

The power of XP Gaia spectrophotometry and Self-Organizing Maps to analyze the evolutionary state and physical properties of Milky Way stars.

Lara Pallas-Quintela¹, Daniel Garabato¹, Minia Manteiga¹, and Carlos Dafonte¹

¹ CIGUS-CITIC, University of A Coruña

Abstract

The Gaia space telescope is able to gather a vast amount of data from each of the sources it explores. Among them, are the Blue and Red Photometer (BP/RP) spectra, also known as XP spectra. Most of those sources (and their physical properties) are easy to identify, but for some others, it is necessary to perform a deeper study on them. More precisely, at the Outlier Analysis (OA) working package, led by the University of A Coruña, it was created an artificial intelligence algorithm able to determine the spectral type of those atypical objects, which is followed by a template-matching procedure.

In this work, it is introduced a method to extract representative sources to be used as templates from a Hertzsprung–Russell diagrams according to the evolutionary stage of the stars of which it is composed. Therefore, we will be able to make a more refined star-classification for the following Gaia’s Data Release.

1 Introduction

On June 13th, the Gaia Space Mission made available to the scientific community its third data release, also known as Gaia DR3 or GDR3. GDR3 made a huge difference with respect to the previous ones by being the first time that spectra gathered from the telescope were published. Among the data products, are BP and RP spectra, which are a low-resolution type of spectra that, nevertheless, are very powerful for extracting the properties of the sources they represent. The Outlier Analysis (OA) working package [4], is in charge of using artificial intelligence (AI) algorithms to group similar BP/RP spectra which are considered to be atypical or outliers, and finally try to find its representative spectral type. Such task is performed following an unsupervised learning process using Self-Organizing Maps (SOM)[3].

Since it is being used an unsupervised AI algorithm, if we want to set the spectral type of each neuron in the trained map, it is important to have a well-characterized subset of sources which are representative of the objects Gaia gathers, also known as templates. In the current work, we are introducing how important is this task and how it is possible to extract templates for stars from Hertzsprung–Russell (HR) diagrams. Before diving into the

template extraction method, it will be made a brief introduction of SOM Maps and the results achieved at OA in DR3.

2 Self-Organizing Maps

The machine learning algorithm implemented at OA is a neural network-based technique which aims to group a set of data in a multidimensional grid (typically 2D) preserving the topological order of the input elements, that is, similar input spectra will be grouped in similar output neurons, the well-known Self-Organizing Maps [3]. It was in GDR3 when the first OA results were published by processing 56 million sources. Such outcomes can be graphically manipulated through the [GUASOM platform](#) [1]. In Figure 1 are shown the neurons which matched any of the reference templates, making a classification according to the evolutionary state of the stars.

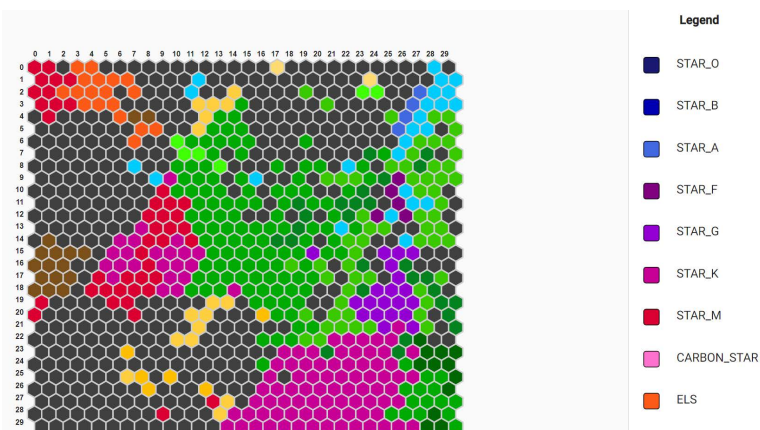


Figure 1: Result SOM Map from OA. Available at the [GUASOM platform](#)

2.1 The importance of selecting representative templates

Since the AI model is trained following an unsupervised classification algorithm, the training dataset does not contain any tagging information about the sources being used. This means that the model is just generated by the similarity of the processed signals. Nonetheless, at OA we want to tag the neurons for their sources to be characterized according to their spectral type. For such, it is necessary to follow a neuron-tagging process once the model is trained.

Among the statistics obtained from the training process, it is computed the “prototype” of each neuron, which is a spectrum representing the average of every BP/RP spectra contained in such neuron. For performing the neuron-tagging process, it is checked which template is more similar to the prototype of the neuron and, afterwards, it is assigned to the neuron the spectral type of such template.

To make the classification as meaningful as possible, it is important that the templates are as representative as possible of the data OA deals with. In order to go on refining the

neuron-tagging process, it was developed a method to extract templates from an HR diagram, that will be introduced in the following section.

3 Template extraction from HR diagrams

In order to achieve a finer star-classification for OA in DR4, we are developing a method to extract template sources from the fiducial line (line traced on highly populated areas across M_g) of a low-extinction HR diagram based on the ones introduced in [2] for DR2. We built our own based on DR3 sources and, in addition to the conditions stated in [2], we set a residual threshold value for extinction of $a_g < 0.003$, getting an HR diagram composed of 16.7 million low-extinction stars. The distribution of such stars is shown in Figure 2.

3.1 Method followed for extracting templates from HR diagrams

Once the general idea was explained, it is time to dive into the method followed, which is summarized in the following four steps:

1. **Dividing the HR diagram into 3 different areas.** The sources under the diagram are divided into *white dwarfs*, *giants* and *main sequence*. Even though white dwarfs were already differentiated from the other stars at OA in DR3, it is also interesting to differentiate between giants and main sequence stars, since, it is also interesting to analyse the feasibility of using XP spectra to differentiate stars according to their evolutionary stages, not necessarily relying on their magnitude and colour values. As shown in Figure 3, for the same spectral type (in this case, G stars), depending on their stage, even though both spectra are quite similar, it is possible to find differences between them, corroborating such differentiations according to the evolutionary stages of the stars.

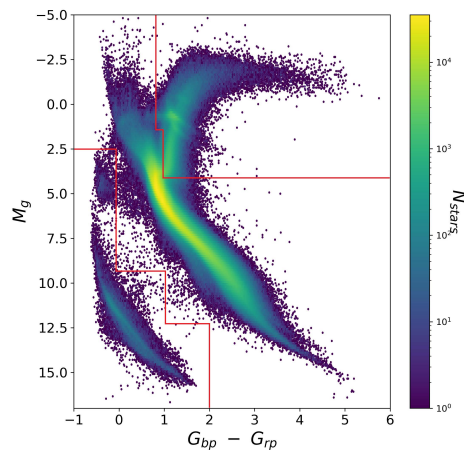


Figure 2: HR diagram used for the template selection, divided in the three main areas.

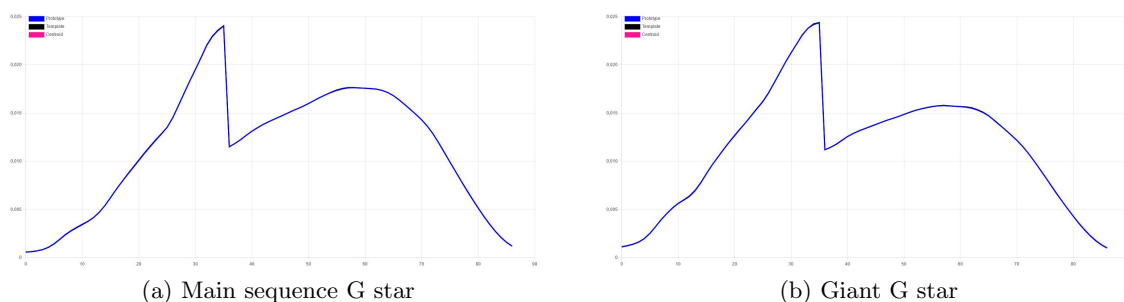


Figure 3: Comparison of the BP/RP spectra of two G stars, one from main sequence (left) and the other one from the giants branch (right).

2. **Building the fiducial line.** The fiducial line represents the most populated regions of each of the previously defined areas in terms of magnitude. Even though this procedure worked really well for both the main sequence and giant branch areas, we were still unable to do so for white dwarfs. In the latter case, for white dwarfs located at very close points in the HR diagrams, we can find major differences among spectra for not to be able to group them in the same spectral type in a reliable way.

To carry be able to create the fiducial lines, the main sequence and the giants area is divided in small regions in terms of magnitude to find its most populated point. Afterwards, such points were interpolated to create a mathematical representation of the most populated regions, as shown in the following Figure:

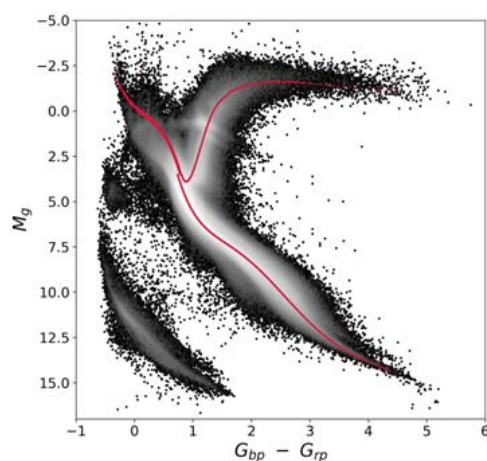


Figure 4: Fiducial lines for *main sequence* and *giant branch* stars.

3. **Selecting templates under the fiducial line.** In this step, there were selected the candidate sources to be used as templates for tagging a SOM Map. For such, biased

spectra were removed and, to simplify the list of templates, we only picked some samples of the ones which were very similar among them, since, as it was previously stated, we only use spectra both for training and labelling the neurons. In this way, we make sure that our templates are built from really representative sample of the stars we are dealing with.

4. **Assigning a spectral type to every template.** Finally, it is assigned a spectral type to each of the templates. Following [5], it was made a colour-Teff transformation and, afterwards, a Teff-spectral type transformation using the tables at [6]. In both cases, it was necessary to take into account if they were main sequence or giant branch stars.

4 Conclusions

With the method explained in this paper, we were able to create a simple and reliable method for extracting templates from an HR diagram following a MK classification. With regard to the next Gaia release, it will be possible for OA to offer to the scientific community a more detailed version of what we did for GDR3.

As it was previously mentioned, this is a reliable algorithm since we are picking spectra from highly populated areas. That is, there is a good sample of stars in the same region with very similar spectra. As a consequence, it is possible to obtain a subset of well-characterized objects that can be used not only for the template-matching procedure followed at OA, but also to explore specific properties that can characterize each type of spectra.

Acknowledgments

We wanted to acknowledge “Xunta de Galicia” for the scholarship granted to the PhD researcher Lara Pallas, with code ED481A 2021/296, for developing her PhD Thesis.

References

- [1] Álvarez, M., Dafonte, C. et al., 2022, *Neural Comput & Applic*, 34, 14
- [2] Babusiaux, B. et al., 2018, *A&A*, 616, 29
- [3] Dafonte, C., Garabato, D. et al., 2018, *Sensors*, 18, 16
- [4] Delchambre, L., Bailer-Jones, C.A.L. et al., 2022, *A&A*, 36
- [5] Mucciarelli, A. and Bellazzini, M., 2020, *A&A*, 616, 29
- [6] University of South Iowa, Spectral type characteristics