Astronomy & Astrophysics manuscript no. manuscrito
April 23, 2024

Letter to the Editor

A determination of the LMC dark matter subhalo mass using the MW halo stars in its gravitational wake.

K. J. Fushimi[1], M. E. Mosquera[1,2], and M. Dominguez[3]

[1] 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

[2] Dept. of Physics, University of La Plata, c.c. 67 1900, La Plata, Argentina.

[3] Instituto de Astronomía Teórica y Experimental, (IATE-CONICET), Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina.CONICET, CCT Córdoba

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 (MW).
Methods. We employed the Gaia Data Release 3 to construct a halo tracers data set consisting of K-Giant stars and RR-Lyrae variables. Additionally, we have compared the data with a theoretical model to estimate the DM subhalo mass.
Results. We have improved the characterisation of the local wake and the collective response due to the LMC orbit. On the other hand, we have estimated for the first time the dark subhalo mass of the Large Magellanic Cloud, **of the order of** 1.7×10^{11} $\mathbf{M}_{\odot}$, comparable to previously reported values in the literature.

Key words. (Galaxies:) Magellanic Clouds– (Cosmology:) dark matter – Galaxy: halo – Galaxies: kinematics and dynamics

1. Introduction

Dark matter (DM) is central to the standard cosmological model (LCDM), providing gravitational support for forming the 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 & Ford 1970), strong and weak gravitational lensing effects (Massey et al. 2010; Clowe et al. 2006), and even the presence of the baryonic acoustic oscillation (Planck Collaboration et al. 2020) in the earlier gravitational wells observed in the cosmic microwave background (CMB).

Despite of these successes, we still lack of precise detection in the laboratory (Bernabei et al. 2022; XENON Collaboration et al. 2023; Barberio et al. 2022; 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 **from physics** beyond the standard model (BSMP), like Weakly Interacting Massive Particles (WIMPS). These massive ($m_{DM} \sim 100$ GeV) particles could weakly interact with nucleons, and therefore signals are looked at by several laboratory and accelerator experiments. Also, its annihilation signals could be detectable through γ-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 limited to small windows in mass (Villanueva-Domingo et al. 2021). Other candidates include massive ultralight particles (ULDM) that could reach masses as low as 10^{-23} eV (Hui et al. 2017). Therefore the possible DM mass range remains unconstrained today. Additionally, DM particles could interact with themselves, have or have no spin, and other properties, including their mass, remain elusive.

Depending on the nature of the dark matter particle, there are relevant changes in the structure and number of dark matter haloes 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 ($m_{DM} \sim 1$ KeV), and others a scale during the non-linear evolution phase where the DM particles self-interact (SIDM) (Tulin & Yu 2018). Both processes change the abundance of dark matter 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) showed, using CDM simulations, that the gravitational pull of dark matter subhalos affects the distribution of stars in galactic haloes, and that could be used to discover dark subhalos (those without star

formation in situ, a precise prediction of the CDM model) and also allows to test the nature of the DM particles itself. This work simulated a passing dark matter subhalo's perturbation to the phase space stellar distribution. Stars are pulled towards the subhalo as it passes, leaving a distinctive feature in 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 are some efforts to quantify the magnitude of phase-space perturbations caused by the passage of dark matter subhalos using simulations and their possible detection (Bazarov et al. 2022).**

Garavito-Camargo et al. (2019) using CDM simulations quantified the impact of the LMC's passage on the density and kinematics of the Milky Way's (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 and a Collective response in both the DM and stellar halos distributions. Such effect was observed for the first time in the Milky Way halo stars by Conroy et al. (2021). The authors studied the effects induced by the Magellanic Clouds system merging on selected samples of Milky Way stars with precise Gaia Satellite measurements. This detection, and the increasing availability of stellar data from Gaia DR3 release, pave the way to a more precise measurement of the effect. Measurements of the wake on the perturber systems of reference will allow pursuit testing the DM particle 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 dark matter halo to the passage of the Large Magellanic Cloud.

Based on the findings of Conroy et al. (2021), we have used the Gaia Data Release 3 to study the DM subhalo of the Magellanic Clouds and our code developments will allow us to apply the methodology to other MW satellites galaxies and Globular clusters and even to develop methods to detect the presence of the dark subhalos predicted by the CDM model.

This work is organised as follows. In Section 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 Section 3, we briefly describe the theoretical model. Meanwhile, we present our results in Section 4. Section 5 presents the conclusions and future perspectives. **Finally, in Appendix A we present a description of the coordinates transformation, in Appendix B the machine learning algorithm used to estimate radial velocities and in Appendix C validation of our method to infer distances.**

2. Data reduction

We studied the gravitational response of the Milky Way's Halo to the passage of the Large Magellanic Cloud in its orbit. To achieve this, we used the data from the Gaia Data Release 3 (Gaia Collaboration et al. 2022, 2016) and created two catalogues of Halo tracers, namely the K-Giants and RR-Lyraes stars. We followed the steps proposed by Conroy et al. (2021) to address this task.

2.1. K-Giant data set

To construct the K-Giant catalogue, we start the analysis with **162240774** sources characterised by *ruwe* values below 1.4, parallax measurements **lower than 0.2 mas and galactic latitude** $|b| > 10^\circ$ **to remove the galactic plane**. To ensure data quality, we performed a series of cleaning procedures. First, we eliminated sources lacking of proper motion and photometric data. To account for dust extinction, we obtained the dustmap from Green (2018) and considered the 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 corrected magnitudes. **Next**, following Riello et al. (2021), we performed the 3σ 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),** we obtained a data set of **490669** sources. **Finally, we restricted our analysis to objects within a galactocentric distance between** 30 **kpc and** 100 **kpc, remaining 245086 sources.** Among them, only **10989** had measured radial velocity from Gaia (Katz et al. 2023).

To estimate the radial velocity for the remaining sources, we employed a machine learning algorithm, specifically a RandomForestRegressor (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's parameters, that are an age of 10 Gyr and metallicity of [Fe/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 distances** masks for **all the globular clusters listed in Harris (1996a) and all the satellite galaxies reported in Drlica-Wagner et al. (2020a)**.

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 (GRAVITY Collaboration et al. 2019), $(V_{R,\odot}, V_{\phi,\odot}, V_{Z,\odot}) = (-12.9, 245.6, 7.78)$ km/s (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 $|B_{Sgr}| < 15^\circ$, where B_{Sgr} is the latitude in the frame of Sagittarius orbital plane, we remove part of the northern arm of the stream by taking out the sources that have $\mu_{\alpha*} > -1.3$ mas/yr, $-0.4 < \mu_\delta < 0.3$ mas/yr 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 is based on proper motions, that is keeping only those that satisfy $\mu^2_{\alpha*}+(\mu_\delta+0.1)^2 < 0.5^2$ (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 data set of K-Giant has **6058** sources.

2.2. RR-Lyrae data set

To build the RR-Lyraes catalogue, we started the process by using the 271779 sources catalogued as RR-Lyraes 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 66610 stars, out 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 use 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 data set (288 points where incorporated) or to complete our data with radial velocity and metallicity.

After the cross-matching process, we ended up with a catalogue of 73598 sources, of which 6523 have measured radial velocities.

To obtain the radial velocity for the remaining sources, we **apply 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 $M_G = 0.32[\mathrm{Fe/H}] + 1.11$ (Muraveva et al. 2018), therefore, one can obtain their distances using the distance modulus relationship.

To ensure consistency, we followed a similar approach used with the K-Giant. To eliminate known objects **from the Harris (1996a) and Drlica-Wagner et al. (2020a) 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 kpc $< R_{gal} <$ 100 kpc. In the final step, we removed stars that did not satisfy the condition $\mu^2_{\alpha*} + (\mu_\delta + 0.1)^2 < 0.5^2$ (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 dark matter subhalo surrounding the Large and Small Magellanic Cloud, we have performed a transformation of coordinates to a new reference system. This particular coordinate reference system is centred in the centre of mass (CM) of the Magellanic Clouds **System (MCS)**, with the x-axis aligned with the direction of the velocity of the centre of mass (see appendix A for details), and it is considered a rest frame. In order to obtain the position and velocity of the centre of mass we have considered that the LMC mass is nine times the SMC mass (Craig et al. 2021).

3. Theoretical model and Likelihood analysis

We used the theoretical model for stellar wakes from dark matter 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 in the perturbed (in our case the centre of mass of the Magellanic Clouds System), with the x-axis in the direction of the velocity of the dark matter 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

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

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

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 σ_v is the velocity dispersion, $\bar{v}_s$ and M_s are the DM subhalo velocity and mass respectively, $R_s(M_s) = 1.62\sqrt{M_s/(10^8\mathrm{M}_\odot)}$ and n_0 is the star density inside the region of interest. This expression can be easily extended for three different velocity dispersion by including v_{0x} (σ_{v_x}), v_{0y} (σ_{v_y}) and v_{0z} (σ_{v_z}). In this case, the distribution function in the subhalo rest frame can be written as

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

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

In order to obtain the mass of the DM subhalo of the MCS, we have 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)

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

where N_d is the number of stars in the region of interest (sphere of radius R centred in the centre of mass of the MCS), N_s is the predicted number of stars in the same region **and θ 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)**

$$N_s(M_s) = \frac{4}{3}\pi R^3 n_0 + \frac{4\pi G M_s n_0}{v_0 v_s}\,\gamma\, F\left(\frac{v_s}{v_0}\right) ,$$
$$\gamma = R^2\sqrt{1+\frac{R_s^2}{R^2}} - R_s^2 \operatorname{arcsinh}\left(\frac{R}{R_s}\right) ,$$
$$F(x) = e^{-x^2}\int_0^x e^{y^2}\,\mathrm{d}y\,.$$

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

$$N_s(M_s) = \frac{4}{3}\pi R^3 n_0 + \frac{4 G M_s n_0}{v_{0x} v_{0y} v_{0z}}\,\gamma\, I\left(\bar{v}_0, v_s\right) ,$$
$$I\left(\bar{v}_0, v_s\right) = \int \frac{\mathrm{d}^3 p}{2\pi}\, e^{-p^2} \cos\left(\frac{2 v_s p_x}{v_{0x}}\right) \left(\left(\frac{p_x}{v_{0x}}\right)^2 + \left(\frac{p_y}{v_{0y}}\right)^2 + \left(\frac{p_z}{v_{0z}}\right)^2\right)^{-1/2} .$$

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

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

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

$$\lambda(M_s) = \ln\left(p_x\left(M_s, \theta\right)\right) , \tag{6}$$

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 have considered 32 walkers and check the convergence every 100 steps. To compute the uncertainties, we have used the percentiles 16^{th}, 50^{th} and 84^{th} 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 data set 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°. The colour bar represents the density contrast, indicating relative density variation from its mean value across the sample. **The past 1 Gyr orbit of the centre of mass of the MCS is shown with a magenta line, computed with the gala package using the MilkyWayPotential (Bovy 2015). The blue and red dotted lines represent the Pisces (Chandra et al. 2023b) and Virgo overdensities (Perottoni 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 stream taken from Chandra et al. (2023a).** The dark blue region corresponds to the masked region representing the galactic plane.

It can be observed, two distinct regions of overdensities. The first one, located in the northern hemisphere, with a longitude range between 225° and 315°, 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 ($-30° < l < 130°$) and exhibits significant prominence **at a longitude of** $50°$ **and a latitude of** $-26°$ approximated. 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° to 90°. 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 the one of the collective response. **The ratio between the counts per pixel of the wake at $l = 50°$, $b = -26°$ and the counts per pixel of the collective response at $l = 279°$, $b = 24°$** (coordinates of the highest overdensity of the collective response) is **1.29** considering the complete data set. As one can see, the **CM of 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, 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 Stream (Chandra et al. 2023a) are found in our dataset.

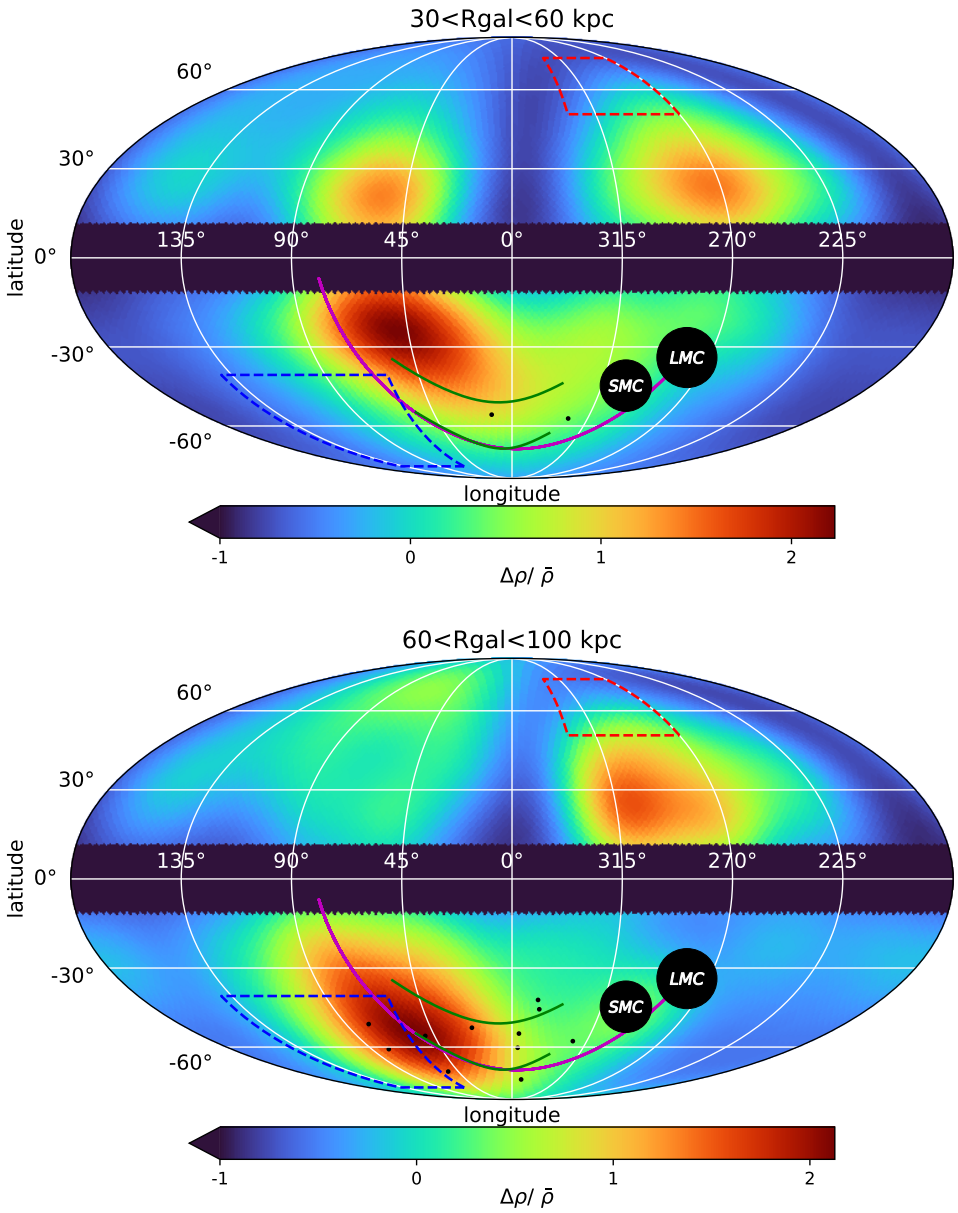

Fig. 1. Density distribution of K-Giant and RR-Lyrae variables (mollweide projection map) with $30 < R_{gal} < 60$ kpc (top panel) and with $60 < R_{gal} < 100$ kpc (bottom panel). The data is smoothed using a Full-Width at Half Maximum (FWHM) of $30°$. The magenta line represents the past orbit of the MCS centre of mass. The blue and red dashed lines represent the Pisces and Virgo overdensities, respectively. The black dots represent the members of the Magellanic Stream. The green lines are the Pisces Plume.

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 data set are ($l = 52°$, $b = -24°$) and ($l = 275°$, $b = 25°$)** respectively, meanwhile for Conroy's data are approximated ($l = 49°$, $b = -54°$) **for the wake** and ($l = 326°$, $b = 54°$) **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 $60 < R_{gal} < 100$ kpc, we successfully replicated the wake's position, with its peak occurring at $l = 57°$, $b = -51°$.** Moreover, when we compared our map with the simulations presented in Conroy et al. (2021), we observed 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 orbit

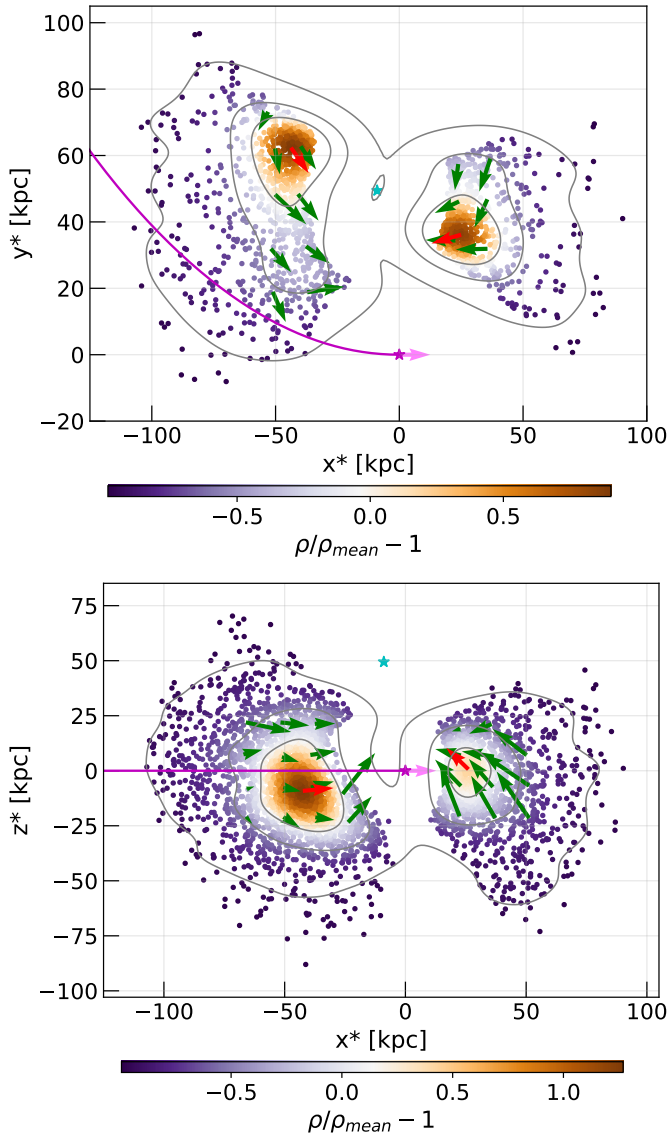

Fig. 2. Overdensity as a function of the position. Magenta line: past orbit of the CM, magenta star: current position of the CM, pink arrow: velocity of the CM. Red arrow: mean velocity of the wake/collective, green arrow: mean velocity of a bin of 15 kpc. Cyan star: position of the galactic centre. Top panel: orbit plane with $|z^*| < 10\,\mathrm{kpc}$; bottom panel: $x^* - z^*$-plane (perpendicular to the orbit plane).

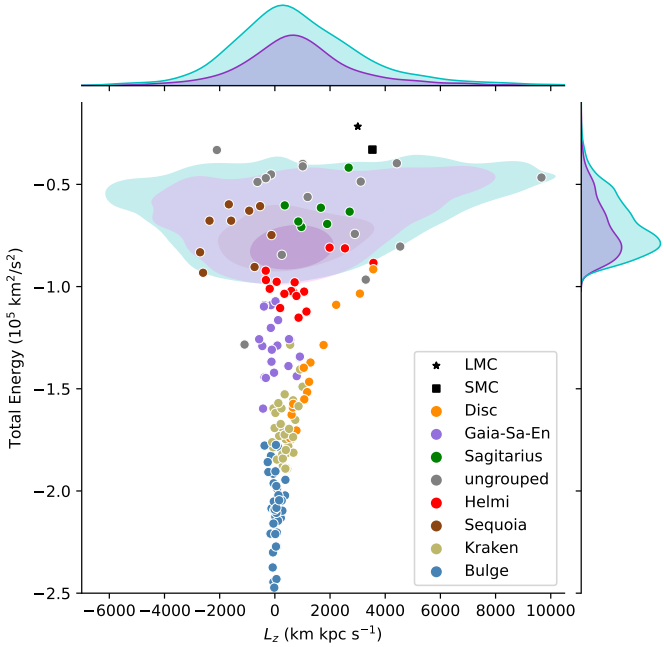

Fig. 3. E-L_z diagram. The mergers data were taken from Callingham et al. (2022). Cyan region: wake, magenta region: collective response.

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 approximated at $x^* = -48.97$ 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 ± 3) km/s. 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 Milky Way'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 performed 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 Milky Way (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 respectively). As one can see, both wake and 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 $E = -0.703 \times 10^5\,\mathrm{km^2/s^2}$, $L_z = 890.433\,\mathrm{km\,kpc\,s^{-1}}$, meanwhile for the collective are $E = -0.716 \times 10^5\,\mathrm{km^2/s^2}$, $L_z = 902.391\,\mathrm{km\,kpc\,s^{-1}}$.

We can interpret the findings presented in Fig. 3 as evidence that the Magellanic clouds are dynamically impacting the current stellar population of the Milky Way's stellar halo. This effect is independent of the stars' origin and impacts populations that were accreted with Sequoia, Saggitarius, Helmi, ED-3-4-5-6, Typhon, L-RL64 (Dodd et al.

2023) and even residing at the disc (Petersen & Peñarrubia 2020). Notice that the mergers are of older origin, consequently, the recent impact of the LMC affects dynamically all these stellar populations. This is according to the idea of a first-infall scenario of the LMC's (Besla et al. 2007; Kallivayalil et al. 2013; Sheng et al. 2024).

Next, we performed the statistical analysis in order to obtain the dark matter subhalo mass of the centre of mass and therefore the DM **subhalo mass** of the LMC. The radius for the region of interest was fixed at $R = 100\,\mathrm{kpc}$, the subhalo velocity v_s was fixed at 314.23 **km/s** (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 have performed a statistical analysis of the data and obtained the velocity standard deviation in each axis. For the analysis using Eqs. (2), we have considered the larger component of $\bar{v}_0$ in the calculations. We have used the described density profile in Section 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 different method indicated with colours**) along with our results (last three values, **with their correspondent statistical errors**).

The dark matter subhalo mass was computed by using the space distribution function (case (a)), the phase-space distribution function of Eq.(3) (case (b)), of 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 3σ with the literature (Vasiliev 2023).

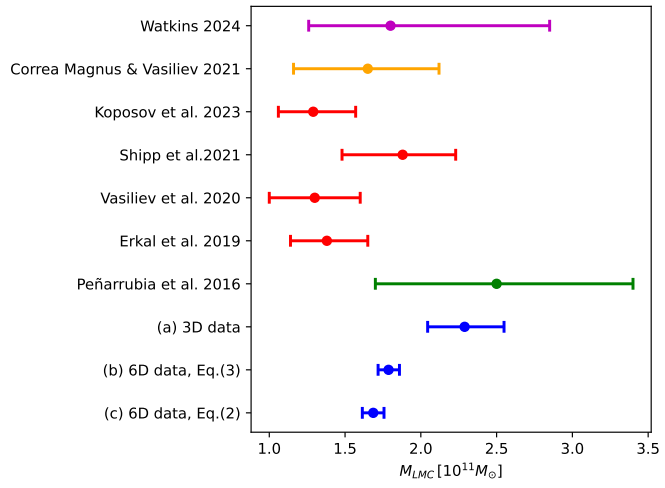

Fig. 4. Dark matter subhalo mass estimation of the LMC. Orange line: kinematic estimation from Milky Way satellites. Red lines: estimation from stellar streams. Green line: estimation based on momentum balance in the Local Group. Blue lines: our results obtained from the likelihood analysis, case (a): $3D$ **data,** $M_{LMC} = 2.289^{+0.260}_{-0.240} \times 10^{11}\mathrm{M}_\odot$**; case (b):** $6D$ **data (Eq.**(3)**),** $M_{LMC} = 1.787^{+0.072}_{-0.069} \times 10^{11}\mathrm{M}_\odot$**; case (c):** $6D$ **data (Eq.**(2)**),** $M_{LMC} = 1.686^{+0.071}_{-0.072} \times 10^{11}\mathrm{M}_\odot$**.**

We have also performed the statistical analysis ($6D$ **Eq.**(2)**) using only the data that have measured radial velocity (Gaia Collaboration et al. 2022; Ahn et al. 2012; Wang et al. 2022). We have found a LMC subhalo mass of** $M_{LMC} = 1.594^{+0.203}_{-0.196} \times 10^{11}\mathrm{M}_\odot$**, that 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-Giants catalogue originally provided by Conroy et al. (2021), and also construct a catalogue for RR-Lyrae stars **both in 6D data**. We 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 centre of mass 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 Milky Way'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 Milky Way's halo, regardless of which past merger event they were accreted.

As per the theoretical proposal made by Buschmann et al. (2018), we were able to estimate the mass of the LMC dark matter subhalo for the first time by using Gaia observational data. We found a reliable estimation of the dark matter 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' mass, our study has successfully determined the mass of the dark matter 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 3σ. 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 multiparameter model such as an ellipsoidal NFW model is used, it is anticipated that the errors will 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 Lerena 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 et al. 2013), Matplotlib (Hunter 2007), Pandas (McKinney 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 TOPCAT (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-Argüello, 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 University Physics Bulletin, 77, 322
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Barberio, E., Baroncelli, T., Bignell, L. J., et al. 2022, arXiv e-prints, arXiv:2205.13849
Bazarov, A., Benito, M., Hütsi, G., et al. 2022, Astronomy and Computing, 41, 100667
Belokurov, V., Deason, A. J., Erkal, D., et al. 2019, Monthly Notices of the Royal Astronomical Society: Letters, 488, L47
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., Bussolotti, A., et al. 2022, in The Fifteenth Marcel Grossmann Meeting on General Relativity. Edited by E. S. Battistelli, ed. E. S. Battistelli, R. T. Jantzen, & R. Ruffini, 1285–1290
Besla, G., Kallivayalil, N., Hernquist, L., et al. 2007, ApJ, 668, 949
Bovy, J. 2015, ApJS, 216, 29
Breiman, L. 2001, Machine Learning, 45, 5
Buehler, R. & Desjacques, V. 2023, Phys. Rev. D, 107, 023516
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, Annual Review of Nuclear and Particle Science, 70, 355
Chandra, V., Naidu, R. P., Conroy, C., et al. 2023a, The Astrophysical Journal, 956, 110
Chandra, V., Naidu, R. P., Conroy, C., et al. 2023b, The Astrophysical Journal, 951, 26
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Clowe, D. et al. 2006, Astrophys. J., 648, L109
Conroy, C., Naidu, R. P., Garavito-Camargo, N., et al. 2021, Nature, 592, 534
Correa Magnus, L. & Vasiliev, E. 2021, Monthly Notices of the Royal Astronomical Society, 511, 2610
Craig, P., Chakrabarti, S., Baum, S., & Lewis, B. T. 2021, arXiv e-prints, arXiv:2107.09791
Crenshaw, J. F., Connolly, A., & Kalmbach, B. 2021, in American Astronomical Society Meeting Abstracts, Vol. 53, American Astronomical Society Meeting Abstracts, 230.01
Crenshaw, J. F. et al. 2024, jfcrenshaw/pzflow: v3.1.3
Cunningham, E. C., Garavito-Camargo, N., Deason, A. J., et al. 2020, The Astrophysical Journal, 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, Research Notes of the American Astronomical Society, 2, 210
Drlica-Wagner, A., Bechtol, K., Mau, S., et al. 2020a, ApJ, 893, 47
Drlica-Wagner, A., Bechtol, K., Mau, S., et al. 2020b, ApJ, 893, 47
Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, Neural Spline Flows
Erkal, D., Belokurov, V., Laporte, C. F. P., et al. 2019, MNRAS, 487, 2685
Foote, H. R., Besla, G., Mocz, P., et al. 2023, arXiv e-prints, arXiv:2307.00053
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Furlanetto, S. R. & Loeb, A. 2002, The Astrophysical Journal, 565, 854
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Vallenari, A., et al. 2022, arXiv e-prints, arXiv:2208.00211
Garavito-Camargo, N., Besla, G., Laporte, C. F. P., et al. 2019, The Astrophysical Journal, 884, 51
GRAVITY Collaboration, Abuter, R., Amorim, A., et al. 2019, A&A, 625, L10
Green, G. 2018, The Journal of Open Source Software, 3, 695
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Harris, W. E. 1996a, AJ, 112, 1487
Harris, W. E. 1996b, 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, Computing in Science and Engineering, 9, 90
Kallivayalil, N., van der Marel, R. P., Besla, G., Anderson, J., & Alcock, C. 2013, ApJ, 764, 161
Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, A&A, 674, A5
Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in IOS Press, 87–90
Koposov, S. E., Erkal, D., Li, T. S., et al. 2023, Monthly Notices of the Royal Astronomical Society, 521, 4936
Kruijssen, J. M. D., Pfeffer, J. L., Chevance, M., et al. 2020, Monthly Notices of the Royal Astronomical Society, 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, Reports on Progress in Physics, 73, 086901
McConnachie, A. W. 2012, AJ, 144, 4
McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51 – 56
Mo, H., van den Bosch, F., & White, S. 2010, Galaxy Formation and Evolution (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, Mon. Not. R. Astron. Soc., 527, 11559
Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3
Peñarrubia, J., Gómez, F. A., Besla, G., Erkal, D., & Ma, Y.-Z. 2016, MNRAS, 456, L54
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Perottoni, H. D., Limberg, G., Amarante, J. A. S., et al. 2022, The Astrophysical Journal Letters, 936, L2
Petersen, M. S. & Peñarrubia, J. 2020, Nature Astronomy, 5, 251
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Price-Whelan, A., Sipőcz, B., Lenz, D., et al. 2020, adrn/gala: v1.3
Price-Whelan, A. M. 2017, The Journal of Open Source Software, 2
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Rubin, V. C. & Ford, W. Kent, J. 1970, Astrophys. J., 159, 379
Sheng, Y., Ting, Y.-S., Xue, X.-X., Chang, J., & Tian, H. 2024, arXiv e-prints, arXiv:2404.08975
Shipp, N., Erkal, D., Drlica-Wagner, A., et al. 2021, The Astrophysical Journal, 923, 149

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
Tulin, S. & Yu, H.-B. 2018, Physics Reports, 730, 1, dark matter self-interactions and small scale structure
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, Frontiers in Astronomy and Space Sciences, 8
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Wang, F., Zhang, H.-W., Xue, X.-X., et al. 2022, Monthly Notices of the Royal Astronomical Society, 513, 1958
Waskom, M. L. 2021, Journal of Open Source Software, 6, 3021
Watkins, L. L., van der Marel, R. P., & Bennet, P. 2024, arXiv e-prints, arXiv:2401.14458
Weinberg, M. D. 1986, Astrophysical Journal, 300, 93
XENON Collaboration, Aprile, E., Abe, K., et al. 2023, arXiv e-prints, arXiv:2303.14729
Zavala, J. & Frenk, C. S. 2019, Galaxies, 7
Zonca, A., Singer, L., Lenz, D., et al. 2019, The Journal of Open Source Software, 4, 1298
Zwicky, F. 1933, Helv. Phys. Acta, 6, 110

Appendix A: LMC rest frame

A.1. CM rest frame

The coordinate system used to compare the data with the theoretical model has its origin in the LMC and SMC centre of mass and the x-axis orientated according to the DM subhalo velocity $\bar{v}_s$. In order to obtain the coordinates of our data set in such rest frame, we have 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

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

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

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

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

where $\left((v_{cm})_x, (v_{cm})_y, (v_{cm})_z\right)$ is the centre of mass 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 Magellanic Clouds System, 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 have to perform a rotation of an angle θ_{orb} according to the following matrix

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

where $\tan\theta_{orb} = \overline{z}_{orb}/\overline{y}_{orb}$, $\overline{z}_{orb}$ ($\overline{y}_{orb}$) is the mean value of the z (y) coordinate in the CM rest frame of the CM 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 for assigning the radial velocities to those stars without such measurements.

B.1. K-Giant radial velocity

Our first sample consists of 245086 K-Giant stars, where 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 magnitude, 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: [$n\_estimators = 4900$, $random\_state = 0$]

B.2. RRLyrae 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 Normalizing Flow (NFs) (Durkan et al. 2019), as 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 have used probabilistic modelling to generate a radial velocity distribution outcome for a significant number of stars (100000). Therefore, the normalizing flow is used to augment and generalise the training data set. We have evaluated the marginal probability of radial velocity given the values of other variables (predictors) and obtained a vector of probabilities. To complement this method, we have 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's 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 is employed by splitting the data into 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 and 0.5 for RR-Lyrae stars 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 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 average velocity for the Sculptor galaxy is due to the lack of any measured radial velocity on this object.

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

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

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 Large Magellanic Cloud.

Appendix C: Validation of estimated distances for K-Giant stars

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

The observational data was extracted from the Gaia Data Release 3 (Gaia Collaboration et al. 2022, 2016), using the option single object searcher to avoid field-contamination. Additionally, the data for the LMC and the SMC was selected on a circular disk in (l, b) centred in the location of each MC, of radius 2°. For Carina, Draco and Sculptor we selected data using an angular mask of 3 times the tidal radius reported in Drlica-Wagner et al. (2020a). We have performed the data reduction/treatment indicated in the text for the K-Giant (that is corrected due to dust extinction, discarded all sources with $E(B-V) > 0.3$, corrected the magnitudes, and performed the 3σ 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 have also restricted the data set in the reported Gaia's parallax.

We calculated the photometric distance for each source and the mean value for each object. The results are shown in Fig. C.1, and, as can be seen, the computed photometric distance and the values reported in the literature (Harris 1996b; Drlica-Wagner et al. 2020b) are in perfect agreement.

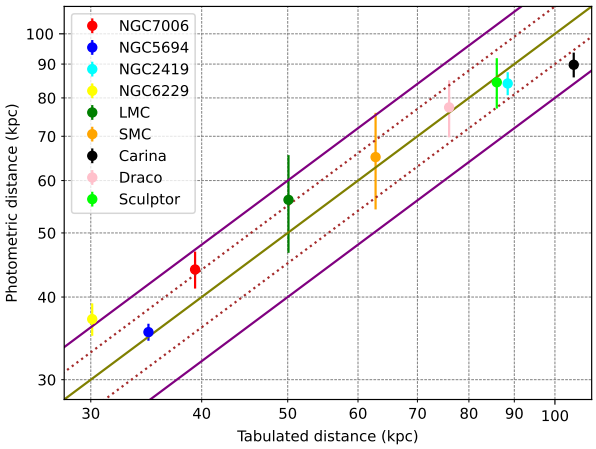

Fig. C.1. Comparison between the photometric distance computed using Eq. (1) and the tabulated distance (Harris 1996b; Drlica-Wagner et al. 2020b). Olive solid line: photometric distance equals tabulated distance; dashed-brown line: ±10% respect to the one-to-one line; purple-solid line: ±20% respect to the one-to-one line. The vertical lines are the statistical errors.

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%.