

Multidimensional Bayesian membership analysis of the Sco OB2 moving group

A. C. Rizzuto,^{★†} M. J. Ireland[†] and J. G. Robertson

Sydney Institute for Astronomy (SIfA), School of Physics, University of Sydney, NSW 2006, Australia

Accepted 2011 June 15. Received 2011 June 15; in original form 2011 March 22

ABSTRACT

We present a new high-mass membership of the nearby Sco OB2 association based on *Hipparcos* positions, proper motions and parallaxes, and radial velocities taken from the 2nd Catalogue of Radial Velocities with Astrometric Data (CRVAD-2). The Bayesian membership selection method developed makes no distinction between subgroups of Sco OB2 and utilizes linear models in calculation of membership probabilities. We select 436 members, 88 of which are new members not included in previous membership selections. We include the classical non-members α -Cru and β -Cru as new members as well as the pre-main-sequence stars HIP 79080 and 79081. We also show that the association is well mixed over distances of 8° on the sky, and hence no determination can be made as to the formation process of the entire association.

Key words: methods: statistical – stars: early-type – stars: kinematics and dynamics – open clusters and associations: individual: Sco-Cen.

1 INTRODUCTION

The Scorpius-Centaurus-Lupus-Crux OB Association (Sco OB2, Sco-Cen) is the nearest location with recent massive star formation. The association was first identified by Kapteyn (1914) during an investigation of the parallaxes of 319 bright OB stars in the region of sky occupied by Sco-Cen. Following this, other kinematic studies confirmed that Sco-Cen is indeed a moving group (Plaskett 1928; Blaauw 1946; Bertiau 1958; Petrie 1962; Jones 1971; de Zeeuw et al. 1999). Since its discovery, Sco-Cen has been classically divided into three distinct subgroups (see Fig. 1), Upper Scorpius (US), Upper Centaurus-Lupus (UCL) and Lower Centaurus-Crux (LCC) (Blaauw 1946), with mean parallaxes of 6.9, 7.1 and 8.5 mas respectively, or distances of 145, 143 and 118 pc (de Zeeuw et al. 1999). UCL and LCC have little interstellar material associated with them, whereas filamentary material can be observed towards US which is connected to the Ophiuchus cloud complex, a region of ongoing star formation (de Geus 1992). Photometry has demonstrated that the Ophiuchus cloud complex is on the near side of US at approximately 125 pc, and isochrone fitting gives ages for the subgroups as 5 Myr for US, 16 Myr for UCL and 16 Myr for LCC (de Geus, de Zeeuw & Lub 1989).

US has received significantly more attention than the other two subgroups of Sco-Cen, primarily due to its relatively compact size. The age spread of US members is tightly bunched around 5 Myr,

which has led to the conclusion that star formation in US was an externally triggered process (Preibisch & Mamajek 2008). The triggering effect has been identified as the shockwave created by several supernova explosions in UCL which occurred approximately 12 Myr ago. Observations of the kinematics of the large H I loops surrounding Sco-Cen suggest that the shockwaves passed US approximately 5 Myr ago, which agrees with the stellar age of the subgroup (de Geus 1992).

UCL and LCC have been subject to less study than US. This is primarily due to their relative lack of concentration on the sky and the closer overlap with the Galactic plane, which makes separation of members and field stars more difficult. Also, large portions of UCL and LCC are not observable from the Northern hemisphere (where many early investigations were conducted), and early investigation into pre-main-sequence stars focused on dark and reflection nebulae, which are less frequent in UCL and LCC (Preibisch & Mamajek 2008). Only in the last two decades have true high-quality measurements of star motions on the sky (*Hipparcos*) become available, which have allowed separation of UCL and LCC members from the field (de Zeeuw et al. 1999). The formation processes which shaped UCL and LCC are as yet unknown, but are expected to be significantly more complicated than that of US.

Elmegreen & Lada (1977) propose that OB associations are formed through successive bursts caused by shockwaves and ionization fronts in molecular clouds. OB stars drive ionization fronts into neutral material in their vicinity. In dense material, the ionization front is preceded by a shock front as it moves into a nearby molecular cloud. Elmegreen & Lada (1977) suggest that the gas accumulated between the two fronts becomes gravitationally unstable and can collapse to form stars. These new stars will then drive

[★]E-mail: aaron.rizzuto@students.mq.edu.au

[†]Present address: Department of Physics and Astronomy, Macquarie University, Sydney 2109, Australia.

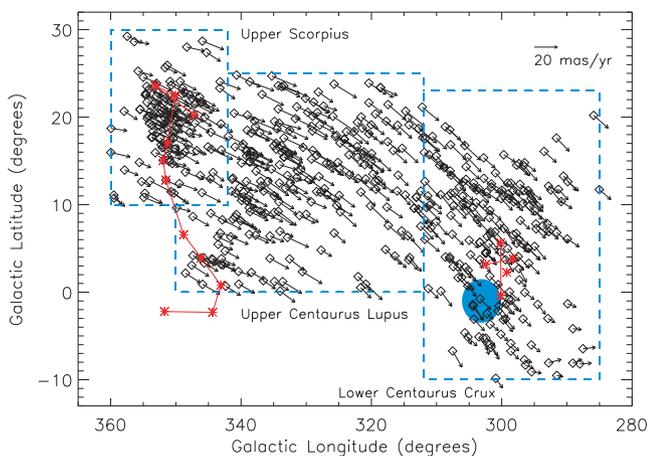


Figure 1. Positions of Sco OB2 on the sky with representative *Hipparcos* proper motion vectors. The members shown in this plot are taken from the de Zeeuw et al. (1999) membership determination, and show the windows (blue) used for each subgroup membership search. The familiar ‘Southern Cross’ and the constellation Scorpius are plotted in red, and the Coalsack nebula is shaded in blue.

the process further into the remaining molecular cloud. This model agrees with the observations that OB associations tend to display a series of distinct subgroups (Ambartsumian 1955). It is possible that Sco-Cen formed in this way.

The latest high-mass membership study of Sco-Cen was carried out by de Zeeuw et al. (1999) who identified a total of 521 members. The search combined two different selection methods, the convergent-point and ‘Spaghetti’ methods (de Bruijne 1999; Hoogerwerf & Aguilar 1999), which use proper motion and parallax to differentiate between members and non-members in distinct regions of sky corresponding to each Sco-Cen subgroup. The convergent-point method has long been a means by which to identify moving group members (Jones 1971), though it does have some clear weaknesses. The method is one-dimensional, with the selection of member stars being based only on proper motion in a specific direction. Additionally, the convergent-point method has a bias towards selecting more distant stars and stars with small proper motion. The ‘Spaghetti’ method is stronger than the convergent-point method, in that it takes into account both proper motion and parallax, though it does not constrain the radial velocity of members. This indicates that the development of a selection method which makes use of radial velocity is a clear avenue for improvement.

The Sco OB2 association is a useful and readily available astrophysical laboratory. It is the largest group of newly formed stars in close proximity to the Sun and provides an ideal testing ground for new star formation and evolution models. Despite, this Sco-Cen is still relatively poorly studied when compared to other associations such as the Pleiades, which is only a fraction of the size of Sco-Cen.

2 MULTIDIMENSIONAL BAYESIAN MEMBERSHIP SELECTION

Since the last membership selection of de Zeeuw et al. (1999), there has been a vast improvement in the available astrometric data for stars in the Sco OB2 field of the sky. The *Hipparcos* (Perryman et al. 1997b) mission data have been reduced in a more accurate way by van Leeuwen (2007), and catalogues of radial velocities

for a large number of the Sco-Cen potential members have become available (Kharchenko et al. 2007). These improvements allow one to develop a membership criterion based not only on more accurate proper motion, but also on parallax, radial velocity and galactic latitude.

In the context of the membership analysis described in this paper, the *Hipparcos* data used were that from van Leeuwen (2007) of stars with $B - V \leq 0.6$, within the area of sky bounded by $(285^\circ \leq l \leq 360^\circ)$ and $(-10^\circ \leq b \leq 60^\circ)$. These are the bright, higher mass objects in the area of sky occupied by the Sco OB2 association. The radial velocities were taken from the 2nd Catalogue of Radial Velocities with Astrometric Data (CRVAD-2; Kharchenko et al. 2007) which is a cross-linking of the All-Sky Compiled Catalogue of 2.5 million stars (Kharchenko 2001), the General Catalogue of Radial Velocities (Barbier-Brossat & Figon 2000) and a number of other, smaller radial velocity catalogues (Kharchenko et al. 2007). Measurements from the RADial Velocity Experiment (RAVE; Steinmetz et al. 2006) are not included in the CRVAD-2 catalogue and, as such, CRVAD-2 can be considered an up-to-date catalogue of stellar radial velocities in the pre-RAVE era. Given that RAVE observations of the Sco OB2 area of sky are not available, CRVAD-2 is the most complete source of radial velocities available for this work. Cross-checking was done between the *Hipparcos* and CRVAD-2 catalogues to find radial velocities for as many of the *Hipparcos* objects as possible. This resulted in 35 per cent of objects having a known radial velocity, with typical uncertainties of $1\text{--}5 \text{ km s}^{-1}$. Objects without a radial velocity were assigned a random radial velocity with an extremely large uncertainty, i.e. the radial velocity is treated as unconstrained in the Bayesian membership analysis described below.

Our membership selection uses the six physical parameters of each candidate star to decide upon a membership probability. These parameters are the three Galactic velocity components and the three position parameters: distance, Galactic longitude and Galactic latitude. In our new membership selection, a member must have velocity, Galactic latitude and distance consistent with the five parameters derived from models described below. The five parameters derived from the models (U , V , W , r and b) will all be functions of Galactic longitude.

The Galactic longitude constraint on membership is a simple cut at each end of the association. The Galactic longitude of members is cut at $l = 360^\circ$ at the US end of the association and at $l = 285^\circ$ at the LCC end. At the US end, active star formation is in play so it is a logical end point for the association. At the LCC end, membership probabilities are typically lower because of the confusion with the Galactic plane, and the lack of very bright high-mass stars at lower Galactic longitudes means there is no clear reason to extend our search further. There may of course still be associated recent star formation beyond LCC, but examining the possibility of this is beyond the scope of this study. In the following subsections, we will describe our model for the association, the computational framework and the Bayesian membership selection.

2.1 Linear models

It is important to first note that the Galactic velocities of the de Zeeuw et al. (1999) members show a clear linear trend with Galactic longitude, with the trend being most pronounced in the component of the velocity pointing out of the Galactic Centre (U) (see Fig. 2a). This trend is also present in the distance of group members, Galactic latitude and the spread of Galactic latitude of de Zeeuw et al. (1999) members. Note that the linear trend in the spread of latitudes was

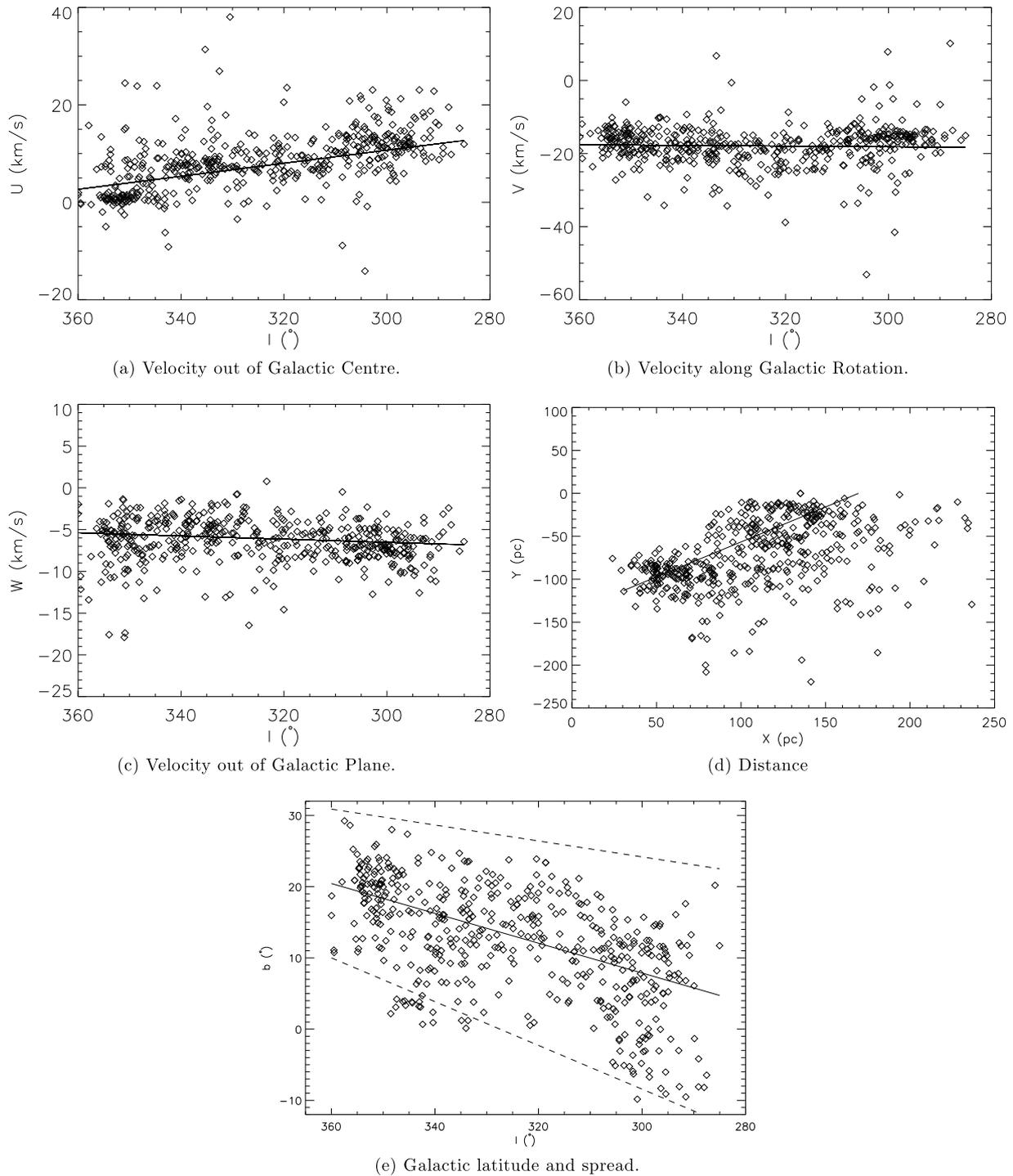


Figure 2. The Galactic velocity components, distance and Galactic latitude of the de Zeeuw et al. (1999) members. The linear trends with Galactic longitude are shown on each graph. Panel (e) shows both the linear trend in Galactic latitude and the linear trend in the spread of Galactic latitude (σ_b) as the dotted lines above and below the solid line. The origin in panel (d) is the position of the Sun, and stars with a greater distance have larger distance errors.

derived by assuming that the window sizes defining the subgroups (see boxes in Fig. 1) were a 2σ coverage of the Sco-Cen association. This produced a standard deviation of Galactic latitude for each subgroup. A linear trend against Galactic longitude was then fitted to these three values. We denote the Galactic latitude standard deviation trend as σ_b .

We characterize the distance of the group members by a line in 3D space, which removes the anthropocentric bias associated with

a simpler model where parallax is a function of galactic longitude. The line fitted to the distance of the de Zeeuw et al. (1999) group members is of the form:

$$r = (A \cos l + B \sin l)^{-1}, \quad (1)$$

where A and B are the fitted parameters and l is the Galactic longitude. Fig. 2(d) displays this fitted line and the de Zeeuw et al. (1999) members. Note that the points with larger distances also

Table 1. The linear model parameters derived for the de Zeeuw et al. (1999) members which were used. The galactic velocity (U , V , W) parameters are in units of km s^{-1} , the galactic latitude (b , σ_b) parameters are in degrees and the distance (r) parameters are in pc^{-1} .

	Slope	Intercept ($l = 0$)
U	-0.13	50.9
V	-0.009	-20.8
W	-0.019	-12.3
b	0.21	-55.0
σ_b	-0.01	45.6
	A	B
r	0.0059	-0.0072

have proportionally larger errors, and so these outlier points did not contribute to the fit in any significant way.

The trends span all three subgroups (US, UCL and LCC) suggesting that perhaps Sco OB2 can be modelled as a continuum rather than three separate subgroups, with the galactic velocity components, distance and Galactic latitude following a linear trend. The linear trend fitted to the de Zeeuw et al. (1999) results will later be used in the calculation of membership probabilities.

The IDL software package MPFIT was used to obtain the parameters of the linear trends for the five parameter space variables. The fitted parameters can be seen in Table 1. Our choice of fitting a linear trend was based on there being no clear observational distinction between the subgroups, and is not a claim of continuous star formation.

2.2 The convergent-point and expected values

2.2.1 The convergent point

The membership selection method we have developed makes use of the concept of the convergent point which was first introduced by Jones (1971) and modernized by de Bruijne (1999). The convergent point is the point on the sky at which a group of stars with a common Galactic velocity will appear to converge to and has formed the basis of one of the search methods used to populate the current membership of Sco-Cen (de Zeeuw et al. 1999). The underlying principle of our new membership selection and its relation to the convergent point are as follows. For each candidate member, expected Galactic velocity components [$U_g(l)$, $V_g(l)$, $W_g(l)$], distance (r_g) and Galactic latitude (b_g) can be calculated from the linear trends as functions of the Galactic longitude (l) of the candidate. From the expected Galactic velocities, a convergent point is calculated (l_{cp} , b_{cp}), and based on this convergent point, expected proper motions and radial velocity can be found (described below). Hence, for a given star to be considered a member, it is required to have proper motions, radial velocity and distance which are consistent with the expected values calculated from the linear models in Galactic longitude.

The significance of the convergent point to the membership selection developed in this paper relates to the velocity coordinate system of the data. The velocities provided by the *Hipparcos* and CRVAD-2 catalogues are two proper motions (equatorial coordinates) and one radial velocity, while the linear models provide Galactic velocity and parallax as a function of Galactic longitude. The concept of the

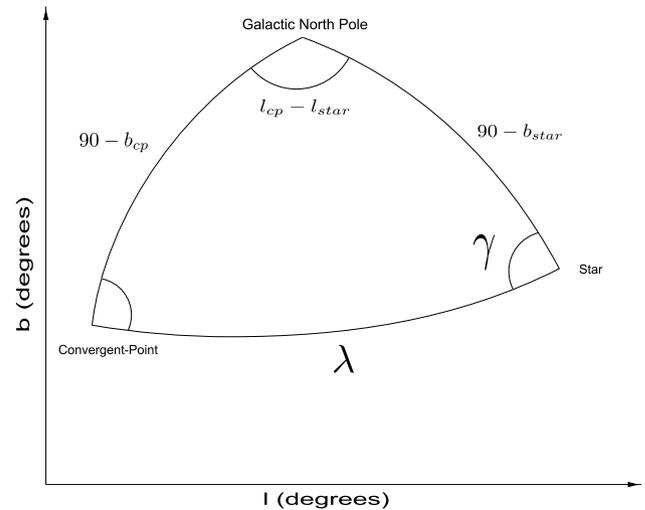


Figure 3. The triangle made from the three great-circles joining the star, convergent point and Galactic north pole. The angles γ and λ are the rotation angle and angular distance between the star and the convergent point respectively. Note that coordinates used here are Galactic coordinates, and also that an analogous diagram can be drawn for equatorial coordinates with the rotation angle θ and the same angular distance λ .

convergent point allows a direct and easily computed conversion from the model Galactic velocities to the velocity coordinate system of the available data. The convergent point can be calculated as follows:

$$\tan(l_{cp}) = \frac{V_g(l)}{-U_g(l)},$$

$$\tan(b_{cp}) = \frac{W_g(l)}{\sqrt{[U_g^2(l) + V_g^2(l)]}}, \quad (2)$$

where l_{cp} and b_{cp} are the coordinates of the convergent point on the sky, and $U_g(l)$, $V_g(l)$ and $W_g(l)$ are the three Galactic velocity components derived from the linear trends and are functions of Galactic longitude. Thus the convergent point is also a function of Galactic longitude.

Once the convergent point for a given candidate member is found, the proper motions provided by the *Hipparcos* catalogue are rotated into a new coordinate system with proper motion along the great circle joining the star and the convergent point and proper motion along the great circle perpendicular to this. The angle of rotation (γ) is shown in Fig. 3 (in Galactic coordinates) and is calculated using trigonometric identities. Note that the rotation angle can also be calculated in equatorial coordinates and will be referred to as θ . The equatorial rotation angle θ was used to rotate the *Hipparcos* proper motions into the new coordinate system described above. The rotation is given by the following equation:

$$\begin{pmatrix} \sin \theta & \cos \theta \\ -\cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} \mu_\alpha \\ \mu_\delta \end{pmatrix} = \begin{pmatrix} \mu_\parallel \\ \mu_\perp \end{pmatrix}, \quad (3)$$

where μ_α and μ_δ are the two *Hipparcos* proper motions, μ_\perp is the proper motion perpendicular to the direction of the convergent point and μ_\parallel is the proper motion towards the convergent point.

The use of the convergent point in our method allows the parameters of the models to be easily converted to the framework of the observational data, rather than manipulating the data into a form comparable to the models. The method also allows uncertainties

to be calculated via a simple rotational transformation of the *Hipparcos* covariance matrix rather than complicated conversions to Galactic velocity coordinates. It is also important to stress that the convergent point is a function of Galactic longitude (although indirectly) and hence each star will have a different convergent point. Hence, we are not using the convergent point in the traditional sense as the point at which association members converge on the sky, but as a means of constructing a new coordinates system which allows us to convert from Galactic velocities to proper motions and radial velocity. Furthermore, this convergent-point-based coordinate system allows a simple membership criterion for μ_{\perp} : this quantity should be small for a member. This criterion is the basis for the classical convergent-point selection method used by Jones (1971).

2.2.2 Expected values

The basis of this new membership selection is that for a candidate star to be a member it must have velocity, distance and position which are consistent with some expected values. The expected values of the five parameters are derived from the linear model described above, and vary for each candidate star with Galactic longitude. The expected Galactic latitude and parallax are taken directly from the linear models, while the expected proper motions and radial velocity require conversion from the linear models for Galactic velocity components [$U_g(l)$, $V_g(l)$, $W_g(l)$].

To convert the expected Galactic velocity from the linear models, unit vectors in the direction of μ_{\parallel} and the radial velocity of each star are needed. These will be denoted by \hat{p} and \hat{r} respectively and can be calculated using spherical trigonometry:

$$\begin{aligned}\hat{p} &= \cos \gamma \hat{b} + \sin \gamma \hat{l}, \\ \hat{r} &= (\cos l \sin b, \sin l \sin b, \cos l),\end{aligned}\quad (4)$$

where \hat{l} and \hat{b} are the Galactic longitude (l) and Galactic latitude (b) unit vectors. The expected Galactic velocity [$\tilde{u}_g(l)$] is then projected in the direction of these two unit vectors yielding the expected velocity along the great circle joining the convergent point and the star, and the expected radial velocity, denoted $v_{\parallel g}$ and $v_{r g}$.

$$\begin{aligned}v_{\parallel g} &= \hat{p} \cdot \tilde{u}_g(l), \\ v_{r g} &= \hat{r} \cdot \tilde{u}_g(l).\end{aligned}\quad (5)$$

In order to compare the expected parallel velocity ($v_{\parallel g}$) to the proper motion in the direction of the convergent point for a given star, it is required that $v_{\parallel g}$ be converted into proper-motion units. This was done by multiplying by the parallax (Π) and $M = 4.74$ km yr s^{-1} which is the ratio of the number of seconds in a Julian year and one astronomical unit in kilometres, yielding the expected parallel proper motion for the given star. This allows ideal membership to be described by equation (6). Note that the expected quantities will vary depending on the position of the star in question, and so will be different for each possible member:

$$\left\{ \begin{array}{l} \mu_{\perp} = 0 \\ \mu_{\parallel} = \Pi M v_{\parallel g}(l_{cp}, b_{cp}) \\ v_r = v_{r g}(l_{cp}, b_{cp}) \\ r = r_g(l) \\ b = b_g(l) \end{array} \right\}.\quad (6)$$

The above five conditions describe the expected parameters of an ideal member star at a given Galactic longitude. It is expected that there is a velocity dispersion around the expected Galactic velocity of each star due to gravitational interaction of Sco-Cen members. The magnitude of the velocity dispersion used in this selection method was 3 km s^{-1} (Kraus & Hillenbrand 2008).

2.3 Preliminary exclusions

Using the expected velocities expressed above and the errors associated with the *Hipparcos* and radial velocity data, a membership probability can be calculated for each star, using Bayes' factors and Monte Carlo integration. Before the probabilities are calculated, a simple and preliminary yes/no selection was made to initially reject stars which would have low membership probabilities. This was done by finding the squared, inverse-variance weighted difference between each velocity component and its respective expected velocity (equation 7), and using a 3 degrees of freedom χ^2 probability density function to calculate the probability, η , of the sum being greater than its calculated value (equation 8). Stars for which a radial velocity was not available in the Karchenko catalogue were treated with a 2 degrees of freedom χ^2 probability density function, with the radial velocity term $R = 0$ (see equation 7). Ideally, model likelihood ratios would be calculated for each star in the field; however, stars with a large probability η will have a small Bayes' factor and will be excluded by the Bayesian process anyway :

$$\begin{aligned}T &= \frac{\mu_{\perp}^2}{\sigma_{\mu_{\perp}}^2 + \sigma_{\text{int}^*}^2}, \\ P &= \frac{(\mu_{\parallel} - M v_{\parallel g} \Pi)^2}{\sigma_{\mu_{\parallel}}^2 + (M \sigma_{\Pi} v_{\parallel g})^2 + \sigma_{\text{int}^*}^2}, \\ R &= \frac{(v_r - v_{r g})^2}{\sigma_{v_r}^2 + \sigma_{\text{int}}^2},\end{aligned}\quad (7)$$

$$\eta = \frac{1}{2^{3/2} \Gamma(3/2)} \int_{T+P+R}^{\infty} x^{1/2} e^{-(x/2)} dx.\quad (8)$$

The cut-off value used for η for stars which progressed to the Bayesian analysis was determined by testing whether setting η lower would change the number of stars chosen by the Bayesian analysis at a 50 per cent threshold. A value of 0.8 was used for η , which was well above the condition described, but still able to significantly reduce the time required for computation. Within Sco-Cen, there is a common velocity among members and a dispersion away from the common velocity caused by gravitational interaction. In equation (7) we have included internal velocity dispersion σ_{int} as a source of deviation from the expected velocities. Conversion to proper motion units is required for μ_{\parallel} and μ_{\perp} using the conversion factor $M = 4.74$ km yr s^{-1} mentioned above and the mean parallax of the de Zeeuw et al. (1999) members. The converted velocity dispersion is denoted σ_{int^*} .

This preliminary selection did not produce membership probabilities, but simply functioned to filter out stars which could not be members based on their velocity characteristics. This is why we do not present a comprehensive analysis of the stars selected, though a brief account is as follows. Approximately four times the number of stars in the de Zeeuw et al. (1999) membership were included by this preliminary selection (2051 stars), which is expected since it was designed to let through all but the most clear non-members (5473 excluded). Note that 15 de Zeeuw et al. (1999) members

were rejected, almost all of which were due to a radial velocity inconsistent with membership, again, this is expected as radial velocity was not used by de Zeeuw et al. (1999). A further discussion and additional details on these stars will be given in the following sections.

2.4 Bayesian membership probabilities

To find a membership probability for each star in the field of Sco OB2, Bayesian model testing was used. Bayes factors can be used to calculate the likelihood ratio (R) between two models, in the case of this membership selection, the models are that a given star is a member of Sco-Cen, denoted M_g , and that a star is a member of the field and not a member of Sco-Cen, denoted M_f . The probability of a model given the data is $R/(R+1)$. The model likelihood ratio (R) is given by equation (9) (Sivia & Skilling 2007), where D denotes the data, the vertical bar denotes ‘given’ and P denotes probability:

$$R = \frac{P(M_g|D)}{P(M_f|D)}. \quad (9)$$

Using Bayes’ theorem (Sivia & Skilling 2007), the model likelihood ratio can be expressed as

$$R = \frac{P(M_g) P(D|M_g)}{P(M_f) P(D|M_f)} = \frac{P(M_g)}{P(M_f)} K, \quad (10)$$

where K is the Bayes’ factor and can be written as

$$K = \frac{P(D|M_g)}{P(D|M_f)} = \frac{\int P(\phi_g|M_g)P(D|\phi_g)d\phi_g}{\int P(\phi_f|M_f)P(D|\phi_f)d\phi_f}, \quad (11)$$

where ϕ_g and ϕ_f are a set of parameters $\{U, V, W, r, b\}$.

The components which make up the five-dimensional integrals in equation (11) can be separated into three uncoupled integrals:

$$P(D|M_g) = \int_{-\infty}^{\infty} P(b_g|M_g)P(D|b_g)db_g \int_{-\infty}^{\infty} P(r_g|M_g)P(D|r_g)dr_g \int P(\{U\}|M_g)P(D|\{U\})d\{U\}, \quad (12)$$

where $\{U\}$ is the set of three group Galactic velocity components $\{U_g, V_g, W_g\}$. These cannot be separated because the data do not constrain these components individually. Note that a completely analogous equation can be written for the field integral [denominator of equation (11)]. First, in the Galactic latitude (b) integral, the term $P(D|b_g)$ is the delta function $\delta(b - b_g)$ because the uncertainties in the *Hipparcos* positions are trivially small. Hence the Galactic latitude integral is analytically computed to be simply $P(b|M_g)$, which is a Gaussian with mean and standard deviation taken from the linear models, and truncated at the edge of our search area ($-10^\circ < b < 60^\circ$). Note that since the mean and standard deviation are function of Galactic latitude, the Gaussian will be different for each star. Similarly, for the field, $P(D|M_f)$ is a Gaussian with a mean of zero and standard deviation taken from the spread of Galactic latitudes in the *Hipparcos* catalogue. Again, this was truncated at the edges of our search area.

A more complicated issue is that of the distance (r) integrals. It is important to note that there is a population effect in the *Hipparcos* catalogue towards more distant stars. This is expected as there are intrinsically more stars per unit distance from the Sun. This effect will thus be intrinsic to both group and field distance distributions. Hence, the field distribution will be defined by this

population effect, and the group distribution will be treated as a Gaussian superimposed on the population effect. The probability distribution of the population effect was determined by binning all the *Hipparcos* stars in our search area into 5 pc bins and calculating the fraction of stars per bin. The population distribution will be denoted by $B(r)$. The terms in the distance integral components for the field are then given by the following:

$$P(r_f|M_f) = B(r), \quad (13)$$

$$P(D|r_f) = \frac{1}{\sqrt{2\pi\sigma_\Pi^2}} \exp\left[-\frac{\left(\frac{1}{r_f} - \Pi\right)^2}{2\sigma_\Pi^2}\right]. \quad (14)$$

The integral of these two components is computed numerically by first defining the 5 pc distance bins and calculating the fractional number of stars per bin. This produces a set of r_f values and $P(r_f|M_f)$, their weightings (one for each bin). For each of these r_f values $P(D|r_f)$ is calculated, multiplied by its corresponding weighting and then totalled to yield the integral. The corresponding group distance distributions are defined in a similar way:

$$P(r_g|M_g) = P(r_g|M_f)C \exp\left\{-\frac{[r_g - \bar{r}_g(l)]^2}{2\sigma_{\bar{r}_g}^2}\right\}, \quad (15)$$

$$P(D|r_g) = \frac{1}{\sqrt{2\pi\sigma_\Pi^2}} \exp\left[-\frac{\left(\frac{1}{r_g} - \Pi\right)^2}{2\sigma_\Pi^2}\right], \quad (16)$$

where $\bar{r}_g(l)$ is the distance taken from the linear trend with Galactic longitude, and C is a normalization constant which is computed numerically. $\sigma_{\bar{r}_g}$ is the standard deviation of the de Zeeuw et al. (1999) members used to fit the linear distance trend. The assumption that the de Zeeuw et al. (1999) members describe the depth of the association well is valid because distance was not constrained in the de Zeeuw et al. (1999) membership selection. Note that, as described above, $P(r_g|M_g)$ is a Gaussian in distance superimposed on the population distribution, and the factor $P(D|r_g)$ will be different for each star since $\bar{r}_g(l)$ is a function of Galactic longitude l . The group distance integral is computed in the same way as the field integral.

Having separated the Galactic longitude and distance components from the integral in equation (11), the remaining integral is the contributions from the Galactic velocity components. The factors involved in this integral are described below.

$P(\phi_g|M_g)$ (the probability of the parameters given the model that a star is in the group) is modelled as the product of three Gaussian distributions and is shown in equation (17). The means of the Gaussians are the expected values of the parameters derived from the linear trends in Galactic longitude $[U_g(l), V_g(l)$ and $W_g(l)]$. The spread of these parameters around the expected values is only influenced by the internal velocity dispersion of the association, thus the Gaussians have a standard deviation equal to σ_{int} :

$$P(\phi_g|M_g) \propto \exp\left\{-\left[\frac{[U - U_g(l)]^2}{2\sigma_{\text{int}}^2} + \frac{[V - V_g(l)]^2}{2\sigma_{\text{int}}^2} + \frac{[W - W_g(l)]^2}{2\sigma_{\text{int}}^2}\right]\right\} \quad (17)$$

$$P(\phi_f|M_f) \propto \exp \left\{ - \left[\frac{(U - U_f)^2}{2\sigma_{U_f}^2} + \frac{(V - V_f)^2}{2\sigma_{V_f}^2} + \frac{(W - W_f)^2}{2\sigma_{W_f}^2} \right] \right\}. \quad (18)$$

The distribution for the field, $P(\phi_f|M_f)$ (equation 18), is defined in an analogous way. However, the Galactic velocities are taken from the Galactic thin disc with values of $(U_f, V_f, W_f) = (9.0, -6.9, -7.0)$ km s⁻¹ and standard deviations of $\sigma_{U_f} = 19.8$, $\sigma_{V_f} = 12.8$, and $\sigma_{W_f} = 8.0$ km s⁻¹ for A-type stars (Robin et al. 2003).

The probability of the data given the parameters from the group, $P(D|\phi_g)$, is also defined as the product of three Gaussian distributions, though, rather than Galactic velocity, the new coordinates calculated from the convergent point using the projections in equation (5) were used, i.e. μ_{\parallel} , μ_{\perp} and v_r .

The final remaining probability distribution to define is $P(D|\phi_f)$, the probability of the data given the model that the star is in the field. This was also modelled as three Gaussians with expected values of μ_{\perp} , μ_{\parallel} and v_r calculated in the same way as was done for the group. Note that this involves finding a convergent point for various points in the parameter space. One may ask why the field stars have a convergent point, and the answer is that the apparent motion of the field stars is dominated by the motion of the Sun and the rotation of the galactic disc, so a convergent point is in fact present for a star at rest with respect to the Galactic disc. However, in this case the convergent point is really an abstract coordinate system used to find μ_{\parallel} and μ_{\perp} . Hence, expected values are calculated for each star as described in the previous section. These two probability functions are shown below in equations (19) and (20):

$$P(D|\phi_g) \propto \exp \left\{ - \left[\frac{\mu_{\perp}^2}{2\sigma_{\mu_{\perp}}^2} + \frac{[\mu_{\parallel} - \Pi M v_{\parallel g}(l)]^2}{2(\sigma_{\mu_{\parallel}}^2 + [M v_{\parallel g}(l) \sigma_{\Pi}]^2)} + \frac{[v_r - v_{r_g}(l)]^2}{2\sigma_{v_r}^2} \right] \right\} \quad (19)$$

$$P(D|\phi_f) \propto \exp \left\{ - \left[\frac{\mu_{\perp}^2}{2\sigma_{\mu_{\perp}}^2} + \frac{(\mu_{\parallel} - A\mu_{\parallel f}\Pi)^2}{2[\sigma_{\mu_{\parallel}}^2 + (A\mu_{\parallel f}\sigma_{\Pi})^2]} + \frac{(v_r - v_{r_f})^2}{2\sigma_{v_r}^2} \right] \right\}. \quad (20)$$

The above was implemented using a Monte Carlo scheme to compute the integrals in equation (11). 10^5 samples were used for the group integral, and 10^6 samples were used for the field to adequately sample the larger parameter space of the field distribution. A brief overview of the process is as follows. From the model distributions (M_g and M_f), a set of randomly sampled parameters are taken to form ϕ_g and ϕ_f . M_g is in fact a normal distribution in each parameter (U, V, W, r, b), with means taken from the linear trends in Galactic longitude. Hence, $\phi_g = \{U, V, W, r, b\}$ is a function of Galactic longitude and will be different for each star. From each parameter sample, a convergent point is calculated and the expected values of μ_{\perp} , μ_{\parallel} and v_r are found with respect to each convergent point. Once this is done, $P(D|\phi_g)$ and $P(D|\phi_f)$ can be found and the integral can be calculated. Note that the integral is over the parameters U, V and W , and not over Galactic longitude; however, these parameters are functions of Galactic longitude for the group integral.

The prior used in the Bayesian selection method, $P(M_g)/P(M_f)$, was generated in an iterative manner. The first prior was the ratio of the number of member stars in the search area to the number of non-member stars in the de Zeeuw et al. (1999) catalogue, with $B - V$ colour less than 0.6 mag. An A-type star with $B - V$ of 0.6 mag at the mean distance of the Sco-Cen association is expected to appear to be 9th magnitude in the visual. Beyond this magnitude the completeness of the *Hipparcos* data greatly declines and so selection of fainter members is not feasible. Subsequent priors were taken as the ratio of the total of the probabilities of membership and the total number of stars in the field. This was repeated until the prior converged (10 iterations). The prior iteration is described in the following equation, with p_n being the membership probabilities, P_k being the k th prior and N_f being the number of stars in the field:

$$P_k = \frac{P(M_g)}{P(M_f)} = \frac{\sum_n^{N_f} p_n}{N_f - \sum_n^{N_f} p_n}. \quad (21)$$

Once the prior converged, stars with membership probabilities of 50 per cent or greater were denoted members of the Sco-Cen association, which has been the historical convention.

3 THE NEW *Hipparcos* MEMBERSHIP OF SCO-CEN OB2

The membership selection method developed in this paper was used to produce an improved membership list for the Sco-Cen OB association. A full list of member stars and their probabilities can be seen in Appendix A. Fig. 4 displays the selected members on the sky with proper motion vectors. Members of Sco-Cen have now been identified in a much larger area on the sky, with new members at significant distances from the central clustering of the association. While there is clear distinction between US and the rest of the association, UCL and LCC now show consistent density and even clumping of members, including in regions that were previously considered sparser boundaries between subgroups.

The selection has produced 436 Sco-Cen members, 348 of which were previous members (de Zeeuw et al. 1999) with the rest being new members. 93 de Zeeuw et al. (1999) members have been rejected as members of Sco-Cen, 11 on the basis of parallax, 29 from radial velocity and 53 from *Hipparcos* proper motions. Of the latter 53, only nine were excluded based on perpendicular proper motion (μ_{\perp}), which was a major basis of the de Zeeuw et al. (1999) selection. The exclusion of previous members based on *Hipparcos* proper motions, especially the exclusions based on μ_{\perp} , emphasizes the significant effect of the newly reduced data in the membership selection. Nine of the rejected previous members (HIP 62002, 64661, 68534, 72099, 74449, 77315, 79610, 80208 and 81316) had membership probabilities greater than 40 per cent, but did not have accurate or available radial velocity measurements. It is likely that these stars can indeed be shown to be members of Sco-Cen with accurate radial velocity measurements.

Importantly, the membership selection has established the classical non-members α -Cru and β -Cru as Sco-Cen stars with membership probabilities of 66 and 73 per cent, respectively. This is the first time these stars have been included as members of Sco-Cen. It is important that they are included as they are bright B-type stars with the correct position and parallax and age to have formed as Sco-Cen members.

Also, we select the classical members η -Cen and μ^1 -Sco which were not included in the last membership survey, and we confirm the membership of HIP 59449, as suggested but unconfirmed as a Sco-Cen member (de Zeeuw et al. 1999).

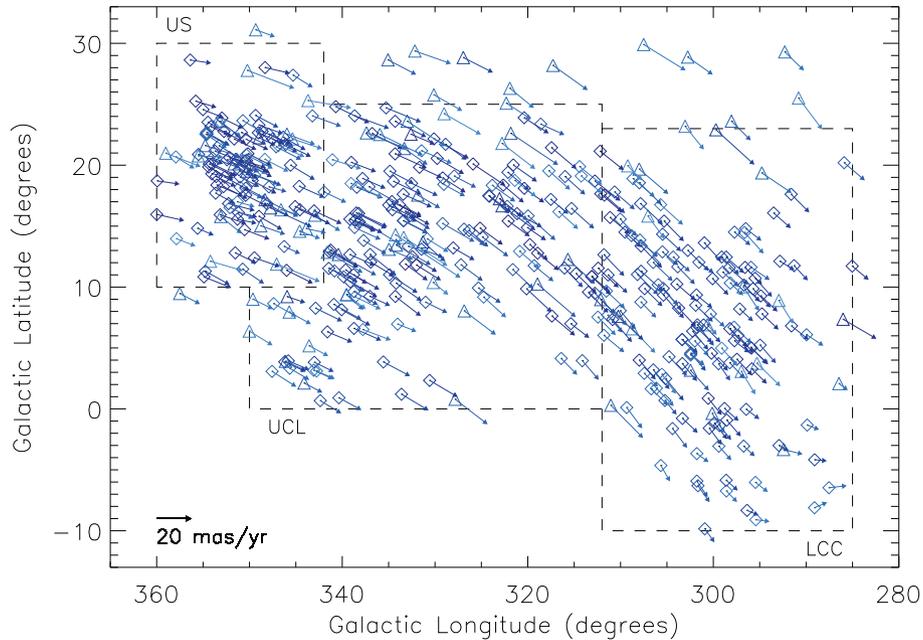


Figure 4. Position and proper motion vectors of the new Sco-Cen members. Note the significantly expanded region of sky within which members have been selected. Diamond shaped points are stars which were included in the last membership (de Zeeuw et al. 1999), and triangular points are new members which were not present in the last membership compilation. Darker blue indicates higher membership probability.

De Zeeuw et al. (1999) suggested, but was unable to confirm, that the pre-main-sequence stars HIP 79080 and HIP 79081 were Sco-Cen members. These stars have now been selected as members with probabilities of 51 and 63 per cent. HIP 78094, another pre-main-sequence star, has consistent proper motions for membership, but cannot be confirmed as a member due to a poor parallax measurement. We also confirm that the classical member δ -Sco has proper motions which are inconsistent with group membership. This star was also excluded from Sco-Cen by de Zeeuw et al. (1999), with the explanation that its binarity, with period of ~ 20 yr, could be responsible for producing a proper motion inconsistent with the group.

4 DISCUSSION

The new membership described here does not require the segmentation of Sco-Cen into the three classical subgroups in order for members to be chosen. Instead Sco-Cen has been treated as a continuum using a linear model to select members and reject non-members. The use of a linear model has removed the need for hard-edged boundaries around selection areas, as seen in the previous membership (de Zeeuw et al. 1999), and has allowed the area in which members can be found to be slightly expanded. Furthermore, the linear trend in the component of the Galactic velocity pointing out of the Galactic Centre (U) demonstrates that the association is indeed expanding. This is expected since OB associations are considered to be loosely bound and is consistent with the idea of a common origin for the entire association.

Fig. 5 displays photometry taken from the Tycho catalogue (Perryman et al. 1997a) in a colour-magnitude diagram of the new Sco-Cen members. The Tycho photometry has been converted to standard UBV using the table found in Bessell (2000). We have plotted three Padova isochrones (Girardi et al. 2002) at 4, 10 and 20 Myr ages which have been reddened to match the Tycho pho-

tometry using the following equation (Aumer & Binney 2009):

$$E(B - V) = \begin{cases} 0 & d < 70 \text{ pc} \\ 0.47(d - 70 \text{ pc})/1 \text{ kpc} & d > 70 \text{ pc}. \end{cases} \quad (22)$$

The value of $E(B - V)$ was calculated to be 0.027 mag and the isochrones were reddened by this amount. From the Hertzsprung–Russell diagram, it can be seen that the youngest stars in Sco-Cen are approximately 6 Myr old, while the older stars appear to be up to 20 Myr old. This is in agreement with historical estimates of the age of the association (Blaauw 1964). Also, US, the youngest section of Sco-Cen, is expected to have formed approximately 5 Myr ago (Preibisch et al. 2002), which agrees with the age estimate of the youngest stars in the new membership. Note that the large spread of members around the isochrones in Fig. 5 is due to both the parallax measurement error, the effects of unresolved binarity on the photometry, and the age spread. There are five stars in Fig. 5 which appear to be above the main sequence, these stars are listed in Table 2. These five stars, based on their spectral types, are significantly more reddened than the rest of the new association members. This could be due to the impingement of the ρ -Oph clouds between the Sun and Sco-Cen (de Geus 1992).

A two-point correlation function was also applied to the new members of Sco-Cen for the purpose of examining the effects of internal velocity dispersion on mixing the primordial fractal structure of the association. This meant totalling the number of other Sco-Cen stars within certain distance bins. The distance bins were $(0^\circ, 0:125)$, $(0:125, 0:5)$, $(0:5, 1^\circ)$, $(1^\circ, 2^\circ)$, $(2^\circ, 4^\circ)$ and $(4^\circ, 8^\circ)$. Each bin was then area-normalized. A histogram of the two-point correlation function can be seen in Fig. 6. At small separation scales bound binary pairing produces a steep power law (Kraus & Hillenbrand 2008). Beyond separations of half a degree, random motions appear to blur out any primordial group structure and produce a shallow slope in the diagram. In this study, it is seen that structure on the order of 8° is blurred out by the group’s velocity dispersion.

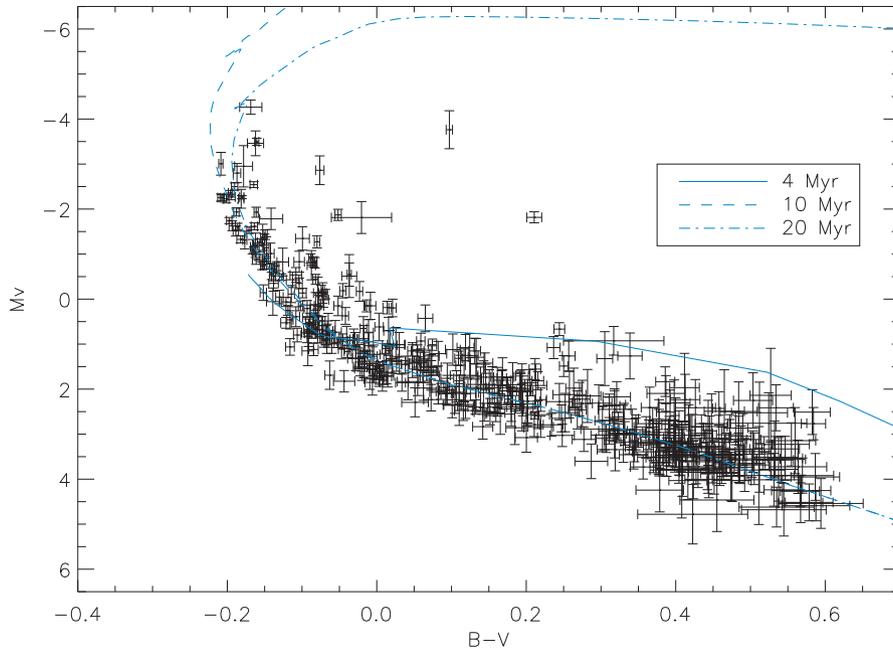


Figure 5. Colour–magnitude diagram of the Sco-Cen members selected by the new method with Padova isochrones for 4 Myr (solid), 10 Myr (dashed) and 20 Myr (dot–dashed). Note the two isolated stars, these are shifted due to differential reddening caused by the ρ -Oph cloud complex which impinges on parts of US.

Table 2. List of stars separated from the main sequence in Fig. 5.

HIP	l	b	<i>Hipparcos</i> spectral type
78820	353.2	23.6	B0.5V
78933	352.7	22.8	B1V
79374	354.6	22.7	B2IV
80112	351.3	17.0	B1III
80569	357.9	20.7	B2Vne

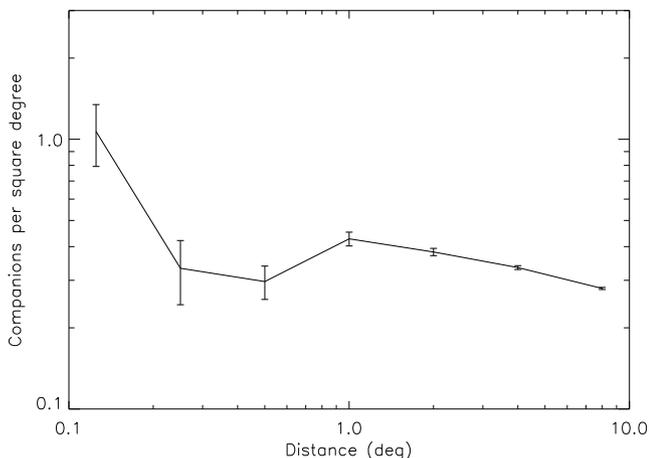


Figure 6. A logarithmic histogram of the two-point correlation function applied to the new Sco-Cen membership list.

This implies that hard-edged boundaries between the subgroups of Sco-Cen cannot be defined in a non-arbitrary way. Additionally, this means that it cannot be determined if Sco-Cen formed via the shock-front mechanism of Elmegreen & Lada (1977), as evidence of separated structures has been effectively mixed out.

4.1 Future work

The new membership of Sco-Cen presented here is for the bluer and higher mass population of Sco-Cen, with a colour $B - V < 0.6$. These are the brightest and most well-known stars in Sco-Cen. It has been proposed that there are many thousands of lower mass stars which are members of Sco-Cen but which are yet to be identified. The membership selection method developed here can be directly applied to these lower mass stars once the required data become available. The only difficulty is deciding upon the use of Bayesian prior. In this study, we have used the de Zeeuw et al. (1999) membership as a starting point in defining the prior which resulted in the new membership. There is currently no equivalent for the fainter, lower mass members. One possibility is to extrapolate the lower mass population based upon the higher mass population of the new membership in Appendix B, and use this extrapolation as a new prior. This would involve the use of accurate models to provide a colour–mass relation for young stars. Along a similar line, spectroscopic techniques can be used to identify lower mass Sco-Cen members. This has been done for a small region of Sco-Cen by Preibisch et al. (2002) using the multi-object spectrograph 2dF, but can be greatly extended with the use of an instrument such as 6dF at the UK Schmidt Telescope which is better matched to the large area of sky covered by the association. As a final point on future membership work, the membership selection method developed here, with some adaptation, can be applied to other OB associations and moving groups, as was done by de Zeeuw et al. (1999) with their selection method.

Age estimation for the various regions of Sco-Cen has traditionally been achieved through isochrone fitting. However, due to parallax uncertainties and the difficulty in determining the temperature of individual stars accurately, previous age estimates have typically been based on a small number of stars (Preibisch et al. 2002). Determination of dynamical masses for a number of Sco-Cen targets can allow a more accurate age estimate, and more precise dating across the association.

5 ACKNOWLEDGMENTS

This project was supported by the Australian Research Council (DP0878674). We thank Adam Kraus for helpful discussions on an early draft, and we thank the anonymous referee for thoughtful suggestions.

REFERENCES

- Ambartsumian V. A., 1955, *The Obser.*, 75, 72
 Aumer M., Binney J. J., 2009, *MNRAS*, 397, 1286
 Barbier-Brossat M., Figon P., 2000, *A&AS*, 142, 217
 Bertiau F. C., 1958, *ApJ*, 128, 533
 Bessell M. S., 2000, *PASP*, 112, 961
 Blaauw A., 1946, *Publ. Kapteyn Astron. Lab. Groningen*, 52, 1
 Blaauw A., 1964, *ARA&A*, 2, 213
 de Bruijne J. H. J., 1999, *MNRAS*, 306, 381
 de Geus E. J., 1992, *A&A*, 262, 258
 de Geus E. J., de Zeeuw P. T., Lub J., 1989, *A&A*, 216, 44
 de Zeeuw P. T., Hoogerwerf R., de Bruijne J. H. J., Brown A. G. A., Blaauw A., 1999, *AJ*, 117, 354
 Elmegreen B. G., Lada C. J., 1977, *ApJ*, 214, 725
 Girardi L., Bertelli G., Bressan A., Chiosi C., Groenewegen M. A. T., Marigo P., Salasnich B., Weiss A., 2002, *A&A*, 391, 195
 Hoogerwerf R., Aguilar L. A., 1999, *MNRAS*, 306, 394
 Jones D. H. P., 1971, *MNRAS*, 152, 231
 Kapteyn J. C., 1914, *ApJ*, 40, 43
 Kharchenko N. V., 2001, *Kinematika i Fizika Nebesnykh Tel*, 17, 409
 Kharchenko N. V., Scholz R., Piskunov A. E., Röser S., Schilbach E., 2007, *Astron. Nachr.*, 328, 889
 Kraus A. L., Hillenbrand L. A., 2008, *ApJ*, 686, L111
 Perryman M. A. C. et al., 1997a, *ESA SP-1200, The Hipparcos and TYCHO Catalogues*. ESA Publ., Noordwijk
 Perryman M. A. C. et al., 1997b, *A&A*, 323, L49
 Petrie R. M., 1962, *MNRAS*, 123, 501
 Plaskett J. S., 1928, *J. R. Astron. Soc. Canada*, 22, 111
 Preibisch T., Mamajek E., 2008, in Reipurth B., ed., *The Handbook of Star Forming Regions*, Vol. II. *Astron. Soc. Pac.*, San Francisco, p. 235

- Preibisch T., Brown A. G. A., Bridges T., Guenther E., Zinnecker H., 2002, *AJ*, 124, 404
 Robin A. C., Reylé C., Derrière S., Picaud S., 2003, *A&A*, 409, 523
 Sivia D. S., Skilling J., 2007, *Data Analysis: A Bayesian Tutorial*, 2nd edn. Oxford Univ. Press, Oxford
 Steinmetz M. et al., 2006, *AJ*, 132, 1645
 van Leeuwen F., 2007, *Astrophysics and Space Science Library*, Vol. 350, *Hipparcos, the New Reduction of the Raw Data*. Springer, Heidelberg

APPENDIX A: MEMBERSHIP TABLE

Table A1. Membership list and probabilities. The full table is available as Supporting Information with the online version of the article.

<i>Hipparcos</i> number	Probability (per cent)
50520	62
50847	62
52357	82
52742	61
53701	68
53771	91

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table A1. Membership list and probabilities.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.