wavelet filtering, specially for faint objects.

- $[Fe/H]$ **estimation**: As expected, the accuracy of the estimation is worse in the case of A stars, since the metal lines are less prominent in such ones. In all cases, the estimation degenerates when metal-poor stars are presented to the algorithm.

- $[\alpha/Fe]$ **estimation**: The algorithm provides accurate results for FGK stars, with residuals below $0.1\,dex$, up to $Grvs = 11$. No systematics are observed in the obtained residuals.

Finally, the tests performed with RV noised spectra indicate that such errors, when follow the expected levels, are too low to affect the ANNs estimation performance in any magnitude level.

4 Future work

A lot of work is projected to be done in the following months. First, the results given by ANNs will be crossed with the ones given by the MATISSE and DEGAS algorithms, in

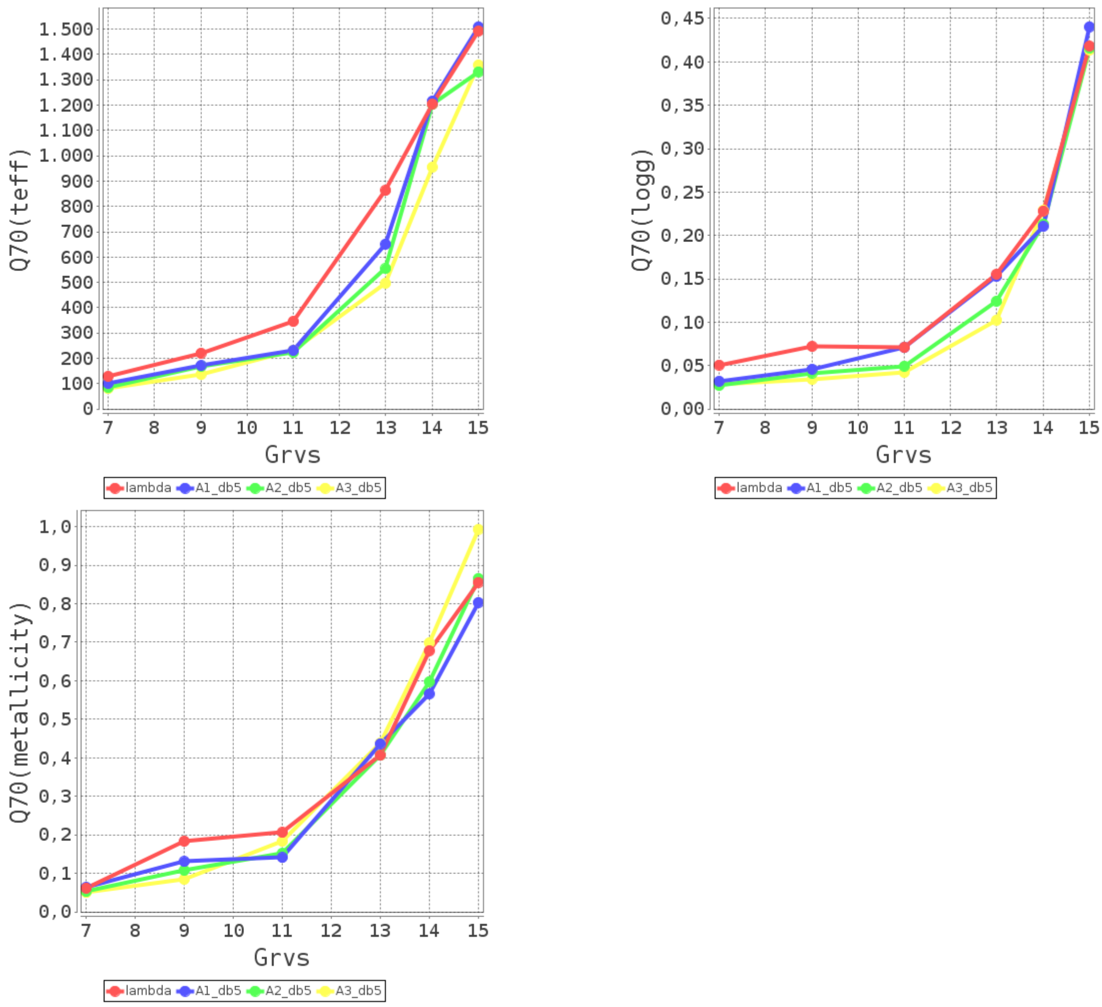

Figure 1: 70th percentile of the absolute residuals (estimated AP value versus real AP value) for A stars. The estimations have been given by ANNs trained with original spectra (called *lambda*), and with the first, second and third level wavelet approximations (called *A*1_*db*5, *A*2_*db*5 and *A*3_*db*5 respectively).

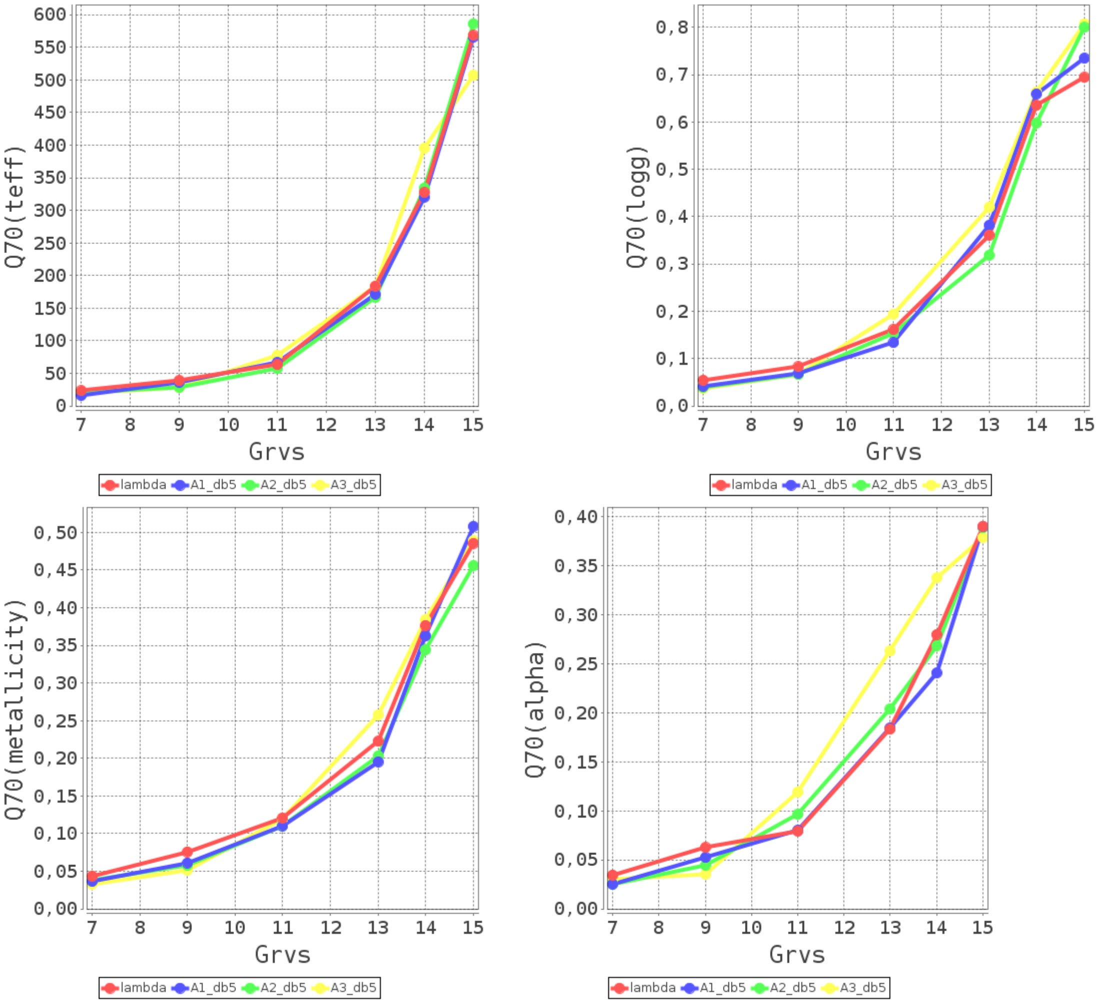

Figure 2: 70th percentile of the absolute residuals (estimated AP value versus real AP value) for FGK stars. The estimations have been given by ANNs trained with original spectra (called *lambda*), and with the first, second and third level Wavelet approximations (called *A1_db5*, *A2_db5* and *A3_db5* respectively).

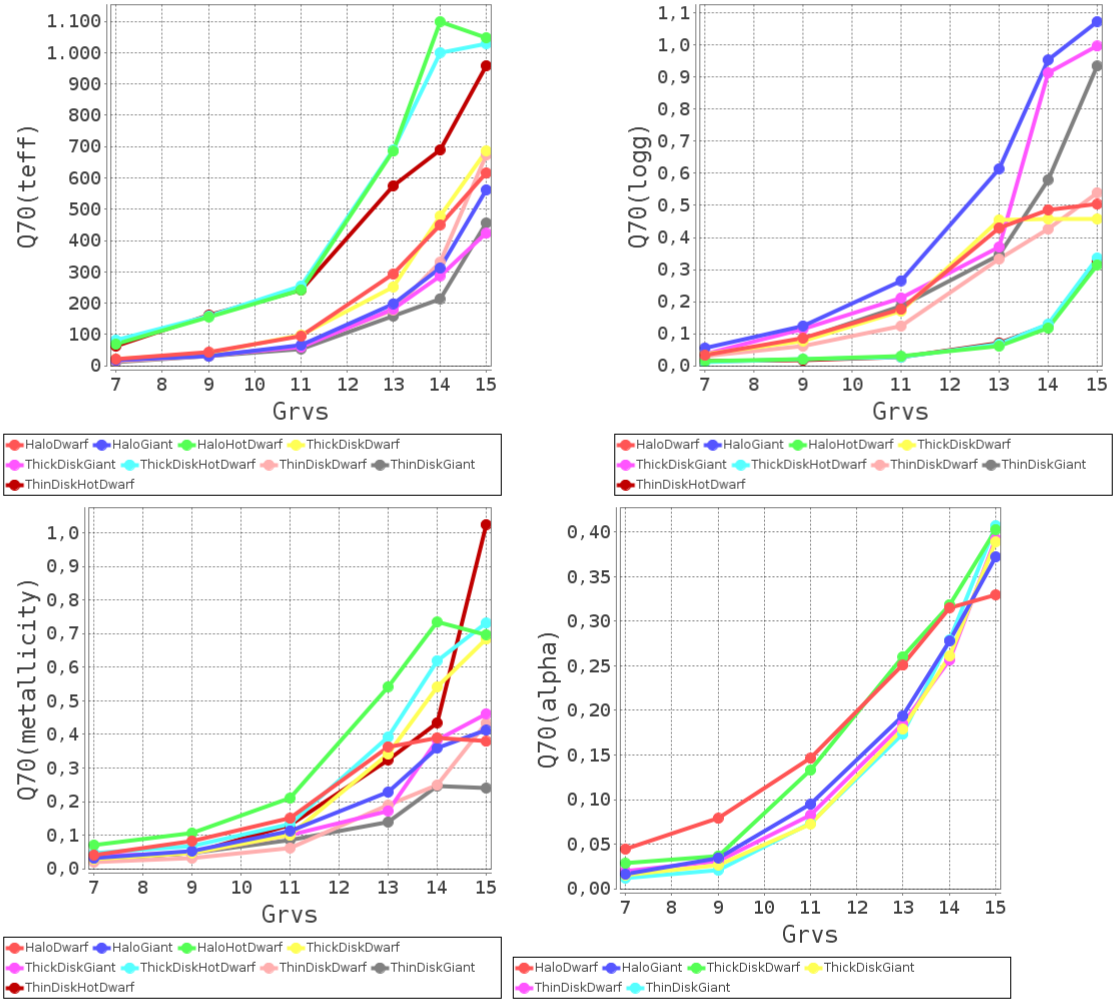

Figure 3: 70th percentile of the absolute residuals for 9 selected stellar populations and the four astrophysical parameters to be estimated.

order to select which algorithm performs best in each case. Then, normalization errors will be simulated to evaluate their impact in the estimations. Furthermore, spectra normalized to the pseudo-continuum will be simulated for comparison with those obtained with the current normalization scheme. Finally, real spectra, coming from the Gaia-ESO survey, will be used for assessing the algorithms performance.

5 Projected publications

An article describing the expected performance of the whole Gaia GSPspec algorithm is being prepared for publication. Its tentative title is “Expected performances on stellar parametrization from Gaia Radial Velocity Spectrometer data”, which is expected to be sent to reviewers in the next month of May.