Stellar libraries for Gaia

R Sordo, A Vallenari, R Tantalo, C Liu, K Smith, F Allard, R Blomme, J-C Bouret, I Brott, P de Laverny, B Edvardsson, Y Frémat, U Heber, E Josselin, O Kochukhov, A Korn, A Lanzafame, C Martayan, E Josselin, O Kochukhov, A Schweitzer, F Thévenin, and J Zorec

1. Gaia

ESA’s Gaia mission, to be launched in 2013, will obtain accurate positions, parallaxes, and proper motions for $10^9$ objects all over the sky to magnitude $G=20$ ($V=20–22$). It will provide an incredibly accurate map of a large portion of our Galaxy. The amount of astrophysical problems that will be addressed with this wealth of data will be enormous considering the large impact of the Hipparcos mission ([1]) on all fields of astronomy. The impact of Gaia data is
strongly related to the precision and accuracy of the astrometric ($\mu$-arcsec) and photometric (millimag) catalog, both connected to the spectrophotometric data Gaia will observe combined with the astrometry. For all observed sources Gaia will obtain low dispersion spectra in the optical range that will not only be used for correcting the astrometry for chromaticity effects, but (possibly) also to characterize the objects. Since no input catalog is foreseen or even possible for Gaia, it would be of lesser impact to create a map of the Galaxy without at least some physical information about the objects. Follow-up observations of a billion of objects is simply impossible. Gaia must therefore be able to characterize the sources with a) a discrete classification of sources (e.g., single stars, galaxies, QSOs, asteroids), and possibly b) their extinction values and Astrophysical Parameters (APs), i.e., $T_{\text{eff}}$, log $g$, etc. for stars, Equivalent Width for Quasars, the Star Formation Rate for Galaxies, etc. Automated algorithms are currently developed using all knowledge for modeling astrophysical objects. Large and up-to-date libraries of stellar spectra are therefore of fundamental importance. A massive effort is ongoing to cover, as far as is currently possible, the H-R diagram for all stars to the best of our present knowledge, producing synthetic datasets of stars from the coolest to the faster evolutionary stages, and to the most evolved stars. Although we will only focus on stars in this paper, the coupling of a stellar spectrum (for all possible evolutionary stages) with stellar evolutionary models allows us to also derive global properties of stellar populations (such as unresolved galaxies) from their integrated light (we quote among others [2, 3, 4, 5]).

2. Spectroscopy on-board Gaia

Gaia will have aboard two different spectrophotometric instruments:

- The Blue and Red Photometer (BP/RP) will measure the spectral energy distribution of all objects, from 300 to 1100 nm with very low resolution (3-30 nm/pix).
- The Radial Velocity Spectrometer (RVS) will take higher resolution spectra of all objects to $V$~16, in a limited wavelength region around the near-IR Ca ii triplet for deriving radial velocity information of the majority of stars (i.e. for cool stars see [6]). The RAVE survey observes the same spectral region with large diagnostic capability ([7]).

The Gaia instruments will convolve the stellar source spectra with their photometric bands. The Gaia community of providers has agreed to compute a database of spectra for all types of sources (we will only discuss stars) with a significantly larger resolving power than the final spectra. The spectra available for Gaia cover with a 3-pixels resolution:

- for BP/RP: 300–1100 nm wavelength range, 0.1 nm step (low resolution);
- for RVS: 840–890 nm wavelength range, 0.001 nm step (high resolution).

The currently available libraries are listed in Table 1. The parameter space ($T_{\text{eff}}$, log $g$, metallicity, etc.) is sampled with discrete steps, providing a sufficiently fine grid for reliable linear interpolation. The steps are chosen based on common knowledge about stellar classification for sampling each stellar subtype: from 250 K for cool stars to 5000 K for very hot stars. Steps of 0.5 dex in log $g$ or metallicity are sufficient for the interpolation ([24]). Further tests are presented in Section 3.1. We present in Fig. 1 the coverage in the $T_{\text{eff}}$-log $g$ plane for all libraries with $T_{\text{eff}}$<10000 K compared to the widely used BASEL library ([25]). The comparison with two overplotted isochrones, for 10 Gyr (blue dotted line) and 6 Myr (black dotted line), underscores our efforts to cover the complete parameter space. These spectral libraries have an unprecedented uniformity (same steps and same metallicity) in parameter space coverage and resolution (see Table 1 for the references). Different codes are used to produce them; the BASEL and Munari libraries cover the entire parameter space. The other codes are strongly optimized for specific purposes or $T_{\text{eff}}$-regimes, and provide a more adequate treatment of all physically important
Table 1. Synthetic stellar libraries computed for Gaia. The separation into two sets is explained in Section 6. UCD indicates Ultra-Cool Dwarfs, while RSG means Red Supergiants.

| Name          | $T_{\text{eff}}$ (K) | $\log g$ | $[\text{Fe}/\text{H}]$ | References |
|---------------|----------------------|----------|--------------------------|------------|
| Munari        | 3500–47500           | 0.0–5.0  | -2.5–1.0                 | [8]        |
| O, B stars    | 15000–50000          | 1.0–5.0  | -5.0–1.0                 | [9]        |
| wind, mass loss |                     |          |                          |            |
| Ap/Bp stars   | 7000–16000           | 4.0      | +0.0                     | [10]       |
| abundance     |                      |          |                          |            |
| A stars       | 6000–16000           | 2.5–4.5  | +0.0                     | [11]       |
| MARCS         | 2800–8000            | -0.5–5.5 | -5.0–1.0                 | [12]       |
| PHOENIX       | 3000–10000           | -0.5–5.5 | -2.5–0.5                 | [13]       |
| UCD           | 400–3000             | -0.5–5.5 | -2.5–0.5                 | [14]       |
| C stars       | 4000–8000            | 0.0–5.0  | -5.0–0.0                 | [12, 15]   |
| Be            | 15000–25000          | 4.0      | +0.0                     | [16]       |
| radius considered; |                  |          |                          |            |
| WR            | 25000–51000          | 2.8–4.0  | +0.0                     | [16]       |
| rate considered; |                   |          |                          |            |
| WD            | 6000–90000           | 7.0–9.0  | -                        | [17]       |
| sdOB          | 26000–10000          | 4.8–6.4  | +0.0                     | [18]       |
| $\Delta \log g = 0.2$ |             |          |                          |            |
| NLTE Cool stars | 4000–6000        | 4.5–5.5  | +0.0                     | [19]       |
| Fast Rotators | 9000–25000           | 3.6–4.2  | +0.0                     | [20]       |
| RSG           | 3000–43000           | -1.0–0.5 | +0.0                     | [21]       |
| MARCS RVS     | 2800–8000           | -0.5–5.5 | -5.0–1.0                 | [22]       |

processes in stellar atmospheres; NLTE effects, dust models for very cool stars, effects of mass-loss, emission from circumstellar envelopes, variations of single element chemical abundances, magnetic field, and $\alpha$-enhancement. These libraries are described in further detail in [26]. It is important to emphasize that, although many codes can compute 3-D stellar atmospheres with very high complexity and accuracy, the time required to calculate a single stellar spectrum is currently prohibitive even for modern computers. For this reason the PHOENIX library for Gaia is for example computed in the 1-D approximation. In the following section we will mainly focus on the low-resolution database. It is however important to emphasize the coherence between the two databases at the different spectral resolutions. Each spectrum in the BP/RP database corresponds to one in the RVS database computed with the same code. It allows us to simultaneously simulate a source in both the BP/RP and the RVS domain, providing consistent spectra.

3. AP algorithms for Gaia
As we explain above, automated algorithms are currently under development for characterizing the observed sources. The algorithms are developed for processing both a given dataset and to exploit the synergy with astrometry. Figure 2 shows BP/RP spectra computed by the Gaia Object Generator (GOG, [27]) after convolving them with the response function of the instrument. The sequences in temperature and gravity are shown for the BASEL library. The impact of using other libraries is shown below.
Figure 1. $T_{\text{eff}}$-log $g$ plane for a sample of libraries presented in Table 1 for $T_{\text{eff}}<10000$ K with solar metallicity. We do not show NLTE Cool Stars and peculiar A-type stars because they cover a very narrow region. The coverage by the BASEL library is shown shaded to facilitate the comparison. The [8] library is not reported since it covers the same region of the BASEL library for $T_{\text{eff}}>3500$ K. The isochrone of an old population (blue dotted line) and of a young population (black dotted line) for solar metallicity from the Padova family ([23]) is overplotted.
Figure 2. GOG simulations in the BP/RP range. Left-hand panel: temperature sequence: $T_{\text{eff}}=4000, 8000, 15,000$ K. Left-hand panel: gravity sequence at $T_{\text{eff}}=3000$ K, from supergiant to dwarf stars.

The Coordination Unit 8 of the Gaia Data Processing and Analysis Consortium is working on a set of algorithms to exploit the different measurements Gaia will provide. The algorithms are both supervised and unsupervised, and will perform physical classification or provide astrophysical parameters. To summarize them:

- the Discrete Source Classifier (DSC) is used to assign an astrophysical class (QSO, star, galaxy) to an object using BP/RP spectra, magnitude, and astrometry. DSC has a modular design with subclassifiers using only photometric data, or in combination with astrometric data or astrophysical priors. The algorithm has been recently used to photometrically classify BHB stars (see [28] for details).

- GSP-Phot: is responsible for the determination of astrophysical parameters from BP/RP spectra and astrometry. Different algorithms are under evaluation for testing the effect of the inclusion of priors in the Bayesian treatment (see for example [29]).

- GSP-Spec is responsible for the analysis of the RVS spectra of FGK stars. Different algorithms are under development, including distance minimization, projection, and pattern recognition.

- the Extended Stellar Parametrizer (ESP) considers hot stars and peculiar objects, such as Emission Line stars, Ultra Cool stars, cool active stars, etc. Hot stars suffer from a poor diagnostic potential at this $T_{\text{eff}}$ of the RVS wavelength region (see Blomme, this Volume).

All the algorithms are developed/trained with synthetic datasets. The available databases of synthetic spectra are used to couple the spectra to predictions from evolutionary track models (yielding mass, radius, age, etc.) for building an ‘emitting source’. The spectrum is then simulated using the Gaia Object Generator (GOG, [27]) to obtain spectroscopic, astrometric,
Figure 3. Results of interpolation tests (see Section 3.1). The rebuilding test for one of the MARCS grid nodes with 1 nm sampling is shown for different interpolators (in 1-D, linear or cubic) and compared to the original spectrum. The interpolation is performed in $T_{\text{eff}}$ (top panel) or $\log g$ (bottom panel).

and photometric measurements. Extinction is applied to the spectra being one of the desiderata of the final Gaia catalog.

3.1. Interpolation procedure
The preparation of training data involves interpolation. The available libraries are computed with fine, but discrete, steps. Each library is simulated by adding the extinction at discrete steps to test the behavior of the simulated spectrum for various parameters. No preferred choice of libraries is made, i.e., all are simulated even in overlapping regions. Next, the libraries are interpolated inside their domain to create a continuous coverage of the AP space. We neither extrapolate nor mix different libraries and linearly interpolation them. As mentioned before, [24] provided an upper error limit of $\sim 10\%$ introduced by this method. However, the tests were limited to the RVS domain for larger resolution. We can therefore consider the error value valid for the RVS wavelength region. To test the interpolation procedure for libraries with 0.1 nm sampling, we computed a subgrid of synthetic spectra with the Linux porting of the Kurucz’s codes ([30, 31] both at the nodes and at half of the nodes for $T_{\text{eff}}$, $\log g$, and $[\text{Fe/H}]$, where
$T_{\text{eff}} = 3500, 5000, 8000, 15000$ K, $\log g = 3.0, 4.5$, and $[\text{Fe/H}] = -2.0, +0.0$. Next, the semi-nodes spectra are reproduced by linear interpolation and compared to the properly computed ones. The interpolation of $[\text{Fe/H}]$ is only tested around solar metallicity. The interpolation introduces an additional uncertainty of at most $\sim 10\%$ around $3500$ K, which rapidly drops to $\sim 4\%$ around $5000$ K, and even less at larger $T_{\text{ef}}$.

A problem with the MARCS library results from missing nodes due to problems with the convergence of the model. It considerably limits the final AP space of the simulations, especially at small $T_{\text{eff}}$ and/or very low gravity, for which model convergence are a problem. We tested the possibility to compute by interpolation at least some of the missing nodes. The test is shown in Fig. 3 for an existing node of $T_{\text{eff}} = 4750$ K, $\log g = 1.5$, and $[\text{Fe/H}] = +0.0$, computed with 1-D linear interpolation (from 2 adjacent existing nodes) or by cubic interpolation (with a parabola using 3 existing nodes). The tests show the very limited error introduced by cubic interpolation in $T_{\text{eff}}$ or $\log g$. If we consider $\lambda > 400$ (relevant for Gaia) the error is less than $10\%$. Tests at lower temperatures (in the Ultra cool dwarfs regime, see Table 1) instead favor linear interpolation. The non-linear effects become too large by mixing spectra with a temperature difference larger than 2 steps (equal to 200 K), which occurs with cubic interpolation.

4. Comparison of the libraries

An inspection of Fig. 1 shows the simultaneous coverage of some regions of the parameter space by different spectral libraries. Different codes adopt different approaches for handling the same physical process in the same physical conditions. Moreover, they can use different atomic or molecular line lists (see Plez, this Volume) or different abundance scales. The BASEL and Munari libraries use Kurucz's 1-D codes with very basic assumptions (even if NLTE effects are introduced in the calculations). Their advantages and drawbacks are well-known in the literature (see among others [8, 32]). Other codes focus on specific AP regions and introduce a more advanced treatment of specific processes, as described above. It is difficult to correctly compare spectra from different libraries. It is generally known that for an observed spectrum an ad hoc tuning of the computation parameters is possible that allows each code to produce a model in good agreement with the observation. A set of reference stars, critically compiled and with well-determined parameters for comparison, is missing. The definition of ‘well-determined parameters’ is ambiguous, also discussed in Heiter’s contribution (this Volume). ‘Who decides which ones are the parameters?’ was asked at this workshop. Binary stars can be used to pinpoint the model atmospheres with some caveats (see [33, 34, 35]), but the question is still open and perhaps awaiting interferometry and asteroseismology. The problems with the differences between different libraries, not only in the Gaia context, have been addressed in various papers, both for details about line reproduction and in terms of broad-band colors. We quote among others: [8] where Main Sequence stars are reproduced using spectra from different libraries and their colors are compared with photometric calibrations. They find that discrepancies are present for all libraries, including peculiar behaviors. [32] test colors and Lick indexes for some of the libraries presented here ([26]) using broad-band colors compared to observations, and find an agreement of up to $0.1$ for optical colors. In Fig. 4 the differences among spectra belonging to different libraries are shown for a cool dwarf ($T_{\text{eff}} = 4000$ K). The upper panel shows the original spectra with full resolution. The BASEL spectrum is only shown for comparison. The huge difference in the slope compared to all synthetic libraries is certainly something to consider; given the fact that the BASEL library is calibrated using broad-band colors of real stars ([25]), the spectral continuum should properly reproduce the actual one used in the literature. The resolving power for the BASEL library is very different from the other libraries and causes a fraction of the apparent discrepancy, however not completely, and it certainly does not explain the different slope of the spectrum at longer wavelengths. The residuals of each library to MARCS are shown slightly smoothed to simplify
Figure 4. Comparison of spectra from different libraries computed for the reported parameters. The BASEL spectrum is plotted for comparison. In the top panels, the synthetic spectra are plotted and the differences with the MARCS spectrum are shown (smoothed for better visualization, using the same color code). The differences are significant, especially in the blue part of the spectrum. The computation of molecular bands is clearly still a problem. The bottom panels show the GOG simulations of the same spectra and the differences with the MARCS spectrum simulations. Due to the lower resolution most of the differences are washed out, but residuals of up to 10% remain.

The plot. The differences are significant for matching molecular bands at all wavelengths. In the red part of the spectrum the slopes agree, but an offset remains. In the lower panels the same comparison is shown for GOG simulations of Gaia spectra. The huge differences in the blue decrease to 10%, while in the red the libraries are very similar. While the cool stars represent
Figure 5. As in Fig. 4, but for a star of larger temperature. Differences are less significant but still present. Note the different slope of the PHOENIX library (top panel).

a problem for the models, the A-type stars are supposed to be well-known. In Fig. 5 the same comparison is shown for $T_{\text{eff}} = 8000$ K. Except for the low-resolution BASEL spectrum, of course, the lines reproduction is in good agreement among the libraries (residuals are computed with respect to the A stars library). However, the slope of the PHOENIX spectrum is very different. The GOG simulations are performed at constant apparent magnitude (i.e. same area). The strong differences in the blue region do not only translate into differences in the blue part of the simulated spectrum, but also affect the overall simulation. It results in the offset of PHOENIX residuals shown in the bottom panel.
Given all of this, Gaia will rely on the available models and the algorithms, however complex, and will classify and derive parameters compared to them. The differences among libraries appear to be small at first glance, while in Fig. 2 variations of temperature and gravity appear to have a stronger effect. However, Gaia is supposed to determine $\log g$ with a precision of 0.5 dex for cool stars, i.e. $1/4$ of the variations shown in the gravity sequence. So, even small differences among the input libraries can have an effect, which must be considered. Tests are now ongoing to determine possible systematic errors due to the mixing of libraries. The DSC algorithm appears to be affected by it, since it assigns to the simulated spectra of the A-stars library a lower rate of misclassification than to the BASEL and MARCS spectra in overlapping $T_{\text{eff}}$ regions. Its causes are still under investigation. A first test with GSP-Phot uses a subsample of the BASEL library simulations for training. Next, the algorithm is applied to the remaining BASEL or MARCS library simulations. The conclusion is that a possible problem exists; the dispersion in the determination of APs (compared to the true values) in the BASEL-MARCS case becomes larger by $\sim30\%$, and worse for K-type stars. Further tests, not involving BASEL with its very low initial resolution, will explore the issue further and give hints to the providers of the libraries, possibly allowing them to focus their efforts on strongly affected AP regions.

5. Single Stellar Population

Gaia will observe $10^7$ unresolved galaxies. Single Stellar Populations (SSPs) are an essential tool to model and interpret the spectral energy distribution of clusters and galaxies. In an SSP, all stars are coeval, share the same chemical composition and the relative proportion among them is set by the Initial Mass Function (IMF). For a given age and metallicity, theoretical stellar models are first used to compute the distribution of stars along the isochrone following an IMF. Next, the appropriate stellar spectrum is assigned to each star and summed up to obtain the integrated light. To simulate a galaxy, SSPs can be combined following a Star Formation History (SFH) and a chemical evolutionary model. For assigning the spectral energy distribution to stars we use the Gaia libraries listed in Table 1. The first part of the table lists the already implemented libraries, while the inclusion of other libraries is ongoing. With the code presented by [37] we compute different SSP datasets, coupling different sets of isochrones (Padova family, [23], [38], [39], and the Teramo family [40]) with the appropriate Gaia libraries according to chemistry (see [41]). Mixing different spectral libraries is always a problem given the above discussed differences. In the literature this is usually a problem of resolution and of coverage. Thanks to the Gaia providers this problem has been overcome. The differences among the different libraries decrease with increasing temperatures. We carefully avoid mixing different libraries at smaller $T_{\text{eff}}$, but instead compute different SSP datasets for a given isochrone. It allows us to further compare the libraries in a different way. Extensive tests on the results (Lick indexes calculation and comparison with observed clusters) are ongoing (see [42] for preliminary results). We find for example in Fig. 6 a very good agreement between the Lick indexes computed on our SSPs (Padova isochrones and Gaia libraries dataset) and the observational data of Globular Clusters from [36]. We show the H$\beta$-Mg2 diagram, where H$\beta$ is a good age indicator for ages above 3 Gyr, while Mg2 is both sensitive to metallicity and $\alpha$-enhancement.

6. Conclusions

We present a homogeneous set of spectral libraries computed with state-of-the-art codes that are currently used in the Gaia community. We compare them for three cases representative of cool, A-type, and hot stars, and find noticeable differences between them. The impact of these differences on the Gaia algorithms that aim to classify and characterize the observed sources, are non-negligible. We are implementing the different spectral libraries into SSP codes, mixing them in different ways but avoiding a mixture inside critical regions of the HR-diagram.
**Figure 6.** Comparison in the Mg2-H$\beta$ diagram between the indexes computed for the SSP grid (Gaia libraries and Padova isochrones) at different ages and metallicities, and with data of Globular Cluster observations of [36].

synthetic spectral libraries and the SSPs will become available to the community through the Italian Space agency Science Data Center website (ASDC, [www.asdc.asi.it](http://www.asdc.asi.it)).

**Acknowledgments**
The authors would like to thank the referee Dr. U. Heiter (Dept. of Physics and Astronomy, Uppsala Univ., Uppsala, Sweden) for the valuable comments.

**References**
[1] Perryman M A C, Lindegren L, Kovalevsky J, Hoeg E, Bastian U, Bernacca P L, Crézé M, Donati F, Grenon M, van Leeuwen F, van der Marel H, Mignard F, Murray C A, Le Poole R S, Schrijver H, Turon C, Arenou F, Froeschlé M and Petersen C S 1997 *A&A* **323** L49–L52
[2] Fioc M and Rocca-Volmerange B 1997 *A&A* **326** 950–962
[3] Vallenari A and Sordo R 2011 *MmSAIS* **18** 205–+
