Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **463,** 4210–4222 (2016) doi:10.1093/mnras/stw2313
Advance Access publication 2016 September 14

An all-sky catalogue of solar-type dwarfs for exoplanetary transit surveys

V. Nascimbeni,[1,2]★ G. Piotto,[1,2] S. Ortolani,[1,2] G. Giuffrida,[3,4] P. M. Marrese,[3,4] D. Magrin,[2] R. Ragazzoni,[2] I. Pagano,[5] H. Rauer,[6,7] J. Cabrera,[6] D. Pollacco,[8] A. M. Heras,[9] M. Deleuil,[10] L. Gizon[11,12] and V. Granata[1,2]

[1]*Dipartimento di Fisica e Astronomia, 'G. Galilei', Università degli Studi di Padova, Vicolo dell'Osservatorio 3, I-35122 Padova, Italy*
[2]*INAF – Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy*
[3]*ASI – Science Data Center, Via del Politecnico snc, I-00133 Rome, Italy*
[4]*INAF – Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monteporzio Catone, Italy*
[5]*INAF – Osservatorio Astrofisico di Catania, via S. Sofia 78, I-95123 Catania, Italy*
[6]*Institute of Planetary Research, German Aerospace Center, Rutherfordstrasse 2, D-12489 Berlin, Germany*
[7]*Department of Astronomy and Astrophysics, Berlin University of Technology, Hardenbergstrasse 36, D-10623 Berlin, Germany*
[8]*Department of Physics, University of Warwick, Coventry CV4 7AL, UK*
[9]*Scientific Support Office, Directorate of Science, European Space Agency, ESTEC/SCI-S, Keplerlaan 1, NL-2201 AZ Noordwijk, the Netherlands*
[10]*Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France*
[11]*Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, D-37077 Göttingen, Germany*
[12]*Institut für Astrophysik, Georg-August-Universität, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany*

Accepted 2016 September 9. Received 2016 September 8; in original form 2016 August 13

ABSTRACT

Most future surveys designed to discover transiting exoplanets, including *TESS* and *PLATO*, will target bright ($V \lesssim 13$) and nearby solar-type stars having a spectral type later than F5. In order to enhance the probability of identifying transits, these surveys must cover a very large area on the sky, because of the intrinsically low areal density of bright targets. Unfortunately, no existing catalogue of stellar parameters is both deep and wide enough to provide a homogeneous input list. As the first *Gaia* data release exploitable for this purpose is expected to be released not earlier than late 2017, we have devised an improved reduced-proper-motion (RPM) method to discriminate late field dwarfs and giants by combining the fourth U.S. Naval Observatory CCD Astrograph Catalog (UCAC4) proper motions with AAVSO Photometric All-Sky Survey DR6 photometry, and relying on Radial Velocity Experiment DR4 as an external calibrator. The output, named UCAC4-RPM, is a publicly available, complete all-sky catalogue of solar-type dwarfs down to $V \simeq 13.5$, plus an extension to $\log g > 3.0$ subgiants. The relatively low amount of contamination (defined as the fraction of false positives; <30 per cent) also makes UCAC4-RPM a useful tool for the past and ongoing ground-based transit surveys, which need to discard candidate signals originating from early-type or giant stars. As an application, we show how UCAC4-RPM may support the preparation of the *TESS* (that will map almost the entire sky) input catalogue and the input catalogue of *PLATO*, planned to survey more than half of the whole sky with exquisite photometric precision.

Key words: methods: statistical – catalogues – planetary systems – stars: solar-type – stars: statistics.

1 INTRODUCTION

Although a couple of thousand exoplanets have been discovered, the general picture on their structure, formation and evolution is still far from being complete. Putting things in context requires a more detailed analysis of individual systems than is currently achievable. This includes, for instance, the estimate of accurate stellar masses and ages through asteroseismology and a detailed characterization of the planetary atmospheres through spectroscopy. Both these tasks require high S/N, and are feasible only when targeting planets hosted by nearby and bright stars, which have been very rare indeed so far. In the next years, two cornerstone, space-based missions that will photometrically detect and partially characterize planetary systems around bright stars will be launched: *TESS* (Ricker et al. 2015), a NASA Explorer mission selected for launch in 2017–2018; and *PLATO* (Rauer et al. 2014), a medium-class mission selected for ESA's M3 launch opportunity (2022–2024).

★ E-mail: valerio.nascimbeni@unipd.it

Aside from being bright ($V \lesssim 13$), the most promising targets for exoplanetary science are solar-type stars, i.e. main-sequence (MS) stars later than spectral type F5, which could include moderately evolved subgiants following a broader definition. As a magnitude-limited sample of field stars is dominated by distant (and intrinsically bright) giants and early-type stars, the only way to access a large sample of bright, nearby solar-type stars (as required by transit exoplanet search surveys) is to dramatically increase the covered sky area up to a significant fraction of the whole celestial sphere. In this case, the selection of a homogeneous target list requires the deepest all-sky stellar classification ever attempted, able to assign at least a spectral type and luminosity class to every star brighter than $V = 13$. This should be regarded as a minimal requirement for carrying out the target selection task. Previous experience from the *CoRoT* and *Kepler* space missions (Deleuil et al. 2009; Brown et al. 2011) has shown that detailed knowledge of the stellar parameters of the targets (such as effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity $[M/{\rm H}]$, stellar mass and radius $R_\star$, $M_\star$, age, etc.) along with the identification and characterization of the background stars (Deleuil et al. 2006) helps to prioritize the targets, and makes the follow-up and the rejection of false alarms much more efficient.

The *Gaia* mission[1] (Perryman et al. 2001), launched in 2013 and currently collecting data, is expected to play a fundamental role in the target selection by performing an unprecedented ultra-high-precision astrometric survey of nearly every source brighter than $V \simeq 20$, along with low-resolution spectrophotometry and radial velocities. An intermediate catalogue which includes stellar parameters from spectrophotometry is expected to be released at the end of 2017.[2] However, we should keep in mind that (i) the *Gaia* stellar classification will be affected by crowding in the densest fields (Bailer-Jones et al. 2013; Recio-Blanco et al. 2016); (ii) there will be some degeneracy among certain parameters, such as temperature and interstellar extinction (Straižys et al. 2006; Bailer-Jones 2010; Bailer-Jones et al. 2013); (iii) other astronomical catalogues (for instance, X/UV/IR/narrow-band photometry and activity diagnostics) will be very complementary to the *Gaia* measurements; (iv) space missions (e.g. *TESS*) and ground-based surveys which are presently in development may require a preliminary target list before 2017, for the performance analysis, to optimize the observing strategy, to fine-tune the spacecraft design, and to begin implementing the foreseen additional observations and coordinated follow-up programmes.

In this paper, we first introduce the basic problem of attempting a large-scale stellar classification by relying only on wide-band photometry and proper motions (Section 2), and review the existing techniques and catalogues designed for that purpose (Section 3). After showing that a new approach has to be devised, in Section 4 we describe how we compiled a brand-new all-sky catalogue of FGK dwarfs and subgiants, called the fourth U.S. Naval Observatory CCD Astrograph Catalog (UCAC4)-RPM. We include a detailed description of the catalogues used as input, the grid-based algorithm exploited to define appropriate selection criteria, and the cross-matching procedure used to estimate the contamination and completeness of the resulting sample. Finally, in Sections 5 and 6, we discuss how UCAC4-RPM can be exploited by the ongoing or forthcoming space- and ground-based transit surveys, and how it could be extended and complemented in the future.

[1] http://sci.esa.int/gaia/
[2] http://www.cosmos.esa.int/web/gaia/release

2 THE BASIC PROBLEM

The most reliable stellar classification is provided by spectra and trigonometric parallaxes from which it is possible to extract the basic parameters of the stellar atmosphere ($T_{\rm eff}$, $\log g$, $[M/{\rm H}]$) and its distance and absolute magnitude in a relatively straightforward way. Once these parameters are known, we can use stellar evolutionary models to derive other important physical quantities such as the radius $R_\star$, the mass $M_\star$ and the age of the star. Spectroscopy and high-precision astrometry are time-consuming. Alternative approaches, based on narrow-band photometry (Árnadóttir, Feltzing & Lundström 2010), are more suited to wide-field surveys (at a price of lower accuracy), but they are still too costly in terms of observing time.

Until *Gaia* releases accurate parallaxes, it is not feasible to characterize hundreds of thousands of stars on an extremely large field of view (FOV) with the techniques mentioned above. The only available, accurate all-sky classifications of this kind are limited to sources at $V \lesssim 8$ or even brighter, based on spectroscopic surveys (HD, MK classifications; Morgan & Keenan 1973; Skiff 2014), narrow-band surveys (Geneva–Copenhagen; Nordström et al. 2004) or the space-based *Hipparcos* catalogue (Perryman & ESA 1997; van Leeuwen 2007). At the present time, we must rely on the only available catalogues that reach the $V \simeq 13$ limit: wide-band photometric catalogues (both in the visible and in the near-infrared) and ground-based astrometric catalogues, which are accurate enough to provide good proper motions, but not trigonometric parallaxes. In what follows, we present the difficulties and limitations of exploiting these sub-optimal data for stellar classification purposes.

For a detailed review of the main standard photometric systems, see Bessell (2005). A summary of calibrated broad-band colours and physical properties of typical MS, solar-metallicity dwarfs is given in tables 15.7– 15.8 by Cox (2000). A revised and constantly updated version of that table is maintained by E. Mamajek[3] (Pecaut & Mamajek 2013).

2.1 Wide-band photometry

Using wide-band photometry to derive stellar parameters is a challenging task for many reasons. First of all, over most of the visible spectrum, the expected differences in colour between cool stars of similar $T_{\rm eff}$ and spectral type (SpT), but different surface gravity $\log g$ and metallicity [Fe/H], are at most a few tenths of a mag, even when exploiting custom-designed narrow- or intermediate-band filters (Zdanavičius 2005, see their figs 1c–1f). When the spectral energy distribution (SED) of those stars is integrated over the wavelength ranges of a typical wide-band photometric system ($\Delta\lambda \approx$ 1000 Å), the differences fade to only a few hundredths of a mag. A notable exception is in the spectral region bluer than the 'Balmer jump', which is sampled by both the Johnson U and the SDSS u' filters (Zdanavičius 1998, 2005). Unluckily, these magnitudes are the most difficult to measure and calibrate from the ground, due to atmospheric effects, standardization issues and low instrumental sensitivity (Bessell 2005). So far, no all-sky, deep catalogue of U/u' magnitudes is available.

Even more importantly, wide-band photometric classifications are known to be highly degenerate on some stellar parameters, such as metallicity and gravity; see for instance the example reported by Belikov & Röser (2008), and Fig. 1 of this paper. By using a B, V, J

[3] http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt

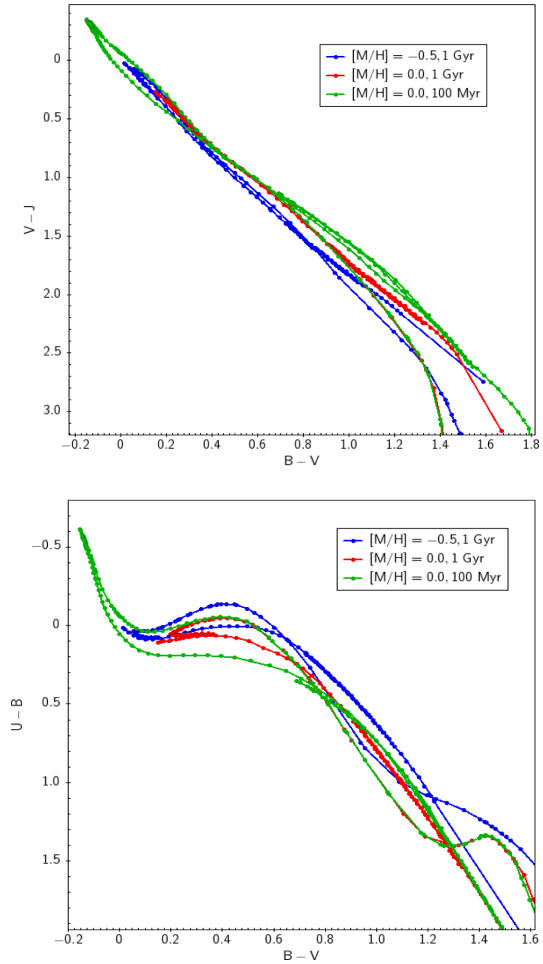

Figure 1. Synthetic $(V-J)$ versus $(B-V)$ (upper panel) and $(U-B)$ versus $(B-V)$ (lower panel) colour–colour diagrams illustrating how retrieving a unique solution for stellar parameters from wide-band photometry can be a complicated task. PARSEC isochrones (PAdova and TRieste Stellar Evolution Code; Bressan et al. 2012) are coded with different colours: $[M/H] = 0.0$ and 1 Gyr age (green), $[M/H] = 0.0$ and 0.1 Gyr (blue), $[M/H] = -0.5$ and 1 Gyr (red). There clearly are multiple cases of degeneracy where isochrones with different metallicities and ages intersect each other, in particular in the FGK region.

two-colour diagram, it is not possible to discriminate between metal-poor dwarfs and solar-metallicity giants, both of G spectral type. F, G and K stars are expected to dominate a magnitude-limited sample at $V < 13$, and they are located in the colour–magnitude diagram (CMD) and colour–colour diagram region where most degeneracy occurs. A further source of degeneracy is interstellar reddening, which is not negligible in a magnitude-limited sample of stars at low and intermediate Galactic latitudes. In particular, $E(B-V)$ (or A_V) and $T_{\rm eff}$ are degenerate: a distant, hot star can be easily misclassified as a closer and cooler star. To a lesser extent, this also holds for near-infrared (NIR) photometry, and even when parallaxes and low-resolution spectra are available (Bailer-Jones 2010, 2011).

The most basic aim of an input catalogue for transit surveys is to distinguish dwarfs from giants for the spectral types ranging from mid-F stars down to the late M ones. The most relevant spectral regions are therefore those sensitive to $\log g$. Aside from the optical ultraviolet (UV) wavelengths, the most prominent features are located on the continuum at 4000–4500 Å ('G band'; mostly for the F-G spectral types) and at the Mg I triplet/MgH absorption feature at 5150–5200 Å (mostly for G-K spectral types, as demonstrated also by Teig 2008). Most narrow- and intermediate-band photometric systems developed for stellar classification purposes, as well as the $G_{\rm red}$ and DDO51 filters originally designed for the Kepler Input Catalog survey (Batalha et al. 2010; Brown et al. 2011) are tailored on these spectral features.

As for wide-band photometry, we note that the three above-mentioned features fall roughly inside the Johnson U, B, V bands, though their contribution is greatly diluted by the bandwidth effect, and it never exceeds a few hundredths of magnitude on the $B-V$ colour. Cousins $R_{\rm c}$ and $I_{\rm c}$ bands are relatively insensitive to $\log g$, except for the spectral class M. Results from synthetic photometry applied to M dwarfs should be approached with caution, as the theoretical uncertainties of their atmospheric opacities result in large differences between calculated and observed SEDs (Zdanavičius 2005). NIR wide-band photometry, that is carried out in the JHK or $JHK_{\rm s}$ bands by ground-based surveys (such as 2MASS; Skrutskie et al. 2006), is particularly effective in discriminating gravity effects on very late spectral types. $H-K$ or $H-K_{\rm s}$ colours, for instance, can differ by a few tenths of a mag between M dwarfs and M giants (Lépine & Shara 2005). This technique can be extended to late K stars. However, NIR colour effects become smaller and very degenerate for earlier types such as F and G. On the other hand, NIR colours are less affected by interstellar extinction, and they can be advantageous on fields at low Galactic latitudes, where reddening effects are not negligible.

2.2 Proper motions

Proper motions, which are provided with a typical accuracy of a few mas yr^{-1} by existing catalogues such as those from the Tycho-2 (Høg et al. 2000) and UCAC4 (Zacharias et al. 2013) surveys, proved to be helpful in discriminating dwarfs from giants when combined with photometric colours. The so-called reduced-proper-motion (RPM) technique exploits a combination of proper motions μ_α, μ_δ and apparent magnitude m as a statistical proxy to the target distance. By crudely assuming that all field stars share the same absolute velocity pointed towards a random direction on the sphere, one concludes that the quantity

$$\mathrm{RPM}(m) = m - 5 + 5\log\sqrt{(\mu_\alpha\cos\delta)^2+\mu_\delta^2}, \quad (1)$$

called *reduced proper motion* is statistically approximating the absolute magnitude of the stars plus a constant offset.[4] RPM diagrams (RPMDs) are therefore similar to CMDs smeared along the vertical axis, and, as such, can be exploited to select specific types of stars.

RPMs are very effective in selecting MS dwarfs among bright ($V \leq 11$) stars belonging to late spectral types (K-M), as demonstrated by Gould & Morgan (2003) and Gontcharov (2009), among others. On fainter stars, and for F and G spectral types, the RPM technique is less effective, and a larger fraction of false positives is expected (Gould & Morgan 2003). RPMs, by themselves, cannot provide stellar parameters other than a very rough estimate of the absolute magnitude or SpT. In addition, selection cuts based on proper motions have to be devised with care, as the resulting sample could be biased in subtle ways towards thick disc or halo targets, which share a higher intrinsic velocity (Lépine & Shara 2005) and different metallicity.

[4] Some authors give a slightly different definition of the RPM quantity, which differs only by an unessential constant.

3 PREVIOUS CLASSIFICATION ATTEMPTS

Beginning in the 2000s, with the advent of accurate all-sky catalogues such as *Hipparcos*, Tycho-2 and 2MASS, some authors have tried to extract stellar parameters using only wide-band photometry and proper motions. Some were driven by the need to extract a target list for exoplanet searches (e.g. Ammons et al. 2006 for the NHK radial velocity survey, Gould & Morgan 2003 for generic transit surveys). Others were trying to provide an all-sky calibration for the *ugriz* passbands (Ofek 2008; Pickles & Depagne 2010), to develop a general method to classify Galactic stellar populations (Bilir et al. 2006; Belikov & Röser 2008), or to distinguish particular classes of stars, such as M dwarfs from M giants (Lépine & Shara 2005). Most of these works exploit the same input catalogues with different algorithms. Usually, they are based on Tycho-2 B_T, V_T and 2MASS JHK_s magnitudes, as they provide uniform, precise all-sky photometry over passbands that contain useful information on [M/H] and log g (see Section 2). Proper motions, when needed, are also extracted from Tycho-2. Unfortunately, most photometric classifications are limited to about $V \leq 11$ by the completeness limit of Tycho-2. While 2MASS provides very good photometry ($\sigma <$ 0.05 mag) down to $V \sim 15$ and Tycho-2 proper motions are also well complemented by the UCAC survey for stars brighter than $V \sim$ 15, no reliable source of visual magnitudes was available for $V \geq 11$ on the whole sky until recently. In what follows, we shortly review a few previous classification attempts which are most relevant to our purposes.

3.1 Template-matching techniques

Ofek (2008) matched the 2MASS and Tycho-2 catalogues, and fitted the resulting $B_TV_TJHK_s$ magnitudes with a set of library spectra computed by Pickles (1998). A best-fitting template was then assigned to each star, and a set of SDSS *griz* synthetic magnitudes was calculated for each template to construct an all-sky catalogue of *griz* magnitudes for calibration purposes. Though not specifically designed to derive stellar parameters, this work yielded spectral types and luminosity classes for ∼1.56 million Tycho-2 entries. On a subset range of spectral types, metal-rich versus metal-poor stars are also differentiated. For unknown reasons, stars having $\delta \simeq 4^\circ$, $54^\circ \leq \delta \leq 59^\circ$, $\delta \geq 80^\circ$ are lacking.

Pickles & Depagne (2010) extended the work done by Ofek by fitting updated spectral templates on a larger set of stellar magnitudes, having complemented the Tycho/2MASS $B_TV_TJHK_s$ with the photographic R_N magnitudes from USNO-B1.0 (through the NOMAD catalogue, compiled by Zacharias et al. 2004). Cuts were performed on 2MASS $J-H$, $H-K_s$ colours and Tycho-2 proper motions, attempting to distinguish giants from dwarfs when the spectral template χ^2-fitting is unable to do it. For our purposes, both the Ofek (2008) and Pickles & Depagne (2010) classifications should be handled with caution. Interstellar reddening is not accounted for, and this is known to let many distant, hot giants to be misclassified as cool dwarfs. If one only selects $V < 11$, SpT > F5 stars, the resulting sample is strongly concentrated towards the Galactic disc, and this is even more striking when selecting only K and M dwarfs, which, at bright magnitudes, should be isotropically distributed on the sky. Moreover, K and M stars dominate the sample along the Galactic plane, while A, F and G stars are expected to do so in a magnitude-limited sample. Summarizing, the classification algorithms above are not to be trusted when working at low Galactic latitudes ($b < 20^\circ$).

A more complex approach to template-matching stellar classification is that attempted by Belikov & Röser (2008), who defined a set of custom extinction-free indices Q_{123} calibrated on the Tycho-2/2MASS photometric systems ($B_TV_TJHK_s$). An interval-cluster analysis was then applied to extract $T_{\rm eff}$, log g, [M/H] by fitting Kurucz models to Q on the whole Tycho-2/2MASS set, after the algorithm has been trained on a subset having known spectroscopic parameters. Adopting Q indices also allowed the authors to constrain the extinction A_V. As a result, bright, nearby dwarfs are homogeneously distributed, except for a few sky regions where very high or anomalous extinction occurs (e.g. the Ophiuchus cloud or at very low b).

3.2 Proper-motion-based techniques

The RPM approach, introduced in Section 2.2, was first pioneered by Kapteyn & van Rhijn (1920) and Hertzsprung (1922), and applied for the first time to the Tycho-2 catalogue by Gould & Morgan (2003), purposely to select targets for transit searches. Since then, wider or improved approaches were attempted, the following one being the most relevant to our purposes.

Ammons et al. (2006) developed an entirely empirical method to identify metal-rich, low-MS stars as targets for N2K, a radial velocity search for hot Jupiters (Fischer et al. 2005). A training set made of 1000 F, G and K stars with both high-resolution spectra from Valenti & Fischer (2005) and photometry from Tycho-2/2MASS was employed to fit polynomials and spline functions to broad-band colours extracted from $B_TV_TJHK_s$ magnitudes and Tycho-2 proper motions. Those analytical functions were then interpolated on all the well-measured Tycho-2 sources with a χ^2-minimization procedure, in order to derive distances and temperatures. For a selected subset of 354 822 FGK dwarfs, [M/H] and probability of multiplicity were also derived with the same technique, while $T_{\rm eff}$ was estimated with a finer polynomial function. On FGK dwarfs with photometric errors $\sigma_V < 0.05$ mag, the temperature and metallicity models give a standard error of 70 K and 0.14 dex, respectively. The binary model can remove 70 per cent of doubles with $1.25 < M_1/M_2 < 3.0$ from a magnitude-limited sample of dwarfs at a cost of cutting 20 per cent of the sample. This technique primarily uses the distance (and hence the absolute magnitude) as a proxy to discriminate dwarfs from giants. It fails when trying to directly estimate log g. The main reasons are: (1) the physical processes that differentiate dwarfs from giants in photometry vary widely as a function of $T_{\rm eff}$: a single polynomial or even a spline cannot be expected to capture all possible effects; (2) some rare kinds of stars are underrepresented by the Valenti & Fischer (2005) training set, namely blue giants and cool red dwarfs. Though reddening is not taken into account, using proper motions to infer the distance minimizes the contamination of the sample by giants, even at low Galactic latitudes.

3.3 The contamination issue

The aim of our work is to select MS dwarfs having a spectral type later than F5 (DLF5 hereafter), possibly also including moderately evolved subgiants within the same SpT range (DSLF5 hereafter). Classification algorithms label catalogue entries as positives (i.e. 'good' targets which meet our requirements; P) or negatives (N), while the terms true (T) and false (F) refer to what these objects actually are. The four possible outcomes within this scheme are therefore TP, TN, FP and FN. Every classification technique which aims at selecting DLF5/DSLF5 stars can unavoidably (1) fail at including some objects which are indeed DLF5/DSLF5, that is it

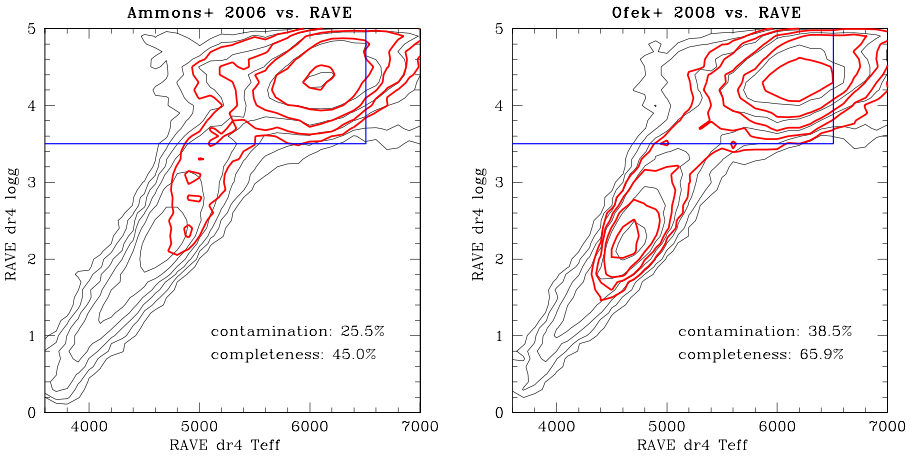

Figure 2. Effective temperature $T_{\rm eff}$ and surface gravity $\log g$ from RAVE DR4 for a subset of stars in common with the AM06 (left-hand panel) and OF08 catalogues (right-hand panel). The distributions of both the overall sample (black contour lines) and for DLF5 subsets (red contour) as classified by photometry alone are shown. The blue lines mark the boundaries $T_{\rm eff} < 6510$ K, $\log g > 3.5$ which define the DLF5 sample; red sources outside this region are therefore classified as contaminants (see Section 3.3).

misses *true negatives*, and (2) include targets (mostly hot dwarfs, evolved giants or non-stellar objects) which are contaminants, otherwise called *false positives*. From here on, we define

(i) *completeness* as the fraction of true positives over the total number of true targets, TP/(TP+TN);
(ii) *contamination* as the fraction of false positives over the total number of positives, FP/(TP+FP).

Ideally, we are searching for a technique which outputs a complete sample with zero contamination.

The DSLF5/DLF5 samples are usually defined in terms of spectral type and luminosity class (Section 1). Translating this definition into a new one based on a range of $T_{\rm eff}$ and $\log g$ would enable an easier, direct and more accurate comparison among different classification schemes. There is no general agreement in the literature on how to link stellar parameters to a given SpT, and second-order effects, such as metallicity and ages, further complicate the problem. We define

(i) the main DSLF5 sample as MS or post-MS stars having $T_{\rm eff} < 6510$ K and $\log g > 3.0$;
(ii) the DLF5 subset as MS stars having $T_{\rm eff} < 6510$ K and $\log g > 3.5$.

The thresholds on $T_{\rm eff}$ and $\log g$ were set according to the empirical calibrations of the F5V SpT published in Cox (2000) and on literature data compiled by E. Mamajek[5] and cross-matched with the PASTEL catalogue of stellar spectroscopic parameters (Soubiran et al. 2010).

In order to assess the completeness and contamination of a sample selected by the existing classification algorithms, we matched catalogues by Ammons et al. (2006) and Ofek (2008, hereafter AM06 and OF08; typical examples of RPM and template-matching-based techniques, respectively) with a reliable catalogue of atmospheric parameters. To that purpose, we chose the fourth and most recent data release (DR4) of the spectroscopic survey RAVE (Steinmetz et al. 2006; Kordopatis et al. 2013). RAVE, Radial Velocity Experiment[6] (Steinmetz et al. 2006), is a spectroscopy survey at $R \simeq 7000$ resolution, based at the Australian Astronomical Observatory, aiming at delivering accurate radial velocities ($\sim$1.5 km s^{-1}), atmospheric parameters and elemental abundances for about half a million stars in the Southern hemisphere. The RAVE sample is magnitude limited at $I < 12$, although not complete.

The DLF5 and DSLF5 subsets were extracted from the OF08 stars according to their 'photometric' $T_{\rm eff}$ and $\log g$, following the definitions given above. As for the AM06 catalogue, the authors provide a pre-selected subset of FGK MS dwarfs, which we further trimmed at $T_{\rm eff} < 6510$ K to obtain a pure DLF5 sample; no DSLF5 sample can be straightforwardly extracted from AM06, because no direct information on $\log g$ is available from this work.

Now, let us focus on the DLF5 subsets of AM06 and OF08. Their distribution (red contours) is compared with that of the general field (black contours) and plotted in Fig. 2 as a function of their spectroscopic parameters $T_{\rm eff}$, $\log g$ as tabulated in RAVE DR4, for both AM06 (left-hand panel) and OF08 (right-hand panel) input catalogues. Following our previous definition, contamination and completeness were estimated with respect to the fraction of stars falling outside the boundaries $\log g > 3.5$ and $T_{\rm eff} < 6510$ K as given by RAVE. The resulting contamination spans $\sim$25–38 per cent for AM06 and OF08, respectively, while completeness spans $\sim$45–66 per cent (Table 1). Contaminants are mostly red giants for OF08, and are equally shared between red giants and earlier MS types for AM06. We also did a similar estimation on the Belikov & Röser (2008) catalogue, which resulted in a $\sim$38.2 per cent contamination, very close to the OF08 result (Table 1). Considering that the vast majority of a magnitude-limited sample is composed of early-type and evolved giants, photometric classification techniques such as these represent an acceptable starting basis for the target selection process.

Unfortunately, all the reviewed catalogues including OF08 and AM06 are magnitude limited ($V \lesssim 11$) because they are based on Tycho-2. If we want to extend our classification to fainter stars ($V \lesssim 13$), while taking advantage of RAVE DR4 as a reliable external calibrator with the aim of improving both completeness and contamination, we must use different input catalogues and devise

[5] https://www.pas.rochester.edu/~emamajek/IV_standards_PASTEL_logg.txt

[6] https://www.rave-survey.org/

Table 1. Sky-averaged contamination and completeness of DLF5 and DSLF5 samples extracted from various stellar classification algorithms, by assuming the RAVE DR4 spectroscopic data base as a reliable source of stellar parameters.

Catalogue	Sample	Subset	Contamination	Completeness	$m_{\rm lim}$
AM06	DLF5[a]	all-sky	25.5 per cent	45.0 per cent	$V < 11$[b]
OF08[c]	DSLF5	all-sky	36.2 per cent	71.8 per cent	$V < 11$
OF08[c]	DLF5	all-sky	38.5 per cent	65.9 per cent	$V < 11$
Belikov & Röser (2008)	DLF5	all-sky	38.2 per cent	–.–[d] per cent	$V < 11$
UCAC4-RPM (this work)	DSLF5	all-sky	28.8 per cent	79.9 per cent	$f < 13.5$
UCAC4-RPM (this work)	DLF5	all-sky	27.7 per cent	80.0 per cent	$f < 13.5$
UCAC4-RPM (this work)	DSLF5	$\|b\| > 20^\circ$	28.0 per cent	71.2 per cent	$f < 13.5$
UCAC4-RPM (this work)	DLF5	$\|b\| > 20^\circ$	27.5 per cent	79.4 per cent	$f < 13.5$
UCAC4-RPM (this work)	DSLF5	$10^\circ < \|b\| < 20^\circ$	28.0 per cent	83.7 per cent	$f < 13.5$
UCAC4-RPM (this work)	DLF5	$10^\circ < \|b\| < 20^\circ$	27.9 per cent	81.8 per cent	$f < 13.5$
UCAC4-RPM (this work)	DSLF5	$\|b\| < 10^\circ$	28.3 per cent	82.0 per cent	$f < 13.5$
UCAC4-RPM (this work)	DLF5	$\|b\| < 10^\circ$	28.0 per cent	80.6 per cent	$f < 13.5$

Notes. Contamination and completeness fractions are defined in Section 3.3.
[a]The AM06 catalogue does not provide $\log g$; instead, a pre-compiled set of FGK dwarfs is available, from which we trimmed a DLF5 sample by imposing $T_{\rm eff} < 6510$ K; no extension to DSLF5 is possible.
[b]The completeness limit of Tycho-2 is here assumed to be $V \lesssim 11$, though slightly depending on SpT and Galactic latitude.
[c]The OF08 DLF5 and DSLF5 sample are selected according to the best-χ^2 Pickles (1998) template.
[d]The only part of the Belikov & Röser (2008) catalogue which is available to us is just a subsample of FGK dwarfs; without knowing the exact composition of the initial sample, it is possible to estimate the contamination but not the completeness.

a new classification algorithm. This is the main driver for the creation of our brand new all-sky catalogue of DLF5 and DSLF5 stars, hereafter named UCAC4-RPM.

4 THE UCAC4-RPM CATALOGUE COMPILATION

4.1 Input catalogues

We chose to adopt UCAC4 (Zacharias et al. 2013) as the starting point to build our new catalogue. UCAC4 is a compiled, all-sky astrometric catalogue designed to provide high-quality CCD positions and proper motions for targets fainter than the limiting magnitude of *Hipparcos* and Tycho-2. The observations are designed to cover the $R = 7.5$–16.3 mag range, and were performed through one passband at approximately 579–643 nm. Most observations were carried out in non-photometric conditions, but UCAC4 native magnitudes $f_{\rm mag}$ have been calibrated against Tycho-2 stars and systematic errors are constrained within 0.1 mag, and sometimes much better. Positional errors are between 15 and 20 mas for stars with $10 \lesssim V \lesssim 14$. UCAC4 is supplemented by proper motions and SuperCosmos (Hambly et al. 2001) and 2MASS NIR photometric data, as well as diagnostic flags. The proper motions of bright stars are based on about 140 catalogues, including *Hipparcos* and Tycho-2. Proper motions of faint stars are based on a re-reduction of early epoch SPM (van Altena 2011) data (at $-90^\circ < \delta < -10^\circ$) plus Schmidt plate data from the SuperCosmos project.

UCAC4 native magnitudes are complemented by the sixth data release (DR6) of the AAVSO Photometric All-Sky Survey[7] (APASS; Henden & Munari 2014) in five filters: Johnson B and V, plus Sloan g', r', i'. Once completed, APASS will cover the magnitude range between $V \simeq 10$ and $V \simeq 17$. It will conveniently link Tycho-2 to SDSS, plus cover the whole sky at the same depth of UCAC4. Johnson B and V were chosen to extend the Tycho-2 calibration to fainter magnitudes, while Sloan g', r', i' will homogeneously extend the much deeper SDSS, SkyMapper, Pan-STARRS surveys on the brighter end.

Three APASS data releases were published after the official release of UCAC4: DR7, DR8 and DR9 (2015 July 29), which in principle should have improved both the completeness and the photometric accuracy of the APASS survey. We merged UCAC4 and APASS DR8 to test this assumption and actually found the opposite: by matching the resulting catalogue with a set of photometric standards, it is easy to see that the photometric homogeneity is much worse with respect to APASS DR6. This is easily explained by some problems, which occurred during the DR7–DR9 observing campaigns and described in the release notes, which include the 'blue' camera malfunctions in the Northern hemisphere and the 'red' camera's poor focus in the Southern hemisphere. Many of these problems are expected to be solved in the forthcoming DR10; meanwhile, we adopt the DR6 magnitudes throughout the present paper.

As a further step, we truncated UCAC4 at $V < 10$, where APASS photometry saturates, and replaced those entries with the corresponding photometry and proper motions from Tycho-2, converting Tycho (B_T, V_T) to Johnson (B, V) through the standard transformations given by Høg et al. (2000). During this process, we identified 176 662 UCAC4 entries at $V > 10$ for which the tabulated B, V magnitudes were mistakenly copied from Tycho-2 (B_T, V_T) without any transformation; we fixed these entries by forcing the correct transformations as above. Finally, APASS magnitudes are not complete down to $f_{\rm mag} = 13.5$ on some particular regions of the northern sky, so for those missing entries, we calibrated UCAC4 $f_{\rm mag}$ and 2MASS J and K against the most recent set of secondary *UBVRI* photometric standards published by P. B. Stetson[8] obtaining

$$\begin{cases} B = f_{\rm mag} + 1.8372\,(J-K) + 0.0627 & (\sigma = 0.15) \\ V = f_{\rm mag} + 0.5659\,(J-K) - 0.1204 & (\sigma = 0.09), \end{cases} \quad (2)$$

where the rms scatter around the best fit is calculated within the range of interest $10 < V < 13$.

[7] https://www.aavso.org/apass

[8] http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/community/STETSON/standards/, Stetson et al. (2000)

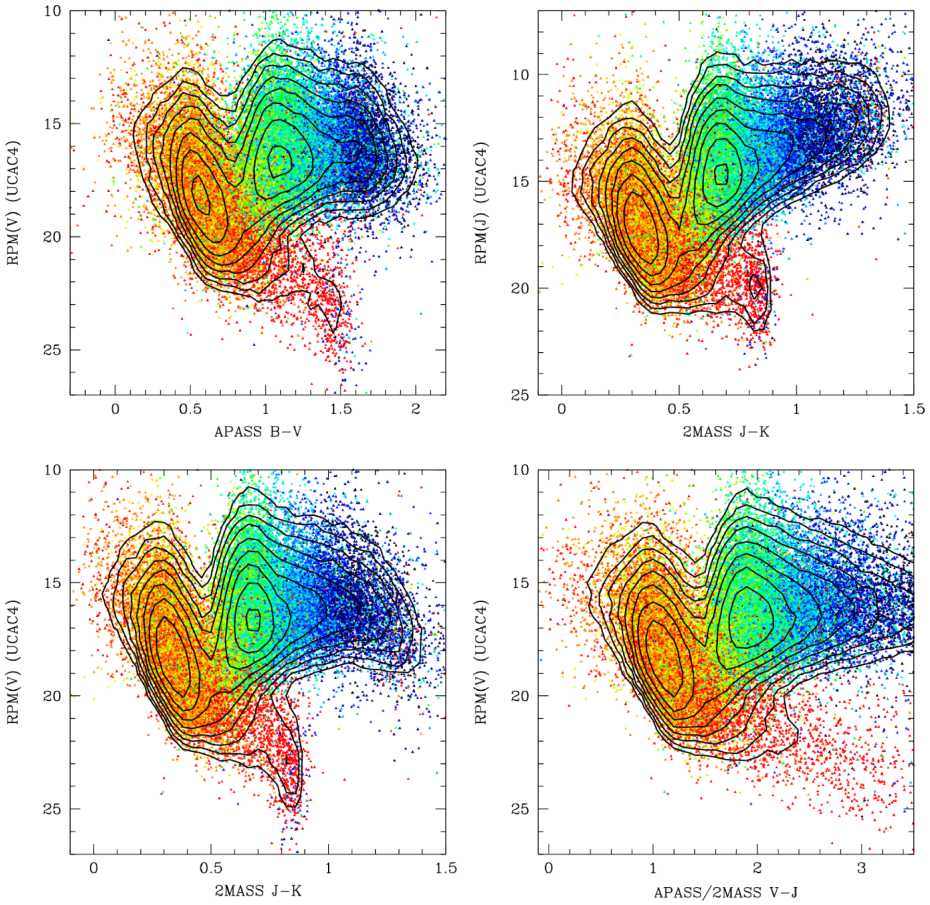

Figure 3. RPMDs for all the UCAC4-RPM sources in common with RAVE DR4, plotted by adopting different permutations of magnitudes: ($B - V$, RPM(V)), ($J - K$, RPM(J)), ($J - K$, RPM(V)), ($V - J$, RPM(V)) in reading order. The number of points is randomly decreased by a factor of 10 to improve clarity, while the isodensity contours (black lines) are evaluated on the full sample. The log g values from RAVE are colour-coded within the range from 0.0 (deep blue) to 5.0 dex (pure red).

The resulting catalogue, trimmed at $f_{\rm mag} < 13.5$ to be more easily manageable, will be called UCAC4-RPM hereafter. Even taking into account colour effects in $V - f_{\rm mag}$, this sample is magnitude-complete at $V < 13$ for FGK spectral types, and constitutes a perfect basis to cherry-pick targets for space-based exoplanet transit search missions such as *TESS* and *PLATO*. Coincidentally, $V = 13$ is also the typical limiting magnitude of the most fruitful wide-field, ground-based transit searches, making our effort of more general use. The overall number of UCAC4-RPM entries is 10 198 407, of which 9928 389 have at least proper motions and B, V photometry either from Tycho2 or from APASS.

RAVE DR4 lists 425 561 stars, of which 412 741 are in common with UCAC4-RPM. 375 203 of them have valid proper motions, B, V magnitudes and atmospheric parameters. The latter subset consists of stars which share a reliable estimate of log g and $T_{\rm eff}$ from RAVE ($\Delta \log g \simeq 0.3$ dex, $\Delta T_{\rm eff} \simeq 150$ K on average), span the full range of Galactic latitude and perfectly overlap the typical magnitude range of our stellar sample. In other words, this is an ideal training set to calibrate RPM-based selections on the whole UCAC4-RPM.

4.2 Defining the RPM selection

Before selecting stars according to their RPMs, one has to choose which photometric bands to employ for the colour (horizontal axis of the RPMD) and the RPM itself (vertical axis). While the proper motion of a given UCAC4-RPM entry is unique, two or three magnitudes have to be chosen from B, V, J, K, the latter two originating from 2MASS and included in UCAC4-RPM through UCAC4.

We ran several tests to see the impact of this choice on the ability of the RPMD to properly separate the MS from the evolved stars: the resulting RPMD for four different test choices, namely ($B - V$, RPM(V)), ($J - K$, RPM(J)), ($J - K$, RPM(V)), ($V - J$, RPM(V)) are plotted in Fig. 3, where the log g from RAVE is colour-coded from 0.0 (blue) to 5.0 dex (red). At first sight, when integrating the results over the whole sky, all four choices are effective in separating the two 'bumps' corresponding to MS stars (centred approximately on unevolved F dwarfs) and red giants. However, the ($B - V$, RPM(V)) diagram performs better at low Galactic latitudes, where interstellar extinction plays a crucial role in the most distant stars. This is easily explained by the fact that $B - V$ is the colour most affected by extinction. While nearby, late-type MS dwarfs share a negligible or very small extinction, distant stars (which are mostly contaminating giants) are moved farther from the MS by the reddening vector, making the dwarf/giant separation easier. For this reason, we adopted the ($B - V$, RPM(V)) plane to perform the following RPM calibrations and selections.

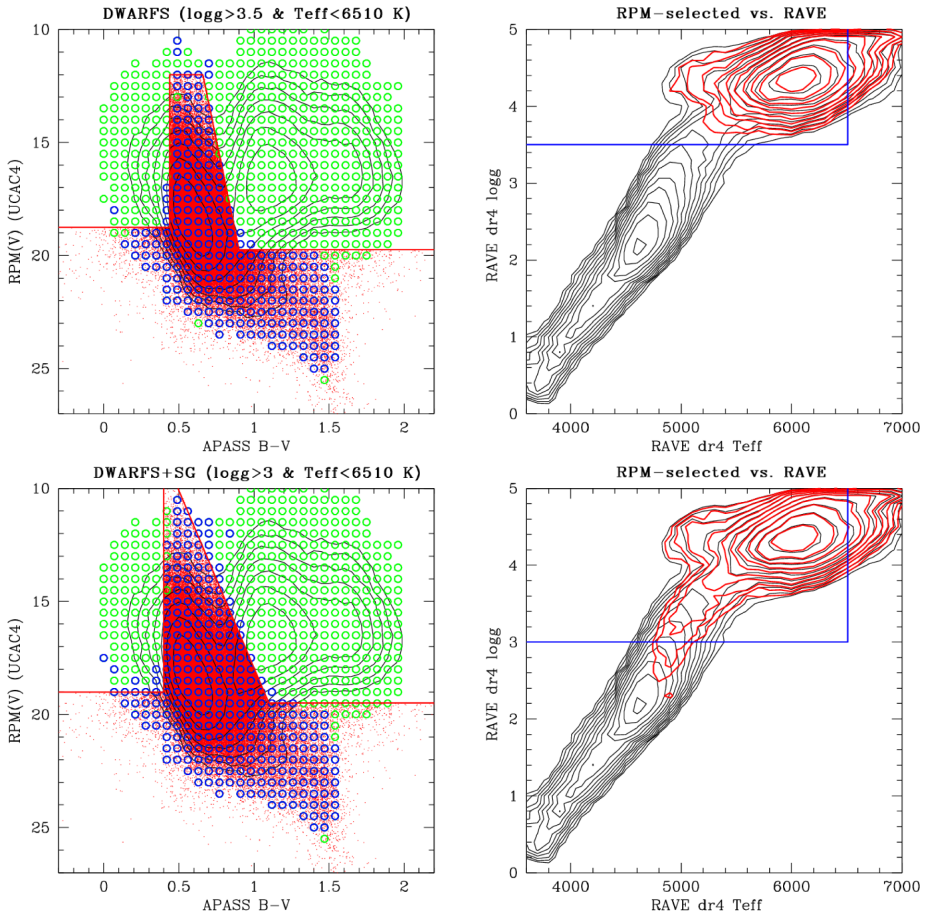

Figure 4. Procedure to extract DLF5 (upper panels) and DSLF5 (lower panels) stars from UCAC4-RPM according to their colour and RPM. Left-hand panels: RPMDs for a subset of stars in common between RAVE DR4 and UCAC4-RPM. Both RPMDs are subdivided into cells (open circles). The stars in each cell are classified as genuine DLF5/DSLF5 targets or as contaminants according to their RAVE $T_{\rm eff}$, $\log g$; then the cell is drawn in blue if at least 50 per cent of them are bona fide DLF5/DSLF5 sources, and green otherwise. Our RPM selection encompassing 'good' cells is approximated by a red polynomial chain, corresponding to the equations (3) and (4) numerical cuts. Right-hand panels: effective temperature $T_{\rm eff}$ and surface gravity $\log g$ from RAVE DR4 for all stars in common with UCAC4-RPM (black contours), and after the RPM selection devised for the DLF5/DSLF5 sample has been applied (red contours). The blue lines mark the boundaries $T_{\rm eff} < 6510$ K, $\log g > 3.5$ (DLF5) or $T_{\rm eff} < 6510$ K, $\log g > 3.0$ (DSLF5); red points outside these regions are therefore classified as contaminants (see the text; compare with Fig. 2 and Table 1).

While one could in principle draw an empirical boundary line between giants and dwarfs on the RPMD by hand (as done, among others, by Gould & Morgan 2003), a more rigorous and quantitative approach is advisable when working with precise requirements on SpT (or, equivalently, on $T_{\rm eff}$ and $\log g$) as in our case. We first subdivided the RPMD plotted with all the stars in common between UCAC4-RPM and RAVE into small rectangular cells having side lengths equal to $\Delta(B-V) = 0.05$ mag and $\Delta{\rm RPM}(V) = 0.5$ mag (open circles in Fig. 4, left-hand panels). Each star in a given cell is classified as a genuine DLF5/DSLF5 target, or as a contaminant, according to its RAVE $T_{\rm eff}$, $\log g$ based on the $T_{\rm eff} < 6510$ K, $\log g > 4.0$ (DLF5) or $T_{\rm eff} < 6510$ K, $\log g > 3.5$ (DSLF5) criterion. The fraction Φ of genuine targets over contaminants is then evaluated for each cell. After some trial-and-error attempts, we define those cells having $\Phi > 50$ per cent as 'good', though this threshold can be changed depending on whether one desires to increase the completeness (lower Φ threshold) or to decrease the contamination (higher Φ) of the final sample. Cells in Fig. 4 are drawn as blue circles if $\Phi > 50$ per cent, and green otherwise. As a final step, with the aim of speeding up the computation, we approximated our RPM selection encompassing 'good' cells through a polynomial chain (red line in Fig. 4), corresponding to the following cuts:

$$\begin{aligned}
&{\rm RPM}(V) > 12 \ \texttt{and} \ (\\
&[(B-V) > 0.7 \ \texttt{and} \ {\rm RPM}(V) > 19.75] \ \texttt{or}\\
&[(B-V) < 0.7 \ \texttt{and} \ {\rm RPM}(V) > 18.75] \ \texttt{or}\\
&[{\rm RPM}(V) > (-10 + 33(B-V)) \ \texttt{and} \ (B-V) > 0.438])
\end{aligned} \tag{3}$$

for the DLF5 selection and

$$\begin{aligned}
&{\rm RPM}(V) > 10 \ \texttt{and} \ (\\
&[(B-V) > 0.7 \ \texttt{and} \ {\rm RPM}(V) > 19.5] \ \texttt{or}\\
&[(B-V) < 0.7 \ \texttt{and} \ {\rm RPM}(V) > 19.0] \ \texttt{or}\\
&[{\rm RPM}(V) > (2 + 16(B-V)) \ \texttt{and} \ (B-V) > 0.40])
\end{aligned} \tag{4}$$

for the DSLF5 selection; the '`and`' and '`or`' logical operators have their usual meaning.

Equation (3) and (4) can be applied to the full UCAC4-RPM sample to get whole-sky, magnitude-limited samples of DLF5 and DSLF5 stars. For instance, the DLF5 subset contains 84 432, 287 914, 899 761, 2627 966 entries at $V < 10$, 11, 12, 13,

Table 2. DLF5/DSLF5 star counts from UCAC4-RPM as a function of V magnitude (left table) and $B - V$ colour (right table), along with the corresponding areal density $N(\square)$, spectral types SpT, absolute magnitude M_V, effective temperature $T_{\rm eff}$ and limiting distance $d_{\rm lim}$ at the limiting magnitude $V = 13$.

V	DLF5	$N(\square)$ (deg^{-2})	DSLF5	$N(\square)$ (deg^{-2})
<8.0	4422	0.10	5818	0.14
8.0–8.5	4995	0.12	6413	0.16
8.5–9.0	11 511	0.27	14 059	0.34
9.0–9.5	21 248	0.51	25 789	0.62
9.5–10.0	42 244	1.02	50 703	1.22
10.0–10.5	71 459	1.73	86 216	2.08
10.5–11.0	131 750	3.19	155 236	3.76
11.0–11.5	227 458	5.51	266 026	6.44
11.5–12.0	383 479	9.29	445 153	10.7
12.0–12.5	647 860	15.7	746 631	18.0
12.5–13.0	1077 810	26.1	1235 974	29.9

$B - V$	DLF5	$N(\square)$ (deg^{-2})	SpT	M_V	$T_{\rm eff}$ (K)	$d_{\rm lim}$ (pc)
<0.45	53 749	1.57	F5V	3.40	6510	831
0.45–0.50	178 269	5.22	F6V	3.70	6340	724
0.50–0.55	243 253	7.13	F8V	4.01	6170	628
0.55–0.60	276 979	8.12	G0V	4.45	5920	513
0.60–0.65	266 499	7.81	G2V	4.79	5770	438
0.65–0.70	235 501	6.90	G4V	4.94	5680	409
0.70–0.75	179 495	5.26	G8V	5.32	5490	343
0.75–0.80	129 239	3.79	G9V	5.55	5340	309
0.80–0.85	87 223	2.55	K0V	5.76	5280	280
0.85–0.90	43 033	1.26	K2V	6.19	5040	230
>0.90	91 697	2.68	K4V	7.04	4620	156

Notes. The left table lists counts from the whole UCAC4-RPM catalogue. The right table is extracted from a subset trimmed at $|b| > 10^\circ$ and $V < 13$ as described in Section 4.4, to be strictly magnitude limited, and to avoid the regions close to the Galactic plane where the contamination rate on the latest spectral types is higher. The quantities SpT, M_V, $T_{\rm eff}$ are the average of each bin calculated through the SpT/$T_{\rm eff}$ calibration by E. Mamajeck described in Section 2.

respectively. A more detailed summary of the UCAC4-RPM star counts as a function of V magnitude and $B - V$ colour is tabulated in Table 2. The full UCAC4-RPM catalogue, augmented with a flag about the DLF5/DSLF5 RPM-based classification, is made publicly available through VizieR and a dedicated web server.[9]

4.3 Estimating completeness and contamination

To estimate the completeness and contamination of our UCAC4-RPM classifications, following the same approach applied in Section 3.3, we can put our subsamples into the RAVE DR4 $\log g$ versus $T_{\rm eff}$ diagram to check how many points lie in the allowed DLF5/DSLF5 region a posteriori, and how many of the RAVE DR4 dwarfs are effectively selected as such by our algorithm. The results are tabulated in Table 1. Our approach results in a comparable amount of contamination (≈28 per cent) and much better completeness (≈80 per cent) with respect to the previous classification attempts (AM06; OF08), i.e. we end up with a much cleaner sample of dwarfs and subgiants at a fainter limiting magnitude. This result is probably close to the intrinsic limit of the RPM technique, which is a statistical method.

In principle, every photometric classification technique is expected to worsen its performances when moving to the densest regions, i.e. close to the Galactic plane. The first reason is because crowding impacts the accuracy of the input catalogues. This could be an issue for APASS (our primary source of optical magnitudes at $V > 10$), because the pixel scale of that survey is 2.57 arcsec pixel^{-1}, and the point spread functions were deliberately defocused to a full width at half-maximum of 1.5–2.0 pixels. On the other hand, our targets are relatively bright ($V < 13$) and only marginally affected by confusion down to a few degrees from the Galactic plane. A second reason is that lines of sight at low Galactic latitudes can reach very high values of interstellar extinction as the distance increases; therefore, early-type stars can be misidentified as later-type dwarfs. This effect is larger at the faint end of the sample, which probes a larger volume of space. But this is counterbalanced by another effect which actually improves the giant/dwarf separation on very reddened lines of sight: evolved stars are moved farther from the MS on the RPMD, as explained in Section 4.2.

To assess how much the proximity to the Galactic plane could be a limiting factor to our classification, we redid the completeness/contamination estimate as above by selecting UCAC4-RPM stars within three different ranges of Galactic latitudes: $|b| < 10$, $10 < |b| < 20$ and $|b| > 20$ (Table 1). The results are pretty clear: on average, the efficiency of our algorithm is not significantly impacted by b. The contamination level is remarkably constant at ≈28 per cent, while completeness actually improves a bit towards the Galactic plane, from ≈71 per cent to ≈81 per cent. By looking at an all-sky chart of DSLF5 stars (Fig. 5), it is easy to notice that their surface density increases by a factor of 2–3 from the Galactic poles to the disc, as predicted by stellar models such as TRILEGAL or Besançon Galactic models (BGM; Robin et al. 2003; Girardi et al. 2005) for DSLF5 stars. The only exception is the presence of a few overdensities at very low b (blue regions in Fig. 5), where the fraction of false positives is much larger than 30 per cent. These regions, however, cover only a negligible fraction of the sky and therefore do not significantly contribute to the overall statistics.

4.4 Comparison with galactic models

As a further validation of our catalogue, we made a more detailed comparison between UCAC4-RPM and GUMS (*Gaia* Universe Model Snapshot; Robin et al. 2012), an all-sky synthetic catalogue based on an updated version of the BGM (Robin et al. 2003) and aimed to a realistic simulation of the *Gaia* performances. Such an analysis should be regarded not only as a validation check for UCAC4-RPM, but also as a test for the BGM and its underlying assumptions, the most relevant of which for our purposes is the luminosity function of the solar neighbourhood (Reid, Gizis & Hawley 2002). UCAC4-RPM is a magnitude-limited sample, as opposed to a volume-limited one, so the local luminosity functions must be first convolved by a Galactic model to get a meaningful comparison. It is worth noting that, at the faint limit considered here, UCAC4-RPM is probing heliocentric distances up to ∼830 pc (assuming M_V(F5V) = 3.4 at $V = 13$), with a median distance of ∼300 pc for a typical

[9] http://groups.dfa.unipd.it/ESPG

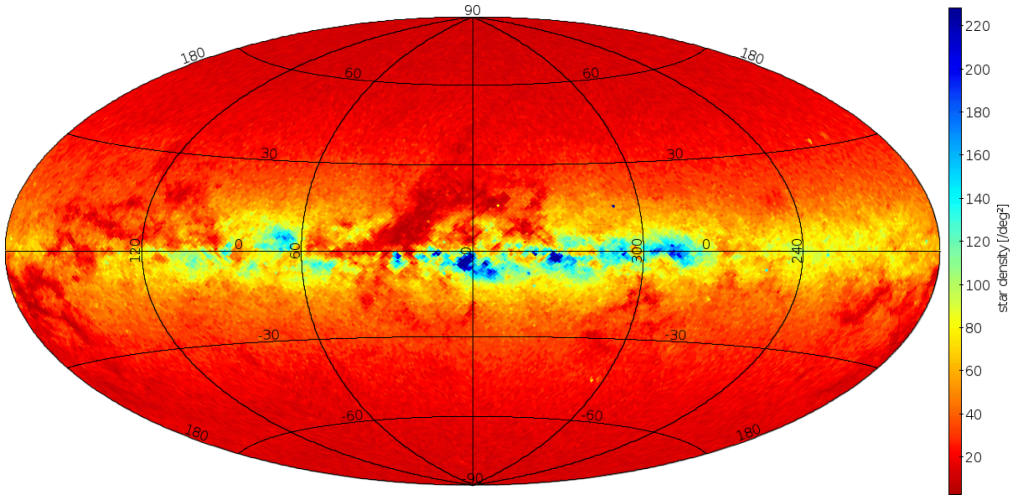

Figure 5. All-sky chart of our UCAC4-RPM catalogue, DSLF5 subset. Areal density is linearly colour-coded from 30 deg^{-2} (pure red) to 80 deg^{-2} (pure yellow) and 220 deg^{-2} (deep blue). Star density varies within a factor of 2–3 as a function of Galactic latitude, as predicted by Galactic models, except for a few regions at $|b| \lesssim 5$ where extreme amounts of crowding and/or interstellar extinction boost the fraction of false positives (see Section 4.3). The underdense regions at $l = 0°$–$30°$, $b = 0°$–$20°$ and $l \sim 175°$, $b \sim -15°$ are the Serpens/Aquila/Ophiuchus rift and the Taurus complex, respectively (Cambrésy 1999).

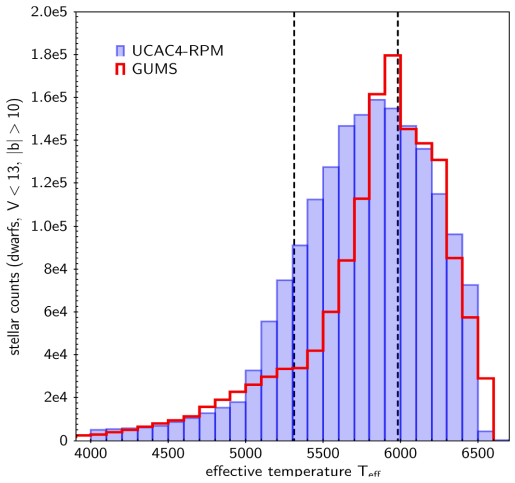

Figure 6. Histogram comparing the temperature distribution of a magnitude-limited sample of dwarf stars from UCAC4-RPM (this work; blue bars) and from the GUMS simulated catalogue (Robin et al. 2012; red bars) based on the BGM (Robin et al. 2003). Both samples are cut at $V < 13$ and $|b| > 10°$. The two dashed vertical lines mark the boundary between G/K and F/G MS spectral types, respectively.

Sun-like star (M_V(G2V) = 4.8 at $V = 12.1$). As a reference, the Reid et al. (2002) luminosity function is based on a *Hipparcos* sample volume limited at 25 pc.

Both UCAC4-RPM and GUMS were first cut at $V < 13$ (where UCAC4-RPM is complete), then only the DLF5 stars were selected by imposing $\log g > 3.5$ and $T_{\rm eff} < 6510$ K following our definition in Section 3.3. In order to perform a fair comparison, a further cut was set at $|b| > 10°$ to exclude regions close to the Galactic plane, where extinction maps (including that by Drimmel, Cabrera-Lavers & López-Corredoira 2003 implemented by GUMS) are known to be unreliable in reproducing the observed stellar counts (Marshall et al. 2006; Schultheis et al. 2014). The resulting histogram as a function of $T_{\rm eff}$ is shown in Fig. 6.

The total number of UCAC4-RPM DLF5 dwarfs at $V < 13$ and $|b| < 10°$ is 1770 693, that is 21 per cent more than the 1460 174 predicted by GUMS. This is a reasonable agreement after considering that the overall contamination rate of UCAC4-RPM is estimated to be at about 28 per cent (Table 1). A closer look at Fig. 6 reveals that while the counts for the F0-F5 subsample ($T_{\rm eff} > 5908$ K) are essentially identical in the two catalogues (598 194 versus 617 643, −3.1 per cent), the stars in excess in UCAC4-RPM come from the G ($5310 < T_{\rm eff} < 5908$ K, 879 558 versus 638 652, +37 per cent) and K subsets ($T_{\rm eff} < 5310$ K, 292 941 versus 199 315, +47 per cent). This is consistent with the hypothesis that the excess counts come from contaminating subgiants, since they are predicted to belong mostly to the G and K spectral types (see Fig. 4, right-hand panels). Summarizing, the UCAC4-RPM results seem to be in a generally good agreement with the current models of the solar neighbourhood and their assumptions.

5 POSSIBLE APPLICATIONS OF UCAC4-RPM FOR EXOPLANET SEARCHES

UCAC4-RPM is nicely suited to build a preliminary input catalogue for future space-based missions, like *TESS* and *PLATO*, as well as ongoing and forthcoming ground-based surveys.

5.1 *TESS*

With its four wide-field optical cameras pointed towards different lines of sight, *TESS* will simultaneously monitor a 24° × 94° strip (i.e. 2300 deg^2) for 27 d, then it will take a 26° turn around the ecliptic pole and repeat this cycle in order to scan nearly the full sky at ecliptic latitude $|\beta| > 6°$ during its two-year nominal mission (Ricker et al. 2015). The temporal coverage will therefore range from 27 d at low ecliptic latitudes to about 1 yr over 900 deg^2 close to the ecliptic poles.

The prime targets for *TESS* are MS dwarfs from F5 to M5; photon noise and confusion set the limiting magnitude at $I_{\rm c} \simeq 12$, or roughly $V \simeq 13$ at K2V. While *TESS* will transmit the full-frame images back to Earth with a 30 min cadence, a highest priority subset of ∼200 000 pre-selected stars will be monitored at a cadence of 2 min. Being magnitude limited at $V \lesssim 13.5$ and focused on DLF5 stars, UCAC4-RPM is perfectly suited to support the *TESS* target selection. It

provides a sample of ∼2600 000 $V < 13$ dwarfs from which the *TESS* target list can be cherry-picked according to a prioritization scheme based on the expected transit detection efficiency. Such an approach should be regarded as complementary to that developed by the *TESS* Target Selection Working Group, which is in charge of building the *TESS* Target Catalog (TTC; Stassun et al. 2014). The TTC is based on a different combination of input catalogues with respect to UCAC4-RPM, and on different selection algorithms and thresholds.

While the full TTC has not been released to the community yet, a public subset of it covering the *K2* fields is named *K2-TESS* and documented by Stassun et al. (2014). *K2-TESS* allows us to probe the overlap region between our catalogue and the TTC. If, for instance, we cross-match the *K2* 'Campaign 2' field with UCAC4-RPM, we get 5662 DSLF5 (of which 4726 are DLF5). Among these, 565 UCAC4-RPM dwarfs are missed by *K2-TESS*, while there are 2234 *K2-TESS* dwarfs missed by UCAC4-RPM, mostly because of their faintness ($V > 13.5$, $J > 11.5$). This is somewhat expected since the two catalogues are based on different methodologies and input catalogues reflecting different scientific priorities. While *K2-TESS* is more focused on cool SpTs (late K and M dwarfs) and mainly based on the 2MASS infrared photometry, UCAC4-RPM is focused on solar-type FGK targets, implying different selection algorithms and thresholds. In other words, the two approaches will be highly synergic in selecting the optimal targets for *TESS* (Stassun et al. 2014).

5.2 *PLATO*

The unique optical design of *PLATO* (Rauer et al. 2014) is driven by the requirement to cover a significant fraction of the sky with a good image quality in order to achieve an unprecedented photometric precision: <34 ppm h^{-1} (parts per million measured over a 1 h time-scale) for the main scientific sample limited at $V < 11$, and even better for the brightest stars of the sample at $V \simeq 6$ (dominated by systematic errors). The present design of *PLATO* implies that the array of 12 cm telescopes is subdivided into four groups, each group pointing towards a direction which is displaced by an angle $\theta = 9.2$ deg from the centre of the overall FOV. Such strategy provides a non-uniform instantaneous coverage of a 2124 deg^2 field.

The observing strategy of *PLATO* consists of a two-step approach in order to maximize the scientific return of the mission. *PLATO* will image two 'long-duration' (LD) fields for more than two years each, and some supplementary fields for a few months each, during the 'step and stare' (S&S) phase. The LD phase will fulfil the main scientific objective of the mission, being sensitive to planets down to the Earth's size orbiting within the habitable zone of solar-type (FGK) stars. The S&S fields will expand the accessible sky area by about one order of magnitude, giving us access to a larger number of bright nearby stars at the expense of probing a smaller range of orbital periods. A simple calculation shows that, at the end of a nominal 6.5 year mission with 2 LD and 10 S&S fields, *PLATO* will have surveyed about two thirds of the whole sky.

The *PLATO* Science Requirements Document[10] defines five complementary stellar samples to be surveyed, listed in decreasing priority and summarized below.

(i) P1: DSLF5 stars brighter than $V = 11$, monitored with a photometric noise level $\sigma \leq 34$ ppm h^{-1}, to be observed during the LD phase.

[10] ID code: ESA-PLATO-ESTEC-SCI-RS-001.

(ii) P2, P3: DSLF5 stars brighter than $V = 8$, monitored at $\sigma \leq 34$ ppm h^{-1}, to be observed during the LD and S&S phase, respectively.

(iii) P4: cool M dwarfs (M0 or later), to be monitored during the LD phase of the mission (at $V < 16$), or during the S&S phase (at $V < 15$). The photometric noise level for both subsamples must be below 800 ppm h^{-1}.

(iv) P5: DSLF5 stars brighter than $V = 13$, to be observed during the LD phase.

Samples P1, P2, P3, P5 are therefore all made of 'solar-like' stars brighter than $V = 13$. Due to telemetry limitations, *PLATO* will not be able to download the full images, collected with cadence between 60 s (normal telescopes) and 2.5 s (for the two 'fast' telescope, dedicated to bright stars, and the only ones equipped with two filters, for colour measurement). Only for a limited subsample of targets, it will be possible to download imagettes. Most of the photometry must be done on board on the selected targets, and also the imagette centres must be pre-selected. In other words, *PLATO* needs an input catalogue (PIC: *PLATO* Input Catalog). It is clear that our new UCAC4-RPM represents the most appropriate tool to preliminarily select and prioritize the target sample. It also is of basic importance for a selection of *PLATO* fields, needed, for engineering reasons, well before *Gaia* catalogues can be used.

5.3 Ground-based surveys

UCAC4-RPM is made publicly available to the astronomical community. Its magnitude range $6 \lesssim V \lesssim 13$ makes it exploitable for the ongoing ground-based transit surveys such as SuperWASP/WASP South (Pollacco et al. 2006; Hellier et al. 2011), HatNet/Hat-South (Bakos et al. 2009a,b), KELT/KELT-South (Pepper et al. 2007, 2012) or NGTS (Chazelas et al. 2012) to vet and prioritize the list of planetary candidates and to discard transit-like signals originating from early-type or giant stars, when no other source of reliable stellar parameters is available. A careful preliminary characterization of the target stars has been proven to greatly speed up the follow-up process, avoiding a loss of observing time and resources by monitoring false positives (Almenara et al. 2009; Bryson et al. 2013).

6 CONCLUSIONS

Throughout the previous sections, we described how we devised a new RPM-based algorithm to assign a luminosity class to field stars by knowing only their proper motions and two optical magnitudes. By applying this optimal algorithm on a new stellar catalogue compiled by matching UCAC4, APASS DR6 and Tycho-2, we ended up with UCAC4-RPM – an all-sky sample of solar-type dwarf stars complete down to at least $V \simeq 13$. We demonstrated that the latter catalogue, once complemented by subgiants within the same spectral type range, meets the requirements set by the *PLATO* team for the target selection of its main stellar samples. In particular, the relatively low level of contamination ($\lesssim$30 per cent) of UCAC4-RPM, together with a $\gtrsim$80 per cent completeness, is well suited to *PLATO* (but also *TESS*), whose telemetry allows us to select many more targets with respect to the nominal requirement of P1 stars, therefore compensating for the fraction lost due to contaminants. UCAC4-RPM proved to be helpful as a starting point to select the (provisional) coordinates of the LD pointing fields, which are needed at this stage to tune the observational strategy, to run engineering tests and to plan an optimal follow-up strategy for the

object of interest to be delivered by *PLATO*. Also ongoing and forthcoming ground-based survey for exoplanet search may benefit by UCAC4-RPM catalogue.

It is possible to take future steps to improve UCAC4-RPM. The most obvious one is the inclusion of newer releases of APASS, to rely on a more accurate, complete and homogeneous source of B and V magnitudes. APASS DR8 and DR9 are already available, but both suffer from photometric inhomogeneity due to the inclusion of more recent data and a filter change; it is expected that APASS DR10, still to be released, will solve most of these problems. Additionally, the training set used in the present work (RAVE) could be improved by including other wide-field spectroscopic surveys, such as SDSS/SEGUE (Yanny et al. 2009) and especially LAMOST/LEGUE (Deng et al. 2012), which will enable us to calibrate our RPM selections on both hemispheres (RAVE is limited to $b \leq 0$) and to fainter magnitudes, i.e. including more dwarfs of K and M spectral types. We performed a preliminary cross-check by mapping the surface gravity listed on the first public release of LAMOST (DR1) on the full UCAC4-RPM sample. As expected, the LAMOST parameters confirm the accuracy of our previous dwarf versus giant separation on the RPMD based on RAVE (Fig. 4). Once the first LAMOST complete release will be made available to the community, we expect to increase the size of our training set by a factor of 10.

The *Gaia* final catalogue, to be released no earlier than 2024, is expected to make UCAC4-RPM obsolete on most of the sky, thanks to its accurate spectrophotometric measurements (BP/RP instrument) and exquisitely precise geometrical parallaxes (ASTRO instrument). However, the algorithms on which UCAC4-RPM is based could help in exploiting the *Gaia* data to estimate luminosity classes at $V > 11$ much earlier than 2024, starting from the second intermediate release (DR2; end of 2017), when just BP/RP integrated magnitudes and proper motions will be available, but not distances or surface gravities. A modified version of the algorithm presented in Section 4 can be easily adapted and calibrated through LAMOST+RAVE to work in the ($G_{\rm BP} - G_{\rm RP}$, RPM(G)) plane.

ACKNOWLEDGEMENTS

VN and GP acknowledge partial support by the Università di Padova through the 'progetto di Ateneo #CPDA103591', and by the Agenzia Spaziale Italiana (ASI) through the contract *PLATO*. VN acknowledges partial support from INAF-OAPd through the grant 'Analysis of HARPS-N data in the framework of GAPS project' (#19/2013) and 'Studio preparatorio per le osservazioni della missione ESA/CHEOPS' (#42/2013). The present work has been carried out following the ASI-INAF agreement number 2015-019-R0, 2015 July 29. VN dedicates this paper to the memory of Anna Scagliarini (1946–2016).

REFERENCES

Almenara J. M. et al., 2009, A&A, 506, 337
Ammons S. M., Robinson S. E., Strader J., Laughlin G., Fischer D., Wolf A., 2006, ApJ, 638, 1004 (AM06)
Árnadóttir A. S., Feltzing S., Lundström I., 2010, A&A, 521, A40
Bailer-Jones C. A. L., 2010, MNRAS, 403, 96
Bailer-Jones C. A. L., 2011, MNRAS, 411, 435
Bailer-Jones C. A. L. et al., 2013, A&A, 559, A74
Bakos G. Á. et al., 2009a, in Pont F., Sasselov D., Holman M. J., eds, Proc. IAU Symp. 253, Transiting Planets. Cambridge Univ. Press, Cambridge, p. 21
Bakos G. et al., 2009b, in Pont F., Sasselov D., Holman M. J., eds, Proc. IAU Symp. 253, Transiting Planets. Cambridge Univ. Press, Cambridge, p. 354
Batalha N. M. et al., 2010, ApJ, 713, L109
Belikov A. N., Röser S., 2008, A&A, 489, 1107
Bessell M. S., 2005, ARA&A, 43, 293
Bilir S., Karaali S., Güver T., Karataş Y., Ak S. G., 2006, Astron. Nachr., 327, 72
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Brown T. M., Latham D. W., Everett M. E., Esquerdo G. A., 2011, AJ, 142, 112
Bryson S. T. et al., 2013, PASP, 125, 889
Cambrésy L., 1999, A&A, 345, 965
Chazelas B. et al., 2012, Proc. SPIE, 8444, 84440E
Cox A. N., 2000, Allen's Astrophysical Quantities, 4th edn. AIP Press, New York
Deleuil M. et al., 2006, in Fridlund M., Baglin A., Lochard J., Conroy L., eds, ESA SP-1306: The CoRoT Mission Pre-Launch Status – Stellar Seismology and Planet Finding. ESA, Noordwijk, p. 341
Deleuil M. et al., 2009, AJ, 138, 649
Deng L.-C. et al., 2012, Res. Astron. Astrophys., 12, 735
Drimmel R., Cabrera-Lavers A., López-Corredoira M., 2003, A&A, 409, 205
Fischer D. A. et al., 2005, ApJ, 620, 481
Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, A&A, 436, 895
Gontcharov G. A., 2009, Astron. Lett., 35, 638
Gould A., Morgan C. W., 2003, ApJ, 585, 1056
Hambly N. C. et al., 2001, MNRAS, 326, 1279
Hellier C. et al., 2011, EPJ Web Conf., 11, 1004
Henden A., Munari U., 2014, Contrib. Astron. Obs. Skalnate Pleso, 43, 518
Hertzsprung E., 1922, Bull. Astron. Inst. Neth., 1, 91
Høg E. et al., 2000, A&A, 355, L27
Kapteyn J. C., van Rhijn P. J., 1920, ApJ, 52, 23
Kordopatis G. et al., 2013, AJ, 146, 134
Lépine S., Shara M. M., 2005, AJ, 129, 1483
Marshall D. J., Robin A. C., Reylé C., Schultheis M., Picaud S., 2006, A&A, 453, 635
Morgan W. W., Keenan P. C., 1973, ARA&A, 11, 29
Nordström B. et al., 2004, A&A, 418, 989
Ofek E. O., 2008, PASP, 120, 1128 (OF08)
Pecaut M. J., Mamajek E. E., 2013, ApJS, 208, 9
Pepper J. et al., 2007, PASP, 119, 923
Pepper J., Kuhn R. B., Siverd R., James D., Stassun K., 2012, PASP, 124, 230
Perryman M. A. C., ESA, eds 1997, ESA SP-1200: The HIPPARCOS and TYCHO Catalogues. ESA, Noordwijk
Perryman M. A. C. et al., 2001, A&A, 369, 339
Pickles A. J., 1998, PASP, 110, 863
Pickles A., Depagne É., 2010, PASP, 122, 1437
Pollacco D. L. et al., 2006, PASP, 118, 1407
Rauer H. et al., 2014, Exp. Astron., 38, 249
Recio-Blanco A. et al., 2016, A&A, 585, A93
Reid I. N., Gizis J. E., Hawley S. L., 2002, AJ, 124, 2721
Ricker G. R. et al., 2015, J. Astron. Telesc. Instrum. Syst., 1, 014003
Robin A. C., Reylé C., Derrière S., Picaud S., 2003, A&A, 409, 523
Robin A. C. et al., 2012, A&A, 543, A100
Schultheis M. et al., 2014, AJ, 148, 24
Skiff B. A., 2014, VizieR Online Data Catalog, 1, 2023
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Soubiran C., Le Campion J.-F., Cayrel de Strobel G., Caillo A., 2010, A&A, 515, A111
Stassun K. G., Pepper J. A., Oelkers R., Paegert M., De Lee N., Sanchis-Ojeda R., 2014, preprint (arXiv:1410.6379)
Steinmetz M. et al., 2006, AJ, 132, 1645
Stetson P. B., 2000, Publ. Astron. Soc. Pac., 112, 925

Straižys V., Lazauskaitė R., Brown A. G. A., Zdanavičius K., 2006, Balt. Astron., 15, 449
Teig M., 2008, PASP, 120, 474
Valenti J. A., Fischer D. A., 2005, ApJS, 159, 141
van Altena W. F., 2011, Rev. Mex. Astron. Astrofis. Ser. Conf., 40, 290
van Leeuwen F., 2007, A&A, 474, 653
Yanny B. et al., 2009, AJ, 137, 4377
Zacharias N., Monet D. G., Levine S. E., Urban S. E., Gaume R., Wycoff G. L., 2004, Am. Astron. Soc. Meeting 205; BAAS, 36, 1418
Zacharias N., Finch C. T., Girard T. M., Henden A., Bartlett J. L., Monet D. G., Zacharias M. I., 2013, AJ, 145, 44
Zdanavičius K., 1998, Balt. Astron., 7, 551
Zdanavičius K., 2005, Balt. Astron., 14, 104

This paper has been typeset from a TEX/LATEX file prepared by the author.