Carbon-enhanced dwarf stars are predominantly a halo population

Jay Farihi [1]* Jason L. Sanders [1] Sophia Lilleengen [2,3] Lewis J. Whitehouse [1] and Denis Erkal [3]

[1] *Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*
[2] *Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK*
[3] *School of Mathematics and Physics, University of Surrey, Guildford, GU2 7XH, UK*

ABSTRACT

This paper reports a Galactic kinematical and dynamical analysis of 1003 main-sequence carbon stars. The sample is drawn from the Sloan Digital Sky Survey, and cross-matched with *Gaia* DR3 to obtain 6-dimensional positions and velocities using a Bayesian framework. The study provides the first reliable distances for a large sample of dwarf carbon stars, which are then analysed using both space motions and actions. The results are combined with dynamical equilibrium models for the three primary Galactic components to assign membership, finding that around 60 per cent belong to the halo, and over 30 per cent originate in the thick disc. Therefore, the results indicate dwarf carbon stars are dominated by a metal-poor halo population, and are thus an excellent resource for stellar archaeology. These stars remain on the main sequence and are relatively nearby, but atmospheric modelling is challenged by their cool effective temperatures and strong molecular features. In light of this, efforts should be made to improve C/O > 1 atmospheric modelling, as the subset of low-mass dwarf carbon stars may numerically dominate the Galactic population of carbon-enriched, metal-poor stars.

Key words: stars: carbon – stars: chemically peculiar – stars: kinematics and dynamics.

1 INTRODUCTION

Ancient, metal-poor stars that were born in the early Galaxy and Universe provide important constraints on the formation and evolution of the first stars, and constrain Galactic chemical enrichment over time (see Beers & Christlieb 2005; Frebel & Norris 2015, and references therein) that eventually produced the necessary ingredients for the solar system (e.g. Jura, Xu & Young 2013; Young 2014).One of the prime challenges in the study of such ancient stars is the identification of their metal poverty, which is gradually being addressed through large surveys using spectroscopy (Christlieb, Wisotzki & Graßhoff 2002; Keller et al. 2007), as well as photometry (York et al. 2000; Starkenburg et al. 2014).

Among the myriad classifications for metal-poor stars, and their corresponding metallicities that span several decades, there is a well-established trend of carbon enrichment (e.g. Lucatello et al. 2005; Yong et al. 2013; Arentsen et al. 2022). While the origin of these carbon-enhanced metal-poor (CEMP) stars remains an outstanding problem in stellar evolution (Frebel & Norris 2015; Yoon et al. 2016), there are essentially two possibilities: mass transfer from a carbon-rich, evolved stellar companion (e.g. Abate et al. 2015), or carbon within natal clouds from which the stars formed (e.g. Cooke et al. 2011). As the number of metal-poor stars identified through continuing observational efforts increases, so too will the number of stars with carbon enhancement.

Dwarf carbon (dC) stars are late-type main-sequence stars with C/O > 1, with spectra often similar to carbon-rich, asymptotic giant branch stars (Dahn et al. 1977). Because carbon is synthesized within the cores of solar- and low-mass stars during their main-sequence lifetime, and only later brought to the stellar surface during the third dredge-up (Iben 1965), the occurrence of these stars cannot readily be accounted for with standard, single-star evolution theories. Intriguingly, the first three confirmed carbon dwarfs have unambiguous halo kinematics and Hertzsprung–Russell (HR) diagram positions well below the main sequence (Harris et al. 1998). Among these visitors from the halo, the prototype dC star, G77-61, is currently at 73.1 pc distance ($\varpi = 13.7 \pm 0.1$ mas; Gaia Collaboration 2023), but is extremely metal-poor ([Fe/H] = −4.0; Plez & Cohen 2005), suggesting a chemodynamical connection to the known CEMP stars located within the Galactic halo. Thus, at least some fraction of dC stars are CEMP stars by definition.

There are approximately 1500 confirmed and candidate dC stars, mostly identified without kinematical bias through large spectroscopic surveys (Green 2013; Li et al. 2024). Although these are sizable samples, they essentially lack characterization such as masses, effective temperatures, metallicities, and binary fractions, with only a handful of exceptions such as the dC stars detected and monitored astrometrically by the US Naval Observatory for over two decades (Harris et al. 2018). For example, only the bright $(g, r) = (14.6, 13.3)$ AB mag, prototype G77-61 has a published T_{eff} and abundances for elements such as C, N, Na, Mg, and Fe (Gass, Liebert & Wehrse 1988); determinations that are challenging in the presence of strong molecular features, even where high-resolution spectroscopy of these faint stars is feasible (Plez & Cohen 2005). While there are no mass estimates for any dC stars as yet, based on the existing (and possibly biased) samples, their positions in the HR diagram suggest temperatures within or near the range 4000–

* E-mail: j.farihi@ucl.ac.uk

4500 K, and thus masses below solar (G77-61 is estimated at $0.3\,M_\odot$; Dearborn et al. 1986).

In the study of carbon-enhanced stars for stellar archaeology, the dC stars have a major advantage in that late-type dwarfs should be numerically dominant in the Galaxy, based on the initial mass function (Salpeter 1955; Reid, Gizis & Hawley 2002; Kirkpatrick et al. 2024). Furthermore, stars remaining on the main sequence are not subject to any possible mixing scenarios, such as those postulated for CEMP stars (Fujimoto, Ikeda & Iben 2000), which are typically on the first ascent giant branch in order to be detected within the distant halo. However, the larger dC population is not yet amenable to spectroscopic abundances studies, where a typical star has $g > 17$ AB mag and $H > 13$ mag, and thus past the magnitude limits of APOGEE in the near-infrared ($H = 12.2$ mag; Majewski et al. 2017). Furthermore, while their $T_{\rm eff} \lesssim 4500\,{\rm K}$ results in the characteristic, strong molecular features of carbon, available atmospheric models are typically carbon-normal at these low temperatures (e.g. Gustafsson et al. 2008, although see Aringer et al. 2016).

Because reliable and robust atmospheric and chemical modelling is not yet possible for dC stars as a class, an alternative approach to understand their origins is to study their kinematics and Galactic orbits. A previous study based on *Gaia* DR1 proper-motion catalogues of several hundred dC stars, and thus lacking reliable distances and instead using absolute magnitude estimates, suggested the population contains a significant halo component of at least 30 per cent, and that the majority are kinematically old, likely metal-poor stars in the thick disc and halo (Farihi et al. 2018). These findings for large numbers of non-kinematically selected stars are consistent with early indications based on a small number of stars detected via high proper motions (Harris et al. 1998).

With *Gaia* DR3 it is now possible to estimate distances to $N \sim 1000$ dC stars and thus obtain full 6-dimensional space positions and velocities. From these, one can calculate Galactic orbits and dynamics. This paper is organized as follows. Section 2 discusses the sample of dC star candidates used in this study and their associated data. Section 3 covers the inferences of distances to these stars using a Bayesian framework, and identifies the subset that can be confidently assigned to the main sequence. Section 4 details the kinematic and action-based analysis of the high-confidence dC stars, and their assignment to one of the three major components of the Galaxy. A brief discussion of the findings and future developments is given in Section 5.

2 CANDIDATE DC STARS AND DATA

Candidate dC stars are sourced from a published catalogue, assembled via cross-correlation between all spectroscopic targets observed during Sloan Digital Sky Survey (SDSS; York et al. 2000) DR7 and DR8, and a series of carbon star spectral templates, yielding a total of 1211 unique objects (Green 2013, table 1). While this identification method is biased towards the specific templates utilized to guide the colour selection criteria, it critically provides a kinematically unbiased sample of stars, which have only been selected for spectroscopic fibres based on their colours. However, because no kinematical cuts were made for the catalogue, it likely contains stars that do not lie on the main sequence, and hence are not dC stars. The removal of such catalogue contaminants is addressed in the next Section.

For each candidate dC star, all primary and ancillary data are sourced from SDSS DR16 (Ahumada et al. 2020), and *Gaia* DR3 (Gaia Collaboration 2023). SDSS *gri* PSF magnitudes are retrieved using ASTROQUERY (Ginsburg et al. 2019) by cross-matching the catalogue of candidate dC stars with the SDSS database using the

J2000 coordinates of each target. With the BESTOBJID and SPECOBJID for each source, radial velocities are adopted as those determined by the SEGUE survey and sourced via SPPPARAMS (Yanny et al. 2009). Parallaxes, ϖ, and proper motions, (μ_α, μ_δ) are taken from *Gaia* DR3, by performing a cone search around each dC source using a radius of 10 arcsec after propagating the *Gaia* astrometry from the reported DR3 epoch J2016.0 back to J2000, the approximate epoch of the SDSS astrometry.

From the initial 1211 candidate dC stars, 1372 matches are returned from the cone search. When more than one *Gaia* source is returned for a given SDSS target, the closest on-sky match is taken and checked by comparing the SDSS *r* brightness to a synthetic *r*-band magnitude computed from the *Gaia* *G*-band mean magnitude and $G_{\rm BP}$–$G_{\rm RP}$ colour (Riello et al. 2021). This yielded appropriate matches, except for the star labelled as SDSS J111324.84+220916.8 by Green (2013), which should be matched to the photometric source SDSS J111324.94+220916.2 (although this star does not enter the final high-confidence dC catalogue). A total of three candidate dC stars lacked counterparts in the *Gaia* DR3 cross-match, reducing the 1211 candidates to 1208. All *Gaia* parallaxes are zero-point corrected (Lindegren et al. 2021).[1]

3 PROBABILISTIC DISTANCE INFERENCE

In order to carry out a Galactic dynamical analysis of the dC star sample, full and reliable 6-dimensional position and velocity data are necessary. It is well known that simply inverting *Gaia* parallax measurements to obtain distances is prone to systematic errors and biases (Luri et al. 2018; Lindegren et al. 2021). An obvious example of these biases is negative parallax (e.g. for faint stars in crowded fields) that yields an unphysical distance if inverted.

Fortunately, there are Bayesian approaches to infer the distances that have been demonstrated to be robust, and which are adopted for the analysis here (Bailer-Jones 2015; Astraatmadja & Bailer-Jones 2016). This inference-based methodology has produced a distance catalogue for all stars that have a parallax in *Gaia* DR3 (Bailer-Jones et al. 2021), including the dC star candidates in this study. A Bayesian approach produces chains (posterior samples) of values for distances for each star based on its observational uncertainties. However, the Bailer-Jones et al. (2021) distance catalogue reports the distances and errors as the median, 16th, and 84th percentiles of these chains but does not provide the chains themselves. Distances need to be drawn from a distribution to perform a statistical analysis of the dC star sample, and it is not possible without the chains and shape of the distance distribution for each star.

Therefore, the 'metric' method is adopted to infer the distances for the sample of dC stars. The observational data from the *Gaia* DR3 astrometry and SDSS photometry are jointly modeled as $\mathbf{y} = [\varpi, \mu_\alpha, \mu_\delta, g, r, i]$. with associated uncertainty covariance matrix $\sigma_{\mathbf{y}}$ (as provided in the *Gaia* DR3 catalogue and assuming photometric errors are uncorrelated). For each star, the goal is to infer the distribution of a set of model parameters $\boldsymbol{\eta} = [d, \mu'_\alpha, \mu'_\delta, g_{\rm mod}, r_{\rm mod}, i_{\rm mod}]$ where d is the distance, $(\mu'_\alpha, \mu'_\delta)$ the true proper motion, and $(g, r, i)_{\rm mod}$ the (extinction-corrected) model magnitudes. The posterior is given by

$$p(\boldsymbol{\eta}|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}|\boldsymbol{\eta})\, p_{\rm Gaia}(d, \mu'_\alpha, \mu'_\delta) p_{\rm CMD}(g, r, i|d), \qquad (1)$$

where the first term is the likelihood, the second term is an astrometric prior, and the third term is a photometric prior. The likelihood is

---

[1] https://pypi.org/project/gaiadr3-zeropoint/

computed as

$$\log \mathcal{L}(\boldsymbol{y}|\boldsymbol{\eta}) \propto -\frac{1}{2}(\boldsymbol{y} - f(\boldsymbol{\eta}))^{\mathrm{T}} (\sigma_{\boldsymbol{y}} + \sigma_{f(\boldsymbol{\eta})})^{-1} (\boldsymbol{y} - f(\boldsymbol{\eta})), \quad (2)$$

where the function $f(\boldsymbol{\eta}) = (1/d, \mu'_\alpha, \mu'_\delta, g + A_g(d), r + A_r(d), i + A_i(d))$, and $A_x(d)$ is the extinction in the x band at distance d. Extinction corrections are computed using 3-dimensional extinction maps (using the DUSTMAPS interface; Green 2018; Green et al. 2019). These are converted to extinction in the (g, r, i) bandpasses using the coefficients (3.651, 2.633, 2.046), which are estimated by scaling the reported Pan-STARRS r-band coefficient by the ratio of the SDSS to Pan-STARRS r-band coefficients, and then ensuring the SDSS colour excess ratios follow that of Pan-STARRS (Schlafly & Finkbeiner 2011). Green et al. (2019) provide uncertainties in the extinction at each distance, which are incorporated into the likelihood by propagating the resulting (correlated) uncertainty in $f(\boldsymbol{\eta})$, given by the covariance matrix $\sigma_{f(\boldsymbol{\eta})}$, whose astrometric components are all zero.

Two prior contributions are introduced in equation (1), p_{Gaia} and p_{CMD}, that define the prior on the *Gaia* astrometry and the prior on the photometry respectively. Each of these priors is now defined.

3.1 Astrometric prior

The astrometric prior is split into two terms: $p_{\mathrm{Gaia}}(d, \mu_\alpha, \mu_\delta) = p(d) p(\mu_\alpha, \mu_\delta | d)$, following Bailer-Jones et al. (2021) and using a gamma distribution for the distance prior

$$p(d) \propto \begin{cases} d^\beta e^{-(d/L)^\alpha} & \text{if } d \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

This distance prior generically takes into account the exponentially decreasing space density of stars beyond some length scale L (Bailer-Jones 2015), although the gamma distribution allows for flexibility around this when looking at stars nearer the Galactic midplane, for example. Bailer-Jones et al. (2021) provide appropriate choices for the $p(d)$ parameters (L, α, β), with on-sky position based on fits to the *Gaia* mock simulation, GeDR3mock. These parameter choices are utilized in the following.

In addition to information from the parallax, there is weak information on the distance from the proper motions. If it is assumed that a typical star is bound to the Galaxy, the distance cannot be so large that, for a fixed proper motion, the star would be unbound. This is imposed by requiring stars have a maximum, heliocentric tangential velocity of $750\,\mathrm{km\,s^{-1}}$ (Bailer-Jones 2017; Luri et al. 2018), where $v_{\mathrm{tan}} = 4.74 d \sqrt{\mu_\alpha^2 + \mu_\delta^2}$ such that $p(\mu_\alpha, \mu_\delta | d) \propto d\, p(v_{\mathrm{tan}})$ and

$$p(v_{\mathrm{tan}}) = \begin{cases} \dfrac{1}{B(a,b)} \left(\dfrac{v_{\mathrm{tan}}}{v_{\mathrm{max}}}\right)^{a-1} \left(1 - \dfrac{v_{\mathrm{tan}}}{v_{\mathrm{max}}}\right)^{b-1} & \text{if } 0 \leq v_{\mathrm{tan}} \leq v_{\mathrm{max}}, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

This is a beta function with the parameters $a = 2, b = 3$ that control the shape of the distribution. This prior favours uniformly distributed velocities at small velocity and the high velocity taper prevents the possibility of arbitrarily large velocity solutions. No priors are placed on the direction of the tangential velocity.

3.2 Photometric prior

Colour (absolute) magnitude diagrams (CMDs) are a tool analogous to the HR diagram for visualising stellar properties and evolution, using colour in place of effective temperature, and absolute magnitude as a proxy for luminosity. The structure of a CMD implies that for

any given colour, a non-uniform probability distribution of absolute magnitude exists. Furthermore, through the distance modulus, the absolute magnitude contains information on the true distance of any given star. Thus, prior constraints can be placed on the true distance of a target based on its colour (Bailer-Jones 2011). This is encoded in the CMD prior, $p_{\mathrm{CMD}}(g, r, i | d)$, the prior probability of finding a star with given (error-free, de-reddened) SDSS photometry given its distance. It should be noted that $p(g, r, i | d)$ is equivalent to $p(M_r, g - r, r - i)$, the probability of a given extinction-corrected absolute r-band magnitude and the two de-reddened colours, $(g - r)$ and $(r - i)$.

The first step toward constructing such a prior is to assemble a high-fidelity CMD. Because the candidate dC stars are sourced from the SDSS, and hence possess multiband photometry, the CMD priors are constructed using stars with excellent astrometry from *Gaia* DR3, and photometry from SDSS DR16. A set of 250 000 stars is generated by randomly querying the *Gaia* DR3 cross-match with the SDSS DR13 best neighbour catalogue[2]. All stars returned by this query are required to have re-normalized unit weight error below 1.2 (Belokurov et al. 2020), and a measured parallax signal-to-noise above 50. Distances to each of the 250 000 stars are taken from the *Gaia* photo-geometric distance catalogue (Bailer-Jones et al. 2021), but the inverted parallaxes are likely also reliable given the quality cuts.

Photometric measurements in the SDSS gri bands for each of the stars with the aforementioned quality cuts are retrieved by querying the DR16 database using the BESTOBJID in CASJOBS. The photometric quality flags for each star in the pool are inspected to ensure that all targets possess good-quality photometric measurements. Extinction corrections are computed using the median of the aforementioned 3-dimensional extinction maps (Green 2018; Green et al. 2019), evaluated at the median *Gaia* photo-geometric distance (using the coefficients described previously). Extinction-corrected SDSS r-band absolute magnitudes, M_r, and the extinction-corrected colours $(g{-}r)$ and $(r{-}i)$ are computed for the 250 000 stars to assemble CMDs. SDSS u-band photometry is omitted because dC stars are faint in this band owing to their cool effective temperatures, exacerbated by strong carbon molecular absorption. Additionally, because of difficulties with ground-based photometry in the red (e.g. a less efficient detector, and telluric features), SDSS z-band photometry is also omitted (Doi et al. 2010). We compute the prior, $p(M_r, g - r, r - i)$, from the density of the 250 000 reference stars using a kernel density estimate evaluated by fast Fourier transform with the KDEPY package[3]. A Gaussian kernel with a standard deviation of 0.1 is adopted. For speed, the density estimate is evaluated on a fine grid of 200 points in each dimension with $-2\,\mathrm{AB\,mag} \leq M_r \leq 16\,\mathrm{AB\,mag}$, $-0.5 \leq (g - r) \leq 2.5$, and $-0.5 \leq g - i \leq 2.0$, which can then be interpolated.

It is important to note that these priors may introduce bias in distance determination, as they pull all posterior samples toward the main sequence. dC stars naturally have different colours to carbon-normal main-sequence stars. Furthermore, there are a handful of dC stars known to exhibit a composite spectrum with a white dwarf, but these represent less than 1 per cent of the sample (Green 2013). However, the candidate dC stars are faint with $\langle r \rangle \approx 18.5\,\mathrm{AB\,mag}$, and have modest- to poor-quality parallax measurements; thus, the use of CMD priors is strongly beneficial to the distance inference.

---

[2] The most recent data release to possess a cross-match provided by the *Gaia* team.

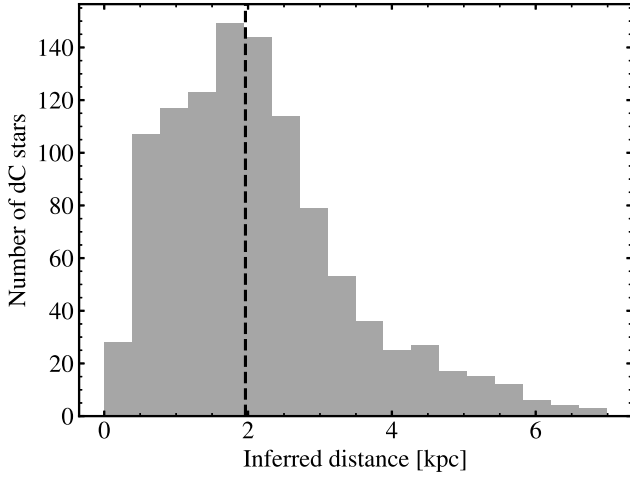[3] https://github.com/tommyod/KDEpy

Figure 1. Distribution of the resulting median distance for 1003 high-confidence dC stars, inferred from the distance modulus based on the results from the CMD diagram analysis. The black dashed line marks the sample median distance at $d = 1.96$ kpc, and the bin width equals the median of the sample distance uncertainty.

3.3 Resulting photogeometric distances for dC stars

The posterior of the model parameters given the data (equation 1) are sampled using the MCMC package EMCEE (Foreman-Mackey et al. 2013). For each dC candidate, an MCMC run starts with 50 walkers, each taking 5000 steps. The walkers are initialized by randomly drawing from a normal distribution centred at the measurement value for $(\mu_\alpha, \mu_\delta, g, r, i)$, and with a standard deviation equal to the measurement uncertainty. The initialization of the distances is done similarly, but with a mean $1/\varpi$ if $\varpi > 0.2$ mas, otherwise 5 kpc and a standard deviation of 30 per cent. All chains are deemed to have converged from the calculation of the autocorrelation times. The initial 1500 steps taken by each walker are discarded, and the median of each sampling chain is taken as the inferred parameter value, with the final error quoted as the 16th and 84th percentile confidence intervals (owing to the non-Gaussian nature of some posterior probability distributions). This results in distances for the dC candidates and produces chains that allow for random sampling in subsequent analysis (Section 4.1).

The distances inferred from the photo-geometric method are shown in Fig. 1, where half are located within 1.96 kpc, and the remaining stars have distances up to 6.5 kpc. These represent the first reliable distances calculated for a statistically compelling sample of $N \sim 1000$ dC stars.

With the distances shown in Fig. 1, the dC stars are plotted on (de-reddened) colour–absolute magnitude diagrams for each of M_r versus $(g–r)$, $(g–i)$, and $(r–i)$. The $(r–i)$ CMD is shown in Fig. 2, and the other CMDs are shown in the appendix (Fig. B1). These CMDs distinguish between the main-sequence and giant branches and thus clean the sample from most evolved stars. Main-sequence candidates are selected as lying within the region where the CMD density prior is above 0.08, as are any stars with $M_r > 7$ AB mag. Stars that fall within this boundary *in all three CMDs* are then classified as high-confidence dC stars. Fig. 2 is shown as an example, where blue points within the border are stars excluded owing to the constraints in the other two colours. This results in a sample of 1003 high-confidence dC stars.
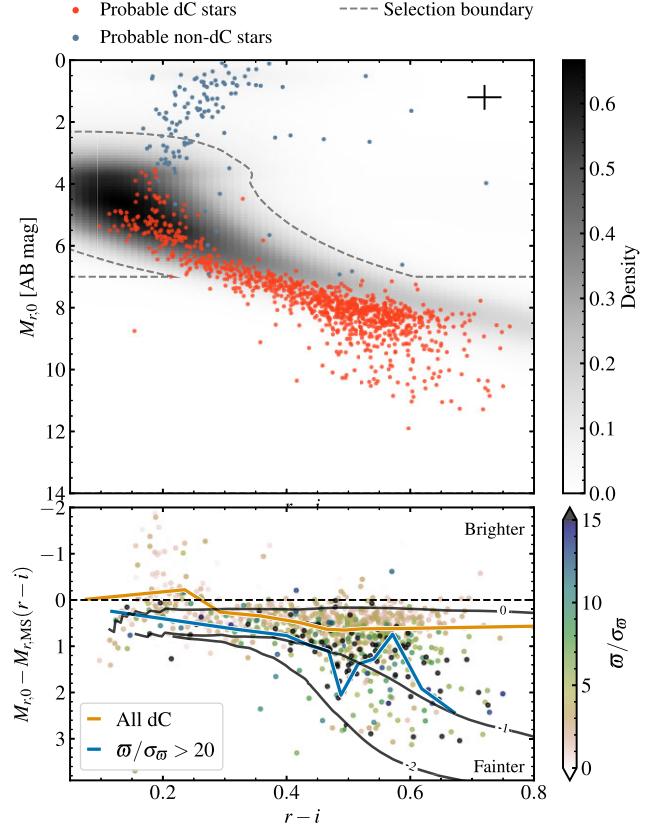

Figure 2. Colour-absolute magnitude diagram of the 1208 candidate dC stars. Top: The x-axis is the de-reddened SDSS $(r–i)$ colour, and the y-axis is the de-reddened absolute magnitude in the SDSS r band, calculated from the inferred distance. The greyscale shading, indicated by the colour bar on the right, is the underlying prior density distribution. The dashed grey line shows the boundary below which stars are considered high-confidence dC stars, defined as where the prior density is above 0.08 or $M_r > 7$ AB mag. The dark orange circles mark the candidates selected as high-confidence dC stars, and the blue circles are rejected (those within the selection boundary here are excluded based on CMDs using $(g–r)$ or $(g–i)$; Fig. B1). The error bar in the top right corner shows the median uncertainties in $(r–i)$ and M_r (using 16th and 84th percentiles) for all candidates. This selection procedure results in 1003 high-confidence dC stars. Bottom: Difference between the de-reddened r-band absolute magnitude and the peak prior density at each $(r − i)$, coloured by the parallax signal-to-noise, ϖ/σ_ϖ. The orange line shows the median for the high-confidence dC stars and the blue the subset with $\varpi/\sigma_\varpi > 20$. The three black lines are model isochrones (Dotter et al. 2008) with metallicities [Fe/H] $= 0, -1, -2$ as labelled.

The distribution of dC stars in absolute magnitude reveals that the population likely possesses a relatively broad continuum of masses and effective temperatures. Furthermore, despite the prior pulling stars toward the stellar locus, a significant fraction of the population lies to the left of the main sequence. This is illustrated in the bottom panel of Fig. 2, which shows the difference in the absolute magnitude with respect to the mode of the prior at each $(r–i)$. Clearly, the bulk of the dC candidates are fainter than the typical main-sequence star at fixed colour. Comparison with a set of model stellar isochrones (Dotter et al. 2008) demonstrates that this behaviour is expected for populations of [Fe/H] ~ -1 dex. Although the isochrones are for carbon-normal stars, the r and i bands are expected to be less affected by the carbon-rich nature than the g band, which overlaps with the series of C_2 Swan bands. This comparison indicates that the

population as a whole is likely to be predominantly metal-poor. As previously mentioned, the only dC star with a constrained metallicity is G77-61 ([Fe/H] $= -4.0$; Gass et al. 1988; Plez & Cohen 2005), and remains one of the most metal-deficient stars known.

A final consideration is whether the adoption of the photometric prior has biased the dC distances. As already noted, it appears the typical dC star is fainter than the typical main-sequence star, so in the absence of strong information from the parallaxes, the distances from this method will be biased towards larger values. The lower panel of Fig. 2 shows the main sequence offset for the full sample coloured by parallax signal-to-noise, ϖ/σ_ϖ, along with median trends for the full sample and those with $\varpi/\sigma_\varpi > 20$. There is a clear trend with ϖ/σ_ϖ, where more accurate parallaxes fall (on average) further below the main sequence (and indeed agree well with the [Fe/H] $= -1$ dex isochrone), while those with lower ϖ/σ_ϖ lie closer to the main-sequence prior. At most, this effect appears to be $\delta M_{r,0} \approx 1.5$ AB mag for stars with $(r-i) \approx 0.7$, which would translate into distance errors of approximately 70 per cent. However, for stars around $(r-i) \approx 0.4$, the bias is of the same order as the random error. This bias will incorrectly place the stars further from the sun and increase the inferred tangential velocity.

4 DYNAMICAL CLASSIFICATION OF DC STARS

Without information on metallicity or abundance for the dC stars, Galactic dynamics is used to constrain the fraction of the population that is likely to be old and metal-poor. The bulk of stars in each major component of the Galaxy have distinct, characteristic ages and chemical properties: broadly, the thin disc is the youngest and most metal-rich, while stars in the halo are, on average, older and metal-poor, with stars in the thick disc somewhere in between. The distribution of the dC stars over these Galactic components should then trace their bulk metallicity and age properties.

4.1 Component assignment using distribution functions

Full knowledge of the 6-dimensional position and velocity, combined with an underlying gravitational potential of the Milky Way, yields the Galactic orbit of any star, which can be labelled by actions. Distribution functions describe the probability that a star belonging to a given component is found on the orbit described by its actions. The three components of interest, the thin disc, thick disc, and stellar halo, are described with published model distributions as follows.

Binney & Vasiliev (2023) present a self-consistent Galaxy model, where the gravitational potential is computed from the subcomponent distribution functions. The Galactic components fitted in that work are three thin discs (where age is used in labelling but not in the modelling), a thick disc, a spheroidal bulge, and spheroidal stellar and dark matter haloes. The stellar discs are described as exponential distributions, and the spheroidal components as double power-laws. The distribution functions of interest are the thin disc (summed over the three model components), the thick disc, and the stellar halo. These distribution functions have been fitted to *Gaia* DR2 data using AGAMA (Vasiliev 2019). One caveat is that for technical reasons related to the density of stars at $J_\phi = 0$, the halo distribution function employed by Binney & Vasiliev (2023) has no significant radial bias contrary to observations. This may weakly bias dC stars on more circular orbits to be assigned to the halo component rather than the thick disc.

To include the observational uncertainties, for each star, 1000 random realizations of $(\alpha, \delta, \mu_\alpha, \mu_\delta)$ are drawn from their correlated

uncertainty distributions, while radial velocities are sampled from the SEGUE data described in Section 2, and distances from their posterior chains (Section 3.3). The radial velocities of all 1003 high-confidence dC stars are found to be reliable based on the SEGUE quality flag `zwarning_elodie`. Note that although the true proper motions are inferred in Section 3, these are not used here. For each realization of each star, the actions (J_r, J_z, $J_\phi = L_z$) and the probability of belonging to each Galactic component are calculated. Unbound posterior samples have undefined actions so their Galactic component membership under an equilibrium distribution function is ambiguous: these posterior samples are defined as belonging to the halo component. The component probabilities are individually normalized by the sum of all three components, where the median, 16th, and 84th percentiles are used to obtain probabilities and actions for each star, with uncertainties. Each dC star is then assigned to the Galactic component with the highest probability.

4.2 Fraction of halo, thick, and thin disc orbits

With the stars assigned to the different Galactic components, their properties are now explored. The right-hand panel of Fig. 3 shows the number of dC stars with the highest assigned probability in each component, where 62.1 ± 0.6 per cent ($N = 624$) of the population belong to the stellar halo, 31.4 ± 0.7 per cent ($N = 320$) are part of the thick disc, and only 6.5 ± 0.4 per cent ($N = 59$) are members of the thin disc (where uncertainties are derived from standard deviations over posterior samples for each star). The median Galactic component posterior probabilities, $\bar{p}_i$, and their corresponding 16th to 84th confidence intervals are shown in Fig. 4. Here, the stars are ordered horizontally by $\bar{p}_{thin} + 2\bar{p}_{thick} + 3\bar{p}_{halo}$ to sort confident thin disc members on the left and confident halo members on the right. It can be seen that a significant fraction of the sample is dynamically classified with high confidence. There is some ambiguity between the thick disc and halo for ≈ 10 per cent of the sample, arising from the propagation of observational uncertainties. On the left-hand side of the diagram, there are stars with tight posteriors owing to small observational uncertainties, for which there is some ambiguity between thin and thick disc membership because of their overlapping dynamical distribution functions.

The distribution functions depend on the actions of each star, which are visualized in Fig. 3. This action-space map of the dC stars, where each object is colour-coded by the probability of belonging to each component, shows how the actions inform the distribution functions. The distribution of thin disc stars is tightly confined to the right corner, with thick disc stars spreading further along the in-plane axis, and toward more radial orbits. The remaining objects in the diagram, and thus the bulk of dC stars, are dominated by those with halo orbits. The analysis in this work has taken care to propagate the uncertainties in the actions (and thus the classification) of each star. In Appendix B, the distribution of the actions along with their uncertainties are shown. Although these uncertainties can be sizeable, it is clear they are sufficiently small for an accurate classification of each star.

5 DISCUSSION

The results here provide the first substantial catalogue of high-confidence dC stars, where the dynamics suggest they are predominantly a halo population, likely implying that they are also a metal-poor or an old population. These results are discussed in relation to previous work, with some caveats and possible implications.
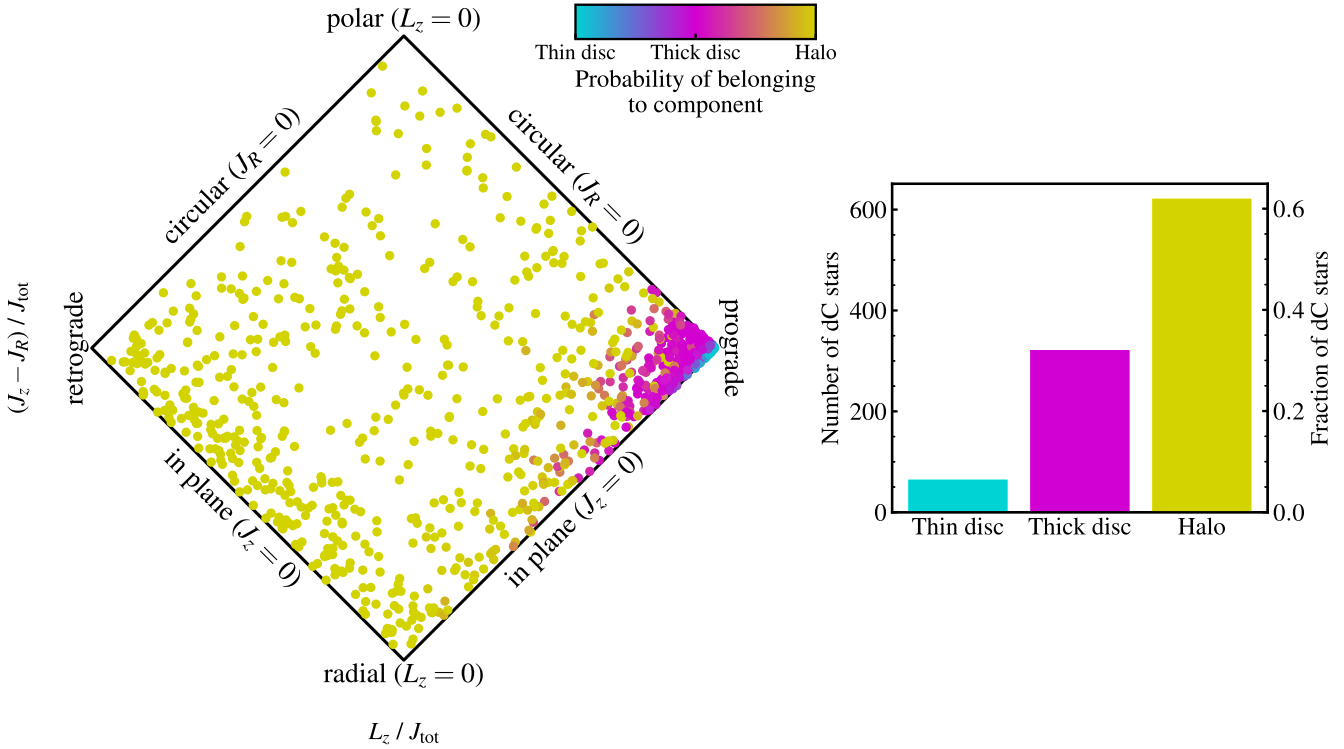
Figure 3. Left: The action-space map of the 1003 high-confidence dC stars. The horizontal axis represents the normalized angular momentum in the (vertical) Galactic Z-direction, ranging from -1 for retrograde to $+1$ for prograde orbits. The vertical axis represents the relation between the vertical and the radial action. Objects on the lower edges move in the Galactic plane, while stars located toward the upper corner move on increasingly polar orbits. Stars located along the upper edges have circular orbits, but these become increasingly radial for diagram positions toward the bottom corner. Therefore, stars located in the right corner move within the disc (prograde, in the plane, and circular), while stars positioned away from this position have thick disc and halo orbits. The dC stars are colour-coded by their median probabilities of belonging to each Galactic component, with a continuous distribution of colours, representing stars on the edge between two components; stars in violet are between the thin and thick disc, and stars in orange are between the thick disc and halo. Right: Distribution of dC stars in the different Galactic components resulting from the action-based classification: the thin disc contains 6.5 per cent (59 stars), the thick disc 31.4 per cent (320 stars), and the halo 62.1 per cent (624 stars) of the dC population. This is a strong indication of an old and metal-poor population.

5.1 Comparison with previous work

Based on the preceding analysis, only a small percentage of dC stars are likely thin disc members, while the remaining fraction have orbits consistent with a metal-poor population, dominated by objects that were formed in the Galactic halo. These results are consistent with a prior study using the same SDSS sample of dC stars as in this work, based on proper motion catalogues generated from *Gaia* DR1 astrometry, finding between 30 and 60 per cent of dC candidate stars belong to the stellar halo (Farihi et al. 2018). That pre-*Gaia* DR2 analysis relies on both reduced proper motion with isochrone fitting, as well as Toomre diagrams, to identify the different Galactic components kinematically (see e.g. Venn et al. 2004), assuming that all dC stars have $M_r = 8.0$ AB mag. This bulk absolute magnitude estimate agrees remarkably well with the results of the sample in this paper, which has a mean (median) absolute magnitude $M_r = 7.7$ AB mag (but a substantial spread, Fig. 2).

More recently and post-*Gaia* DR2, Roulston et al. (2022) use a Toomre diagram for estimating the age for a subsample of the SDSS dC stars that are also analyzed here. In that study, the dC stars are compared to a sample of carbon-normal K and M dwarfs that are reported to represent the thin and thick disc, and the authors subsequently argue that the resulting poor match implies the dC stars have either a thick disc or halo origin. To underscore the success of the dynamical analysis shown in Fig. 3 as compared to a purely

kinematic selection, the 1003 dC stars are plotted in a Toomre diagram in Fig. 5, colour-coded by their component probabilities. The components roughly segregate along the curves of constant velocity, with some overlap between. Notably, however, the sample distributions in Fig. 5 and those in Roulston et al. (2022) do not match; e.g. their figure 1 does not display any stars in the retrograde disc (i.e. at low $\sqrt{U^2 + W^2}$ and negative V). This could be due to an incorrect coordinate transformation of their data.

It has been suggested that the kinematics of dC stars might be biased owing to radial velocities that include a significant orbital motion component for those in close binaries, such as the handful of systems consistent with a post-common envelope configuration (Margon et al. 2018; Whitehouse et al. 2021; Roulston et al. 2022). However, existing radial velocity measurements are consistent with circular orbits, where the average orbital speed along the line of sight is zero, and there is an equal chance to decrease or increase the total observed line-of-sight velocity. Furthermore, the orbital velocity semi-amplitudes for dC stars in these binaries are typically smaller than $50\,\mathrm{km\,s^{-1}}$, with a few exceptions closer to $100\,\mathrm{km\,s^{-1}}$, and a random measurement will yield no more than 0.72 of these extrema. The dominant halo population of dC stars have space velocities that are significantly larger than the average range of orbital speeds measured along the line of sight, and therefore, based on this and the aforementioned caveats, orbital motion cannot significantly inflate the inferred kinematics of dC stars.
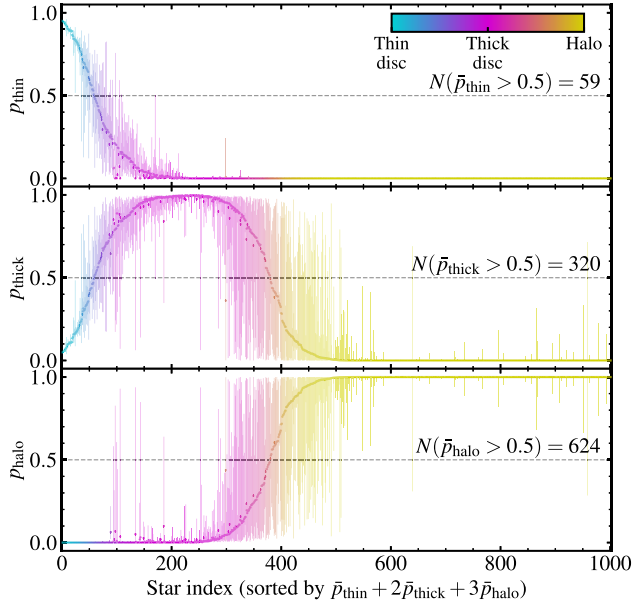
Figure 4. Individual Galactic component posteriors for the 1003 high-confidence dC stars. Each panel shows a dot and bar per star representing the posterior median and the 16th to 84th percentile confidence range, respectively, for membership in each component (top: thin disc; middle: thick disc; bottom: halo). The stars are ordered horizontally by the sum of median probabilities, $\bar{p}_{\rm thin} + 2\bar{p}_{\rm thick} + 3\bar{p}_{\rm halo}$, and coloured by their median classification as per the colour bar. The annotation in each panel gives the number of stars with a median classification probability greater than 0.5 for each component. It can be seen that a large fraction of stars classified as halo members have an almost 100 per cent probability of belonging to the halo, and almost zero probability of belonging to the two disc components. In contrast, the distribution of stars classified as thin disc is not as clean, where many objects have a non-negligible probability of belonging to the thick disc. Distinguishing between the thin and thick disc solely based on dynamics is not straightforward, where chemical abundances would improve upon this purely dynamical analysis. However, the classification is reliable for thick disc and particularly halo stars.
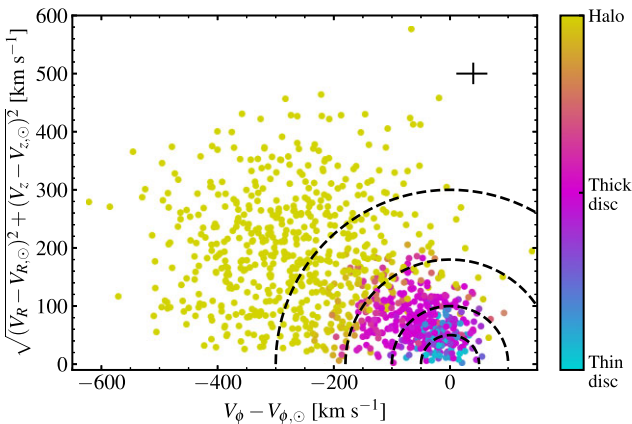

Figure 5. Toomre diagram of the high-confidence dC stars colour-coded as in Fig. 3. The diagram is corrected to the local standard of rest, so that the thin disc has $V_\phi - V_{\phi,\odot} = 0\,{\rm km\,s^{-1}}$. The dashed semi-circles trace constant total space velocities of 50, 100, 180, and 300 km s^{-1}. The error bar in the upper right shows the median uncertainty. While the distribution of dC stars in these Galactic components roughly follows the semi-circular benchmarks, owing to the broad spatial distribution, several stars are in locations where they would be attributed to a different component if the selection is done only kinematically using the Toomre diagram.
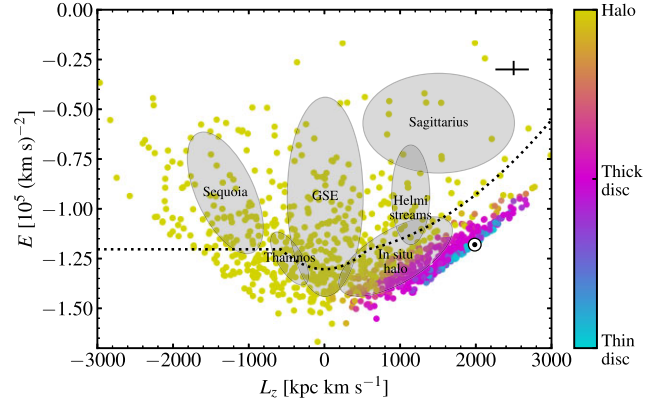

Figure 6. Energy–angular momentum diagram of the 1003 high-confidence dC stars colour-coded as in Fig. 3. The conventional sun symbol plots the location of a circular orbit at the solar Galacto-centric radius. The error bar in the upper right is the median uncertainty of the sample. The dashed black line marks the approximate separation between *in situ* and accreted globular clusters (Belokurov & Kravtsov 2024), and the grey labelled ellipses show sub-structures (Naidu et al. 2020).

The results found here for dC stars passing through the solar neighbourhood are directly comparable to a kinematical study of 644 CEMP giant stars situated closer to and within the Galactic halo (Zepeda et al. 2023, their fig. 8 parallels Figs 3 and 6). While that study leverages chemical abundance data that are not available for the dC sample, they find the CEMP stars are distributed over the entirety of orbital action space, similar to the dC population with the exception of its disk component. Abundance information allows co-identification with halo substructures for a few dozen CEMP stars, which is not possible here, but nevertheless the dC halo stars overlap with known substructures in a similar manner. Future data and modelling are necessary for definitive substructure association for any dC stars.

5.2 Selection function of the sample

To assess the impact of this spatial selection, the component membership probabilities are computed from the Binney & Vasiliev (2023) model using only the spatial locations of the sample (i.e. the density fractions of each component). For the entire sample, the ratio of thin:thick:halo membership based purely on spatial location is found to be 0.17 : 0.67 : 0.16, while additionally using the kinematics gives 0.07 : 0.31 : 0.62. There is indeed a bias toward observing thick disc and halo objects given the SEGUE selection but despite this, the dynamic classification shows the sample has a stronger tendency toward the thick disc and halo than expected purely on the basis of the selection. This is evidence that the bulk of dC stars indeed have thick disc or halo kinematics, and are thus metal-poor or old, but fuller modelling of the selection function would be required to conclude this definitively.

5.3 Do thick disc and halo kinematics imply metal-poor or old?

The dC sample has been shown to have kinematics dominated by stars consistent with membership in the thick disc or halo. This was assessed using purely dynamical distribution functions designed to give the best representation of the density and kinematics of the Galaxy with no explicit consideration of the properties of the stars (Binney & Vasiliev 2023). However, it is well known that the

average star in these Galactic components *is* metal-poor and old, with [Fe/H] $\lesssim -0.5$ (consistent with the isochrone comparison in Fig. 2), and age $\gtrsim 10$ Gyr (e.g. Ivezić et al. 2008; Kilic et al. 2017), thus the association between metallicity-age and kinematics.

Recent work and the arrival of *Gaia* data have enabled a richer and deeper understanding of the Milky Way stellar halo. Fig. 6 plots the energy-angular momentum diagram for the high-confidence dC stars along with the *in situ*-accreted boundary for globular clusters proposed by Belokurov & Kravtsov (2024), and the approximate locations of the structures classified by Naidu et al. (2020). It can be seen that there is no particularly strong tendency toward any particular halo sub-component in the sample and the stars classified as halo are distributed across both accreted and *in situ* parts of the diagram.

The association of dC stars with different halo sub-components would have implications for their metallicities and ages. For example, the GSE stars are known to have $-1.28 <$ [Fe/H] < -1.18 with a low metallicity tail down to at least [Fe/H] $= -2.1$ (Amarante et al. 2022; Carrillo et al. 2024), and are predominantly old with a possible tail of younger stars (Horta et al. 2024). In contrast, the *in situ* halo component (Bonaca et al. 2017; Belokurov et al. 2020) can contain old stars with [Fe/H] ≈ -0.7, and highly radial and possibly even retrograde orbits. Further study is warranted to more accurately classify the sub-structure membership of the sample studied here, but based on the energy–angular momentum diagram and comparison with known sub-structures, the dC stars seem fairly representative of the average local halo star. It is thus likely that, on average, the sample is old and metal-poor.

A further tool for assessing more explicitly the metallicity and ages of the stars is using an *extended* distribution function (e.g. Sanders & Binney 2015; Binney & Vasiliev 2024). These models attempt to reproduce both the dynamical *and* chemical structure of the Galaxy, so one can construct e.g. $p([\text{Fe/H}]|\,x,\,v)$. Alternatively, data-driven label transfer techniques that attempt to match kinematics as a function of auxiliary parameters could be employed (Zhang et al. 2025). These are both potentially fruitful directions for further work with this sample.

5.4 The origin of dC halo stars

Metal-deficient stars are more susceptible to atmospheric pollution via mass transfer (e.g. metal-free white dwarfs), because only a fraction of carbon-rich material is sufficient to increase the C/O ratio beyond 1, relative to solar metallicity counterparts (De Kool & Green 1995), all else being equal. This is part of the consensus interpretation for CH and CEMP-s stars, which are low-metallicity binaries where the secondary star is carbon-enhanced post-mass transfer (e.g. Lucatello et al. 2005; Jorissen et al. 2016). The likely metal-poor nature of the dC population implied by the dynamical analysis presented in this work is consistent with a similar, binary origin, at least for some fraction, and where a growing body of work suggests dC stars are often found in binaries where past mass transfer is likely or plausible (Whitehouse et al. 2018; Harris et al. 2018; Margon et al. 2018; Roulston et al. 2021).

It remains unknown if there are dC stars analogous to the CEMP-no population, which may contain intrinsically carbon-rich stars that are not correlated with their binary properties (similar to the nearby field population; Starkenburg et al. 2014; Yoon et al. 2016). Only G77-61 has detailed abundances for such a comparison, and it is often classified as CEMP-no based on the non-detection of barium, and its extreme metal poverty, but it should be noted the upper limit on s- and r-process elements is not as robust as for typical (giant) CEMP

stars (Plez & Cohen 2005; Arentsen et al. 2019). Based on Fig. 2, the dC stars in this study span a mass range that likely overlaps with the known CEMP population (which are giants owing to luminosity bias), but also extends to lower masses that will not leave the main sequence in a Hubble time.

Nevertheless, the results have also demonstrated that there is a subdominant thin disc component in the dC sample. Thin disc stars are characteristically solar metallicity so it is significantly more difficult to produce a carbon-rich atmosphere through mass transfer for these objects. However, the results suggest it is possible in rare cases (especially considering the high ratio of solar metallicity to metal-poor star populations). Further study of dC stars in more metal-rich environments and a more detailed characterization of the selection function can help constrain their binary properties and the efficiency of mass transfer as a function of metallicity.

6 CONCLUSIONS AND OUTLOOK

This study has analyzed 1003 high-confidence dC stars to determine their distances, 3-dimensional space motions, and Galactic orbits, using a combination of SDSS DR16 photometry and radial velocity data, and state-of-the-art *Gaia* DR3 astrometry. These properties are used to dynamically classify the dC stars and assign component membership likelihood, using action-based distribution functions. This is achieved by first using Bayesian inference to obtain accurate distances for these faint stars, using a set of astrometric and photometric priors based on expected HR diagram positions. The results are summarized as follows:

(i) Probabilistic distances are inferred for 1208 candidate dC stars from Green (2013), using astrometry from *Gaia* DR3 and photometry from SDSS DR16 (Fig. 1). With these distances, colour–absolute magnitude diagrams are computed for all candidate stars. By selecting regions in the $(g-r)$, $(g-i)$, and $(r-i)$ colours that cover the main sequence and below, a total of 1003 high-confidence dC stars are identified (Fig. 2).

(ii) The actions of the 1003 dC stars with reliable astrometry and spectroscopy are calculated, and Galactic component membership is assigned based on the distribution functions from Binney & Vasiliev (2023). Each star is thus assigned to the thin disc, the thick disc, or the stellar halo. Roughly 60 per cent of the dC population are found to have dynamics consistent with halo membership and around 30 per cent with thick disc orbits (Figs 3 and 4). It is demonstrated that the conclusion of predominantly halo membership is not a result of the spatial selection function of the sample.

(iii) The high-confidence dC stars preferentially fall below (to the left of) the main sequence in the M_r versus $(r-i)$ diagram. Comparison with carbon-normal isochrones in this plane suggests the dC population is metal poor with [Fe/H] ~ -1. The dynamical association with the thick disc and halo yields further evidence that the dC stars are metal-poor or old. Comparison with known Galactic halo sub-structures does not reveal a tendency toward any known sub-component, suggesting they are broadly representative of the average metal-poor and old stellar halo (Section 5.4).

These results demonstrate the significance of further investigation of these rare, poorly understood stars. The next generation of large spectroscopic surveys (e.g. 4MOST) will likely observe many known dC stars and possibly discover new examples. Furthermore, the results suggest that a more dedicated spectroscopic follow-up campaign of dC stars should be undertaken. Spectroscopic analysis to measure metallicities and abundances of these stars remains a challenge, but the rewards of overcoming current limitations are

tempting: a detailed spectroscopic study would provide conclusive evidence on the metallicity of the population as a whole, possibly opening up new avenues to discover and study very (or extremely) metal-poor stars.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data used in this paper are available in public archives from the Sloan Digital Sky Survey and the European Space Agency. The data set produced in this study is made available alongside this publication with the data columns described in Table A1. The table and the code to reproduce all calculations and plots in the paper are provided at https://github.com/jls713/dc.

REFERENCES

Abate C., Pols O. R., Stancliffe R. J., Izzard R. G., Karakas A. I., Beers T. C., Lee Y. S., 2015, A&A, 581, A62
Ahumada R. et al., 2020, ApJS, 249, 3
Amarante J. A. S., Debattista V. P., Beraldo e Silva L., Laporte C. F. P., Deg N., 2022, ApJ, 937, 12
Arentsen A., Starkenburg E., Shetrone M. D., Venn K. A., Depagne É., McConnachie A. W., 2019, A&A, 621, A108
Arentsen A., Placco V. M., Lee Y. S., Aguado D. S., Martin N. F., Starkenburg E., Yoon J., 2022, MNRAS, 515, 4082
Aringer B., Girardi L., Nowotny W., Marigo P., Bressan A., 2016, MNRAS, 457, 3611
Astraatmadja T. L., Bailer-Jones C. A. L., 2016, ApJ, 832, 137
Bailer-Jones C. A. L., 2011, MNRAS, 411, 435
Bailer-Jones C. A. L., 2015, PASP, 127, 994
Bailer-Jones C., 2017, Joint inference from parallax and proper motions, Provided by ESA, https://dms.cosmos.esa.int/COSMOS/doc_fetch.php?id = 3504158
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Beers T. C., Christlieb N., 2005, ARA&A, 43, 531
Belokurov V., Kravtsov A., 2024, MNRAS, 528, 3198
Belokurov V. et al., 2020, MNRAS, 496, 1922
Binney J., Vasiliev E., 2023, MNRAS, 520, 1832
Binney J., Vasiliev E., 2024, MNRAS, 527, 1915
Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, ApJ, 845, 101
Carrillo A., Deason A. J., Fattahi A., Callingham T. M., Grand R. J. J., 2024, MNRAS, 527, 2165
Christlieb N., Wisotzki L., Graßhoff G., 2002, A&A, 391, 397
Cooke R., Pettini M., Steidel C. C., Rudie G. C., Jorgenson R. A., 2011, MNRAS, 412, 1047
Dahn C. C., Liebert J., Kron R. G., Spinrad H., Hintzen P. M., 1977, ApJ, 216, 757
De Kool M., Green P. J., 1995, ApJ, 449, 236
Dearborn D. S. P., Liebert J., Aaronson M., Dahn C. C., Harrington R., Mould J., Greenstein J. L., 1986, ApJ, 300, 314

Doi M. et al., 2010, AJ, 139, 1628
Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, ApJS, 178, 89
Farihi J., Arendt A. R., Machado H. S., Whitehouse L. J., 2018, MNRAS, 477, 3801
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Frebel A., Norris J. E., 2015, ARA&A, 53, 631
Fujimoto M. Y., Ikeda Y., Iben Icko J., 2000, ApJ, 529, L25
Gaia Collaboration et al., 2023, A&A, 674, A1
Gass H., Liebert J., Wehrse R., 1988, A&A, 189, 194
Ginsburg A. et al., 2019, AJ, 157, 98
Green P., 2013, ApJ, 765, 12
Green G., 2018, J. Open Source Softw., 3, 695
Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, ApJ, 887, 93
Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund Å., Plez B., 2008, A&A, 486, 951
Harris H. C. et al., 1998, ApJ, 502, 437
Harris H. C. et al., 2018, AJ, 155, 252
Horta D., Lu Y. L., Ness M. K., Lisanti M., Price-Whelan A. M., 2024, ApJ, 971, 170
Iben Icko J., 1965, ApJ, 142, 1447
Ivezić Ž. et al., 2008, ApJ, 684, 287
Jorissen A. et al., 2016, A&A, 586, A158
Jura M., Xu S., Young E. D., 2013, ApJ, 775, L41
Keller S. C. et al., 2007, PASA, 24, 1
Kilic M., Munn J. A., Harris H. C., von Hippel T., Liebert J. W., Williams K. A., Jeffery E., DeGennaro S., 2017, ApJ, 837, 162
Kirkpatrick J. D. et al., 2024, ApJS, 271, 55
Li L. et al., 2024, ApJS, 271, 12
Lindegren L. et al., 2021, A&A, 649, A4
Lucatello S., Tsangarides S., Beers T. C., Carretta E., Gratton R. G., Ryan S. G., 2005, ApJ, 625, 825
Luri X. et al., 2018, A&A, 616, A9
Majewski S. R. et al., 2017, AJ, 154, 94
Margon B., Kupfer T., Burdge K., Prince T. A., Kulkarni S. R., Shupe D. L., 2018, ApJ, 856, L2
Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, ApJ, 901, 48
Plez B., Cohen J. G., 2005, A&A, 434, 1117
Reid I. N., Gizis J. E., Hawley S. L., 2002, AJ, 124, 2721
Riello M. et al., 2021, A&A, 649, A3
Roulston B. R., Green P. J., Toonen S., Hermes J. J., 2021, ApJ, 922, 33
Roulston B. R. et al., 2022, ApJ, 926, 210
Salpeter E. E., 1955, ApJ, 121, 161
Sanders J. L., Binney J., 2015, MNRAS, 449, 3479
Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103
Starkenburg E., Shetrone M. D., McConnachie A. W., Venn K. A., 2014, MNRAS, 441, 1217
Vasiliev E., 2019, MNRAS, 482, 1525
Venn K. A., Irwin M., Shetrone M. D., Tout C. A., Hill V., Tolstoy E., 2004, AJ, 128, 1177
Whitehouse L. J., Farihi J., Green P. J., Wilson T. G., Subasavage J. P., 2018, MNRAS, 479, 3873
Whitehouse L. J., Farihi J., Howarth I. D., Mancino S., Walters N., Swan A., Wilson T. G., Guo J., 2021, MNRAS, 506, 4877
Yanny B. et al., 2009, AJ, 137, 4377
Yong D. et al., 2013, ApJ, 762, 27
Yoon J. et al., 2016, ApJ, 833, 20
York D. G. et al., 2000, AJ, 120, 1579
Young E. D., 2014, Earth Planet. Sci. Lett., 392, 16
Zepeda J. et al., 2023, ApJ, 947, 23
Zhang H., Iorio G., Belokurov V., Evans N. W., Bobrick A., D'Orazi V., 2025, preprint (arXiv:2504.06720)

APPENDIX A: DATA TABLE

The data for the high-confidence dC stars are provided in Table A1.

Table A1. Description of the columns for the provided processed candidate dC stars. The columns e_X_high and e_X_low give the upper and lower 1σ uncertainties of columns X. Energy, actions, and component probabilities are calculated using the model from Binney & Vasiliev (2023). Halo probabilities can be computed as e.g. probhalo = 1-probthin-probthick.

Column	Description
name	SDSS name from Green (2013)
source_id	*Gaia* DR3 source id
ra	*Gaia* DR3 right ascension (deg)
dec	*Gaia* DR3 declination (deg)
angular_separation	Angular separation between SDSS DR16 and *Gaia* DR3 positions (arcsec)
M_r	Extinction-corrected r-band absolute magnitude
g_r	de-reddened (g–r) colour
g_i	de-reddened (g–i) colour
r_i	de-reddened (r–i) colour
distance	Distance (kpc)
A_r	r-band extinction
R	Galactocentric radius (kpc)
phi	Galactocentric azimuth (rad)
z	Galactocentric vertical height (kpc)
vR	Galactocentric radial velocity (km s^{-1})
vphi	Galactocentric azimuthal velocity (km s^{-1})
vz	Galactocentric vertical velocity (km s^{-1})
energy	Energy (km s^{-1})^2
JR	Radial action (kpc km s^{-1})
Jz	Vertical action (kpc km s^{-1})
Jphi	z-component of angular momentum (kpc km s^{-1})
probthin	Thin disc probability from $(\boldsymbol{x}, \boldsymbol{v})$
probthick	Thick disc probability from $(\boldsymbol{x}, \boldsymbol{v})$
comp	Most-likely component membership (1 = thin, 2 = thick, 3 = halo)
probthin_x	Thin disc probability from $\boldsymbol{x}$
probthick_x	Thick disc probability from $\boldsymbol{x}$
dC	Dwarf carbon star boolean (True = dC)

APPENDIX B: ADDITIONAL PLOTS

This appendix provides additional supporting plots. Fig. 2 shows the M_r versus (r–i) colour–magnitude diagram used for classifying high-confidence dC stars. Fig. B1 shows M_r versus. (g–r) and M_r vs. (g–i) diagrams which have also been used as part of the classification. It is expected that the g band photometry of dC stars is most distinct from that of carbon-normal stars owing to the presence of strong C$_2$ Swan absorption in this wavelength range, although the r band is also moderately affected (see fig. 1 of Green 2013). One additional interesting feature of these plots compared to Fig. 2 is that the dC stars with $\varpi/\sigma_\varpi > 20$ that are located near the main-sequence turn-off region are redder than the prior (and correspondingly the low parallax signal-to-noise dC candidates). Again, this is an indication of the carbon sensitivity of the g band.

Fig. B2 shows the action distribution of the high-confidence dC stars along with their uncertainties. The uncertainties can be substantial, particularly for the stars classified as halo, but sufficiently small to reliably distinguish the different components.
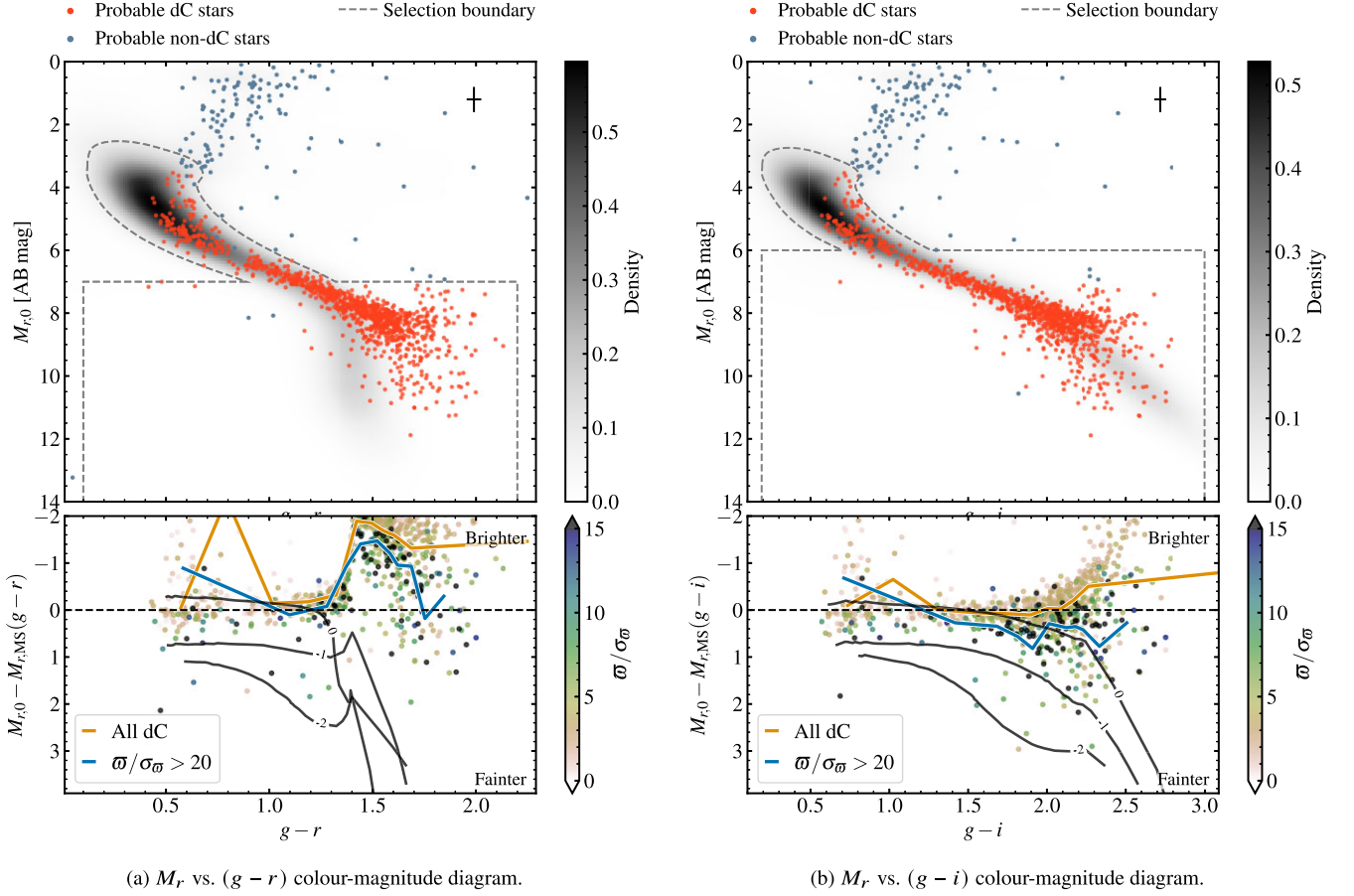
(a) M_r vs. $(g - r)$ colour-magnitude diagram.

(b) M_r vs. $(g - i)$ colour-magnitude diagram.

Figure B1. Additional colour–(absolute) magnitude diagrams for the candidate dC stars. See Fig. 2 for a description of the figures.
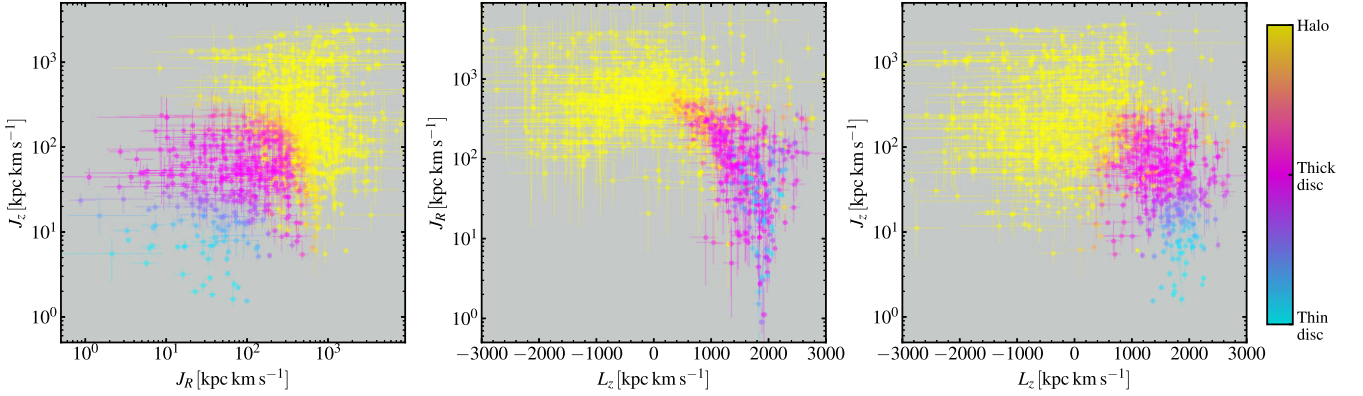

Figure B2. Action distributions for the high-confidence dC stars. The points are coloured by their assigned component probability.

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