MINIMUM DISTANCE METHOD OF CLASSIFICATION APPLIED TO GAIA SIMULATED PHOTOMETRIC DATA

V. Malyuto

Tartu Observatory, 61602 Tartumaa, Tõravere, Estonia

ABSTRACT

The minimum distance method of classification is applied here to the most recent simulated photometric data for the Gaia-2 configuration. The classification accuracies of astrophysical parameters have been estimated for some selected stars (they are among the so-called ‘scientific targets’ for which the photometric system will be designed). It has been found that for reliable classification it is useful to include not only the neighbours of a star in the photometric data space but also their neighbours in the parametric space. The method will be applied to future simulated data based on a modified Gaia instrument design and for the finally proposed photometric system.

Key words: Gaia; Photometry; Mathematical procedures; Stars: fundamental parameters; Stars: classification.

1. INTRODUCTION

Gaia will gather photometric data (in selected filters) which will be used for classifying stars across the entire HR diagram. Reliable classification methods should be applied to these data for determining at least four astrophysical parameters (APs): effective temperatures, gravity, metal abundance and reddening for all stars brighter than the limiting magnitude (some authors call this classification process parametrization). A commonly-used method of classification is now the minimum distance method, MDM (see, for example, Bailer-Jones 2003). Comparison of methods and their modifications is certainly important and may help with choosing optimal classification algorithms.

2. TARTU CLASSIFICATION PERFORMANCE USING MDM

To classify stars in Gaia, we should have photometric parameters (PPs) for programme stars (which we would like to classify) and for standard stars (templates) with the known APs. At least 10 or more PPs per object should be measured; photometric measurements are dependent on at least four APs: $T_{\rm eff}, \log g$, [M/H] and E_{B-V}. A grid of templates should cover the entire HR diagram.

We begin with the same approach as in Bridzius & Vansevičius (2002). A weighted metric distance between a programme star and templates in the photometric data space is calculated using a classic formula:

$$\Delta = \sum_{i=1}^{N} w_i (P_{0i} - P_i)^2 / \sum_{i=1}^{N} w_i \qquad (1)$$

where N is the total number of PPs, P_i are PPs for the programme star, P_{0i} are PPs from the grid of templates, w_i is the weight of each individual PP. The weights are assumed to be equal to $1/\sigma^2{}_i$, where σ_i are the standard errors of the corresponding PPs for the programme star. We try here, as well as in Malyuto & Shvelidze (2004), an approach which we call here the *standard version of MDM* which contains the following steps:

1. Calculating the metric distances between the programme star and every template and finding the nearest neighbour (NN) which has the shortest distance in the photometric data space (we call it the nearest photometric neighbour) among all templates.

2. Adding some more photometric neighbours whose metric distances are within a fixed neighbourhood of the distance for the NN. This neighbourhood is defined as the shortest distance + the shortest distance *K, where K is a factor called the neighbourhood size, this K should be determined by trials.

3. For obtaining the classification result we simply average the appropriate APs for all chosen photometric neighbours with their weights (the weighting factors are inversely proportional to metric distances and are normalized in such a way that their sum equals to 1).

This standard version of MDM is similar to Soubiran’s approach as described in Brown (2003). Some classification results obtained with the use of the standard version of MDM for selected stars have been presented and discussed by Malyuto & Shvelidze (2004).

However sometimes (especially for small K) we may deal with the degenerated cases when one or more APs for all chosen photometric neighbours are exactly the same as for the NN. Therefore there is no real interpolation between APs when the APs are averaged in such cases and the classification results could be fictitious. To avoid such degeneration of the classification algorithm, we try here also the *extended version of MDM* which contains the following steps:

1. Repeating steps 1 and 2 from the standard version of MDM.

2. For every photometric neighbour (chosen at the previous step) we detect its neighbours in the parametric space (81 in all), we call them parametric neighbours.

3. For every photometric neighbour we obtain the classification result by weighted averaging of the APs for its photometric neighbour and corresponding parametric neighbours.

4. Calculating the final classification result by the weighted averaging of the classification results obtained at the previous step.

3. APPLICATION TO THE GAIA-2 CONFIGURATION

The most recent grids of simulated photometric data for Gaia-2 configuration are used: 1) 116 144 templates (Jordi et al. 2003a), regular grid; 2) 20 000 programme stars with the known APs (Jordi et al. 2003b), non-regular grid, there are 20 ‘observations’ per every star. The classification is performed for the 1X photometric system with G = 18 mag.

For the present analysis we have selected some stars from Jordi et al. (2003b) having wide ranges of effective temperatures and gravities; their APs are chosen to be close to those of some scientific targets (STs) for which the photometric system will be designed (Jordi et al. 2003c), we call these selected stars STs too. Classification with two versions of MDM (described in the previous Section) has been performed for every ‘observation’ of the selected STs, the classification accuracies are estimated as the r.m.s. differences between the calculated and known APs for each ST. Some typical examples are presented in Figures 1-3, spectral types are taken according to the calibration: spectral MK types versus APs from Straizys (1992). Mostly both versions of MDM provide rather similar classification accuracies. However at least in some cases the standard version of MDM certainly provides fictitious classification accuracies because of degeneracy of the classification algorithm described in the previous Section. The most convincing example is for the ST with $T_{\rm eff}$=13797 K, $\log g$=3.28, [M/H] =0.14, E_{B-V}=0.0 (Figure 1) where the classification accuracy of E_{B-V} is only 0.005 at the neighbourhood size K=0.1 (0.03 at K=0.4). However closer consideration shows that for 19 ‘observations’ of 20 (11 of 20 at K=0.4) the classification results are based on the value E_{B-V}=0.00 which is the same for all photometric neighbours; therefore the classification accuracy is fictitious. The classification accuracy of E_{B-V} obtained with the use of the extended version of MDM is 0.03 at K=0.1 (0.04 at K=0.4) and should be more realistic. The choice of a universal neighbourhood size (the same for all STs and providing the best classification accuracies) is somewhat uncertain. Judging on Figure 3, the best choice for K is about 0.5; judging on Figure 2, it is about 0.2, and the choice is very uncertain for Figure 1. We accept now the best value K=0.4 as a compromise.

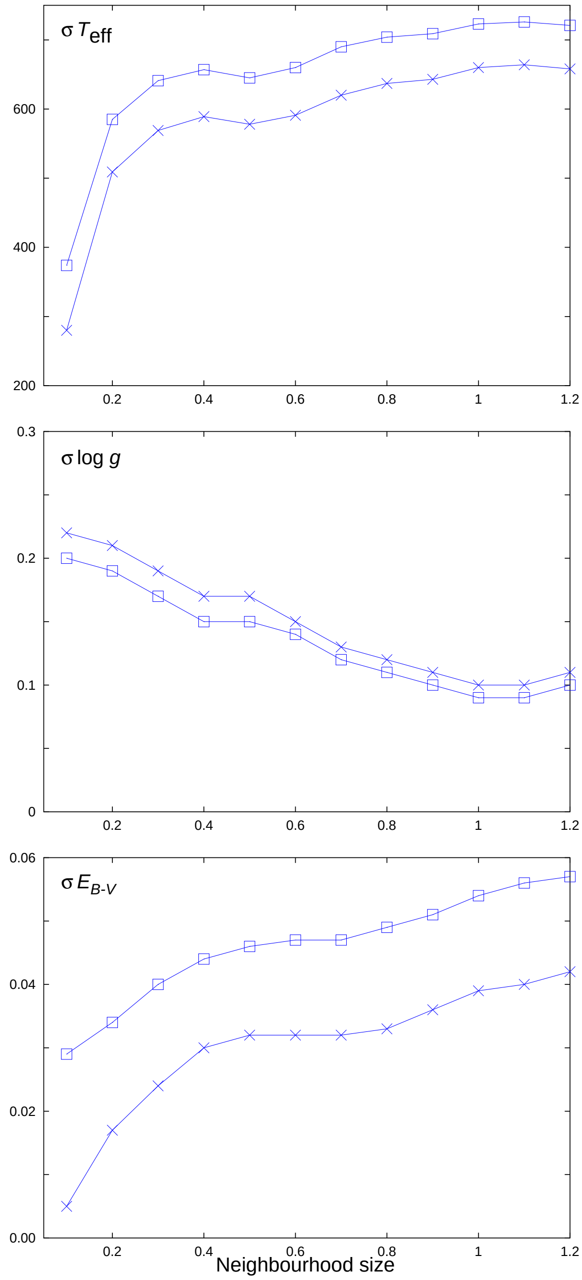

Figure 1. Classification accuracies obtained with the use of MDM for the ST with $T_{\rm eff}$ = 13 797 K, $\log g$ = 3.28, [M/H] = 0.14, E_{B-V} = 0.0 (spectral type B6III*) at different neighbourhood sizes. Crosses correspond to the standard version of MDM, open squares correspond to the extended version of MDM.*

4. CONCLUSIONS

In an attempt to improve the minimum distance method of classification, a conception of photometric and parametric neighbours (in the photometric data space and in the parametric space, respectively) among templates has been considered. For some selected stars the classification has been performed using the photometric neighbours only (the standard version of MDM) and using the photometric and parametric neighbours both (the extended version of MDM). We argue that the extended version of MDM provides more realistic results and should be preferable. We plan to apply the method to future simulated data based on a modified Gaia instrument design and for the finally proposed photometric system.

ACKNOWLEDGMENTS

Support from the Estonian Science Foundation (grant No. 4702) is gratefully acknowledged.

REFERENCES

Bailer-Jones, C.A.L., 2003, Gaia Spectroscopy, Science and Technology, U.Munari ed., ASP Conference Series, Vol. 298, 199

Bridzius, A., Vansevičius, V., 2002, ApSS, 280, 41

Brown, A.G.A., 2003, Gaia technical report, Gaia-ICAP-AB-003

Jordi, J.M., Carrasco, J.M., Figueras, F., 2003a, Gaia Technical Report, UB-PWG-013

Jordi, J.M., Carrasco, J.M., Figueras, F., et al., 2003b, Gaia technical report, UB-PWG-014.

Jordi, J.M., Figueras, F., Carrasco, J.M., et al., 2003c, Gaia technical report, UB-PWG-009

Malyuto, V., Shvelidze, T., 2004, Baltic Astronomy, 9, accepted for publication

Straižys, V., 1992, Multicolor Stellar Photometry, Pachart Publishing House,Tucson, Arizona

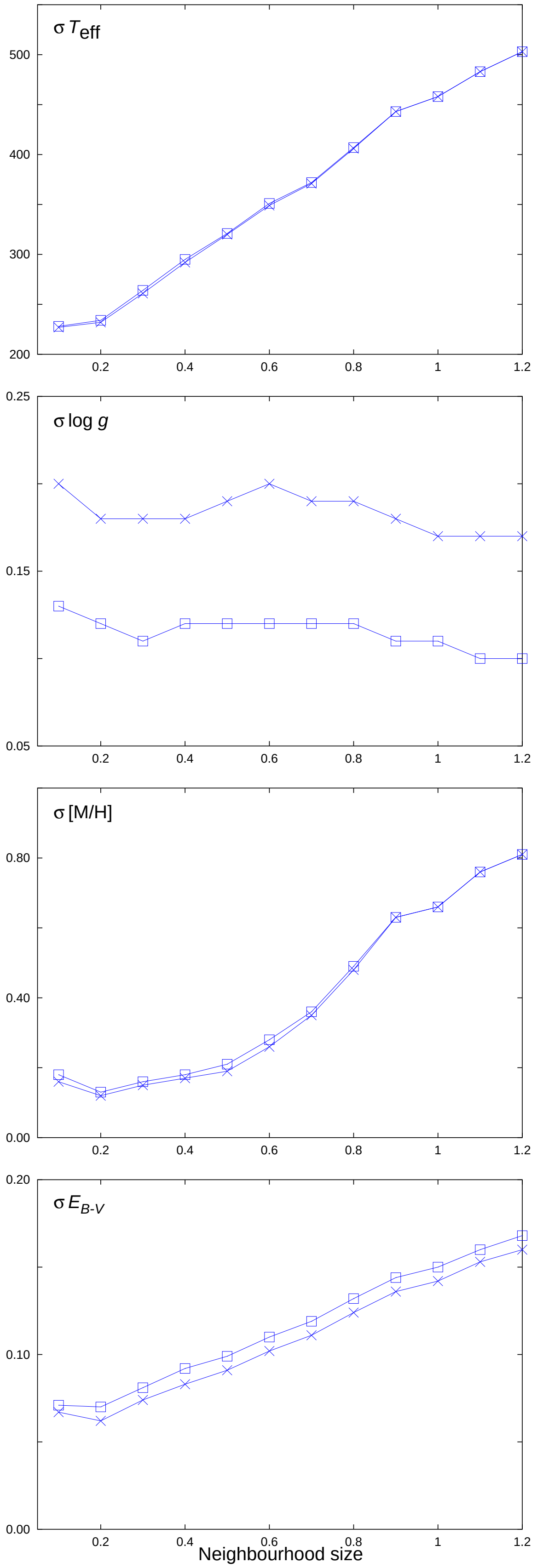

Figure 2. *The same as in Figure 1, but for the ST with* $T_{\rm eff} = 7039$ *K,* $\log g = 4.65$*,* $[{\rm M/H}] = 0.01$*,* $E_{B-V} = 0.0$ *(spectral type* F2V*).*

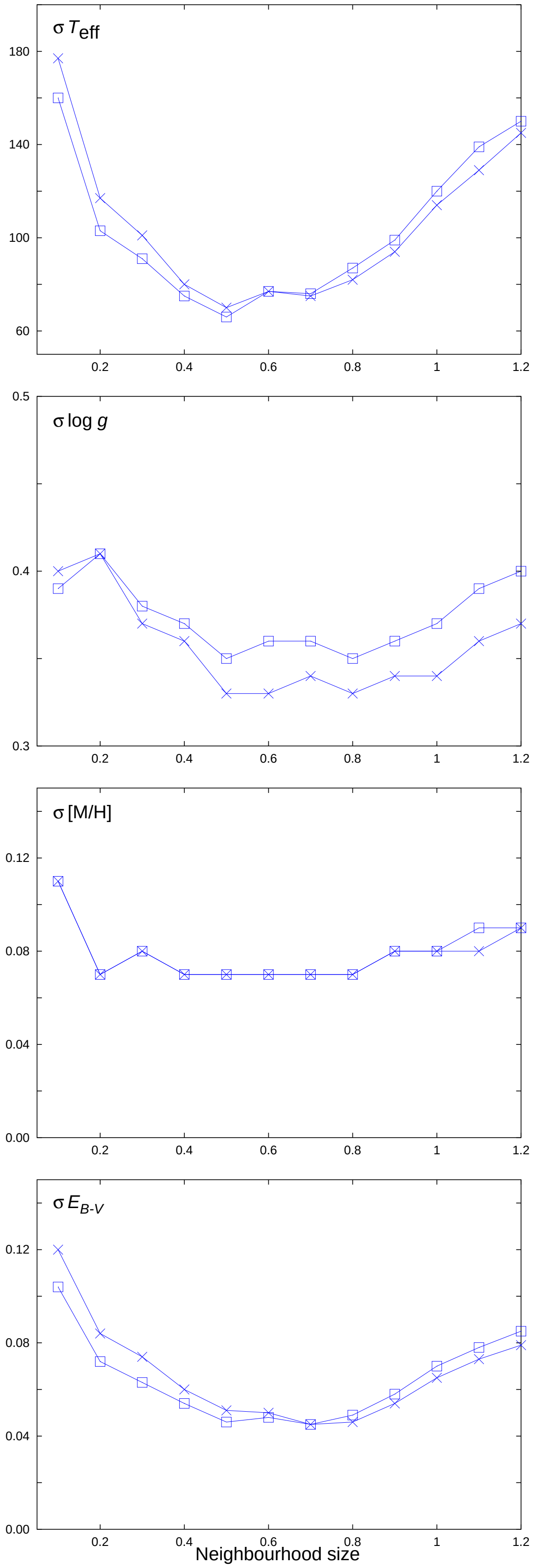

Figure 3. The same as in Figure 1, but for the ST with $T_{\rm eff} = 6034$ *K,* $\log g = 4.68$*,* $[{\rm M/H}] = 0.08$*,* $E_{B-V} = 0.20$ *(spectral type* G0V*).*