III.5

Stellar classification in CoRoT faint stars fields

C. Damiani[1,2], J.-C. Meunier[1], C. Moutou[1], M. Deleuil[1], and F. Baudin[2]

[1] Aix-Marseille Université, CNRS, Laboratoire d'Astrophysique de Marseille, UMR 7326, 13388, Marseille, France
e-mail: cilia.damiani@ias.u-psud.fr
[2] Université Paris-Sud, CNRS, Institut d'Astrophysique Spatiale, UMR8617, 91405 Orsay Cedex, France

1. Introduction

The extremely precise photometric time-series provided by the faint stars channel of the CoRoT mission serve a very large number of scientific objectives. In addition to major contributions in the faint stars field, let us recall also that the CoRoT faint stars chanel has allowed to produce important results in stellar physics (see the numerous contributions in this book). For either the core or additional programs, both the selection of the targets and the exploitation of the light curve require at least a basic knowledge of the targets, such as their colours or spectral type and luminosity class.

This is one of the reasons why the Exodat (Deleuil et al. 2009) database was built. Its prime objective was to provide position, colours and stellar classification for the stars in the observable zones of CoRoT, the so-called "CoRoT eyes". Exodat is an information system and a database that gathers an extended body of data of various origin, and delivers it to the community through a user friendly web interface[1] (Agneray et al. 2014). It contains data for more than 51 million stars providing information for both the potential targets of the exoplanet program (with $11 \leq r \leq 16$), and also the fainter background sources. The later are needed in order to estimate the targets' level of light contamination. To help the selection of the targets and the preparation of the observations, a first order stellar classification (FOSC) was conducted using broad-band multi-colour photometry available through Exodat. Any detailed analysis of targets of particular interest would require complementary observations for a better characterisation. Indeed it has been done on a target to target basis, especially for planet hosting stars, but also on more consequent samples (see Sect. 4). Overall, no more than 15% of the targets observed over the mission lifetime were the object of complementary observations. Notwithstanding, stellar parameters are essential to the scientific exploitation of the CoRoT light curves. For the vast majority of targets, to this day, and at least until the third data release of the GAIA mission, scheduled in late 2017/2018[2], the only source of information concerning the spectral classification of all stars observed in the exoplanet channel is Exodat.

A final refined treatment of the light curves has been recently applied to the complete CoRoT data set and a new and ultimate version of N2 products has been released. The spectral classification has also been revised for this last version, taking into account the particularities of each CoRoT runs. While an overview of the FOSC method is given in Deleuil et al. (2009), they do not provide an estimation of the reliability of the final results. In this paper, we give a thorough description of the method in Sect. 2. In Sect. 3 we give the statistical error expected from the FOSC as estimated using a synthetic survey of the Milky Way. In Sect. 4 we compare our results to the classification obtained by spectroscopic campaigns. We give our conclusions in Sect. 5.

2. Classification method

The FOSC method is adapted from the automatic classification technique developed by Hatziminaoglou et al. (2002) for celestial objects classification and quasar identification. It consists of fitting the apparent broad-band photometry magnitudes of the target to the spectral energy distribution (SED) of template stars of different spectral types. The templates are taken from the Pickles stellar spectral flux library UVKLIB (Pickles 1998). This library encompasses spectral types from O to M and luminosity classes from supergiants to dwarfs with good completeness and uniformity. To avoid introducing too many degeneracies, only solar abundance templates are used in the FOSC. Hence, the classification is done over 106 different spectral types including dwarfs, subgiants, giants, bright giants and supergiants with effective temperatures (T_{eff}) comprised between $3.398 \leq \log T_{eff} \leq 4.600$. There are 36 templates for dwarf stars, which roughly correspond to a sampling of about one template every two subclasses in the main-sequence. Templates of extragalactic sources, white dwarfs and brown dwarfs are not used in the FOSC because it is expected that the number of such objects in the considered range of magnitudes and galactic regions is small enough to be safely neglected. Templates of binary stars are not used neither despite the fact that the proportion of binary stars in the sample is expected to be significant. Again, this limitation is conceded to avoid too many degeneracies, because broad-band photometry does not allow to clearly distinguish the binary nature of the target. If the spectral

---

[1] http://cesam.lam.fr/exodat/
[2] see http://www.cosmos.esa.int/web/gaia/release

type of the two components are similar, the resulting spectral classification would give the same result when using the template of a single star, since it is insensitive to the absolute magnitude of the target. On the other hand, if the spectral type of the two components are very different, the colours of the binary would largely be undistinguishable from the colours of the brightest component, due to the extremely different luminosities of the two components in the considered wavelength range. In those cases, the spectral type of a binary determined with the FOSC method will thus correspond to the one of the primary star. In intermediate cases, we expect that the spectral type determination will be degraded when the target is a binary. However to ensure the reliability of the result, the binary nature of the target should be known before attempting the classification. This was not the case prior to the mission, but the results of the planetary transit search in CoRoTlight curves has identified a number of eclipsing binaries. In this case a second order classification, suited for binary targets, would produce better results. This is out of the scope of this paper, and will be tackled in a dedicated study of CoRoT binaries.

To take into account the effect of interstellar absorption, the SEDs of the UVKLIB library are used to produce a library of reddened templates. They were created for a range of colour excess from $E_{B-V} = 0$ to $E_{B-V} = 4$ in steps of 0.05. The extinction law is taken from Fitzpatrick (1999) using the average value of the ratio of total to selective extinction $R_V = \frac{A_V}{E_{B-V}} = 3.1$. The actual value of R_V does not matter here given that in the wavelength range considered in this study (350 nm $\leq \lambda \leq 2.2\,\mu$m), the extinction curve is insensitive to R_V variations. For a given value of the colour excess E_{B-V}, the reddened flux distribution is then multiplied with the appropriate relative spectral response curves of the different bands.

The bands used are Harris B and V and Sloan-Gunn r' and i' for about two thirds of the targets, RGO U for only 3% of them, and 2MASS J, H, and K_s bands for almost all of them (98%). We will refer to this set of magnitude as the "Obscat" catalogue. When the B, r' or i' bands are not available, the PPMXL catalogue (Roeser et al. 2010) was used instead. It combines the USNO-B1.0 blue, red and infrared magnitudes with the 2MASS magnitudes and provides accurate position and mean motion in the ICRS. The photometric calibration of the USNO-B1.0 catalogue is of marginal quality. As detailed in Monet et al. (2003), the root of the problem is the lack of suitable faint photometric standards. Moreover, the original catalogue does not provide an error for each individual measurement, but it is estimated to have an accuracy of about 0.3 mag. Acknowledging that the magnitudes from USNO-B1.0 should be used with care, we used a relative error of 3% for all filters. Consequently, the quality of the fit when using USNO-B1.0 photometry is expected to be degraded. This concerns about a third of all the CoRoT targets.

Integrating the transmitted flux over the filter's wavelength range provides the reddened template fluxes in the different bands. It was chosen to use magnitudes in the Vega system, and observed magnitudes are converted to this system if required. The spectral energy distribution of α Lyrae (Vega) was treated in the same way to obtain the reference flux in every band.

For each target, the apparent magnitudes are converted to fluxes which are normalised to the flux in the band where the photometric error is the smallest. The photometric error ϵ_i is also converted to an error in flux σ_i, taking into account the error on the zero point ϵ_{0i} for the corresponding filter i, using the relationship[3]

$$\sigma_i = \frac{2.3}{2.5}\sqrt{\epsilon_i^2 + \epsilon_{0i}^2}\,. \tag{1}$$

The observed flux is then confronted with the computed fluxes for all the reddened templates by computing

$$\chi_j^2(E_{B-V}) = \sum_{i=1}^{n} \frac{\left(F_{\rm obs}^i - F_{\rm mod_j}^i(E_{B-V})\right)^2}{\sigma_i^2} \tag{2}$$

where $F_{\rm obs}^i$ is the observed flux in the i band, $F_{\rm mod_j}^i(E_{B-V})$ is the flux in the same band for the template j reddened with a colour excess E_{B-V}, σ_i is the estimated error of the observed flux in this band and n is the number of bands that are used.

Finding the minimum χ^2 then leads to the attribution of a spectral type and a colour excess that best reproduce the observed magnitudes. The reduced χ^2 is used as quality index QI

$$\text{QI} = \chi^2/(\text{n} - 3). \tag{3}$$

As already pointed out by Hatziminaoglou et al. (2002), a classification scheme based on the χ^2 technique may lead to degeneracies, which are class-dependent. Multiple minima may occur in the parameter space and such degeneracies can only be solved by including additional information in the classification procedure. In the following we describe the strategy used to avoid the main degeneracies.

2.1. Dwarfs and giants degeneracy

The effect of luminosity on stellar energy distributions is often seen in some metallic lines that are usually shallow and/or narrow. It is thus very difficult to distinguish dwarf and giant stars when adjusting SEDs to broad-band photometry only. To avoid strong degeneracies with the FOSC method, we first use a preselection tool of the luminosity class based on a colour-magnitude diagram (CMD) before the identification of the best-fitting stellar template. For each observed field, the J vs. $J - K$ CMD plane is divided in two regions defined by a linear function that is adapted to the mean extinction of the field. Stars are hitherto identified as dwarfs or giants, and their subsequent classification is done separately. The method is particularly reliable close to the Galactic plane where the typical extinction is enhancing the gap between the bluer dwarfs and redder giants. Specifically, the bluer part of the diagram should not contain any giants, while the redder may well contain cool dwarfs, especially at faint magnitudes (see also Sect. 3). Following this separation, the targets are then distributed to two separate pipelines: one only uses "dwarfs" templates, which includes luminosity classes IV and V, while the other only uses "giants" templates, which includes luminosity

---

[3] The numerical factor in the the right-hand-side of Eq. (1) comes from the Taylor expansion of the decimal logarithm used in the computations of magnitudes

classes I, II, and III. If the 2MASS magnitudes are not available, the classification is done using both dwarf and giant templates. In some rare instances, a particular target may have been observed by CoRoT in 2 different runs (or more) and it might result in different pre-classifications, because in regions of steep extinction gradients, the mean extinction may be different even for overlapping runs. In this case, the classification that has the minimum χ^2 is selected and the corresponding spectral type and luminosity class are adopted for all the runs.

2.2. Interstellar reddening

The main problem one faces when dealing with stellar classification is the role of interstellar extinction, which modifies the observed colours of stars. This problem is particularly important in our study, since the CoRoT fields are located within a few degrees from the Galactic Plane. A distant, strongly reddened early-type star could be identified as a closer cooler (intrinsically redder) star. One way to handle this would be to use an independent measurement of the extinction along the line of sight and set it as fixed parameter in the fitting procedure. However, this would require to know the distance to the target, which is not the case here. Consequently, we let the reddening be a free parameter in the classification scheme, but limit it within a reasonable range. Dust emission is one of the major foregrounds hampering the study of the cosmic microwave background (CMB). Thus one of the products associated with the 2013 release of data from the Planck mission[4] was a new parametrisation of dust emission that covers the whole sky, based on data from the High Frequency Instrument (HFI) (Planck Collaboration et al. 2014). We use the τ_{353} maps to constrain the range of E_{B-V} in our stellar classification. For every run, the region of the CoRoT CCD was extracted from the τ_{353} maps, and converted to E_{B-V} using the scaling factor $E_{B-V}/\tau_{353} = 1.49 \times 10^4$ (Planck Collaboration et al. 2014). We must stress that adopting the scaling factor estimated using the correlations with quasars in diffuse directions could systematically overestimate E_{B-V} in denser regions. Thus we took the value of the 90% percentile of the distribution of E_{B-V} for a given run to set the value of the corresponding upper bound of the range of E_{B-V} values allowed in the fitting procedure. For two runs in the centre direction located near dense clouds (SRc01 and LRc03), this value was too high to produce a significant constraint for the fitting procedure, and we set the upper bound to the more reasonable value of $E_{B-V} \leq 4$.

3. Comparison with synthetic galactic populations

Given that the method is based on χ^2 minimisation, it is expected that the greater source of uncertainty will come from the intrinsic degeneracy between spectral type and reddening. However, even if we set the reddening to its actual values, the intrinsic limitations of the method (such as the sampling in effective temperature, in luminosity classes, or overlooking metallicity) must introduce systematic errors.

In order to evaluate the systematics of the method over the entire space of parameter covered by the classification we used the Padova isochrones (Marigo et al. 2008). We computed a synthetic catalogue containing stars with any combination of spectral type and colour excess, and that is not limited by the selection of targets introduced by the instrument. This shows that in favourable conditions, the FOSC can return a spectral type that is correct to about a few subclasses. However, if the magnitudes are affected by severe reddening, or if some bands are missing, the accuracy of the result can be severely limited, especially for early-type stars lacking a measurement of U.

Stars observed by CoRoT are limited in apparent magnitudes by the instrumental capacities. In this case, it is reasonable to expect that reddening and spectral type will be correlated to some extent. Moreover, the magnitudes used for the FOSC come from several catalogues, with heterogeneous properties. Thus to estimate the average error made on the spectral type of CoRoT targets, it is necessary to include those effects into the synthetic sample. The spatial distribution of CoRoT targets spans over a wide area and different points have very different galactic longitudes, latitudes and interstellar extinction. We used the Galaxia code (Sharma et al. 2011) to generate synthetic galactic populations in each of the directions observed during the 26 runs of CoRoT. Galaxia is a publicly available code based on the Besanon model of the Galaxy (Robin et al. 2003) that is particularly suited to generate synthetic surveys over wide areas. We made no modification to the source code and use it as it is described in Sharma et al. (2011). The synthetic model of galaxy produces a population of stars that has the same statistical properties as stars of the Milky way in terms of distance to the sun, age, mass and heavy elements distribution, as a function of the pointing direction. The only major unknown is interstellar absorption. This may induce systematic differences in stellar properties because the observed sample is magnitude limited. For each run, we use the CCD footprints of CoRoT to select in the model the portion of the sky that exactly corresponds to the observed area and we adjust the synthetic magnitudes to the observed ones. This is done by introducing a correction factor that scales the synthetic E_{B-V} produced by Galaxia. This factor is obtained by minimising the difference between the synthetic distribution of $J - K$ and the ones of the input catalogue for the observed field. Next, the errors on synthetic magnitudes are generated to reproduce the systematics and random behaviour of the errors in the input catalogue. The magnitudes and errors of the synthetic catalogue is then confronted to the input catalogue by performing a Kolmogorov-Smirnov test. In all instances, the null hypothesis of the two catalogues being drawn form the same population cannot be rejected at the 0.5% level.

3.1. Luminosity class

Here, we will not attempt to assess the validity of the FOSC results in terms of individual luminosity classes, but we will examine the basic distinction between dwarfs and giants. As previously described, this is not in itself a result of the χ^2 minimisation of the FOSC, but it is based on the position of the target in a colour-magnitude diagram. The result of this classification is binary, with stars being either dwarfs or
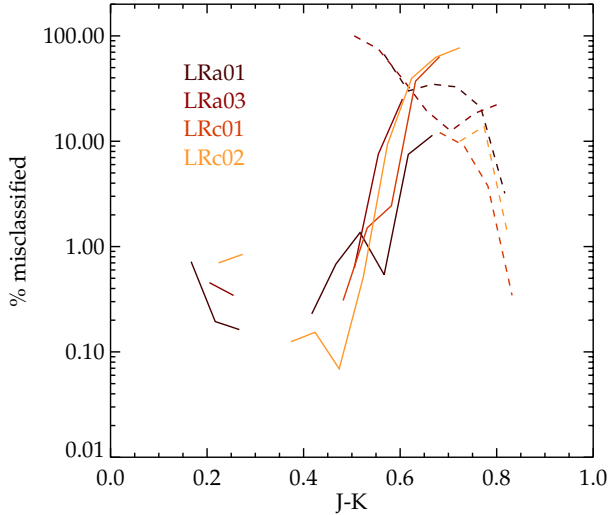
Fig. III.5.1. Percentage of misclassified targets as a function of $J - K$ for different runs. The percentage of dwarfs (stars with $\log g \geq 3.5$) wrongly classified as giants (luminosity class I, II or III) is given in solid lines, and the percentage of giants (stars with $\log g > 3.5$) misclassified as dwarfs (luminosity class IV or V) is given in dotted lines. The colours of the lines correspond to the different runs, as indicated on the plot.

giants. The significance of this classification can be assessed using the synthetic catalogue by producing a $(J - K, J)$ CMD and apply the selection criterion used for the targets. We can then compare the classification as dwarf of giants to the $\log g$ of the synthetic stars. Results are displayed in Fig. III.5.1 for a few representative runs.

We see that overall, the classification is more often wrong for giants than for dwarfs, as expected. For $J - K < 0.5$ and $J - K > 0.8$, there are less than 1% of the stars in a class that are misclassified, because there is no overlap of the two populations in this part of the diagram. In this case, the luminosity classification can be considered correct to a very high level of significance. The greater uncertainty evidently lies within the stars that have a $J - K$ close to the criterion used for the selection, some of which being almost certainly misclassified. However, due to the gap between dwarfs and giants in a $(J - K, J)$ CMD, this actually represents a very small percentage of the total sample. Integrating over the values of $J - K$ yields between 1 and 10% of the misclassified stars in the dwarf sample. Respectively, between 3 and 15% of stars classified as giants are actually dwarfs. Overall, the percentage of wrongly classified stars in the entire sample is less than 7%. Note finally that when the J or K magnitude, or both, are missing, the preselection between dwarfs and giants can not be performed. In this case, the FOSC is run using both dwarfs and giants templates, and the luminosity class is thus decided by the χ^2 minimisation. This results in severely degraded performances, with 10 to 50% of the stars classified as dwarfs that are actually giants, and from 30 to 65% of stars misclassified stars in giants, depending on mean extinction. Moreover, the spectral classification is also affected, with typical errors multiplied a few times. Extreme caution should be taken when dealing with those few stars for which infrared magnitudes are unavailable. Let us stress again that this concerns only a very small proportion of the whole sample of CoRoT/Exoplanet targets.

3.2. Effective temperature of CoRoT targets

The majority of targets have Obscat magnitudes, which produce a smaller uncertainty on the spectral type compared to what is obtained when using the lower quality USNO-B1.0 magnitudes. We have created a population representative of the whole sample by merging the catalogues obtained for the different runs, and performing the FOSC taking magnitudes from the two catalogues in proportion of what they are in the population of targets. The stars are randomly attributed to one or the other catalogue, and the process is repeated 100 times. Each time, the median of the relative error is computed. As a measure of the spread of the distribution of the relative error, we compute the difference between the values of the 15th percentile and the 85th, which we note 2σ. Then we take the average of those values found over the 100 trials. Overall, we find an average median absolute temperature difference $|\Delta T_{\text{eff}}| = 533 \pm 6$ K for the whole sample of stars classified as dwarfs and $|\Delta T_{\text{eff}}| = 280 \pm 3$ K for the whole sample of giant stars. The corresponding standard deviation is of about 925 ± 5 K for dwarfs and 304 ± 4 K for giants. This is the performance expected from classification methods using broad-band photometry. For example Farmer et al. (2013) give a median difference of 500 K for the dwarfs of the Kepler input catalogue, but they perform better for giants with a reported median ΔT_{eff} of 50 K. This is probably due their use of different filters, and especially their custom D51 filter, which is sensitive to surface gravity (2011). Nevertheless, we can safely conclude that the spectral classification obtained by the FOSC method is generally good up to half a spectral class for the majority of CoRoT targets.

4. Comparison with spectroscopic data

To complement the statistical study presented in the previous section, we compare the result of the FOSC to the spectral type obtained by independent measurement such as mid or high-resolution spectroscopy. We have chosen, in the available literature, catalogues of spectral types containing a large number of targets and obtained with homogenous methods. As a result, we compare the FOSC with two different spectral catalogues. The first one contains the fundamental stellar parameters for 1127 stars in the LRa01, LRc01 and SRc01 CoRoT fields (Gazzano et al. 2010, hereafter Gaz10). The second one gives an estimation of the spectral type obtained for 11466 stars in the IRa01, LRa01, LRa02 and LRa06 fields (Sebastian et al. 2012; Guenther et al. 2012, hereafter jointly cited as SG12).

In terms of luminosity class, the agreement between the FOSC and spectral methods is very good for main-sequence stars. We find that more than 97% of stars classified as dwarfs are also attributed a $\log g \geq 3.5$ by Ga10, and 85% of dwarfs are also classified as such by SG12. On the other hand, there are only 75% of stars classified as giants by the FOSC that also have $\log g < 3.5$ from the Ga10 analysis and 70% that agree with the classification of SG12. This is a bit lower than what was estimated from the synthetic sample, for which the proportion of correctly classified giants was between 85 and 97 %, but still consistent within the uncertainty given by Ga10 and SG12.
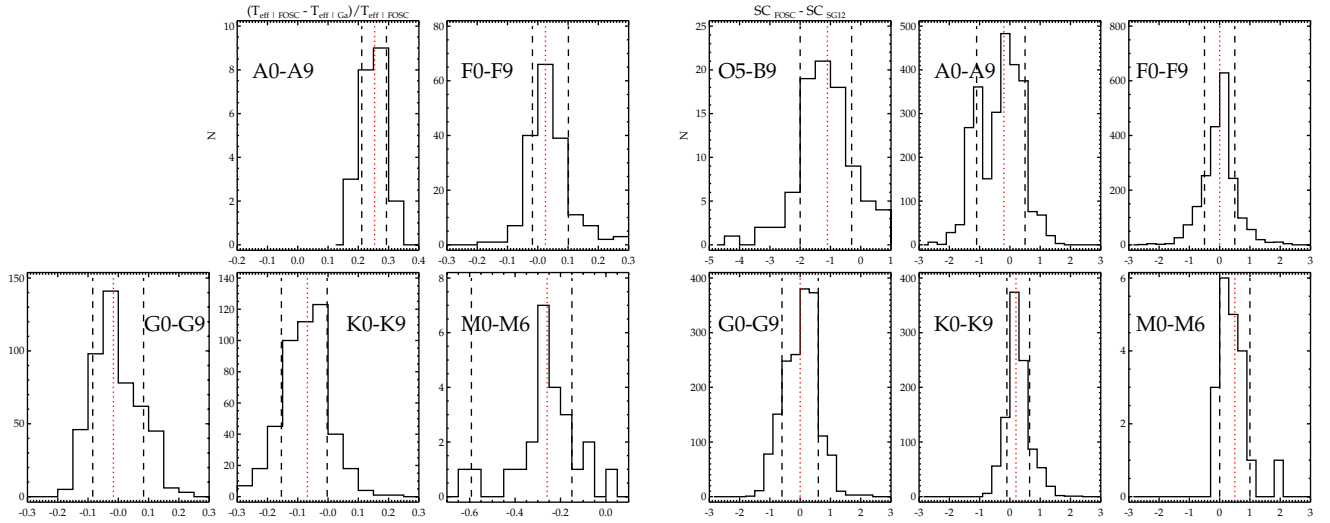
Fig. III.5.2. *Left*: distribution of relative temperature difference found by Ga10 and the FOSC. *Right*: distribution of difference in spectral class found by SG12 and the FOSC. The spectral class alphabetical scale has been converted to a numerical scale, attributing numbers from 0 to 8 to the classes O through M. Furthermore, the sub-class is converted as a decimal value. Each plot corresponds to a subsample selected by spectral class as determined by the FOSC and labeled on the plot. The vertical red dotted line shows the median value and the black dashed lines give the value of the 15th and 85th percentiles.

Regarding the spectral type, as can be seen in Fig. III.5.2, there are consistent systematics when comparing the FOSC to Ga10 and SG12. In both cases, early-types found by the FOSC tend to overestimate the effective temperature and later-types tend to underestimate it [5]. This trend was clearly found with the synthetic catalogue for hot stars. Except for M-type stars, the precision and accuracy of the FOSC are similar when estimated with the synthetic galactic population and with results from spectroscopy. Typically for late-type stars, the median error is less than about $\pm5\%$ and the interval between the 15th and 85th percentiles has a spread of 10 to 20% in effective temperature, or about 1 spectral class. Those values are doubled for earlier type-stars. The comparison with Ga10 indicates that the uncertainty for the effective temperature of M-type stars is of the same order, but it is difficult to see if it is also the case when comparing with SG12 due to a very small sample size. It is possible that the greater error in the classification of M stars is related to their metallicity. We recommend exercising caution when using the sample of stars classified as M-type with the FOSC.

5. Conclusions

In this paper, we have described the procedure used to estimate the spectral type of CoRoT targets. This information is delivered together with the CoRoT data, and is available online through the Exodat database. Due to the faintness and the number of the targets, we use a method based on broad-band photometry. The photometry obtained during a mission-dedicated observational campaign are used whenever possible, but it was necessary to resort to poorer quality photometry for about a third of the targets. Using a synthetic catalogue, we were able to quantify how the performance of the classification are affected by the quality of the photometry. We also provide a statistical estimate of the precision and accuracy of the classification. We find consistent results by confronting the result of the FOSC to other independent estimates of the spectral type. We find that the luminosity class is estimated with great significance, and that the typical uncertainty on the spectral type given by the FOSC is of half a spectral class.

Our method cannot provide a precise error bar for individual CoRoT targets but is valid in a statistical way. Let us note that in particular there are two CoRoT runs pointed in directions of particularly high interstellar absorption (SRc01 and LRc03) where the validity of these results may be limited. However, the strong reddening in this case dims the magnitude of distant targets to a point where they exit the fainter magnitude limits of CoRoT targets. This effectively removes almost all giants of the target sample, and filters out the dwarfs that would be too severely reddened to ensure good performances of the FOSC. It can be expected thus that the performances of the FOSC on the remaining targets, not strongly reddened, would be equivalent to what is shown here. Finally, this study has made clear that interstellar absorption is the main source of uncertainty on the spectral type. The performances of the FOSC are significantly improved by limiting the range of possible values for the reddening. This was only performed in the latest version of the FOSC, and previous versions are expected to produce significantly more uncertain results.

---

[5] This results in negative and positive spectral class difference respectively because they are sorted by order of decreasing effective temperature

References

Agneray, F., Moreau, C., Chabaud, P., Damiani, C., & Deleuil, M. 2014, in Astronomical Society of the Pacific Conference Series, Vol. 485, Astronomical Data Analysis Software and Systems XXIII, eds. N. Manset, & P. Forshay, 203

Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, A&AS, 106, 275

Deleuil, M., Meunier, J. C., Moutou, C., et al. 2009, AJ, 138, 649

Farmer, R., Kolb, U., & Norton, A. J. 2013, MNRAS, 433, 1133

Fitzpatrick, E. L. 1999, PASP, 111, 63

Gazzano, J.-C., de Laverny, P., Deleuil, M., et al. 2010, A&A, 523, A91

Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, A&AS, 141, 371

Guenther, E. W., Gandolfi, D., Sebastian, D., et al. 2012, A&A, 543, A125

Hatziminaoglou, E., Groenewegen, M. A. T., da Costa, L., et al. 2002, A&A, 384, 81

Marigo, P., & Girardi, L. 2007, A&A, 469, 239

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984

Pickles, A. J. 1998, PASP, 110, 863

Planck Collaboration, Abergel, A., Ade, P. A. R., et al. 2014, A&A, 571, A11

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440

Sebastian, D., Guenther, E. W., Schaffenroth, V., et al. 2012, A&A, 541, A34

Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, ApJ, 730, 3