

## CHROMOSPHERICALLY ACTIVE STARS IN THE RADIAL VELOCITY EXPERIMENT (RAVE) SURVEY. I. THE CATALOG

M. ŽERJAL<sup>1</sup>, T. ZWITTER<sup>1,2</sup>, G. MATIJEVIČ<sup>3</sup>, K. G. STRASSMEIER<sup>4</sup>, O. BIENAYMÉ<sup>5</sup>, J. BLAND-HAWTHORN<sup>6</sup>, C. BOECHE<sup>7</sup>,  
K. C. FREEMAN<sup>8</sup>, E. K. GREBEL<sup>7</sup>, G. KORDOPATIS<sup>9</sup>, U. MUNARI<sup>10</sup>, J. F. NAVARRO<sup>11,18</sup>, Q. A. PARKER<sup>12,13,14</sup>,  
W. REID<sup>12,13</sup>, G. SEABROKE<sup>15</sup>, A. SIVIERO<sup>4,16</sup>, M. STEINMETZ<sup>4</sup>, AND R. F. G. WYSE<sup>17</sup>

<sup>1</sup> Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia; [marusa.zerjal@fmf.uni-lj.si](mailto:marusa.zerjal@fmf.uni-lj.si)

<sup>2</sup> Center of Excellence SPACE-SI, Aškerčeva cesta 12, 1000 Ljubljana, Slovenia

<sup>3</sup> Department of Astronomy and Astrophysics, Villanova University, 800 E Lancaster Avenue, Villanova, PA 19085, USA

<sup>4</sup> Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany

<sup>5</sup> Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, 11 rue de l'Université, F-67000 Strasbourg, France

<sup>6</sup> Sydney Institute for Astronomy, School of Physics A28, Sydney, NSW 2006, Australia

<sup>7</sup> Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr. 12-14, D-69120 Heidelberg, Germany

<sup>8</sup> Research School of Astronomy and Astrophysics, Australia National University, Weston Creek, Canberra, ACT 2611, Australia

<sup>9</sup> Institute of Astronomy, Cambridge University, Madingley Road, Cambridge CB3 0HA, UK

<sup>10</sup> INAF Osservatorio Astronomico di Padova, I-36012 Asiago, Italy

<sup>11</sup> Department of Physics and Astronomy, University of Victoria, Victoria BC, V8P 5C2, Canada

<sup>12</sup> Department of Physics and Astronomy, Macquarie University, Sydney, NSW 2109, Australia

<sup>13</sup> Research Centre for Astronomy, Astrophysics and Astrophotonics, Macquarie University, Sydney, NSW 2109, Australia

<sup>14</sup> Australian Astronomical Observatory, P.O. Box 296 Epping, NSW 1710, Australia

<sup>15</sup> Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, RH5 6NT, UK

<sup>16</sup> Department of Physics and Astronomy, Padova University, Vicolo dell'Osservatorio 2, I-35122 Padova, Italy

<sup>17</sup> Johns Hopkins University, Homewood Campus, 3400 North Charles Street, Baltimore, MD 21218, USA

*Received 2013 April 1; accepted 2013 August 14; published 2013 October 7*

### ABSTRACT

RAVE, the unbiased magnitude limited survey of southern sky stars, contained 456,676 medium-resolution spectra at the time of our analysis. Spectra cover the Ca II infrared triplet (IRT) range, which is a known indicator of chromospheric activity. Our previous work classified all spectra using locally linear embedding. It identified 53,347 cases with a suggested emission component in calcium lines. Here, we use a spectral subtraction technique to measure the properties of this emission. Synthetic templates are replaced by the observed spectra of non-active stars to bypass the difficult computations of non-local thermal equilibrium profiles of the line cores and stellar parameter dependence. We derive both the equivalent width of the excess emission for each calcium line on a 5 Å wide interval and their sum  $EW_{IRT}$  for  $\sim 44,000$  candidate active dwarf stars with signal-to-noise ratio  $>20$ , with no cuts on the basis of the source of their emission flux. From these,  $\sim 14,000$  show a detectable chromospheric flux with at least a  $2\sigma$  confidence level. Our set of active stars vastly enlarges previously known samples. Atmospheric parameters and, in some cases, radial velocities of active stars derived from automatic pipelines suffer from systematic shifts due to their shallower calcium lines. We re-estimate the effective temperature, metallicity, and radial velocities for candidate active stars. The overall distribution of activity levels shows a bimodal shape, with the first peak coinciding with non-active stars and the second with the pre-main-sequence cases. The catalog will be made publicly available with the next RAVE public data releases.

*Key words:* stars: activity – stars: emission-line, Be

### 1. INTRODUCTION

RAVE (RADial Velocity Experiment; Steinmetz et al. 2006; Zwitter et al. 2008; Siebert et al. 2011; Kordopatis et al. 2013) is an ongoing spectroscopic southern sky survey with an unbiased magnitude-limited selection of stars ( $9 < I < 12$ ). The aim of the survey is a determination of radial velocities and an estimation of atmospheric parameters (effective temperature, surface gravity, and metallicity) of Galactic stellar populations. The typical signal-to-noise ratio (S/N) of the measured spectra is  $\sim 40$ ; their typical resolving power is in the mid-range ( $R \sim 7500$ ). All spectra in the sample are continuum normalized and shifted to zero heliocentric radial velocity (RV). The survey covers the near-infrared range from  $\sim 8410$  Å to  $\sim 8795$  Å where the singly ionized calcium triplet ( $\lambda\lambda = 8498, 8542, 8662$  Å) dominates the spectral shape for a broad range of atmospheric parameters, with temperature being the most important. There

is practically no contribution of telluric lines in that part of the spectrum; the only significant spectral signature of the interstellar medium is a diffuse interstellar band at 8620 Å (Munari et al. 2008). All spectra used in this study were obtained between 2004 April 8 and 2012 October 20. These stars ( $\sim 350,000$ ) have been classified with the locally linear embedding (LLE) method by Matijevič et al. (2012) in order to explore the morphology of the spectra, reveal problematic examples, and discover peculiar stars. The sample, based on the results of the LLE method, consists of  $\sim 90\%$ – $95\%$  normal single stars. However, two large distinct groups of peculiars were found. One of them, active stars, show chromospheric emission in calcium lines ( $\sim 50,000$  stars).

The phenomenon of chromospheric emission occurs in young main-sequence solar-type or cooler dwarfs with convective envelopes and affects strong spectral lines with the largest impact on  $H\alpha$ , Ca II H and K (3969, 3934 Å), and the Ca II IR triplet. The latter is covered by the RAVE survey. There is strong emission in the Mg II H and K cores, but it occurs at 2800 Å,

<sup>18</sup> Senior CIFAR Fellow.

which is not observable from the ground. The chromospheric emission component is seen as an additional flux at the central wavelengths of strong lines. It can also be Doppler shifted with respect to the RV of the stellar photosphere. Younger stars appear to be more active, and since chromospheric activity level correlates well with a diminishing stellar rotation rate through coupling by the magnetic dynamo effect, it can be used as an age diagnostic tool for the nearest relatively young stars (0.6–4.5 Gyr; Mamajek & Hillenbrand 2008). The nature of chromospheric activity evolution has been questioned by Pace (2013), Zhao et al. (2011), and Lyra & Porto de Mello (2005). Hall (2008) presents a review of recent advances in stellar chromospheric activity research and Soderblom (2010) reviews its connection to stellar ages while Strassmeier (2009) discusses its connection to starspots.

The majority of emission datasets in the literature cover the Ca II H and K lines. The “HK Project” at Mount Wilson Observatory operated from 1966 through 2003 and gathered a large collection of multiple observations of  $\sim 1300$  stars over 40 yr. This collection offers an invaluable opportunity to study cyclical variations of activity (Wilson 1978; Duncan et al. 1991; Baliunas et al. 1995).

Other large datasets were described by Henry et al. (1996,  $\sim 800$  stars), Strassmeier et al. (2000,  $\sim 1000$  stars), Wright et al. (2004,  $\sim 1200$  stars), Gray et al. (2006,  $\sim 1700$  stars), Isaacson & Fischer (2010,  $\sim 2600$  main-sequence and giant stars) and Zhao et al. (2013, over 13,000 F, G, and K disk stars).

Chromospheric emission in Ca II infrared triplet (IRT), discussed below, has not been investigated yet to a large extent. The line at  $8542 \text{ \AA}$  was investigated for 49 stars by Linsky et al. (1979), who concluded that this line is a good diagnostic of chromospheric activity. Dempsey et al. (1993) analyzed Ca II IRT lines of 45 active F2–M5 stars. Soderblom et al. (1993) used the line at  $8542 \text{ \AA}$  to disentangle the activity of Pleiades cluster members. Mallik (1997) studied 146 stars brighter than  $V = +7.0$ . Jackson & Jeffries (2010) analyzed 237 stars in NGC 2516.

Here, we describe a sample of newly discovered  $\sim 44,000$  candidate active stars drawn from an unbiased magnitude-limited spectroscopic survey. Candidate active stars cover a wide range of activity levels, from very strong emission objects to numerous examples with only a marginal detection or no detection at all. We note that the presence of an emission component does not imply a chromospheric origin. In fact, some of the most active candidate stars may well be examples of other kinds of activity, for example, Strassmeier et al. (2005) report the emission component in classical T Tauri stars where the emission partly comes from accretion shocks. Tomasella et al. (2010) present an atlas of different peculiar stellar types in the RAVE wavelength range, many of them with an emission in the Ca II IRT. However, since the chromospheric origin holds for the vast majority of cases, we use this reference throughout the paper.

A detailed description of our sample of candidate active stars is given in Section 2. The summary of the activity analysis procedure and new parameter estimation is provided in Section 3. In Section 4, we comment on the distribution of stars with different activity levels, while a general discussion and conclusions are given in Section 5. A description of the catalog of chromospherically active candidate stars, which is the main result of this paper, is given in Appendix A.

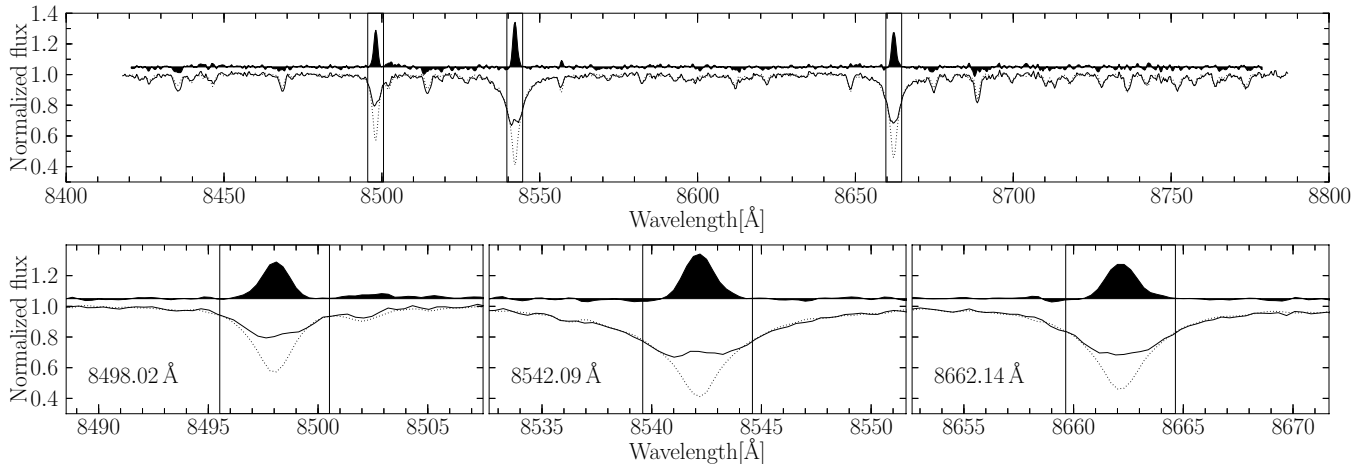
## 2. CATALOG OF CANDIDATE ACTIVE STARS

The selection of candidate active stars from the entire RAVE database of over 500,000 spectra is based on the classification

by the LLE method. The LLE method is a general dimensionality reduction algorithm with the property of conserving locally linear patches of the high-dimensional manifold. So, the neighbors in low-dimensional projected space are also the neighbors in the original high-dimensional space. The method is executed in three separate steps. First, it finds the nearest neighbors for each data point. Then, it calculates the weights that best represent a given data point as a linear combination of its neighbors in the original space. Finally, it finds a low-dimensional space in which each data point is still best represented by the same set of weights that were calculated in the previous step. For a more detailed description of the method, see Roweis & Saul (2000), Vanderplas & Connolly (2009), or Matijević et al. (2012) and references therein.

Matijević et al. (2012) extended the method for spectral classification purposes. Because the method is unable to handle large data samples in a reasonable amount of time, the procedure was broken down into the following steps. A base set of 5000 spectra was randomly selected from the entire RAVE database and each spectrum from the set was rebinned to a common wavelength range with 1200 bins (providing 5000 data points in the 1200-dimensional wavelength space). Using the LLE method with the number of nearest neighbors set to 20, the optimal two-dimensional space was found. While the dimensionality of the projected space might not reveal all the potentially interesting features of the underlying spectra, it is sufficient to make a distinction between various spectral groups. After the low-dimensional space was calculated, the rest of the spectra from the database of several hundred thousand objects were rebinned and projected onto this space. Due to the lack of rare peculiar morphologies in the base set (spectra in the base set were selected randomly and it is therefore unlikely that less common types were picked), the two-dimensional projection of such objects is misleading and they are usually not grouped together properly. To overcome this problem, we reduced the projected sample to a new set of 5000 points by removing many points from the densely populated areas full of very common normal stars while preserving the outliers representing rare morphologies. This ensures that the new set includes more peculiars relative to normal stars. The newly generated sample served as the new base set for the next iteration of calculating the optimal low-dimensional space, projecting the spectra, and reducing the projection. This was repeated several times to ensure that the final base set represented all spectral morphologies found in the database as uniformly as possible. Spectra in the final base set were manually classified into 11 classes. The rest of the spectra were then automatically classified by finding the 20 nearest neighbors from the base set and calculating the linear weights. This provided a quantitative measure of how much signal of each of the base neighboring spectra is represented in each spectrum. The closest 20 neighbors were arranged in a sequence corresponding to the order of the weights, with the base spectrum having the largest weight coming first and so on.

Our catalog of candidate active stars described in this paper includes all objects from the RAVE database that have at least one chromospherically active neighbor within its first six neighbors. The number of neighbors to sufficiently describe the morphology of the spectrum is chosen as the limit where the contribution of the next successor to the information about the spectrum is negligible. Furthermore, we kept only those objects with S/N greater than 20 because under this limit the re-estimation of atmospheric parameters, RV (not based on calcium lines but on weaker Fe I and Ti I lines), and activity levels are unreliable. We lose  $\sim 18\%$  of stars with this S/N cut, but since



**Figure 1.** Top panel: an example for a RAVE spectrum of a candidate active star (HD 220054). The solid line is the spectrum itself and the dotted line is its nearest normal neighbor that is used to determine the subtracted spectrum (black area on the top). The three bands are the 5 Å wavelength ranges that are not taken into account when searching for the nearest neighbor but are used in determining the stellar activity rate. Bottom panels: enlarged sections with calcium lines. Central wavelengths for each line are given in the left bottom corner.

we have a large database, the overall quality of the rest of the database is better. Our final database of candidate active spectra contains  $\sim 44,000$  stars that were recognized as candidates for chromospheric activity by the LLE classification (Matijević et al. 2012). Most of them are main-sequence stars of K spectral type. The emission flux of the sample stars continuously ranges from an undetectable amount in almost inactive stars to examples with an emission component completely filling the calcium line and even extending above the continuum level in some cases (see Figure 1). Our catalog contains candidates for chromospherically active stars. However, since only one active spectrum is required within the first six neighbors for stars in our catalog, other neighbors can be of any type. Consequently, the results should be treated with caution because we analyzed all objects with affected calcium lines, paying no respect to the origin of the additional emission flux. For this reason, the term “active star” refers to a candidate active star hereafter in this paper.

In addition to the database of active stars, we use another database of measured “normal” reference spectra that are used to compute the subtracted emission flux in this paper. Single-lined, main-sequence dwarfs without any peculiarities (our requirement was that at least the first seven LLE flags indicate that there are no peculiarities), with high S/N ( $> 30$ ), temperatures below 7000 K, and  $\log g > 3.5$  are denoted “normal stars” within this paper for simplicity. Our database of selected normal stars is called the “normal database”; its selection procedure is given in the next chapter. Kordopatis et al. (2013, hereafter DR4) recently computed an updated list of stellar parameters for normal stars. Their computation is based on the DEGAS and MATISSE algorithms using the Kordopatis et al. (2011) grid. So, we decided to use the values of stellar parameters from Kordopatis for the normal database. As we explain below, the parameter values pertaining to active stars are therefore compatible with this recent re-estimation of stellar parameters.

All RAVE pipelines are designed for the most abundant group of spectra ( $\sim 90\%$ – $95\%$  of the entire dataset)—i.e., stars that show no peculiarities. Radial velocities and stellar atmospheric parameters for active spectra might suffer from systematic offsets because activity makes the three calcium lines—a dominant feature in RAVE spectra—shallower. One

of our goals is therefore a re-estimation of stellar parameters for these stars.

### 3. ESTIMATION OF THE CHROMOSPHERIC ACTIVITY LEVEL

The most commonly used proxies for Ca II H and K chromospheric activity are the dimensionless  $S$  and  $R'_{\text{HK}}$  indices. The  $S$  index was introduced by Vaughan et al. (1978) as the strength of chromospheric emission in both line cores. It is proportional to the sum of fluxes in the H and K lines and normalized with respect to the nearby continuum:  $S \propto (\mathcal{F}_H + \mathcal{F}_K)/(\mathcal{F}_R + \mathcal{F}_V)$ . The full width at half maximum of both the K and H bands with a triangular profile is  $\sim 1 \text{ \AA}$ ; the fluxes  $\mathcal{F}_R$  and  $\mathcal{F}_V$  are the red and violet continuum bands from 3991–4011 Å and 3891–3911 Å, respectively. The  $S$  index has a color term and includes a contribution from the stellar photosphere. The second and most commonly used activity measure is  $R'_{\text{HK}}$ , which was introduced by Noyes et al. (1984). The  $R'_{\text{HK}}$  index is the ratio of the emission flux in the cores of the calcium H and K lines (the bandwidth is 1 Å for each line core) to the total bolometric flux of the star.

A possible proxy for the Ca II IRT part is a central line depression based on the flux difference between the observed central flux level and non-local thermodynamic equilibrium (LTE) synthetic spectrum. It was described by Andretta et al. (2005). Soderblom et al. (1993) and Jackson & Jeffries (2010) adopted the chromospheric flux, while Mallik (1997) used the sum of equivalent widths of the three emission components. The empirical flux scales for spectral types A to early M for luminosity classes I to V, including the IRT range, were first derived by Hall (1996). An analogous measure to the  $R'_{\text{HK}}$  for the IRT part is given in Marsden et al. (2009) as  $R'_{\text{IRT}} = F'_{\text{IRT}}/\sigma T_{\text{eff}}^4$ , where  $F'_{\text{IRT}}$  is the surface flux emitted by the IRT calcium lines. The surface flux is derived by a conversion of the equivalent width of the emission excess. The latter is revealed by a spectral subtraction technique. The idea is to find a synthetic counterpart of the observed active spectrum and to remove the photospheric contribution by subtracting the two. The subtracted spectrum represents the chromospheric emission flux.

Disadvantages of this method are the requirement of precise values of the atmospheric parameters for the active star and the

**Table 1**  
Selection Criteria for Normal Inactive Observed Spectra (See the Text).

No.	$N_n$	$T_{\text{eff}}$ (K)	S/N	$N$
1	7	$< 4500$	30	406
2	7	$4500 \leq T_{\text{eff}} < 5000$	40	832
3	7	$4500 \leq T_{\text{eff}} < 5500$	40	3648
4	7	$5500 \leq T_{\text{eff}} < 6000$	50	5379
5	10	$6000 \leq T_{\text{eff}} < 7000$	80	2187

**Notes.** The database is built from five subsets and contains 12,452 objects. The second column ( $N_n$ ) gives the number of the first nearest neighbors of the spectrum that were required to be normal (LLE-based selection) and the S/N indicates the minimal required value.  $N$  is the number of spectra in the particular subset. The most relaxed requirements are used for the coolest stars.

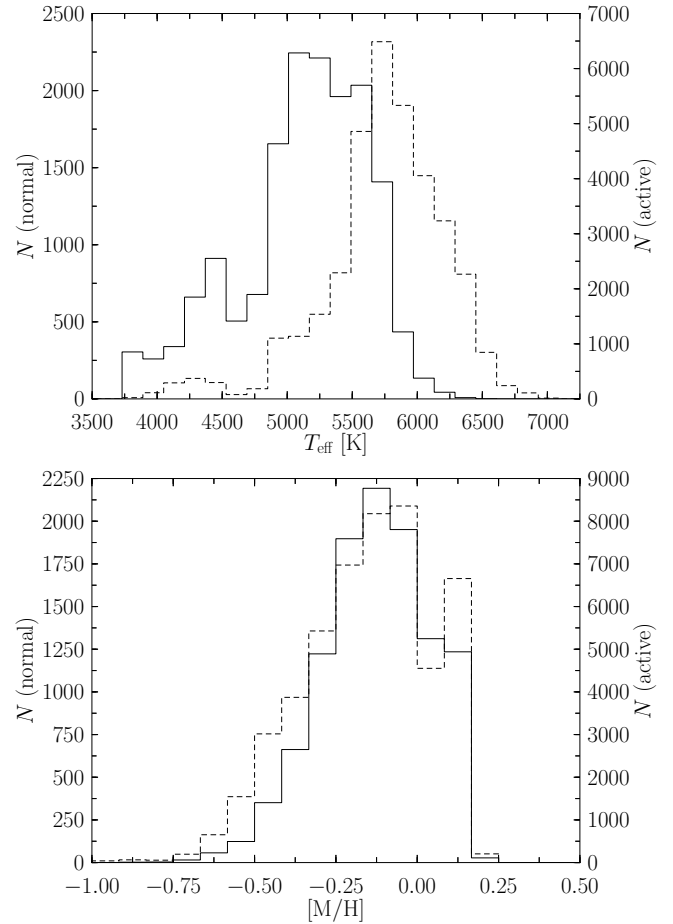
non-LTE nature of the core of the line, which exactly overlaps with the expected emission flux. The synthetic spectrum should also be artificially broadened due to stellar rotation (a poorly determined quantity for RAVE spectra) and appropriately adapted to the resolving power of the instrument (to be precise, the instrumental point-spread function, the spectral purity, and scattered light within the spectrograph might also play a role here).

Since the majority of stars in the RAVE database are normal stars (38% of them are dwarfs with  $\log g \geq 3.5$ ), we seized the opportunity and used observed normal high S/N spectra instead of synthetic templates to compute the subtracted spectra. Our normal database contains 12,452 observed spectra. It is large enough to cover the entire parameter space of candidate active stars—effective temperature and metallicity variations, together with gravity, possible variations of resolving power along the spectrum, and rotational velocity.

The best matching templates for each of the active spectra are found by the  $\chi^2$  nearest neighbor search where the cores of the calcium lines are not taken into account (see Figure 1). When the nearest normal neighbor is found, we assume that it represents the active spectrum itself in a quiet state with the same stellar and instrumental parameters. One may ask if the selected normal stars are completely inactive or not. We checked that their activity levels are below our detection limits, so they can serve our purposes well. The equivalent widths of the chromospheric flux in the calcium lines are computed from the subtracted spectra. The details of the procedure are given below.

### 3.1. Normal Database Selection

A large database of normal spectra assumed to represent active spectra in a completely quiet state was built to achieve nearly uniform coverage of the entire stellar parameter space. Since chromospheric emission occurs in main-sequence dwarfs, we set the upper temperature limit of the database to 7000 K and the gravity limit to  $\log g \geq 3.5$ . The lower limit in the temperature is 3500 K because this is the limit of the synthetic grid (Munari et al. 2005) used in the RAVE DR3 pipeline (Siebert et al. 2011) and there is only a small number of stars with temperatures below this limit because this survey is magnitude limited; therefore, its low luminosity stars lie in a small volume in the vicinity of the Sun. Most stars with true values lower than 3500 K actually fall in the 3500–3600 K range for the DR3 pipeline. The determination of parameters in the DR4 pipeline (Kordopatis et al. 2013) for normal spectra is based on a synthetic grid with a step size of 250 K for spectra between 4000 and 8000 K and 200 K for spectra between 3000 and 4000 K. The first group ranges from 0.0–5.0 dex with a step



**Figure 2.** Parameter distribution for surface temperature and metallicity. The solid line represents candidate active stars while the dashed one indicates inactive dwarfs (normal database). Bin heights are given on the right and left, respectively. Values for both sets on plots were re-estimated in this paper as an average of values for the five closest inactive spectral tracings, in the  $\chi^2$  sense.

size of 0.5 dex in  $\log g$  and from 0.0–5.5 dex in the second group. The metallicity ranges from  $-5.0$  to  $+1.0$  dex with a step size of 0.25 dex. Typical internal uncertainties of the RAVE pipeline output are 170 K in temperature, 0.27 dex in  $\log g$ , and 0.16 dex in metallicity.

A spectrum was recognized as normal in the same manner as in the case of the selection of active stars. Due to a much larger set of normal spectra, the requirements for the nearest neighbors' quality and confidence were more rigorous. Details are given in Table 1. We note that the density of normal stars is very low outside the main sequence where possible active outliers (pre-main-sequence stars) might be found. So, we relaxed the selection criteria in regions scarcely populated with normal stars. In particular, we were forced to accept also observed spectra with a lower S/N ratio. In total, we used five different subsets that were merged into the final normal database (see Table 1).

We culled the database in the next step to reduce the number of objects in extremely dense parts of parameter subspace (surface temperature, gravity, and metallicity). We divided the parameter subspace into smaller volume units and kept deleting randomly selected stars until the number density in each volume unit dropped below the predefined density. Finally, a set of 12,452 normal spectra remained and we declared these objects to be our normal database. The distribution of effective temperatures and metallicities of the stars in this database is shown in Figure 2, together with the parameters for active stars.

### 3.2. Template Spectra

For each active spectrum, we need to select a few best matching template spectra from the normal database. The match was evaluated using  $\chi^2$  statistics, ignoring the cores of the calcium lines. We decided to cut the range of  $\lambda_c \pm 2.5 \text{ \AA}$  around the central wavelength ( $\lambda_c$ ) of each calcium line, so we removed three  $5 \text{ \AA}$ -wide ranges (that is, 18 pixels). The width of the cut was determined by visual inspection of the most active spectra where the emission peak is the most prominent (well above the continuum level), and hence the widest (see Figure 1; the omitted wavelength range is marked with bands). Most emission features are typically shallower and affect only the cores of the calcium lines. Although all three cuts intervene far into the wings of the Ca lines, there are other relatively strong spectral lines in the spectra (mostly Fe I and Ti I lines) so the  $\chi^2$  nearest neighbor search can work. Another reason for the relatively wide omitted range is the quite numerous cases of moderate velocity shifts of the emission components.

The distance between two spectra was characterized as

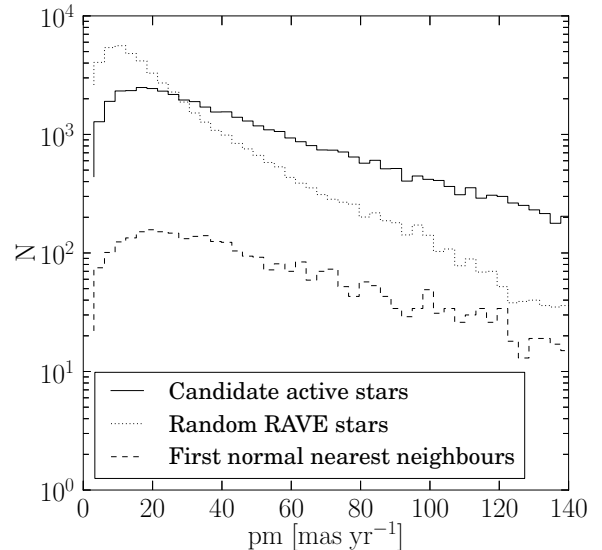
$$\chi^2 = \sum_i (f_a^i - f_n^i)^2, \quad (1)$$

where  $f_a^i$  and  $f_n^i$  denote flux values for the  $i$ th wavelength bin of the active and normal spectrum, respectively.

Equation (1) assumes both spectra share a common wavelength scale. This way, all the spectra in both databases were linearly interpolated to the same wavelength bins. We chose the same interpolation wavelength range and step size as was used in the LLE computation:  $8422 \text{ \AA} \leq \lambda \leq 8780 \text{ \AA}$  with 1200 bins linear in wavelength, i.e., with  $\sim 0.3 \text{ \AA}$  spacing. This represents a moderate oversampling over the observed  $\sim 0.37 \text{ \AA}$  sampling. Lower and upper interpolation wavelength limits were chosen as typical observed spectrum limits. However, a small fraction of RAVE spectra exist where the interpolation wavelength range is not completely covered. Due to this fact, we compare only the intersection of both spectra each time.

All RAVE spectra are continuum normalized. However, the normalization might in some cases not be ideal and the continuum levels of the two spectra may differ. All normal spectra in the database were multiplied so that their median level matched the median of an active spectrum. All medians were derived ignoring the windows around the calcium lines. This was done for each active spectrum from the database.

A basic underlying assumption of the  $\chi^2$  comparison is that all RAVE spectra are shifted to a zero heliocentric RV. Since the selection of all spectra from a normal database is based on the results of the LLE method (where the same assumption holds true), their velocity values are highly reliable. However, this may not be the case for active spectra because emission components of the calcium line could affect the parameter determination and cause RV jitter (Martínez-Arnáiz et al. 2010). Hence, the nearest neighbor search was performed iteratively: in each iteration, the best matching template was found and the active spectrum was shifted in velocity space with respect to the template. Due to missing calcium lines, only a part of the entire spectrum ( $8668 \text{ \AA} \leq \lambda \leq 8780 \text{ \AA}$ ) with relatively strong Fe I spectral lines ( $8674.75, 8688.63 \text{ \AA}$ ) was used to compute cross-correlation functions and derive radial velocities. Spectra were shifted iteratively until the last correction decreased below  $1 \text{ km s}^{-1}$  (that is, 10% of a pixel width, which is  $10.5 \text{ km s}^{-1}$ ). The errors of the computed radial velocities are based on the dispersion of



**Figure 3.** Comparison of proper motions of candidate active stars (solid line,  $\sim 50,000$  stars) with their nearest normal neighbors (dashed line) and randomly chosen stars from the overall RAVE sample (dotted line,  $\sim 50,000$  stars). The median values of the distributions are  $\sim 17 \text{ mas yr}^{-1}$  for the entire sample and  $\sim 40 \text{ mas yr}^{-1}$  for candidate active stars. The distributions are shown on a logarithmic scale. The comparison supports the assumption that the active sample consists mainly of dwarf stars.

data points of the cross-correlation function (standard deviation, denoted by  $\sigma$ ) around a fitted parabola and correspond to the half-width of the parabola at a  $1\sigma$  level below the maximum of the function. If the Fe lines are too weak, the cross-correlation function peak is wide and contributes larger errors.

The overall procedure for each active spectrum is as follows.

1. Interpolate all spectra (both normal and active) to match the same wavelength bins.
2. Cut calcium lines from all spectra (both active and normal) at  $\lambda_c \pm 2.5 \text{ \AA}$ .
3. Renormalize spectra based on their median value comparison.
4. Apply a nearest neighbors search by  $\chi^2$ .
5. Compute a potential RV correction of the spectrum—take only the selected part of spectrum.
6. If an RV correction is applied, repeat the loop from step 2.
7. If the RV correction is below  $1 \text{ km s}^{-1}$ , exit the loop.

Step 1 was done only once at the beginning to save time and was not repeated each time in a loop due to its independence of any variable parameters.

#### 3.2.1. New Stellar Parameters

Each candidate active spectrum was assigned new stellar parameters (effective temperature, metallicity, and gravity). Although the signs of chromospheric activity were already observed in giants as well (see, for example, Dupree et al. 1999), we assumed that all normal and candidate active spectra are dwarfs with  $\log g > 3.5$ , and hence they share similar values of surface gravity due to the fact that this is a very weak and unbound parameter in the RAVE sample. Partial support for our assumption comes from Figure 3, which compares proper motions of three sets of data: candidate active stars ( $\sim 50,000$ ), their nearest normal neighbors ( $\sim 3600$  because some candidate active stars share the same normal neighbors), and randomly chosen stars ( $\sim 50,000$ ) from the entire RAVE

**Table 2**  
Description of the Output Catalog of  $\sim 44,000$  RAVE Candidate Active Stars

Acronym	Type	Unit	Description
rave_obs_id	String	...	Target designation
RAdeg	Float	deg	R.A. (J2000.0)
DEdeg	Float	deg	Decl. (J2000.0)
STN_K	Float	...	S/N
EWirt	Float	Å	EW <sub>IRT</sub>
eEWirt	Float	Å	EW <sub>IRT</sub> error
nnid00	String	...	Target designation of star's first normal neighbor
nnid01	String	...	Target designation of star's second normal neighbor
nnid02	String	...	Target designation of star's third normal neighbor
nnid03	String	...	Target designation of star's fourth normal neighbor
nnid04	String	...	Target designation of star's fifth normal neighbor
Teff_K2	Float	K	New effective temperature
eTeff_K2	Float	K	Error of new effective temperature
Logg_K2	Float	dex	New log $g$
eLogg_K2	Float	dex	Error of new log $g$
Met_N_K2	Float	dex	New [M/H]
eMet_N_K2	Float	dex	Error of new [M/H]
EW8498	Float	Å	EW <sub>8498</sub>
eEW8498	Float	Å	EW <sub>8498</sub> error
EW8542	Float	Å	EW <sub>8542</sub>
eEW8542	Float	Å	EW <sub>8542</sub> error
EW8662	Float	Å	EW <sub>8662</sub>
eEW8662	Float	Å	EW <sub>8662</sub> error
RVshift	Float	km s <sup>-1</sup>	Our RV correction due to activity
eRVshift	Float	km s <sup>-1</sup>	Error of RV correction due to activity
RV	Float	km s <sup>-1</sup>	Total heliocentric RV
eRV	Float	km s <sup>-1</sup>	Error of total heliocentric RV

**Note.** The main features are the activity levels (EWirt), newly estimated stellar parameters (Teff\_K2, Logg\_K2, Met\_N\_K2) and corrected radial velocities.

sample. It is clearly seen that candidate active stars and their nearest neighbors (these are normal stars with  $\log g > 3.5$ ) show a similar distribution shape while the overall sample significantly differs. Stars in the candidate active sample and their neighbors contain a larger fraction of stars with larger proper motions than the overall sample (the median values of the distributions are  $\sim 17$  mas yr<sup>-1</sup> for the entire sample and  $\sim 40$  mas yr<sup>-1</sup> for candidate active stars). Approximately half of the entire RAVE sample is represented by giant stars. Since a fraction of detected giants reaches larger distances due to their larger luminosities compared with dwarfs, these stars lower the median value of the distribution of the proper motion because the average measured proper motion of more distant star is smaller. The larger median proper motion for candidate active stars, together with a distribution similar to the part of the normal database, indicates that the candidate active sample is dominated by dwarfs and that the number of giants is small.

New stellar parameters were defined as the arithmetic mean of the original parameters (from the RAVE pipeline) of its first five nearest normal neighbors. The scatter of these new parameters was estimated as their standard deviation, where the internal uncertainty of each parameter was ignored.

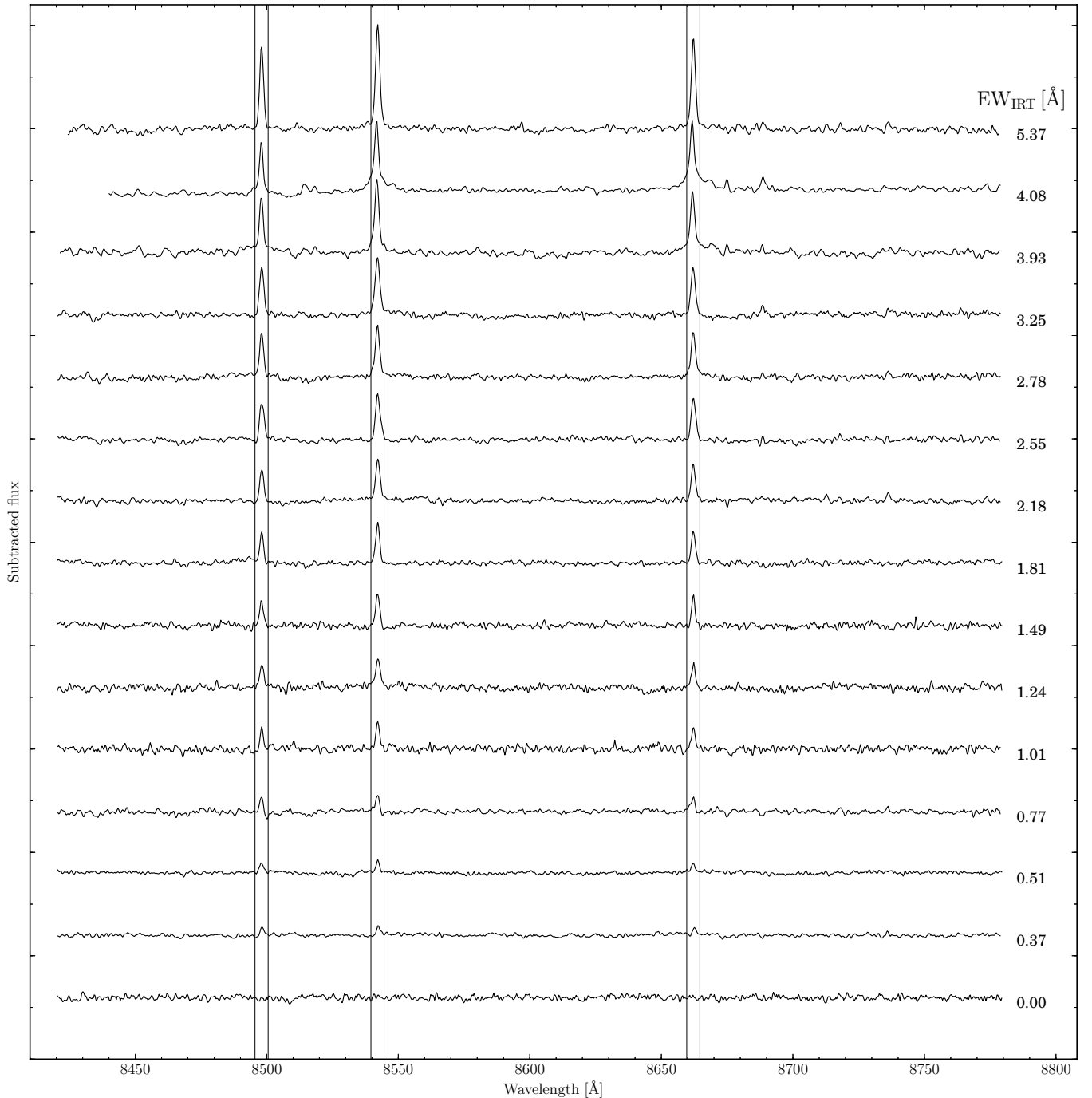
All values of the new parameters of course fall within the normal database range. Figure 2 shows the distribution of the temperature and metallicity for both normal and active spectra (the newly defined parameters). The temperature distribution of active spectra is centered around 5000 K and the average is lower than in the normal database. The distribution of the active objects' metallicities matches the distribution of the normal database.

Nonzero RV shifts of both the normal and active spectra were investigated (normal spectra were not shifted; only a potential shift was calculated for reference). The mean value of shifts for the normal database is zero with a standard deviation of 2.5 km s<sup>-1</sup>. Corrections were applied to  $\sim 61\%$  of all active spectra. The distribution of the RV correction for active spectra has a zero mean value with a standard deviation of 19 km s<sup>-1</sup>. The amplitude of the RV correction for 1812 spectra (that is, 6% of all nonzero corrections) is larger than  $1 \sigma_{RV}$  and 173 spectra cross the  $3 \sigma_{RV}$  level. Here,  $\sigma_{RV}$  is the uncertainty on the derived RV. The median value of all uncertainties is 5.8 km s<sup>-1</sup>. Five percent of the stars with nonzero corrections have velocities  $\geq 5.25$  km s<sup>-1</sup> (half a pixel). Ninety-nine percent of nonzero RV corrections for active stars fall within  $\pm 10.5$  km s<sup>-1</sup> (1 pixel) and there are 101 stars with values greater than  $\pm 21$  km s<sup>-1</sup> (2 pixels).

However, there are  $\sim 20$  very active stars with a displacement of several hundreds of km s<sup>-1</sup>. The reason is that the most prominent features in these spectra are calcium emission lines while the RAVE RV pipeline, which was designed for absorption spectra, applied wrong and very large shifts. RV corrections allowed us to discover the most active spectra in our sample. RV corrections are given in Table 2 in the catalog.

### 3.3. Subtracted Spectra

The chromospheric emission flux was extracted by a subtraction method. The nearest normal neighbor was subtracted from an active spectrum to remove the photospheric contribution to the total flux. The subtracted spectrum contains only the emission flux of the object; the rest oscillates around zero due to the



**Figure 4.** Sequence of emission component spectra arranged from an inactive star at the bottom to one of the most active stars on the top. The labels on the right give the  $EW_{IRT}$  value. The three bands indicate wavelength integration ranges used to derive the equivalent widths of the emission components.

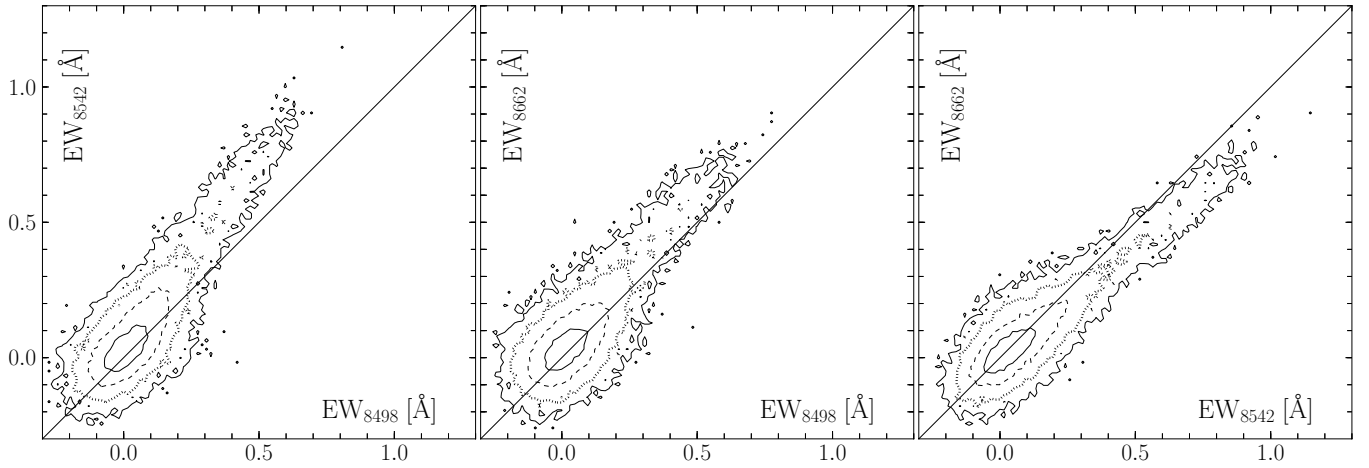
noise. The exceptions are wavelengths of calcium line cores with chromospheric emission. In the case of a nearly quiet activity state, the emission peaks are buried in the noise. With increased activity, they become more and more apparent (Figure 4).

Fifty-eight percent of the subtracted spectra have the highest pixel value among the 18 pixels within the strongest calcium line (at 8542 Å) crossing the  $2\sigma_S$  level and 15% of them cross the  $5\sigma_S$  level. Here,  $\sigma_S$  is the standard deviation of the subtracted spectrum outside the calcium line ranges.

#### 4. ACTIVITY

In this paper, the activity was computed for each of the three calcium lines first as an equivalent width  $EW_\lambda$  in the subtracted

spectrum using the same wavelength range as for the calcium cuts (width of 5 Å, or 18 pixels for each line). The computation was done five times by subtracting in turn each of the five nearest neighbors from the normal database. It turned out that an average  $EW_\lambda$  computed with the first five nearest neighbors was not significantly different from an average using the 10, 15, or 20 nearest neighbors, so we kept the value derived using the five nearest neighbors. The stability of  $EW_\lambda$  confirms the assumption that our normal database completely and sufficiently covers the entire parameter space. However, it may happen in rare cases that the number of nearest neighbors that adequately fits the active spectrum is fewer than five and that their successors have slightly different spectra. In that case, we still take the first five



**Figure 5.** Comparison of emission flux equivalent widths for all three calcium lines. Different linestyles indicate the number of objects per bin in a particular contour. Contour values are two stars for the outer solid line, 11 stars for the dotted line, 55 stars for the dashed line, and 181 stars for the inner solid line. The bin size is  $0.017 \text{ \AA}$ . The scale is logarithmic.

neighbors, but the errors (the scatter) of the equivalent widths, which are estimated as the standard deviation of activity levels within that five neighbors, are higher. The typical error on the equivalent width of the emission component of each calcium line is  $0.05 \text{ \AA}$ .

The equivalent width of the chromospheric flux for each particular line ranges from  $\sim -0.2 \text{ \AA}$  to  $\sim 1 \text{ \AA}$ , or even more in the case of the most active stars. The reason for negative values is a template mismatch in the case of marginally active or inactive stars. Another possibility is that the chromospheric flux is dominated by noise. Yet another reason could be that there are no non-active stars even in the normal database since every star shows at least minimal signs of activity, but the detection of such low levels of activity is well below our limits in the RAVE database.

Figure 5 shows correlations of activity between all three lines. The most obvious characteristic of all three plots is the distribution of activity levels that are concentrated between  $\sim -0.1 \text{ \AA}$  and  $\sim 0.25 \text{ \AA}$ . Such a result is expected because of the selection criteria. Our active database consists of all stars that have at least one active neighbor within the first six neighbors (LLE based classification). This is a relatively weak criterion as we wish to include marginally active candidate stars as well. However, stars with no statistically significant activity can be discarded later for the purposes of further analyses.

The second feature visible on the diagrams is the tightly correlated strengths of lines compared with each other. One can easily see that the  $8542 \text{ \AA}$  line is the strongest and that the  $8498 \text{ \AA}$  line is the weakest. A good correlation is another test for consistency of the results and excludes the possibility that the emission signal comes from another source such as cosmic rays or similar errors. The individual panels of Figure 5 compare two of the lines. The error estimation for each calcium line was computed as a scatter (a standard deviation) among the equivalent widths obtained by subtracting the first five nearest normal neighbors ( $0.05 \text{ \AA}$  in all three cases). Similarly, the dispersion around zero for normal stars for each line is  $0.06 \text{ \AA}$ .

To reduce the dimensionality of the problem and to improve the significance of the results, we introduce the sum of equivalent widths

$$EW_{\text{IRT}} = EW_{8498} + EW_{8542} + EW_{8662}$$

for all three calcium lines for each star. The overall distribution of activity ( $EW_{\text{IRT}}$ ) is shown in Figure 6 and clearly shows a bimodal distribution with two peaks at  $\sim 0.1$  and  $\sim 1.5 \text{ \AA}$ . For reference, the activity of normal stars was derived with the same method (we took second–sixth nearest neighbors since the first was the spectrum itself). As expected, the distribution of values in the normal database is Gaussian shaped around  $-0.05 \text{ \AA}$  with a standard deviation of  $\sigma = 0.16 \text{ \AA}$ . This value is adopted as a general uncertainty in any  $EW_{\text{IRT}}$  value for both normal and active stars. The second reference sample includes pre-main-sequence stars—RAVE stars that are classified as such in the SIMBAD database. This distribution overlaps with the second peak of active stars and perhaps implies a discovery of pre-main-sequence stars in the RAVE survey. This topic will be explored in our next papers.

Another source of uncertainty could be the time variability of active spectra. The HK Project contains 40 yr of data that showed the variability of activity. Our results may reflect the activity level at the time of observation, but this can change with time.

The abscissa along the top of Figure 6 shows the measure of probability  $p_{\log}$  that a star with a given  $EW_{\text{IRT}}$  and  $\sigma$  is more active than a similar inactive normal star. If we have two Gaussian distributions with  $(EW_{\text{IRT}}^0, \sigma_0)$  and  $(EW_{\text{IRT}}^1, \sigma_1)$ , where the equivalent width denotes the mean value and  $\sigma^2$  is the variance of the distribution, and we randomly pick one sample from each distribution, the probability that  $EW_{\text{IRT}}^1 > EW_{\text{IRT}}^0$  is

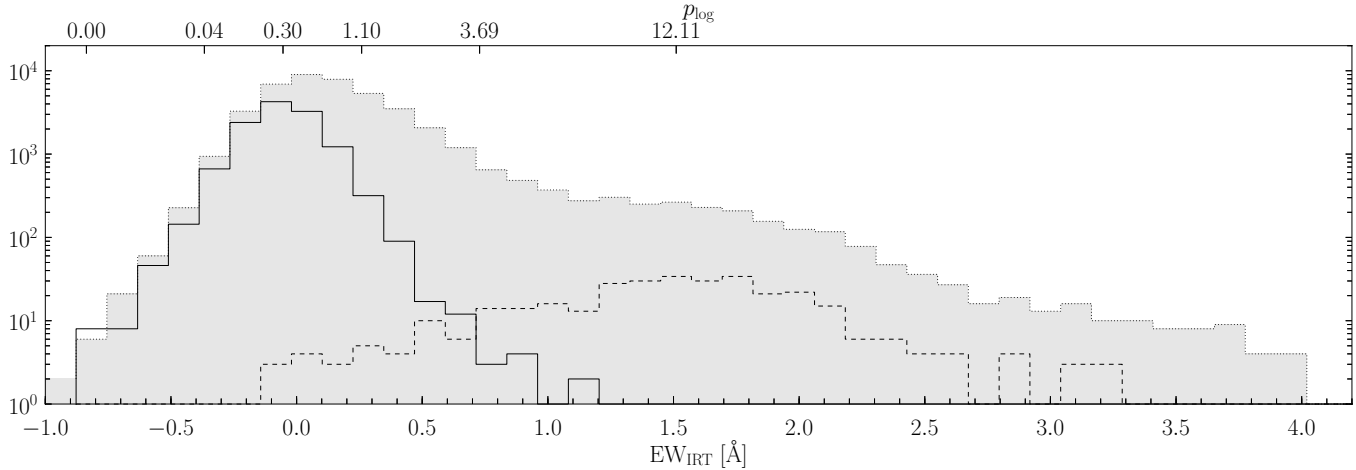
$$P(EW_{\text{IRT}}^1 > EW_{\text{IRT}}^0) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{EW_{\text{IRT}}^1 - EW_{\text{IRT}}^0}{\sqrt{2(\sigma_0^2 + \sigma_1^2)}} \right) \right]$$

(for a derivation, see Matijević et al. 2011). Because the error function  $\operatorname{erf}(z)$  for a given number  $z$  reaches values very close to 1 at  $z = 2$ , we introduced

$$p_{\log} = -\log_{10}(1 - P)$$

to make the differences more distinguishable. However, due to a limited computational precision, we had to stop at  $P \approx 1 - 10^{-13}$ , where  $p_{\log} \approx 12$ . If stars have exactly the same equivalent widths, then  $P = 1/2$  and  $p_{\log} = 0.3$ . Higher  $P$  values imply higher  $p_{\log}$  values. Values on the plot were computed with





**Figure 6.** Distribution of  $EW_{\text{IRT}}$  for active stars (gray histogram), normal stars (which are assumed to be inactive; solid line), and pre-main-sequence stars (selection is based on the SIMBAD classification of RAVE stars; dashed line). The scale is logarithmic.  $p_{\log}$  is a measure of the probability that a star with given  $EW_{\text{IRT}}$  differs from an inactive spectrum.  $p_{\log}$  values correspond to  $5\sigma$  and  $2\sigma$  below zero and  $2\sigma$ ,  $5\sigma$ , and  $10\sigma$  above zero. A  $p_{\log}$  value of 0.3 corresponds to  $EW_{\text{IRT}} = 0 \text{ \AA}$ .

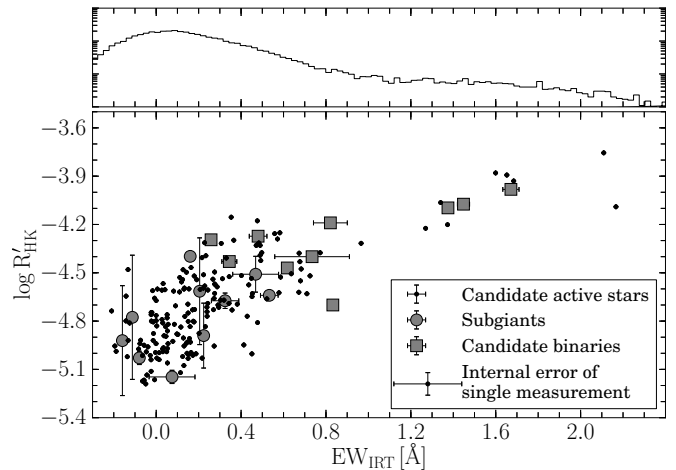
$\sigma_0 = \sigma_1 = 0.16 \text{ \AA}$  and  $EW_{\text{IRT}}^0 = -0.05 \text{ \AA}$  (the mean value of the distribution of normal stars).

Forty-six percent of stars in our active database have  $EW_{\text{IRT}}$  below  $1\sigma$  and 54% of stars have  $EW_{\text{IRT}}$  above  $1\sigma$ . Furthermore, there are 32% of stars with  $EW_{\text{IRT}} > 2\sigma$  ( $\sim 14,200$  stars,  $p_{\log} = 1.10$ ), 19% with  $EW_{\text{IRT}} > 3\sigma$  ( $\sim 8400$  stars,  $p_{\log} = 1.77$ ), and 12% with  $EW_{\text{IRT}} > 4\sigma$  ( $\sim 5200$  stars,  $p_{\log} = 2.63$ ).

## 5. DISCUSSION AND CONCLUSIONS

This paper quantitatively describes our catalog of candidate active RAVE stars. It significantly enlarges previously known samples. It will become public and be part of the next data release. Our catalog is based on the LLE classification of stellar spectra (Matijević et al. 2012) and contains all spectra that have at least one chromospherically active nearest neighbor within its first six neighbors. However, since the other five neighbors can be of any type and because we analyzed all stars with affected calcium lines independent of the origin of the additional emission flux (furthermore, besides the possibility of problems with continuum normalization, in some very rare cases with underestimated S/N, a large noise may fill the calcium lines and mimic the emission), the active star candidates we derive should be treated with caution as they might not be error free. The estimation for activity levels is independent of the RAVE pipeline version because our method is based on a direct comparison of fluxes and omits the use of stellar parameters.

A large database of normal spectra (high S/N, single-lined, main-sequence dwarfs without any peculiarities) was used to apply a spectral subtraction technique on candidate active stars in order to extract the equivalent widths of emission flux in the Ca II IRT lines. Although all spectra were previously shifted to zero heliocentric RV within the RAVE pipeline, it turned out that moderate and very active candidate spectra still suffered from RV displacements due to their shallower or even emission-like calcium lines that could not be treated properly in the pipeline. Therefore, RV corrections had to be applied to 61% of candidate active spectra ( $\geq 5.25 \text{ km s}^{-1}$  for 5% of candidate active stars with nonzero corrections). For more than 99% of stars with nonzero corrections, the amplitude of the shift was up to  $10.5 \text{ km s}^{-1}$ . However, there were  $\sim 20$  exceptional cases of very active candidate stars with emission-like spectra and with displacements of several hundred  $\text{km s}^{-1}$  due to the improper

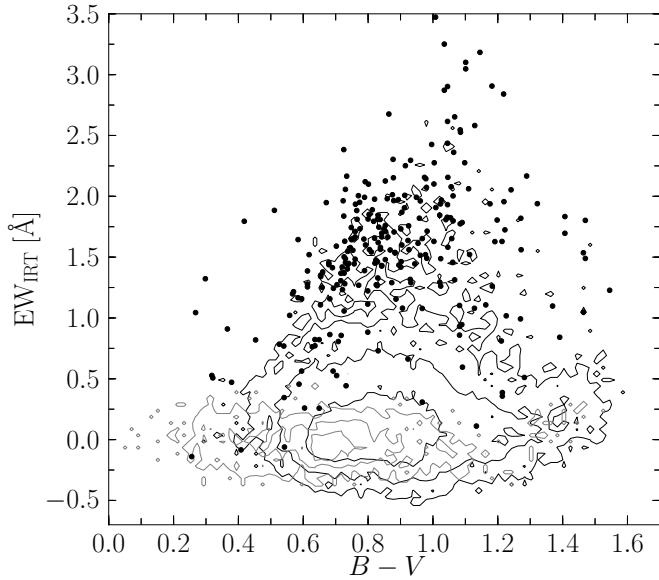


**Figure 7.**  $\log R'_{\text{HK}}$  vs.  $EW_{\text{IRT}}$  values for 211 RAVE stars matched with a set of available online catalogs (they can contain different populations of stars because their selection criteria might differ). Some of the stars were measured multiple times by various authors; in this case, we plot an average value for the star. The typical error of  $EW_{\text{IRT}}$  is  $0.16 \text{ \AA}$ . Our estimation of the  $\log R'_{\text{HK}}$  scatter is 0.07. The large circles indicate “subgiants”—stars that are classified as luminosity class III, IV, or IV–V in the SIMBAD database. The squares are binary candidates found by the LLE classification. The internal error of a single measurement marks the internal error for  $EW_{\text{IRT}}$  while its ordinate error value is the  $\log R'_{\text{HK}}$  scatter. The histogram in the top panel shows the distribution of  $EW_{\text{IRT}}$  values for the entire sample of active RAVE stars on a logarithmic scale.

previous treatment in the RAVE pipeline caused by the strong emission nature of the spectra.

Our analysis shows a good correlation of activity between all three lines (Figure 5). A distribution of stellar activity rates  $EW_{\text{IRT}}$  has a clear bimodal shape where the most abundant peak matches the inactive sample and the most active peak arises at activity levels of pre-main-sequence stars (Figure 6). The most active candidate stars reach equivalent widths of  $\sim 4 \text{ \AA}$ . While the majority of the sample is composed of shallower calcium lines, clear additional emission peaks appear within the lines in moderate and very active candidate stars. Some of these stars even show Doppler shifts that will be discussed in our future paper.

The comparison of our  $EW_{\text{IRT}}$  and  $\log R'_{\text{HK}}$  (for stars from the literature) forms another supplementary table to the catalog. Five hundred and forty three measurements of 211 matching



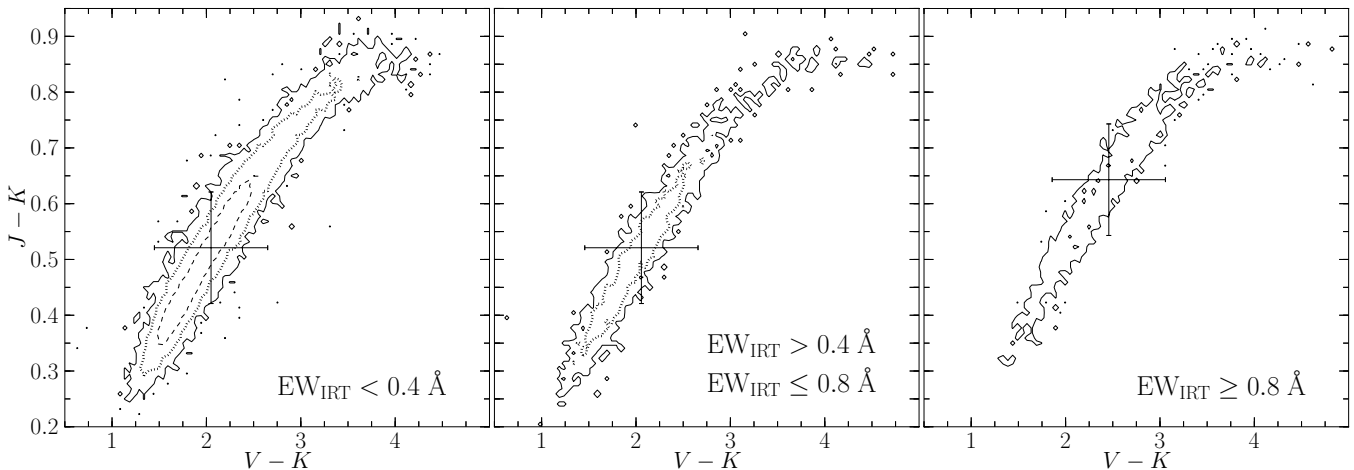
**Figure 8.**  $EW_{\text{IRT}}$  vs.  $(B - V)$ . The gray contours correspond to the normal database (2, 7, and 20 stars per bin; note that a fraction of normal stars lack  $B - V$  data) and the black lines indicate contours for active stars (2, 11, and 55 stars per bin). RAVE pre-main-sequence stars (SIMBAD classification) are added for reference (black dots). The bin size is 0.02 in color and 0.05 Å in  $EW_{\text{IRT}}$ .

stars were found in 21 different catalogs from the northern and southern sky (Duncan et al. 1991; Baliunas et al. 1995; Henry et al. 1996; Rocha-Pinto & Maciel 1998; Strassmeier et al. 2000; Tinney et al. 2002; Gray et al. 2003; Wright et al. 2004; Gray et al. 2006; Jenkins et al. 2006; Cincunegui et al. 2007; Hall et al. 2007; Lockwood et al. 2007; White et al. 2007; Schröder et al. 2009; Isaacson & Fischer 2010; López-Santiago et al. 2010; Martínez-Arnáiz et al. 2010; Arriagada 2011; Jenkins et al. 2011; Wright et al. 2011). In the case of Duncan et al. (1991), Baliunas et al. (1995), Gray et al. (2003, Table 5), Cincunegui et al. (2007), and López-Santiago et al. (2010), the  $S$  index was converted to  $\log R'_{\text{HK}}$  using the relation introduced by Noyes et al. (1984).  $B - V$  colors used in the derivation of  $\log R'_{\text{HK}}$  for stars from Duncan et al. (1991) and López-Santiago et al. (2010) were derived using SIMBAD values. We also checked Knutson

et al. (2010) but found no stars from the RAVE database. Figure 7 shows the correlation between  $EW_{\text{IRT}}$  and  $\log R'_{\text{HK}}$ . There exists a strong correlation between these two quantities. However, most stars are scattered around a less active area and therefore we assume that this sample is subject to relatively high intrinsic variability and measurement errors. This sample consists of single measurements and stars with multiple observations (either in time or in more than one catalog); in these cases, the average values are plotted. The typical scatter of  $\log R'_{\text{HK}}$  in Figure 7 is 0.07. This is the median value of standard deviations of multiple observations. Uncertainties on  $\log R'_{\text{HK}}$  from Henry et al. (1996) are comparable with our estimation from multiple observations. It seems that the sample of 211 matching catalog stars contains mostly stars with a lower activity rate than the RAVE objects (around  $\log R'_{\text{HK}} = -5.1$ ). However, the most active stars from various authors reached  $\log R'_{\text{HK}} \approx -3.8$  with the second (most active) peak at  $\log R'_{\text{HK}} \approx -4.5$ . Figure 7 shows that  $\log R'_{\text{HK}} \sim -5$  is undetectable in the RAVE survey by our measurement technique. Most of the stars in our catalog therefore have only a marginal detection of activity. However, on the other hand, there are numerous very active candidate objects discovered that merit further detailed analysis. Although various authors report a gap in the bimodal  $\log R'_{\text{HK}}$  distribution at  $\log R'_{\text{HK}} \approx -4.75$  (a rough estimate; the gap is called the Vaughan–Preston gap), we suspect that the possible existence of such a gap in our distribution is difficult to detect.

Based on the 58 age estimations of 46 matched stars from 3 of the catalogs (Wright et al. 2004; Isaacson & Fischer 2010; Arriagada 2011), the ages of stars vary from 100 Myr at relatively high activity levels ( $\log R'_{\text{HK}} \sim -4.3$ ) to 10 Gyr at  $\log R'_{\text{HK}} \sim -5.2$ .

This catalog provides an opportunity to further study the possible parameter dependence of activity, which will be explored in our next paper. Line asymmetries could be studied and since the RAVE sample is large, a discovery of peculiars and particularly interesting stars among the active sample is possible. A. Siviero et al. (2013, in preparation) are working on the ongoing observational project of RAVE peculiars (at Asiago Observatory), including chromospherically active stars, that will provide additional observations. Another topic to be discussed in our ongoing project is the dependence of activity on stellar populations and photospheric physics. As an example, Figures 8 and 9 show



**Figure 9.** Color-color diagram ( $J - K$  vs.  $V - K$ ) for candidate active stars.  $J$  and  $K$  are Two Micron All Sky Survey magnitudes and  $V$  is from the APASS catalog. The entire candidate active dataset is divided into three subsets according to  $EW_{\text{IRT}}$ . The median of the distribution (cross on each diagram) moves from  $(J - K, V - K) = (0.52, 2.05)$  for stars with a weak activity to a much redder value of  $(0.64, 2.46)$  for the most active candidate stars. The bin size is 0.05 for  $V - K$  and 0.01 for  $J - K$ . Contour levels are the number of stars per bin: 2, 11, and 67.

the dependence of activity on color ( $B - V$ ) and in color–color diagrams where the most active candidate stars move to the reddest parts of the plot. In Figure 9, we used the APASS magnitudes as given in the DR6 public data release (Henden et al. 2012). Furthermore, Gray et al. (2006) discovered an absence of the Vaughan–Preston gap for  $[M/H] < -0.2$ , implying a question of metallicity dependence on activity level that could be investigated.

It is well known that chromospheric emission anti-correlates with stellar ages (e.g., Soderblom 2010). We plan to investigate and use RAVE stars, which are members of stellar clusters (Conrad et al. 2013), to calibrate the RAVE relation of  $EW_{\text{IRT}}$  with age and to explore the spatial distribution, the distribution with distance from the Galactic plane, and the kinematics of candidate active stars in our Galaxy.

Since the RAVE catalog of candidate active stars consists of a relatively large number of objects, it is a suitable and convenient database to compare with the results of the upcoming Gaia space mission (covering Ca II IRT; Katz et al. 2004; Wilkinson et al. 2005) and HERMES-GALAH (covering  $H\alpha$  and  $H\beta$ ; Zucker et al. 2012).

We would like to acknowledge the very relevant and helpful comments of the anonymous referee, which substantially improved the quality and presentation of the paper.

Funding for RAVE has been provided by the Anglo-Australian Observatory, the Leibniz-Institut fuer Astrophysik Potsdam, the Australian National University, the Australian Research Council, the French National Research Agency, the German Research foundation, the Istituto Nazionale di Astrofisica at Padova, The Johns Hopkins University, the National Science Foundation of the USA (AST-0908326), the W. M. Keck foundation, the Macquarie University, the Netherlands Research School for Astronomy, the Natural Sciences and Engineering Research Council of Canada, the Slovenian Research Agency, Center of Excellence Space.si, the Swiss National Science Foundation, the Science & Technology Facilities Council of the UK, Opticon, Strasbourg Observatory, and the Universities of Groningen, Heidelberg and Sydney. The RAVE Web site is [http://www.rave\\_survey.org](http://www.rave_survey.org).

This research has made use of the SIMBAD database and the VizieR catalog access tool, operated at CDS, Strasbourg, France.

G.M. acknowledges support from NASA ADAP grant number NNX13AF32G.

## APPENDIX

This section is a description of the catalog of candidate active RAVE stars and includes two tables of newly derived parameters. Appendix A is the main catalog of this paper and contains the  $EW_{\text{IRT}}$  of 53,347 stars.

Table 3 in Appendix B provides both RAVE  $EW_{\text{IRT}}$  and  $\log R'_{\text{HK}}$  values available from the literature to compare the results (211 stars). Tables 2 and 3 are both based on RAVE data obtained between 2004 April 8 and 2012 October 20. Since RAVE is an ongoing project, new data will be added in the future.

## APPENDIX A THE CATALOG

This is the description of the output catalog of  $\sim 44,000$  RAVE candidate active stars that will be accessible through VizieR

**Table 3**

Description of the List Containing Both  $\log R'_{\text{HK}}$  and  $EW_{\text{IRT}}$  Values for 543 Measurements of 211 RAVE Stars Found in the Literature and Observed by Other Authors in Their Ca II H and K Lines

Acronym	Type	Unit	Description
rave_obs_id	String	...	Target designation
RAdeg	Float	deg	R.A. (J2000.0)
DEdeg	Float	deg	Decl. (J2000.0)
STN_K	Float	...	S/N
EWirt	Float	Å	$EW_{\text{IRT}}$
eEWirt	Float	Å	$EW_{\text{IRT}}$ error
R'HK	Float	...	$\log R'_{\text{HK}}$
eR'HK	Float	...	Error of $\log R'_{\text{HK}}$
logAge	Float	...	$\log_{10}$ age
Ref	String	...	$\log R'_{\text{HK}}$ reference
RefA	String	...	Age reference
fLLE	Char	...	Average flag from LLE classification
Teff_K2	Float	K	New effective temperature
eTeff_K2	Float	K	Error of new effective temperature
Logg_K2	Float	dex	New $\log g$
eLogg_K2	Float	dex	Error of new $\log g$
Met_N_K2	Float	dex	New $[M/H]$
eMet_N_K2	Float	dex	Error of new $[M/H]$

(Table 2). Among the usual and obvious values, it contains newly estimated stellar parameters (Teff\_K2, Logg\_K2, and Met\_N\_K2) that are the arithmetic mean of the parameters of a star's first five nearest normal neighbors. The scatter of such estimates is derived as the standard deviation of these five values. RVshift is our correction to the original value (only values that exceed  $3\sigma_{\text{RV}}$  due to the S/N limitations) and RV is the corrected radial velocity. EW8498, EW8542, and EW8662 are average values for each line (averaged over five neighbors); their errors are derived in the same way as in the previous case. The same holds true for EWirt, which denotes  $EW_{\text{IRT}}$  (Table 2).

## APPENDIX B

### COMPARISON OF $\log R'_{\text{HK}}$ AND $EW_{\text{IRT}}$

Table 3 is a list of 543 measurements of 211 RAVE stars that were found in the literature and observed by other authors in their Ca II H and K lines (Table 3). It contains  $\log R'_{\text{HK}}$  and  $EW_{\text{IRT}}$  values that were compared in Figure 7.  $S$  indices from Baliunas et al. 1995 and Cincunegui et al. 2007 were converted to  $\log R'_{\text{HK}}$  using  $B - V$  values from the same catalogs, while colors for stars from Duncan et al. (1991), López-Santiago et al. (2010), and Table 5 from Gray et al. (2003) were obtained with the SIMBAD database. LogAge is the logarithm of age of the star (in years), where available (Wright et al. 2004; Isaacson & Fischer 2010; Arriagada 2011). If a star is found in catalogs multiple times, we list the average value (averaged over time or/and different catalogs). The other parameters are a subset of the entire list of parameters from the main catalog, as described in Appendix A.

## REFERENCES

- Andretta, V., Busà, I., Gomez, M. T., & Terranegra, L. 2005, *A&A*, 430, 669  
 Arriagada, P. 2011, *ApJ*, 734, 70  
 Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. 1995, *ApJ*, 438, 269  
 Cincunegui, C., Díaz, R. F., & Mauas, P. J. D. 2007, *A&A*, 469, 309  
 Conrad, C., Scholz, R.-D., & Kharchenko, N. 2013, *A&A*, submitted (arXiv:1309.4325)  
 Dempsey, R. C., Bopp, B. W., Henry, G. W., & Hall, D. S. 1993, *ApJS*, 86, 293  
 Duncan, D. K., Vaughan, A. H., Wilson, O. C., et al. 1991, *ApJS*, 76, 383

- Dupree, A. K., Whitney, B. A., & Pasquini, L. 1999, *ApJ*, **520**, 751
- Gray, R. O., Corbally, C. J., Garrison, R. F., McFadden, M. T., & Robinson, P. E. 2003, *AJ*, **126**, 2048
- Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, *AJ*, **132**, 161
- Hall, J. C. 1996, *PASP*, **108**, 313
- Hall, J. C. 2008, *LRSP*, **5**, 2
- Hall, J. C., Lockwood, G. W., & Skiff, B. A. 2007, *AJ*, **133**, 862
- Henden, A. A., Levine, S. E., Terrell, D., Smith, T. C., & Welch, D. 2012, *JAVSO*, **40**, 430
- Henry, T. J., Soderblom, D. R., Donahue, R. A., & Baliunas, S. L. 1996, *AJ*, **111**, 439
- Isaacson, H., & Fischer, D. 2010, *ApJ*, **725**, 875
- Jackson, R. J., & Jeffries, R. D. 2010, *MNRAS*, **407**, 465
- Jenkins, J. S., Jones, H. R. A., Tinney, C. G., et al. 2006, *MNRAS*, **372**, 163
- Jenkins, J. S., Murgas, F., Rojo, P., et al. 2011, *A&A*, **531**, A8
- Katz, D., Munari, U., Cropper, M., et al. 2004, *MNRAS*, **354**, 1223
- Knutson, H. A., Howard, A. W., & Isaacson, H. 2010, *ApJ*, **720**, 1569
- Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, arXiv:1309.4284
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011, *A&A*, **535**, A106
- Linsky, J. L., Hunten, D. M., Sowell, R., Glackin, D. L., & Kelch, W. L. 1979, *ApJS*, **41**, 481
- Lockwood, G. W., Skiff, B. A., Henry, G. W., et al. 2007, *ApJS*, **171**, 260
- López-Santiago, J., Montes, D., Gálvez-Ortiz, M. C., et al. 2010, *A&A*, **514**, A97
- Lyra, W., & Porto de Mello, G. F. 2005, *A&A*, **431**, 329
- Mallik, S. V. 1997, *A&AS*, **124**, 359
- Mamajek, E. E., & Hillenbrand, L. A. 2008, *ApJ*, **687**, 1264
- Marsden, S. C., Carter, B. D., & Donati, J.-F. 2009, *MNRAS*, **399**, 888
- Martínez-Arnáiz, R., Maldonado, J., Montes, D., Eiroa, C., & Montesinos, B. 2010, *A&A*, **520**, A79
- Matijević, G., Zwitter, T., Bienaymé, O., et al. 2011, *AJ*, **141**, 200
- Matijević, G., Zwitter, T., Bienaymé, O., et al. 2012, *ApJS*, **200**, 14
- Munari, U., Sordo, R., Castelli, F., & Zwitter, T. 2005, *A&A*, **442**, 1127
- Munari, U., Tomasella, L., Fiorucci, M., et al. 2008, *A&A*, **488**, 969
- Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., & Vaughan, A. H. 1984, *ApJ*, **279**, 763
- Pace, G. 2013, *A&A*, **551**, L8
- Rocha-Pinto, H. J., & Maciel, W. J. 1998, *MNRAS*, **298**, 332
- Roweis, S., & Saul, L. K. 2000, *Sci*, **190**, 2323
- Schröder, C., Reiners, A., & Schmitt, J. H. M. M. 2009, *A&A*, **493**, 1099
- Siebert, A., Williams, M. E. K., Siviero, A., et al. 2011, *AJ*, **141**, 187
- Soderblom, D. R. 2010, *ARA&A*, **48**, 581
- Soderblom, D. R., Stauffer, J. R., Hudon, J. D., & Jones, B. F. 1993, *ApJS*, **85**, 315
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, **132**, 1645
- Strassmeier, K., Washuettl, A., Granzer, T., Scheck, M., & Weber, M. 2000, *A&AS*, **142**, 275
- Strassmeier, K. G. 2009, *A&ARv*, **17**, 251
- Strassmeier, K. G., Rice, J. B., Ritter, A., et al. 2005, *A&A*, **440**, 1105
- Tinney, C. G., McCarthy, C., Jones, H. R. A., et al. 2002, *MNRAS*, **332**, 759
- Tomasella, L., Munari, U., & Zwitter, T. 2010, *AJ*, **140**, 1758
- Vanderplas, J., & Connolly, A. 2009, *AJ*, **138**, 1365
- Vaughan, A. H., Preston, G. W., & Wilson, O. C. 1978, *PASP*, **90**, 267
- White, R. J., Gabor, J. M., & Hillenbrand, L. A. 2007, *AJ*, **133**, 2524
- Wilkinson, M. I., Vallenari, A., Turon, C., et al. 2005, *MNRAS*, **359**, 1306
- Wilson, O. C. 1978, *ApJ*, **226**, 379
- Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, *PASP*, **123**, 412
- Wright, J. T., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2004, *ApJS*, **152**, 261
- Zhao, J. K., Oswalt, T. D., Rudkin, M., Zhao, G., & Chen, Y. Q. 2011, *AJ*, **141**, 107
- Zhao, J. K., Oswalt, T. D., Zhao, G., et al. 2013, *AJ*, **145**, 140
- Zucker, D. B., de Silva, G., Freeman, K., Bland-Hawthorn, J., & Hermes Team, 2012, in ASP Conf. Ser. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto (San Francisco, CA: ASP), 421
- Zwitter, T., Siebert, A., Munari, U., et al. 2008, *AJ*, **136**, 421