Validation of GAIA BP-RP deblended spectra by Self Organizing Maps

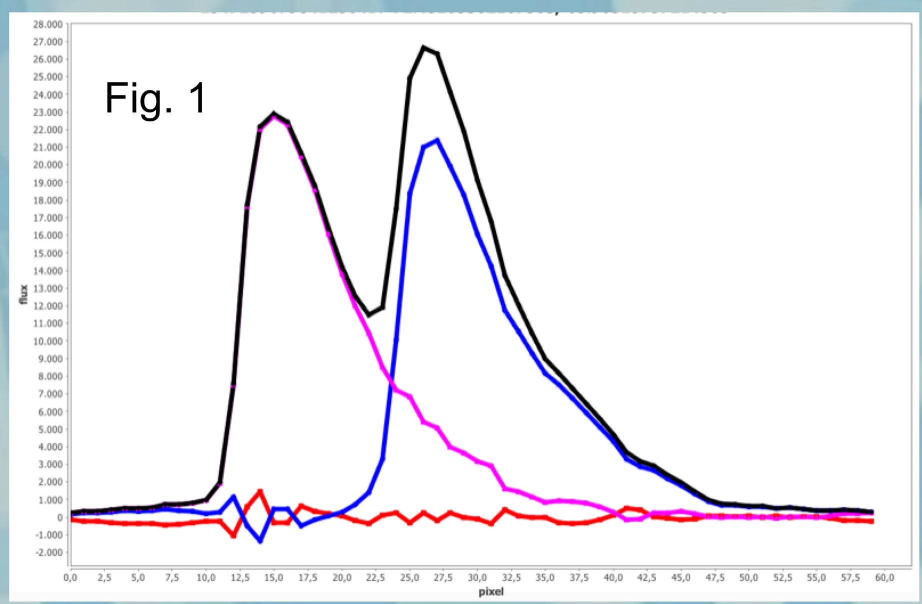
Luigi Pulone[1], Marco Castellani[1], Giuliano Giuffrida[2]

(1) INAF-Osservatorio Astronomico di Roma, Via di Frascati, 00078 Monte Porzio Catone, Roma, Italy
(2) Biblioteca Apostolica Vaticana, INAF Associated

ABSTRACT

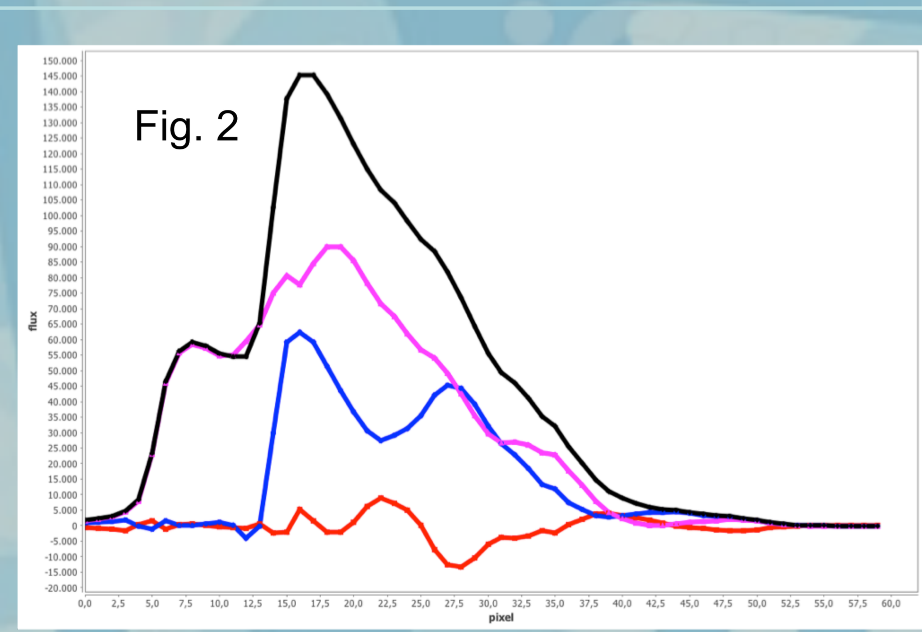
GAIA is the ESA mission to chart a three-dimensional map of the Galaxy. GAIA will provide unprecedented positional, proper motions, and radial velocity measurements with the accuracies needed to produce a stereoscopic and kinematics census of more than one billion stars in our Galaxy and throughout the Local Group.
One of the most challenging issues with GAIA is the analysis of crowded fields. More than 20% of Blue and Red low dispersion spectra of blended stars could be recovered if a proper deblending is included in the photometric processing.

Our group has developed a code to disentagle deblended spectra in crowded fields which could be eventually implemented in the next GAIA Data Release DR3. We show how a new validation technique based on Self Organizing Maps (SOM) could be able to classify deblended BP-RP spectra in order to discriminate good results from the bad ones in a very efficient way. Preliminary tests have been conducted on GAIA simulated transits coming from the Tycho2 catalogue of double stars



Depending on the quality of the transit, not all the deblending solutions are so effective and a good method to validate the process is mandatory. Sometime the deb lending algorithm based on Principal Component Analysis provides bad results with low residuals.

Figure 1 Example of a good fit of a double star spectrum. The x axis indicates the 60 bins in wavelength of the spectrum, while the y axis show the counts (electrons) of the flux.
In black the observed blended transit, in pink the first extracted spectrum, in blue the second extracted spectrum, and in red the residuals from the flux of transit and that one of the two extracted spectra.
Figure 2: Example of a bad fit with low residuals.



The ultimate goal is to obtain a realiable "quality index" coupled with each deblended spectrum, in order to safely discard the great majority of non physical situations. Apparently good results of the deblending (evaluated by the very small residuals obtained) do not necessarily imply the good quality of the fit. As a matter of fact, the adopted algorithm based upon Principal Component Analysis can obtain good deblending even with non physical shape. Moreover, the validation based on Root Mean Square has further limitations, because it assumes that the theoretical reference templates cover the full range of the spectral features and do not account for the possibility that a spectrum may share features with more than one template. Another drawback consists in eventual non-gaussian uncertainties which may cause an RMS matching not reliable.

Therefore, additional criteria are needed in order to assess the quality of the fit, besides the simple evaluation of the residuals. In the following we describe a further approach for validating the deblending outputs based on Self Organizing Maps (SOM).

The SOMs (Kohonen T., Schroeder M. R. & Huang, T. S., 2001, Self-Organizing Maps, 3rd edn, Springer Verlag) are unsupervised neural networks, which map high-dimensional data to a low dimensional space maintaining the closeness relationships the data have in the multi-dimensional space. The low-dimensional space consists of a bi-dimensional grid of neurons connected with each other. Similar input patterns are expected to be mapped to neighboring neurons in the output grid.

The idea is to build up an NxM map in which the BP, and RP spectra (input layer) can be clustered on the basis of their shapes. This can be done directly by feeding the Kohonen map by a set of deblended spectra without any comparison with theoretical templates. After the subdivision of spectra in classes, the selection of good spectra becomes an easy task.
Our two SOMS, one for BP and the other for RP spectra , have two different layers of cells. A sixty dimensional input layer (the spectra) and a competitive two dimensional (12x6) layers. The learning accomplishes both the clustering of the input data, and the spatial ordering onto the 2D map in order to maintain the distance that the input have in the multi-dimensional space. Moreover, SOMs classify the dataset without supervision, and there is no needing of target vectors, and backpropagation of errors as in the Feed Forward NNs.

In our simulation more than 3000 simulated transits of Tycho 2 double stars have been deblended. Thus, the input layer has been made up of 6000 spectra for each channel. The BP, and RP SOMs have been built up in the R environment, and the embedded “kohonen” library package has been used. Figure 3, and 4 represents the SOMs for the two spectral channel. The classes in which the spectra are subdivided are embedded in bold defined regions. It is possible to discriminate good and bad spectra on the basis of visual inspection of the neurons which show ta representative spectrum of the class in which the results are classified

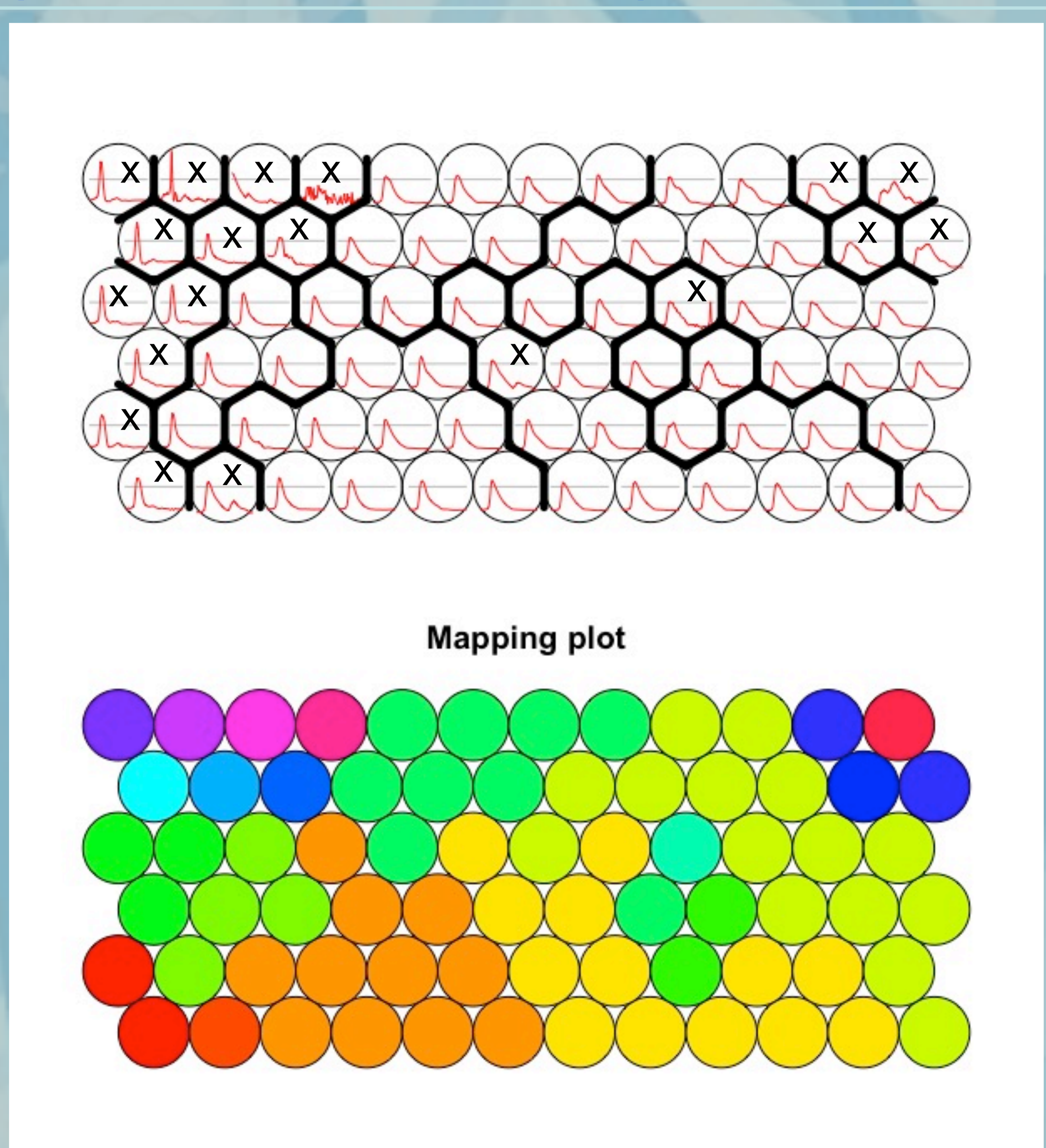
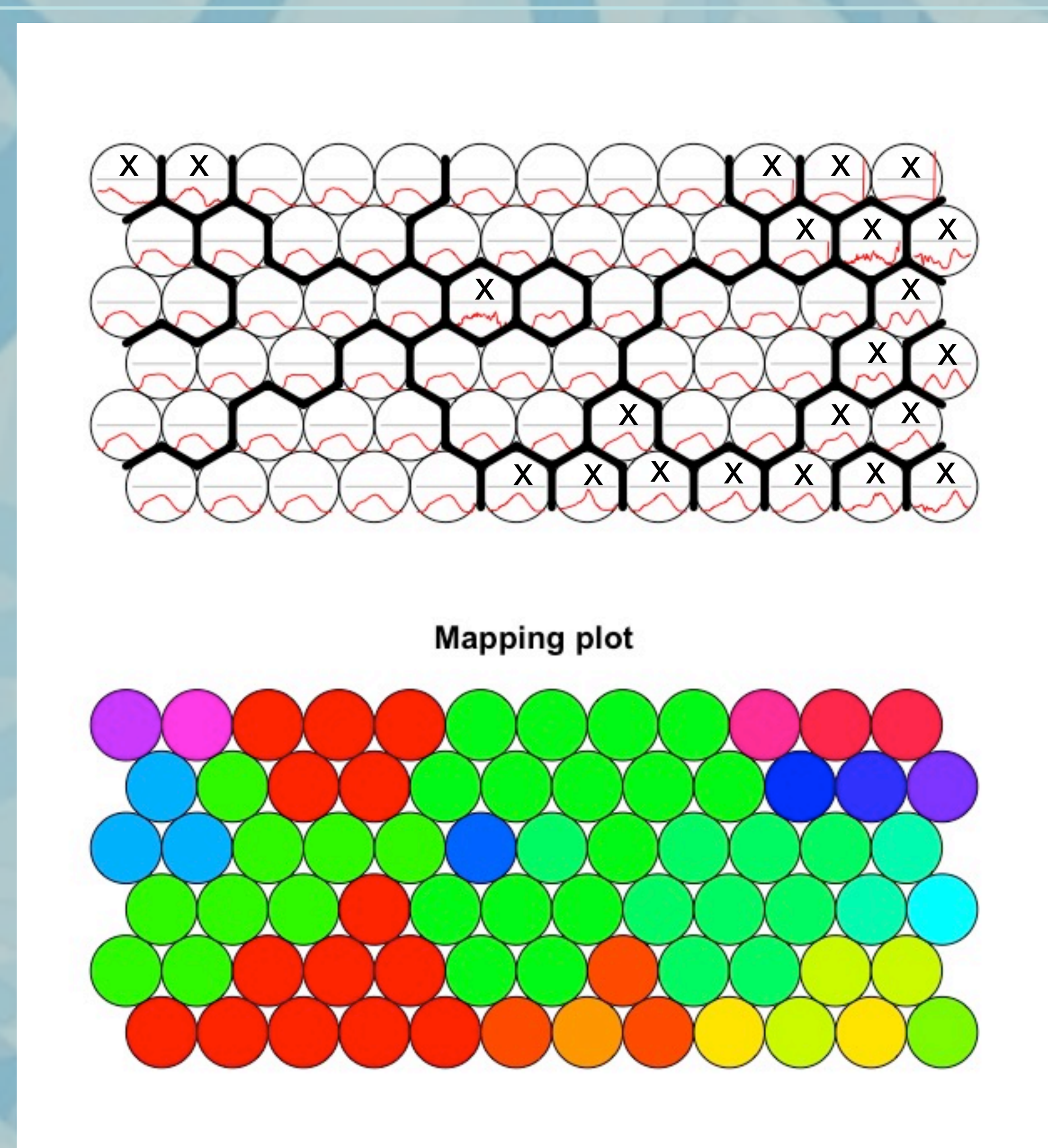


Fig 3, (BP spectra) and 4 (RP spectra) shows the 12x6 SOMs in which all the deblended spectra are subdivided in 20 classes. The codebook vector are shown for each cell. The net realizes a sort of classification in terms of similarity among the flux distributions (i.e. astrophysical parameters).
Exploiting this classification map the exclusion of bad spectra becomes an easy task. The mapping plots on the bottom represent in different colors the 20 classes found by the SOMs.

A visual inspection of the BP SOM permits to identify 89% of good deblending for the BP and RP channels.

“X” symbol indicates bad results.



The experiments in validating BP, and RP deblended spectra of simulated transits by mean of SOM used in a predictive way seems to be an optimal and fast technique