


# A determination of the Large Magellanic Cloud dark matter subhalo mass using the Milky Way halo stars in its gravitational wake

K. J. Fushimi<sup>1</sup> , M. E. Mosquera<sup>1,2</sup>, and M. Dominguez<sup>3</sup>

<sup>1</sup> Facultad de Ciencias Astronómicas y Geofísicas, University of La Plata, Paseo del Bosque S/N 1900, La Plata, Argentina  
e-mail: [kfushimi@fcaglp.unlp.edu.ar](mailto:kfushimi@fcaglp.unlp.edu.ar)

<sup>2</sup> Dept. of Physics, University of La Plata, c.c. 67 1900, La Plata, Argentina

<sup>3</sup> Instituto de Astronomía Teórica y Experimental, (IATE-CONICET), Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, CONICET, CCT Córdoba, Laprida 854, X5000BGR Córdoba, Argentina

Received 12 July 2023 / Accepted 10 May 2024

## ABSTRACT

**Aims.** Our goal is to study the gravitational effects caused by the passage of the Large Magellanic Cloud (LMC) in its orbit on the stellar halo of the Milky Way.

**Methods.** We employed *Gaia* Data Release 3 to construct a halo tracers dataset consisting of K-giant stars and RR-Lyrae variables. Additionally, we compared the data with a theoretical model to estimate the dark matter subhalo mass.

**Results.** We have improved the characterisation of the local wake and the collective response due to the LMC's orbit. We have also estimated for the first time the dark subhalo mass of the LMC to be of the order of  $1.7 \times 10^{11} M_{\odot}$ , which is comparable to previously reported values in the literature.

**Key words.** Galaxy: halo – galaxies: kinematics and dynamics – Magellanic Clouds – dark matter

## 1. Introduction

Dark matter (DM) is central to the standard cosmological model (ΛCDM), providing gravitational support for the formation of galaxies and systems of galaxies (Mo et al. 2010). Its existence is backed by a plethora of observational data, including galaxies' rotation curves (Zwicky 1933; Rubin et al. 1970), strong and weak gravitational lensing effects (Massey et al. 2010; Clowe et al. 2006), and even the presence of baryonic acoustic oscillation (Planck Collaboration VI. 2020) in the earlier gravitational wells observed in the cosmic microwave background (CMB).

Despite these successes, we still lack a precise detection either in the laboratory (Bernabei et al. 2022; Aprile et al. 2023; Barberio et al. 2023; Amaré et al. 2022) or by indirect astrophysical observation (Abdalla et al. 2022; Acharyya et al. 2023; Abe et al. 2023). Many theoretical candidates arise out of physics beyond the standard model (BSMP), like weakly interacting massive particles (WIMPs). These massive ( $m_{\text{DM}} \sim 100 \text{ GeV}$ ) particles could weakly interact with nucleons, and therefore their signals have been looked at by several laboratory and accelerator experiments. Also, their annihilation signals could be detectable through  $\gamma$ -ray emission by high-energy telescopes. Despite a massive experimental effort, DM remains a theoretical hypothesis, albeit one with impressive empirical support.

Other DM candidates could be more massive, such as primordial black holes (PBHs) (Carr & Kühnel 2020), which were recently constrained with a series of consistency tests. Nowadays, there is still room to be an essential contributor to the DM content but this is limited to small windows in mass (Villanueva-Domingo et al. 2021). Other candidates include massive ultralight particles (ULDMS) that could reach

masses as low as  $10^{-23} \text{ eV}$  (Hui et al. 2017). Therefore, the possible DM mass range remains unconstrained today. Additionally, DM particles could interact with themselves, and have or not have spin, and other properties, including their mass, remain elusive.

Depending on the nature of the DM particle, there are relevant changes in the structure and number of DM halos and subhalos (Zavala & Frenk 2019). For example, some candidates, like warm dark matter (WDM), introduce a cut-off scale in the initial power spectrum of mass fluctuations ( $m_{\text{DM}} \sim 1 \text{ KeV}$ ), and others a scale during the non-linear evolution phase in which the DM particles self-interact (SIDM) (Tulin & Yu 2018). Both processes change the abundance of DM subhalos and the density profiles of the DM halos in comparison to the predictions of the cold dark matter (CDM) model.

Buschmann et al. (2018) show, using CDM simulations, that the gravitational pull of DM subhalos affects the distribution of stars in galactic halos, and could be used to discover dark subhalos (those without star formation in situ, a precise prediction of the CDM model) and also allow one to test the nature of the DM particles itself. This work simulated a passing DM subhalo's perturbation to the phase space stellar distribution. Stars are pulled towards the subhalo as it passes, leaving a distinctive feature in the halo stars' velocity and number density, known as a wake. This phenomenon was previously analytically described by Weinberg (1986) due to the gravitational friction that provokes the orbital decay of the satellite galaxies that inhabit the DM subhalo. There have been some efforts to quantify the magnitude of phase-space perturbations caused by the passage of DM subhalos using simulations and their possible detection (Bazarov et al. 2022).

Garavito-Camargo et al. (2019) used CDM simulations to quantify the impact of the LMC's passage on the density and

kinematics of the Milky Way (MW) DM halo and the observability of these structures in the MW's stellar halo. Their results indicated a pronounced wake, which could be decomposed in a transient response and a collective one in both the DM and stellar halo distributions. Such an effect was observed for the first time in the MW halo stars by [Conroy et al. \(2021\)](#). The authors studied the effects induced by the Magellanic Clouds System (MCS) merging in selected samples of MW stars with precise *Gaia* Satellite measurements. This detection, and the increasing availability of stellar data from *Gaia* DR3 release, has paved the way for a more precise measurement of the effect. Measurements of the wake on the perturber systems of reference will allow pursuit further testing of the DM particle's nature using this merger data, as was proposed recently by [Foote et al. \(2023\)](#) (in the context of systems of galaxies see [Furlanetto & Loeb 2002](#); [Buehler & Desjacques 2023](#)). Furthermore, [Aguilar-Argüello et al. \(2022\)](#) and [Cunningham et al. \(2020\)](#) conducted studies on the decomposition into spherical harmonics of both density and velocity, respectively, in order to quantify the response of the DM halo to the passage of the Large Magellanic Cloud (LMC).

Based on the findings of [Conroy et al. \(2021\)](#), we have used *Gaia* Data Release 3 to study the DM subhalo of the Magellanic Clouds. Our code developments will also allow us to apply the methodology to other MW satellite galaxies and globular clusters and even to develop methods of detecting the presence of the dark subhalos predicted by the CDM model.

This work is organised as follows. In Sect. 2, we present the data samples selection methodology used to identify the effects of the DM subhalo of the LMC on the MW stellar halo. In Sect. 3, we briefly describe the theoretical model. Meanwhile, we present our results in Sect. 4. Section 5 presents the conclusions and future perspectives. Finally, we present in Appendix A a description of the coordinates' transformation, in Appendix B the machine learning algorithm used to estimate radial velocities, and in Appendix C a validation of our method of inferring distances.

## 2. Data reduction

We studied the gravitational response of the MW's halo to the passage of the LMC in its orbit. To achieve this, we used the data from *Gaia* Data Release 3 ([Gaia Collaboration 2023, 2016](#)) and created two catalogues of halo tracers; namely, K-giant stars and RR-Lyrae ones. We followed the steps proposed by [Conroy et al. \(2021\)](#) to address this task.

### 2.1. K-giant dataset

To construct the K-giant catalogue, we started the analysis with 162 240 774 sources characterised by *ruwe* values below 1.4, parallax measurements lower than 0.2 mas, and galactic latitudes of  $|b| > 10^\circ$  to remove the galactic plane. To ensure data quality, we performed a series of cleaning procedures. First, we eliminated sources lacking proper motion and photometric data. To account for dust extinction, we obtained the dust map from [Green \(2018\)](#) and considered the Schlegel, Finkbeiner and Davis (SFD) map to derive the excess colour,  $E(B - V)$ . Subsequently, we discarded all sources with  $E(B - V) > 0.3$ . To obtain the corrected magnitudes, we considered the following coefficients:  $A_G/E(B - V) = 2.4$ ,  $A_{BP}/E(B - V) = 2.58$ , and  $A_{RP}/E(B - V) = 1.65$ . To focus solely on the giant branch, we restricted the selection to sources satisfying the condition  $1.4 < (BP^* - RP^*) < 2$ , where  $BP^*$  and  $RP^*$  represent the cor-

rected magnitudes. Next, following [Riello et al. \(2021\)](#), we performed the  $3\sigma$  cut upon the corrected BP and RP flux excess factor ( $C^*$ ). After completing the data-cleaning process, to ensure the purity of our catalogue specifically for K-giant stars, we performed a cross-match with the spectral types provided by *Gaia* (*gaiadr3*.astrophysical parameters) and obtained a dataset of 490 669 sources. Finally, we restricted our analysis to objects within a galactocentric distance between 30 kpc and 100 kpc, leaving 245 086 sources. Among them, only 10 989 had radial velocities measured by *Gaia* ([Katz et al. 2023](#)).

To estimate the radial velocity for the remaining sources, we employed a machine learning algorithm, specifically a Random-ForestRegressor ([Pedregosa et al. 2011](#)). The accuracy of our model is 87.0% (see Appendix B for details).

To determine the photometric distance, we used the MIST code ([Dotter 2016](#); [Choi et al. 2016](#); [Paxton et al. 2011](#)) to generate an isochrone with the specific LMC parameters, which are an age of 10 Gyr and a metallicity of  $[\text{Fe}/\text{H}] = -1.5$ . We restricted the isochrone to an effective temperature from 3800 K to 4400 K, and fitted the polynomial equation

$$M_G = 2.8894(BP^* - RP^*)^2 - 11.9263(BP^* - RP^*) + 8.7151. \quad (1)$$

To validate our distance inference method, we compared our calculated distances with reference values for some globular clusters, the Magellanic Clouds (LMC and SMC), and some satellite galaxies (see Appendix C for details). Our method successfully reproduced tabulated distances, since the mean calculated distances differed by less than 20%.

Afterwards, we implemented several masks to exclude known objects from our analysis. Specifically, we applied angular and distance masks to all the globular clusters listed in [Harris \(1996\)](#) and all the satellite galaxies reported in [Drlica-Wagner et al. \(2020\)](#).

Following the methodology proposed by [Conroy et al. \(2021\)](#), we employed proper motions to eliminate structures linked to the Sagittarius stream. To achieve this, we initially correct the proper motions due to the solar reflex motion, with the *Gala* package ([Price-Whelan 2017](#); [Price-Whelan et al. 2020](#)). The used parameters were  $R_\odot = 8.122$  kpc ([Abuter 2019](#)),  $(V_{R,\odot}, V_{\phi,\odot}, V_{Z,\odot}) = (-12.9, 245.6, 7.78)$  km s $^{-1}$  ([Drimmel & Poggio 2018](#)), and the distance of the Sun from the Galactic mid-plane,  $Z_\odot = 20.8$  pc ([Bennett & Bovy 2019](#)). For  $b > 10^\circ$  and  $|B_{\text{Sgr}}| < 15^\circ$ , where  $B_{\text{Sgr}}$  is the latitude in the frame of the Sagittarius orbital plane, we removed part of the northern arm of the stream by taking out the sources that have  $\mu_{\alpha^*} > -1.3$  mas yr $^{-1}$ ,  $-0.4 < \mu_\delta < 0.3$  mas yr $^{-1}$ , and  $\mu_\delta > 1.7\mu_{\alpha^*} + 0.4$ . To eliminate the rest of the north arm, we applied a mask to the region with coordinates  $b > 0^\circ$  and  $180^\circ < l < 210^\circ$ . The final selection of sources was based on proper motions; that is, we kept only those that satisfied  $\mu_{\alpha^*}^2 + (\mu_\delta + 0.1)^2 < 0.5^2$  ([Conroy et al. 2021](#)). By implementing this criterion, one effectively excludes disk stars, stars belonging to the Large and Small Magellanic Clouds, the Sagittarius dwarf spheroidal, and other Sagittarius arms. After this matching, our final dataset of K giants had 6058 sources.

### 2.2. RR-Lyrae dataset

To build the RR-Lyrae catalogue, we started the process by using the 271 779 sources catalogued as RR-Lyrae variables by *Gaia*. Initially, we performed data cleaning by keeping stars with *ruwe* < 1.4, and excluding those lacking metallicity, as well as those with errors in metallicity exceeding the absolute value of the metallicity itself. We then discarded all sources with

$E(B - V) > 0.3$ , and the galactic plane  $|b| < 10^\circ$ . These cuts yielded a set of 66 610 stars, of which only 2440 had radial velocity measurements provided by *Gaia*.

To increase the amount of data with measured radial velocities and metallicities, we performed a series of data cross-matching steps. We used the SEGUE catalogue (Ahn et al. 2012) to complete the radial velocities and metallicities of our dataset, using the ones measured by Sloan (cross-match with ID table *sdss\_dr17\_x1p5\_specobj\_gaia\_dr3\_gaia\_source*). In this case, no extra points were incorporated.

Additionally, we utilised the RR-Lyrae catalogue provided by Wang et al. (2022) to complete our catalogue with the sources not present in our dataset (288 points were incorporated) or to complete our radial velocity and metallicity data. After the cross-matching process, we ended up with a catalogue of 73 598 sources, of which 6523 have measured radial velocities. To obtain the radial velocity for the remaining sources, we applied a combined algorithm of data augmentation + random-forest regressor. In this case, the accuracy of our model is 49.0% (see Appendix B for details). Similar to the K-giant approach, we corrected the magnitudes to account for dust extinction. The absolute magnitude is connected to the metallicity through  $M_G = 0.32[\text{Fe}/\text{H}] + 1.11$  (Muraveva et al. 2018); therefore, one can obtain their distances using the distance modulus relationship.

To ensure consistency, we followed a similar approach to that used with the K-giants. To eliminate known objects from the Harris (1996) and Drlica-Wagner et al. (2020) catalogues, we applied an angular mask, taking into account the distance to the sources. Additionally, to exclude the Sagittarius stream from our analysis, we employed the cut-off criterion of  $|B_{Sgr}| < 15^\circ$  for  $b > 0^\circ$ . Then, we focused our analysis on objects with galactic distances between  $30 \text{ kpc} < R_{\text{gal}} < 100 \text{ kpc}$ . In the final step, we removed stars that did not satisfy the condition  $\mu_{\alpha^*}^2 + (\mu_\delta + 0.1)^2 < 0.5^2$  (Conroy et al. 2021) after performing the solar reflex motion correction to the proper motions. Therefore, the final sample has 2446 sources.

Since our aim is to extract the mass of the DM subhalo surrounding the Large and Small Magellanic Clouds, we performed a transformation of coordinates to a new reference system. This particular coordinate reference system is centred on the centre of mass (CM) of the MCS, with the  $x$  axis aligned with the direction of the velocity of the CM (see Appendix A for details), and it is considered a rest frame. In order to obtain the position and velocity of the CM, we considered the LMC mass to be nine times the SMC mass (Craig et al. 2022).

### 3. Theoretical model and likelihood analysis

We used the theoretical model for stellar wakes from DM subhalos proposed by Buschmann et al. (2018), where the authors assumed a Plummer sphere for the density profile of the DM subhalo. The reference system used in this section corresponds to the one centred on the CM of the MCS, with the  $x$  axis in the direction of the velocity of the DM subhalo mass (see Appendix A.1 for details). From the collisionless Boltzmann equation, they derived the time-independent phase-space distribution function in the subhalo rest frame

$$f(\bar{r}, \bar{v}, M_s) = f_0(\bar{v}) \left( 1 + \frac{2GM_s}{v_0^2} (\bar{v} + \bar{v}_s) \cdot \bar{\alpha} \right), \quad (2)$$

$$\bar{\alpha}(\bar{r}, \bar{v}, M_s) = \frac{1}{rv} \frac{\sqrt{1 + \frac{R_s^2}{r^2} \frac{\bar{v}}{v} - \frac{\bar{r}}{r}}}{\sqrt{1 + \frac{R_s^2}{r^2} - \frac{\bar{v} \cdot \bar{r}}{rv}}}.$$

In the equations,  $f_0(\bar{v}) = \frac{n_0 e^{-(\bar{v} + \bar{v}_s)^2/v_0^2}}{\pi^{3/2} v_0^3}$ ,  $v_0 = \sqrt{2}\sigma_v$  where  $\sigma_v$  is the velocity dispersion,  $\bar{v}_s$  and  $M_s$  are the DM subhalo velocity and mass, respectively,  $R_s(M_s) = 1.62 \sqrt{M_s/(10^8 M_\odot)}$ , and  $n_0$  is the star density inside the region of interest. This expression can be easily extended for three different velocity dispersions by including  $v_{0x}(\sigma_{v_x})$ ,  $v_{0y}(\sigma_{v_y})$ , and  $v_{0z}(\sigma_{v_z})$ . In this case, the distribution function in the subhalo rest frame can be written as

$$f(\bar{r}, \bar{v}, M_s) = \frac{n_0 e^{-\left(\frac{v_x + v_s}{v_{0x}}\right)^2 - \left(\frac{v_y}{v_{0y}}\right)^2 - \left(\frac{v_z}{v_{0z}}\right)^2}}{\pi^{3/2} v_{0x} v_{0y} v_{0z}} (1 + 2GM_s \beta), \quad (3)$$

$$\beta(\bar{r}, \bar{v}, M_s) = \left( \frac{v_x + v_s}{v_{0x}^2} \hat{i} + \frac{v_y}{v_{0y}^2} \hat{j} + \frac{v_z}{v_{0z}^2} \hat{k} \right) \cdot \bar{\alpha}.$$

In order to obtain the mass of the DM subhalo of the MCS, we performed a statistical analysis using the likelihood function to compare observational data in the new reference system and the theoretical model. This analysis was performed by using only the space data (3D) and the phase-space data (6D). The un-binned likelihood function for the 6D analysis is (Buschmann et al. 2018)

$$p_{6D}(M_s, \theta) = e^{-N_s(M_s)} \prod_{k=1}^{N_d} f(\bar{r}_k, \bar{v}_k, M_s), \quad (4)$$

where  $N_d$  is the number of stars in the region of interest (sphere of radius  $R$  centred on the CM of the MCS),  $N_s$  is the predicted number of stars in the same region, and  $\theta$  are the fixed parameters of our model; that is,  $n_0$ ,  $v_0$ ,  $R_s$ , and  $\bar{v}_s$ . For the Plummer sphere and for the distribution function of Eq. (2),  $N_s(M_s)$  can be computed as (Buschmann et al. 2018)

$$N_s(M_s) = \frac{4}{3} \pi R^3 n_0 + \frac{4\pi G M_s n_0}{v_0 v_s} \gamma F\left(\frac{v_s}{v_0}\right),$$

$$\gamma = R^2 \sqrt{1 + \frac{R_s^2}{R^2}} - R_s^2 \operatorname{arcsinh}\left(\frac{R}{R_s}\right),$$

$$F(x) = e^{-x^2} \int_0^x e^{y^2} dy.$$

For the Plummer sphere and for the distribution function of Eq. (3), the predicted number of stars is

$$N_s(M_s) = \frac{4}{3} \pi R^3 n_0 + \frac{4GM_s n_0}{v_{0x} v_{0y} v_{0z}} \gamma I(\bar{v}_0, v_s),$$

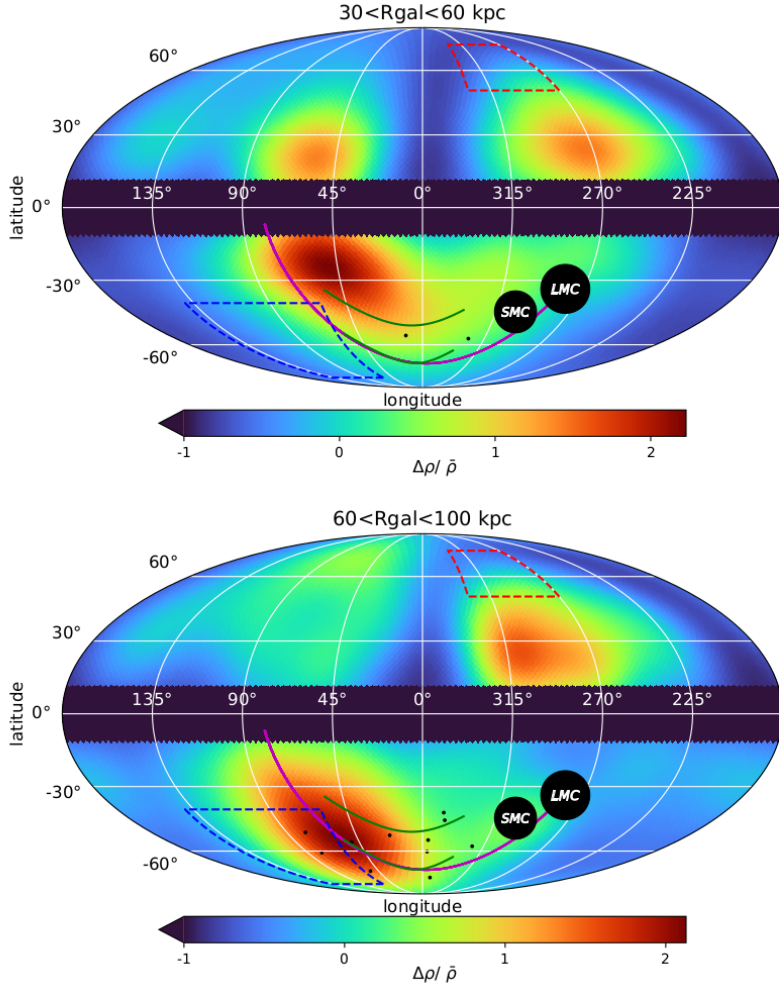
$$I(\bar{v}_0, v_s) = \int \frac{d^3 p}{2\pi} e^{-p^2} \cos\left(\frac{2v_s p_x}{v_{0x}}\right) \left( \left(\frac{p_x}{v_{0x}}\right)^2 + \left(\frac{p_y}{v_{0y}}\right)^2 + \left(\frac{p_z}{v_{0z}}\right)^2 \right)^{-1/2}.$$

For the 3D data, the un-binned likelihood function is

$$p_{3D}(M_s, \theta) = e^{-N_s(M_s)} \prod_{k=1}^{N_d} \int d^3 v f(\bar{r}_k, \bar{v}, M_s). \quad (5)$$

To determine the DM subhalo mass, we used a Markov-chain Monte Carlo (MCMC) method. For this purpose, we utilised the emcee package (Foreman-Mackey et al. 2013). We considered the function

$$\lambda(M_s) = \ln(p_x(M_s, \theta)), \quad (6)$$



**Fig. 1.** Density distribution of K-giant and RR-Lyrae variables (Mollweide projection map) with  $30 < R_{\text{gal}} < 60$  kpc (top panel) and  $60 < R_{\text{gal}} < 100$  kpc (bottom panel). The data was smoothed using an FWHM of  $30^\circ$ . The magenta line represents the past orbit of the MCS CM. The dashed blue and red lines represent the Pisces and Virgo overdensities, respectively. The black dots represent the members of the Magellanic Stellar Stream. The green lines are the Pisces Plume.

where  $x$  stands for the  $3D$  or  $6D$  analysis. The prior used was  $10 < \log_{10} M_s < 12$ . The functions,  $p_x$ , are presented in Eqs. (4) and (5). We considered 32 walkers and checked the convergence every 100 steps. To compute the uncertainties, we used the 16th, 50th, and 84th percentiles of the samples in the marginalised distributions (Foreman-Mackey et al. 2013).

#### 4. Results

In Fig. 1, we present a Mollweide projection map displaying the distribution of our final dataset of 8504 stars in galactic coordinates (6058 K giants and 2446 RR Lyraes). To enhance the visual representation, the map has been smoothed using Gaussian functions with a full width at half maximum (FWHM) of  $30^\circ$ . The colour bar represents the density contrast, indicating the relative density variation from its mean value across the sample. The past 1 Gyr orbit of the CM of the MCS is shown with a magenta line, computed with the gala package using the MilkyWayPotential (Bovy 2015). The dotted blue and red lines represent the Pisces (Chandra et al. 2023b) and Virgo overdensities (Perotoni et al. 2022), respectively. The green lines are the polynomials adjusted by Chandra et al. (2023b) that limit the Pisces Plume. The black dots represent the members of the Magellanic Stellar Stream taken from Chandra et al. (2023a). The dark blue region corresponds to the masked region representing the galactic plane.

Two distinct regions of overdensities can be observed. The first one, located in the northern hemisphere, with a longitude

range between  $225^\circ$  and  $315^\circ$ , is associated with the collective response, also known as the global response. On the other hand, the southern feature appears to cover a larger area ( $-30^\circ < l < 130^\circ$ ) and exhibits significant prominence at a longitude of  $50^\circ$  and a latitude of approximately  $-26^\circ$ . This overdensity is associated with the local wake. A more minor component is also present in the northern hemisphere, within a longitude range of  $30^\circ$ – $90^\circ$ . It appears separate from the southern component due to the masking of the galactic plane, implemented to prevent contamination. The intensity of the wake is greater than that of the collective response. The ratio between the counts per pixel of the wake at  $l = 50^\circ$ ,  $b = -26^\circ$  and the counts per pixel of the collective response at  $l = 279^\circ$ ,  $b = 24^\circ$  (the coordinates of the highest overdensity of the collective response) is 1.29, considering the complete dataset. As one can see, the CM of the MCS past orbit is located over the local wake, and the deviations could be signalling the effect of the DM mass of the wake according to the results of Foote et al. (2023).

In the upper panel of Fig. 1, an inner region of the halo is shown (sources between 30 and 60 kpc from the galactic centre). It can be noted that both the wake and the collective do not correspond to either the Pisces or the Virgo overdensities. In previous works (Belokurov et al. 2019; Conroy et al. 2021), it was noticed that there exists a subregion produced by the Magellanic Clouds called the Pisces Plume. The southern overdensity that we identified as the wake does not fall within this region.

On the other hand, in the lower panel, we plot the outer region of the halo, between 60 and 100 kpc, a region studied

by Belokurov et al. (2019) and Conroy et al. (2021). Once again, the collective does not belong to the Virgo overdensity. Nevertheless, the global response could be truncated due to the masking of the galactic plane and the Sagittarius stream. However, part of the wake lies on the edge of the Pisces overdensity, and the maximum of the wake is indeed located in the Pisces Plume. Additionally, it has been verified that none of the catalogued points belonging to the Magellanic Stellar Stream (Chandra et al. 2023a) are found in our dataset.

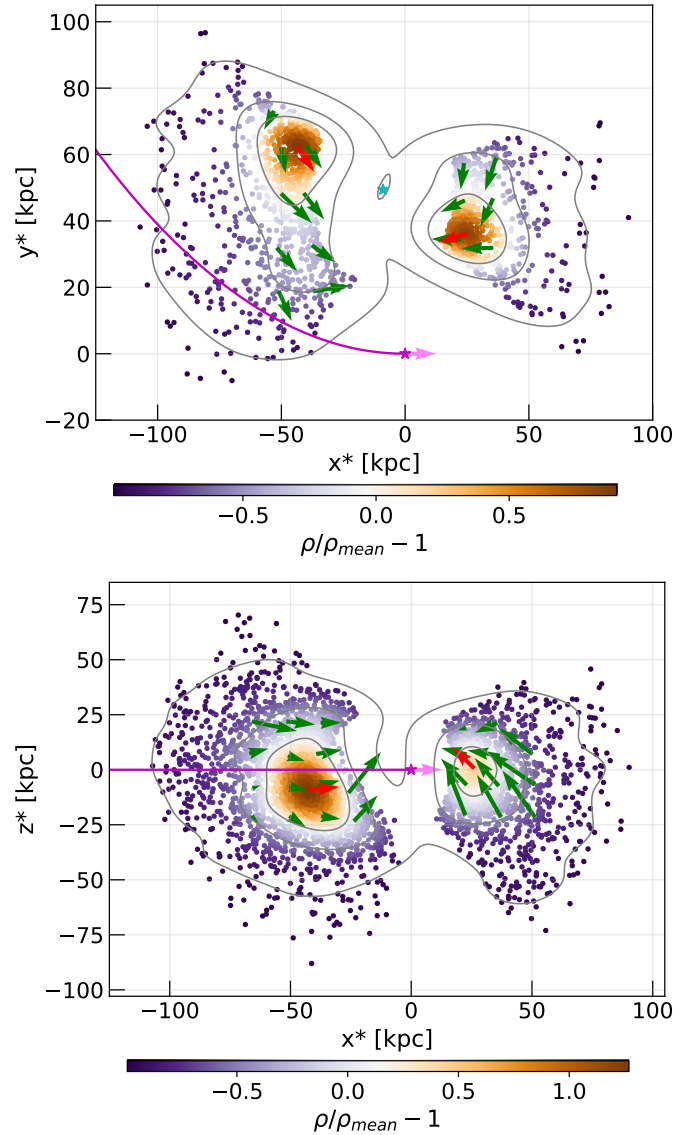
Comparing our results with the ones obtained by Conroy et al. (2021), we observe some slight differences in the locations of the overdensities. Specifically, our sample's maximum southern overdensity is slightly displaced further north. Similarly, the maximum northern overdensity in our sample is slightly shifted towards the south and east. In particular, we have also compared only our K-giant sample results with the final public catalogue developed by Conroy et al. (2021). The coordinates of the maximum overdensities of the local wake and the collective response for our dataset are ( $l = 52^\circ$ ,  $b = -24^\circ$ ) and ( $l = 275^\circ$ ,  $b = 25^\circ$ ), respectively; meanwhile, for Conroy's data they are approximately ( $l = 49^\circ$ ,  $b = -54^\circ$ ) for the wake and ( $l = 326^\circ$ ,  $b = 54^\circ$ ) for the collective. Our definition of the local wake is larger than the one proposed by Conroy et al. (2021), and the ratio between the counts per pixel of the wake and the counts per pixel of the collective response at each maximum is 1.5 for our data and 1.33 for Conroy's data-set. However, if we consider only K giants located within  $60 < R_{\text{gal}} < 100$  kpc, we successfully replicated the wake's position, with its peak occurring at  $l = 57^\circ$ ,  $b = -51^\circ$ . Moreover, when we compare our map with the simulations presented in Conroy et al. (2021), we observe quite an agreement regarding the positions of the overdensities.

In Fig. 2, we plot the superficial overdensity of the wake along with the past orbit of the CM (magenta line), in a new coordinate frame; namely, the orbit frame (see Appendix A.2). In this frame, the plane  $x^*-y^*$  contains the CM orbit and the  $z^*$  axis is perpendicular to the CM orbit. The origin of this new coordinate frame is the current location of the CM, and the  $x^*$  axis is coincident with the direction of the DM subhalo mass velocity. As one can see, the wake is located in  $x^* < 0$  and it moves towards the perturber. In particular, the lower panel of Fig. 2 is in good agreement with the results presented in Fig. 1 of Buschmann et al. (2018).

The maximum value of the density is approximately  $x^* = -48.97$  kpc,  $y^* = 55.65$  kpc, and  $z^* = -7.94$  kpc in the new orbit frame. Using the stars in a 10 kpc neighbourhood (177 stars), we computed the velocity dispersion, resulting in  $(48 \pm 3)$  km s $^{-1}$ . The characterisation of its complete stellar population and dynamical properties will be addressed in a forthcoming work.

The *Gaia* satellite data and spectroscopic surveys have unveiled the structure, composition, and formation history of the MW's stellar halo in recent years (Helmi et al. 2018; Belokurov et al. 2019; Kruijssen et al. 2020; Callingham et al. 2022). Satellite galaxies, globular clusters, stars, and streams are now associated with different halo components, which constitute the remains of our galaxy's past merger events.

In order to study the possible origin of the stellar populations of the wake and collective response, we have made an  $E$  vs.  $L_z$  diagram (see Fig. 3). We have taken the same coordinate system convention as the one adopted by Callingham et al. (2022). In this diagram, we have discriminated the different known mergers of the MW (colour points, data extracted from Callingham et al. 2022), the LMC (star), the SMC (squared), and the wake and collective response (cyan and magenta regions,

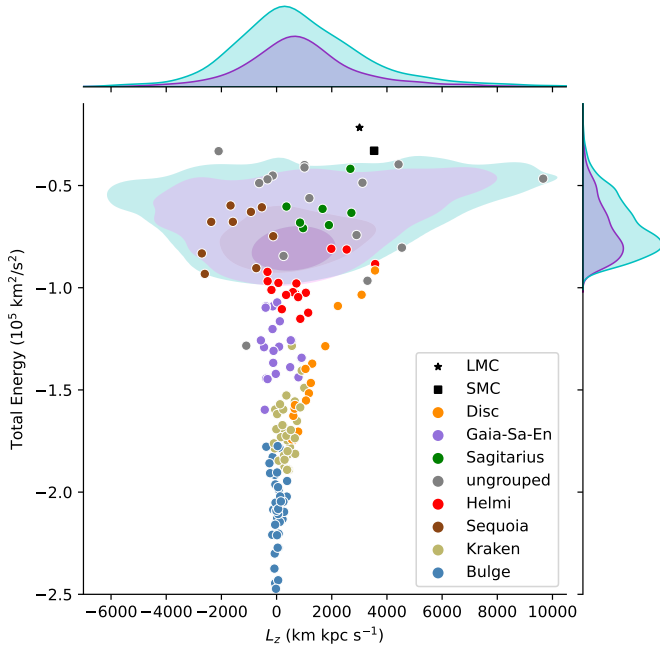


**Fig. 2.** Overdensity as a function of the position. Magenta line: past orbit of the CM. Magenta star: current position of the CM. Pink arrow: velocity of the CM. Red arrow: mean velocity of the wake or collective. Green arrow: mean velocity of a bin of 15 kpc. Cyan star: position of the galactic centre. Top panel: orbit plane with  $|z^*| < 10$  kpc. Bottom panel:  $x^*-z^*$  plane (perpendicular to the orbit plane).

respectively). As one can see, both the wake and the collective are extended regions in the diagram without a defined sign of  $L_z$  but limited in energy; however, it is not in the range of the Gaia-Sausage-Enceladus energy. The mean values for the wake are  $E = -0.703 \times 10^5$  km $^2$  s $^{-2}$  and  $L_z = 890.433$  km kpc s $^{-1}$ ; meanwhile, for the collective, they are  $E = -0.716 \times 10^5$  km $^2$  s $^{-2}$  and  $L_z = 902.391$  km kpc s $^{-1}$ .

It should be noted that the stellar population of the MW's Halo was accreted by mergers of older origin such as Sequoia, Sagittarius, Helmi, ED-3-4-5-6, Typhon, and L-RL64 (Dodd et al. 2023). Consequently, the recent impact of the LMC could dynamically affect all these stellar populations; however, more studies are needed.

Next, we performed a statistical analysis in order to obtain the DM subhalo mass of the CM, and therefore the DM subhalo mass of the LMC. The radius for the region



**Fig. 3.**  $E-L_z$  diagram. The mergers data were taken from [Callingham et al. \(2022\)](#). Cyan region: wake. Magenta region: collective response.

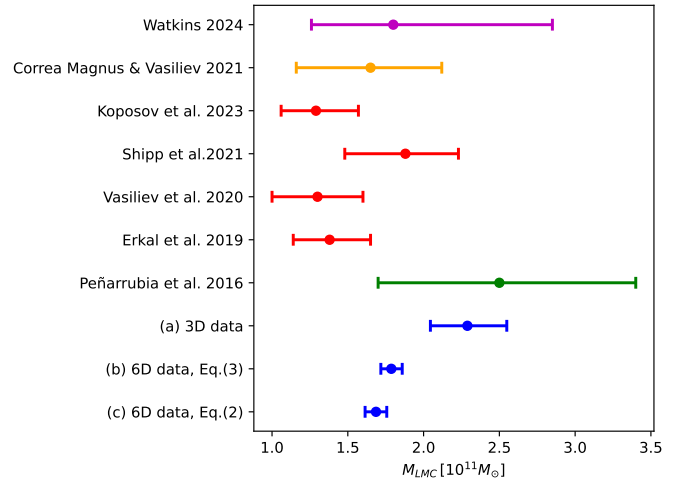
of interest was fixed at  $R = 100$  kpc, the subhalo velocity,  $v_s$ , was fixed at  $314.23 \text{ km s}^{-1}$  ([van der Marel et al. 2002](#); [Martínez-Delgado et al. 2019](#)), and  $n_0$  was obtained from the reduced observational data described in the previous section. For the velocity dispersion, we performed a statistical analysis of the data and obtained the velocity standard deviation in each axis. For the analysis using Eq. (2), we considered the larger component of  $\bar{v}_0$  in the calculations. We used the described density profile in Sect. 3 and we present our results in Fig. 4, along with the DM subhalo mass estimation of the LMC published in the literature ([Watkins et al. 2024](#); [Koposov et al. 2023](#); [Shipp et al. 2021](#); [Vasiliev et al. 2021](#); [Erkal et al. 2019](#); [Peñarrubia et al. 2016](#)); using a different method, indicated with colours) along with our results (the last three values, with their corresponding statistical errors).

The DM subhalo mass was computed by using the space distribution function (case (a)) and the phase-space distribution function of Eq. (3) (case (b)) and Eq. (2) (case (c)). Our results are consistent, despite the distribution function used in the analysis. However, the fit obtained using only the space data is slightly higher than the 6D analysis results. Furthermore, our findings agree within  $3\sigma$  with the literature ([Vasiliev 2023](#)).

We also performed the statistical analysis (6D Eq. (2)) using only the data with measured radial velocity ([Gaia Collaboration 2023](#); [Ahn et al. 2012](#); [Wang et al. 2022](#)). We found an LMC subhalo mass of  $M_{\text{LMC}} = 1.594^{+0.203}_{-0.196} \times 10^{11} M_{\odot}$ , which is in good agreement with our results.

## 5. Conclusions

In this work, we employed the recently published *Gaia* Data Release 3, which improves the precision of proper motions along with the Segue catalogue ([Ahn et al. 2012](#)) and the one provided by [Wang et al. \(2022\)](#). This enabled us to extend the K-giant catalogue originally provided by [Conroy et al. \(2021\)](#), and also to construct a catalogue for RR-Lyrae stars, both in 6D data. We



**Fig. 4.** Dark matter subhalo mass estimation of the LMC. Orange line: kinematic estimation from MW satellites. Red lines: estimation from stellar streams. Green line: estimation based on momentum balance in the Local Group. Blue lines: our results obtained from the likelihood analysis. Case (a): 3D data,  $M_{\text{LMC}} = 2.289^{+0.260}_{-0.240} \times 10^{11} M_{\odot}$ . Case (b): 6D data (Eq. (3)),  $M_{\text{LMC}} = 1.787^{+0.072}_{-0.069} \times 10^{11} M_{\odot}$ . Case (c): 6D data (Eq. (2)),  $M_{\text{LMC}} = 1.686^{+0.071}_{-0.072} \times 10^{11} M_{\odot}$ . The error bars of our results are purely statistical and based on a restricted one-parameter model.

reproduced the previously published results and identified the overdensities associated with a wake and the collective response using these two halo tracers. A notable finding of this study is the extension of the southern overdensity towards lower galactocentric distances; that is, between 30 and 100 kpc. Moreover, we were able to show that the southern overdensity, identified as the wake, trails the CM of Magellanic Clouds (see Fig. 2).

We have confirmed that the Pisces plume overdensity, described in [Belokurov et al. \(2019\)](#), is associated with the wake of the Magellanic Clouds in the outer regions of the MW's halo. Furthermore, we have discovered that the overdensity on the halo's stellar population, caused by the Magellanic Clouds' wake and the global response, affects the stars in the MW's halo, regardless of which past merger event they were accreted from.

As per the theoretical proposal made by [Buschmann et al. \(2018\)](#), we were able to estimate the mass of the LMC DM subhalo for the first time by using *Gaia* observational data. We found a reliable estimation of the DM halo surrounding the LMC by performing two different analyses, using only the space distribution data and using both the phase and space data. Considering a relationship between the Large and Small Magellanic Clouds' masses, our study has successfully determined the mass of the DM subhalo of the larger cloud. Even more, our findings are in agreement with prior results ([Correa Magnus & Vasiliev 2021](#); [Koposov et al. 2023](#); [Shipp et al. 2021](#); [Vasiliev et al. 2021](#); [Erkal et al. 2019](#); [Peñarrubia et al. 2016](#)), within  $3\sigma$ . This consistency with previous studies indicates the reliability of our methodology. Additionally, this method gives competitive errors compared to different mass determination methods. It is important to point out that the errors mentioned were calculated using the assumptions of a Plummer profile for a spherical subhalo. However, if a more complex multi-parameter model such as an ellipsoidal Navarro, Frenk and White (NFW) model were used, it is anticipated that the errors would increase.

*Acknowledgements.* This work was partially supported by grants from the National Research Council of Argentina (CONICET PIP 616). K.J.F. is a Post

Doctoral fellow of the CONICET. M.E.M. and M.D. are members of the Scientific Research Career of the CONICET. M.D. work was supported by the Preparing for Astrophysics with LSST Program, funded by the Heising Simons Foundation through grant 2021-2975, and administered by Las Cumbres Observatory. We would like to express our gratitude to the referee for providing us with their valuable comments and suggestions. M.D. thanks valuable suggestions by Nicolás Garavito-Camargo and the support of the CCA and the Flatiron Institute. M.D. also thanks the possibility of virtual participation in the following programs: “Building a physical understanding of galaxy evolution with data-driven astronomy” at the KITP-UCB, the Chicago “*Gaia* DR3 Sprint” and “Streams22: Community Atlas of Tidal Streams.” at the Carnegie Observatories and the Flatiron Institute. K.J.F. wants to thank Alejo Molina Lera and Martín Gamboa Lereña for the useful discussions. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC) (<https://www.cosmos.esa.int/web/gaia/dpac/consortium>), and code developed by the *Gaia* Project Scientist Support Team. Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the *Gaia* Multilateral Agreement. This research has utilized the following software: Astropy (Astropy Collaboration 2013), Matplotlib (Hunter 2007), Pandas (McKinney et al. 2010), Seaborn (Waskom 2021), Healpy (Zonca et al. 2019), SciPy (Virtanen et al. 2020), NumPy (Harris et al. 2020), emcee (Foreman-Mackey et al. 2013), gala (Price-Whelan 2017; Price-Whelan et al. 2020), The Jupyter Notebook (Kluyver et al. 2016), Scikit-learn (Pedregosa et al. 2011), Pzflow (Crenshaw et al. 2024), and TOP-CAT (Taylor 2005).

## References

- Abdalla, H., Aharonian, F., Benkhali, F. A., et al. 2022, *Phys. Rev. Lett.*, **129**, 111101
- Abe, H., Abe, S., Acciari, V. A., et al. 2023, *Phys. Rev. Lett.*, **130**, 061002
- Acharyya, A., Archer, A., Bangale, P., et al. 2023, *ApJ*, **945**, 101
- Aguilar-Arquiello, G., Valenzuela, O., & Trelles, A. 2022, *A&A*, **663**, A93
- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, *ApJS*, **203**, 21
- Amaré, J., Cebrián, S., Cintas, D., et al. 2022, *Moscow Univ. Phys. Bull.*, **77**, 322
- Aprile, E., Abe, K., Agostini, F., et al. 2023, *Phys. Rev. Lett.*, **131**, 041003
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Barberio, E., Baroncelli, T., Bignell, L. J., et al. 2023, *Eur. Phys. J. C*, **83**, 878
- Bazarov, A., Benito, M., Hütsi, G., et al. 2022, *Astron. Comput.*, **41**, 100667
- Belokurov, V., Deason, A. J., Erkal, D., et al. 2019, *MNRAS*, **488**, L47
- Bennett, M., & Bovy, J. 2019, *MNRAS*, **482**, 1417
- Bernabei, R., Belli, P., Bussoletti, A., et al. 2022, in *The Fifteenth Marcel Grossmann Meeting on General Relativity*, eds. E. S. Battistelli, R. T. Jantzen, & R. Ruffini, 1285
- Bovy, J. 2015, *ApJS*, **216**, 29
- Breiman, L. 2001, *Mach. Learn.*, **45**, 5
- Buehler, R., & Desjacques, V. 2023, *Phys. Rev. D*, **107**, 023516P
- Buschmann, M., Kopp, J., Safdi, B. R., & Wu, C.-L. 2018, *Phys. Rev. Lett.*, **120**, 211101
- Callingham, T. M., Cautun, M., Deason, A. J., et al. 2022, *MNRAS*, **513**, 4107
- Carr, B., & Kühnel, F. 2020, *Ann. Rev. Nucl. Part. Sci.*, **70**, 355
- Chandra, V., Naidu, R. P., Conroy, C., et al. 2023a, *ApJ*, **956**, 110
- Chandra, V., Naidu, R. P., Conroy, C., et al. 2023b, *ApJ*, **951**, 26
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
- Clowe, D., Bradac, M., Gonzalez, A. H., et al. 2006, *ApJ*, **648**, L109
- Conroy, C., Naidu, R. P., Garavito-Camargo, N., et al. 2021, *Nature*, **592**, 534
- Correa Magnus, L., & Vasiliev, E. 2021, *MNRAS*, **511**, 2610
- Craig, P. A., Chakrabarti, S., Baum, S., & Lewis, B. T. 2022, *MNRAS*, **517**, 1737
- Crenshaw, J. F., Connolly, A., & Kalmbach, B. 2021, *Am. Astron. Soc. Meet. Abstr.*, **53**, 230.01
- Crenshaw, J. F., Connolly, A., & Kalmbach, B. 2024, <https://doi.org/10.5281/zenodo.10710271>
- Cunningham, E. C., Garavito-Camargo, N., Deason, A. J., et al. 2020, *ApJ*, **898**, 4
- Dodd, E., Callingham, T. M., Helmi, A., et al. 2023, *A&A*, **670**, L2
- Dotter, A. 2016, *ApJS*, **222**, 8
- Drimmel, R., & Poggio, E. 2018, *Res. Notes Am. Astron. Soc.*, **2**, 210
- Drlica-Wagner, A., Bechtol, K., Mau, S., et al. 2020, *ApJ*, **893**, 47
- Durkan, C., Bekasov, A., Murray, L., & Papamakarios, G. 2019, arXiv e-prints [arXiv:1906.04032]
- Erkal, D., Belokurov, V., Laporte, C. F. P., et al. 2019, *MNRAS*, **487**, 2685
- Foote, H. R., Besla, G., Mocz, P., et al. 2023, *ApJ*, **954**, 163
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Furlanetto, S. R., & Loeb, A. 2002, *ApJ*, **565**, 854
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, **674**, A1
- Garavito-Camargo, N., Besla, G., Laporte, C. F. P., et al. 2019, *ApJ*, **884**, 51
- GRAVITY Collaboration (Abuter, R., et al.) 2019, *A&A*, **625**, L10
- Green, G. 2018, *J. Open Source Softw.*, **3**, 695
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, **585**, 357
- Harris, W. E. 1996, *AJ*, **112**, 1487
- Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, *Nature*, **563**, 85
- Hui, L., Ostriker, J. P., Tremaine, S., & Witten, E. 2017, *Phys. Rev. D*, **95**, 043541
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, *A&A*, **674**, A5
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, *Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows* (IOS Press), 87
- Koposov, S. E., Erkal, D., Li, T. S., et al. 2023, *MNRAS*, **521**, 4936
- Kruijssen, J. M. D., Pfeffer, J. L., Chevance, M., et al. 2020, *MNRAS*, **498**, 2472
- Martínez-Delgado, D., Vivas, A. K., Grebel, E. K., et al. 2019, *A&A*, **631**, A98
- Massey, R., Kitching, T., & Richard, J. 2010, *Rep. Progr. Phys.*, **73**, 086901
- McConnachie, A. W. 2012, *AJ*, **144**, 4
- McKinney, W. 2010, in *Proceedings of the 9th Python in Science Conference*, eds. S. van der Walt, & J. Millman, 51
- Mo, H., van den Bosch, F., & White, S. 2010, *Galaxy Formation and Evolution* (Cambridge: Cambridge University Press)
- Muraveva, T., Delgado, H. E., Clementini, G., Sarro, L. M., & Garofalo, A. 2018, *MNRAS*, **481**, 1195
- Naik, A. P., & Widmark, A. 2023, *MNRAS*, **527**, 11559
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, **192**, 3
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Peñarrubia, J., Gómez, F. A., Besla, G., Erkal, D., & Ma, Y.-Z. 2016, *MNRAS*, **456**, L54
- Perotoni, H. D., Limberg, G., Amarante, J. A. S., et al. 2022, *ApJ*, **936**, L2
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Price-Whelan, A. M. 2017, *J. Open Source Softw.*, **2**, 388
- Price-Whelan, A., Sipőcz, B., Lenz, D., et al. 2020, <https://doi.org/10.5281/zenodo.4159870>
- Riello, M., De Angeli, F., Evans, D. W., et al. 2021, *A&A*, **649**, A3
- Rubin, V. C., Ford, W., & Kent, J. 1970, *ApJ*, **159**, 379
- Shipp, N., Erkal, D., Drlica-Wagner, A., et al. 2021, *ApJ*, **923**, 149
- Taylor, M. B. 2005, in *Astronomical Data Analysis Software and Systems XIV*, eds. P. Shopbell, M. Britton, & R. Ebert, *ASP Conf. Ser.*, **347**, 29
- Tulin, S., & Yu, H.-B. 2018, *Phys. Rep.*, **730**, 1
- van der Marel, R. P., Alves, D. R., Hardy, E., & Suntzeff, N. B. 2002, *AJ*, **124**, 2639
- Vasiliev, E. 2023, *Galaxies*, **11**, 59
- Vasiliev, E., Belokurov, V., & Erkal, D. 2021, *MNRAS*, **501**, 2279
- Villanueva-Domingo, P., Mena, O., & Palomares-Ruiz, S. 2021, *Front. Astron. Space Sci.*, **8**
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, **17**, 261
- Wang, F., Zhang, H.-W., Xue, X.-X., et al. 2022, *MNRAS*, **513**, 1958
- Waskom, M. L. 2021, *J. Open Source Softw.*, **6**, 3021
- Watkins, L. L., van der Marel, R. P., & Bennet, P. 2024, *ApJ*, **963**, 84
- Weinberg, M. D. 1986, *ApJ*, **300**, 93
- Zavala, J., & Frenk, C. S. 2019, *Galaxies*, **7**, 81
- Zonca, A., Singer, L., Lenz, D., et al. 2019, *J. Open Source Softw.*, **4**, 1298
- Zwicky, F. 1933, *Helv. Phys. Acta*, **6**, 110

## Appendix A: Large Magellanic Cloud rest frame

### A.1. Centre-of-mass rest frame

The coordinate system used to compare the data with the theoretical model has its origin in the LMC and SMC CM and the  $x$  axis orientated according to the DM subhalo velocity,  $\bar{v}_s$ . In order to obtain the coordinates of our dataset in such a rest frame, we performed the following steps:

1. we boosted the data to the new frame by  $\bar{r}_{boost} = \bar{r}_{obs} - \bar{r}_{cm}$ ;
2. we performed the rotations upon the boosted data using the matrix

$$M = \begin{pmatrix} \cos \theta_1 \cos \theta_2 & \sin \theta_1 \cos \theta_2 & \sin \theta_2 \\ -\sin \theta_1 & \cos \theta_1 & 0 \\ -\cos \theta_1 \sin \theta_2 & -\sin \theta_1 \sin \theta_2 & \cos \theta_2 \end{pmatrix}, \quad (\text{A.1})$$

to obtain the coordinate  $(x', y', z')$  in the new rest frame. The velocity had to be transformed as well and the final velocity had to be boosted using  $\bar{v}_s = v_s \hat{l}'$ . The angles are defined as

$$\tan \theta_1 = \frac{(v_{cm})_y}{(v_{cm})_x}, \quad (\text{A.2})$$

$$\tan \theta_2 = \frac{(v_{cm})_z}{\sqrt{(v_{cm})_x^2 + (v_{cm})_y^2}}, \quad (\text{A.3})$$

where  $((v_{cm})_x, (v_{cm})_y, (v_{cm})_z)$  is the CM velocity in the solar coordinate system.

### A.2. Orbit frame

The coordinate system used to compare our results with the Fig. 1 presented by Buschmann et al. (2018) has its origin in the CM of the MCS, its  $z^*$  axis perpendicular to the orbital plane, and the  $x^*$  axis orientated in the direction of the velocity of the DM subhalo. To obtain these new coordinates from the CM rest frame, we had to perform a rotation of an angle,  $\theta_{orb}$ , according to the following matrix:

$$M_{orb} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{orb} & \sin \theta_{orb} \\ 0 & -\sin \theta_{orb} & \cos \theta_{orb} \end{pmatrix}, \quad (\text{A.4})$$

where  $\tan \theta_{orb} = \bar{z}_{orb} / \bar{y}_{orb}$ ,  $\bar{z}_{orb}$  ( $\bar{y}_{orb}$ ) is the mean value of the  $z$  ( $y$ ) coordinate in the CM rest frame of the CM's past orbit.

## Appendix B: Estimation of radial velocities using machine learning

To complete the phase information for all the halo stars in our sample, we applied two machine learning techniques to assign radial velocities to those stars without such measurements.

### B.1. K-giant radial velocity

Our first sample consists of 245086 K-giant stars, of which 10989 have measured radial velocities. The last subsample was used to train a random forest regressor (RF) (Breiman 2001). The chosen predictors for this study were the angular coordinates, proper motions, G magnitudes, BP and RP colours, distances to the Sun and the galactic centre, and the Galactocentric Cartesian coordinates. To prevent overfitting, a standard cross-validation analysis was performed. The RF hyper-parameters were tuned using GridSearchCV from the Scikit Learn library, resulting in the following values: [ $n\_estimators = 4900$ ,  $random\_state = 0$ ]

**Table B.1.** Comparison between the mean radial velocity inferred with machine learning and the values tabulated for MW satellite galaxies (McConnachie 2012).

Galaxy	$V_r$ tabulated [km/s]	$V_r$ inferred [km/s]
LMC	262.2	250.62±30.40
SMC	145.6	133.64±30.51
Carina	222.9	213.00±41.37
Draco	-291.0	-186.10±45.91
Sculptor	111.4	-0.88±149.37

### B.2. RR-Lyrae radial velocity

We observed a decrease in the number of measured radial velocities for RR-Lyrae stars, resulting in a drop in the fraction of measured radial velocities to (5510/67276), considering galactocentric distances between 10 to 100 kpc. To tackle this issue, we aimed to model the spatial distribution of radial velocities. To achieve this, we utilised normalising flows (NFs) (Durkan et al. 2019), as have been implemented by Crenshaw et al. (2021), to model the joint posterior probability of radial velocities and predictors using a subsample of the features described earlier. However, we limited the feature space to prevent any bias against less luminous stars. Therefore, we did not consider the magnitude and colours of stars as predictors in this case.

We used probabilistic modelling to generate a radial velocity distribution outcome for a significant number of stars (100000). Therefore, the normalizing flow was used to augment and generalise the training dataset. We evaluated the marginal probability of the radial velocity given the values of other variables (predictors) and obtained a vector of probabilities. To complement this method, we used the RF algorithm to map the posterior conditional radial velocity distribution to the real measured value in the training sample, similar to the method used for K-giant stars. The RF hyper-parameters were tuned using GridSearchCV from the Scikit Learn library, resulting in the following values: [ $max\_depth = 50$ ,  $max\_features = 8$ ,  $min\_samples\_leaf = 1$ ,  $min\_samples\_split = 6$ ,  $n\_estimators = 800$ ]. It is important to note that the result of the NFs is a vector that represents the conditional radial velocity. The vector is measured on a 1000-dimensional grid that samples velocities ranging from -700 to 700 km/s. To prevent overfitting, cross-validation was employed by splitting the data in two and using 80% for training and 20% for validation, similar to the previous case.

### B.3. Goodness of fit in radial velocity regression

R-Squared ( $R^2$ ), or the coefficient of determination, is a statistical measure used to determine the proportion of variance in the dependent variable that can be explained by the independent variable in a regression model. The statistics were calculated for our two samples of stars, resulting in values of 0.86 for K-giant stars and 0.5 for RR-Lyrae ones, respectively. These values are comparable to those recently reported by Naik & Widmark (2023) using Bayesian neural networks, and provide us with a complete sample of halo stars in the phase space.

We tested the inferred stars' radial velocities by measuring the mean radial velocities of satellite galaxies and compared them with the tabulated velocity values (McConnachie 2012). The results are presented in Table. B.1. Our machine-learning results show that the average velocity of the stars on these satellite galaxies is well reproduced. The relevant error of the aver-



age velocity for the Sculptor galaxy is due to the lack of any measured radial velocity on this object.

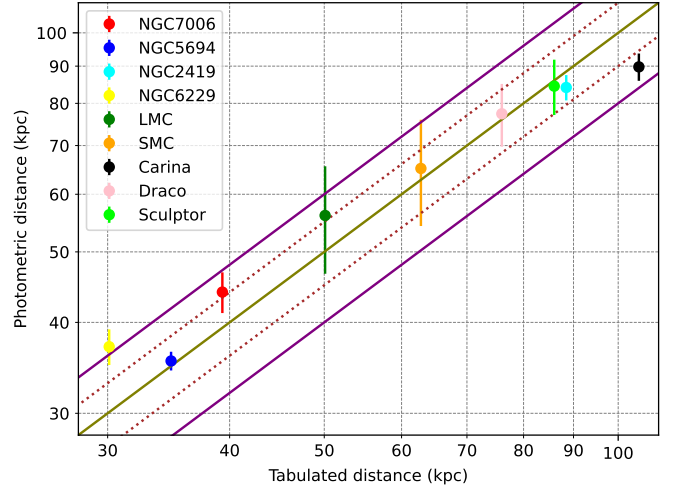
Using this method, we can estimate the mean velocity field in the halo, which is necessary for the likelihood estimation of the subhalo mass of the LMC.

### Appendix C: Validation of estimated distances for K-giant stars

To test the photometric distance obtained for the K giants, we computed the photometric distances of the globular clusters NGC7006, NGC5694, NGC2419, and NGC6229, and of the LMC, SMC, Carina, Draco, and Sculptor.

The observational data was extracted from *Gaia* Data Release 3 (Gaia Collaboration 2023, 2016), using the option single object searcher to avoid field-contamination. Additionally, the data for the LMC and the SMC were selected on a circular disk in  $(l, b)$  centred on the location of each MC, with a radius of  $2^\circ$ . For Carina, Draco, and Sculptor we selected data using an angular mask of three times the tidal radius reported in Drlica-Wagner et al. (2020). We performed the data reduction or treatment indicated in the text for the K giant (that is, corrected for dust extinction, discarded all sources with  $E(B - V) > 0.3$ , corrected the magnitudes, and performed the  $3\sigma$  cut upon the corrected BP and RP flux excess factor ( $C^*$ ). Finally, we selected the giant branch and performed the cross-match with the K-giant catalogue given by *Gaia* (*gaiadr3.astrophysical* parameters). For the LMC and SMC, we also restricted the dataset in the reported *Gaia* parallax.

We calculated the photometric distance of each source and the mean value of each object. The results are shown in



**Fig. C.1.** Comparison between the photometric distances computed using Eq. (1) and the tabulated distances (Harris 1996; Drlica-Wagner et al. 2020). Solid olive line: photometric distance equals tabulated distance. Dashed brown line:  $\pm 10\%$  with respect to the one-to-one line. Solid purple line:  $\pm 20\%$  with respect to the one-to-one line. The vertical lines are the statistical errors.

Fig. C.1 and, as can be seen, the computed photometric distances and the values reported in the literature (Harris 1996; Drlica-Wagner et al. 2020) are in good agreement.

We have also applied our fit to K-giant stars from the Conroy et al. (2021) catalogue and found that our calculated distances are in agreement with the distances reported by Conroy et al. (2021) within 10%.