

# Distance determination for RAVE stars using stellar models

M.A. Breddels

supervisors: A.Helmi, M.C. Smith

November 16, 2007

## Abstract

The Radial Velocity Experiment (RAVE) is an ongoing project which will measure radial velocities and stellar atmosphere parameters (temperature, metallicity, surface gravity and rotational velocity) of up to one million stars. Since its start in 2003 until the internal data release of June 2007,  $\sim 200\,000$  observations have been made, from which 50 994 spectra have been reduced, resulting in 21 032 observations having astrophysical parameters.

This new dataset will be interesting for Galaxy structure studies. However, there is a crucial piece of information missing: The distance to the stars. When combined with radial velocities from the RAVE survey, proper motions (when available) from Starnet2, Tycho2 and UCAC2 catalogues this dataset will provide the full 6d phase space for each star.

In this report, we present a method to derive the distance to a star using stellar models and astrophysical parameters such as effective temperature ( $T_{eff}$ ), surface gravity ( $\log(g)$ ) and metallicity ( $[Fe/H]$ ) as provided by RAVE. This method is tested with artificial data and a subset of the RAVE data for which distances are known, yielding consistent results. When we apply our method to a set of 17 434 stars for which we are able to derive their distance for the first time, there are 4 048 stars with relative distance errors  $< 28\%$ , 6 921 and 12 412 whose relative errors are smaller than 37% and 46% respectively.

Using the obtained distances in combination with sky coordinates, radial and proper motions we are able to derive the full 6d phase space for these stars. We recover the known gradients in the metallicity of stars as we move away from the Galactic plane. Velocity for stars in the Solar neighbourhood are analysed, signatures of a non rotating halo is found and the UV velocity ellipsoid shows the well known vertex deviation.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Stellar evolution . . . . .	5
2.2	Statistics . . . . .	10
<b>3</b>	<b>Analysis</b>	<b>14</b>
3.1	Description of the method . . . . .	14
3.2	Testing the method . . . . .	15
3.3	Testing the method on GC-survey stars observed by RAVE .	17
3.4	Distances for stars in the RAVE dataset . . . . .	18
3.5	6d phase space coordinates for stars in the RAVE dataset . .	18
3.6	Calculation time . . . . .	23
<b>4</b>	<b>Results</b>	<b>24</b>
<b>5</b>	<b>Discussion and conclusion</b>	<b>30</b>
<b>A</b>	<b>Description of RAVE catalogue with phase space coordinates</b>	<b>31</b>
<b>B</b>	<b>Software manual</b>	<b>33</b>

## 1 Introduction

During the past few decades, cosmology has changed from philosophy to precision physics. We now believe to have a good understanding of the universe we live in. This is a flat universe, dominated by dark energy which causes the expansion to accelerate (Spergel et al., 2007). We further believe that most matter is in the form of cold dark matter, cold referring to non-relativistic. Zwicky (1937) was the first to introduce dark matter to explain the missing mass in nebulae. By this definition dark matter does not interact with normal matter through the electro-magnetic force, but does interact through the gravitational force. This popular and successful cosmological picture is referred to as  $\Lambda$ CDM cosmology and predicts a hierarchical formation of structure in the Universe. In this framework, small structures grow first, to later form bigger structures. The larger structures we see today are the filaments and super clusters as observed by galaxy surveys like the Sloan Digital Sky Survey (SDSS). At smaller scales we find galaxies, ellipticals, spirals, and finally dwarf galaxies.

The power spectrum is used to measure the power of density perturbations on a certain scale. The current cosmological model predicts a Harrison-Zel'dovich spectrum ( $P(k) \propto k$ , where  $k$  is the wavenumber). This power spectrum predicts more power on small scale, so there should be more dwarf galaxies than ellipticals and spirals. In this cosmological model it is likely that many of these dwarf galaxies have merged to form larger galaxies. If these dwarf galaxies have merged with the Milky Way, how would we be able to detect this? We see evidence of an ongoing merger between our Galaxy and the Sagittarius stream (Majewski et al., 2003). However, we might also be able to detect already merged galaxies using the kinematics of stars in the Milky Way (Helmi et al., 2006).

From numerical simulations we have learnt that stars from a common progenitor, when merged with the Milky Way, move on similar orbits (Helmi and de Zeeuw, 2000). Helmi et al. (2006) used the apocenter (A), pericenter (P) and the  $z$ -angular momentum ( $L_z$ ) to find substructure, which they refer to as APL space. To construct the APL space the full 6d phase space is needed, 3d positions and 3d velocities. Helmi et al. (2006) used the Geneva-Copenhagen survey (GC-survey) (Nordström et al., 2004), which has around 14 000 stars with full phase space information. The sample is magnitude complete to  $V = 7.7$ , i.e. it contains very bright stars.

In the future the GAIA satellite (Perryman et al., 2001) is expected to observe  $10^9$  stars of the Milky Way, which is 1% of the total of  $10^{11}$ . One of the objectives of the GAIA mission is to create a 6d phase space map of stars in our Galaxy. This enormous dataset will be highly valuable to learn more about the formation of our own Galaxy. GAIA will measure multi-colour photometry and spectroscopy with a limiting magnitude of  $V = 20$ . Trigonometric parallax and proper motions will also be measured, and

combined with the radial velocities from spectra will then give us 6d phase space coordinates of these stars. The mission will start in 2011 and last for about 5 years.

In the meantime the Radial Velocity Experiment (RAVE) is measuring radial velocities and stellar atmosphere parameters (temperature, metallicity, surface gravity and rotational velocity) of up to one million stars. Spectra are taken using the 6dF spectrograph on the 1.2m UK Schmidt Telescope of the Anglo-Australian Observatory. Observed stars are drawn from the Tycho-2 and SuperCOSMOS catalogues in the magnitude range  $9 < I < 12$ . Since the start in 2003 until the internal data release of June 2007,  $\sim 200\,000$  observations are made, from which 50 994 spectra are reduced. This new dataset will be interesting for Galaxy structure studies. However, there is a crucial piece of information missing: The distance to the stars. When combined with radial velocities from the RAVE survey, proper motions (when available) from Starnet2, Tycho2 and UCAC2 catalogues, this dataset will provide the full 6d phase space coordinates for each star.

A common way to measure the distance to a star is the trigonometric parallax method. A different method, also used by Nordström et al. (2004) for the GC-survey, is distance determination by photometric parallax. Main sequence stars show a tight correlation between their colour and absolute magnitude. Therefore an empirical calibration can be used to derive a relation between photometric colours, metallicity and absolute magnitude (Crawford, 1975).

In this report, we present a method to obtain the distance to a star using stellar models (isochrones). We infer the probability of an absolute magnitude given its effective temperature ( $T_{eff}$ ), surface gravity ( $\log(g)$ ) and metallicity ( $[Fe/H]$ ). With some caution<sup>1</sup> also colour indices can be used. The distance to a star can then be calculated using the distance modulus.

In section 2, we present a general introduction. We will discuss the connection between stellar evolution theory, stellar tracks and isochrones to gain insight in these topics before presenting the statistical methods for the distance determination. Section 3 will test the method by using synthetic data and a subset of the RAVE data for which distances are known. The application of the method to the RAVE dataset will also be discussed. Results about the velocity and metallicity distributions will be presented in section 4.

---

<sup>1</sup>Extinction can modify colour indices, this can be corrected for, or infra-red (IR) bands can be used.

## 2 Background

### 2.1 Stellar evolution

Stellar evolution is relatively well understood. The evolution of a star is determined by its mass, initial chemical composition and a set of equations, such as hydrostatic equilibrium, mass continuity, conservation of energy... (Salaris and Cassisi, 2005). Given these initial conditions, the evolution of a star in time is fully determined by these equations. Since they have no solution in an analytic form, the equations have to be solved using numerical techniques. This means that the stellar evolutionary tracks are usually released as tables of solutions. For instance the Yonsei-Yale ( $Y^2$ ) evolutionary stellar tracks are calculated for different alpha-enhancement ( $[\alpha/Fe]$ ), metallicity ( $[Fe/H]$  or  $Z$ ) and mass ( $m$ ). For certain values of  $[\alpha/Fe]$ ,  $Z$  and  $m$ , the evolution from pre-main-sequence till the helium core flash is tabulated. The tabulated data for the  $Y^2$  stellar tracks are listed in table 1. Other properties such as the radius and surface gravity can also be calculated. The radius  $R$  of a star can be found using:

$$L = 4\pi R^2 \sigma T_{eff}^4, \quad (1)$$

where  $L$  is the luminosity,  $\sigma$  the Stephan-Boltzmann constant,  $T_{eff}$  the effective temperature. Using the radius, we find the surface gravity:

$$g = \frac{Gm}{R^2}, \quad (2)$$

where  $G$  is the gravitational constant and  $m$  the mass of the star. It is common to use  $\log(g)$  instead of  $g$  in literature, where  $g$  is in cgs units.

column	explanation
Age	in Gyr, the free variable
$\log(T_{eff})$	logarithm of effective temperature in Kelvin
$\log(L/L_{sun})$	logarithm of luminosity over solar luminosity
$Y_{core}$	mass fraction of helium in the core
$M_{core}$	mass of the core

**Table 1:** Tabulated values for the  $Y^2$  stellar tracks.

In astronomy, the mass fraction of hydrogen, helium and all heavier elements (usually called metals) are denoted by  $X$ ,  $Y$  and  $Z$  respectively, where  $X + Y + Z = 1$ , such that given two fractions, the other can be calculated. Since iron lines are easy to measure from a spectra, the traditional metal abundance indicator is the iron abundance, and is defined as:

$$[Fe/H] \equiv \log \left( \frac{N(Fe)}{N(H)} \right)_* - \log \left( \frac{N(Fe)}{N(H)} \right)_\odot, \quad (3)$$

where  $n(E)$  is the number of atoms of element  $E$  in the star and the subscript  $\odot$  refers to our Sun. The units are referred to as dex. Rather than the number of atoms it is also possible to use mass fractions in Eq. (3). If solar abundance is assumed, the iron abundance can also be expressed in the mass fraction  $Z$  (all metals):

$$[Fe/H] = \log \left( \frac{Z}{X} \right)_* - \log \left( \frac{Z}{X} \right)_{\odot}, \quad (4)$$

If the solar abundance is not assumed, the right hand side of Eq. (4) gives the abundance of all metals to hydrogen, and write is as:

$$[M/H] = \log \left( \frac{Z}{X} \right)_* - \log \left( \frac{Z}{X} \right)_{\odot}, \quad (5)$$

where  $M$  refers to all metals. For stars having non solar abundances distribution, other ratios can be interesting, and the abundance of certain metals with respect to others is expressed as in Eq. (3). For example calcium ( $Ca$ ) to iron ( $Fe$ ):

$$[Ca/Fe] \equiv \log \left( \frac{N(Ca)}{N(Fe)} \right)_* - \log \left( \frac{N(Ca)}{N(Fe)} \right)_{\odot}. \quad (6)$$

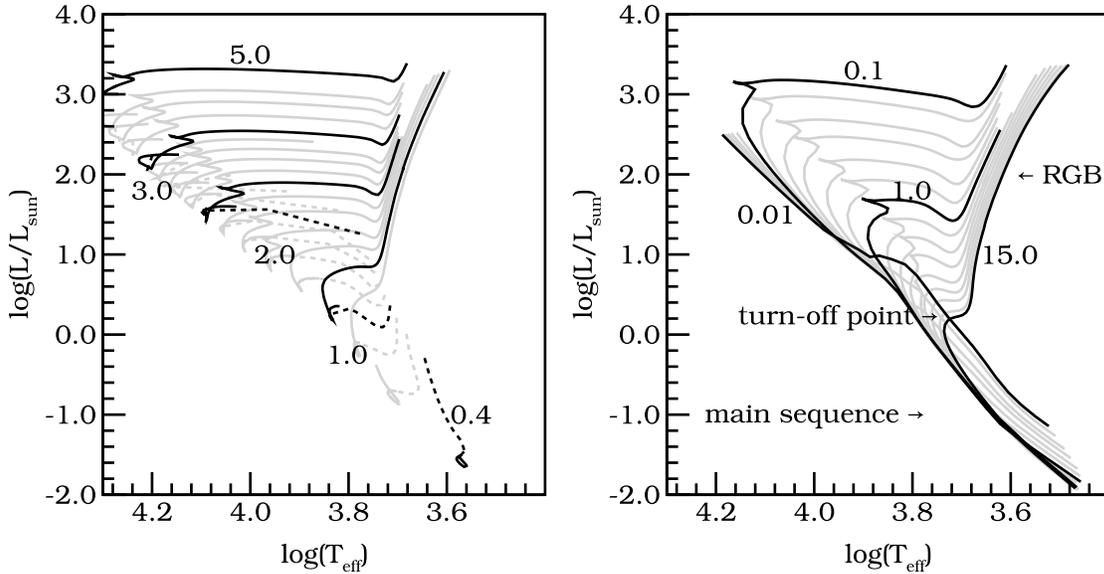
The abundance of the  $\alpha$  elements ( $O$ ,  $Ne$ ,  $Mg$ ,  $Si$ ,  $S$ ,  $Ca$  and  $Ti$ ) compared to iron ( $Fe$ ) is denoted as  $[\alpha/Fe]$ . When  $[\alpha/Fe] > 0$ , a star is said to be ‘alpha-enhanced’. The metal and iron abundance are approximately related by (Salaris and Cassisi, 2005):

$$[M/H] \approx [Fe/H] + \log(10^{[\alpha/Fe]} 0.694 + 0.306). \quad (7)$$

The difference in  $[\alpha/Fe]$  is often attributed to the different contribution of Type Ia and Type II supernovae. Type Ia inject more iron elements into the inter stellar medium (ISM) compared to Type II, while Type II explosions are typically more rich in  $\alpha$  elements.

Colour indices from stars are easy to measure and therefore we would also like to have this in the stellar models. If stars behave like a blackbody, the theoretical absolute magnitude in different bands can be calculated from the effective temperature and response function (‘shape’) of a colour filter since the shape of the spectrum is completely determined by the Planck curve. From these absolute magnitudes, colour indices can be constructed, which is defined as the difference between the magnitudes in two different bands.

Stars do not behave like blackbodies, therefore theoretical spectra have to be calculated. From the chemical composition, the surface gravity and the temperature, different absorption and emission lines can be modelled to construct a complete spectrum. Although the  $Y^2$  stellar tracks do not



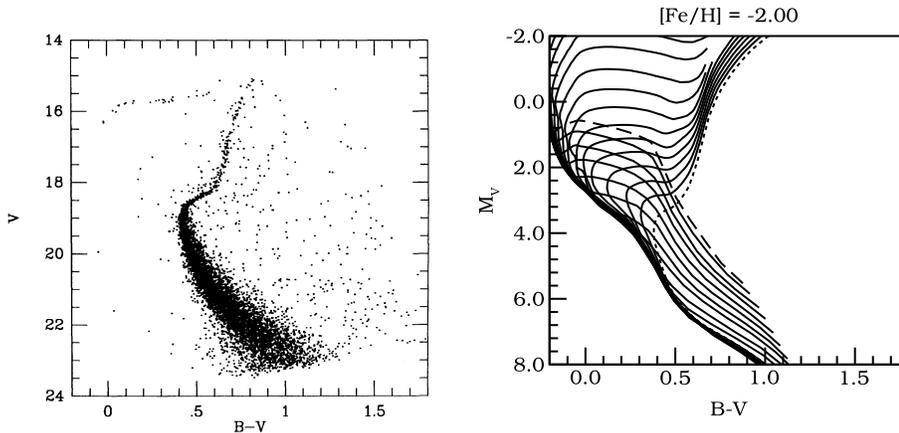
**Figure 1:** **left:**  $Y^2$  stellar tracks for 0.4, 1.0, 2.0, 3.0 and 5.0  $M_{\odot}$  (black), as indicated by the labels. Pre main sequence tracks are shown dotted **right:**  $Y^2$  isochrones for 0.01, 0.1, 1.0 and 15 Gyr (black) as indicated by the labels. For stars on the 0.01 Gyr isochrone, all stars are still on the main sequence. For older isochrones we can clearly see the turnoff points. On the red giant branch (RGB) the bright and relative cold giants can be found.

include colour indices, they are available for the isochrones. To compute the colour indices, precalculated tables are used which are also correlated to observations (Yi et al., 2001).

Figure 1 shows  $\log(L/L_{\odot})$  versus  $\log(T_{\text{eff}})$  and is called the Hertzsprung-Russell (HR) Diagram. This diagram is comparable to plots of absolute magnitude (e.g.  $M_V$ ) versus colour (e.g.  $B - V$ ), since effective temperature and colour are correlated, and the luminosity is related to the magnitude. A diagram showing apparent magnitude (e.g.  $V$ ) versus colour (e.g.  $B - V$ ) is referred to as a Colour Magnitude Diagram (CMD).

The left panel of Fig. 1 shows stellar tracks with masses ranging from  $m = 0.4M_{\odot} - 5M_{\odot}$ . In the upper left region we find the high mass stars as indicated by the labels. A high mass star of  $m = 5M_{\odot}$  arrives at the helium flash at an age of  $\tau \approx 0.06\text{Gyr}$  while a low mass star of  $m = 0.4M_{\odot}$  will start burning helium at an age of  $\tau \approx 118\text{Gyr}$ .

The stellar tracks form the basis of isochrones. Isochrones (meaning ‘same age’), are similar to stellar tracks. But instead of keeping mass constant, and age as a free variable, age is kept constant and mass is a free variable. To generate an isochrone of a certain age, the stellar tracks from different masses are interpolated such that lines of constant age are gener-



**Figure 2:** **left:** Colour Magnitude Diagram (CMD) of M68 (from Walker, 1994) **right:** CMD with theoretical  $Y^2$  isochrones for similar metallicity.

ated. The right panel Fig. 1 shows isochrones of similar chemical composition and can be compared to the stellar tracks shown on the left panel. All stars on the 0.01 Gyr isochrone are still on the main sequence. For older isochrones we can clearly see the main sequence turnoff points, the point at which the star stops burning hydrogen in its core. After this stage, the star expands and cools down proceeding onto the red giant branch (RGB) sequence.

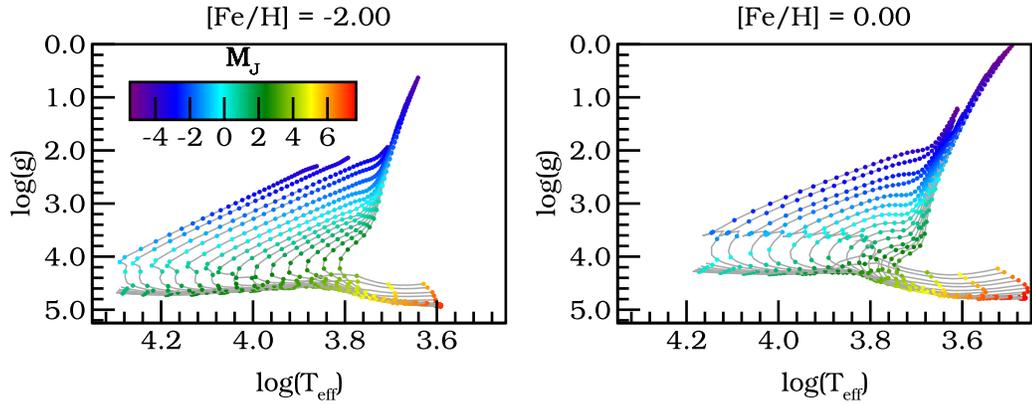
In the left panel of Fig. 2, we see a CMD of the globular cluster Messier 68 (M68), plotted as colour index  $B - V$  against apparent magnitude  $V$ . We can clearly see that there is one dominant stellar population. This is usually referred to as a single (or simple) stellar population (SSP). When we compare the isochrones in the right panel of Fig. 2 to M68, we can already see this is a very old population.

In the plot with theoretical isochrones, the absolute magnitude is plotted, and in the data from M68 the apparent magnitude. The difference between these magnitudes, the distance modulus ( $\mu$ ), is due to distance:

$$\mu = m - M = 5 \log(d) - 5, \quad (8)$$

where  $m$  is the apparent magnitude,  $M$  the absolute magnitude, and  $d$  the distance in parsec. In practise the effect of extinction complicates matters since interstellar matter can absorb part of the light, and an extra term must be added to Eq. (8).

The stellar tracks and isochrones can also be seen (in a mathematical sense) as a function ( $\mathcal{F}$ ) of alpha-enhancement ( $[\alpha/Fe]$ ), metallicity ( $Z$ ), mass ( $m$ ) and age ( $\tau$ ) that maps to absolute magnitude ( $M_\lambda$ ), surface gravity ( $\log(g)$ ), effective temperature ( $T_{eff}$ ), and colours:



**Figure 3:**  $\log(g)$  versus  $\log(T_{eff})$  plot for isochrones from 0.01-15 Gyr spaced logarithmically,  $[\alpha/Fe] = 0$ . Colour indicates the absolute magnitude in the J band. **left:** Isochrones for  $[Fe/H] = -2.50$  dex. **right:** Isochrones for  $[Fe/H] = 0.00$  dex. Overall shape of the isochrones is quite similar for both metallicities.

$$\begin{aligned} \mathcal{F}([\alpha/Fe], Z, \tau, m) &\mapsto (M_\lambda, \log(g), T_{eff}, colours\dots), \\ \mathcal{I}(m) = \mathcal{F}(m)|_{[\alpha/Fe], Z, \tau} &\mapsto (M_\lambda, \log(g), T_{eff}, colours\dots). \end{aligned} \quad (9)$$

For example, the isochrones are a function  $\mathcal{I}(m)$  of mass, which is obtained from  $\mathcal{F}$  by keeping all other variables constant.

Assuming solar  $\alpha$ -enhancement,  $[\alpha/Fe] = 0.0$ , we define the function  $\mathcal{S}(Z, \tau, m)$  as  $\mathcal{F}$ , keeping  $[\alpha/Fe]$  constant

$$\mathcal{S}(Z, \tau, m) = \mathcal{F}(Z, \tau, m)|_{[\alpha/Fe]} \mapsto (M_\lambda, \log(g), T_{eff}, colours\dots). \quad (10)$$

Therefore the isochrones or stellar tracks from  $Y^2$  can be seen as a samples from the theoretical stars defined by  $\mathcal{S}(Z, \tau, m)$ . A sample of these theoretical ‘model stars’ are shown in Fig. 3 for two different metallicities. Each model star is represented as a dot, and the connecting lines correspond to the isochrones of different ages.

The metallicity, surface gravity, and effective temperature of a star can be obtained from a spectrum and its colours can be measured photometrically. Extinction can affect colour, but the near-IR and IR colours used by the RAVE survey are less affected. If the function  $\mathcal{S}$  at fixed  $Z$  from age and mass to e.g.  $\log(g)$  and  $T_{eff}$  is injective<sup>2</sup>, we can invert it, and determine  $M_\lambda$ . But the function is not injective, meaning that for a given  $\log(g)$  and  $T_{eff}$  there is not a single mass and age. This overlap can be seen best on the right of Fig. 3 around  $\log(T_{eff}) = 3.8, \log(g) = 4$ , where the isochrones

<sup>2</sup>Also referred to as one-to-one.

can be seen to overlap. However in a statistical sense, it is still possible to infer the probability of a mass, age, and therefore absolute magnitude.

Figure 4 shows the same isochrones as Fig. 3, illustrating the relation between  $M_J$  and  $T_{eff}$ , and between  $M_J$  and  $\log(g)$  separately. From this figure we can see how errors in  $\log(g)$  and  $T_{eff}$  affects the uncertainty of the absolute magnitude ( $M_J$  in this example). The middle row in Fig. 4 shows that the value of  $M_J$  is better defined by  $\log(g)$  for RGB (top right) than for main sequence (bottom left) stars, independent of metallicity. On the other hand, the bottom row of Fig. 4 shows that  $T_{eff}$  essentially determines  $M_J$  for main sequence stars, again independent of metallicity. We therefore expect that a small error in  $\log(g)$  will give better absolute magnitude estimates for RGB stars, while a small error in  $T_{eff}$  will have a similar effect on main sequence stars. We also expect this to be not strongly dependent on metallicity.

## 2.2 Statistics

Descriptive statistics is the mathematical tool we have for summarising all the data we collect in a few numbers, which we may call a statistic. Instead of working with images, it is easier to work with quantitative data such as the mean magnitude of a star and its error. Instead of a full spectrum, we would rather know the stars physical parameters such as  $[Fe/H]$  and  $[\alpha/Fe]$ . But we also use statistics to draw conclusions from the data, known as inferential statistics, or inference for short. For example, having statistics about the weather and knowing that dark clouds cover the sky, we want to infer the probability that it will rain or not.

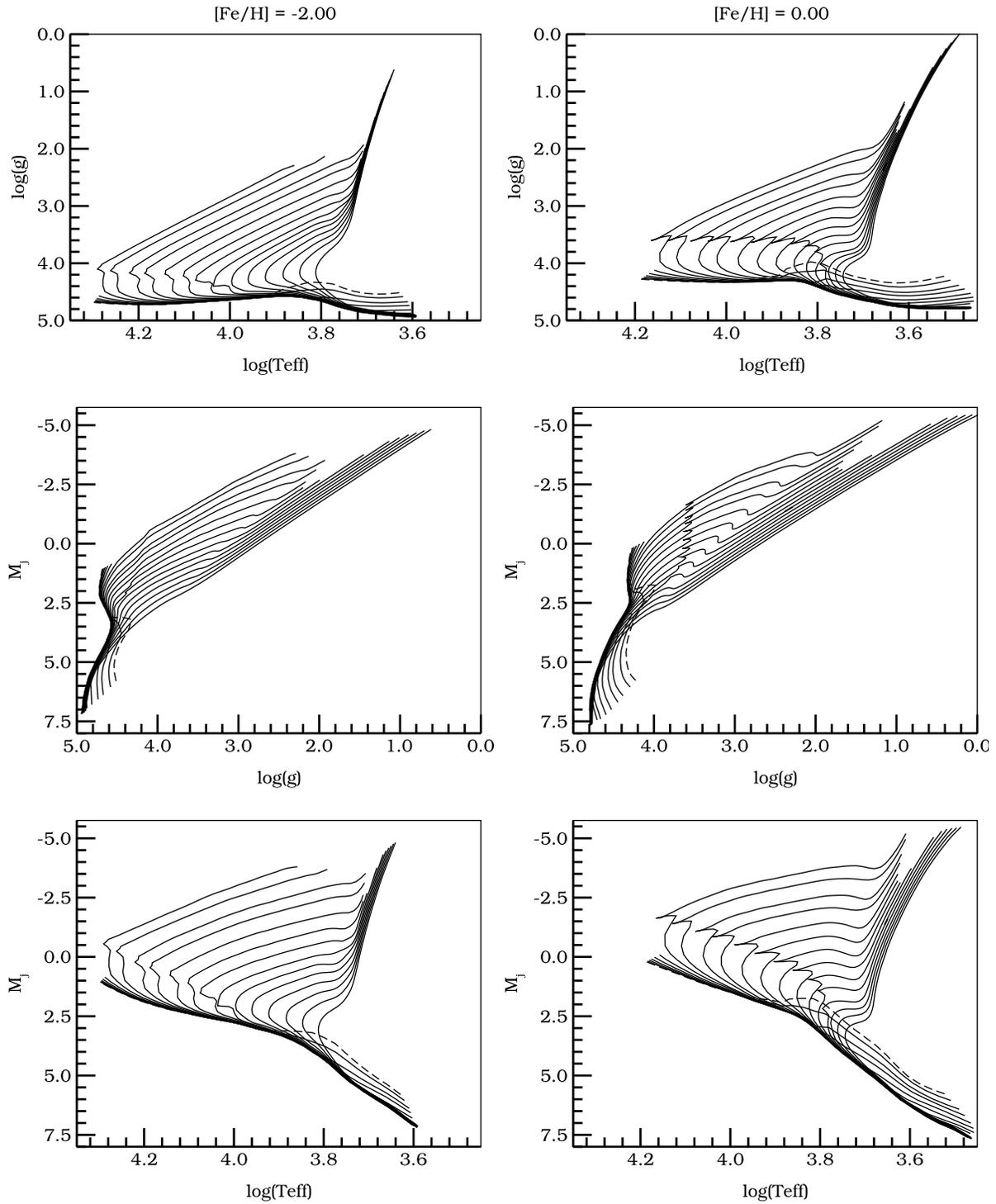
Let us suppose we observe a star with a telescope and obtain its spectrum. From this, we derive  $T_{eff}$ ,  $\log(g)$  and  $[Fe/H]$  with an uncertainty because our measurements have errors, which we assume are Gaussian. We search for the most likely star, by requiring that

$$\chi_{model}^2 = \frac{(T_{eff} - T_{eff,model})^2}{2\sigma_{T_{eff}}^2} + \frac{(\log(g) - \log(g)_{model})^2}{2\sigma_{\log(g)}^2} + \frac{([Fe/H] - [Fe/H]_{model})^2}{2\sigma_{[Fe/H]}^2}, \quad (11)$$

is a minimum. The subscript *model* refers to values from the model star obtained from the isochrones. Using our most likely model star and the measurement errors, we can construct a probability distribution function (pdf) which describes the probability of measuring  $T_{eff}$ ,  $\log(g)$  and  $[Fe/H]$ :

$$P(T_{eff}, \log(g), [Fe/H] | \overline{T_{eff}}, \overline{\log(g)}, \overline{[Fe/H]}) = \frac{1}{2\pi^{3/2}\sigma_{T_{eff}}\sigma_{\log(g)}\sigma_{[Fe/H]}} e^{-\chi^2}, \quad (12)$$

$$\chi^2 = \frac{(T_{eff} - \overline{T_{eff}})^2}{2\sigma_{T_{eff}}^2} + \frac{(\log(g) - \overline{\log(g)})^2}{2\sigma_{\log(g)}^2} + \frac{([Fe/H] - \overline{[Fe/H]})^2}{2\sigma_{[Fe/H]}^2}, \quad (13)$$

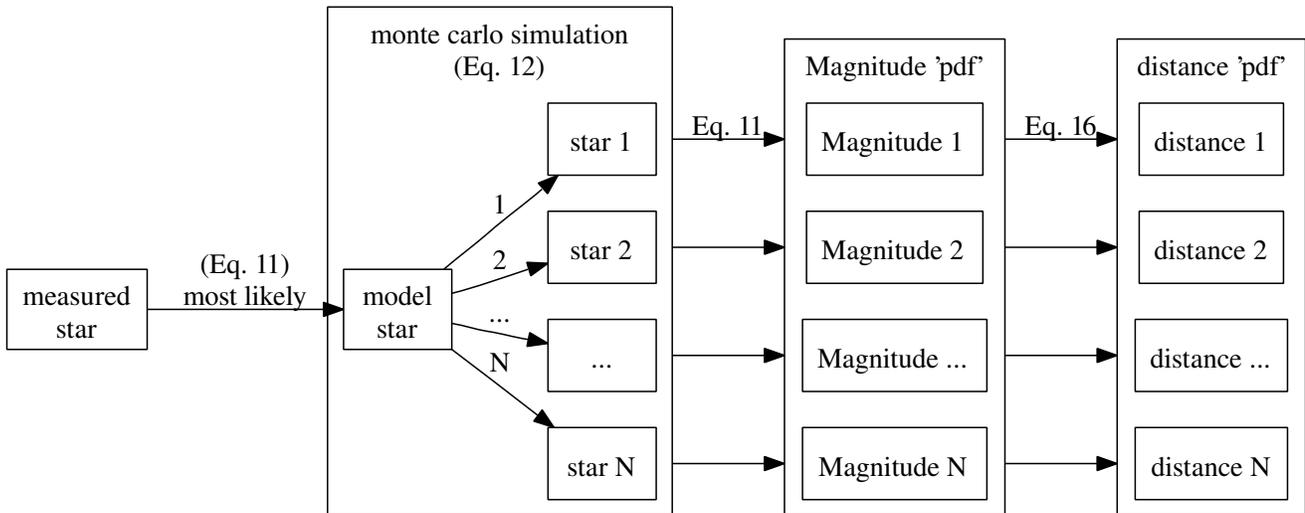


**Figure 4:** Isochrones for  $[\alpha/Fe] = 0$ ,  $[Fe/H] = -2.50$  (left column) and  $[Fe/H] = 0.00$  (right column), ages ranging from 0.01-15 Gyr spaced logarithmically. Dashed line indicates the youngest (0.01 Gyr) isochrone. **top row:** Similar to figure 3, shown for completeness. **middle row:**  $M_J$  is best restricted by  $\log(g)$  for RGB stars (top right). **bottom row:**  $M_J$  is best restricted by  $T_{eff}$  for main sequence stars (bottom right).

where the overline  $\overline{T_{eff}}$  indicates the values from our most likely model star. Observing the mean values has the highest probability, and decreases with the distance from the mean value, as measured by  $\chi^2$ . In a hypothetical situation where we know what our model star is, we could simulate the observation by drawing random samples from the pdf defined by Eq. (12). This is the basic idea behind of Monte Carlo simulations.

We want to find the pdf of  $M_J$ , so we can calculate statistics from it such as the mean value and the variance. We therefore simulate many observations which we draw from the pdf defined in Eq. (12) assuming our most likely model star is the correct one. From these simulated observations, we again find the best matching model star from the isochrones (see Fig. 3), which is the model for which the distance measured by Eq. (11) is a minimum. This model star provides us with the corresponding value of  $M_J$ . The recipe for the method is pictured in Fig. 5, and a description can also be found in Press 1988, chapter 15.6. When many simulations are done, the histogram of  $M_J$  values obtained using this method will give us our best estimate of its pdf. Note that the absolute magnitude of the best matching model star will be the most likely, which will not necessarily be equal to the mean absolute magnitude like in the case of a Gaussian pdf.

For the RAVE data,  $T_{eff}$  is determined only from the spectra, and not photometrically. The  $T_{eff}$  and a colour index are therefore uncorrelated in the sense that they are independently drawn. Both  $T_{eff}$  and one colour index can be used for the Monte Carlo simulation. If multiple colour indices are used, using the same band, samples should be properly drawn from a correlated multivariate distribution (Wall and Jenkins, 2003, chapter 6.5).



**Figure 5:** Overview of the Monte Carlo method used to derive absolute magnitudes and distances for the RAVE stars. The first step is to find the most likely model star which will be used as the true model. From this model stars,  $N$  samples are created by applying Gaussian noise to the observables. For each sample we will find the best matching model star, giving us estimates for the absolute magnitude. The  $N$  samples are our best estimate for the pdf of the absolute magnitude. We can calculate statistics from this pdf, such as the mean and variance, and we can also use the samples to do further calculations such as the distance.

### 3 Analysis

In this section we will first describe the exact method that we will use to calculate distances for the RAVE data. We will test its performance and dependence on the magnitude of the errors on our observables in §3.2, with emphasis on errors typical for the RAVE data. We will also test the method on a subset of the RAVE data for which distances have been determined in the GC-survey in §3.3. Finally, in §3.4, the method is applied to the full RAVE dataset.

#### 3.1 Description of the method

The RAVE dataset provides  $[M/H]$ , J and K band magnitudes,  $\log(g)$  and  $T_{eff}$ . These measured quantities can be used for the isochrone fitting routine to derive the absolute magnitude for each star. These bands, J ( $1.25 \pm 0.12 \mu\text{m}$ ) and K ( $2.22 \pm 0.22 \mu\text{m}$ ), are in the infrared (IR) and are less affected by dust than visual bands. The J and K band magnitudes in the RAVE database come from the DENIS and 2MASS survey. The magnitudes used for the fitting routine are the weighted average of the available magnitudes, where the weight is the reciprocal of the measurement error:

$$X_{weighted} = \frac{\sum_j w_j X_j}{\sum_j w_j}, \quad (14)$$

where  $X_j$  are the measured values and  $w_j = 1/\sigma_j^2$  the corresponding weight. The error in the average is calculated as:

$$\sigma_{weighted}^2 = \frac{1}{\sum_j 1/\sigma_j^2}. \quad (15)$$

Errors in  $[M/H]$ ,  $T_{eff}$  and  $\log(g)$  are estimated at 0.25 dex, 300K and 0.3 dex respectively (Zwitter, private comm.), better error estimates are under development. The current internal data release of June 2007 has two metal abundances, one uncalibrated, determined from the spectra alone ( $[m/H]$ ), and a calibrated value ( $[M/H]$ ). The latter is calibrated using a subset of stars with accurate metallicity estimates and is the one that is used for the fitting routine. We only use the isochrones with  $[\alpha/Fe] = 0$  because the RAVE measurements of  $[\alpha/Fe]$  are not accurate enough. We also assume  $[Fe/H] = [M/H]$  which will be valid only when  $[\alpha/Fe] = 0$  (Eq 7). To estimate the absolute magnitude, the most likely model star is found first, as described in 2.2, calculating the minimum  $\chi_{model}^2$  using  $[Fe/H]$ ,  $\log(g)$ ,  $T_{eff}$  and  $J - K$  with errors as described above. From the most likely star, 5000 samples are generated by convolving these most likely values with Gaussian errors as described in §2.2. For every generated sample, the nearest model star as measured by the  $\chi_{model}^2$  is found. This gives 5000 samples which is

our representation of the pdf for the absolute magnitude. The sample mean and standard deviation are calculated and we will use these values for the analysis in the following sections.

The isochrones used for the fitting routine are the  $Y^2$  isochrones. The isochrones can be downloaded from the  $Y^2$  website<sup>3</sup>, where also an interpolation routine is available, called `YYmix2`. We generated a new set of 40 isochrones using the `YYmix2` code with ages between 0.01 and 15.0 Gyr, spaced logarithmically, for every  $[Fe/H]$  between -2.5 and 1.0 dex, with 0.25 dex separation (corresponding to 1 sigma in  $[M/H]$ ). This resulted in 13 files, for every metallicity one file, each containing 40 isochrones. The separation between the points of the isochrones has been visually inspected, and is in general smaller than the errors in  $T_{eff}$  and  $\log(g)$ .

### 3.2 Testing the method

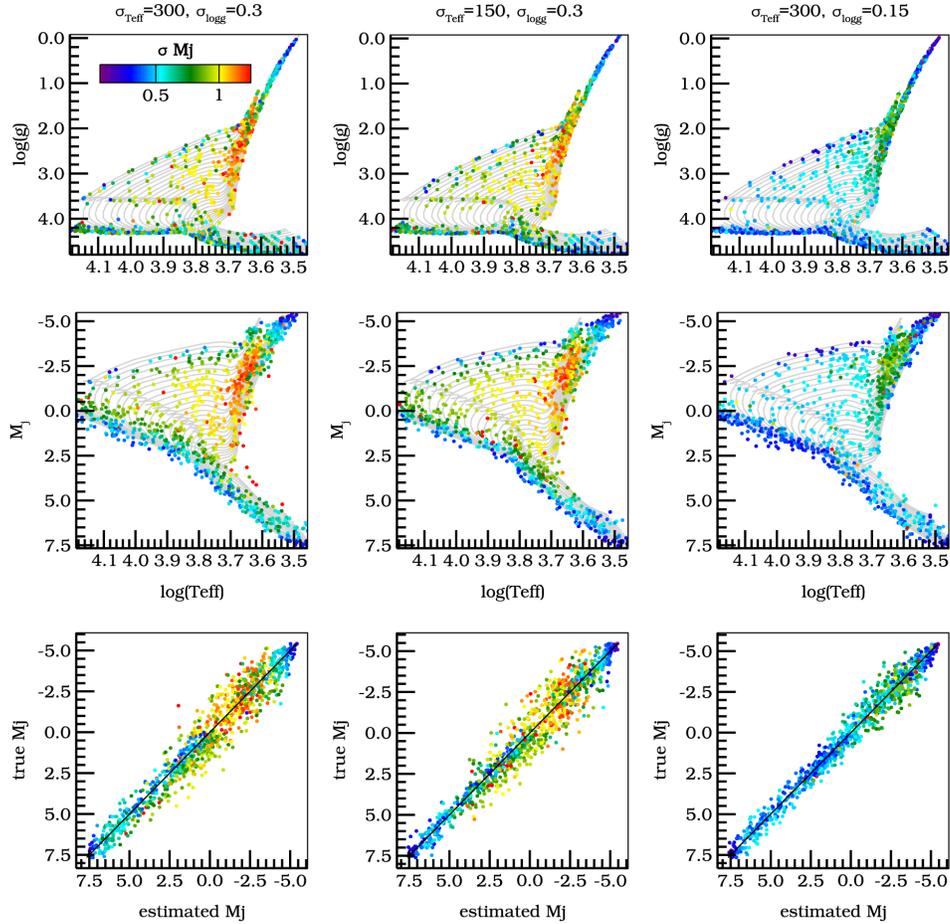
To test the method, we take a sample of 1000 random model stars. This set is large enough, that one can be confident the method is working and to determine for what kind of stars the model works best. We draw this set from one metallicity, randomly uniform in  $\log(\tau)$  and randomly uniform in mass index. Now we convolve  $[Fe/H]$ ,  $T_{eff}$ ,  $\log(g)$  and the colour indices with a Gaussian corresponding to the error in the RAVE survey to mimic our measurements. We run the method on this set of 1000 stars and analyse its results.

The reason for choosing a fixed metallicity is twofold. In §2.1 we have seen that different metallicities should give similar results in terms of the accuracy with which the absolute magnitude can be derived. It should therefore be sufficient to only analyse the method for one metallicity. Secondly, it also means that the results only have to be compared to one set of isochrones, making it easier to analyse the figures. Note that although one metallicity is used to generate the sample, after error convolution, isochrones for all metallicities are used for the fitting routine.

In the left column of Fig. 6, we plotted the results for the 1000 random model stars. The colours indicate the errors on  $M_J$  and are clipped to a value  $\sigma_{M_J} = 1.25$ . The middle row shows the results on an HR diagram. Stars on the main sequence and on the RGB appear to have the smallest errors as expected (see §2.1). In the bottom row, the estimated magnitude is plotted against the input magnitude of the model star from which the estimate was derived, showing the deviation from the real absolute magnitude grows with  $\sigma_{M_J}$ , as expected. The method appears to be give sane results, showing no systematic errors.

We now run the method again, decreasing the error in  $T_{eff}$  to 150 Kelvin. The results are shown in the middle column in Fig. 6. The errors in  $M_J$  do

<sup>3</sup><http://www-astro.physics.ox.ac.uk/~yi/yyiso.html>



**Figure 6:** Effect of the errors of  $\log(g)$  and  $T_{\text{eff}}$  on the estimated absolute magnitude  $M_J$ . The main sequence and RGB stars perform best. Reducing the errors in  $\log(g)$  has the largest effect. **Left column:** Errors similar to the RAVE dataset,  $\sigma_{T_{\text{eff}}} = 300\text{K}$  and  $\sigma_{\log(g)} = 0.3$ . **Middle column:** Reducing the errors in effective temperature,  $\sigma_{T_{\text{eff}}} = 150\text{K}$ . **Right column:** Reducing the errors in surface gravity,  $\sigma_{\log(g)} = 0.15$ . **Top row:** The random sample of 1000 stars, with colours indicating errors, clipped to a value of  $\sigma_{M_J} = 1.25$ . **Middle row:** HR Diagram with colours indicating the same errors as the bottom row. **Bottom row:** Estimated versus real (model) absolute magnitude ( $M_J$ ). The line indicates the when they are equal. Colours indicates the same errors as the top row. The error between the real and estimated absolute magnitude grows as the variance grows (colour towards red).

not seem to have changed much, except for the main sequence stars (slightly more blue and purple). Instead of decreasing the error in  $T_{eff}$ , we decrease the error in  $\log(g)$  to 0.15 dex and run the method again. The results are shown in the right column in Fig. 6 and shows that the accuracy with which we can determine  $M_J$  has increased significantly. Therefore reducing the error in  $\log(g)$  will result in significant improvements in the estimate of the absolute magnitude.

We can now apply Eq. (8) to find the distance  $d$  to a star:

$$d = 10^{(J-M_J+5)/5} = 10^{(\mu+5)/5}, \quad (16)$$

where  $d$  is in parsec. Using error propagation, we can estimate the error in the distance:

$$\sigma_d^2 = \left( \frac{\partial d(\mu)}{\partial \mu} \right)^2 \sigma_\mu^2 = \left( \frac{\ln(10)}{5} d \sigma_\mu \right)^2, \quad (17)$$

$$\frac{\sigma_d}{d} = \frac{\ln(10)}{5} \sigma_\mu \approx 0.46 \sigma_\mu = 0.46 \sqrt{\sigma_J^2 + \sigma_{M_J}^2}, \quad (18)$$

where we find that an error of 1 magnitude in the distance modulus corresponds to a 46% error in distance. If the error in apparent magnitude is negligible compared to the error in absolute magnitude, we find for the relative error in distance:

$$\frac{\sigma_d}{d} \approx 0.46 \sigma_\mu \approx 0.46 \sigma_{M_J}. \quad (19)$$

The left column of Fig. 6 shows that we expect for the main sequence and RGB stars in the RAVE data set a relative error in distance in the order of 25% (blue colors), and for the other stars around 50-60% (green colors).

### 3.3 Testing the method on GC-survey stars observed by RAVE

To test the method on real data, we use a subset of 64 stars from the RAVE dataset whose distances are known from the GC-survey. Of these stars, 35 have distances derived from trigonometric parallaxes (Hipparcos satellite). For the other 29 stars, this was not available or accurate enough and therefore a photometric parallax was determined by Nordström et al. (2004), having an uncertainty of about 13%.

In the top left panel of Fig. 7 we show the locations in a  $\log(g)$  versus  $\log(T_{eff})$  plot. Overlaid is a set of isochrones with  $[Fe/H] = 0$ . In the top right panel of Fig. 7 the distances derived using our method (and  $1\sigma$  errors) are plotted versus the distance as given by the GC-survey. The green circles indicate stars with distance from Hipparcos, red squares and blue triangles from photometric parallax for respectively F and G stars. Relative errors are in the range of 50%, as expected. A few stars are more than  $3\sigma$  away

from the one-to-one correspondance (shown as the black line). In the bottom row of Fig. 7 the corresponding HR diagrams are shown, where on the left the absolute magnitudes are used from the GC-survey, and on the right the estimated absolute magnitudes from our method using the astrophysical parameters ( $[M/H]$ ,  $\log(g)$ ,  $T_{eff}$ , and magnitudes) from the RAVE survey. Stars having their distance overestimated by 50 parsec with respect to the GC-survey distances are indicated by a magenta cross, underestimation by 50 parsec is indicated by yellow cross. The top left panel of Fig. 7 shows that the over and underestimation are caused by the errors in  $\log(g)$ .

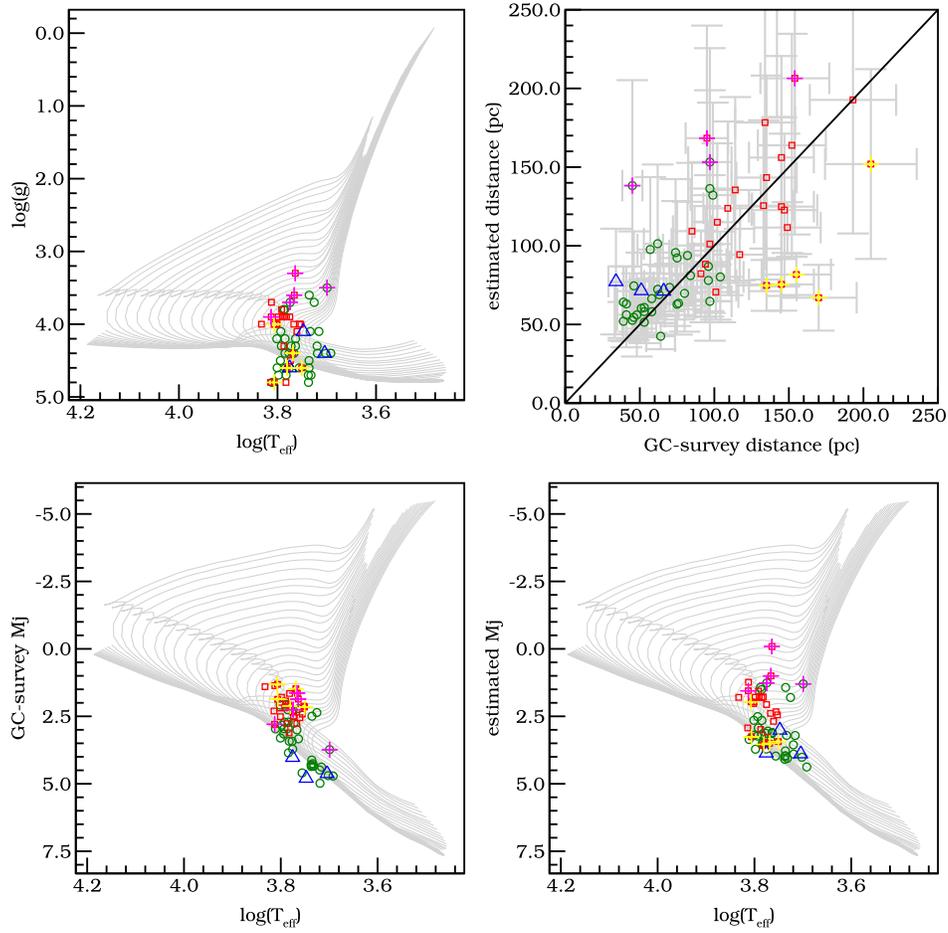
### 3.4 Distances for stars in the RAVE dataset

The internal data release of June 2007 contains 50 994 observations, of which 21 032 have astrophysical parameters. We first clean up the dataset according to the constraints of table 2. For some stars multiple observations are available, these are grouped by their id, and a weighted average (Eq. 14) and corresponding error (Eq. 15) for all radial velocity, proper motions and apparent magnitudes are calculated. The astrophysical parameters ( $[M/H]$ ,  $\log(g)$  and  $T_{eff}$ ) have nominal errors as described in §3.1. For these parameters an unweighted average is calculated and the error in the average is kept equal to the nominal errors of a single measurement. The resulting source count, after applying these constraints and grouping the multiple observations, is 18 663. After this step, the best model star is first found as described in §2.2. If it has a  $\chi_{model}^2 \geq 6$  (Eq. 13, including J-K) it is not considered further. This last step gets rid of stars which do not fit to any isochrone very well, and results in 17 434 sources which are used for the isochrone fitting method. The maximum  $\chi_{model}^2$  value of 6 is chosen after visual inspection of the outliers in the  $\log(g)$  vs  $T_{eff}$  plot.

We now apply the isochrone fitting routine to the 17 434 sources, resulting in 5000 estimates of  $M_J$  for each star. In the next step we generate 5000 apparent magnitudes ( $J$ ) for each star by convolving it with its errors assuming it is Gaussian. We now combine the absolute and apparent magnitudes giving us 5000 distances (Eq. 16). From these, we calculate the mean distance and the standard deviation for each star.

### 3.5 6d phase space coordinates for stars in the RAVE dataset

To calculate the full 6d phase space coordinates, the radial velocity, proper motion and sky coordinates are also needed. From the radial velocity and proper motion, 5000 samples are created by convolving them with their errors, assuming they are Gaussian. These 5000 samples are combined with the 5000 distances from the previous step. We now have 5000 distances, radial velocities and proper motions (in right ascension and declination direction). Combined with the sky coordinates we now calculate the 6d phase



**Figure 7:** Results from the RAVE subset for stars with known distances from the GC-survey. The grey lines are isochrones for  $[Fe/H] = 0$ . The green circles indicate stars with distance from Hipparcos, red squares and blue triangles from photometric parallax for respectively F and G stars. Stars having their distance overestimated by 50 parsec with respect to the GC-survey distances are indicated by a magenta cross, underestimation by 50 parsec is indicated by yellow cross. **Top left:** Overview of the location of the overlapping subset of RAVE and GC dataset in a  $\log(T_{eff})$  versus  $\log(g)$  plot. **Top right:** GC distance versus estimated distance from the isochrone fitting method. **Bottom left:** HR diagram using the absolute magnitude from the GC-survey. **Bottom right:** HR diagram using the absolute magnitude from the isochrone fitting method.

column name	constraint	
spectrum quality flag	empty (equals '-')	only stars which have no problem with the spectrum, no cosmic rays, no binaries etc.
signal to noise ratio	> 10	
$\chi^2$	< 2000	
proper motions	$\neq$ 9999.9	proper motions should be available (9999.9 means not available)
[M/H]	$\neq$ 9.9	metallicity abundance should be available
error in proper motion	$\neq$ 0.0	unable to calculate weighted averages
error in K magnitude	$\neq$ 0.0	unable to calculate weighted averages

**Table 2:** Constraints on the RAVE dataset.

space coordinates, using code provided by A. Helmi (private communication) based on Johnson and Soderblom (1987).

The  $x', y', z'$  coordinate system we use is a right handed Cartesian coordinate system centred on the Sun indicating positions, with the  $x'$  axis pointing from the Sun to the Galactic Centre (GC), the  $y'$  axis pointing in the direction of rotation and the  $z'$  axis pointing towards the Northern Galactic Pole (NGP). The  $x, y, z$  coordinate system is similar to the  $x', y', z'$  coordinate system, but centred on the GC, assuming the Sun is at  $(x, y, z) = (-8, 0, 0)$ . An overview can be found in Fig. 8 with Galactic longitude (l) and latitude (b) shown for completeness.

The velocities with respect to the Sun in the directions of  $x', y', z'$  are  $U, V, W$  respectively. For velocities of nearby stars, a Cartesian coordinate system will be sufficient, but for large distances, a cylindrical coordinate system makes more sense for the disk stars. To calculate these coordinates, we first have to transform the  $U, V, W$  velocities to the Galactic rest frame, indicated by  $v_x, v_y, v_z$  as shown in Fig. 9. Assuming a local standard of rest (LSR) of  $v_{lsr} = 220$  km/s, and the velocity of the Sun with respect to the LSR of  $(U_\odot, V_\odot, W_\odot) = (10.4\text{km/s}, 5.25\text{km/s}, 7.17\text{km/s})$  (Dehnen and Binney, 1998) we find:

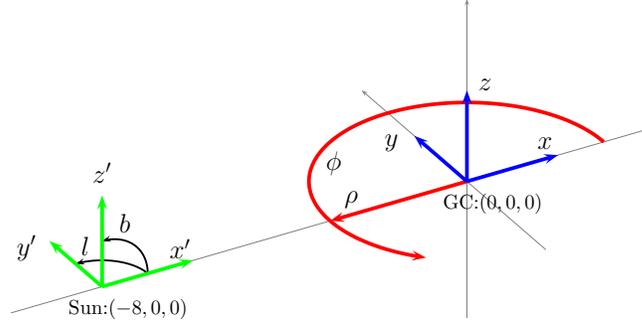
$$v_x = U + U_\odot, \quad (20)$$

$$v_y = V + V_\odot + v_{lsr}, \quad (21)$$

$$v_z = W + W_\odot. \quad (22)$$

$$(23)$$

The relations between Cartesian  $(x, y, z)$  and cylindrical coordinates



**Figure 8:** Overview of the Galactic coordinates. The Sun is found at  $(x, y, z) = (-8, 0, 0)$ .

$(\rho, \phi, z)$  are:

$$x = \rho \cos(\phi), \quad (24)$$

$$y = \rho \sin(\phi), \quad (25)$$

$$z = z, \quad (26)$$

$$\rho^2 = x^2 + y^2, \quad (27)$$

$$\tan(\phi) = \frac{y}{x}. \quad (28)$$

We can use this to find the velocities in the directions of  $\rho$  and  $\phi$ :

$$v_\rho = \frac{d\rho}{dt} = \frac{xv_x + yv_y}{\rho}, \quad (29)$$

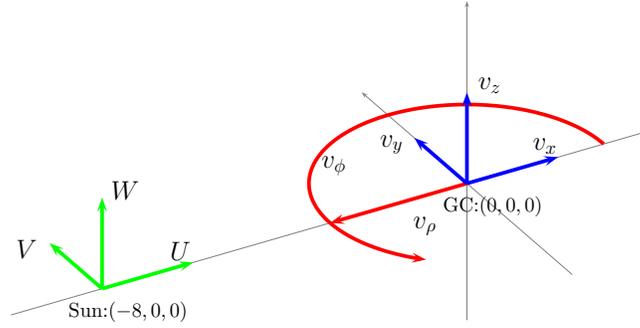
$$v_\phi = \rho \frac{d\phi}{dt} = \frac{xv_y - yv_x}{\rho}. \quad (30)$$

$$(31)$$

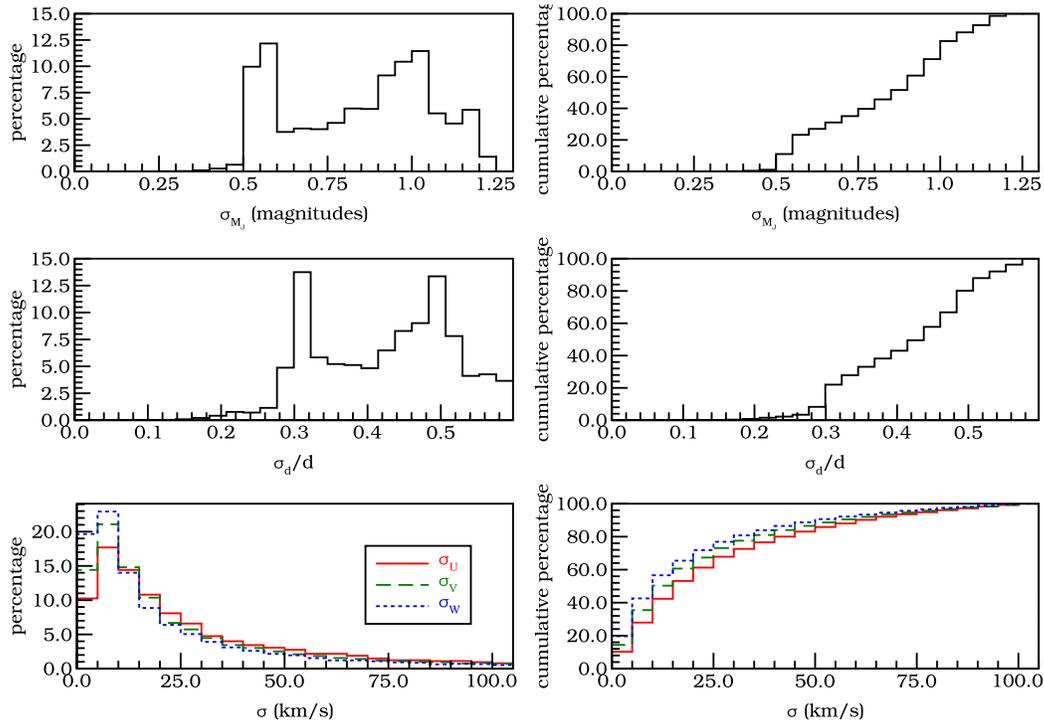
Note that the direction of  $\phi$  is anti-clockwise, meaning that the LSR is at  $(v_\rho, v_\phi, v_z) = (0, -220, 0)$ .

We calculated 5000 Cartesian coordinates positions and 5000 velocities  $(U, V, W, v_x, v_y, v_z, v_\rho, v_\phi)$  for each star. From these, we calculate the mean and standard deviation for each position and velocity component.

An overview of the errors is found in Fig. 10. In the top row, we plot the distributions of errors for  $M_J$  on the left, and cumulative errors on the right. In the middle row, the relative distance errors are plotted. The x-axis is scaled by the factor  $\frac{\ln(10)}{5} \approx 0.46$  (found in Eq. 18) with respect to the top row such that the top two rows can be compared. Although the shapes of the two histograms are similar, the errors in relative distance are larger than those in  $M_J$ , because of the error in the apparent magnitude  $J$ . In the bottom row, errors in velocities are plotted for the three components, all showing similar error behaviour.



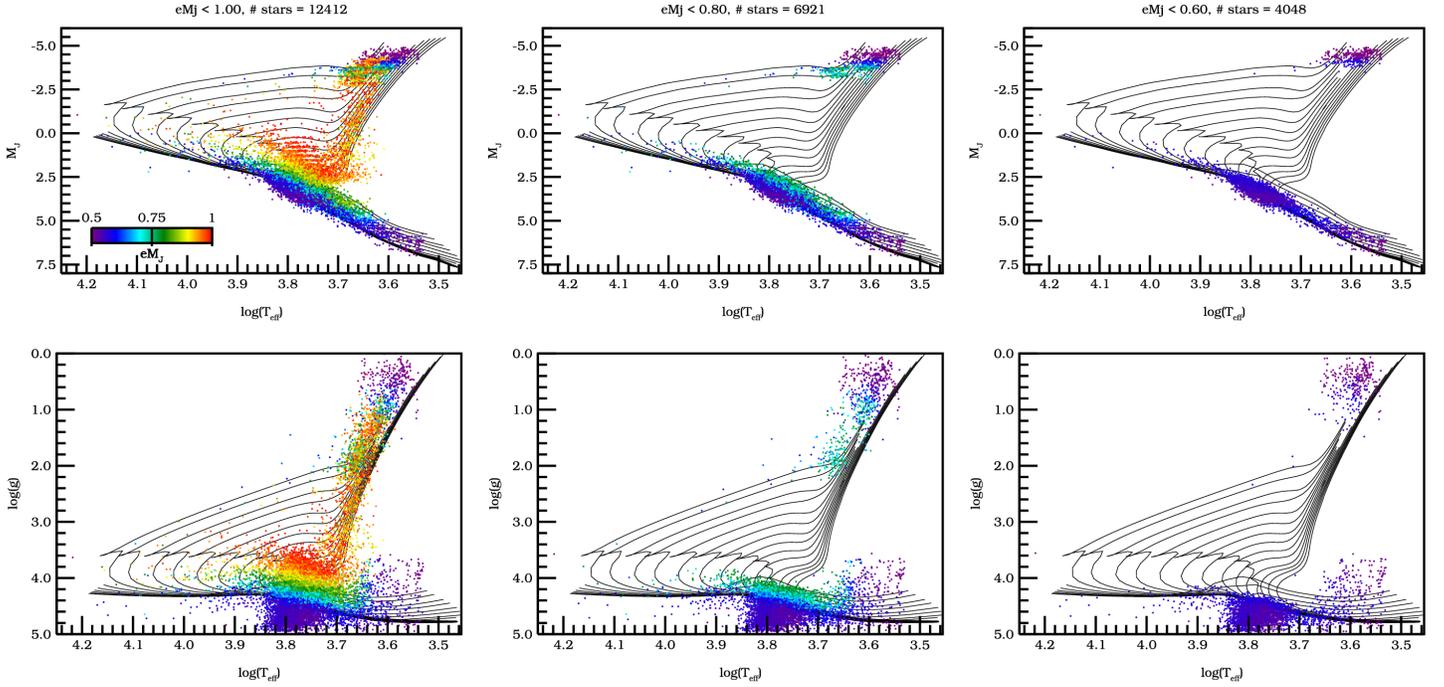
**Figure 9:** Overview of Galactic velocity coordinate systems.  $U, V, W$  velocities are with respect to the Sun and are aligned with the  $x', y', z'$  coordinate system.  $v_x, v_y, v_z$  are Cartesian velocities, and  $v_\rho, v_\phi$  are cylindrical velocities, both with respect to the Galactic rest frame.



**Figure 10:** Error distribution and cumulative plot for  $M_J$ , relative distance, and  $U, V, W$ . **Top row:** Errors for  $M_J$ . **Middle row:** Relative distance errors, x-axis is scaled by the factor  $\frac{\ln(10)}{5} \approx 0.46$  (found in equation 18) with respect to the top row such that the top two rows can be compared. **Bottom row:** Errors for  $U$  (red solid),  $V$  (green dashed) and  $W$  (blue dotted).

### 3.6 Calculation time

For the sample of 17 434 stars, it takes around 100 minutes to estimate the absolute magnitudes of the stars on a cluster of around 13 Dual core Intel Pentium (D) 2.8 Ghz pc, one 2x Dual Core AMD Opteron Processor 275 (2.2 Ghz) pc and a quad core Intel Xeon CPU 5150 (2.7 Ghz) pc. The calculations of the 6d phase space coordinates takes a few minutes on the same machines. If the total of  $10^6$  sources need to be calculated, this would take around 60 times longer,  $\sim 4$  days for the estimation of the absolute magnitudes and a few hours for the 6d phase space calculations.



**Figure 11:** Results for applying the isochrone fitting method to the RAVE data. Colours indicate the magnitude of the error in  $M_J$  (left  $\epsilon_{M_J} < 1$ , middle  $\epsilon_{M_J} < 0.8$ , right  $\epsilon_{M_J} < 0.6$ ). Isochrones for  $[Fe/H] = 0$  are plotted for comparison. **Top row:** HR diagram of RAVE dataset showing that the stars on the main sequence and RGB stars have the smallest errors. **Bottom row:**  $\log(T_{eff})$  versus  $\log(g)$ , colour coding as in the top panels. The typical error in  $\log(g)$  is 0.3 dex and in  $T_{eff}$  is 300 K.

## 4 Results

The results of the isochrone fitting method applied to the RAVE dataset are plotted in Fig. 11 for different limits on the maximum error in absolute magnitude. 12 412 stars have error  $\sigma_{M_J} < 1$ , for  $\sigma_{M_J} < 0.8$ , 6 921 sources, and for  $\sigma_{M_J} < 0.6$ , only 4 048 sources will be left. These correspond to relative distance errors less than 46%, 37% and 28%.

Our Galaxy consist mainly of three components, a bulge, a halo and the thin and thick disks (Binney and Tremaine, 1987). The thin disk has a scale height of around 300 pc, while the thick disk scale height is approximately 1 kpc. The disk is known to be dominated by metal rich stars, while halo stars are in general metal poor. To see if this is reflected in the RAVE data, we will now focus on how the metallicity changes as function a of distance from the plane. In Fig. 12 we show the spatial distribution of stars in the RAVE dataset. In the top left panel we can see a rather strong concentration of

stars towards  $z = 0$ , i.e. the Galactic mid plane. The bottom left panel on the other hand shows a more axisymmetric distribution. This figure also shows that the RAVE survey probes deep into the halo, and we therefore can look for a metallicity change as function of height above the Galactic plane ( $z$ ).

Many stars have high uncertainties in their position, and due to a high density of stars in the plane, the halo gets ‘polluted’ by disk stars. Therefore we will from now on only work with stars having  $\sigma_{M_J} < 0.8$ . The spacial distribution of this cleaned up sample is plotted in Fig. 13, showing less halo stars, and a more pronounced disk than Fig. 12.

In the left panel of Fig. 14 we have plotted the metallicity distribution for this sample for three ranges of  $z$ . The red solid line for  $0 \leq |z| < 1$  kpc, the green dashed line for  $1 \leq |z| < 3$  kpc and the blue dotted line for  $3 \leq |z| < 8$  kpc. Note that most of the stars, as expected, are in the thin disk, and have a mean metallicity of 0.0 dex. In the middle panel the histograms are normalised such that the area of a single histogram is 100% to make it easier to compare the metallicity distributions. As we move to higher  $|z|$ , the mean metallicity decreases, but even for our farthest bin we appear to be dominated by the thin disk. Note however that a metal poor tail is clearly visible in the central panel of Fig. 14. In the right panel of Fig. 14 the mean metallicity is plotted as function of  $|z|$ , showing a decrease in the mean metallicity till at least 6 kpc. A similar trend is seen in Ak et al. (2007a) and Ak et al. (2007b), although our gradient in Fig. 14 is much shallower, possibly due to contamination of disk stars having large distance errors.

We will now analyse the velocities of the stars in the RAVE dataset. For the tangential velocity, we have:

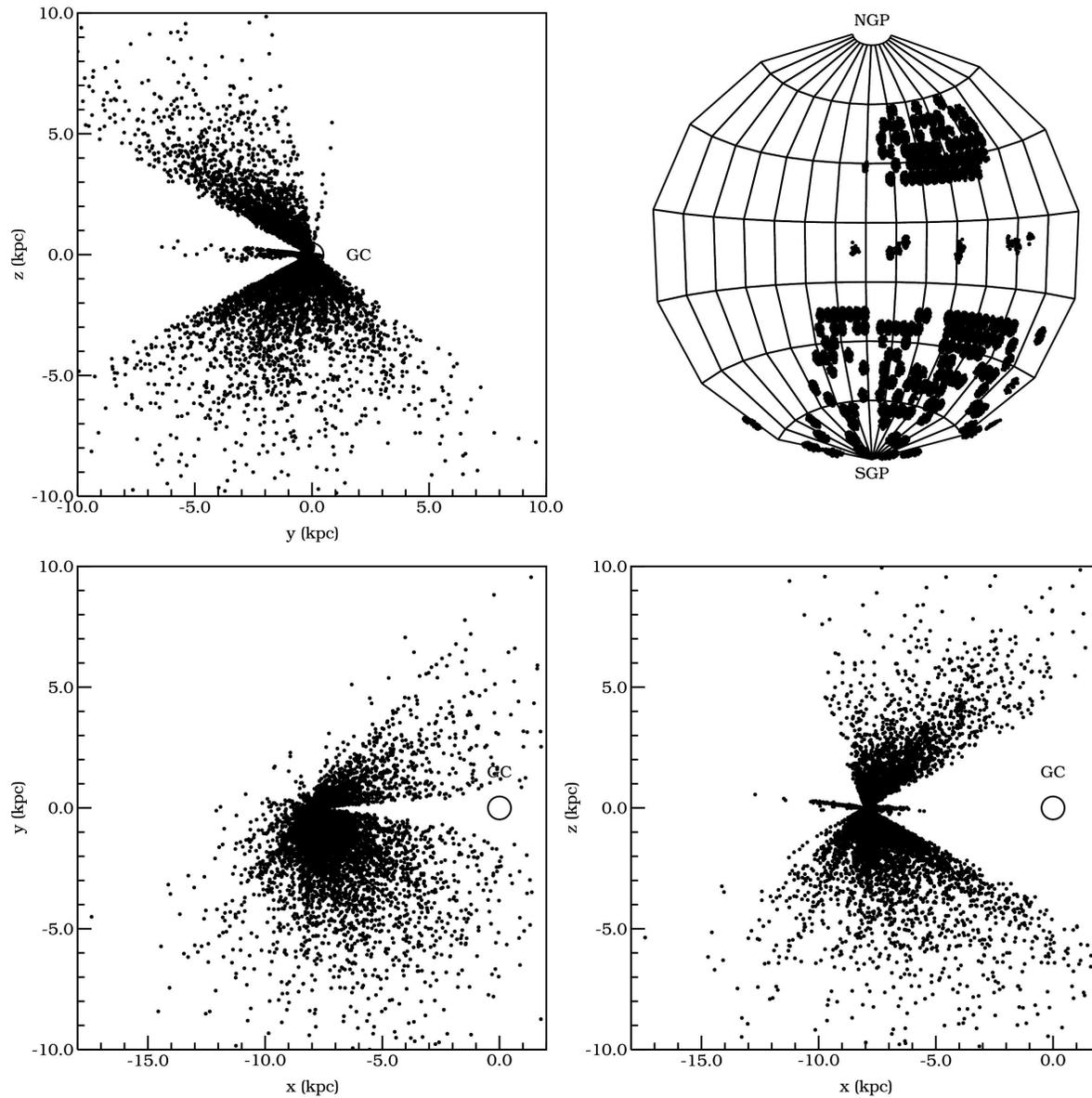
$$v_{\perp} \propto \mu d, \quad (32)$$

where  $\mu$  is the proper motion and  $d$  the distance. For the error in  $v_{\perp}$  we get:

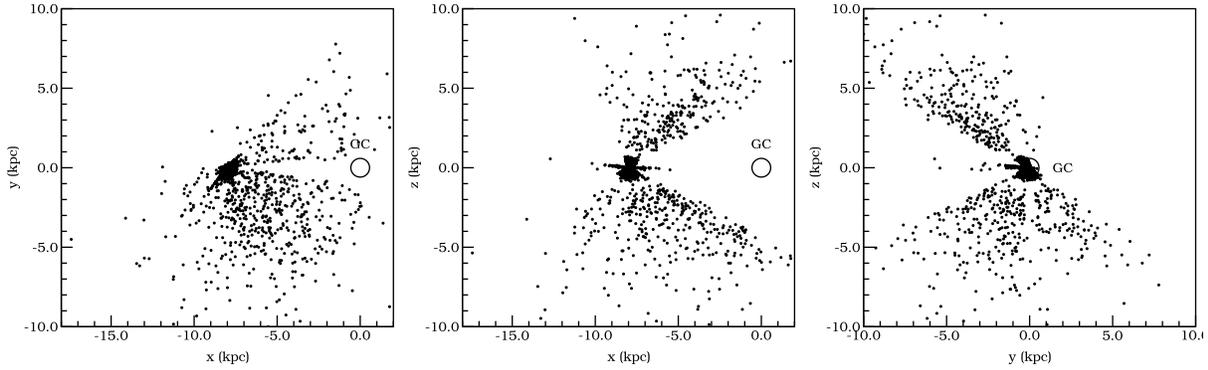
$$\sigma_{v_{\perp}}^2 \propto \mu^2 \sigma_d^2 + \sigma_{\mu}^2 d^2 \propto v_{\perp}^2 \sigma_{M_J}^2 + \sigma_{\mu}^2 d^2. \quad (33)$$

If we want to analyse velocities, we want them to have small error. Therefore we select a subsample of stars with small errors in  $M_J$  ( $\sigma_{M_J} < 0.8$ ), in proper motion ( $\sigma_{\mu, RA} < 5$  milli arc second per year (mas/year),  $\sigma_{\mu, DEC} < 5$  mas/year) and radial velocities ( $\sigma_{v_R} < 5$  km/s). We do not yet select stars with small distances since we also want to analyse the halo stars. This subset contains 5 029 stars.

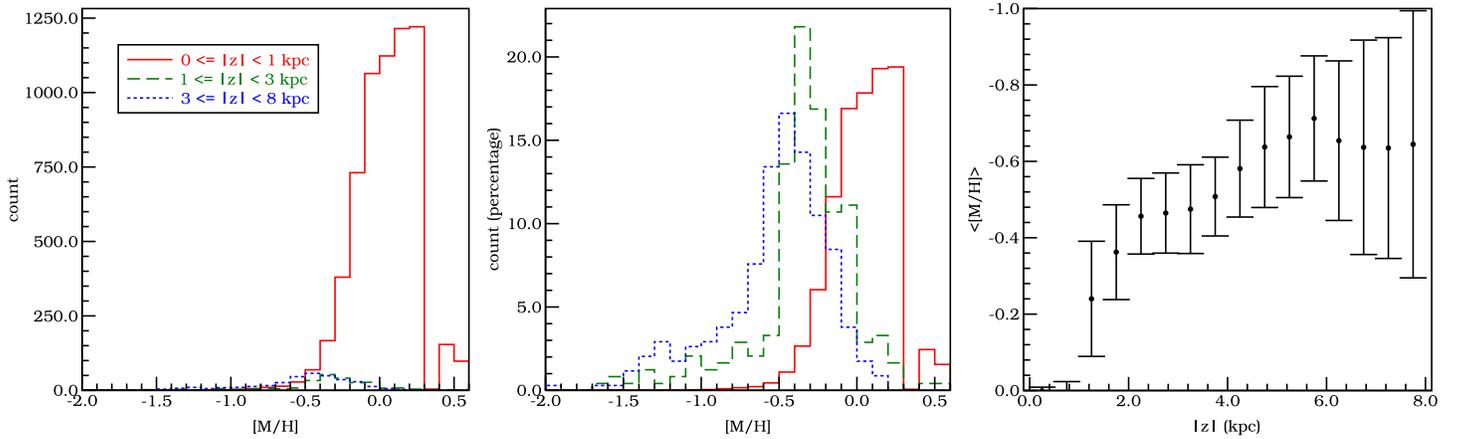
In Fig. 15 we have plotted the average angular velocities in different bins of  $|z|$ . It shows a decreasing rotational velocity as we move away from the Galactic plane which can be explained by a fast rotating thin disk, a less fast rotating thick disk, and an almost non rotating halo.



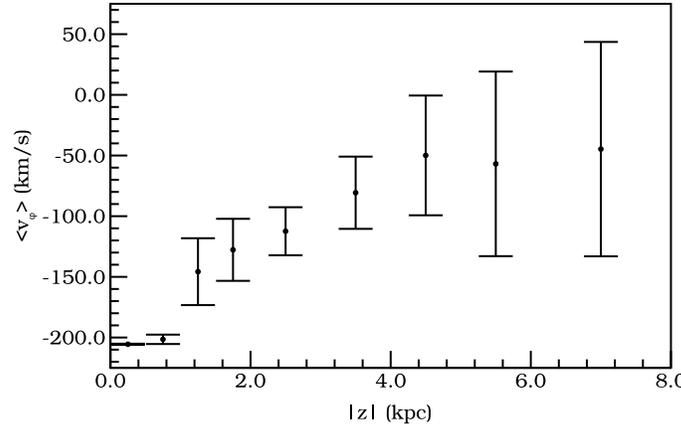
**Figure 12:** The RAVE stars in galactic coordinates, the circle with label GC indicates the galactic centre. **Bottom left:** Face on view of our Galaxy. **Bottom right:** Edge on view of our Galaxy, with our Sun at the left, and the galactic centre on the right. **Top left:** Edge on view, looking through the galactic centre at our Sun. **Top right:** Galactic sky coordinates, with the Northern Galactic Pole (NGP) at the top, showing the sky coordinates for the RAVE observations.



**Figure 13:** Similar to Fig. 12, except only showing stars with  $\sigma_{M_J} < 0.8$  ( $\sigma_d/d = 38\%$ ).



**Figure 14: Left and middle:** Metallicity distribution for stars in different bins of height above the Galactic plane. Stars further away from the Galactic plane are more metal poor. **Left:** Absolute count for different metallicity bins. **Middle:** Normalised histograms such that the area of a single histogram is 100%. **Right:** Mean metallicity as function of distance from the Galactic plane showing a decrease in metallicity till 6 kpc. Errorbars indicate  $1\sigma$  errors in the means.

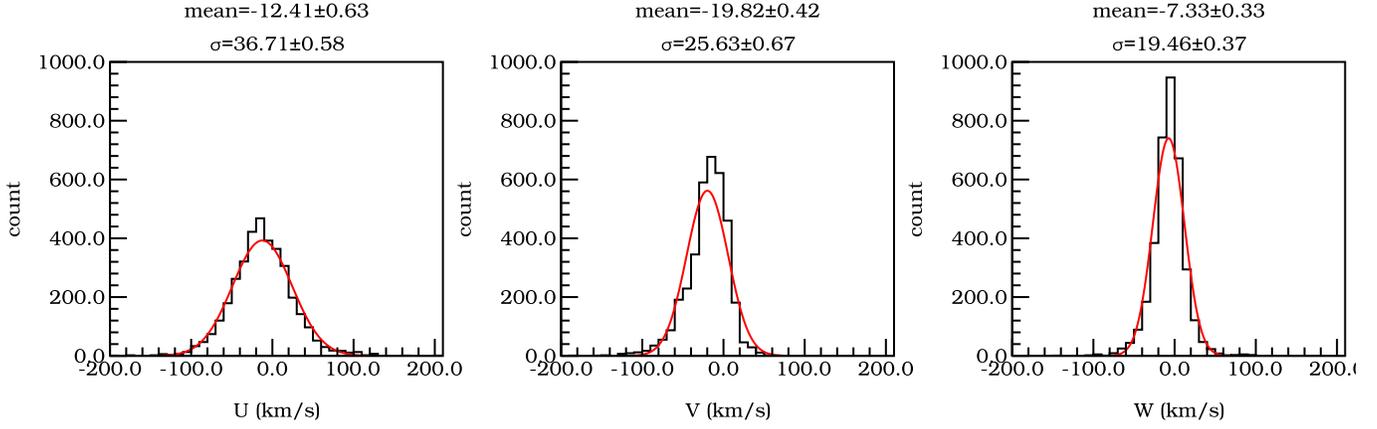


**Figure 15:** Rotational velocity decreasing as function of  $|z|$ , errorbars indicate  $1\sigma$  errors in the means. Note that negative  $v_\phi$  means clockwise rotation.

If both the errors and the velocity distribution were Gaussian, then the observed velocity dispersion would be the result of the convolution of the intrinsic velocity dispersion and the measurement error. And if the measurement error ( $\epsilon$ ) was the same for each star, then  $\sigma_{obs}^2 = \sigma_{int}^2 + \epsilon^2$ . If we know  $\epsilon$ , we can calculate  $\sigma_{int}$ , but if the  $\epsilon$  are much smaller compared to  $\sigma_{int}$ , the effect on  $\sigma_{obs}$  is negligible.

Close to the Galactic plane, and not too far from the Sun, the errors in velocity are small, therefore apart from the restrictions described above, we will restrict ourselves to a small volume around the Sun to analyse the velocity dispersions. In Fig. 16 the velocity distributions for 2 188 stars in a cylinder centred around the Sun with a radius of 500pc and a height of 600pc (300 above and below the Galactic plane) are plotted. This sample has average errors of  $\epsilon_U = 8.0$  km/s,  $\epsilon_V = 6.5$  km/s,  $\epsilon_W = 5.4$  km/s. The solid lines are best fit Gaussians to the velocity distributions and although the average errors are small compared to the velocity dispersions, this figure shows that the wings are slightly wider, which causes the estimates of  $\sigma$  to be somewhat larger. This can be understood as the effect of the errors in velocity which are proportional to the magnitude of the velocity itself, creating these wings. To be able to give accurate and unbiased values for velocity dispersions a more sophisticated model than a Gaussian is needed, which will not be treated in this report.

For this sample, we compute the mean velocities, the full velocity dis-



**Figure 16:** Velocity distributions for the U, V and W components (histogram) and the best Gaussian fit (solid line). **Left:** The velocity distribution for U is highly symmetric, showing a slight negative mean U. **Middle:** The V component shows a slight asymmetry, having a longer tail towards the lower part. **Right:** The velocity distribution of W shows a lower dispersion than U and V, and is symmetric.

persion tensor  $\sigma_{ij}$  and the vertex deviation:

$$\bar{U} = -12.43 \pm 0.60 \text{ km/s} \quad (34)$$

$$\bar{V} = -19.81 \pm 0.42 \text{ km/s} \quad (35)$$

$$\bar{W} = -7.33 \pm 0.32 \text{ km/s} \quad (36)$$

$$\sigma_U = 36.72 \pm 0.55 \text{ km/s} \quad (37)$$

$$\sigma_V = 25.62 \pm 0.66 \text{ km/s} \quad (38)$$

$$\sigma_W = 19.44 \pm 0.39 \text{ km/s} \quad (39)$$

$$\sigma_{UV} = 10.94 \pm 1.00 \text{ km/s} \quad (40)$$

$$\sigma_{UW}^2 = -6.41 \pm 17.13 \text{ km/s} \quad (41)$$

$$\sigma_{VW}^2 = 19.12 \pm 14.71 \text{ km/s} \quad (42)$$

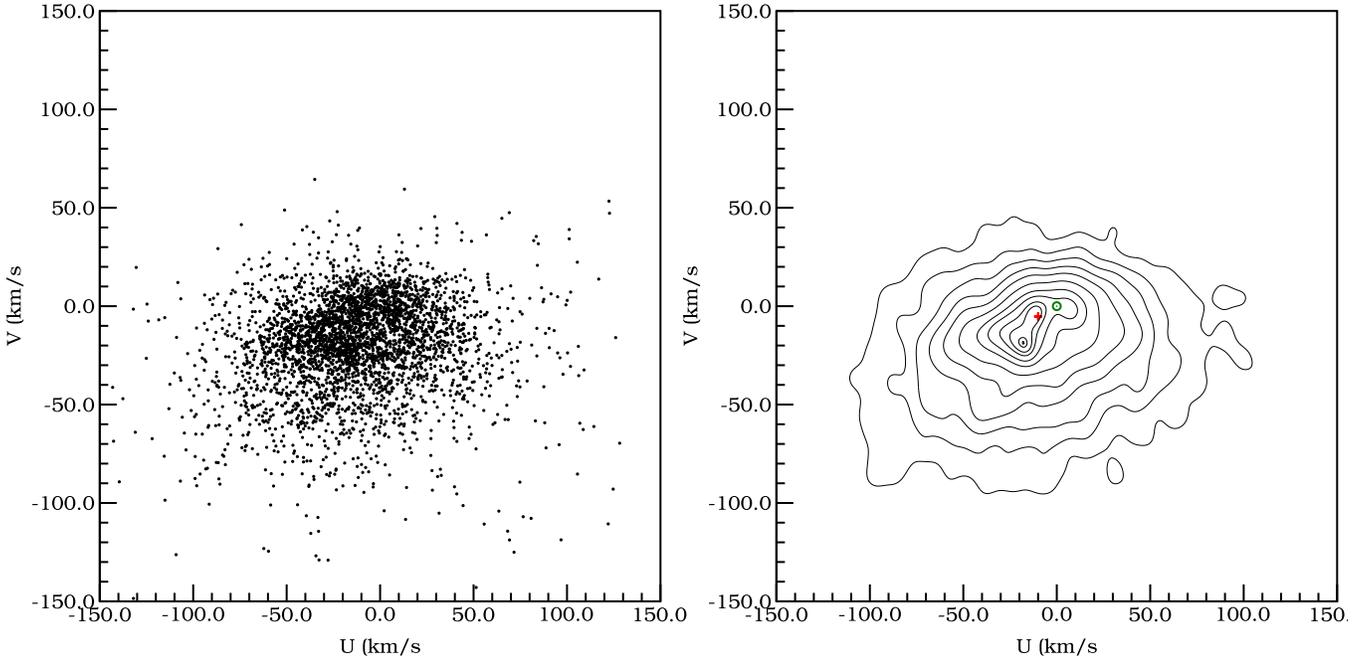
$$l_v = 9.61 \pm 1.62^\circ \quad (43)$$

where  $l_v$  is the vertex deviation, which is defined as:

$$l_v = \frac{1}{2} \arctan \left( 2 \frac{\sigma_{UV}^2}{\sigma_U^2 - \sigma_V^2} \right), \quad (44)$$

and is a measure for the orientation of the UV velocity ellipsoid. The velocities and dispersions are in good agreement with Famaey et al. (2005) and Dehnen and Binney (1998) using Hipparcos data. The  $\sigma_{UW}^2$  and  $\sigma_{VW}^2$  do not indicate a significant asymmetry in these directions.

Close inspection of the middle panel of Fig. 16 shows an asymmetric distribution for the V component, with a longer tail towards lower velocities.



**Figure 17:** **Left:** UV plane shows asymmetry and a vertex deviation. **Right:** Isodensity contour lines for the UV plane. The red + symbol marks the LSR the green  $\odot$  symbol marks the Solar velocity (0,0). The contour lines contain 2, 6, 12, 21, 33, 50, 68, 80, 90, 99 and 99.9 percent of the stars.

This is due to two effects. The first is that we are seeing the asymmetric drift, stars showing negative rotational velocities with respect to the LSR. Velocity dispersion and stellar density increase towards the GC (Binney and Tremaine, 1987) such that at a fixed radius, we find more stars near their apocenter than their pericenter, meaning we see more visitors on orbits closer to the GC than we see visitors on larger orbits. Secondly the tail towards low  $V$  can also be seen in Fig. 17 where each star is plotted in UV-plane in the left panel, and the density contours are shown in the right panel. A slight over density of stars around  $U \approx -50$ ,  $V \approx -50$  can be seen which will affect the symmetry of the  $U$  velocity component. This over density is called the Hercules stream, and is thought to be an effect of the bar of our Galaxy (Famaey et al., 2005).

## 5 Discussion and conclusion

We presented a method to derive absolute magnitudes, and therefore distances, for RAVE stars. It is based on the use of isochrone fitting in the metallicity,  $\log(g)$ ,  $T_{eff}$  and colour space. For this method we used the  $Y^2$  isochrones.

The errors in the estimated absolute magnitudes for RGB stars are found to depend mainly on the error in  $\log(g)$  while for main sequence stars the accuracy of  $T_{eff}$  is also important (§3.2). The errors in  $\log(g)$  and  $T_{eff}$  for the RAVE data give rise to relative error in distance in the range 30%-50%, but as seen in the results (§4), the data does reflect the known properties of the halo and disk stars of the Milky Way.

A metallicity gradient found in the direction away from the Galactic plane corresponding to an increase in the fraction of metal poor halo stars. The decrease in the  $v_\phi$  component as we move away from the Galactic plane can be explained by the transition from thin disk to thick disk to halo stars. The asymmetry in the  $V$  direction and vertex deviations are clearly visible in the UV plane and are most likely a combination of asymmetric drift and structures such as the Hercules streams. The vertex deviation is in agreement with previously found values.

Helmi et al. (2006) studied the kinematics of our Galaxy using the Geneva Copenhagen data. The relative error in distance for this survey is around  $\sim 13\%$  for  $\sim 14\,000$  stars. For the RAVE survey the errors are in the order  $\sim 30\text{-}50\%$  for  $\sim 12\,000$  stars but this survey is  $\sim 4$  magnitudes deeper and hence probes further away. In the future the RAVE survey will increase its size by a factor of  $\sim 20$  providing an interesting and significantly larger dataset before the GAIA mission will fly.

## A Description of RAVE catalogue with phase space coordinates

The resulting dataset is provided as a comma separated values (CSV) file, with headers. The columns are described in the table below. See Steinmetz et al. (2006) for a more detailed description about the RAVE data.

Field name	Units	Type	Description
OBJECT_ID		string	RAVE internal identifier
RA	deg	float	Right ascension (J2000)
DE	deg	float	Declination (J2000)
Glon	deg	float	Galactic longitude
Glat	deg	float	Galactic latitude
RV	km/s	float	Weighted mean of available radial velocities
eRV	km/s	float	Weighted error of available radial velocities
pmRA	km/s	float	Weighted mean of available proper motions
pmDE	km/s	float	Weighted mean of available proper motions
epmRA	km/s	float	Weighted error of available proper motions
epmDE	km/s	float	Weighted error of available proper motions
Teff	Kelvin	float	Arithmetic mean of available temperatures
nTeff		int	Number of observations having $T_{eff}$
logg	$\log(\frac{cm}{s^2})$	float	Arithmetic mean of available surface gravities
nlogg		int	Number of observations having $\log(g)$
MH	dex	float	Arithmetic mean of RAVE uncalibrated metallicity abundance
nMH		int	Number of observations having MH
MHcalib	dex	float	Arithmetic mean of RAVE calibrated metallicity abundance
nMHcalib		int	Number of observations having MHcalib
AM	dex	float	Arithmetic mean of RAVE alpha enhancement
nAM		int	Number of observations having AM
Jmag	mag	float	Weighted mean from available magnitudes
eJmag	mag	float	Weighted error from available magnitudes
Kmag	mag	float	Weighted mean from available magnitudes
eKmag	mag	float	Weighted error from available magnitudes
Mj	mag	float	Absolute magnitude in J band (from fitting routine)
eMj	mag	float	Error in MJ
distance	kpc	float	Distance from MJ and Jmag
edistance	kpc	float	Error in distance
age	Gyr	float	Age estimate from isochrone fitting <sup>1</sup>
eage	Gyr	float	Error in age
mass	in $M_{\odot}$	float	Mass estimate from isochrone fitting <sup>1</sup>
emass	in $M_{\odot}$	float	Error in mass

---

<sup>1</sup>Not analysed or tested.

Field name	Units	Type	Description
xGal	kpc	float	Galactic x coordinate <sup>2</sup>
exGal	kpc	float	Error in x
yGal	kpc	float	Galactic y coordinate <sup>2</sup>
eyGal	kpc	float	Error in y
zGal	kpc	float	Galactic z coordinate <sup>2</sup>
ezGal	kpc	float	Error in z
U	km/s	float	Galactic velocity in $x'$ direction w.r.t the Sun <sup>2</sup>
eU	km/s	float	Error in U
V	km/s	float	Galactic velocity in $y'$ direction w.r.t the Sun <sup>2</sup>
eV	km/s	float	Error in V
W	km/s	float	Galactic velocity in $z'$ direction w.r.t the Sun <sup>2</sup>
eW	km/s	float	Error in W
vxGal	km/s	float	Galactic velocity ( $v_x$ ) in $x$ direction in Galactic rest frame <sup>2</sup>
evxGal	km/s	float	Error in vxGal ( $v_x$ )
vyGal	km/s	float	Galactic velocity ( $v_y$ ) in $y$ direction in Galactic rest frame <sup>2</sup>
evyGal	km/s	float	Error in vyGal ( $v_y$ )
vzGal	km/s	float	Galactic velocity ( $v_z$ ) in $z$ direction in Galactic rest frame <sup>2</sup>
evzGal	km/s	float	Error in vzGal ( $v_z$ )
Vr	km/s	float	Galactic velocity ( $v_\rho$ ) in $\rho$ direction in Galactic rest frame <sup>2</sup>
eVr	km/s	float	Error in Vr ( $v_\rho$ )
Vphi	km/s	float	Galactic velocity ( $v_\phi$ ) in $\phi$ direction in Galactic rest frame <sup>2</sup>
eVphi	km/s	float	Error in Vphi ( $v_\phi$ )

## B Software manual

The fitting routine software package is provided as a zipped tarball, ‘ko-rave-yyyy-mm-dd.tar.gz’ containing all software to obtain the full 6d phase space coordinates, where yyyy-mm-dd refers to year, month and day of month at which the released is made. It contains a small manual (filename README) which is included below.

**Author:** M.A. Breddels

### About

This document is about setting up and running the isochrone fitting routine for the RAVE dataset. It outputs position and velocity components. Most

---

<sup>2</sup>See §3.5 for a description.

tasks are simplified by making use of 'Makefiles'. All commands, like:

```
# make rave_pre
```

should be executed from the same directory as this help file. The command:

```
# make
```

Without arguments should give you some quick help info.

## Requirements

- Python 2.3 or above
- c compiler (gcc)

## Optional

- Pyrex (when changes are made to *korave/isofind.pyx*)

## Includes

- Python code + module to do the fitting *korave/\**
- Python cluster module to run the software on multiple machines *mab/cluster.py*
- Sample isochrones *data/isochrones/\** (you can change the location in the *Makefile*)
- Sample RAVE data *data/rave.dat* (you can change the name in the *Makefile*)

This means that you don't have to add any files to test the whole package.

## Step 1: Setting up

### Python code

Part of the Python code is done in a module. Normally it's written in *c*, but I used a wrapper program for it called *Pyrex*. Fortunately this program generates *c-code* (which I included) so there is no need to install it. But, if you make changes to *korave/isofind.pyx* you need to generate the corresponding *korave/isofind.c* by using the pyrex program. **If** you do, run this:

```
# pyrex korave/isofind.pyx
```

**Note**, this is not required, I included the *korave/isofind.c* file. To compile the module, I use distutils, which is always included with Python, to make it easier I added a Makefile target, so:

```
# make pycode
```

should do the trick, ignore the compiler warnings.

## Isochrones

I included the isochrones in *data/isochrones*, I advise not to generate new isochrone until you got everything working. You can change the location in the *Makefile*.

## RAVE data

I included a sample of 10 sources in *data/rave.dat*, you can change the location in the *Makefile*

## Step 2: Pre isochrone fitting

### Generate the csv file

Whatever format you want to feed into the fitting routine, reformat it to a *comma separated values* (csv) file. All routines are working with *csv* files. I included a file *korave/rave\_csv.py* that can at least reformat the internal data release file of July 2007 to a csv file, but not without problems. There is some problem with the number of headers found and the number of columns (they seem to be space separated, but some columns are stucked together, therefore I recommend csv files). Hold your breath and execute:

```
# make rave_csv
```

If you used the included file (*data/rave.dat*), ignore the warning, the output should give:

```
header length = 68
first line length = 68
header length doesn't match, they will be truncated, this is ok if
it's only the last column, check it!
```

If you used you own file **check** the number of header etc, do a manual check so that you **absolutely** sure the csv file is correct.

## Preprocessing

We're now gonna add multiple observation together, and throw away observation as described in the report. The following make command will execute *korave/rave\_pre.py*:

```
# make rave_pre
```

If you used the included file (*data/rave.data*), you will (hopefully get):

```
python korave/rave_pre.py data/rave.csv > data/rave.pre.csv
input: 10
output: 6
```

Meaning, out of the 10 sources, 6 are left in the file *data/rave.pre.csv*

## Cleaning up

Now we are throwing away the  $\text{chisq} > 6$  stars:

```
# make rave_clean
```

If you used the included file (*data/rave.data*), you should get the following output:

```
python ko/rave_clean.py data/rave.pre.csv
      data/isochrones/yy* > data/rave.clean.csv
loading RAVE db...
input: 6
loading isochrones...
calculating chisqs...
precount: 6
output: 6
```

We have now the file *data/rave.clean.csv* which will be ready for the isochrone fitting routine.

## Step 3: Isochrone fitting

This takes quite a while to compute, but I coded a Python module that can execute the same program on multiple computers using ssh, and communicating via STDIN/STDOUT. That means that it's almost zero-config cluster software, and should run out of the box. It does have some requirements:

- You should be able to connect to it with ssh without a password, therefore make sure that the command *ssh virgo02* connects to this machine instantly, without asking for a password. If you don't know how, see the Kapteyn HOWTO database, or google on *ssh private/public key*.

- The files on each node of the clusters should see the same directory structure, meaning that */some/dir/isochrone* should be the same from each computer the program will run on.
- In case of mixed 32 and 64 bit machines, and when the Python module is compiled on a 32 bit machine there may be missing 32 libraries (you can manually copy them and make sure LD\_LIBRARY\_PATH points to them). When the module is compiled on a 64 bit machine, it will likely crash on the 32 bit machines.

I recommend compiling it on a 32 bit machine, and make sure the 64 bit machines have the correct 32 bit libraries. The *hypatia* and *hercules03* machines missed the following libraries:

```
libblas.so      libblas.so.3.0.3  liblapack.so    liblapack.so.3.0.3
libblas.so.3   libg2c.so.0       liblapack.so.3
libblas.so.3.0 libg2c.so.0.0.0   liblapack.so.3.0
```

When you run `python -c "import korave.isofind"` on *hypatia* it will complain that it cannot find these libraries.

I copied these files from the (local) `/usr/lib/` directory from a 32 bit machine to a network disk directory (`/Users/users/breddels/bulk/lib32`), and added this path to the LD\_LIBRARY\_PATH using the following code in my `~/cshrc`:

```
switch ( $HOST )
case hercules03:
case hypatia:
    setenv LD_LIBRARY_PATH /Users/users/breddels/bulk/lib32
    breaksw
endsw
```

Now `python -c "import korave.isofind"` should run fine on these two machines.

## Preparing the program

Open the file `korave/isofit.py`, the last lines should look something like:

```
node = mab.cluster.createNode(worker=worker, processors=2, parser=parser)

for i in range(2,16):
    if i not in [2,4]:
        node.addRemote("virgo%02d" % i, processors=2)
```

```
node.addRemote("hercules03", processors=3, forceUse=True)
node.addRemote("hypatia", processors=3)
node.start()
```

`node.addRemote` takes 1 required argument, the machine hostname. The optional (keyword) argument `processors` is the number of threads that will be started on the remote machine, and therefore should equal the number of processors (see next section). The optional (keyword) argument `forceUse`, is a boolean which defaults to *False*. When someone is locally logged in using *X*, the program chooses not to use the machine, when `forceUse` is *True* it will ignore this, and will always use the machine. Note that adding *virgo02-virgo15* is done in a loop, and excluding machines (e.g. when they are down) is doing by an exclusion list (*[2,4]* in this case).

*hercules03* and *hypatia* are 64 bit machines, make sure that 32 system libraries are available.

### Listing the machines

The program takes a *-help* argument:

```
$ python korave/isofit.py --help
options:
  -h, --help            show this help message and exit
  --cmd=CMD             execute 'cmd' on remote machines
  --test               like --noremote, but without multiple threads
  --noremote           doesn't use remote slaves
  --who                shows who is logged in on the remote machine
  --listfree           list the 'free' machines (no X logins)
  --cpucount           get cpu count for remote machines
  --cpuinfo            get cpu info for remote machines
  -q, --quiet          no debug output
  --slave              puts machine in slave mode(used internally)
  --processors=PROCESSORS
                       # processors for machine(used internally)

  --eTeff=ETEFF
  --eLogg=ELOGG
  --testjob            run test stuff
```

The *-cpucount* option is useful for getting the number of cpu's on remote machines (instead of manually checking `/proc/cpuinfo` on all machines). The *-cpuinfo* gives information about the type of cpu's the remote machines have.

## Running the fit

When the program crashes, it is kind of hard to kill (due to multi threading), use `kill -9 <pid>` where pid is the process id obtained from `ps` for instance. During testing it is useful to run the program with the `-test` argument, when given, the program gracefully dies when crashes, and it is possible to kill it with Ctrl-C.

The make command `rave_isofit_test` does exactly this, it is recommended to run this first to see if the program is working properly (not crashing):

```
$ make rave_isofit_test
```

If this works ok, try running the program normally:

```
$ make rave_isofit
```

This should create one file for each star in the database to the `data/rave/` directory. These files contain the 5000 fitted magnitudes, and some statistics (mass, age, and errors).

The command is executed by the `time` command so that after completion it will list the time it took to run. Don't worry if it crashes near the end, restart, and it will continue where it stopped.

## Step 4: To phase space

About the same process as the previous step, edit `korave/rave_6d.py` for the machines you want to use.

First test if everything is working:

```
$ make rave_6d_test
```

If ok, run the whole batch:

```
$ make rave_6d
```

This should give you the file `data/rave.final.csv`

## Step 5: Post processing

Any post processing can be done in `ko/rave_post.py`, the default is to do nothing:

```
$ make rave_post
```

This will output the file `data/rave.final.post.csv`

And post-post process is done by `ko/rave_post_vel.py`, which makes cuts to select stars for velocity analysis:

```
$ make rave_post_vel
```

This will output the file `data/rave.final.vel.csv`

## References

- S. Ak, S. Bilir, S. Karaali, and R. Buser. Estimation of galactic model parameters with the Sloan Digital Sky Survey and the metallicity distribution in two fields in the anti-centre direction of the Galaxy. *Astronomische Nachrichten*, 328:169–+, February 2007a.
- S. Ak, S. Bilir, S. Karaali, R. Buser, and A. Cabrera-Lavers. The metallicity distributions in high-latitudes with SDSS. *New Astronomy*, 12:605–612, November 2007b.
- J.J. Binney and S. Tremaine. *Galactic Dynamics*. Princeton University Press, 1987.
- D. L. Crawford. Empirical calibration of the ubvy,beta systems. I. The F-type stars. *AJ*, 80:955–971, November 1975.
- W. Dehnen and J. J. Binney. Local stellar kinematics from HIPPARCOS data. *MNRAS*, 298:387–394, August 1998.
- B. Famaey, A. Jorissen, X. Luri, M. Mayor, S. Udry, H. Dejonghe, and C. Turon. Local kinematics of K and M giants from CORAVEL/Hipparcos/Tycho-2 data. Revisiting the concept of superclusters. *A&A*, 430:165–186, January 2005.
- A. Helmi and P. T. de Zeeuw. Mapping the substructure in the Galactic halo with the next generation of astrometric satellites. *MNRAS*, 319:657–665, December 2000.
- A. Helmi, J. F. Navarro, B. Nordström, J. Holmberg, M. G. Abadi, and M. Steinmetz. Pieces of the puzzle: ancient substructure in the Galactic disc. *MNRAS*, 365:1309–1323, February 2006.
- D. R. H. Johnson and D. R. Soderblom. Calculating galactic space velocities and their uncertainties, with an application to the Ursa Major group. *AJ*, 93:864–867, April 1987.
- S. R. Majewski, M. F. Skrutskie, M. D. Weinberg, and J. C. Ostheimer. A Two Micron All Sky Survey View of the Sagittarius Dwarf Galaxy. I. Morphology of the Sagittarius Core and Tidal Arms. *ApJ*, 599:1082–1115, December 2003.
- B. Nordström, M. Mayor, J. Andersen, J. Holmberg, F. Pont, B. R. Jørgensen, E. H. Olsen, S. Udry, and N. Mowlavi. The Geneva-Copenhagen survey of the Solar neighbourhood. Ages, metallicities, and kinematic properties of  $\sim 14000$  F and G dwarfs. *A&A*, 418:989–1019, May 2004.

- M. A. C. Perryman, K. S. de Boer, G. Gilmore, E. Høg, M. G. Lattanzi, L. Lindegren, X. Luri, F. Mignard, O. Pace, and P. T. de Zeeuw. GAIA: Composition, formation and evolution of the Galaxy. *A&A*, 369:339–363, April 2001.
- W. H. Press. *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press, second edition, 1988.
- M. Salaris and S. Cassisi. *Evolution of Stars and Stellar Populations*. Evolution of Stars and Stellar Populations, by Maurizio Salaris, Santi Cassisi, pp. 400. ISBN 0-470-09220-3. Wiley-VCH, December 2005., 2005.
- D. N. Spergel, R. Bean, O. Doré, M. R.olta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, June 2007.
- M. Steinmetz, T. Zwitter, A. Siebert, F. G. Watson, K. C. Freeman, U. Munari, R. Campbell, M. Williams, G. M. Seabroke, R. F. G. Wyse, Q. A. Parker, O. Bienaymé, S. Roeser, B. K. Gibson, G. Gilmore, E. K. Grebel, A. Helmi, J. F. Navarro, D. Burton, C. J. P. Cass, J. A. Dawe, K. Fiegert, M. Hartley, K. S. Russell, W. Saunders, H. Enke, J. Bailin, J. Binney, J. Bland-Hawthorn, C. Boeche, W. Dehnen, D. J. Eisenstein, N. W. Evans, M. Fiorucci, J. P. Fulbright, O. Gerhard, U. Jauregi, A. Kelz, L. Mijović, I. Minchev, G. Parmentier, J. Peñarrubia, A. C. Quillen, M. A. Read, G. Ruchti, R.-D. Scholz, A. Siviero, M. C. Smith, R. Sordo, L. Veltz, S. Vidrih, R. von Berlepsch, B. J. Boyle, and E. Schilbach. The Radial Velocity Experiment (RAVE): First Data Release. *AJ*, 132:1645–1668, October 2006.
- A. R. Walker. BVI CCD photometry of galactic globular clusters. 2: M68. *AJ*, 108:555–584, August 1994.
- J. V. Wall and C. R. Jenkins. *Practical Statistics for Astronomers*. Princeton Series in Astrophysics, November 2003.
- S. Yi, P. Demarque, Y.-C. Kim, Y.-W. Lee, C. H. Ree, T. Lejeune, and S. Barnes. Toward Better Age Estimates for Stellar Populations: The  $Y^2$  Isochrones for Solar Mixture. *ApJS*, 136:417–437, October 2001.
- F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. *ApJ*, 86: 217–+, October 1937.